ADVANCED THEORY OF STATISTICS

1. Distribution Theory
2. Estimation
3. Tests of Hypotheses

Lectures delivered by D. G. Chapman

(Stat 611-612)

1958-1959

North Carolina State College

Notes recorded by R. C. Taeuber

Institute of Statistics

Mimeo Series No. 241

September, 1959

TABLE OF CONTENTS

CHAPTER I - PRELIMINARIES

I: Preliminaries:

Set: A collection of points in R_k (Euclidian k-dimensional space) — S

Def. 1/ $S_1 + S_2$ is the set of points in either or both sets.

$S_1 + S_2$ is the set of points in both S_1 and S_2.

if S_1 is contained in S_2 ($S_1 \subset S_2$) then $S_2 - S_1$ is the set of points in S_2 but not in S_1.

Exercise 1/ Show that $S_1 + S_2 = S_2 + S_1$

$$S_1 S_2 = S_2 S_1$$

There is also the obvious extension of these definitions of set addition and multiplication to 3 or more sets.

$$\sum_{n=1}^{\infty} S_n = \text{ the set of points in at least one of the } S_n$$

$$\prod_{n=1}^{\infty} S_n = \text{ the set of points common to all the } S_n$$

S^* is defined as the complement of S and is the same as $R_k - S$.

Lemma 1/ $(S_1 + S_2)^* = S_1^* \, S_2^*$

Proof: Let ε denote "is an element of"

$x \varepsilon (S_1 + S_2)^*$ means that x is not a member of either S_1 or S_2,

i.e. $x \not\varepsilon S_1$, $x \not\varepsilon S_2$,

therefore $x \varepsilon S_1^*$, $x \varepsilon S_2^*$

since x is common to both S_1^*, S_2^*, $x \varepsilon S_1^* S_2^*$.

To complete the proof

$$x \varepsilon S_1^* S_2^* \implies x \varepsilon S_1^* \text{ and } x \varepsilon S_2^*$$
$$\implies x \not\varepsilon S_1 \text{ and } x \not\varepsilon S_2$$
$$\implies x \not\varepsilon (S_1 + S_2)$$
$$\implies x \varepsilon (S_1 + S_2)^*.$$

Exercise 2/ Show that $S_2 - S_1 = S_2 S_1^*$.

Exercise 3/ In R_2 define $S_1 = \{x,y : x^2 + y^2 \leq 1\}$

i.e. the set of points x,y subject to the restriction $x^2 + y^2 \leq 1$

$$S_2 = \{x,y : |x| \leq .8,\ |y| \leq .8\}$$
$$S_3 = \{x,y : x = 0\}$$

Represent graphically $S_1 + S_2$, $S_1^* S_2$, $S_3 S_2 S_1$, $S_1 S_2 - S_1 S_3$.

Def. 2/ If $S_1 \subset S_2 \subset S_3 \ldots\ldots$ (an exploding family)

We define: $$\lim_{n \to \infty} S_n = \sum_{n=1}^{\infty} S_n$$

And if $S_1 \supset S_2 \supset S_3 \ldots\ldots$ (a nested family)

We define: $$\lim_{n \to \infty} S_n = \prod_{n=1}^{\infty} . S_n$$

Such sequences of sets are called monotone sets.

Exercise 4/

(a) Show that the closed interval in R_2 $\{x,y: |x| \leq 1, |y| \leq 1$ may be represented as an infihite product of a set of open intervals:

$$\text{Ans: } S_n = \{x,y: |x| < 1 + \tfrac{1}{n}, |y| < 1 + \tfrac{1}{n}$$

(b) Show that the open interval in R_2 $\{x,y: |x| < 1, |y| < 1$ can be represented as an infinite sum of closed intervals:

Ans: $S_n = \{x,y: |x| \leq 1 - \frac{1}{n}, |y| \leq 1 - \frac{1}{n}$

Probability is generally thought of in terms of sets, which is why we study sets.

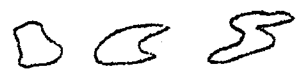
<u>Def. 3/</u> <u>Borel Sets</u> — the family of sets which can be obtained from the family of intervals in R_k by a finite or enumerable sequence of operation of set addition, multiplication, or complementation are called Borel Sets.

The word multiplication could be deleted, since multiplication can be performed by complementation, e.g.:

$$(S_1 + S_2)^* = S_1^* S_2^* \quad (S_1^* + S_2^*)^* = S_1 S_2 \quad (S^*)^* = S$$

<u>Def. 4/</u> A(S) is an additive set function if
1/ for each Borel Set A(S) is a real number, and
2/ if $S_1, S_2, \ldots$ are a sequence of disjoint sets

$$A\left(\sum_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} A(S_n).$$

Examples: — area is a set function 

— in R_1 $A(S) = \int_S f(x)\,dx$ $\qquad A_1(S) = \int_{S_1} x\,dx$

<u>Def. 5/</u> P(S) is a probability measure on R_k if

1/ P is an additive set function
2/ P is non-negative
3/ $P(R_k) = 1$

$\emptyset$ will denote the empty set which contains no points, i.e. $\emptyset = R_k^*$; $\emptyset + R_k = R_k$.

<u>Ex. 5/</u> $P(\emptyset) = 0$

<u>Ex. 6/</u> if $S_1 \subset S_2$ then $P(S_1) \leq P(S_2)$

<u>Lemma 2/</u> $P(S_1 + S_2 + \ldots) \leq P(S_1) + P(S_2) + \ldots$

<u>Problem 1:</u> Prove lemma 2.

<u>Lemma 3</u>/ $P(\lim_{n\to\infty} S_n) = \lim_{n\to\infty} P(S_n)$ if S_n is a monotone sequence.

Proof: case I — $S_1 \subset S_2 \subset S, \ldots\ldots$

Define: $S_1' = S_1$

$S_2' = S_2 - S_1$

$S_3' = S_3 - S_2$ etc.

These sets S_n' are disjoint; also $\sum_{n=1}^{\infty} S_n = \sum_{n=1}^{\infty} S_n'$

$$P(\lim_{n\to\infty} S_n) = P(\sum_{n=1}^{\infty} S_n)$$

$$= P(\sum_{n=1}^{\infty} S_n')$$

$$= \sum_{n=1}^{\infty} P(S_n') \qquad \text{the additive property of P}$$

$$= P(S_1') + P(S_2') + P(S_3') + \ldots$$

$$= P(S_1) + P(S_2 - S_1) + P(S_3 - S_2) \ldots$$

$$= P(S_1)$$
$$- P(S_1) + P(S_2)$$
$$- P(S_2) + P(S_3)$$
etc.

$= P(S_n)$ after n steps

$= \lim_{n\to\infty} P(S_n)$

Case 2: $S_1 \supset S_2 \supset S_3 \ldots\ldots$

<u>Problem 2</u>/ Prove lemma 3 for case 2.

<u>Def. 6</u>/ Associated with any probability measure $P(S)$ on R_1 there is a point function $F(x)$ defined by

$$F(x) = P(-\infty, x).$$

$F(x)$ is called a distribution function — d.f.

<u>Theorem 1</u>/ Any distribution function, $F(x)$, has the following properties:

1. It is a monotone, non-decreasing sequence.
2. $F(-\infty) = 0$; $F(+\infty) = +1$
3. $F(x)$ is continuous on the right.

Proof: 1/ For $x_1 < x_2$ we have to show that $F(x_1) \leq F(x_2)$

$F(x_1) = P(-\infty, x_1)$ $\quad$ $F(x_2) = P(-\infty, x_2)$

The interval $(-\infty, x_1) \subset$ the interval $(-\infty, x_2)$

From exercise 6 we have that $P(I_1) \leq P(I_2)$

Therefore $F(x_1) \leq F(x_2)$

2/a/ If we define G_n as the interval $(-\infty, -n)$ $\quad n=1,2,3,\ldots\ldots$

Then $G_1 \supset G_2 \supset G_3 \ldots\ldots$

$G = \lim_{n\to\infty} G_n = \emptyset$ $\quad$ (the empty set)

$\lim_{n\to -\infty} F(n) = \lim_{n\to\infty} P(G_n) = P(\lim_{n\to\infty} G_n)$ $\quad$ From lemma 3

$= P(G) = P(\emptyset) = 0$

b/ Follows in a similar fashion by defining $G_n = (-\infty, n)$

3/ Pick a point a — for this point we want to show

$$\lim_{\substack{x\to a\\ x>a}} F(x) = F(a),$$

Consider a nested sequence $\varepsilon_n \to 0$, $\varepsilon_n > 0$.

Then $\lim_{n\to\infty} F(a + \varepsilon_n) = F(a)$ is the property to be shown.

If we define $H_n = (-\infty, a + \varepsilon_n)$ $\quad n = 1,2,3,\ldots\ldots$

$\lim_{n\to\infty} H_n = H =$ interval $(-\infty, a)$

$\lim P(H_n) = P(\lim H_n) = P(H)$ $\quad$ lemma 3

Therefore $\lim_{n\to\infty} F(a + \varepsilon_n) = F(a)$

Problem 3/ Show that

$$F(a) = \lim_{\substack{x \to a \\ x < a}} F(x) + P\{[a]\}$$ Where $[a]$ is the set whose only point is a.

Or in familiar terms $F(a) = F(x - 0) + Pr(x = a)$

Where $F(x - 0)$ is the limit from the left.

Theorem 2/ To any point function $F(x)$ satisfying properties 1, 2, and 3 of theorem 1, there corresponds a probability measure $P(S)$ defined for all Borel Sets such that for any interval $(-\infty, x)$

$$P(-\infty, x) = F(x).$$

Proof omitted — see Cramer p. 53 referring to p. 22.

Theorem 3/ A distribution function $F(x)$ has at most a countable number of discontinuities.

Proof: Let v_n be the number of points of discontinuity with a jump $> \frac{1}{n}$ then $v_n \leq n$ which is what we have to show.

Suppose the contrary holds, i.e. $v_n > n$.

Then if we let Sn be the set of such discontinuities, we have

$$1 = P(R_1) \geq P(S_n) > \frac{1}{n}(n) \geq 1$$ which is a contradiction.

Therefore, the total number of discontinuities $= \sum_{n=1}^{\infty} v_n < \sum_{1}^{\infty} n$

where $\sum_{1}^{\infty} n$ is the sum of the integers which is a countable sum.

Notation:

[ ] square brackets — means the end point is not included in the interval — i.e., an open interval.

( ) round brackets — mean the end points are included in the interval, i.e., a closed interval.

(a, b] is the interval a to b, including a but not b.

Def. 7: In R_k to each probability measure $P(S)$ there corresponds a unique distributic function

$$F(x) = F(x_1, x_2, \ldots, x_k)$$

$$= P\left[\text{interval } (-\infty, -\infty, \ldots; x_1, x_2, \ldots, x_k)\right]$$

The interval is the set of points in R_k

$$-\infty < X_i \leq x_i \qquad i = 1, 2, \ldots, k$$

<u>Theorem 4</u>: $F(x_1, x_2, \ldots, x_k)$ has the following properties:

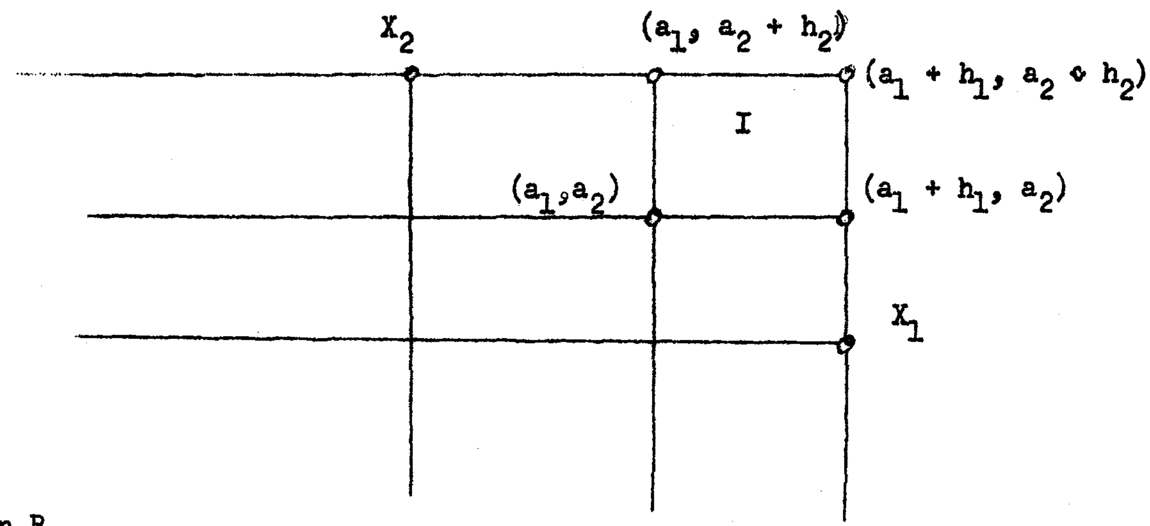
1. It is continuous from the right in each variable
2. $F(-\infty, x_2, \ldots, x_k) = F(x_1, -\infty, x_3, \ldots, x_k) = \ldots = 0$
   $F(+\infty, +\infty, \ldots, +\infty) = 1$
3. $\Delta_k F(a_1, a_2, \ldots, a_k) \geqslant 0$ see p. 79 Cramer

   i.e., the P measure of any interval is non-negative.

Conversely if $F(x_1, x_2, \ldots, x_k)$ has these properties, then there is a unique P-measure defined by

$$P(I) = F(x_1, x_2, \ldots, x_k).$$

That is, I is the interval $[-\infty, -\infty, \ldots; x_1, x_2, \ldots, x_k)$.



In R_2

$$P(I) = F(a_1 + h_1, a_2 + h_2) - F(a_1, a_2 + h_2) - F(a_1 + h_1, a_2) + F(a_1, a_2)$$

$$= \Delta_2 F(a_1, a_2)$$

in R_3

$$P(I) = F(a_1 + h_1, a_2 + h_2, a_3 + h_3)$$
$$-F(a_1, a_2 + h_2, a_3 + h_3) - F(a_1 + h_1, a_2, a_3 + h_3)$$
$$-F(a_1 + h_1, a_2 + h_2, a_3)$$
$$+F(a_1, a_2, a_3 + h_3) + F(a_1, a_2 + h_2, a_3) + F(a_1 + h_1, a_2, a_3)$$
$$-F(a_1, a_2, a_3)$$
$$= \Delta_3 F(a_1, a_2, a_3)$$

The proof of theorem 4 is by analogy with the linear case (theorem 1),

$$F(x_1, +\infty, \ldots, +\infty) = P(\,[-\infty, x_1)\,)$$
$$= F_1(x_1)$$
$= \text{the marginal distribution of } x_1$

(similarly for other dimensions)

<u>Def. 8:</u>

If F is continuous and differentiable in all variables, then

$$\frac{\partial^k F}{\partial x_1, \partial x_2, \ldots, \partial x_k} = f(x_1, x_2, \ldots x_k)$$

is the density function of $x_1, x_2, \ldots, x_k$.

<u>Exercise 7:</u>

In R_2 $\quad F(x, y) = 0 \quad \text{if} \quad x \leq 0 \quad \text{or} \quad y \leq 0$

$$= \frac{(x + y)}{2} \quad \text{for} \quad 0 < x \leq 1, \; 0 < y \leq 1$$

$$= 1 \text{ for } x > 1, y > 1$$

Can this be a distribution function in R_2?

How can the definitions be completed?

Solution -- consider the marginal distribution of x

$$F_1\ (x) = F(x, +\infty) \geqslant F(x, 1) = \frac{x + 1}{2}$$

$$F_1\ (0) = \frac{(0 + 1)}{2} = \frac{1}{2}$$

But in fact $F(0, 0) = 0$

$F(0, y) = 0$ for all y

$F_1\ (0) = 0$

Therefore there is a contradiction. F cannot be a proper distribution function.

If $F(x, y)$ is a proper distribution function, then the two marginal distributions

$$F_1(x) = F(x, +\infty)$$

$F_2(y) = F(+\infty, y)$ must be proper and in this case they break down.

Problem 4: If we define $f(x, y) = x + y \qquad 0 \leq x \leq 1$

$0 \leq y \leq 1$

$= 0$ elsewhere

Find $F(x, y)$; $F_1(x)$; and $F_2(y)$

Show that $F(x, y)$ satisfies the properties of a distribution function.

Def. 9: Random Variable

We assume we have experiments which yield a vector valued set of observations

$\underline{X} = (X_1, X_2, \ldots, X_k)$ with the properties:

1. For each Borel set S in R_k there is a probability measure P(S) which is the probability that the whole vector $\underline{X}$ falls in the set in S
   (P(S) is non-negative, additive, and $P(R_k) = 1$). Cramer p. 152-4 axioms 1 and 2

2. If $\underline{X}_1, \ldots, \underline{X}_n$ are random variables in $R_{k_1}, R_{k_2}, \ldots, R_{k_n}$
   then the combined vector $(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n)$ is also a random variable
   in $R_{k_1 + k_2 + k_3 \ldots + k_n}$.

Conditional Distribution

$(\underline{X}, \underline{Y})$ are random variables in R_{k_1}, R_{k_2}.

Let S and T be sets in R_{k_1}, R_{k_2}.

Def. 10: If P(X belongs to S) > 0, then we define conditional probability

$$P(Y \subset T \mid X \subset S) = \frac{P(Y \subset T, X \subset S)}{P(X \subset S)}.$$

We show that $P(Y \subset T \mid X \subset S)$ does satisfy the requirements of a probability measure

1- It is non-negative since $P(Y \subset T, X \subset S)$ is non-negative.
2- It is additive since $P(Y \subset T, X \subset S)$ is additive in R_{k_2}.

$$P(Y \subset T_1 / X \subset S) + P(Y \subset T_2 / X \subset S) = P(Y \subset T_1 + T_2 \mid X \subset S)$$

3- $$\frac{P(Y \subset R_k, X \subset S)}{P(X \subset S)} = \frac{P(X \subset S)}{P(X \subset S)} = 1$$

If $P(Y \subset T) > 0$ we could also define

$$P(X \subset S \mid Y \subset T) = \frac{P(X \subset S, Y \subset T)}{P(Y \subset T)}$$

In familiar terminology what we are saying is that

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)} \quad \text{or} \quad \Pr(A, B) = \Pr(A \mid B)\Pr(B).$$

If we have the corresponding distribution functions

$$F(x, y);\ F_1(x) = F(x, +\infty);\ \text{and } F_2(y) = F(+\infty, y) \text{ then;}$$

Def. 11: X and Y are independent random variables if the joint distribution function $F(x, y)$ factors into $F_1(x)\, F_2(y)$.

See p. 160 Cramer -- he goes first to probability measures, then to d.f.

notation:

-- Capital latin letters used for random variables in general.
-- Small latin letters used for observations or specific values of the random variables.

$$\Pr(X \leq x) = F(x)$$

Def. 11 -- extension:

In the case of n random variables, $X_1, X_2, \ldots, X_n$ these are independent if

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)\, F_2(x_2) \ldots F_n(x_n).$$

Note: Three variables may be pairwise independent, but may not be (mutually) independent -- see the example on p. 162 of Cramer.

If density functions exist, then $X_1, X_2, \ldots, X_n$ independent means that

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)\, f_2(x_2) \ldots f_n(x_n).$$

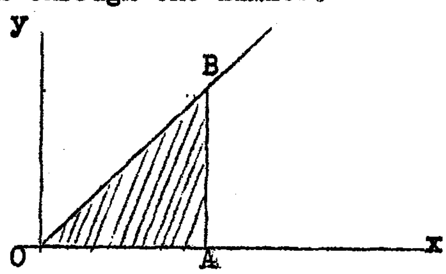
Note: The fact that the density functions factor does not necessarily mean independence since dependence may be brought in through the limits.

e.g. X and Y are distributed uniformly on OAB

$f(x, y) = 2$

$0 \leq x \leq 1$

$0 \leq y \leq x$



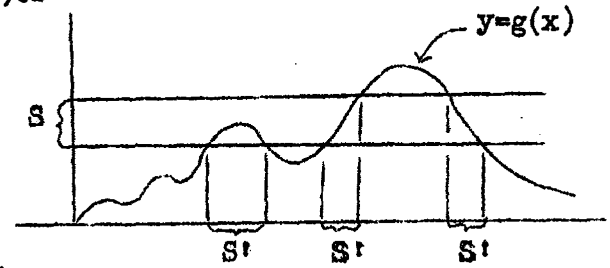
Exercise 8: Find $F(x, y)$; $F_1(x)$; and $F_2(y)$.

Functions of Random Variables:

g is a Borel measurable function if for each Borel set S, there is a set S' such that

$$x \in S' \Longrightarrow g(x) \in S$$

is also a Borel set. e.g.



A notation sometimes used is that $S' = g^{-1}(S)$ -- that S' is the inverse image of S under the mapping g.

Now consider $Y = g(X)$ where X is a random variable.

$$Pr(y \subset S) = Pr(x \subset S') = P(S')$$

Therefore, any Borel measureable function, $Y = g(X)$ of a random variable, X, is itself a random variable.

$$Pr(y \subset S) = P\left[g^{-1}(S)\right]$$

This extends readily to k dimensions.

Transformations:

We have X and Y which are random variables with distribution function $F(x, y)$ and a density function $f(x, y)$.

Let $\alpha = \phi_1(X, Y) \qquad \beta = \phi_2(X, Y)$

Where ϕ_1 and ϕ_2 are 1 to 1 with α, β; are continuous, differentiable; and the Jacobian of the transformation is non-vanishing.

$$J = \begin{vmatrix} \frac{\partial X}{\partial \alpha} & \frac{\partial X}{\partial \beta} \\ \frac{\partial Y}{\partial \alpha} & \frac{\partial Y}{\partial \beta} \end{vmatrix}$$

We then have the inverse functions $X = \Psi_1(\alpha, \beta)$

$$Y = \Psi_2(\alpha, \beta)$$

The density function $f(a, b)$ of the random variables α, β is

$$f\left[\Psi_1(a, b); \Psi_2(a, b)\right] |J|$$

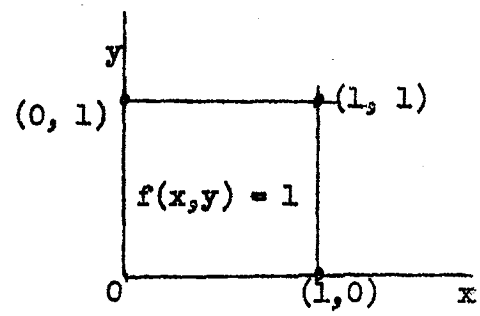
However under the transformation the limits of the variables will be changed and these have to be worked out in each individual case. (See Anderson and Bancroft.)

Problem 5: X, Y are uniformly and independently distributed on (0, 1).
Find the distribution of $Z = XY$ and $-2 \ln XY$.

Example: For X, Y as in problem 5, find the distribution function of $Z = X + Y$.

Solution: Consider also $W = X - Y$

Consider the joint distribution of (Z, W)



$$Z = X + Y$$
$$W = X - Y$$
$$\therefore \frac{Z + W}{2} = X \quad \frac{Z - W}{2} = Y$$

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

Density of Z, W = $f(x, y)\ |J| = 1 \cdot \frac{1}{2}$

The limits of Z and W are dependent

$$Z = X + Y$$
$$W = X - Y$$

If $Z = z$, then W takes on values from $(0 - z)$ thru 0 to $z - 0$, so that

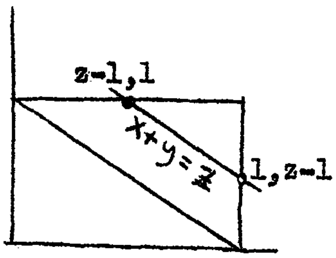
for $Z = z \leq 1 \qquad -z \leq W \leq z$

$$\therefore \quad f(z,W) = \frac{1}{2} \qquad \text{with limits } \begin{matrix} -z \leq W \leq +z \\ z \leq 1 \end{matrix}$$

Since we started with only Z, and "artificially" added W to get a solution, we must now get the marginal distribution of Z (this being what we desire).

$$F_1(z) = \int_{-z}^{+z} |J| \ dw = \int_{-z}^{+z} \frac{1}{2}\, dw = \frac{1}{2}w \Big]_{-z}^{+z} = z$$

$$F(z) = \int_0^z F(t)dt = \frac{z^2}{2} \qquad 0 < z \leq 1$$



If $X = z > 1$, then ε takes on values from $(z-1)-1$ to $1-(z-1)$ or from $(z-2)$ to $(2-z)$

$$f_1(z) = \int_{z-2}^{2-z} \frac{1}{2}\, d\varepsilon = 2 - z$$

$$F(z) = \frac{1}{2} + \int_1^z (2-z)\, dz = \frac{1}{2} - \frac{(2-z)^2}{2}\Bigg]_1^z$$

$$= \frac{1}{2} - \frac{(2-z)^2}{2} + \frac{1}{2} = 1 - \frac{(2-z)^2}{2}$$

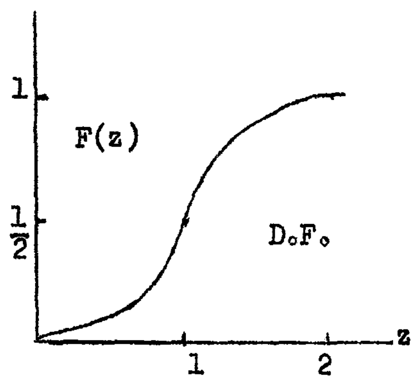
$$\therefore\ F(z) = \frac{z^2}{2} \qquad 0 \leq z \leq 1$$

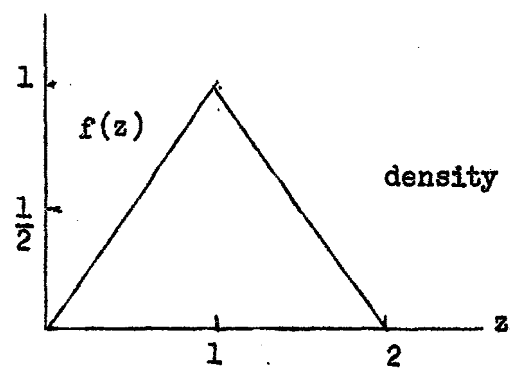
$$= 1 - \frac{(2-z)^2}{2} \qquad 1 \leq z \leq 2$$

$$f(z) = z \qquad 0 \leq z \leq 1$$

$$= 2-z \qquad 1 \leq z \leq 2$$

See p. 245, 6 in Cramer





Joint density of Z, W $f(z, w)$ $= \frac{1}{2}$ $\qquad 0 \leq z \leq 1$, $\quad -z \leq w \leq z$

$= \frac{1}{2}$ $\qquad 1 \leq z \leq 2$, $\quad z-2 \leq w \leq 2-z$

If the transformation is not 1 to 1 (that is $J = 0$) then the usual devi e to avoid the difficulties that may arise is to divide R_k into regions in each of which the transformation is 1 to 1, and then work separately in each region.

i.e., consider in R_1 $Y = X^2$ We should consider 2 separate cases:

$X < 0$

$X \geqslant 0$

Cramer p. 167

Riemann-Stieltjes Integral:

Let F(x) be a d.f., with at most a finite number of discontinuities in (a,b) and let g(x) be a continuous function, then we can define

$$\int_a^b g(x)\, dF(x) \quad \text{as follows:}$$

Cramer p. 71-74

Divide (a, b) into n sub-intervals $x_1, x_2, \ldots, x_n$ of length $\leq \Delta$

$$\text{Let } \underline{S}_n = \sum_{i=1}^{n} \left[\inf_{x_{i-1} \leq x \leq x_i} g(x)\right] \left[F(x_i) - F(x_{i-1})\right]$$

$$\overline{S}_n = \sum_{i=1}^{n} \left[\sup_{x_{i-1} \leq x \leq x_i} g(x)\right] \left[F(x_i) - F(x_{i-1})\right]$$

$\underline{S}_n < \overline{S}_n$ but as $n \to \infty$, $\Delta \to 0$ $\underline{S}_n$ is increasing, $\overline{S}_n$ is decreasing

They can be shown to have a common limit.

So the common limit is called the R-S integral,

$$\text{Also define} \quad \int_{-\infty}^{+\infty} g(x)\, dF(x) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b g(x)\, dF(x)$$

provided the limit exists

and in general $\int_a^b g(x)\, dF(x)$ has all the usual properties of the familiar Riemann integral.

If F(x) has density f(x) which is continuous except at a finite number of points, then

$$F(x_i) - F(x_{i-1}) = f(x')\,(x_i - x_{i-1}) \qquad x_{i-1} < x' < x_i$$

$$= f(x')\,\Delta_i(x)$$

$$\underline{S}_n = \sum \left[\inf g(x)\right] \left[f(x')\right] \Delta x$$

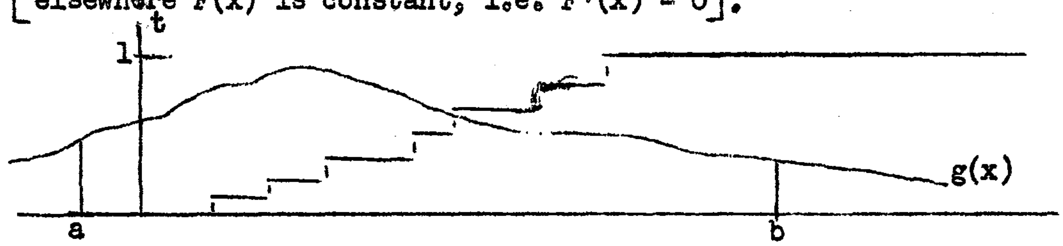
$$\lim \underline{S}_n = \int_a^b g(x)\,dF(x) = \int_a^b g(x)\,\underline{f}(x)\,dx = \text{ ordinary Riemann integral.}$$

If F(x) has only jumps at $x_1, x_2, \ldots, x_n$ and elsewhere is constant

$$\int_a^b g(x)\,dF(x) = \sum g(x_i)\,\underbrace{Pr\left[x = x_i\right]}_{\text{the jump}}$$

If g(x) is continuous then this limit (the R-S Integral) exists. Also, if g(x) has at most a finite number of discontinuities and so does F(x) and they don't coincide, then the R-S integral exists.

__Def. 12:__ X is a discrete random variable if there exists a countable set of points, $x_1, x_2, \ldots, x_n$ with $Pr(X = x_i) = p_i$ and $\sum(p_i) = 1$ [elsewhere F(x) is constant, i.e. $F'(x) = 0$].



for such a discrete random variable, the R-S integral reduces to a sum:

$$\int_a^b g(x)\,dF(x) = \lim_{n\to\infty} \sum g(x_i'')\left[F(x_i') - F(x_{i-1}')\right]$$

where the x_i' are points of division of (a, b) and x_i'' is an intermediate point in the i^{th} interval

$$= \lim_{n \to \infty} \sum g(x_i)\, p_i$$

$$= \sum g(x_i)\, p_i$$ summed over the set of points x_i in (a, b) -- the points where there is some probability.

<u>Def. 13:</u> X is a continuous random variable if F(x) is continuous and has a derivative f(x) continuous except at a countable number of points.

$$F(x_i) - F(x_{i-1}) = f(x_i'')\left[x_i - x_{i-1}\right] \quad \text{(the theorem of the mean)}$$

$$x_{i-1} \leq x_i'' \leq x_i$$

$$\int_a^b g(x)\, dF(x) = \lim_{n \to \infty} \sum g(x_i') \left[F(x_i) - F(x_{i-1})\right]$$

$$= \lim_{n \to \infty} \sum g(x_i')\, f(x_i'')\, \Delta_i$$

$$= \int_a^b g(x)\, f(x)\, dx$$

We can extend this definition to k-dimensions readily by writing:

$$\int_a^b g(x_1, \ldots, x_k)\, d_{x_1, \ldots, x_k} F(x_1, \ldots, x_k)$$

$$= \lim_{n \to \infty} \sum g(x_1', \ldots, x_k')\, \Delta_k F(x_1, \ldots, x_k)$$

For a def. of Δ_k see p. 8

If $F(x_1, x_2, \ldots, x_k)$ is continuous and the density $f(x_1, x_2, \ldots, x_k)$ exists and is continuous, then

$$\int_a^b g(x_1, \ldots, x_k)\, d_{x_1 \ldots x_k} F(x_1, \ldots, x_k) =$$

$$\int_{a_1}^{b_1} \ldots \int_{a_k}^{b_k} g(x_1, \ldots, x_k)\, f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k$$

In R_1 $\quad \int_{-\infty}^{x} d\,F(x) = F(x) - F(-\infty) = F(x)$

$$\int_a^b dF(x) = F(b) - F(a)$$

If we let $b \to +\infty$, $a \to -\infty$

$$\int_{-\infty}^{+\infty} dF(x) = 1$$

and this extends easily to the k-dimensional case, so that we have:

$$\int_{-\infty}^{+\infty} d_{x_1 \ldots x_k} F(x_1, \ldots, x_k) = 1$$

Consider $k = 2$, and the marginal distributions

$$F_1(x_1) = F(x_1, +\infty) = \int_{-\infty}^{+\infty} dF_{x_2}(x_1, x_2)$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b dF_{x_2}(x_1, x_2)$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \left[F_1(x_1, b) - F_1(x_1, a) \right]$$

$$= F_1(x_1, +\infty) - F_1(x_1, -\infty)$$

$$= F_1(x, +\infty) - 0$$

This also extends readily to R_k

$$F_1(x_1) = F(x_1, +\infty_2, \ldots, +\infty_k)$$

$$= \int_{-\infty}^{\infty} d_{x_2, x_3, \ldots, x_k} F(x_1, x_2, \ldots, x_k)$$

with x_1 held fixed.

If the density function exists, then this reduces to a k-1 integral

$$f_1(x_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots, x_k)\, dx_2, \ldots, dx_k$$

Problem 6:

if $X_1, \ldots, X_n$ have independent, uniform on (0, 1), distributions,

show that $-2 \sum_{1}^{n} \ln X_i$ has a χ^2 distribution with 2n d.f.

and indicate the statistical application of this.

see: Snedecor ch. 9
Fisher -- about p. 100
Anderson + Bancroft -- last chapter of section 1

From problem 5b $Z = -2 \ln X Y = -2 \ln X + -2 \ln Y$

or is the sum of 2 χ^2 with 2 d.f. each

References on integrals:

-- Cramer pp. 39-40

-- Sokolnikoff -- Advanced Calculus -- ch. 4

Chapter II

Properties of Univariate Distributions; Characteristic Functions

Standard Distributions:

A. Trivial or point mass (discrete)

$$\Pr\left[X = a\right] = 1 \qquad F(x) = 0 \quad x < a; \qquad F(x) = 1 \quad x \geq a$$

B. Uniform (continuous)

$$F(x) = 0 \quad x < 0$$
$$F(x) = x \quad 0 \leq x \leq 1$$
$$F(x) = 1 \quad x > 1$$

C. Binomial (discrete)

$$\Pr\left[X = k\right] = \binom{n}{k} p^k (1 - p)^{n-k} \qquad k = 1, 2, \ldots, n \qquad 0 \leq p \leq 1$$

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = \left[(1-p) + p\right]^n = 1^n = 1$$

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \equiv 1 \text{ is an identity in } p, n$$

D. Poisson (discrete)

$$\Pr\left[X = k\right] = e^{-\lambda} \frac{\lambda^k}{k!} \qquad k = 1, 2, \ldots, \infty \qquad \lambda > 0$$

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

E. Negative binomial (discrete)

$$\Pr\left[X = k\right] = \binom{r+k-1}{r-1} p^r (1 - p)^k \qquad k = 1, 2, \ldots, \infty$$

$0 < p < 1$

r is an integer

Example: Draw from an urn with proportion p of red balls, with replacement, until we get r reds out. The random variable in this situation is the number of non-reds drawn in the process; to have k black balls means that in the first r + k - 1 trials, we got r - 1 reds and k blacks, and then on the last trial got a red; the probability of this is

$$\binom{r+k-1}{r-1} p^{r-1} (1-p)^k \quad x \, p.$$

This is what is referred to as inverse sampling in that the number of defectives is specified rather than specifying the sample size which is then scrutinized for the number of defectives.

F. Normal distribution (continuous)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \begin{array}{l} \sigma > 0 \\ -\infty < \mu < \infty \\ -\infty < x < \infty \end{array}$$

Problem 7: Prove that

$$\sum_{k=0}^{\infty} \binom{r+k-1}{r-1} p^r (1-p)^k = 1$$

i.e., is an identity in r, p

Hint: is in the name - express $(a+b)^{-n}$ in an infinite series.

Def. 14: If X is a random variable with distribution F(x) and if $\int_{-\infty}^{\infty} g(x)\, dF(x)$ exists, then we define the expectation of g(X) as

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, dF(x)$$

this being the R - S integral,

if X is continuous $\quad E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$

if X is discrete $\quad E\left[g(X)\right] = \sum_{v=0}^{-\infty} g(x_v)\, p_v$

Problem 8: Given

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\, dt & x > 0 \end{cases}$$

(This is a censored normal distribution -- i.e., all the negative values are concentrated at the origin)

Find: E(X)

Def. 15: If $E\left[X - E(X)\right]^k$ exists it is defined to be the k^{th} central moment and is denoted μ_k.

If $E(X)^k$ exists it is defined as the k^{th} moment about the origin, and is denoted α_k.

Exercise 9: Find E(X) for each of the standard distributions.

Theorem 5: $E\left[g(X) + h(Y)\right] = E\left[g(X)\right] + E\left[h(Y)\right]$

Proof: Let F(x, y) be the joint d.f. of X, Y and F_1 and F_2 the marginal d.f.

$$E\left[g(X) + h(Y)\right] = \iint_{-\infty}^{\infty} \left[g(x) + h(y)\right] d_{xy}F(x,y) = \iint g(x) d_{xy}F(x,y) + \iint h(y) d_{xy}F(x,y)$$
$$= \int g(x) d_x F_1(x) + \int h(y) d_y F_2(y) = E\left[g(X)\right] + E\left[h(Y)\right]$$

Theorem 6: If X, Y are independent random variables, then

$$E\left[g(X)\, h(Y)\right] = E\left[g(X)\right] E\left[h(Y)\right]$$

Proof: See Cramer, p. 173.

Corollary: If X and Y are independent random variables, then

$$Var(X + Y) = Var(X) + Var(Y)$$

Moments:

$$\alpha_1 = E(X) = \text{mean}$$
$$\mu_2 = E(X - \mu)^2 = \text{variance}$$
$$\mu_3 = E(X - \mu)^3$$
$$\mu_4 = E(X - \mu)^4$$

etc.

for the normal distribution -- N(0, 1)

$$f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$E(X^k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx$$

all odd moments (k odd) = 0 by a "symmetry" argument.

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = 1 \quad \text{from integration by parts}$$

$$E(X^4) = 3$$

or in general

$$E(X^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X^{2n} e^{-x^2/2} dx = 1 \cdot 3 \cdot 5 \cdot \ldots x\ (2n - 1)$$

which can be shown by induction.

<u>Theorem 7:</u> Let $\alpha_0 = 1, \alpha_1, \alpha_2, \ldots$ be the moments of a distribution function $F(x)$, i.e.,

$$\alpha_k = \int_{-\infty}^{+\infty} X^k \, d\, F(x)$$

then if for some $r > 0$, $\sum_{k=0}^{\infty} \frac{\alpha_k r^k}{k!}$ converges absolutely then $F(x)$ is the only distribution with these moments.

Proof: See Cramer, p. 176.

Example: $N(0, 1)$

$$\alpha_{2k+1} = 0$$

$$\alpha_{2k} = \frac{(2k-1)!}{2^{k-1}(k-1)!} = 1 \cdot 3 \cdot 5 \cdot \ldots (2k - 1)$$

$$\sum_{k=0}^{\infty} \frac{\alpha_k r^k}{k!} = \sum_{k=0}^{\infty} \frac{(2k-1)!}{2^k\,(k-1)!} \frac{r^{2k}}{(2k)!} \quad \text{since odd terms drop out.}$$

$$= \sum_{k=0}^{\infty} \frac{(r^2)^k}{2^{k-1}\,(k-1)!\,.2k} = \sum \frac{(r^2)^k}{2^k\,k!}$$

$$= \sum \frac{1}{k!}\left(\frac{r^2}{2}\right)^k \qquad \text{now} \sum_0^{\infty} \frac{x^k}{k!} = e^x$$

= exponential series

$$= e^{-r^2/2}$$

$\therefore \sum_{k=0}^{\infty} \frac{\alpha_k r^k}{k!}$ converges absolutely for all r, therefore the only distribution with these moments is the d.f. with the density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{i.e., the normal}$$

Problem 9: Find the moments of the uniform distribution and show that this is the only d.f. with these moments.

Theorem 8: (Tchebycheff's)

If g(X) is a non-negative function of X then for every $K > 0$

$$Pr\left[g(X) \geq K\right] \leq \frac{E\left[g(X)\right]}{K}$$

Proof: $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, dF(x)$

Let S be the set of values of X where $g(X) \geq K$

$$Pr\left[g(X) \geq K\right] \geq \int_S g(x)\, dF(x) \qquad \text{since the smallest value of } g(X) \text{ in } S \text{ is } K$$

$$\geq K \int_S dF(x) = K\, Pr\left[g(X) \geq K\right]$$

$$\therefore Pr\left[g(X) \geq K\right] \leq \frac{E\left[g(X)\right]}{K}$$

Corollary: The above (th. 8) converts readily into the more familiar form

$$Pr\left[|X - \mu| \geq k\sigma\right] \leq \frac{1}{k^2}$$

Proof: (See p. 182 in Cramer) setting

$$g(X) = (X - \mu)^2 \qquad K = k^2\sigma^2 \qquad E\left[g(X)\right] = \sigma^2$$

$$Pr\left[(X - \mu)^2 \geq (k\,\sigma)^2\right] \leq \frac{\sigma^2}{k^2\sigma^2}$$

taking the square root of the left hand side

$$Pr\left[|X - \mu| \geqslant k\sigma\right] \leqslant \frac{1}{k^2}$$

<u>Theorem 9</u>: If X_n is a sequence of binomial random variables with parameters n, p, then given any $\varepsilon > 0$, $\delta > 0$ there exists an n_o such that for $n > n_o$

$$Pr\left[|X_n/n - p| \geqslant \varepsilon\right] \leqslant \delta$$

(which says that if you take larger and larger samples, then the observed ratio X/n approaches the true value)

X_n = number of successes in n independent trials with a probability of success in each trial = p

Proof: $\sigma^2_{X_n} = np(1-p)$ $\qquad \sigma^2(X_n/n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \leqslant \frac{1}{4n}$

From corollary to theorem 8

$$Pr\left[\left|\frac{X_n}{n} - p\right| \geqslant k\sigma\right] \leqslant \frac{1}{k^2}$$

Choose: $\frac{1}{k^2} = \delta$ or $k = \frac{1}{\sqrt{\delta}}$

$$k\sqrt{\frac{p(1-p)}{n}} = \varepsilon \quad \text{or} \quad n = \frac{p(1-p)}{\delta\,\varepsilon^2}$$

Hence if n is chosen this large, from the corollary to theorem 8 the stated probability inequality follows.

Note: Theorem 9 could be rewritten

$$Pr\left[\left|\frac{X_n}{n} - p\right| \geqslant \varepsilon\right] \leqslant \frac{1}{4n\,\varepsilon^2} \qquad n = \frac{p(1-p)}{\delta\,\varepsilon^2}$$

$$\delta = \frac{p(1-p)}{n\,\varepsilon^2} \leqslant \frac{1}{4n\,\varepsilon^2}$$

<u>Characteristic Functions</u>:

<u>Def. 16</u>: Characteristic function: $\phi_X(t) = \int_{-\infty}^{\infty} e^{ixt}\, dF(x)$

or since $e^{ixt} = \cos xt + i \sin xt$

$$\phi_X(t) = \int_{-\infty}^{\infty} \cos xt\, dF(x) + i \int_{-\infty}^{\infty} \sin xt\, dF(x)$$

the successive derivatives of $\phi(t)$ when evaluated at $t = 0$ yield the moments of $F(x)$ except for a factor of a power of "i".

$$\frac{d}{dt}\phi(t) = \phi'(t) = \int_{-\infty}^{\infty} ix\, e^{ixt}\, dF(x)$$

$$\phi'(0) = i\int_{-\infty}^{\infty} x\, e^{0}\, dF(x) = i\,\mu$$

$$\phi''(t) = \int_{-\infty}^{\infty} (ix)^2\, e^{ixt}\, dF(x)$$

$$\phi''(0) = i^2 \int_{-\infty}^{\infty} x^2\, dF(x) = i^2\alpha_2$$

in general

$$\phi^k(0) = i^k E(X^k) = i^k\alpha_k$$

the moment generating function operates in the same manner, except it does not include the factor "i", and is therefore not as general in application,

$$MGF = M(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt}\, dF(x) \quad \text{if this integral exists}$$

and operates by evaluating successive derivatives with respect to t at $t = 0$

Lemma: if $E(X^k)$ exists, then $\phi^k(t)$ exists and is continuous. The converse is also true.

<u>Examples:</u>

1. Trivial distribution: $Pr\left[X = a\right] = 1$

$$\phi(t) = \int_{-\infty}^{\infty} e^{ixt}\, dF(x) = e^{iat} \times 1 = e^{iat}$$

2. Binomial:

$$\phi(t) = \sum_{k=0}^{n} e^{itk}\binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum \binom{n}{k}(pe^{it})^k (1-p)^{n-k} = \left[pe^{it} + (1-p)\right]^n$$

3. Poisson: $$\phi(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= \sum e^{-\lambda} \frac{(\lambda e^{it})^k}{k!}$$

$$= e^{-\lambda} e^{\lambda e^{it}}$$

$$= e^{\lambda (e^{it} - 1)}$$

4. Normal (0, 1):

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2 - 2itx}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2 - 2itx + (it)^2}{2} + \frac{(it)^2}{2}} dx$$

setting $y = x - it$*

$$= \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right\} e^{-t^2/2}$$

$$= e^{-t^2/2}$$

(Note the term in curly brackets is the integral of a normal density and equals one.)

* The validity of this complex transformation has to be justified. See Problem 11.

If X is N(0, 1)

$$\phi(t) = e^{-t^2/2}$$

$$\phi'(t) = -\frac{2t}{2} e^{-t^2/2}$$

$$\phi''(t) = -e^{-t^2/2} + t^2 e^{-t^2/2}$$

$$\phi''(0) = -1$$

$$E(X^2) = i^2 \phi'' = 1$$

Problem 10: Prove

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, dx = 1$$

Problem 11: Show that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} \, e^{-x^2/2} \, dx = e^{-t^2/2}$$

without using the transformation used in class

Hint: $e^{itx} = \cos tx + i \sin tx$

Theorem 10:

$$\phi_{AX+B}(t) = e^{iBt} \phi_X(At)$$

where X is a random variable with a d.f. F(x) and a characteristic function $\phi(t)$

Proof: $Y = AX + B$ where A, B are constants

$$\phi_Y(t) = E(e^{itY}) = E\left[e^{it(AX+B)}\right]$$

$$= E(e^{itAX} e^{itB}) = e^{itB} \underbrace{E(e^{i(At)X})}_{\phi_X(At)}$$

$$= e^{iBt} \phi_X(At)$$

if X is $N(0, 1)$ then $Y = \sigma X + \mu$ is $N(\mu, \sigma^2)$

$$\phi_Y(t) = e^{it\mu} \phi_X(\sigma t)$$

$$= e^{it\mu} e^{-\frac{(\sigma t)^2}{2}} = e^{it\mu - \frac{(\sigma t)^2}{2}}$$

Def. 17: The cumulant generating function, K(t), is defined to be:

$$K(t) = \ln \phi(t)$$

Example: For Y which is $N(\mu, \sigma^2)$

$$K(t) = it\mu - \frac{\sigma^2 t^2}{2}$$

Note: For further discussion of cumulants see Cramer or Kendall.

Note: Originally cumulants = semi-invariants (British school name = Scandanavian school name) -- however, semi-invariants have been extended so that now cumulants are a special case.

Theorem 11: If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\phi_{\sum X_i}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

(the c.f. of a sum = the product of the individual c.f.)

Proof:

$$E\left[e^{(it \sum X_i)}\right] = E\left[e^{itX_1} e^{itX_2} \ldots e^{itX_n}\right]$$

$$= E\left[e^{itX_1}\right] E\left[e^{itX_2}\right] \ldots E\left[e^{itX_n}\right] \text{ by independence}$$

$$= \phi_{X_1}(t)\ \phi_{X_2}(t) \ldots \phi_{X_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

Example: If X_i are NID (μ_i, σ_i^2), then the c.f. of $Y = \sum_1^n X_i$

$$= \prod_1^n \left(e^{it\mu_i - \frac{\sigma_i^2 t^2}{2}}\right)$$

$$= e^{it \sum \mu_i - \frac{(\sum \sigma_i^2)}{2} t^2}$$

therefore we could say that Y is $N(\sum_1^n \mu_i, \sum_1^n \sigma_i^2)$

To justify this last step we need to show the converse of $F(X) \longrightarrow \phi_X(t)$ i.e., that $\phi_X(t) \longrightarrow F(X)$. Therefore, we need the following lemma and theorem.

Lemma:

$$\lim_{T \to \infty} \frac{2}{\pi} \int_0^T \frac{\sin ht}{t}\, dt = \begin{cases} -1 & h < 0 \\ 0 & h = 0 \\ +1 & h > 0 \end{cases}$$

Proof: $$J(\alpha, \beta) = \int_0^\infty e^{-\alpha u} \frac{\sin \beta u}{u}\, du \qquad \alpha \geq 0$$

$$\frac{\partial J}{\partial \beta} = \int_0^\infty e^{-\alpha u} \cos \beta u\, du$$

Note: Differentiation under the integral can be justified.

$$= \frac{\alpha}{\alpha^2+\beta^2} \qquad \text{see tables or integrate by parts twice}$$

$$\int \frac{\partial J}{\partial \beta}\, d\beta = \int \frac{\alpha}{\alpha^2+\beta^2}\, d\beta = \text{arc tan}\,\frac{\beta}{\alpha} + C$$

Let $\beta \to 0$ then $J(\alpha, 0) = e^{\alpha u} \cdot 0 \cdot du = 0$

$$\text{arc tan } 0 + C = 0$$
$$0 + C = 0$$
$$\therefore C = 0$$

$$J(\alpha, \beta) = \text{arc tan}\,\frac{\beta}{\alpha}$$

Let $\alpha \to 0$, and put $\beta = h$, then

$$\lim_{\alpha \to 0} J(\alpha, h) = \text{arc tan} \pm \infty \text{ depending on } h + \text{ or } -$$
$$= \frac{\pi}{2} \qquad h > 0$$
$$= -\frac{\pi}{2} \qquad h < 0$$

<u>Theorem 12</u>: If $F(x)$ is continuous at $a - h$, $a + h$, then

$$F(a+h) - F(a-h) = \lim_{T \to \infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t}\, e^{-ita}\, \phi(t)\, dt$$

If $\int_{-\infty}^{\infty} |\phi(t)|\, dt < \infty$ then $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, \phi(t)\, dt$

Note: Recall that $\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$ [def. 16]

Combining this with the above theorem means that given $\phi(t)$ or $f(x)$ we can determine the other.

Proof:

Define $J = \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t}\, e^{-ita}\, \phi(t)\, dt$

$$= \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t}\, e^{-ita} \left[\int_{-\infty}^{\infty} e^{itx}\, d\,F(x) \right] dt$$

Interchanging integrals (reversing the order of integration) which can be justified this becomes

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{-T}^{T} \frac{\sin ht}{t} \underbrace{e^{-ita} \; e^{itx}}_{e^{it(x-a)}} \, dt \right\} d\,F(x)$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{-T}^{T} \frac{\sin ht}{t} \Big[\cos t(x-a) + i \sin t(x-a) \Big] \, dt \right\} dF(x)$$

Note: The $\frac{\sin}{t}$ sin term is an odd function $\therefore \int_{-T}^{T} = 0$

The $\frac{\sin}{t}$ cos term is an even function $\therefore \int_{-T}^{T} = 2\int_{0}^{T}$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} 2 \int_{0}^{T} \frac{\sin ht}{t} \cos t(x-a) \, dt \, dF(x)$$

Note: $2 \cos A \sin B = \sin (A+B) - \sin (A - B)$

$$J = \frac{1}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{0}^{T} \frac{\sin t(x-a+h) - \sin t(x-a-h)}{t} \, dt \right\} d\,F(x)$$

now take the limit as $T \to \infty$

using the lemma just proved, with that h = (x-a+h) here

		$x=a-h$		$x=a+h$	
For $\sin t(x-a+h)$	$x-a+h<0$		$x-a+h>0$		
For $\sin t(x-a-h)$			$x-a-h<0$		$x-a-h>0$
For x in each region	both $= -\frac{\pi}{2}$		1st part $\frac{\pi}{2}$ 2nd part $-\frac{\pi}{2}$		both $= \frac{\pi}{2}$
Whole integral (in brackets)	0		π		0

i.e., in the region $a - h \leq x \leq a + h$

$$\int_0^T \frac{1}{t} \sin t(x - a + h)\, dt - \int_0^T \frac{1}{t} \sin t(x - a - h)\, dt = \pi$$

elsewhere $= 0$

There
$$J = \frac{1}{\pi} \int_{a-h}^{a+h} \pi \, dF(x) + \int_{-\infty}^{a-h} 0 .\, dF(x) + \int_{a+h}^{\infty} 0 .\, dF(x)$$
$$= F(a + h) - F(a - h)$$

Proof of the second statement in the theorem:

$$\frac{F(a + h) - F(a - h)}{2h} = \frac{1}{2h\,\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{t} e^{-ita} \phi(t)\, dt$$

taking the limit of both sides as $h \to 0$

$$f(a) = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} \underbrace{\lim_{h \to 0} \frac{\sin ht}{ht}}_{= 1} e^{-ita} \phi(t)\, dt$$

therefore

$$f(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ita} \phi(t)\, dt$$

<u>Problem 12</u>: Let X_i $(i = 1, 2, \ldots, n)$ have a density function given by

$$f(x) = a\, x^{a-1} \qquad a > 0 \qquad 0 \leq x \leq 1$$

Find the density of $Y = \prod_{i=1}^{n} X_i$ $\qquad X_i$ are independent

(may need the result of Cramer p. 126)

<u>Problem 13</u>: Define a factorial moment

$$E(x_{[r]}) = E\left[x(x-1) \ldots (x-r+1)\right]$$

Define

$F^*(t) = \int_{-\infty}^{\infty} (1 + t)^x \, dF(x)$ as the factorial moment generating function.

Find

$F^*(t)$ for the binomial and use this to get the factorial moments.

Problem 14: If $\phi(t) = e^{-|t|}$ find the density of $f(x)$ corresponding to ϕ. Find the distribution of the mean of n independent variables with this d.f.

Theorem 13: A necessary and sufficient condition that a sequence of distribution functions F_n tend to F at every point of continuity of F is that ϕ_n, the characteristic function corresponding to F_n, tend to a function, $\phi(t)$ which is continuous at $t = 0$ [or tends to a function $\phi(t)$ which is itself a characteristic function].

Proof: Omitted -- see Cramer p. 96-98

This theorem says, if we have
$$\begin{matrix} F_1 & F_2 & F_3 & \cdots\cdots & F \\ \phi_1 & \phi_2 & \phi_3 & \cdots\cdots & \phi \end{matrix}$$

we go from the F_i to the ϕ_i -- observe that the ϕ_i tend to a limit [for exampl the normal approximation to the binomial] -- observe that the limit is itself a characteristic function [of the normal] -- then go to F by previously discussed techniques.

Theorem 14: (Central Limit Theorem)

If $X_1, X_2, X_3 \cdots\cdots$ are a sequence of independently and identically distributed random variables with a distribution function $F(x)$ with finite first and second moments [say mean μ and variance σ^2] then:

1-- the distribution of $Y_n = \sqrt{n}\left(\frac{1}{n}\sum X_i - \mu\right)$ tends, as $n \to \infty$, to the normal distribution with mean 0 and variance σ^2

2-- for any interval (a, b)

$$\lim_{n\to\infty} \Pr\left[a < Y_n < b\right] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_a^b e^{-t^2/2\sigma^2}\,dt$$

3-- the sequence $\left[Y_n\right]$ is asymptotically $N(0, \sigma^2)$

Proof: Denote the c.f. of X_i as $\phi(t)$

to get the c.f. of $Y_n = \sqrt{n}\ \dfrac{\sum(X_i - \mu)}{n} = \dfrac{\sum(X_i - \mu)}{\sqrt{n}}$

$$\phi_{Y_n}(t) = \left[e^{-\frac{\mu}{\sqrt{n}}it}\ \phi\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

if we expand $\phi(t)$ in a Taylor series, we have

$$\phi(t) = 1 + i\alpha_1 t + i^2\alpha_2 t^2/2 + \text{remainder} \qquad \alpha_1 = \mu; \quad \alpha_2 = \mu^2 + \sigma^2$$

$$= 1 + i\mu t - (\mu^2 + \sigma^2)\, t^2/2 + R(t)$$

where $\frac{R(t)}{t^2} \to 0$ as $t \to 0$

$$\ln \phi_{Y_n}(t) = n\left[\ln e^{-i\frac{\mu}{\sqrt{n}}t} + \ln \phi\left(\frac{t}{\sqrt{n}}\right)\right]$$

$$= i\sqrt{n}\,\mu t + n \ln \phi\left(\frac{t}{\sqrt{n}}\right)$$

$$= i\mu\sqrt{n}\, t + n \ln\left[1 + \frac{i\mu t}{\sqrt{n}} - \frac{(\mu^2 + \sigma^2)t^2}{2n} + R\left(\frac{t}{\sqrt{n}}\right)\right]$$

Note: $\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \dots\dots$

$$= x - \frac{x^2}{2} + R'$$

where $\frac{R'}{x^2} \to 0$ as $x \to 0$

setting $x = i\mu\frac{t}{\sqrt{n}} - (\mu^2 + \sigma^2)\frac{t^2}{2n} + R\left(\frac{t}{\sqrt{n}}\right)$ we get

$$\ln \phi_{Y_n}(t) = -i\sqrt{n}\,\mu t + n\left[i\mu\frac{t}{\sqrt{n}} - (\mu^2 + \sigma^2)\frac{t^2}{2n} + R\left(\frac{t}{\sqrt{n}}\right) - \frac{1}{2}\left(i\mu\frac{t}{\sqrt{n}}\right)^2 + \underbrace{\frac{t^3}{n^{3/2}}A_n + \frac{R''(t)}{n^2}}_{\text{denote this by } \frac{R'''}{n^{3/2}}}\right]$$

$$= -\frac{\sigma^2 t^2}{2} + n R\left(\frac{t}{\sqrt{n}}\right) + \frac{R'''}{\sqrt{n}}$$

$$\lim_{n\to\infty} \ln \phi_{Y_n}(t) = -\frac{\sigma^2 t^2}{2} \qquad \text{since } \lim_{n\to\infty} \frac{R'''}{\sqrt{n}} = 0$$

$$\lim_{n\to\infty} \frac{R\left(\frac{t}{\sqrt{n}}\right) t^2}{\left(\frac{t}{\sqrt{n}}\right)^2} = 0$$

$$\therefore \lim_{n\to\infty} \phi_{Y_n}(t) = e^{-\frac{\sigma^2 t^2}{2}}$$

since as $n \to \infty$, $\frac{t}{\sqrt{n}} \to 0$

which is the characteristic function of the normal distribution $N(0, \sigma^2)$

Note: See Cramer p. 214-215.

The theorem says, for any given ε there is some sufficiently large $n(\varepsilon, a, b)$ such that

$$\left| \Pr\left[a < Y_n < b\right] - \frac{1}{\sqrt{2\pi}\,\sigma} \int_a^b e^{-t^2/2\sigma^2}\, dt \right| < \varepsilon$$

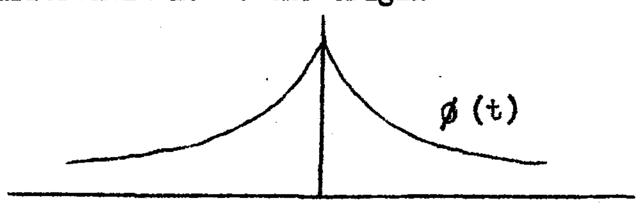
<u>Problem 15</u>: Let X be a Poisson random variable with parameter λ. Show that for any (a, b)

$$\lim_{\lambda \to \infty} \Pr\left[a < \frac{X - \lambda}{(\lambda)^{1/2}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt$$

a. directly -- use Sterling's approximation
b. by characteristic functions see Feller ch. VII

Problem 15 -- says that the standardized Poisson random variable is asymptotically normally distributed N(0, 1)

Cauchy Distribution -- see problem 14 --- has no mean or variance since $e^{-|t|}$ is not differentiable at the origin



$$\lim_{\substack{b \to \infty \\ a \to \infty}} \frac{1}{\pi} \int_a^b \frac{x}{1+x^2}\, dx = \frac{1}{2\pi} \ln(1+x^2) \Big|_a^b$$

$$= \infty - \infty$$

therefore E(x) does not exist for a Cauchy random variable, even though the distribution is symmetric about the origin.

<u>Theorem 15</u>: (Liapounoff)

If $X_1, X_2, \ldots X_n$ are independent random variables with means μ_i and variances σ_i^2 and with

$$\rho_i^3 = E\,|X_i - \varepsilon_i|^3 < \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{\sum_1^n \rho_i^3}{\left(\sum \sigma_i^2\right)^{3/2}} = 0$$

then

$$Y = \frac{\sum X_i - \sum \mu_i}{\left(\sum \sigma_i^2\right)^{1/2}} \text{ is asymptotically } N(0, 1)$$

Proof: See Cramer p. 216-217.

Problem 16: Define $M^*(t) = E(X^t)$ as the Mellin Transform

If X has density $f(x) = k\,x^{k-1}$ $0 \leq x \leq 1$

then $M^*(t) = \frac{k}{k+t}$

If X has density $f(x) = -k^2\,x^{k-1}\ln x$ $0 \leq x \leq 1$

then $M^*(t) = \left(\frac{k}{k+t}\right)^2$

Use this to find the density of $Y = X_1 X_2$ where the X_i are independent and have density $f(x) = k\,x^{k-1}$

State a theorem necessary to validate this approach.

Laplace Transform:

$$E\left[e^{-sx}\right] = f(\hat{s}) = \int_0^\infty e^{-sx}\,f(x)\,dx \qquad \text{if } f(x) = 0 \quad x < 0$$

-- extensive uses in differential equations
-- extensive tables of the L.T. in the literature -- tables for passing from the transform to the fuction and vice versa -- see Doetsch, Tables of the Laplace Transform.

Note: Replacing (s) by (-t) gives the m.g.f.; or by (-it) gives the c.f.

Fourier Transform -- the mathematical name for the characteristic function

$$= E\left[e^{itx}\right]$$

We have previously noted that if X, Y are NID with means μ_i and variances σ_i^2 then $Z = X + Y$ is normal $N(\mu_1 + \mu_2;\ \sigma_1^2 + \sigma_2^2)$.

There is a converse to this "addition theorem for normal variates" to the effect that:

if Z = X + Y where X, Y are independent and Z is normal, then X, Y are both normal --- see Cramer p. 213.

Derived Distributions (from the Normal and others):

name	density	c.f.	E(x)	σ^2
χ_n^2	$\frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})}\,e^{-x/2}\,x^{\frac{n}{2}-1}$	$(1-2it)^{-\frac{n}{2}}$	n	2n

-- if $X_1, \ldots, X_n$ are NID(0, 1), $\sum_1^n X_i^2$ is χ_n^2

-- it has the additive property $\chi_{m+n}^2 = \chi_m^2 + \chi_n^2$

Gamma

$\frac{a^\lambda}{\Gamma(\lambda)} e^{-ax} x^{\lambda-1}$ $\quad x \geq 0$ $\qquad (1 - \frac{it}{a})^{-\lambda}$ $\qquad \frac{\lambda}{a}$ $\qquad \frac{\lambda}{a^2}$

-- $\chi_{(1)}^2$ is a special case of the gamma

-- = Pearson's Type III distribution with starting point at the origin.

Student's t (n) $\quad \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(\frac{n}{2})} \frac{1}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}}$ $\qquad$ -- $\qquad 0$ $(n>1)$ $\qquad \frac{n}{n-2}$ $(n>2)$

-- $t = \frac{X\sqrt{n}}{Y}$ where X, Y are independent, X is N(0, 1), Y is χ_n^2

-- t with 1 d.f. is a Cauchy distribution

F(m,n) $\quad \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\ \Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{n/2} \frac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{\frac{m+n}{2}}}$ $\quad x>0$ $\qquad$ -- $\qquad \frac{n}{n-2}$ $\qquad \frac{2n^2(m+n-2)}{m(n-2)^2\ (n-4)}$

-- if X, Y are independent, X is χ_m^2, Y is χ_n^2, then $F = \frac{X/m}{Y/n}$

Fisher's z is defined by $F = e^{2z}$ see Cramer p. 243.

Beta

$\beta(p,q)$ $\quad \frac{\Gamma(p+q)}{\Gamma p\,\Gamma q} x^{p-1} (1-x)^{q-1}$ $\quad 0 \leq x \leq 1$ $\qquad$ -- $\qquad \frac{p}{p+q}$ $\qquad \frac{pq}{(p+q)^2(p+q+1)}$

-- putting $p = \frac{m}{2}$. $q = \frac{n}{2}$, we get $\beta = \frac{\frac{m}{n} F}{1 + \frac{m}{n} F}$

-- see Kendall for the relation to the Binomial.

Gamma function:

$$\Gamma(p) = \int_0^\infty e^{-x} x^{n-1} dx \qquad n>0$$

$$\Gamma(p) = (p-1)\,\Gamma(p-1)$$

if p is a positive integer $\Gamma(p) = (p-1)!$

$$\lim_{p\to\infty} \frac{\Gamma(p+1)}{p^p e^{-p} \sqrt{2\pi p}} = 1 \qquad \text{Stirling's Formula}$$

$$\lim_{p\to\infty} \frac{\Gamma(p+h)}{p^h \Gamma(p)} = 1 \qquad \text{h Fixed}$$

Problem 17: X and Y are NID $(0, \sigma^2)$. Find the marginal distribution of

1. $r = \sqrt{X^2 + Y^2}$

2. $\theta = \arc\tan Y/X$

Problem 18:

Cramer p. 319 No. 10.

Chapter III

Convergence

Convergence of Distributions:

a. Central Limit Theorem (theorem 14)

b. Poisson distribution -- (problem 15) -- if X is a Poisson r.v., then

$$\lim_{\lambda \to \infty} \Pr\left[a < \frac{X - \lambda}{\sqrt{\lambda}} < b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt$$

Proof: 1. By c.f. is straightforward -- see sol. to problem 15, or p. 250.

2. Direct

$$P_{(a,b)} = \sum_{\lambda + a\sqrt{\lambda}}^{\lambda + b\sqrt{\lambda}} e^{-\lambda} \frac{\lambda^K}{K!}$$

where $P_{(a,b)}$ is the above probability statement.

Let $K = \lambda + \sqrt{\lambda}\, x$

$$x = \frac{K - \lambda}{\sqrt{\lambda}}$$

$$x + \Delta x = \frac{K + 1 - \lambda}{\sqrt{\lambda}}$$

$$\Delta x = \frac{1}{\sqrt{\lambda}}$$

$$P_{(a,b)} = \sum_{x=a}^{b} e^{-\lambda} \frac{\lambda^{\lambda + x\sqrt{\lambda}}}{(\lambda + x\sqrt{\lambda})!}$$

using Stirling's Formula:

$$\frac{n!}{n^n e^{-n} (2\pi n)^{1/2}} = 1 + \theta(n)$$

where $\theta(n) \to 0$ as $n \to \infty$

$$= \sum_a^b e^{-\lambda} \frac{\lambda^{\lambda + x\sqrt{\lambda}}}{(2\pi)^{1/2} (\lambda + x\sqrt{\lambda})^{\lambda + x\sqrt{\lambda} + 1/2}\, e^{-(\lambda + x\sqrt{\lambda})}}$$

$$= \frac{1}{\sqrt{2\pi}} \sum_a^b \frac{e^{x\sqrt{\lambda}}}{(1 + \frac{x}{\sqrt{\lambda}})^{\lambda}} \cdot \frac{1}{(1 + \frac{x}{\sqrt{\lambda}})^{x\sqrt{\lambda}}} \frac{1}{\sqrt{\lambda}} \frac{1}{(1 + \frac{x}{\sqrt{\lambda}})^{1/2}} \left[1 + \theta(\lambda) \right]$$

Notes: $\lim\limits_{\lambda \to \infty} \dfrac{1}{(1+ \frac{x}{\sqrt{\lambda}})^{x\sqrt{\lambda}}} = \lim\limits_{x\sqrt{\lambda} = u \to \infty} \dfrac{1}{(1+\frac{x^2}{u})^u} = e^{-x^2}$

$$\lim_{\lambda \to \infty} \ln \frac{e^{x\sqrt{\lambda}}}{(1+ \frac{x}{\sqrt{\lambda}})^{\lambda}} = \lim_{\lambda \to \infty} \left\{ x\sqrt{\lambda} - \lambda \ln \left[1 + \frac{x}{\lambda} \right] \right\}$$

$$= \lim_{\lambda \to \infty} \left\{ \left[x\sqrt{\lambda} - \lambda \left(\frac{x}{\sqrt{\lambda}} - \frac{x^2}{2\lambda} + R\left(\frac{1}{\lambda^{1/2}} \right) \right) \right] \right\}$$

$$= \frac{x^2}{2}$$

$$P_{(a,b)} = \frac{1}{\sqrt{2\pi}} \sum_a^b \frac{e^{x^2/2}}{e^{x^2}} \Delta x(1 + \theta'(\lambda) \qquad \text{where } \theta'(\lambda) \to 0 \text{ as } \lambda \to \infty$$

hence the $\lim\limits_{\lambda \to \infty} P_{(a,b)} = \lim \left[\text{Riemann sum } \frac{1}{2\pi} \sum_a^b e^{-x^2/2} \Delta x \right]$

$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Note: For a similar proof for the binomial case, see J. Neyman: First Course in Statistics, Chapter 4.

c. If X has the χ^2 distribution, then

$$\frac{X - n}{\sqrt{2n}} \text{ is A } N(0,1) \text{ as } n \to \infty$$

Proof: See Cramer, p. 250.

d. If X has the Student's distribution with n d.f., then

$$\frac{X - 0}{\sqrt{\frac{n}{n-2}}} \text{ is A } N(0,1) \text{ as } n \to \infty$$

Proof: Deferred for now; a proof working with densities is given by Cramer, p. 250.

e. If X has the Beta distribution with parameters p, q, then the standardized variate is A N(0,1) as $p \to \infty$, $q \to \infty$, and p/q remains finite

Note: X has mean $\dfrac{p}{p+q} = \dfrac{1}{1 + q/p}$

Proof: Omitted

Problem 19: X is a negative binomial r.v. (r,p)

a. Find the limiting d.f. of X as $p \to 1$; $q \to 0$; $rq \to \lambda$ (finite).
b. State and prove a theorem that shows that under certain conditions a linear function of X is A $N(0,1)$.

Problem 20: X,Y have a joint density $f(x,y)$.

Define $E\left[g(X) \mid Y\right] = \int_{-\infty}^{\infty} g(x)\, f(x|y)dx$

where $f(x \mid y) = f(x,y)/f_1(y)$. $f_1(y)$ is the marginal density of y.

Show that $E\left[g(X)\right] = \underset{Y}{E}\ \underset{X}{E}\left[g(X)|Y\right]$

Use this to find the unconditional distribution of X if $X|Y$ is $B(Y,p)$ while Y is Poisson (λ).

Convergence in Probability

Def. 18: A sequence of r.v., $X_1, X_2, \ldots, X_n$ is said to converge in probability to a r.v. X if given ε, δ there exists a number $N(\varepsilon,\delta)$ such that for $n > N(\varepsilon,\delta)$

$$Pr\left[\,|X_n - X| > \varepsilon\right] < \delta$$

which is written $X_n \xrightarrow[P]{} X$ "$\xrightarrow[P]{}$" indicates converging in probability.

As a special case of this we have $X_n \xrightarrow[P]{} c$ if given ε & δ there exists N such that for $n > N$

$$Pr\left[\,|X_n - c| > \varepsilon\right] < \delta$$

Further, in this notation, theorem 9 may be written:
if X_n is a binomial r.v. with parameter n, then

$$\frac{X_n}{n} \xrightarrow[P]{} p$$

Theorem 16: If $X_1, X_2, \ldots, X_n$ are a sequence of independent random variables with means μ_i and variances σ_i^2 then if

$$\frac{\sum_1^n \sigma_i^2}{n^2} \to 0$$

then

$$Y_n = \frac{1}{n}\sum_1^n (X_i - \mu_i) \xrightarrow[P]{} 0$$

Proof: By the corollary to the Tchebycheff Theorem (thm.8)

$$\Pr\left[\left|\frac{1}{n}\sum_1^n (X_i - \mu_i)\right| > K\,\sigma_{Y_n}\right] \leq \frac{1}{K^2}$$

given ε, δ choose $\frac{1}{K^2} = \frac{\delta}{2} < \delta$

Now $\sigma^2_{Y_n} = \frac{1}{n^2}\sum_1^n \sigma_i^2 \longrightarrow 0 \quad$ as $n \longrightarrow \infty$

Put $K = \sqrt{\frac{2}{\delta}}$ and take n sufficiently large so that

$$\frac{\sqrt{2}}{\sqrt{\delta}}\left(\frac{1}{n^2}\sum_1^n \sigma_i^2\right)^{1/2} < \varepsilon$$

with this choice of n

$$\Pr\left[\left|\frac{1}{n}\sum_1^n (X_i - \mu_i)\right| > \varepsilon\right] < \delta$$

If $X_1, X_2, X_3, \ldots\ldots$ have the same distribution, then $\sigma^2_{Y_n} = \frac{\sigma^2}{n}$

and $\sigma^2_{Y_n} = \frac{\sigma^2}{n} \rightarrow 0 \quad$ as $n \longrightarrow \infty$

so that in this particular case, $\bar{X} \xrightarrow[P]{} \mu$.

Theorem 17: (Khintchine) (Weak Law of Large Numbers)

If $X_1, X_2, \ldots, X_n$ are independently and identically distributed r.v. with mean μ, then $\bar{X} \xrightarrow[P]{} \mu$.

Proof: (See Cramer, p. 253-4)

$$\phi_X(t) = E(e^{iX_n t})$$

$$\phi_{\bar{X}}(t) = E(e^{i(\frac{1}{n}\sum X_i)t}) = \left[\phi_X(\tfrac{t}{n})\right]^n$$

$$\ln \phi_{\bar{X}}(t) = n \ln \phi(\tfrac{t}{n}) = n \ln\left[1 + i\mu\frac{t}{n} + R\right]$$

where $\frac{R}{t} \rightarrow 0 \quad$ as $t \rightarrow 0$

$$\phi_{\bar{X}-\mu}(t) = e^{-i\mu t}\phi_{\bar{X}}(t)$$

$$\ln \phi_{\bar{X}-\mu}(t) = -i\mu t + n \ln(1 + i\mu\frac{t}{n} + R)$$

Note: $\ln(1+x) = x + R(x)$

where $\frac{R(x)}{x} \to 0$ as $x \to 0$

$$= -i\mu t + n\left[i\mu\frac{t}{n} + R(\tfrac{t}{n}) + R'\right]$$

$$= \frac{R''(\frac{t}{n})}{\frac{t}{n}}\, t \text{ which as } n \to \infty \quad \frac{t}{n} \to 0$$

$$\text{so } \frac{R''(\frac{t}{n})}{\frac{t}{n}} \to 0$$

hence $\phi_{\bar{X}-\mu}(t) \to 1$ as $n \to \infty$

but if $\phi(t) = 1$ $\quad F(x) = 0 \quad x < 0$

$\qquad\qquad\qquad\qquad = 1 \quad x \geq 0$

or, in other words, the limiting distribution of $\bar{X} - \mu$ is the trivial d.f. which takes the value 0 with probability 1.

Question:

If $X_1, X_2, X_3, \ldots, X_n \xrightarrow[P]{} X$

Do $\mu_1 \ \mu_2 \ \mu_3 \ \ldots \ \mu_n \to \mu$?

Do $\sigma_1^2 \ \sigma_2^2 \ \sigma_3^2 \ \ldots \ \sigma_n^2 \to \sigma^2$?

i.e., does $X_n \xrightarrow[P]{} X$ imply $\mu_n \to \mu$?

Not necessarily, as shown by the following example.

Example:

Let X_n be defined as follows:

$X_n = n^2$ with probability $\frac{1}{n}$

$\quad = 0$ with probability $1 - \frac{1}{n}$

$\therefore X_n \xrightarrow[p]{} 0$

$E\left[X_n\right] = 0(1 - \frac{1}{n}) + n^2(\frac{1}{n}) = n$

as $n \to \infty$ $\quad E\left[X_n\right] \to \infty$

Problem 21: Let X be the r.v. defined as the length of the first run in a series of binomial trials with $Pr\left[\text{success}\right] = p$. Find the distri bution of X, $E\left[X\right]$, and $\sigma^2(X)$.

Problem 22: Let $X_1, X_2, \ldots, X_n$ be independently uniformly distributed (on 0,1 Let $Y = \min(X_1, X_2, \ldots, X_n)$. Find the d.f. of Y, $E\left[Y\right]$, and $\sigma^2(Y)$.

Find the asymptotic d.f. of nY and of $\frac{Y - E(Y)}{\sigma_Y}$.

Is $\frac{Y - E(Y)}{\sigma_Y}$ A N(0,1) as $n \to \infty$?

Theorem 18: Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables with distributi functions $F_1(x), F_2(x), \ldots, F_n(x) \to F(x)$. Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of random variables tending in probability to c.

Define: $U_n = X_n + Y_n$; $\quad V_n = X_n Y_n$; and $W_n = X_n/Y_n$.

a) the d.f. of $U_n \to F(x-c)$; and if $c > 0$

b) the d.f. of $V_n \to F(x/c)$

c) the d.f. of $W_n \to F(xc)$

Proof: All three parts are similar — see Cramer p. 254-5 for proof of the third part.

Proof of the first statement (a):

Assume $x - c$ is a point of continuity of F. Let ε be small so that $x - c \pm \varepsilon$ is an interval of continuity.

Let S_1 = the set of points such that

$$X_n + Y_n \leq x; \quad |Y_n - c| \leq \varepsilon$$

S_2 = the set of points such that

$$X_n + Y_n \leq x; \quad |Y_n - c| > \varepsilon$$

$S = S_1 + S_2$ = the set of points such that

$$X_n + Y_n \leq x$$

$P(S_2) = Pr\left[X_n + Y_n \leqq x, \; |Y_n - c| > \varepsilon\right] \leqq Pr\left[|Y_n - c| > \varepsilon\right]$ which tends to 0 as $n \to \infty$, therefore we can choose n_1 so that $n > n_1$ implies $P(S_2) < \frac{\delta}{3}$

In S_1: $c - \varepsilon \leqq Y_n \leqq c + \varepsilon$, thus

$$F(x - c - \varepsilon) - \frac{\delta}{3} < F(x - c - \varepsilon) = P\left[X_n \leqq x - (c + \varepsilon)\right] \leqq$$

$$P(X_n + Y_n \leqq x) = P(X_n = x - Y_n)$$

$$\leqq P\left[X_n \leqq x - (c - \varepsilon)\right] = F_{n_2}\left[x - c + \varepsilon\right] < F(x - c + \varepsilon) - \frac{\delta}{3}$$

where n_2 is chosen so that when $n > n_2$ $F_n(x) - F(x) < \frac{\delta}{3}$ in the vicinity of c.

Therefore, in S_1

$$F(x - c - \varepsilon) - \frac{\delta}{3} < P(X_n + Y_n \leqq x) < F(x - c + \varepsilon) + \frac{\delta}{3}$$

ε can be chosen so that $F(x - c + \varepsilon) - F(x - c - \varepsilon) < \frac{\delta}{3}$ for $n > \max(n_1, n_2)$.

Noting that $Pr\left[U_n \leq x\right] = Pr\left[X_n + Y_n \leqq x\right] = P(S_1) + P(S_2)$, we can write:

$$-\delta \leqq -\frac{\delta}{3} - \frac{\delta}{3} \leqq F(x - c - \varepsilon) - \frac{\delta}{3} + 0 - F(x - c) \leqq$$

$$P(S_1) + P(S_2) - F(x - c) = Pr\left[X_n \leq x\right] - F(x - c)$$

$$\leqq F(x - c + \varepsilon) + \frac{\delta}{3} + \frac{\delta}{3} - F(x - c) \leqq \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} \leqq \delta$$

which makes use of $F(x - c + \varepsilon) - F(x - c) = \frac{\delta}{3}$

$$F(x - c - \varepsilon) - F(x - c) = -\frac{\delta}{3}$$

This then reduces to $-\delta \leqq Pr\left[X_n \leqq x\right] - F(x - c) \leqq \delta$

hence $\left|Pr\left[X_n \leqq x\right] - F(x - c)\right| \leqq \delta$ for $n > \max(n_1, n_2)$

which is what we set out to prove in the first place.

Theorem 19: If $X_n \xrightarrow[P]{} c$ and if $g(x)$ is a function continuous at $x = c$,

$g(X_n) \xrightarrow[P]{} g(c)$

Problem 23: Prove Theorem 19. (work with fact that $g(x)$ is continuous)

Example of Theorem 18:

Show t_n is A N(0,1)

$$t_n = \frac{\sqrt{n}(\bar{X} - \mu)}{s_n}$$ where $x_1, x_2, \ldots$ are independent with mean μ, var σ^2

$$s_n^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

$$t_n = \frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} \cdot \frac{\sigma}{s_n} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \Big/ \frac{s_n}{\sigma}$$ which is in the W_n form

$$X_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \qquad Y_n = \frac{s_n}{\sigma}$$

X_n is A N(0,1) by the central limit theorem.
Hence, if we show that

$$Y_n = \frac{s_n}{\sigma} \xrightarrow[P]{} 1$$

then the statement that t is A N(0,1) follows from theorem 18(c).

$$\frac{s_n^2}{\sigma^2} = \frac{\sum_1^n (X_i - \bar{X})^2}{(n - 1)\sigma^2} = \frac{\sum_1^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2}{\sigma^2(n - 1)}$$

$$= \frac{n}{n - 1} \frac{\sum (X_i - \mu)^2}{n\sigma^2} - \frac{n}{n-1} \frac{(\bar{X} - \mu)^2}{\sigma^2}$$

$$\frac{s_n^2}{\sigma^2} = \frac{n}{n-1} \left[\frac{1/n \sum (X_i - \mu)^2}{\sigma^2} - \frac{(\bar{X} - \mu)^2}{\sigma^2} \right]$$

by Khintchine's theorem (No. 17) this sample mean tends to σ^2

i.e., $\frac{1}{n}\sum(\bar{X}_i - \mu)^2 \xrightarrow[P]{} \sigma^2$

therefore, the first term $\xrightarrow[P]{} 1$

$\frac{n}{n-1} \rightarrow 1$ as $n \rightarrow \infty$

hence, it remains to show that

$$\frac{(\bar{X} - \mu)^2}{\sigma^2} \xrightarrow[P]{} 0$$

but we know that $\left|\bar{X} - \mu\right| \xrightarrow[P]{} 0$ as $n \rightarrow \infty$ by Khintchine's theorem (No. 17).

Therefore, $\frac{(\bar{X} - \mu)^2}{\sigma^2} \xrightarrow[P]{} 0$ and $\frac{s_n^2}{\sigma^2} \xrightarrow[P]{} 1$

and thus by another application of Theorem 19 $\frac{s_n}{\sigma} \xrightarrow[P]{} 1$

Note: The t_n in this case is Student's distribution if the x are independently and normally distributed -- however, this is for general t_n.

Re Theorem 19, see: Mann and Wald; Annals of Math. Stat., 1943; "On Stochastic Order Relationships" -- for extending ordinary limit properties to probability limit properties.

Misc. remarks: On the Taylor series remainder term as used in the proofs of theorems 14, 17, and on p. 37 -- see Cramer p. 122.

If f(x) is continuous and a derivative exists, then we can write

$$f(x) = f(a) + (x - a)\, f'\left[a + \theta\,(x - a)\right] \qquad 0 \leq \theta \leq 1$$

$$= f(a) + (x - a)\left[f'(a) + f'(a + \theta(x - a)) - f'(a)\right]$$

$$= f(a) + (x - a)f'(a) + R$$

where $R = (x - a)\left[f'(a + \theta(x - a)) - f'(a)\right]$

$$\frac{R}{x - a} = f'\left[a + \Theta(x - a)\right] - f'(a)$$

$$f'\left[a + \Theta(x - a)\right] - f'(a) \longrightarrow 0$$

then if f' is continuous, as $x \longrightarrow a$

$$\lim_{x \to a} \frac{R}{x - a} = 0$$ i.e., R converges to 0 faster than $x - a$

<u>Remark</u>: If there exists an A such that $\int_{-\infty}^{A} g(x)\, dF_n(x) < \varepsilon$ and $\int_{A}^{\infty} g(x) dF_n(x) < \varepsilon$

for $n = 1, 2, 3, \ldots$ then if $F_n \xrightarrow[P]{} F$ Ref. Cramer, p. 74

$$\int_{-\infty}^{\infty} g(x) dF_n(x) \xrightarrow[P]{} \int_{-\infty}^{\infty} g(x) dF(x)$$

<u>Question</u>: Under what conditions does $E(t_n) = 0$ for all n

or does $E(t_n) \longrightarrow 0$ as $n \longrightarrow \infty$?

(t_n is defined as in the example illustrating theorem 18)

Counter-examples:

Define $p_n = \frac{1}{\sqrt{2\pi}} \int_{1-\frac{1}{n}}^{1} e^{-x^2/2}$

X_n is normal (0,1) except on the interval $(1 - \frac{1}{n}, 1)$

and $X_n = 1$ with probability p_n.

Then with probability $(p_n)^n$ $X_1, X_2, \ldots, X_n = 1$

in which case $t_n \quad \frac{\sqrt{n}(1 - \mu)}{0} = \infty$

therefore $E(t_n)$ is not defined.

<u>Problem 24</u>:

X is Poisson λ, then $Y = \frac{(X - \lambda)^2}{X}$

is asymptotically $\chi^2_{(1)}$ as $\lambda \longrightarrow \infty$

Convergence Almost Everywhere:

Def. 19: A sequence of r.v., $X_1, X_2, \ldots \rightarrow X$ a.e. if given ε, δ there exists as N such that

$$\Pr\left[\,|X_j - X| < \varepsilon \qquad j = N, N+1, N+2, \ldots\ldots\right] \geq 1 - \delta$$

Ref.: Feller, Chapter 9.

Note: Convergence almost everywhere is sometimes called "strong convergence". Convergence in probability is sometimes called "weak convergence".

Example: (of a case when $X_n \xrightarrow[P]{} c$ but $X_n \not\rightarrow c$ a.e.)

$X_n = 0$ with probability $1 - \frac{1}{n}$

$X_n = 1$ with probability $\frac{1}{n}$

the X's are independent.

(1) To show $X_n \xrightarrow[P]{} 0$

$$\Pr\left[\,|X_n - 0| > \varepsilon\right] = \frac{1}{n} \text{ for any } \varepsilon < 1$$

and $\frac{1}{n}$ can be made arbitrarily small by increasing n.

(2) $X_n \not\rightarrow 0$ a.e.

$$\Pr\left[\,|X_n - 0| < \varepsilon \quad n = N, N+1, N+2, \ldots\right]$$

$$= \prod_{n=N}^{\infty} \Pr\left[X_n < \varepsilon\right] = \prod_{j=0}^{\infty}\left(1 - \frac{1}{N+j}\right)$$

Note: $1 - x < e^{-x} \qquad 0 \leq x \leq 1$

$$\leq \prod_{j=0}^{\infty} e^{-\frac{1}{N+j}} = \exp\left\{-\sum_{j=0}^{\infty} \frac{1}{N+j}\right\}$$ which series is divergent

$$\leq e^{-\infty} = 0$$

therefore $Pr\left[|X_n - 0| < \varepsilon \quad n = N, N+1, N+2, \ldots\ldots \right] = 0$

therefore $X_n \not\rightarrow 0$ a.e.

Problem 25:

If $X_n = 0$ with probability $1 - \frac{1}{2^n}$

$X_n = 1$ with probability $\frac{1}{2^n}$

then $X_n \rightarrow 0$ a.e.

Problem 26:

$\phi_X(t) = \cos ta$

1. What is the d.f. of X?
2. Is $\bar{X}$ A. N. as $n \rightarrow \infty$ (suitably normalized) ?
3. To waht does $\bar{X}$ converge as $a \rightarrow 0$?

Problem 27:

X_i and Y_i are independent and identically distributed random variables with means μ, ν and variances σ_1^2, σ_2^2. Find and prove the asymptotic d.f. of $\bar{X}_n \bar{Y}_n$ (suitably normalized).

Kolmogorov inequality: Let $\{X_i\}$ be a sequence of r.v. with means μ_i and variances σ_i^2.

Then

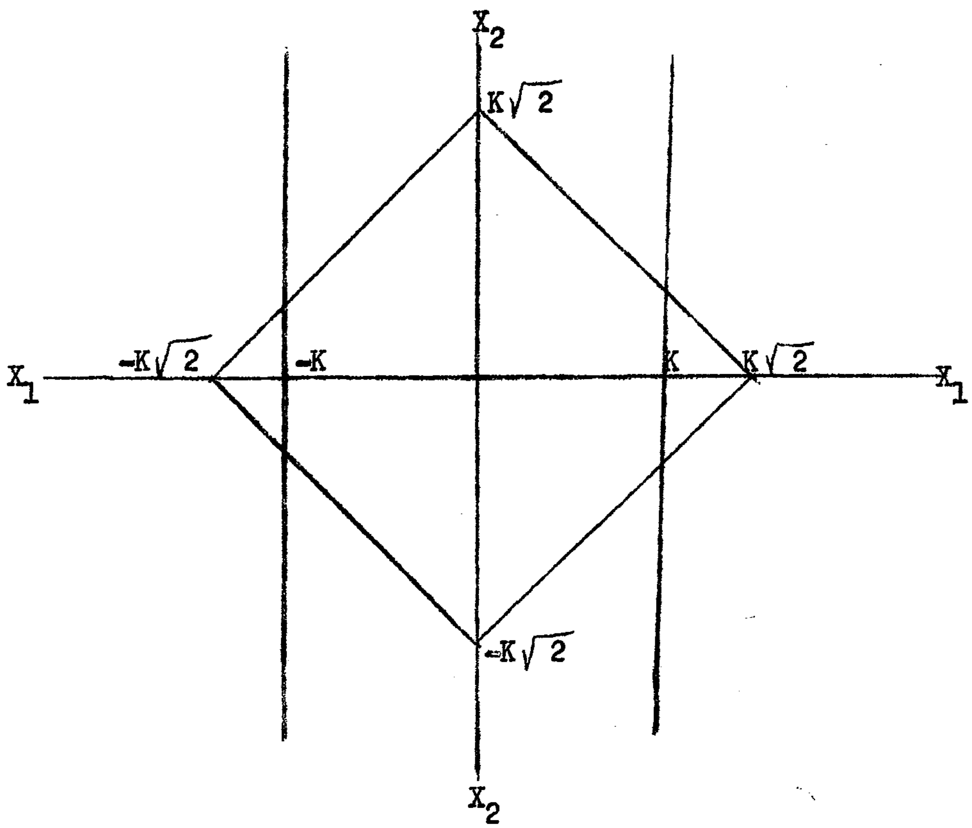
$$Pr\left[\frac{\left| \sum_1^n X_i - \sum_1^n \mu_i \right|}{\left(\sum_1^n \sigma_i^2 \right)^{1/2}} < K \qquad K = 1, 2, \ldots, m \right] > 1 - \frac{1}{K^2}$$

Example:

$X_1 \quad 0 \quad 1$

$X_2 \quad 0 \quad 1$

$Pr\left[|X_1| < K, \quad |X_1 + X_2| < \sqrt{2}K \right] > 1 - \frac{1}{K^2}$



Theorem 20:

If X_1, X_2, X_3, ... are independent random variables with means μ_i and variances σ_i^2, then, if $\sum_1^\infty \sigma_i^2/i^2$ converges

$$\left[\overline{X}_n - \overline{\mu}_n\right] \rightarrow 0 \text{ a.e.}$$

or we say that $\overline{X}$ obeys the strong law of large numbers

where $\overline{\mu} = \frac{1}{n}\sum_1^n \mu_i$

Proof of Theorem 20:

Let A_j be the event that for some n in the interval $2^{j-1} < n < 2^j$ $\left|\overline{X}_n - \mu_n\right| > \varepsilon$ (violating the definition of convergence a.e.)

$$\Pr\left[A_j\right] = \Pr\left[\left|\bar{X}_n - \mu_n\right| > \varepsilon \quad \text{for some } n\right]$$

$$= \Pr\left[\left|\sum_1^n (\bar{X}_n - \mu_n)\right| > n\varepsilon\right]$$

$$\leq \Pr\left[\left|\sum (X_n - \mu_n)\right| > 2^{j-1}\varepsilon\right] \quad \text{inserting the lower bound on } n$$

$$\leq \Pr\left[\frac{\left|\sum (X_n - \mu_n)\right|}{(\sum \sigma_i^2)^{1/2}} > \frac{2^{j-1}\varepsilon}{(\sum \sigma_i^2)^{1/2}}\right]$$

from the Kolmogorov inequality:

$$\Pr\left[\frac{X_1 - \mu_1}{\sigma_1} < K; \quad \frac{(X_1 + X_2) - (\mu_1 + \mu_2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} < K; \quad \ldots\ldots; \quad \frac{\left|\sum (X_i - \mu_i)\right|}{(\sum \sigma_i^2)^{1/2}} < K\right] > 1 - \frac{1}{K^2}$$

hence for the event A

$$\Pr\left[A_j\right] \leq \frac{\sum_1^{2^j} \sigma_i^2}{2^{2(j-1)}\varepsilon^2} \qquad \text{letting } k = \frac{2^{j-1}\varepsilon}{(\sum \sigma_i^2)^{1/2}}$$

and inserting the upper bound of n in the summation

$$\leq \frac{4\sum^{2^j} \sigma_i^2}{2^{2j}\,\varepsilon^2}$$

$$\sum_{j=1}^{\infty} \Pr\left[A_j\right] \leq \frac{4}{\varepsilon^2} \sum_{j=1}^{\infty} \frac{1}{2^{2j}} \sum_{i=1}^{2^j} \sigma_i^2$$

Interchanging the order of summation

$$\leq \frac{4}{\varepsilon^2} \sum_{i=1}^{\infty} \sigma_i^2 \sum_{\substack{j \\ 2j > i}} \frac{1}{2^{2j}}$$

Note that $\displaystyle\sum_{\substack{j \\ 2j > i}} \frac{1}{2^{2j}} = \frac{\frac{1}{i^2}}{1 - \frac{1}{2^2}}$ since it is a geometric series.

$$\leq \frac{4}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} \cdot \frac{4}{3} = \frac{16}{3\varepsilon^2} \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}$$

Now this sum is finite by hypothesis — hence $\sum_j \Pr\left[A_j\right]$ converges and we can choose N so that

$$\sum_{j=1}^{\infty} \Pr\left[A_j\right] < \delta$$

hence definition 19 for convergence a.e. is satisfied.

Corollary: (to theorem 20)

If X_i are independent and identically distributed $(i = 1,2,3,\ldots..)$ with mean μ and variance σ^2 then

$$\bar{X} \longrightarrow \mu \quad \text{a.e.}$$

Proof: is immediate since $\sum \frac{\sigma_i^2}{i^2} = \sigma^2 \sum \frac{1}{i^2}$ which converges, i.e. $< \infty$.

Other Types of Convergence:

1. Convergence in the mean

$$\text{l.i.m.} \quad X_n = X \quad \text{if} \lim_{n \to \infty} E\left[X_n - X\right]^2 \longrightarrow 0$$

Note: l.i.m. = limit in the mean

Implies convergence in probability but not convergence a.e.

Ref: Cramer — Annals of Math. Stat. 1947.

2. Law of the iterated logarithm

Ref.: Feller, Chapter 8

3. St. Petersburg paradox

$X = 2^n$ with probability $\frac{1}{2^n}$ $n = 1, 2, 3, \ldots,$ Note: $\sum_1^\infty \frac{1}{2^n} = 1$

e.g. Toss a coin until heads comes up -- count the total number of tosses required (= n) -- bank pays 2^n to the player.

$$E(x) = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum_1^{\infty} 1 = \infty$$

for a fair game, the entry fee should be equal to the expected gain; therefore, this game presents a problem in the determinat[ion] of the "fair" entry fee.

Ref.: Feller, p. 235-7 -- he shows

$$Pr\left[\left|\frac{\sum_1^n X_i}{n \ln n} - 1\right| > \varepsilon\right] < \delta$$

that is, the game "becomes fair" if the entry fee is n ln n.

CHAPTER IV

ESTIMATION (Point):

Ref: E. L. Lehman, "Notes on the Theory of Estimation", U. of Cal. Bookstore
Cramer -- ch. 32-3

$F(x, \theta)$ -- a family of distributions

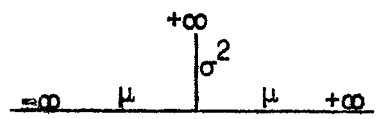
X, θ may be vector valued, in which case they will be denoted $\underline{X}, \underline{\theta}$

$X \varepsilon R_k$

$\theta \; \varepsilon \Omega$ -- parameter space

Example: 1: if $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2) then

Ω consists of all possible μ and all positive σ^2



2: $F(x)$ $F \varepsilon \mathcal{F}$ the family of all continuous distributions

Ω is then the space of all continuous d.f.

-- this is the non-parametric case
-- might wish to estimate $E_F(x) = \int_{-\infty}^{\infty} x \, dF(x)$

provided that we add the restriction that $E_F(X) < \infty$

Estimate of $g(\theta)$ is some function of $\underline{X}$ from R_k to Ω which in some sense comes close to $g(\theta)$

Or in the "general decision theory" point of view (Wald)

an estimate of $g(\theta)$ is a decision function $d(\underline{X})$ and we have associated with each decision function a loss function $W\left[d(\underline{X}), \theta\right]$ with $W = 0$ whenever $d(\underline{X}) = g(\theta)$

The choice of loss function is arbitrary, but we frequently choose

$$W\left[d(\underline{X}), \theta\right] = \left[d(\underline{X}) - g(\theta)\right]^2$$

Def. 20: Risk Function is defined as

$$R(d, \underline{\theta}) = E\left\{W\left[d(\underline{X}), \underline{\theta}\right]\right\} = \int_{-\infty}^{\infty} W\left[d(\underline{X}), \underline{\theta}\right] dF(\underline{x}, \underline{\theta})$$

Example: $X_1, X_2, \ldots, X_n$ are NID(μ, 1)

$$d^*(X) = \bar{X}$$

$$R(d^*, \mu) = E(\bar{X} - \mu)^2 = \frac{1}{n}$$

A "best" estimate might be defined as one which minimizes $R(d, \theta)$ with respect to d uniformly in θ.

$R(d, \theta) \leq R(d^*, \theta)$ for all θ with d^* any other estimator.

consider the estimate d(x) of $g(\theta)$ defined as $d(x) = g(\theta_0)$ $R(d, \theta_0) = 0$

Hence a uniformly (supposedly) minimum risk estimate can be found only if there exists a d(x) such that $R\left[d(x), \theta\right] = 0$

An example would be similar to asking which is better for estimating time -- a stopped clock which is right twice a day, or one that runs five minutes slow.

Since a uniformly best estimate is virtually impossible to find, we want to consider alternative

WAYS to formulate the problem of obtaining best estimates:

I. by restricting the class of estimates
 1. unbiased estimates

Def. 21: $d(\underline{X})$ is unbiased if $E\left[d(\underline{X})\right] = g(\theta)$

$d(\underline{X})$ is a minimum variance unbiased estimate (m.v.u.e.) if $E\left[d(\underline{X}) - g(\theta)\right]^2$ is minimized over unbiased estimates d

 2. invariant estimates

Let h(X) be the transformation of a real line into itself which induces a transformation h on parameter space. If $d\left[h(X)\right] = \bar{h}\left[d(X)\right]$ then d(X) is invariant under this transformation.

Example: family of d.f. with $E(X) < \infty$

Problem is to estimate E(X)

$h(x) = ax + b$ $\bar{h}\left[E(X)\right] = a\,E(X) + b$

An estimate d(X) of μ is invariant if $d\left[h(X)\right] = \bar{h}\left[d(X)\right]$

$d(aX + b) = a\,d(X) + b$

Therefore $\bar{X}$ is an invariant estimate of μ under this transformation.

Note that $d(X) = \mu_0$ is not invariant.

3. Best linear unbiased estimates (b.l.u.e.)

Def. 22: Estimates of $g(\Theta)$ which are unbiased, linear in the X_i and which among such estimates have minimum variance are b.l.u.e.

Problem 28: $X_1, X_2, \ldots, X_n$ are independent random variables with mean μ and and variance σ^2

show that $\bar{X}$ is the b.l.u.e. of μ.

Problem 29: $X_1, X_2, \ldots, X_n$ are NID (μ, σ^2)

$$W[d(X), \Theta] = b[d(X) - \mu] \text{ if } d(X) > \mu$$
$$= c[d(X) - \mu] \text{ if } d(X) < \mu$$

a. $d(X) = \bar{X}$ find $R(\bar{X}, \mu, \sigma)$

(note: the answer depends on the loss function constants only, not μ)

b. $d(X) = \bar{X} + a$ -- show how to determine a such that $R(d, \mu, \sigma)$ is minimized

[note: the answer involves $\phi(z)$ which is defined as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt \Big]$$

Comment on this problem:

An orthogonal transformation

-- is a rotation or reflection

-- $\underline{y} = A\underline{x}$ where A is orthogonal

-- $|J| = 1$

-- $\sum y_i^2 = \sum x_i^2$

-- For $y_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n$; $\sum_{j=1}^{n} a_{ij}^2 = 1$; $\sum_j a_{ij}a_{kj} = 0 \quad i \neq k$

In general if $d(\underline{X})$ is a function of $T(X_1, X_2, \ldots, X_n)$

$$R(d, \Theta) = E[W(d, \Theta)]$$
$$= \int \cdots \int W[T(x), \Theta]\, dF(\underline{x})$$

Making the transformation $y = T(x)$

$$\left.\begin{array}{l} y = \\ \cdot \\ \cdot \\ \cdot \\ y = \end{array}\right\} \text{n-1 functions independent of the first one}$$

Then

$$R\left[d(T), \Theta\right] = \int W\left[T(x), \Theta\right] \underbrace{dF_1(y_1)}_{\substack{\text{marginal distribution} \\ \text{of } y_1 = T(x)}} \underbrace{\int \cdots \int dF(y_2, y_3, \ldots, y_n)}_{\substack{\text{conditional distribution} \\ \text{of } y_2, \ldots, y_n \\ \text{given } y_1 = 1}}$$

$$= \int W\left[y_1, \Theta\right] dF(y_1)$$

$$= \int W\left[T(x), \Theta\right] dF\left[T(x)\right]$$

II. Optimum Properties in the Large

1. Bayes Estimates

Def. 23: If Θ has a known "a priori" distribution $H(\Theta)$ then the Bayes estimate of $g(\Theta)$ is that $d(\Theta)$ which minimizes

$$\int_{\Omega} R(d, \Theta)\, dH(\Theta) \quad \text{with respect to } d(x)$$

Example: X is $B(n, p)$ and p is uniform on (0, 1)

Let $W\left[d(X), \Theta\right] = \left[d(X) - p\right]^2$ and minimize this with respect to d

$$R(d, \Theta) = \sum_{x=0}^{n} \left[d(x) - p\right]^2 \binom{n}{x} p^x (1 - p)^{n-x}$$

Average risk = risk function averaged with respect to p

$$= \int_0^1 R(d, \Theta)\, dp$$

$$= \sum_{x=0}^{n} \int_0^1 \left[d^2(x) - 2pd(x) + p^2\right] \frac{n!}{x!(n-x)!} p^x(1, - p)^{n-x}\, dp$$

$$\text{Note: } \int_0^1 p^a(1-p)^b\, dp = \frac{a!\, b!}{(a+b+1)!}$$

Using this evaluation on each part separately we have

$$= \sum_{x=0}^{n} \left[d^2(x)\frac{n!}{x!(n-x)!} \cdot \frac{x!(n-x)!}{(n+1)!} - \frac{2d(x)n!}{(x)!(n-x)!} \cdot \frac{(x+1)!(n-x)!}{(n+2)!} + \frac{n!}{x!(n-x)!} \circ \frac{(x+2)!(n-x)!}{(n+3)!} \right]$$

$$= \sum_{x=0}^{n} \left[\frac{d^2(x)}{n+1} - \frac{2d(x)(x+1)}{(n+1)(n+2)} + \frac{(x+1)(x+2)}{(n+1)(n+2)(n+3)} \right]$$

$$= \frac{1}{n+1} \sum_{x=0}^{n} \left[d^2(x) - \frac{2d(x)(x+1)}{(n+2)} + \left(\frac{x+1}{n+2}\right)^2 + \frac{(x+1)(x+2)}{(n+2)(n+3)} - \left(\frac{x+1}{n+2}\right)^2 \right]$$

$$= \frac{1}{n+1} \sum_{x=0}^{n} \left[\left\{ d(x) - \frac{x+1}{n+2} \right\}^2 + \frac{x+1}{n+2} \left\{ \frac{n+1-x}{(n+2)(n+3)} \right\} \right]$$

This is certainly minimized with respect to d(x) if each term in the first summation is zero -- i.e. if

$$d(X) = \frac{X+1}{n+2}$$

<u>Problem 30</u>: if $d(X) = \frac{X}{n}$ find R(d, p) as a function of p and also

$$\text{average } R = \int_0^1 R(d, p)\, dp$$

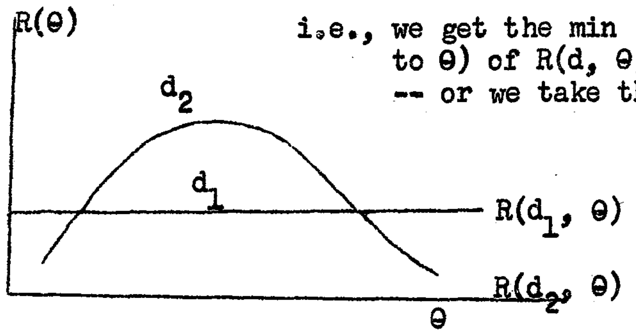
— if p is uniformly distributed on (0, 1)

— $R(d, p) = E\left[d(X) - p\right]^2$

2. Minimax Estimate

Def. 24: $d(X)$ is a minimax estimate if $d(X)$ minimizes $\sup_\theta R(d, \theta)$ in comparison to any other estimate $d^*(X)$

i.e., we get the min (with respect to d) of the max (with respect to θ) of $R(d, \theta)$
-- or we take the inf (d) sup (θ) of $R(d, \theta)$



$d_1(x)$ is minimax estimate since it has a minimum maximum point

3. Constant Risk Estimates

Def. 25: A constant risk estimate is one for which $R(d, \theta)$ is constant with respect to θ

Problem 31: Find a constant risk estimate among linear estimates of p if X is $B(n,p)$

III -- By dealing only with large sample (asymptotic) properties of estimates

Def. 26: Consistent Estimates -- $d_n(\underline{X})$ is consistent if:

$$d_n(\underline{X}) \xrightarrow[p]{} \theta$$

(it does not necessarily follow that $E(d_n) \longrightarrow \theta$ or that $\sigma^2(d_n) \longrightarrow 0$)

Problem 32: If $E(d_n) \longrightarrow \theta$ and $\sigma^2(d_n) \longrightarrow 0$ then $d_n(X)$ is consistent.

(these are the sufficient conditions for consistency)

Def. 27: Best Asymptotically Normal Estimates (B.A.N.E.) -- $d(X)$ is a B.A.N. estimate if:

1. $\dfrac{d_n - E(d_n)}{\sigma(d_n)}$ is A. $N(0, 1)$

2. if d_n^* is any other A. N. estimate, then

$$\lim_{n \to \infty} \frac{\sigma^2(d_n)}{\sigma^2(d_n)} \leq 1$$

METHODS OF ESTIMATION

A -- Methods of Moments (K. Pearson)

$\underline{\Theta} = \Theta_1, \Theta_2, \ldots, \Theta_k$ -- equate the first k sample moments to k population moments (expressed as functions of $\Theta_1, \Theta_2, \ldots, \Theta_k$). Solve these equations for $\Theta_1, \Theta_2, \ldots, \Theta_k$ and these are the moment estimates.

example: $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2)

first population moment = μ

second population moment = σ^2

then $\bar{X} = \tilde{\mu}$; $\dfrac{\sum_1^n (X_i - \bar{X})^2}{n} = \tilde{\sigma}^2$ ("n" being the divisor used by K. Pearson)

note: This method yields poor estimates in many cases -- has very few optimum properties

B -- Method of Least Squares (Gauss -- Markov)

let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a r.v. with $E(\underline{X}) = A\underline{\Theta}$ $\quad \underline{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_s)$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{ns} \end{pmatrix} \qquad s \leq n \qquad A \text{ is of rank } s$$

i.e., both $\underline{X}$ and $\underline{\Theta}$ are column vectors

or we have $E(X_i) = \sum_{j=1}^{s} a_{ij}\Theta_j \qquad i = 1, 2, \ldots, n$

also $\sigma^2_{(X_i, X_j)} = \delta_{ij}\sigma^2$ i.e., the covariances = 0

Def. 28: $\underline{\Theta}^*$ is a least squares estimate of $\underline{\Theta}$ if $\underline{\Theta}^*$ minimizes

$$(\underline{X} - A\underline{\Theta})' (\underline{X} - A\underline{\Theta}) = \sum_{i=1}^{n} (X_i - a_{i1}\Theta_1 - \cdots - a_{is}\Theta_s)^2$$

Theorem 21: (Gauss -- Markov)

With the given conditions on $X_1, X_2, \ldots, X_n$ the least squares estimate $\underline{\Theta}^*$ is a best linear unbiased estimate (b.l.u.e.) of $\underline{\Theta}$.

Proof Theorem 21: ref: Placket: Biometrika, 1949, p. 458
Lehman

we first show that $\underline{\theta}^* = C^{-1} A' \underline{X}$ where $C = A'A$ $C' = C$ since it is symetric

if we write $\underline{\theta} = C^{-1} A' \underline{X} + \underline{y}$ then we are trying to minimize

with respect to y, $\Phi = (\underline{X} - A\,\underline{\theta})'(\underline{X} - A\,\underline{\theta})$

$$= (\underline{X} - AC^{-1} A' \underline{X} - A\underline{y})' (\underline{X} - AC^{-1} A'\underline{X} - A\underline{y})$$

$$= (\underline{X} - AC^{-1} A' \underline{X})' (\underline{X} - AC^{-1} A'\underline{X}) - (\underline{X} - AC^{-1} A'\underline{X})' A\underline{y}$$

$$- (A\underline{y})' (\underline{X} - AC^{-1} A'\underline{X}) + (A\underline{y})'(A\underline{y})$$

now the cross-product terms equal zero

e.g. $-\underline{X}'A\underline{y} + \underline{X}'A(C^{-1})'A'A\underline{y} = -\underline{X}'A\underline{y} + \underline{X}'A\underline{y} = 0$

since $(C^{-1})' = C^{-1}$ $C = A'A$

$$(C^{-1})'A'A = C^{-1}C = I$$

similarly the second cross-product also = 0

hence Φ is minimized with respect to y if $(A\underline{y})'(A\underline{y})$ is minimized which will happen if $A\underline{y} = 0$ since A has maximum rank s this will happen only if y=0

or writing $\Phi = \sum_{i=1}^{n} (X_i - a_{i1}\theta_1 - \ldots - a_{ij}\theta_j - \ldots - a_{is}\theta_s)^2$

formally minimizing Φ in the usual fashion by differentiating with respect to the θ_j

$$\frac{\partial \Phi}{\partial \theta_j} = -2\sum_{i=1}^{n} a_{ij}(X_i - a_{i1}\,\theta_1 - \ldots - a_{ij}\theta_j - \ldots - a_{is}\theta_s) = 0$$

$j = 1, 2, \ldots, s$

solving these equations

$$\sum_{i=1}^{n} a_{ij}x_j = (\sum_{i=1}^{n} a_{ij}a_{i1})\,\theta_1 + \ldots + (\sum_{i=1}^{n} a_{ij}a_{is})\,\theta_s$$

or $A'\underline{X} = (A'A)\,\underline{\theta}$

To show that $\underline{\theta}^*$ so defined (i.e., $= C^{-1} A'\underline{X}$) is B.L.U.E.

consider a linear estimate $B_{s\times n}\underline{X}$

$E[B\underline{X}] = \underline{\theta} \qquad B\,E[\underline{X}] = \underline{\theta} \qquad B A\underline{\theta} = \underline{\theta}$ thus $BA = I$

note that $\sigma^2 BB'$ is the covariance matrix of $B\underline{X}$ -- the elements in the diagonal are the variances of the estimates $\underline{\theta}$ -- we thus wish to minimize these diagonal elements (by proper choice of B) subject to the restriction that $BA = I$

$$(B - C^{-1}A')(B - C^{-1}A')' = BB' - B(A')'(C^{-1})' - C^{-1}A'B' + C^{-1}A'(A')'(C^{-1})'$$

using the relationships that $BA = I \quad A'A = C \quad C' = C$ or $(C^{-1})' = C^{-1}$

$$(B-C^{-1}A')(B-C^{-1}A')' = BB' - (C^{-1})' - C^{-1} + C^{-1} = BB' - C^{-1}$$

thus $BB' = C^{-1} + (B - C^{-1}A')(B - C^{-1}A')'$

minimization will occur if $(B - C^{-1}A') = 0$ or if $B = C^{-1}A' \quad (\underline{\theta} = C^{-1}A'\underline{X})$

hence $\underline{\theta}^*$ are B.L.U.E.

example: if $X_1, X_2, \ldots, X_n$ are uncorrelated with $E(X_i) = \mu$ and common variance σ^2 then the least squares estimate of μ is $\bar{X}$

let $A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ here $s = 1$ (we have a 1xn matrix of 1's)

$C = A'A = n \qquad C^{-1} = \frac{1}{n}$

$$\mu^* = C^{-1}A'\underline{X} = \frac{1}{n}(1, 1, \ldots, 1)\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Problem 33: $E(X_i) = \alpha + \beta t_i \qquad i = 1, 2, \ldots, n$

t_i are known constants and assume $\sum_1^n t_i = 0$

find b.l.u.e. of α, β using theorem 21.

Aiken extended this result in 1934 to the case where the X_i are correlated and we know the correlation matrix V (up to an arbitrary multiplier) -- b.l.u.e. are also least squares estimates which are obtained by minimizing

$$(\underline{X} - A'\underline{\theta})' V^{-1} (\underline{X} - A'\underline{\theta})$$

ref: Plackett: Biometrika, 1949, p.45

C. Maximum Likelihood Estimates (ref. Cramer ch. 33)

Def. 29: Likelihood function

$$L = f(X_1, X_2, \ldots, X_n, \underline{\theta}) \text{ if the } X\text{'s are continuous}$$

$$= p(X_1, X_2, \ldots, X_n, \underline{\theta}) \text{ where the } p\text{'s are discrete probabilities if the } X\text{'s are discrete}$$

if the X_i are continuous, independent, and identically distributed

$$L = \prod_1^n f(X_i, \underline{\theta}) \quad \text{or} \quad \ln L = \sum_1^n \ln f(X_i, \underline{\theta})$$

Def. 30: A Maximum Likelihood Estimate (M.L.E.) is that value of $\underline{\theta}$ (denoted $\hat{\underline{\theta}}$) which maximizes the function L (or $\ln L$)

It may happen (from the third case in def. 29) that $\hat{\underline{\theta}}$ is the solution of the set of equations

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \qquad i = 1, 2, \ldots, s \qquad \underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$$

Regularity conditions needed in the maximum likelihood derivations (ref: Cramer 500-504)

1. The X_1 are continuous, independent, and identically distributed.
   We will assume first that θ is a scalar.
2. $\dfrac{\partial \ln f(X_i, \theta)}{\partial \theta}$ is a function of X_i and hence is a random variable.

   we assume that $\dfrac{\partial^k \ln f(X_i, \theta)}{\partial \theta^k}$ exists for $k = 1, 2, 3$
3. Ω is an interval and θ_o (the true value of θ) is an interior point
4. $\dfrac{\partial^k \ln f(X_i, \theta)}{\partial \theta^k} < F_k(X_i)$ which is integrable over $(-\infty, \infty)$

$E\left[F_3(X_i)\right] < M$ for all θ -- i.e., it is bounded

5. $E\left[\frac{\partial \ln f_i}{\partial \theta}\right]^2 = \int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2 f(x_i, \theta)\, dx_i = k^2 \qquad 0 < k^2 < \infty$

__Theorem 22:__ If f, Ω satisfy the regularity conditions, and if L, or ln L, has a unique maximum, then

1- the maximum likelihood estimate $\hat{\theta}$ is the solution of the equation

$$\frac{\partial \ln L}{\partial \theta} = 0$$

2- $\hat{\theta}$ is consistent

3- $\sqrt{n}\,(\hat{\theta} - \theta_0)$ is asymptotically normal with mean 0 and variance $\dfrac{1}{E\left[\frac{\partial \ln f_i}{\partial \theta}\right]^2}$

Proof: 1- since $\frac{\partial \ln L}{\partial \theta}$ is continuous with a continuous derivative, if

$$\ln L\left(= \sum_1^n \ln f(X_i, \theta)\right) \text{ has a maximum, } \frac{\partial \ln L}{\partial \theta} = 0 \text{ at this max.}$$

2-- to show that $\hat{\theta}$ is consistent

$$f_i = f(X_i, \theta)$$

$$\frac{\partial \ln f_i}{\partial \theta} = \left.\frac{\partial \ln f_i}{\partial \theta}\right]_{\theta=\theta_0} + \left.\frac{\partial^2 \ln f_i}{\partial \theta^2}\right]_{\theta=\theta_0}(\theta - \theta_0) + \left.\frac{\partial^3 \ln f_i}{\partial \theta^3}\right]_{\theta=\theta_0+\varepsilon(\theta-\theta_0)}\frac{(\theta-\theta_0)^2}{2}$$

summing each term on both sides of the equality, dividing by n, and doing some substituting

$$\frac{1}{n}\frac{\partial \ln L}{\partial \theta} = \frac{1}{n}\sum_{i=1}^n \left.\frac{\partial \ln f_i}{\partial \theta}\right]_{\theta=\theta_0} + \frac{1}{n}\sum_{i=1}^n \left\{\left.\frac{\partial^2 \ln f_i}{\partial \theta^2}\right]_{\theta=\theta_0}(\theta-\theta_0)\right\} + \frac{z}{n}\sum_{i=1}^n F_3(X_i)\,\frac{(\theta-\theta_0)^2}{2}$$

the term in the third derivative is replaced by the term in regularity condition 4a -- and multiplied by the factor $z\,(0 \le z \le 1)$ to restore the equality

this equation can be written:

$$\frac{1}{n}\frac{\partial \ln L}{\partial \theta} = B_0 + B_1(\theta - \theta_0) + z\,B_2\frac{(\theta-\theta_0)^2}{2}$$

---

note that we have: $$\int_{-\infty}^{\infty} f(x_i, \Theta)\, dx_i = 1 \qquad [1]$$

from which we can show:

$$\frac{\partial}{\partial\Theta}\int f(x_i, \Theta)\, dx_i = 0 \iff \int \frac{\partial f(x_i, \Theta)}{\partial \Theta}\, dx_i = 0 \iff$$

$$\int\left(\frac{1}{f(x_i, \Theta)}\frac{\partial f_i}{\partial\Theta}\right) f(x_i, \Theta)\, dx_i = 0 \iff \int\left(\frac{\partial \ln f_i}{\partial\Theta}\right) f(x_i, \Theta)\, dx_i = 0$$

therefore: $$E\left[\frac{\partial \ln f_i}{\partial \Theta}\right] = 0 \qquad [4]$$

also, differentiating a second time we have:

$$\frac{\partial^2}{\partial\Theta^2}\int_{-\infty}^{\infty} f(x_i, \Theta)\, dx_i = 0 \qquad [2]$$

or $$\int \frac{\partial^2 f(x_i, \Theta)}{\partial \Theta^2} = 0$$

$$E\left[\frac{\partial^2 \ln f_i}{\partial \Theta^2}\right] = \int_{-\infty}^{\infty} \frac{\partial^2 \ln f_i}{\partial\Theta^2} f(x_i, \Theta)\, dx_i = \int_{-\infty}^{\infty}\frac{\partial}{\partial\Theta}\left[\frac{1}{f(x_i, \Theta)}\frac{\partial f_i}{\partial\Theta}\right] f(x_i,\Theta)\, dx_i$$

$$= \int_{-\infty}^{\infty}\left\{\underbrace{\frac{1}{f(x_i, \Theta)}\frac{\partial^2 f_i}{\partial\Theta^2} f(x_i, \Theta)}_{= 0 \text{ by } [2] \text{ above}} - \frac{1}{f_i^2}\left(\frac{\partial f_i}{\partial\Theta}\right)^2 f(x_i, \Theta)\right\} dx_i$$

$$= -\int_{-\infty}^{\infty}\left(\frac{1}{f_i}\frac{\partial f_i}{\partial\Theta}\right)^2 f_i\, dx_i = -\int_{-\infty}^{\infty}\left(\frac{\partial \ln f_i}{\partial \Theta}\right)^2 f_i\, dx_i$$

$$= -E\left[\frac{\partial \ln f_i}{\partial \Theta}\right]^2 = -k^2$$

$$\text{or } -E\left[\frac{\partial^2 \ln f(x_i, \theta)}{\partial \theta^2}\right] = E\left[\frac{\partial \ln f(x_i, \theta)}{\partial \theta}\right]^2 = k^2 \qquad [3]$$

Thus, by Khintchine's theorem (no. 17)

$$B_0 \xrightarrow[p]{} 0$$

$$B_1 \xrightarrow[p]{} -k^2$$

$$B_2 \xrightarrow[p]{} E\left[F_3(x_i)\right] < M$$

Let S = the set of points where

$$|B_0| < \delta^2 \qquad B_1 \leq \frac{1}{2} k^2 \qquad |B_2| < 2M .$$

We can find an n, given ε, δ, such that $\Pr[S] > 1 - \varepsilon$.

In S the right hand side (r.h.s.) of the Taylor series expansion is to be considered.

Consider: $\theta = \theta_0 + \delta$

$$\text{r.h.s.} = B_0 + B_1\delta + \frac{1}{2}z\, B_2\, \delta^2$$

$$\leq \delta^2(1 + M) - \frac{1}{2} k^2 \delta \qquad \text{letting } z = 1$$

So that, if $\delta < \frac{k^2}{2(1+M)}$ the r.h.s. ≤ 0 .

Considering: $\theta = \theta_0 - \delta$

$$\text{r.h.s.} = B_0 - B_1\, \delta + \frac{1}{2}\, z\, B_2\delta^2$$

$$\geq -\delta^2 + \frac{1}{2} k^2\, \delta^2 - M\delta^2 = -(M + 1)\delta^2 + \frac{1}{2} k^2\delta$$

$$\geq 0 \qquad \text{for the same } \delta < \frac{k^2}{2(1+M)}$$

Hence, in S, which occurs with probability $> (1 - \varepsilon)$, $\frac{1}{n}\frac{\partial \ln L}{\partial \theta} = 0$ has a root in the interval $(\theta_0 - \delta, \theta_0 + \delta)$ and $\ln L$ has a maximum in the interval (at the root).

$\hat{\theta}$ is then the maximum likelihood estimate and the solution of the equation

$$\frac{1}{n}\frac{\partial \ln L}{\partial \theta} = B_o + B_1(\hat{\theta} - \theta_o) + \frac{1}{2} z B_2(\hat{\theta} - \theta_o)^2 = 0$$

which yields $\hat{\theta} - \theta_o = - \dfrac{B_o}{B_1 + \frac{z}{2} B_2(\hat{\theta} - \theta_o)}$

Multiplying both sides by $k\sqrt{n}$

$$(\hat{\theta} - \theta_o)\ k\sqrt{n} = - \frac{\dfrac{B_o\sqrt{n}}{k}}{\dfrac{B_1}{k^2} + \dfrac{z}{2} \cdot \dfrac{B_2}{k^2}(\hat{\theta} - \theta_o)}$$

We know that: $B_o = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \dfrac{\partial \ln f_i}{\partial \theta}\Bigg|_{\theta_o}$

$$E\left[B_o\right] = 0 \qquad V\left[B_o\right] = \frac{1}{n} E\left[\frac{\partial \ln f_i}{\partial \theta}\right]^2 = \frac{k^2}{n}$$

therefore $\dfrac{\sqrt{n}\,B_o}{k}$ is A.N.(0, 1)

$B_1 \xrightarrow[p]{} -k^2 \qquad B_2 < M$ (i.e., it is bounded)

$$\hat{\theta} - \theta_o \xrightarrow[p]{} 0$$

Thus, by the relationships we have just stated, and by use of theorem 18

$$k\sqrt{n}\,(\hat{\theta} - \theta_o) \text{ is A.N.}(0, 1)$$

or: $\sqrt{n}\,(\hat{\theta} - \theta_o)$ is A.N. $\left(0, \dfrac{1}{k^2} = \dfrac{1}{E\left[\dfrac{\partial \ln f_i}{\partial \theta}\right]^2}\right)$.

Example: $f(x) = a\,e^{-ax} \qquad x \geq 0$

Find the m.l.e. of a, and its asymptotic distribution.

$$L = a^n e^{-a \sum_1^n x_i}$$

$$\ln L = n \ln(a) - a \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} - \sum_{1}^{n} x_i = 0$$

$$\text{therefore } \hat{a} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

We can easily verify that $E(x) = \frac{1}{a}$; $E(\bar{x}) = \frac{1}{a}$ or $\tilde{a} = \frac{1}{\bar{x}}$ by the method of moments.

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{a^2} = \frac{n\, \partial^2 \ln f_i}{\partial a^2}$$

$$\therefore -\frac{\partial^2 \ln f_i}{\partial a^2} = \left(\frac{\partial \ln f_i}{\partial a}\right)^2 = \frac{1}{a^2}$$

$$\text{Variance} = \frac{1}{E\left[\frac{\partial \ln f_i}{\partial a}\right]^2} = \frac{1}{1/a^2} = a^2$$

Hence: $\sqrt{n}\,(\frac{1}{\bar{x}} - a)$ is A $N(0, a^2)$

or $\frac{\sqrt{n}\,(\frac{1}{\bar{x}} - a)}{a}$ is A $N(0, 1)$.

Problem 34: $f(x) = a^2 e^{-a^2 x} \qquad x > 0$

-- find the m.l.e. of a and its asymptotic distribution.

-- verify that the same result could be obtained by a Taylor Series expansion.

Problem 35: X_i is Poisson $\lambda \qquad (i = 1, 2, \ldots, n)$

-- find the m.l.e. of λ and its asymptotic distribution.

Problem 36: X is uniform on the interval $(0, a)$.

-- Find the m.l.e. of a [not by differentiating] $(\hat{a})$.

-- Correct it for bias and find the asymptotic distribution of the unbiased estimate $(\hat{a}')$.

-- Another unbiased estimate is $a^* = 2\bar{x}$.

Compare the actual variances of $\hat{a}'$, a^*.

Note: a^* is a moment estimate -- for another comparison of the method of moments with the method of maximum likelihood, see Cramer p. 505.

Note: The m.l.e. is invariant under single valued functional transformations i.e., the m.l.e. of $g(\theta)$ is $g(\hat{\theta})$.

Remark: If $d_n(X)$ is A $N(\mu, \sigma^2/n)$ and $g[d]$ is continuous with continuous first and second derivatives and the first derivative $\neq 0$ at $x=\mu$ then $\sqrt{n}\,[g(d_n) - g(\mu)]$ is A. $N(0, [g'(\mu)]^2\ \sigma^2)$.

Proof: by use of a Taylor series expansion:

$$g[d_n(x)] = g(\mu) + g'(\mu)\ [d_n - \mu] + g''(\mu)\ \frac{(d_n-\mu)^2}{2}$$

From which we can get

$$\frac{\sqrt{n}\,[g(d) - g(\mu)]}{\sigma} = g'(\mu)\underbrace{\frac{\sqrt{n}\,(d-\mu)}{\sigma}}_{AN(0,\ 1)} + \underbrace{\underbrace{\frac{\sqrt{n}\,g''(\mu)\ (d-\mu)}{2\,\sigma}}_{\text{bounded}}\ \underbrace{(d-\mu)}_{\xrightarrow[p]{} 0}}_{\xrightarrow[p]{} 0}$$

Hence by use of theorem 18

$$\frac{\sqrt{n}\,[g(d) - g(\mu)]}{\sigma\, g'(\mu)} \text{ is A } N(0,\ 1).$$

Multiparameter case:

Theorem 23: Under generalized regularity conditions of theorem 22, if $\underline{\theta}' = (\theta_1, \theta_2, \ldots, \theta_n)$, then for sufficiently large n and with probability $1 - \varepsilon$, the m.l.e. of $\underline{\theta}$ is given as the solution of the equations

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \qquad i = 1, 2, \ldots, s$$

and further: $\hat{\theta}_1^o, \hat{\theta}_2^o, \ldots, \hat{\theta}_s^o$ has asymptotically a joint normal distribution with means $\theta_1, \theta_2, \ldots, \theta_s$ and variance-covariance matrix V^{-1}

where
$$V = -n \begin{pmatrix} E\left\{\frac{\partial^2 \ln f}{\partial \theta_1^2}\right\} & E\left\{\frac{\partial^2 \ln f}{\partial \theta_1 \theta_2}\right\} & \ldots & E\left\{\frac{\partial^2 \ln f}{\partial \theta_1 \theta_s}\right\} \\ \vdots & \vdots & & \vdots \\ E\left\{\frac{\partial^2 \ln f}{\partial \theta_s \theta_1}\right\} & E\left\{\frac{\partial^2 \ln f}{\partial \theta_s \theta_2}\right\} & \ldots & E\left\{\frac{\partial^2 \ln f}{\partial \theta_s^2}\right\} \end{pmatrix}$$

This is the so-called <u>information</u> matrix used in multiparameter estimation.

Sketch of part of the proof:

$$0 = \frac{1}{n} \frac{\partial \ln L}{\partial \theta_i} = \frac{1}{n} \frac{\partial \ln L}{\partial \theta_i}\bigg|_{\underline{\theta}^o} + \frac{1}{n} \sum_{j=1}^{s} (\theta_j - \theta_j^o) \frac{\partial^2 \ln L}{\partial \theta_j \theta_i}\bigg|_{\underline{\theta}^o}$$

+ second and higher degree terms $i = 1, 2, \ldots, s$

(Note: $\ln L$ terms can be replaced by $\sum \ln f_i = n \ln f_i$ terms.)

From theorem 22: $\ln L = \sum_{1}^{n} \ln f(x_i, \underline{\theta})$

$$E\left(\frac{\partial \ln L}{\partial \theta_j}\right) = E\left(\sum \left[\frac{\partial \ln f_i}{\partial \theta_j}\right]\right) = 0$$

$$-E\left(\frac{\partial^2 \ln f_i}{\partial \theta_j^2}\right) = E\left(\frac{\partial \ln f_i}{\partial \theta_j}\right)^2$$

We also need a set of covariance terms:

$$-E\left(\frac{\partial^2 \ln f_i}{\partial \theta_j \theta_k}\right) = E\left[\left(\frac{\partial \ln f_i}{\partial \theta_j}\right)\left(\frac{\partial \ln f_i}{\partial \theta_k}\right)\right]$$

which follows in a manner similar to the derivation of the variance expression in theorem 22.

Now we can write:

$$\beta_{jo} = \frac{1}{n} \sum_{1}^{n} \frac{\partial \ln f(x_i, \underline{\theta})}{\partial \theta_j} \Bigg|_{\underline{\theta}^o} \xrightarrow[p]{} 0$$

$$b_{jk} = \frac{1}{n} \sum_{1}^{n} \frac{\partial^2 \ln f(x_i, \underline{\theta})}{\partial \theta_j \theta_k} \Bigg|_{\underline{\theta}^o}$$

The maximum likelihood equations can be written (ignoring the second degree terms in the expansions):

$$-\underline{\beta_o} = B\,(\underline{\hat{\theta} - \theta^o}) \qquad \text{in matrix form or completely}$$

written out as: $-\beta_{1o} = (\hat{\theta}_1 - \theta_1^o)\, b_{11} + \ldots + (\hat{\theta}_s - \theta_s^o)\, b_{1s}$

$$\vdots \qquad\qquad \vdots$$

$$-\beta_{so} = (\hat{\theta}_1 - \theta_1^o)\, b_{s1} + \ldots + (\hat{\theta}_s - \theta_s^o)\, b_{ss}$$

$B \xrightarrow[p]{} E[B]$ (i.e., each element b_{jk} replaced by $E[b_{jk}]$)

For large n: $B\,[E(B)]^{-1} \xrightarrow[p]{} 1$

$$-B_o = B\,[E(B)]^{-1}\,[E(B)]\;(\underline{\hat{\theta} - \theta^o}) = [E(B)](\underline{\hat{\theta} - \theta^o})$$

or: $(\underline{\hat{\theta} - \theta^o}) = -[E(B)]^{-1}\, B_o$.

$[E(B)] = V$ and is non-singular.

Variance-covariance matrix of $\underline{\hat{\theta} - \theta^o} = E(B)^{-1}\,[\text{var-cov } \underline{B_o}]\,\{[E(B)]^{-1}\}'$

$$= V^{-1} \qquad V \qquad V^{-1}$$

$$= V^{-1}$$

Example of the information matrix used in the multiparameter estimation:

$$f(x;\, a,\, b) = \frac{a^b}{\Gamma(b)}\, e^{-ax}\, x^{b-1} \qquad x > 0 \quad a > 0 \quad b > 0$$

$$\ln L = n\, b \ln a - n \ln \Gamma(b) - a \sum_{1}^{n} x_i - (b-1)\, \bar{x}_L\, n$$

$$\text{defining } \bar{x}_L = \frac{\sum \ln x_i}{n}$$

$$\frac{1}{n}\frac{\partial \ln L}{\partial a} = \frac{b}{a} - \bar{x} = 0$$

$$\frac{1}{n}\frac{\partial \ln L}{\partial b} = \ln a - \frac{\Gamma'(b)}{\Gamma(b)} + \bar{x}_L = 0$$

From the first equation $\hat{a} = \frac{b}{\bar{x}}$

Substituting this in the second equation we get:

$$\ln\frac{b}{\bar{x}} - F_2(b) + \bar{x}_L = 0 \qquad \text{defining } F_2(b) = \frac{\Gamma'(b)}{\Gamma(b)} = \frac{d}{db}\left(\ln \Gamma(b) \right)$$

$$\ln b - F_2(b) = \ln \bar{x} - \bar{x}_L$$

Note: Pearson has compiled tables for $F_2(b)$ which he called the di-gamma function

Thus we have:

$$\frac{\partial^2 \ln L}{\partial a^2} = n\left(-\frac{b}{a^2}\right)$$

$$\frac{\partial^2 \ln L}{\partial a\,\partial b} = n\left(\frac{1}{a}\right)$$

$$\frac{\partial^2 \ln L}{\partial b^2} = n\left[\frac{d^2}{db^2} - \ln \Gamma(b)\right] = n\left(-F_2'(b)\right)$$

$$V = n\begin{bmatrix} \frac{b}{a^2} & \frac{1}{a} \\ \frac{1}{a} & F_2'(b) \end{bmatrix}$$

$$|V| = n\left[+\frac{b}{a^2}F_2'(b) - \frac{1}{a^2}\right] \qquad \text{note: } F_2'(b) \text{ is called the tri-gamma}$$

The asymptotic variances or covariances are thus:

$$\text{of } \hat{a} = \frac{F_2'(b)}{|V|} = \frac{F_2'(b)\,a^2}{n\left[bF_2'(b) - 1\right]}$$

$$\text{of } \hat{a}, \hat{b} = \frac{1/a}{|V|} = \frac{a}{n\left[bF_2'(b) - 1\right]}$$

of $\hat{b} = \frac{b/a^2}{|V|} = \frac{b}{n\left[bF_2'(b) - 1\right]}$

$\sqrt{n}\ (\hat{b} - b)$; $\sqrt{n}\ (\hat{a} - a)$ thus have a joint normal asymptotic distribution with means 0 and variance-covariance matrix:

$$\left(\frac{V}{n}\right)^{-1} = \begin{bmatrix} \frac{a^2 F_2'(b)}{b\,F_2'(b) - 1} & \frac{a}{b\,F_2'(b) - 1} \\ \frac{a}{b\,F_2'(b) - 1} & \frac{b}{b\,F_2'(b) - 1} \end{bmatrix}$$

<u>Exercise</u>: $f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

-- Find the m.l.e. of μ and σ^2 and find the information matrix.

-- The m.l.e. of $\hat{\mu}$ and $\hat{\sigma}^2$ are $\bar{x}$, $\frac{\sum_1^n (x_i - \bar{x})^2}{n}$

-- $\hat{\mu}$, $\hat{\sigma}^2$ are independent, therefore the covariance terms in the information matrix are = 0 .

<u>Problem 37</u>: $f(x) = \frac{1}{2}\ e^{-|x-\mu|}$ $\qquad -\infty < x < \infty$

Note: This is the so-called Laplace distribution.
It is an example of finite theory -- the m.l.e. is not a minimum-variance estimate for any finite sample size.

a) Find the m.l.e. of μ (based on n independent observations).

Does it satisfy the conditions of theorem 22? why not?

b) Is $\bar{x}$ an unbiased estimate of μ? find $\sigma^2_{\bar{x}}$.

<u>Problem 38</u>: X is Poisson λ.

Y is Poisson $\lambda\mu$.

We have a single observation.

Find the m.l.e. of λ, μ and also their information matrix.

Check that the same results are obtained by expanding $\hat{\mu}$ in a Taylor series about μ, hence this result is asymptotic as $\lambda \longrightarrow \infty$

$\mu\lambda \longrightarrow \infty$.

<u>Problem 39</u>: $\frac{X_i}{a}$ is χ_1^2; $\frac{Y_j}{a+b}$ is χ_1^2 X, Y are independent

$a > 0$ $b > 0$

$i = 1,2,\ldots,m$ $j = 1,2,\ldots,n$

Find the m.l.e. of a, b and also the asymptotic variance-covariance matrix.

D. UNBIASED ESTIMATION:

Theorem 24: Information Theorem (Cramer-Rao)

If $d(X)$ is a regular estimate of θ and $E\left[d(X)\right] = \theta + b(\theta)$ (where $b(\theta)$ is a possible bias factor) then

$$\sigma_d^2 \geqslant \frac{\left[1 + b'(\theta)\right]^2}{n k^2} \qquad \text{for all } n$$

where

$$k^2 = E\left[\frac{\partial \ln f(X)}{\partial \theta}\right]^2 = - E\left[\frac{\partial^2 \ln f(X)}{\partial \theta^2}\right]$$

-- the equality holds if and only if $d(X)$ is a linear function of $\frac{\partial \ln f}{\partial \theta}$.

Regularity conditions for the Information Theorem:

1. θ is a scalar in open interval.

   $X_1, X_2, \ldots, X_n$ are independently and identically distributed with density $f(X, \theta)$.

2. $\frac{\partial}{\partial \theta}$ exists for almost all x (the set where $\frac{\partial f}{\partial \theta}$ does not exist must not depend on θ).

   (Note: Problem 36 where $f(x) = \frac{1}{a}$ (i.e. uniform on (0, a)) would be an exception to this condition).

3. $\int f(x, \theta)\, dx$ can be differentiatedunder the integral sign with respect to θ.

4. $\int d(x) f(x, \theta)\, dx$ can be differentiated under the integral sign with respect to θ.

5. $0 < k^2 < \infty$

Proof: From the proof of theorem 22 we remember that

$$E\left(\frac{\partial \ln f}{\partial \theta}\right) = 0 \qquad - E\left(\frac{\partial^2 \ln f}{\partial \theta^2}\right) = E\left(\frac{\partial \ln f}{\partial \theta}\right)^2 = k^2$$

Differentiating both sides of this equation, we get

$$\int d(x) \frac{1}{f} \frac{\partial f}{\partial \theta} f(x)\, dx = 1 + b'(\theta)$$

which can also be written $E\left[d(\underline{X}) \frac{\partial \ln f}{\partial \theta}\right] = 1 + b'(\theta)$.

The correlation coefficient of $S(\underline{X}) = \frac{\partial \ln f(\underline{X})}{\partial \theta}$ and $d(\underline{X}) \leq 1$

$$\frac{(E\left[S(\underline{X})\ d(\underline{X})\right])^2}{\sigma^2_{d(\underline{X})}\ \sigma^2_{S(\underline{X})}} \leq 1$$

Note: $E\left[S(\underline{X})\right]\ E\left[d(\underline{X})\right]$ term vanishes since $E\left[S(\underline{X})\right] = 0$

or
$$\sigma^2_d \geq \frac{\left[1 + b'(\theta)\right]^2}{\sigma^2_{S(\underline{X})}}$$

now
$$\sigma^2_{S(\underline{X})} = \sum_1^n E\left(\frac{\partial \ln f(X_i)}{\partial \theta}\right)^2 = n\ k^2$$

therefore
$$\sigma^2_{d(\underline{X})} \geq \frac{\left[1 + b'(\theta)\right]^2}{n\ k^2} = \frac{\left[1 + b'(\theta)\right]^2}{n\ E\left[\frac{\partial \ln f(\underline{X})}{\partial \theta}\right]^2}$$

Since $r^2 = 1$ if and only if the random variables are linearly related it follows that the equality in this result holds if and only if $d(\underline{X})$ is a linear function of $S(\underline{X})$.

<u>Example</u>: $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2), σ^2 known.

Find the C-R lower bound for the variance of unbiased estimates of μ.

$$L = \frac{1}{(2\ \pi\ \sigma)^{n/2}}\ e^{-\frac{\sum (X_i - \mu)^2}{2\ \sigma^2}}$$

$$\ln\ L = K - \frac{\sum (X_i - \mu)^2}{2\ \sigma^2}$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{\sum (X_i - \mu)}{\sigma^2}$$

$$-\frac{\partial^2 \ln L}{\partial \mu^2} = +\frac{n}{\sigma^2}$$

$$k^2 = \frac{1}{\sigma^2}$$

Hence any regular unbiased estimate $d(\underline{X})$ of μ must satisfy $\sigma_d^2 \gtrsim \frac{\sigma^2}{n}$

but $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ hence $\bar{x}$ is a minimum variance unbiased estimate (m.v.u.e.).

<u>Example (Discrete)</u>: Binomial

x is B(n, p) -- find the C-R lower bound for the unbiased estimates of p.

$$L = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\ln L = K + x \ln p + (n-x) \ln (1-p)$$ easier way.

$$\frac{\partial \ln L}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0 = x\left(\frac{1}{p} + \frac{1}{1-p}\right) - \frac{n}{1-p} = \frac{x}{p(1-p)} - \frac{n}{(1-p)}$$

$$-\frac{\partial^2 \ln L}{\partial p^2} = \frac{x}{p^2} + \frac{n-x}{(1-p)^2} \qquad \frac{n}{p(1-p)}\left[\frac{x}{n} - \frac{np}{n}\right] =$$

$$-E\left[\frac{\partial^2 \ln L}{\partial p^2}\right] = \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} = \frac{n}{p(1-p)} = n\,k^2 \qquad \therefore CLP = \frac{1}{\frac{n}{p(1-p)}} = \frac{p(1-p)}{n}$$

Hence $\sigma_d^2 \geqslant \frac{p(1-p)}{n}$ but $\sigma^2_{(\frac{X}{n})} = \frac{p(1-p)}{n}$ so that $\frac{X}{n}$ is m.v.u.e.

<u>Note</u>: Recall that the C-R bound is achieved if and only if $d(\underline{X})$ is a linear function of $\frac{\partial \ln f}{\partial \theta}$ -- if $\hat{\theta}$ is not a linear function of $\frac{\partial \ln f}{\partial \theta}$ then $\hat{\theta}$ will not achieve the C-R lower bound.

<u>Problem 40</u>: Find the C-R lower bound for the unbiased estimates of the Poisson parameter λ based on independent observations $X_1, X_2, \ldots, X_n$.

<u>Example</u>: Negative Binomial (Ref. Lehman Ch. 2, p. 2-21, 22)

$L = Pr(x) = pq^x$ i.e., sample until a single success occurs

$$\ln L = \ln p + x \ln (1-p)$$

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p} - \frac{x}{1-p}$$

$$-\frac{\partial^2 \ln L}{\partial p^2} = \frac{1}{p^2} + \frac{x}{(1-p)^2} \qquad E(x) = \frac{1-p}{p}$$

$$-E\left[\frac{\partial^2 \ln L}{\partial p^2}\right] = \frac{1}{p^2} + \frac{1}{p(1-p)} = \frac{1}{p^2 (1-p)} = n k^2$$

thus $\sigma_d^2 \geqslant p^2 (1 - p)$

To find the M.L.E. of p:

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p} - \frac{x}{1-p} \text{ or } \hat{p} = \frac{1}{1+X}$$

-- it can be shown that this estimate of p is biased.

Can we find an unbiased estimate? Yes -- by solving the equation $E\left[d(X)\right] = p$.

To do this we observe

X:	0	1	2	3	... n	.....
Prob:	p	pq	pq^2	pq^3	...pq^n	.....

$$E\left[d(X)\right] = d(0)p + d(1)pq + d(2)pq^2 + \dots + d(n)pq^n + \dots \equiv p$$

or

$$d_o + d_1 q + d_2 q^2 + \dots + d_n q^n + \dots \equiv 1$$

If this power series is to be an identity in q we can equate the coefficients on the left and on the right

$$d_o = 1 \qquad d_1 = 0 \qquad d_2 = 0 \qquad \dots$$

The unbiased estimate is thus $p^* = 1$ if $x = 0$

$= 0$ if $x \geqslant 1$

i.e., the decision as to the status of the whole lot is based on the first observation.

This is unbiased, but .....

(this example is used to show why we don't always want an unbiased estimate)

$\sigma_{p^*}^2 = pq \geqslant p^2 q$ (the C-R lower bound) -- thus the C-R lower bound can't be met

∵ since this estimate is the only unbiased estimate by the uniqueness of power series.

<u>To find an unbiased estimator</u> we can try to solve the equation

$$\int_{-\infty}^{\infty} d(\underline{x})\ f(\underline{x}, \theta)\ d\underline{x} = \theta \quad \text{for the continuous case.}$$

or

$$\sum d(n)\ p_n(\theta) = \theta \quad \text{for the discrete case.}$$

These equations, however, are in general rather messy and difficult to handle.

<u>Problem 41</u>: $f(x) = k\, x^{k-1} \qquad 0 \leq x \leq 1 \qquad k > 0$

1. Find the C-R lower bound.
2. Is it attained by the M.L.E.?

<u>Multiparameter extension of the Information (C-R) Theorem</u>

We have $f(\underline{x};\ \theta_1, \theta_2, \ldots, \theta_k)$.

Assume the same regularity conditions as for Theorem 24 in all the θ's.

Denote: $S_i = \dfrac{\partial \ln L}{\partial \theta_i} \qquad E\left[S_i\ S_j\right] = \lambda_{ij}$

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{pmatrix} \qquad \Lambda \text{ is non-singular}$$

$E\left[S_i\right] = 0$ as in theorems 22, 24

Let $d(\underline{X})$ be an unbiased estimate of θ_1

so that $\int d(\underline{x})\ f(\underline{x};\ \theta_1, \theta_2, \ldots, \theta_k)\ d\underline{x} = \theta_1$.

By differentiation with respect to θ_1

$E\left[d\ S_1\right] = 1$.

By differentiation with respect to θ_j

$E\left[d\ S_j\right] = 0$.

<u>Theorem 25</u>: Under the regularity conditions stated

$$\sigma^2_{d_1} \geqslant I_1' \Lambda^{-1} I_1 \qquad \text{where } I_1' = (1, 0, \ldots, 0) \;\; k \text{ components}$$

or

$$\sigma^2_{d_1} \geqslant \frac{\begin{vmatrix} \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & & \vdots \\ \lambda_{k2} & \cdots & \lambda_{kk} \end{vmatrix}}{\begin{vmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{vmatrix}}$$

Proof: Observe that $E\left[d_1 - \theta_1 - \sum_1^k a_i\, S_i\right]^2 \geqslant 0$ where the a's are arbitrary constants to be determined.

$$E\left[d_1 - \theta_1\right]^2 - 2\,E\left[(d_1 - \theta_1) \sum a_i\, S_i\right] + E\left[\sum a_i\, S_i\right]^2 \geqslant 0$$

From the above relationships this becomes

$$\sigma^2_{d_1} - 2a_1 + \sum_{i=1}^k \sum_{j=1}^k a_i a_j\, E\left[S_i S_j\right] \geqslant 0$$

which then becomes

$$\sigma^2_{d_1} \geqslant 2\,a_1 - \sum_{i=1}^k \sum_{j=1}^k a_i a_j\, \lambda_{ij} \qquad \text{for all real } a_1, a_2, \ldots, a_k\,.$$

Call the right hand side of this inequality $\varphi(\underline{a})$

We wish to maximize $\varphi(\underline{a})$ with respect to $\underline{a}$ to get the best possible statement about the bound of the variance.

$$\varphi(\underline{a}) = 2a_1 - \sum_{i=1}^k a_i^2\, \lambda_{ii} - \sum_{i \neq j}\sum a_i a_j\, \lambda_{ij}$$

$$\frac{\partial \varphi}{\partial a_1} = 2 - 2a_1 \lambda_{11} - 2 \sum_{j=2}^{k} a_j \lambda_{1j} = 0$$

or

$$\sum_{j=1}^{k} a_j \lambda_{1j} = 1$$

$$\frac{\partial \varphi}{\partial a_i} = - 2a_i \lambda_{ii} - 2 \sum_{j=1, j\neq i}^{k} a_j \lambda_{ij} = 0$$

or

$$\sum_{j=1}^{k} a_j \lambda_{ij} = 0 \qquad i = 2, 3, \ldots, k$$

Hence the equations in $\underline{a}$ to maximize $\varphi(\underline{a})$ are

$$\sum_{j=1}^{k} a_j^o \lambda_{1j} = 1$$

$$\sum_{j=1}^{k} a_j^o \lambda_{ij} = 0 \qquad i = 2, 3, \ldots, k$$

where a_j^o is a maximizing value of a_j.

How, multiply the i^{th} equation by a_i^o and add all the equations.

The left hand side of the sum so obtained is

$$\sum_{i=1}^{k} a_i^o \sum_{j=1}^{k} a_j^o \lambda_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i^o a_j^o \lambda_{ij}$$

The right hand side is a_1^o.

Hence $\varphi_{max}(\underline{a}) = 2 a_1^o - a_1^o = a_1^o$

Therefore

$$\sigma_{d_1}^2 \geqslant a_1^o \qquad \text{where: } \Lambda \underline{a}^o = I_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\underline{a}^o = \Lambda^{-1} I_1$$

$$a_1^o = I_1' \Lambda^{-1} I_1$$

or, as stated in the theorem:

$$\sigma^2_{d_1} \geqslant I_1' \Lambda^{-1} I_1 = \frac{\begin{vmatrix} \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & & \vdots \\ \lambda_{k2} & \cdots & \lambda_{kk} \end{vmatrix}}{\begin{vmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{vmatrix}}$$

Remember: $\lambda_{ij} = E\left[S_i S_j\right]$ $S_i = \frac{\partial \ln L}{\partial \theta_i}$

<u>Example</u>: $f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1}$ $x > 0$

-- the generalized Gamma or Pearson's Type III

-- a, b are the unknown parameters

Find the lower bound for the variance of the unbiased estimate of a.

$$\ln L = n \ln a - n \ln \Gamma(b) - a\bar{x} + n(b-1) \ln x$$

$$S_1 = \frac{\partial \ln L}{\partial a} = \frac{nb}{a} - \bar{x}$$

$$S_2 = \frac{\partial \ln L}{\partial b} = n \ln a - n \frac{\partial}{\partial b}\left[\ln \Gamma(b)\right] + n \ln x$$

denote $\frac{\partial}{\partial b}\left[\ln \Gamma(b)\right] = F_2(b)$

$$\lambda_{11} = E\left[S_1^2\right] = -\frac{\partial^2 \ln L}{\partial a^2} = \frac{nb}{a^2}$$

$$\lambda_{22} = E\left[S_2^2\right] = -\frac{\partial^2 \ln L}{\partial b^2} = n F_2'(b)$$

$$\lambda_{12} = \lambda_{21} = E\left[S_1 S_2\right] = -\frac{\partial^2 \ln L}{\partial a \partial b} = -\frac{n}{a}$$

$$\Lambda = n \begin{pmatrix} \frac{b}{a^2} & -\frac{1}{a} \\ -\frac{1}{a} & F_2'(b) \end{pmatrix}$$

$$\sigma^2_{\hat{a}} \geqslant \frac{F_2'(b)}{n\left(\frac{b\,F_2'(b)}{a^2} - \frac{1}{a^2}\right)} = \frac{a^2\,F_2'(b)}{n\left[b\,F_2'(b) - 1\right]} = \frac{a^2}{n\left[b - \frac{1}{F_2'(b)}\right]}$$

Note: See the example for Theorem 23.

<u>Note</u>: We started the proof with $E\left[d_1 - \theta_1 - \sum_{i=1}^{k} a_i\, S_i\right]^2 \geqslant 0.$

The strict inequality will hold unless $\left[d_1 - \theta_1 - \sum_{1}^{k} a_i\, S_i\right]$ is essentially constant.

Therefore: If the multiparameter C-R lower bound is attained, d_1 is a linear function of $\sum_{i=1}^{k} a_i\, S_i$

and, as before, the M.L.E. (corrected for bias if necessary) attains the lower bound (<u>IF there is any</u> estimate that attains that lower bound).

<u>Problem 42</u>: $X_1, \ldots, X_n$ are negative binomial

$$\Pr\left[X = x\right] = \binom{r + x - 1}{r - 1} p^r q^x \qquad p + q = 1 \qquad x = 0, 1, 2, \ldots$$

Find the C-R lower bounds for the unbiased estimates of p

1. If r is known

2. If r is unknown

Problem 42 (b): Show that if $X | \lambda$ is Poisson (λ) and λ is distributed with density

$$f(\lambda) = \frac{a^b}{\Gamma(b)} e^{-a\lambda} \lambda^{b-1}$$

then X is negative binomial. Identify functions of a, b with p, r

---

Summary on usages re m.l.e.:

We have the following criteria for estimates:

1. Efficiency (Cramer) -- attains the C-R lower bound.
2. (Asymptotic) Efficiency (Fisher) -- (n) (variance) $\longrightarrow \frac{1}{k^2}$ in the limit.
3. Minimum variance unbiased estimates.
4. Best asymptotic normal (b.a.n.) -- among A.N. estimates, no other has a smaller asymptotic variance.

Properties of m.l.e.

1. If there is an efficient (Cramer) estimate (i.e., if the estimate is a linear function of $\frac{\partial \ln L}{\partial \theta}$) then the m.l.e. is efficient (Cramer).
2. If the m.l.e. is efficient (Cramer) and unbiased it is m.v.u.e. Other m.l.e. may or may not be m.v.u.e.
3. Under the general regularity conditions, the m.l.e. is asymptotically efficient (Fisher).
4. Among the class of unbiased estimates, then the m.l.e. are b.a.n., otherwise (i.e., if the m.l.e. are biased) we can not say that the m.l.e. are necessar b.a.n.

   ref: Le Cam; Univ. of Calif. Publications in Statistics; Vol. I, No. 11, 1953

-- The class of efficient (Cramer) estimates is contained in the class of m.v.u.e. -- which is the real reason we are interested in attaining the C-R lower bound.

-- 1, 2 are finite results, i.e. for any n.

-- 3, 4 are finite (asymptotic) results -- for large n only.

E. m.v.u.e. -- COMPLETE SUFFICIENT STATISTICS:

-- Let X be a random variable with density $f(\underline{x}, \underline{\theta})$, $\underline{\theta} \in \Omega$.

-- T(X) is a statistic, i.e., a function of the random variable $(\underline{X})$.

-- Assume that the conditional d.f. of X, given T = t, exists.

Def. 31: $T(\underline{X})$ is a sufficient statistic for $\underline{\theta}$ if the distribution of X given T = t is (mathematically) independent of $\underline{\theta}$, that is, if in

$$f(\underline{x}, \underline{\theta}) = g(T, \underline{\theta}) \; h(x|T)$$

h is independent of $\underline{\theta}$.

Examples:

No. 1 -- suppose $X_1, X_2, \ldots, X_n$ are $N(\mu, 1)$

$$f(x, \mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum(x_i - \mu)^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum x_i^2} e^{\mu\sum x_i} e^{-\frac{n\mu^2}{2}}$$

$$T(\underline{x}) = \sum_1^n x_i = n\bar{x}$$

$$f(\underline{x}, \mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum(x_i - \bar{x})^2}{2}} e^{-\frac{n(\bar{x} - \mu)^2}{2}}$$

$\bar{x}$ is thus sufficient for μ.

-- distribution of $x|\bar{x}$ is $N(\bar{x}, 1)$

No. 2 -- Poisson: $f(x_1, x_2, \ldots, x_n) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!}$

$$f(\underline{x}) = \underbrace{e^{-n\lambda} \frac{(n\lambda)^{\sum x_i}}{(\sum x_i)!}}_{g(T, \lambda)} \underbrace{\frac{(\sum x_i)!}{\prod x_i!} \left(\frac{1}{n}\right)^{x_1} \cdots \left(\frac{1}{n}\right)^{x_n}}_{h(X|T)}$$

$\sum x_i$ is Poisson $(n\lambda)$; given $\sum x_i$, the individual observations are multinomial and $T = \sum x_i = n\bar{x}$ is sufficient for λ.

No. 3 -- Normal

$$f(\underline{x}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}}$$

$$= \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}} e^{-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}}$$

Here sufficient statistics for μ, σ^2 are $\bar{x}$; $\sum_1^n (x_i - \bar{x})^2$.

<u>Problem 43</u>: Find sufficient statistics (if any) for the parameters of the following distributions:

a-- Gamma $\quad f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1} \qquad x > 0$

b-- power $\quad f(x) = k\, x^{k-1} \qquad 0 \leq x \leq 1$

c-- Beta $\quad f(x) = \frac{1}{\beta(a, b)} x^{a-1} (1 - x)^{b-1} \qquad 0 \leq x \leq 1$

d-- Cauchy $\quad f(x) = \frac{1}{\pi} \left(\frac{1}{1 + (x - \mu)^2} \right)$

e-- neg. binomial $\quad p(x) = \binom{x + r - 1}{r - 1} p^r q^x \qquad x = 0, 1, 2, \ldots$

f-- normal $\quad$ mean: $\alpha + \beta t_i \qquad t_i$ known

variance: σ^2

<u>Lemma</u>: for any real number c

$$E\left[X - c\right]^2 \geq E_T\left[E_X(X|T) - c\right]^2$$

Proof: $E\left[X^2 \mid t\right] \geq \left[E(X \mid t)\right]^2$ where t is any particular value of the statistic T.

$$E[X] = E_T\left[E_X(X \mid T)\right]$$

$$E\left[X^2\right] = E_T\left[E_X(X^2 \mid T)\right] \geq E_T\left[E(X \mid T)\right]^2$$

replacing X by $(X - c)$ we get

$$E\left[X - c\right]^2 \geq E_T\left[E_X(X - c \mid T)\right]^2$$

$$\geq E_T\left[E_X(X \mid T) - c\right]^2$$

Remark: the equality holds only if X is a function of T since the equality holds at step 1 if $X \mid t$ is a constant.

Theorem 26: Let $\underline{X}$ have density $f(\underline{x}, \underline{\theta})$.

Let T be a sufficient statistic for $\underline{\theta}$.

If $d(\underline{X})$ is any unbiased estimate of $g(\underline{\theta})$, then $\psi(T) = E_X\left[d(\underline{X}) \mid T\right]$ is an unbiased estimate of $g(\underline{\theta})$ and $\sigma_\psi^2 \leq \sigma_d^2$ with the equality holding only if $\psi(T) = d(\underline{x})$.

Proof: Since T is sufficient then the distribution of $d(\underline{X}) \mid T$ is independent of $\underline{\theta}$ and hence

$\psi(T) = E_X\left[d(\underline{X}) \mid T\right]$ is independent of $\underline{\theta}$.

1- $E\left[\psi(T)\right] = E_T\left\{E_X\left[d(\underline{X}) \mid T\right]\right\} = E\left[d(\underline{X})\right] = g(\underline{\theta})$

2- In the result of the lemma put $d(\underline{X}) = X$ $\quad g(\underline{\theta}) = c$

then it follows immediately that

$$\sigma_d^2 = E\left[d(\underline{x}) - g(\underline{\theta})\right]^2 \geq E_T\left[\psi(T) - g(\underline{\theta})\right]^2$$

$$\geq \sigma_\psi^2$$

Examples:

1.-- Binomial X_i is $B(1, p)$ $\quad i = 1, 2, \ldots, n$

$$f(x_1, x_2, \ldots, x_n) = p^X(1 - p)^{n-X} \quad \text{where } X = \sum_1^n x_i$$

(since $f(\underline{x})$ is ordered, the coefficient term is omitted)

X is a sufficient statistic for p.

$E(X_1) = p$ so that X_1 is an unbiased estimate of p.

$$\sigma^2_{x_1} = p(1 - p)$$

Define $\psi(X) = E[X_1 | X]$ -- by the theorem $\psi(X)$ will be unbiased and have smaller variance.

$X_1 = 1$ if success on trial 1
$\quad = 0$ if failure on trial 1

in n trials we have X successes

$$Pr[X_1 = \text{success} \mid X] = X/n$$

$$Pr[X_1 = 0 | X] = 1 - \frac{X}{n}$$

$$Pr[X_1 = 1 / X] = \frac{X}{n}$$

therefore

$$\psi(X) = E[X_1 | X] = \frac{X}{n}$$

2-
Problem 44: Negative binomial; r known

show that
1. X is sufficient for p.
2. $(1 - X_1)$ is an unbiased estimate of p.

note: $X_i = 1$ if failure on trial i
$\quad = 0$ otherwise

find $E[X_1 | X]$.

$$X = \sum X_i$$

Example No. 3: Uniform distribution on $(0, \theta)$

$Z = \max(X_1, X_2, \ldots, X_n)$ is sufficient for θ.

$E[X_i] = \frac{\theta}{2}$ so that $2X_i$ is an unbiased estimate of θ.

$\psi(Z) = E[2X_i | Z]$ will be unbiased and have smaller variance.

proof: If X_i is one of the observations less than Z then X_i is uniform on $(0, Z)$,

then $E[2X_i | X_i < Z] = 2(\frac{Z}{2}) = Z.$

If X_i is the maximum (i.e., $= Z$) then $E[2X_i | Z] = 2Z.$

Thus:

$$E[2X_i | Z] = \frac{n-1}{n} Z + \frac{1}{n} 2Z = (1 - \frac{1}{n})Z + \frac{2}{n} Z$$

$$= (1 + \frac{1}{n}) Z = (\frac{n+1}{n}) Z$$

4- Given that X is a Poisson random variable, estimate $e^{-\lambda}$ (i.e., estimate the probability of getting a zero observation).

Given $X_1, X_2, \ldots, X_n$;

$$p(X_1, \ldots, X_n) = e^{-\lambda n} \frac{(\lambda)^{\sum X_i}}{\prod X_i !}$$

we know that $T = \sum_1^n X_i$ is sufficient for λ and

$$p(T) = e^{-n\lambda} \frac{(n\lambda)^T}{T!}$$ (see example No. 2 on p. 86)

(also $\bar{X}$ is m.l.e. for λ and $e^{-\bar{X}}$ is m.l.e. of $e^{-\lambda}$).

What is an unbiased estimate of $e^{-\lambda}$??

Let n_0 = the number of zeros among $X_1, X_2, \ldots, X_n$.

$$E\left[\frac{n_0}{n}\right] = e^{-\lambda} = Pr[X = 0]$$

so that $\frac{n_0}{n}$ is an unbiased estimate of $e^{-\lambda}$.

Define the estimate of $e^{-\lambda}$

$$\psi(T) = E\left[\frac{n_0}{n} \middle| T\right]$$

Let: $Y_i = 1$ if $X_i = 0$

$= 0$ if $X_i > 0$

then $n_0 = \sum_1^n Y_i$

$$E\left[Y_i \mid T\right] = Pr\left[Y_i = 1 \mid T\right] = Pr\left[X_i = 0 \mid T\right] = (1 - \frac{1}{n})^T$$

$$E\left[\frac{n_0}{n} \mid T\right] = \frac{1}{n} \sum_{1}^{n} E\left[Y_i / T\right] = \frac{1}{n}\left[n(1 - \frac{1}{n})^T\right]$$

To show that $\psi(T) = (1 - \frac{1}{n})^T$ is unbiased:

$$E\left[\psi(T)\right] = \sum_{T=0}^{\infty} (1 - \frac{1}{n})^T \; e^{-n\lambda} \; \frac{(n\lambda)^T}{T!}$$

$$= \sum e^{n\lambda} \frac{\left[(1 - \frac{1}{n})(n\lambda)\right]^T}{T!} = e^{-n\lambda} \; e^{n\lambda - \lambda} = e^{-\lambda}$$

<u>Problem 45</u>: Find the variance of $\psi(T) = (1 - \frac{1}{n})^T$

Compare it with the C.R. lower bound for unbiased estimates of $e^{-\lambda}$.

Sufficient statistics are not unique!

Example 1: Poisson $T = \sum X_i$ is sufficient, but $\frac{T}{n} = \bar{X}$ is equally good.

Example 2: $X_1, X_2, \ldots, X_n$ are $N(\mu, 1)$

$$f(\underline{x}) = (\frac{1}{\sqrt{2\pi}})^n \; e^{-\frac{\sum (X_i - \mu)^2}{2}} = (\frac{1}{\sqrt{2\pi}})^n \; e^{-\frac{\sum X_i^2 + 2\mu \sum X_i - n \mu^2}{2}}$$

$\sum X_i^2$, $\bar{X}$ are sufficient for μ, however $\sum X_i^2$ is not necessary and gives no more information about μ than does $\bar{X}$

Example 3: $f(\underline{x}; \mu, \sigma^2) = (\frac{1}{\sqrt{2\pi} \; \sigma})^n \; e^{-\frac{\sum (X_i - \mu)^2}{2\sigma^2}}$

$$T_1 = \bar{X} \qquad T_2 = \sum (X_i - \bar{X})^2$$

or also $T_1 = \bar{X}$ $\quad T_2 = \sum X_i^2$ — this set gives the same information (need to combine them to estimate σ^2)

Example 4: Same as No. 3, but μ is known,

then $T = \sum (X_i - \mu)^2$ is sufficient for σ^2.

Remark: (aimed at problem 43, but holds in general)

The m.l.e. is a function of the sufficient statistic(s)

since $L = g(\underline{T}, \underline{\theta})\, h(\underline{X}, \underline{T})$

$$\ln L = \ln g + \ln h$$

$$\frac{\partial \ln L}{\partial \theta_i} = \frac{1}{g(\underline{T}, \underline{\theta})} \frac{\partial g}{\partial \theta_i} = 0 \qquad \frac{\partial \ln L}{\partial \theta_i} = 0$$ since it is independent of θ

The solution of this equation (which gives the m.l.e.) obviously depends only on T.

Def. 32: A sufficient statistic is called complete if

$E_\theta\left[h(\underline{T})\right] \equiv_\theta 0$ implies $h(\underline{T}) = 0$. $\qquad \equiv_\theta$ denotes identically equal in θ

i.e. $\int h(\underline{T})\, f(\underline{T}, \underline{\theta})\, d\underline{T} \equiv_\theta 0$

Theorem 27: If T is a complete sufficient statistic and $d(T)$ is an unbiased estimate of $g(\theta)$ then $d(T)$ is an essentially unique minimum variance unbiased estimate (m.v.u.e.).

Proof: Let $d_1(T)$ be any other unbiased estimate of $g(\theta)$; then we know that

$$E_\theta\left[d(T)\right] = g(\theta)$$

$$E_\theta\left[d_1(T)\right] = g(\theta)$$

thus by subtraction $E_\theta\left[d(T) - d_1(T)\right] = 0$ for all θ.

The completeness of T implies that $d(T) - d_1(T) = 0$ (see def. 32)

or that $d(T) = d_1(T)$.

Further -- let $d'(\underline{X})$ be any other unbiased estimator. We know from theorem 26 that $E\left[d'(\underline{X}/T)\right] = \psi(T)$ is unbiased and $\sigma^2_\psi \leq \sigma^2_{d'}$ with the equality holding only if d' is a function of T.

But, by the first part of the proof $\psi(T) = d(T)$,

Hence, if we start with an unbiased estimator not a function of T, we can improve it. If we start with an unbiased estimator that is a function of T, it is $d(T)$. This is the contention of the theorem.

Example No. 1: Binomial X is $B(n, p)$ $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

For completeness of X, which has previously been shown to be a sufficient statistic, we need

$$\sum_{x=0}^{n} h(X) \binom{n}{x} p^x (1-p)^{n-x} \underset{p}{\equiv} 0 \implies h(X) = 0 .$$

The left hand side is a polynomial in p of degree n.

$$P_n(p) \equiv 0 \qquad \text{i.e.} \quad a_n p^n + a_{n-1} p^{n-1} + \ldots + a_0 \equiv 0$$

For this to be identically zero in p implies that all coefficients are zero, which means $h(X) = 0$ for every $x = 0, 1, 2, \ldots, n$,

therefore $h(X) \underset{p}{\equiv} 0$ and X is a complete sufficient statistic,

hence $\frac{X}{n}$ is m.v.u.e.

Example No. 2: Poisson $X_1, X_2, \ldots, X_n$ are Poisson (λ).

$T = \sum X_i$ is sufficient and is Poisson $(n\lambda)$.

For completeness, we must have

$$\sum_{T=0}^{\infty} h(T)\, e^{-n\lambda} \frac{(n\lambda)^T}{T!} \underset{\lambda}{\equiv} 0 \quad \text{implying} \quad h(T) = 0 \text{ on the integers}$$

$$\text{or} \quad \sum_{T=0}^{\infty} h(T)\,\frac{(n\lambda)^T}{T!} \underset{\lambda}{\equiv} 0$$

$$h_0 + h_1 \frac{n\lambda}{1!} + h_2 \frac{(n\lambda)^2}{2!} + \cdots \underset{\lambda}{\equiv} 0$$

Such a power series identically zero means each coefficient = 0 or $h(T) = 0$, so that T is complete -- therefore $\frac{T}{n}$ is m.v.u.e. of λ,

$(1 - \frac{1}{n})^T$ is m.v.u.e. of $e^{-\lambda}$.

Example No. 3: Normal $X_1, X_2, \ldots, X_n$ are $N(\mu, 1)$.

$\bar{X}$ is sufficient and is $N(\mu, \frac{1}{n})$.

$$E_\mu\left[h(\bar{X})\right] = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(\bar{x})\, e^{-\frac{n(\bar{x}-\mu)^2}{2}}\, d\bar{x} = \sqrt{\frac{n}{2\pi}} \int h(\bar{x})\, e^{-\frac{n\bar{x}^2}{2}}\, e^{-n\bar{x}\mu} e^{-\frac{n\mu^2}{2}}\, d\bar{x}$$

$$= \sqrt{\frac{n}{2\pi}}\, e^{-\frac{n\mu^2}{2}} \int \left[h(\bar{X})\, e^{-\frac{n\bar{x}^2}{2}}\right] e^{-n\bar{x}\mu}\, d\bar{x} \underset{\mu}{\equiv} 0$$

(this is a bilateral Laplace transform)

By the theory of Laplace transforms (ref: Widder's book) if the Laplace transform $\equiv 0$ identically, then the function = 0 or

$$h(\bar{X})\, e^{-\frac{n\bar{X}^2}{2}} = 0 \quad \text{which implies that} \quad h(\bar{X}) = 0,$$

therefore $\bar{X}$ is complete.

If $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ then by a similar argument $(\bar{X}, s^2)$ are complete for μ, σ^2.

Remark: Any estimate which is unbiased and a function of $\bar{X}$, s^2 is m.v.u.e. of $g(\mu, \sigma^2)$.

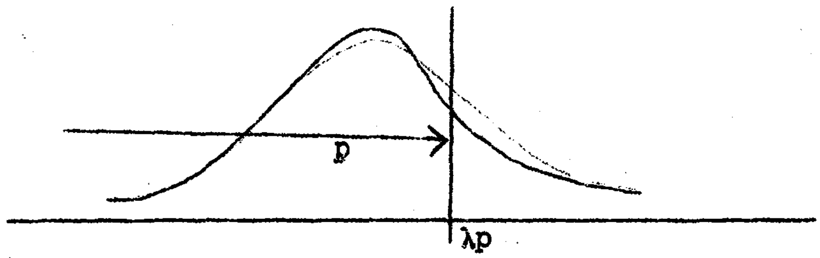
In particular, if $g(\mu, \sigma^2) = \mu^2 + \sigma^2 = E\left[X^2\right]$,

$\frac{1}{n} \sum X_i^2$ is an unbiased estimate of $\mu^2 + \sigma^2$.

$$\frac{1}{n} \sum X_i^2 = \frac{(n-1)s^2 + n\bar{X}^2}{n} \quad \text{so} \quad \frac{1}{n} \sum X_i^2 \text{ is m.v.u.e. of } \mu^2 + \sigma^2$$

Problem 46: Find m.v.u.e. of λ_p where $Pr[x < \lambda_p] = p$,

and $X_1, X_2, \ldots, X_n$ are $NID(\mu, \sigma^2)$



Example: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$

$Y_1, Y_2, \ldots, Y_n$ are $N(\nu, \sigma^2)$

Sufficient statistics are $(\bar{X}, \bar{Y}, s_x^2, s_y^2)$

(i.e., we have four sufficient statistics for three parameters),
thus this sufficient statistic is not complete.

Remark: for completeness, the sufficient statistic vector must have the same number of components as the parameter vector.

Remark: for an example of what can happen by forcing the criteria of unbiasedness on an estimate, see Lehmann p. 3-13, 14.

Non-parametric Estimation (m.v.u.e.)

Let $X_1, X_2, \ldots, X_n$ be independent random variables with continuous d.f. $F(x)$.
The problem is to estimate some function $g(F)$, e.g.

$$g(F) = \int_{-\infty}^{\infty} x \, dF(x) = E[X] \quad \text{provided } E[X] < \infty$$

$$g(F) = \sigma_x^2$$

$g(F) = F(x)$ i.e., we want to estimate the density function

$$g(F) = F(a) = Pr[X \leq a]$$

$$g(F) = F(b) - F(a)$$

$g(F) = (X_1, X_n)$ such that $F(X_n) - F(X_1) \geq 1 - \alpha$

(two-sided tolerance limit problem)

<u>Theorem 28</u>: If $d(X_1, X_2, \ldots, X_n)$ is symmetric in $X_1, X_2, \ldots, X_n$ and $E[d(\underline{X})] = g(F)$ then $d(\underline{X})$ is m.v.u.e. of $g(F)$.

Proof: Sufficient statistic is $T = (\sum X_i, \sum X_i^2, \ldots, \sum X_i^n)$.

Consider the n equations:

$$\sum X_i = t_1$$

$$\sum X_i^2 = t_2$$

$$\ldots$$

$$\sum X_i^n = t_n$$

These equations have at most $n!$ solutions.

Assume (as is true with probability 1) that all the X's are distinct -- if $X_1, X_2, \ldots, X_n$ is a solution, so is any permutation of the X's.

There are $n!$ permutations of the X's so these give a complete set of solutions.

Since the sufficient statistic may be regarded as a set of observations, order disregarded, any function of $\underline{T}$ is symetric in $X_1, X_2, \ldots, X_n$.

To show completeness, we must show

$$\int g(T)\, dF(T) \underset{F}{\equiv} 0 \implies g(T) = 0 .$$

Consider the sub-family of densities $C(\theta_1, \theta_2, \ldots, \theta_n) e^{-x^{2n} - \theta_1 x - \theta_2 x^2 - \ldots \theta_n x^n}$

$$f(\underline{x}) = C^n e^{-\sum x_i^{2n} - \theta_1 \sum x_i - \theta_2 \sum x_i^2 - \ldots - \theta_n \sum x_i^n}$$

$$= C^n e^{-\sum x_i^{2n} - \theta_1 t_1 - \theta_2 t_2 - \ldots - \theta_n t_n}$$

$$f(\underline{x}) = h(\underline{x}) \exp(-\theta_1 t_1 - \theta_2 t_2 - \ldots - \theta_n t_n)$$

$$T = (t_1, t_2, \ldots, t_n)$$

$$f(\underline{T}) = f(t_1, t_2, \ldots, t_n) = C \exp(-\theta_1 t_1 - \theta_2 t_2 - \ldots - \theta_n t_n)\, p(t_1, \ldots, t_n)$$

where: $p(t_1, t_2, \ldots, t_n)$ is polynomial in the t's obtained

from the Jacobian -- it does not involve the θ's and is non-negative.

Hence, if $$\int_{-\infty}^{\infty} g(\underline{T})\, e^{-\sum_{j=1}^{n} \theta_j t_j}\, p(\underline{T}) \underset{\theta_1, \ldots, \theta_n}{\equiv} 0$$

by the same type of theory as in the normal case (uniqueness of the bilateral Laplace transform) this implies

$$g(\underline{T})\, p(\underline{T}) = 0 \qquad \text{or} \qquad g(\underline{T}) = 0 \qquad \text{q.e.d.}$$

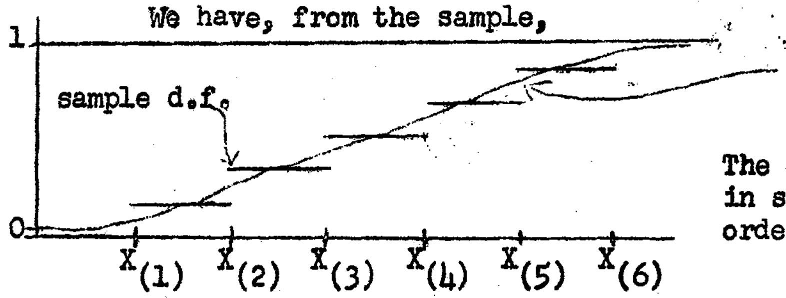
Example 1: Estimate $E(X)$.

$\bar{X}$ is symetric function of $X_1, X_2, \ldots, X_n$,

further, it is unbiased so that by theorem 28 it is m.v.u.e.

Example 2: Estimate $F(x)$.

We have, from the sample,



The observations here are ordered in size but not by observation order (chronological).

$$F_n(x) = \frac{1}{n}(\# x_i \le x) = \frac{1}{n} \sum_{i=1}^{n} \Psi_i(x) \quad \text{where } \Psi_i(x) = 1 \text{ if } x_i \le x$$
$$= 0 \text{ if } x_i > x$$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \Psi_i(x)$$ indicates symmetry in the x's.

It then remains to show that it is unbiased, i.e., for each x: $E\left[F_n(x)\right] = F(x)$.

For each fixed x the number of $x_i \le x$ is a binomial random variable with parameters $[n, F(x)]$ -- and hence, since the sample frequency is an unbiased estimate of the binomial parameter, we have the required unbiasedness.

Remark: $Fn(x) \xrightarrow[p]{} F(x)$ for all x [Proof given later]

Problem 47: Find the m.v.u.e. of $Pr[X \leq a]$ (a known) given observations $X_1, X_2, \ldots, X_n$ with d.f. $F(x)$.

Problem 48: Given $X_1, X_2, \ldots, X_n$ are $NID(0, \sigma^2)$ and

$S = \sum X_i^2$ is sufficient for σ^2,

a) find the minimum risk estimate of σ^2 among functions of the form aS.

b) is there a constant risk estimator of σ^2 of the form $aS + b$?

F. χ^2-estimation

-- most generally applied to multinomial situations.
-- usually associated with Karl Pearson.

ref: Ferguson -- Annals of Math. Stat. -- Dec. 58

Multinomial Distribution:

Given -- a series of trials with:

possible outcomes of each independent trial $E_1 \; E_2 \ldots E_k$

probabilities for a given outcome: $p_1 \; p_2 \ldots p_k$

and n experiments result in: $v_1 \; v_2 \ldots v_k$

with the restrictions that $\sum p_i = 1 \quad \sum v_i = n$

$$Pr[v_1, v_2, \ldots, v_k] = \frac{n!}{v_1! \, v_2! \ldots v_k!} \, p_1^{v_1} p_2^{v_2} \ldots p_k^{v_k}$$

[this is a term of the multinomial expansion of $(p_1 + p_2 + \ldots + p_k)^n$]

Characteristic function:

$$\varphi_{v_1, v_2, \ldots, v_k}(t_1, t_2, \ldots, t_k) = \sum \frac{n!}{v_1! \, v_2! \ldots v_k!} (p_1 e^{t_1})^{v_1} \ldots (p_k e^{t_k})^{v_k}$$

$$= (p_1 e^{t_1} + \ldots + p_k e^{t_k})^n$$

$$\varphi_{v_1, v_2, \ldots, v_k}(t_1, 0, \ldots, 0) = (p_1 e^{t_1} + p_2 + \cdots + p_k)^n$$

$$= [p_1 e^{t_1} + (1 - p_1)]^n$$

i.e., any one variable in a multinomial situation is is binomial $B(n, p_i)$.

Moments: $E[v_i] = np_i$ $\qquad \sigma_{v_i}^2 = np_i(1 - p_i)$

$$\sigma_{v_i v_j} = -np_i p_j = \left.\frac{\partial \varphi}{\partial t_i \partial t_j}\right|_{t=0} - E[v_i]\ E[v_j]$$

Asymptotic distributions:

a) $z_i = \dfrac{v_i - np_i}{\sqrt{np_i(1-p_i)}}$ is $A\ N(0, 1)$ as $n \longrightarrow \infty$.

b) $\sum_{i=1}^{k} z_i^2 = \sum_{i=1}^{k} \dfrac{(v_i - np_i)^2}{np_i}$ has a χ^2-distribution with k-1 d.f. as $n \longrightarrow \infty$

(see Cramer for the transformation from k dependent to k-1 independent variables)

c) as $n \longrightarrow \infty$, $np_i \longrightarrow \lambda_i$ $\qquad i = 1, 2, \ldots, k-1$

$v_1, v_2, \ldots, v_{k-1}$ have a limiting multi-Poisson distribution with

parameters $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$ ($v_1, v_2, \ldots, v_{k-1}$ are independent in the limit).

Example: (Homogeneity)

We have 3 sets of multinomial trials with possible outcomes E_1, E_2, E_3,

with probabilities $\theta_1, \theta_2, 1 - \theta_1 - \theta_2$ for each set of trials.

Therefore we have the following outcomes

v_{11}	v_{12}	v_{13}	n_1
v_{21}	v_{22}	v_{23}	n_2
v_{31}	v_{32}	v_{33}	n_3
$v_{\circ 1}$	$v_{\circ 2}$	$v_{\circ 3}$	N

$$\ln L = \ln\left[p_{11}^{v_{11}} \; p_{12}^{v_{12}} \; p_{13}^{v_{13}} \; p_{21}^{v_{21}} \; p_{22}^{v_{22}} \; p_{23}^{v_{23}} \; p_{31}^{v_{31}} \; p_{32}^{v_{32}} \; p_{33}^{v_{33}} \right] + \ln K$$

$$= v_{\circ 1} \ln \theta_1 + v_{\circ 2} \ln \theta_2 + v_{\circ 3} \ln (1 - \theta_1 - \theta_2) + \ln K$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{v_{\circ 1}}{\theta_1} - \frac{v_{\circ 3}}{1-\theta_1-\theta_2} = 0$$

$$\frac{\partial \ln L}{\partial \theta_2} = \frac{v_{\circ 2}}{\theta_2} - \frac{v_{\circ 3}}{1-\theta_1-\theta_2} = 0$$

These give the two equations:

(1) $(v_{\circ 1} + v_{\circ 3}) \theta_1 + v_{\circ 1} \theta_2 = v_{\circ 1}$

(2) $v_{\circ 2} \theta_1 + (v_{\circ 2} + v_{\circ 3}) \theta_2 = v_{\circ 2}$

which when added together yield

(3) $N \theta_1 + N \theta_2 = v_{\circ 1} + v_{\circ 2} = N - v_{\circ 3}$

Multiplying (3) by $\frac{v_{\circ 2}}{N}$ we get

(4) $v_{\circ 2} \theta_1 + v_{\circ 2} \theta_2 = \frac{v_{\circ 2}(v_{\circ 1} + v_{\circ 2})}{N}$

(2) - (4) yields $v_{\circ 3} \theta_2 = \frac{v_{\circ 2} \; v_{\circ 3}}{N}$

or $\hat{\theta}_2 = \frac{v_{\circ 2}}{N}$

Similarly: $\hat{\theta}_1 = \frac{v_{\cdot 1}}{N}$.

It can also be found that $$V(\hat{\theta}_2) = \frac{\theta_2 (1-\theta_2)}{N}$$

$$\text{cov}(\hat{\theta}_1, \hat{\theta}_2) = - \frac{\theta_1 \theta_2}{N}$$

Problem 49: (Independence)

Consider one sequence of N trials that result in one of the rc events $E_{11}, E_{12}, \ldots, E_{rc}$ with probabilities $p_{ij} = \rho_i \tau_j$

where $\sum_{i=1}^{r} \rho_i = 1$; $\sum_{j=1}^{c} \tau_j = 1$.

Find the m.l.e. of ρ_i, τ_j and also their variances and covariances.

χ^2-estimation: general case

Given:--s series of n_i trials, each trial resulting in one of the events $E_1, \ldots, E_k$

-The probability of E_j occuring on any trial in the i^{th} series is p_{ij} ; $\sum_{j=1}^{k} p_{ij} = 1$.

-the random variable in the problem is the number of occurrences of each event:

$$v_{i1}, v_{i2}, \ldots, v_{ik}, \quad \sum_{j=1}^{k} v_{ij} = n_i .$$

-the p_{ij} are functions of $\underline{\theta}$ (continuous with continuous first and second derivatives).

-expectation is with respect to "j" -- i.e., $E[v_{ij}] = n_i p_{ij}$ -- the use of the "i" subscript is a convenience, we could consider the s trials as one big trial.

The following methods of estimating $\underline{\theta}$ are asymptotically equivalent, i.e.,

- (i) they are consistent
- (ii) they are asymptotically normal
- (iii) they have equal asymptotic variances

1. maximum likelihood
2. minimum χ^2
3. modified minimum χ^2
4. transformed minimum χ^2

1. __Maximum Likelihood Estimation:__

Maximize $K \prod_{i=1}^{s} \prod_{j=1}^{k} \left[p_{ij}(\underline{\theta})\right]^{v_{ij}}$ or $\sum_{i=1}^{s} \sum_{j=1}^{k} v_{ij} \ln p_{ij}(\underline{\theta})$

subject to $\sum_{j=1}^{k} p_{ij}(\underline{\theta}) = 1$ for each i, with respect to θ;

or actually solve the equations:

$$\sum_i \sum_j \frac{v_{ij}}{p_{ij}(\underline{\theta})} \frac{\partial p_{ij}}{\partial \underline{\theta}} = 0 \qquad [1]$$

(provided suitable regularity conditions are imposed as given in theorems 22, 23).

2. __Minimum χ^2__

$$\sum_i \sum_j \frac{(v_{ij} - n_i p_{ij})^2}{n_i\, p_{ij}}$$ is asymptotically distributed as χ^2 with $s(k-1)$ d.f.

The method is to minimize this expression with respect to $\underline{\theta}$ or to solve the equations:

$$-2 \sum_i \sum_j \frac{n_i(v_{ij} - n_i p_{ij})}{n_i p_{ij}} \frac{\partial p_{ij}}{\partial \underline{\theta}} - \sum \sum \frac{(v_{ij} - n_i p_{ij})^2}{n_i p_{ij}^2} \frac{\partial p_{ij}}{\partial \underline{\theta}} = 0 \qquad [2]$$

This set of equations could be expressed as:

$$\sum_i \sum_j \frac{v_{ij}}{p_{ij}(\underline{\theta})} \frac{\partial p_{ij}}{\partial \underline{\theta}} - \sum_i n_i \sum_j \frac{\partial p_{ij}}{\partial \underline{\theta}} + \underbrace{\frac{1}{2} \sum_i \sum_j \frac{(v_{ij} - n_i p_{ij})^2}{n_i p_{ij}^2} \frac{\partial p_{ij}}{\partial \underline{\theta}}}_{R} = 0$$

$$\sum_j p_{ij} = 1 \qquad \sum_j \frac{\partial p_{ij}}{\partial \underline{\theta}} = 0$$

therefore the second term equals 0

R = the third term $\xrightarrow[p]{} 0$ as $N \longrightarrow \infty$

Hence, equation [2] is the same as equation [1] except for R and thus it is seen that under suitable regularity conditions the solutions of equation [2] tend in probability to the solutions of equation [1].

3. Modified Minimum χ^2

Replace p_{ij} in the denominator of the regular minimum χ^2 equation with its estimated value, and then minimize with respect to $\underline{\theta}$:

$$\chi_m^2 = \sum_i \sum_j \frac{(v_{ij} - n_i p_{ij})^2}{v_{ij}}$$

subject to $\sum_j p_{ij} = 1$ for $i = 1, 2, \ldots, s$.

Comparing the two χ^2- methods

$$\chi^2 - \chi_m^2 = \sum_i \sum_j (v_{ij} - n_i p_{ij})^2 \left(\frac{1}{n_i p_{ij}} - \frac{1}{v_{ij}} \right)$$

$$= \sum_i \sum_j (v_{ij} - n_i p_{ij})^2 \left\langle \frac{v_{ij} - n_i p_{ij}}{n_i p_{ij} v_{ij}} \right\rangle$$

$$= \sum_i \sum_j \frac{(v_{ij} - n_i p_{ij})^3}{n_i p_{ij} v_{ij}} = R$$

R $\xrightarrow[p]{} 0$ faster than does χ^2 ------ therefore methods (2) and (3) are asymptotically equivalent.

The equations to be solved in this case are:

$$\sum \sum \frac{n_i (v_{ij} - n_i p_{ij})}{v_{ij}} \frac{\partial p_{ij}}{\partial \underline{\theta}} = 0 \qquad [3]$$

4. Transformed Minimum χ^2

Recall that if $\frac{\sqrt{n}(X_n - \mu)}{\sigma_n}$ is asymptotically normal, then $\frac{g(X_n) - g(\mu)}{\sigma\, g'(\mu)}$ is A.N. $\sigma = \frac{\sigma_n}{\sqrt{n}}$

and the asymptotic variance of $g(X_n)$ is $\sigma^2 [g'(\mu)]^2$.

Before writing the necessary equation, write

$$q_{ij} = \frac{v_{ij}}{n_i} \qquad \text{thus the } q_{ij} \text{ are random variables.}$$

$$q_{ij} \xrightarrow[p]{} p_{ij} \qquad E\left[q_{ij}\right] = p_{ij}$$

Thus the modified minimum χ^2 equation can be written:

$$\chi_m^2 = \sum_i \sum_j \frac{n_i(q_{ij} - p_{ij})^2}{q_{ij}}$$

The equivalent estimation procedure is thus to minimize

$$\sum_i \sum_j \frac{n_i\left[g(q_{ij}) - g(p_{ij})\right]^2}{q_{ij}\left[g'(q_{ij})\right]^2}$$

(the numberator should be $p_{ij}\left[g'(p_{ij})\right]^2$ but the difference occasioned by the modification $\longrightarrow 0$)

for g's which are continuous with continuous first and second derivatives, and with $g'(p_{ij})$ bounded away from zero.

This method will be useful if the transformation, $g\left[p_{ij}(\underline{\theta})\right]$ is simple.

The equations to be solved are thus:

$$\sum_i \sum_j \frac{n_i\left[g(q_{ij}) - g(p_{ij})\right]}{q_{ij}\left[g'(q_{ij})\right]^2} \; g'(p_{ij}) \; \frac{\partial p_{ij}}{\partial \underline{\theta}} = 0 \qquad [4]$$

<u>Summary</u>:

Method (1) [m.l.e.] -- is most useful if the p_{ij} can be expressed as products, i.e., $p_{ij} = p_i \tau_j$ as in problem 49.

Method (2) [minimum χ^2] -- most useful if p_{ij} can be expressed as a sum of probabilities, i.e., $p_{ij} = \alpha_i + \beta_j$.

Method (3) [modified χ^2] -- same as (2).

Method (4) [transformed χ^2] -- may be most useful in some special cases.

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \begin{matrix} n_1 \\ n_2 \\ n_3 \end{matrix}$$

Examples of minimum - χ^2 procedures:

1. Linear Trend: $n_i = 2 \qquad p = 3$

obs.	prob.	obs.	prob.	total
v_{11}	$p_{11} = p - \Delta$	v_{12}	$p_{12} = 1 - p + \Delta$	n_1
v_{21}	$p_{21} = p$	v_{22}	$p_{22} = 1 - p$	n_2
v_{31}	$p_{31} = p + \Delta$	v_{32}	$p_{32} = 1 - p - \Delta$	n_3

Problem is to estimate both p and Δ.

Using the modified minimum-χ^2 procedure:

$$\chi^2 = \left[v_{11} - n_1(p-\Delta)\right]^2\left[\frac{1}{v_{11}} + \frac{1}{v_{12}}\right] + \left[v_{21} - n_2 p\right]^2\left[\frac{1}{v_{21}} + \frac{1}{v_{22}}\right] + \left[v_{31} - n_3(p+\Delta)\right]^2\left[\frac{1}{v_{31}} + \frac{1}{v_{32}}\right]$$

$$-\frac{1}{2}\frac{\partial \chi^2}{\partial p} = a_1\left[v_{11} - n_1(p-\Delta)\right] + a_2\left[v_{21} - n_2 p\right] + a_3\left[v_{31} - n_3(p+\Delta)\right] = 0$$

$$\text{where } a_i = n_i\left(\frac{1}{v_{i1}} + \frac{1}{v_{i2}}\right)$$

$$-\frac{1}{2}\frac{\partial \chi^2}{\partial \Delta} = -a_1\left[v_{11} - n_1(p-\Delta)\right] + a_3\left[v_{31} - n_3(p+\Delta)\right] = 0$$

These two equations yield the following two equations which can be solved simultaneously for p and Δ:

$$p\sum n_i a_i + (a_3 n_3 - a_1 n_1)\Delta = \sum a_i v_{i1}$$

$$(a_3 n_3 - a_1 n_1)\,p + (a_1 n_1 + a_3 n_3)\Delta = a_3 v_{31} - a_1 v_{11}$$

Note: remember that this procedure is asymptotically equivalent to the m.l.e. -- therefore the asymptotic variance-covariance matrix can be found by evaluating

$\frac{1}{2}\frac{\partial^2 \chi^2}{\partial p^2}$; $\frac{1}{2}\frac{\partial^2 \chi^2}{\partial \Delta^2}$; $\frac{1}{2}\frac{\partial \chi^2}{\partial p\,\partial \Delta}$ at the expectations since $-\frac{\chi^2}{2}$ is the exponent of the asymptotic normal distribution.

$$\frac{1}{2}\frac{\partial \chi^2}{\partial p^2} = \sum a_i n_i \qquad \frac{1}{2}\frac{\partial^2 \chi^2}{\partial p\,\partial \Delta} = a_3 n_3 - a_1 n_1 \qquad \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \Delta^2} = a_1 n_1 + a_3 n_3$$

Recall $a_1 = n_1\left(\frac{1}{v_{11}} + \frac{1}{v_{12}}\right)$ $\quad E\left[v_{11}\right] = n_1(p-\Delta)$ $\quad E\left[v_{12}\right] = n_1(1-p+\Delta)$

Replacing the random variables in a_1 by their expectations yields

$$a_1 \xrightarrow[p]{} \left(\frac{1}{p-\Delta} + \frac{1}{1-p+\Delta}\right) = \frac{1}{(p-\Delta)(1-p+\Delta)}$$

Similarly:

$$a_2 \xrightarrow[p]{} \frac{1}{p(1-p)}$$

$$a_3 \xrightarrow[p]{} \frac{1}{(p+\Delta)(1-p-\Delta)}$$

thus the inverse of the A. V-COV matrix is

$$V^{-1} = \begin{pmatrix} a_1n_1 + a_2n_2 + a_3n_3 & -a_1n_1 + a_3n_3 \\ -a_1n_1 + a_3n_3 & a_1n_1 + a_3n_3 \end{pmatrix}$$

---

2. Logistic (Bio-assay) problem (Bergson)

-- applying greater dosages produces greater kill (or reaction)

s = general $\qquad n_i = 2$

$x_1, x_2, \ldots, x_s$ dosages

$n_1, n_2, \ldots, n_s$ receive dosages

$v_1, v_2, \ldots, v_s$ die (or react)

$n_1 - v_1, \ldots, n_s - v_s$ survive

$$p_i = \frac{1}{1+e^{-(\alpha+\beta x_i)}} \qquad 1 - p_i = \frac{e^{-(\alpha+\beta x_i)}}{1+e^{-(\alpha+\beta x_i)}}$$

$$\frac{1-p_i}{p_i} = e^{-(\alpha+\beta x_i)} \qquad q_i = \frac{v_i}{n_i} = \text{proportion dying or reacting}$$

$$-\ln\left(\frac{1-p_i}{p_i}\right) = \alpha + \beta x_i$$

Applying the transformed-χ^2 method to estimate α, β and using the transformation

$$g(x) = -\ln\left(\frac{1-x}{x}\right) = \ln\left(\frac{x}{1-x}\right)$$

$$g'(x) = \frac{1-x}{x}\;\frac{(1+x)-(x)}{(1-x)^2} = \frac{1}{x(1-x)}$$

The transformed-χ^2 is given by:

$$\chi_T^2 = \sum_{i=1}^{s} n_i \left[\frac{[g(q_i) - g(p_i)]^2}{q_i[g'(q_i)]^2} + \frac{[g(1-q_i) - g(1-p_i)]^2}{(1-q_i)\;[g'(1-q_i)]^2} \right]$$

$$= \sum_{i=1}^{s} n_i \left[[g(q_i) - g(p_i)]^2 \left(\frac{1}{q_i[g'(q_i)]^2} + \frac{1}{(1-q_i)\,[g'(1-q_i)]^2} \right) \right]$$

Putting in the values of g, g', we have

$$\chi_T^2 = \sum_{i=1}^{s} n_i \left[\ln\frac{q_i}{1-q_i} - (\alpha + \beta x_i)\right]^2 \left[\frac{[q_i(1-q_i)]^2}{q_i} + \frac{[q_i(1-q_i)]^2}{1-q_i} \right]$$

$$= \sum_{i=1}^{s} n_i \left[\ln\frac{q_i}{1-q_i} - (\alpha + \beta x_i)\right]^2 \left[q_i(1-q_i) \right]$$

Problem 50:

Consider rc sequences of n trials which may result in E or $\tilde{E}$ where in trial (ij)

$$\Pr[E] = \alpha_i + \beta_j \qquad \Pr[\tilde{E}] = 1 - \alpha_i - \beta_j \qquad \begin{matrix} i = 1,2,\ldots,r \\ j = 1,2,\ldots,c \end{matrix}$$

Set up equations to estimate α_i, β_j by

1. maximum likelihood.
2. minimum modified χ^2

   based on observations v_{ij}, $n - v_{ij}$.

Problem as stated produces $r+c$ equations in $r+c$ unknowns, but their matrix is singular -- i.e., the equations are not independent by virtue of the fact that the sum of the r equations for the α_i equals the sum of the c equations for the β_j.

Therefore, reduce the number of sequences to 4,

and add the restriction that $p_{22} = p_{12} + p_{21} - p_{11}$

$$\text{i.e.:}\quad p_{11} = \alpha_1 + \beta_1$$

$$p_{12} = \alpha_1 + \beta_2$$

$$p_{21} = \alpha_2 + \beta_1$$

$$p_{22} = \alpha_2 + \beta_2 = p_{12} + p_{21} - p_{11}$$

i.e., reparameterize and find the equations necessary to solve for the new parameters.

Problem 51:

Given s sequences of n_i trials may result in E or $\tilde{E}$ where in sequence i

$$\Pr\left[E\right] = e^{-\alpha_i \lambda} \qquad \Pr\left[\tilde{E}\right] = 1 - e^{-\alpha_i \lambda} \qquad (\alpha_i \text{ known; } i = 1,2,\ldots,s)$$

(let the number of E's observed be denoted by v_i).

Set up the equations to estimate λ by

1. maximum likelihood.

2. minimum modified χ^2.

and find the asymptotic variance of $\hat{\lambda}$.

(note: Ferguson discusses this problem in his Dec. 58 article in the Annals)

G. Minimax estimation:

Recall that $d(\underline{X})$ is a minimax estimate of $g(\underline{\theta})$ if

$$\sup_{\theta} R\left[d, \theta\right] \text{ is minimized by } d(\underline{X}).$$

$$R\left[d, \theta\right] = E\left[d(\underline{X}) - g(\underline{\theta})\right]^2 = \int \left[d(\underline{x}) - g(\underline{\theta})\right]^2 f(\underline{x})\, d\underline{x}$$

[see def. 24 on p. 54]

Methods to try to find minimax estimates:

1. Cramer-Rao

$$\sigma_d^2 \geq \frac{[1+b'(\theta)]^2}{nE\left[\frac{\partial \ln f(x)}{\partial \theta}\right]^2} = \frac{[1+b'(\theta)]^2}{nE\left[-\frac{\partial^2 \ln f(x)}{\partial \theta^2}\right]}$$

$$E\left[d(\underline{X}) - g(\underline{\theta})\right]^2 = E\left[d(\underline{X}) - E[d(\underline{X})] + E[d(\underline{X})] - g(\underline{\theta})\right]^2$$
$$= \sigma_d^2 + [b(\theta)]^2$$

$$R[d,\theta] \geq \frac{[1+b'(\theta)]^2}{nE\left[\frac{\partial \ln f(x)}{\partial \theta}\right]^2} + b^2(\theta) = k(\theta)$$

If we have an estimate $d(\underline{X})$ which actually has the risk $k(\theta)$ then any other estimator with risk less than $k(\theta)$ leads to a contradiction.

Example: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$

and we know $R(\bar{x}) = \frac{\sigma^2}{n}$

Lehmann shows that $\frac{[1+b'(\theta)]^2}{n/\sigma^2} + b^2(\theta) \leq \frac{\sigma^2}{n}$

which implies that $b(\theta) = 0$

Therefore $\bar{X}$ is a minimax estimator of μ.

2. This method based on the following theorem which will be stated without proof.

<u>Theorem 29</u>: If a Bayes estimator [def. 23] has constant risk [def. 25] then it is minimax.

$d(\underline{X})$ is constant risk if $E\left[d(\underline{X}) - g(\theta)\right]^2$ is constant.

$d(\underline{X})$ is Bayes if $d(\underline{X})$ minimizes $\int \left[d(\underline{x}) - g(\theta)\right] f(x, \theta)\, dG(\theta)$

where G is the "a priori" distribution of θ.

ref: Lehmann -- section 4, p. 19-21

<u>Example</u>: Binomial -- X is binomial $B(n, p)$.

Recall that we found that $d(X) = \frac{\sqrt{n}}{1+\sqrt{n}} \frac{x}{n} + \frac{1}{2(1+\sqrt{n})}$

is a constant risk estimator of p [see problem 31, p. 54].

If p has a prior Beta distribution

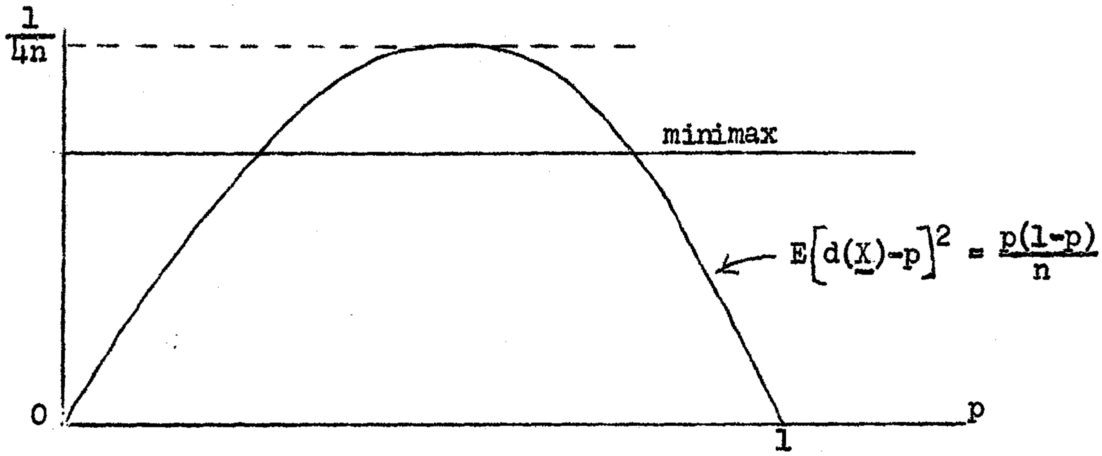
$$f(p) = K\, p^{\frac{\sqrt{n}}{2} - 1} (1 - p)^{\frac{\sqrt{2}}{2} - 1}$$

then $d(X)$ is Bayes;

hence the minimax estimate of the binomial parameter p is

$$\frac{\sqrt{n}}{1+\sqrt{n}} \frac{X}{n} + \frac{1}{2(1+\sqrt{n})} .$$

Graphically we have the following risk functions:



Problem 52: Compare $R_{\hat{d}}(p)$ and $R_{d_m}(p)$ for $n = 25$.

$$(\hat{d} = \frac{X}{n} ; \quad d_m = \text{minimax estimate})$$

H. Wolfowitz's Minimum Distance Estimation

$X_1, X_2, \ldots, X_n$ are observations from the distribution $F(x, \theta)$.

We are also given some measure of distance between 2 distributions $\rho(F, G)$.

e.g.: $$\rho(F, G) = \sup_{-\infty < x < \infty} \left| F(x) - G(x) \right|$$

$$\rho_1 (F, G) = \int \left[F(x) - G(x)\right]^2 d\frac{F(x) + G(x)}{2}$$

Also we have that the sample d.f., $F_n(x) = \frac{\text{no. of } X\text{'s} < x}{n}$ (see p. 97).

$\tilde{\theta}$ is a minimum distance estimate of θ if $\rho\left[F(x, \theta), F_n(x)\right]$ is minimized by choosing $\theta = \tilde{\theta}$.

Note: We want the whole sample d.f. to agree with the whole theoretical distribution -- not just the means or the variances agreeing.

In particular, if we use the "sup-distance" then $\hat{\theta}$ is that estimate of θ which yields

$$\min_{\theta} \sup_{-\infty < x < \infty} \left| F(x, \theta) - F_n(x) \right| .$$

<u>Remark:</u> $\tilde{\theta}$ is a consistent estimate of θ.

<u>Proof:</u> Suppose, for all sufficiently large n, $\tilde{\theta}$ differs from θ by more than ε.

Then for some δ

$$\sup \left| F(x, \tilde{\theta}) - F(x, \theta_o) \right| > \delta$$

and for some n_1

$$\sup \left| F_{n_1}(x) - F(x, \theta_o) \right| < \frac{\delta}{3}$$

We want to show that

$$\sup_{-\infty<x<\infty} \left| F_n(x) - F(x, \theta_o) \right| < \sup_{-\infty<x<\infty} \left| F_n(x) - F(x, \tilde{\theta}) \right| .$$

Let x' be the x such that

$$F(x', \tilde{\theta}) - F(x', \theta_o) = \delta$$

$$F_{n_1}(x') - F(x', \theta_o) < \frac{\delta}{3}$$

therefore $F(x', \tilde{\theta}) - F_{n_1}(x') > \frac{2}{3}\delta$

thus $\sup \left| F(x', \theta) - F_{n_1}(x') \right| > \frac{2}{3}\delta$

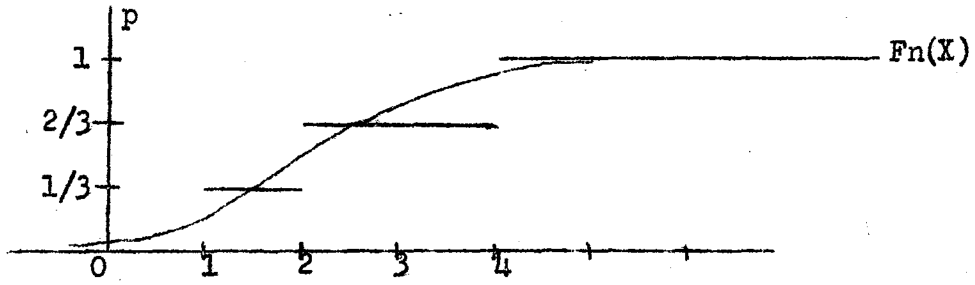
$$> \sup \left| F_n(x) - F(x, \theta_o) \right|$$

but this contradicts the fact that θ minimizes $\sup \left| F_n(x) - F(x, \theta) \right|$

therefore $\tilde{\theta}$ is consistent.

Example: Find the minimum distance estimate for μ given X_1, X_2, X_3 are $N(\mu, 1)$

$X_1 = 1$ $X_2 = 2$ $X_3 = 4$



Note: The "sup-distance" will obviously occur at one of the jump points on $F_n(x)$.

To find the minimum distance estimate of μ an itterative procedure was used, which started by guessing at μ and then finding $F(x, \mu)$ at each X_i.

μ	$F_n(x)$ =	X_1 .33	X_2 .67	X_3 1.00	sup
2.4	$z_i = x_i - \mu$ =	-1.4	-0.4	1.6	
	$\varphi(z_i)$ =	.081	.345	.945	.325 (.67 - .345)
2.3	z_i =	-1.3	-0.3	1.7	
	$\varphi(z_i)$ =	.097	.382	.955	.288
2.33	$\varphi(z_i)$ =	-	.371	.952	.299
2.29	$\varphi(z_i)$ =	-	.386	.956	.284 (.67 - .386)

Other values of μ were tried, but the min sup-distance = .284,

therefore $\tilde{\mu} = 2.29$. ($\bar{\bar{x}} = 2.33$)

Problem 53: Take 4 observations from a table of normal deviates (add an arbitrary factor if desired) and find the minimum distance estimate of μ.

Chapter V: TESTING OF HYPOTHESES -- DISTRIBUTION FREE TESTS

I. Basic concepts.

$X_1, X_2, \ldots, X_n$ have distribution function $F(x)$ -- continuous
-- absolute continuous
[density $f(x)$]

The hypothesis H specifies that $F \varepsilon \mathcal{F}_0$ (some family of d.f.)

Alternative H specifies that $F \varepsilon \mathcal{F} - \mathcal{F}_0$

Def. 33: A test is a function $\varphi(\underline{x})$ taking on values between 0 and 1

i.e., if $\varphi(\underline{x}) = 1$ reject H with probability 1
$= k$ " H " " k
$= 0$ " H " " 0

(note that this considers a test as a function instead of the usual consideration of regions)

Def. 34: A test is of size α if $E[\varphi(\underline{x})]$ for $F \varepsilon \mathcal{F}_0 \leq \alpha$

$$E[\varphi(\underline{x})] = \int_{-\infty}^{\infty} \varphi(\underline{x})\, dF(\underline{x})$$

(this says that the probability of rejecting H when it is true $\leq \alpha$)

Def. 35: Power of a test is $E[\varphi(\underline{x})]$ for $F \varepsilon \mathcal{F} - \mathcal{F}_0$ and is denoted $\beta_{\varphi}(F)$

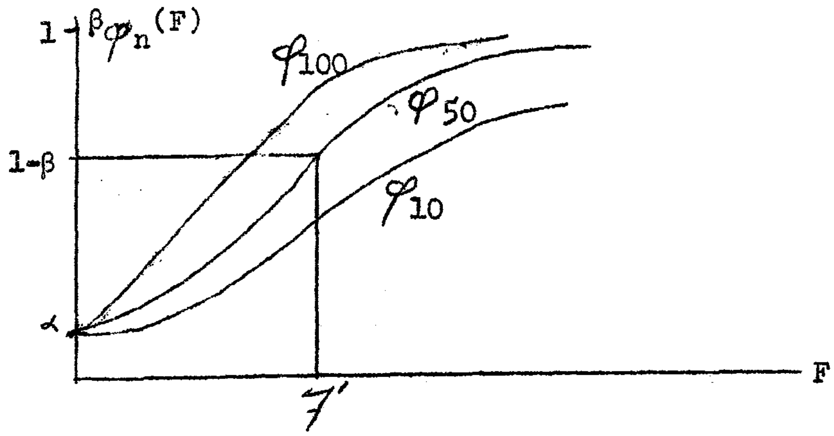
Def. 36: $\{\varphi_n\}$ is a consistent sequence of tests if for $F \varepsilon \mathcal{F} - \mathcal{F}_0$

$$\beta_{\varphi_n}(F) \longrightarrow 1 \quad \text{as } n \longrightarrow \infty$$

Def. 37: Index of a sequence of tests:

$N(\varphi_n, \mathcal{F}', \alpha, \beta)$ is the least integer n such that, given that

φ is of size α, $\beta_{\varphi_n}(F \mid F \varepsilon \mathcal{F}') \geq 1 - \beta$



Def. 38: Asymptotic Relative Efficiency (A.R.E.)

Let $p(\mathcal{F}', \mathcal{F}_0)$ be some measure of the distance of $\mathcal{F}'$ from $\mathcal{F}_0$

Let φ_n and $\tilde{\tilde{\varphi}}_n$ be two sequences of consistent tests

then the A.R.E. of $\tilde{\tilde{\varphi}}$ to φ_n is defined to be

$$\lim_{p \to 0} \frac{N(\varphi_n, \mathcal{F}', \alpha, \beta)}{N(\tilde{\tilde{\varphi}}_n, \mathcal{F}', \alpha, \beta)}$$ provided the limit exists

Problem 54: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ (σ known)

H: $\mu = 0$ alternative: $\mu > 0$

Consider two tests of H:

1. the mean test: reject H if $\bar{X} > \frac{z_{1-\alpha}}{\sqrt{n}} \sigma$

$$\left[\varphi(\bar{x}) = 1 \text{ if } \bar{X} > \frac{z_{1-\alpha}}{\sqrt{n}} \sigma\right.$$
$$= 0 \text{ if } \bar{X} \le \frac{z_{1-\alpha}}{\sqrt{n}} \sigma$$

2. the median test: [use the large sample distribution of the median, i.e the sample median is asymptotically normal with mean = the population median and variance = $\pi\sigma^2/2n$]

$\tilde{\varphi}$ test: reject H if the sample median $> z_{1-\alpha} \frac{\sigma\sqrt{\pi}}{\sqrt{2n}}$

a) find the index for each test for the alternative μ'

b) find the A.R.E., i.e.

$$\lim_{\mu' \to 0} \frac{N(\varphi, \mu', \alpha, \beta)}{N(\tilde{\varphi}, \mu', \alpha, \beta)}$$

II. Distribution Free Tests

refs: Siegel -- Non-parametric Statistics

Fraser -- Non-parametric Methods in Statistics

Savage -- Bibliography in the Dec. 1953 J.A.S.A.

A. Quantile and Median Tests:

Def. 39: Let the solution of the equation $F(\lambda_p) = p$ define the quantile λ_p.

If the solution is not unique, define $\lambda_p = \inf_x [F(x) = p]$

Given the problem to test H_o: $\lambda_p = \lambda_o$ against the alternative H_1: $\lambda_q = \lambda_o$

note: the alternative could be stated $\lambda_p \neq \lambda_o$ but this statement precludes consideration of the power of the test since the power in this case woul be undetermined by virtue of the unknown behavior of $F(x)$.

Test: X = the number of $x_1, x_2, \ldots, x_n \leq \lambda_o$

under H_o: X is $B(n, p)$

for the two sided test $(q \neq p)$ at level α using the normal approximation

$$\text{reject } H \text{ if: } \frac{|x - np| - \frac{1}{2}}{\sqrt{np(1-p)}} \geq z_{1-\alpha/2}$$

Problem 55: Compute the power of the test (using the normal approximation) as a function of q for $n = 100$ and $p = 0.5$

Example: for paired comparisons, to test if the median is equal to zero set up the series of observations $d_1, d_2, \ldots, d_n$, $d_i = x_i - y_i$

$$\text{where } x_i \leq y_i \iff d_i \leq 0$$

$$x_i > y_i \iff d_i > 0$$

This test that the median is zero is often referred to as the sign test, sinc X in this case is merely the number of d_i with a negative sign.

Def. 40: $\hat{\lambda}_p$, the sample quantile, is defined as the solution of $F_n(\hat{\lambda}_p) = p$

Theorem 30: $\hat{\lambda}_p$ has density function

$$f(\hat{\lambda}_p) = h(x) = \frac{n!}{\mu!(n-\mu-1)!} \left[F(x)\right]^{\mu} \left[1 - F(x)\right]^{n-\mu-1} f(x)$$

where $\mu = [np]$ which denotes the greatest integer in np

Proof: ref: Cramer p. 368

$$\Pr\left[\hat{\lambda}_p \leq x\right] = \Pr\left[\mu + 1 \text{ or more observations} \leq x\right]$$

$$= \sum_{j=\mu+1}^{n} \binom{n}{j} \left[F(x)\right]^{j} \left[1 - F(x)\right]^{n-j} = F(\hat{\lambda}_p)$$

To get the density h(x) [assuming that F(x) has a density f(x)] we differentiate the summation, getting the following two summations [from each part of the product that occurs in each term summed]

$$h(x) = \sum_{j=\mu+1}^{n} \binom{n}{j} j\left[F(x)\right]^{j-1} \left[1 - F(x)\right]^{n-j} f(x)$$

$$- \sum_{j=\mu+1}^{n} \binom{n}{j} (n - j) \left[F(x)\right]^{j} \left[1 - F(x)\right]^{n-j-1} f(x)$$

The corresponding terms [in $F^a(1 - F)^b$] cancel each other except for the first $(j = \mu + 1)$ term in the first summation which has no corresponding term in the second sum.

Thus: $$h(x) = \frac{n!}{\mu!(n-\mu-1)!} \left[F(x)\right]^{\mu} \left[1 - F(x)\right]^{n-\mu-1} f(x)$$

By the usual limiting process applied to density functions, $\hat{\lambda}_p$ is asymptotically normal, with mean λ_p and variance $\frac{1}{f^2(\lambda_p)} \frac{p(1-p)}{n}$

i.e. λ_p is $A.N.\left(\lambda_p, \frac{p(1-p)}{n\, f^2(\lambda_p)}\right)$

Problem 56: Let x have density $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

where $\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad \int_{-\infty}^{\infty} x\,f(x)\,dx = 0 \qquad \int_{-\infty}^{\infty} x^2 f(x)\,dx = 1$

and $f(x)$ is symmetric about 0.

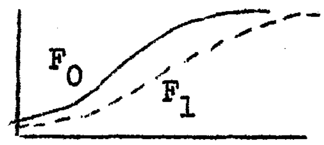
H_0: $\mu = 0$ $\qquad H_1$: $\mu = \mu_1 > 0$

Find the condition on $f(x)$ such that the A.R.E. of the median to the mean is greater than 1. Use large sample normal approximations for both tests.

Specialize this result to $f(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}} \quad -\infty < x < \infty$

III. One Sample Tests

H_0: $F = F_0$ completely specified $\qquad H_1$: $F = F_1 < F_0$



the smaller distribution has larger observations.

H_1': $F = F_1 \neq F_0$

problems: 1) goodness of fit (usually a wider problem since F_0 is usually not completely specified)

2) "slippage test"

3) combination of tests -- ref: A Birnbaum, JASA, Sept. 1954

4) randomness in time or space -- ref: Bartholomew, Barton, and David; Biometrika; circa 1955-6

Randomness in Time:

Assume: $\Pr[n$ events in $(0,t)$ (the time interval 0 to $t)] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$

i.e., is Poisson with parameter λt

Probability of event occuring in one time interval is independent of an occurrence in any other time interval.

$$0 \quad T_1 \quad T_2 \quad T_3 \quad \ldots \quad T_n \quad T$$

Let T_i be the time at which the i^{th} event occurred.

$$t_i = T_i - T_{i-1} \qquad T_i = \sum_{j=1}^{i} t_j$$

$$\Pr\left[t_i > t\right] = \Pr\left[\text{no event occurs in time } (0,t)\right] = e^{-\lambda t}$$

$f(t)$ = density of the $t_i = \lambda e^{-\lambda t}$ (the exponential distribution)

$f(t_1, t_2, \ldots, t_n) = \lambda^n e^{-\lambda \sum t_i}$ since the t's are independent

Distribution of $T_1, T_2, \ldots, T_n$ is obtained by transformation

$$T_1 = t_1$$

$$T_2 = t_1 + t_2 \qquad\qquad |J| = 1$$

$$\ldots$$

$$T_n = t_1 + t_2 + \ldots + t_n$$

$$f(T_1, T_2, \ldots, T_n) = \lambda^n e^{-\lambda T_n}$$

but $0 \leq T_1 \leq T_2 \leq \ldots \leq T_n$

Therefore, let us find the conditional distribution of $T_1, T_2, \ldots, T_n$ given that n events occurred in the fixed time interval $(0, T)$.

$f(T_1, T_2, \ldots, T_n ; n)$ = density of $T_1, T_2, \ldots, T_n$ and also the probability that no events occur in the time interval (T_n, T)

$$= \lambda^n e^{-\lambda T_n} \; e^{\lambda(T - T_n)}$$

$$= \lambda^n e^{-\lambda T}$$

$$f(n) = e^{-\lambda T} \frac{(\lambda T)^n}{n!}$$

therefore $f(T_1, T_2, \ldots, T_n \quad n) = n!\left(\frac{1}{T}\right)^n$

This distribution is the distribution of n ordered uniform independent random variables on $(0, T)$.

Ordered Observations:

$X_1, X_2, \ldots, X_n$ have density $f(x)$ and are independent.

$Y_1, Y_2, \ldots, Y_n$ are the ordered X's.

Joint density of the Y's is $n!\, f(y_1)\, f(y_2) \ldots f(y_n)$.

$$-\infty < y_1 \leq y_2 \leq \ldots \leq y_n < \infty$$

If the X's have a continuous distribution then we may disregard any equalities between the X's or the Y's.

The marginal distribution of the Y_i can be obtained by three methods, i.e.

1. Integration:

$$= n! \int_{y_{n-1}}^{\infty} f(y_n)\, dy_n \ldots \int_{y_i}^{\infty} f(y_{i+1})\, dy_{i+1} \times f(y_i) \times$$

$$\int_{-\infty}^{y_i} f(y_{i-1})\, dy_{i-1} \ldots \int_{-\infty}^{y_3} f(y_2)\, dy_2 \int_{-\infty}^{y_2} f(y_1)\, dy_1$$

note: -- the observations above y_i are constrained by the next lower observation, since this is an ordered sequence

-- the observations below y_i are constrained from above

integrating this we get

$g_i(y)$ = the marginal density of Y_i

$$= \frac{n!}{(n-i)!\,(i-1)!} \left[1 - F(y_i)\right]^{n-i} \left[F(y_i)\right]^{i-1} f(y_i)$$

2. By differentiation:

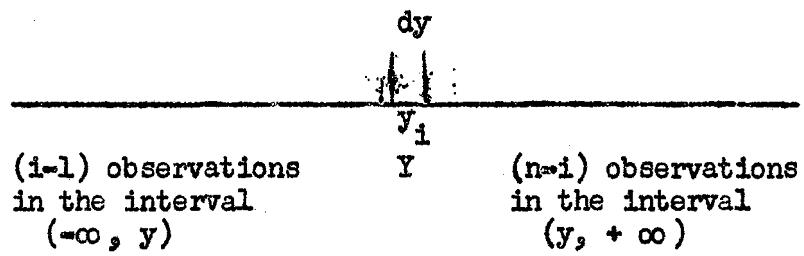
$$\Pr\left[Y_i < y\right] = \Pr\left[i \text{ or more observations fall to the left of } y\right]$$

$$= \sum_{j=i}^{n} \binom{n}{j} \left[F(y)\right]^j \left[1 - F(y)\right]^{n-j}$$

$$g_i(y) = \frac{\partial}{\partial y} \sum \binom{n}{j} \left[F(y)\right]^j \left[1 - F(y)\right]^{n-j}$$

$$= \frac{n!}{(n-i)!(i-1)!}\left[F(y)\right]^{i-1}\left[1 - F(y)\right]^{n-i}\ f(y)$$

3. Heuristic Method (Wilks):



The probability of this is essentially a trinomial distribution, therefore

$$g_i(y) = \frac{n!}{(i-1)!\ 1!(n-i)!}\left[F(y)\right]^{i-1}\left[f(y)\right]^{1}\left[1 - F(y)\right]^{n-i}$$

$$= \Pr\left[Y_i \text{ is in the interval } dy\right]$$

Example: In particular, if we have the uniform distribution:

$f(y) = 1 \qquad F(y) = y$

$g_i(y) = \frac{n!}{(n-i)!(i-1)!}\, y^{i-1}\ (1 - y)^{n-i}$ (which is a Beta distribution)

and $E\left[Y\right] = \frac{i}{n+1}$

We could also define the spacings: $S_i = Y_i - Y_{i-1} \qquad i = 1, 2, \ldots, n+1$

$$E\left[S_i\right] = \frac{1}{n+1} \qquad \sum_{i=1}^{n+1} S_i = 1$$

put $Y_{n+1} = 1 \qquad Y_0 = 0$

Problem 57: Show that each S_i has the same density.

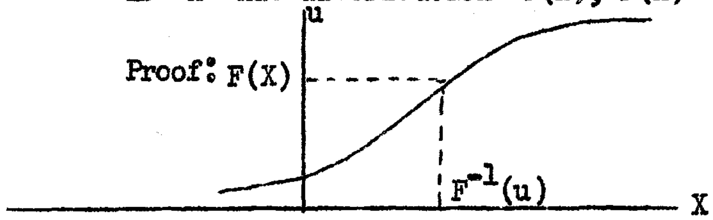
a) find this density.

b) find i) $E\left[\sum S_i^2\right]$ ii) $E\left[\sum_{i=1}^{n+1}\left|S_i - \frac{1}{n+1}\right|\right]$

Lemma: Probability Transformation:

If X has distribution $F(x)$, $F(X)$ has a uniform distribution on $(0, 1)$.

Proof:



define the inverse function as
$$F^{-1}(u) = \inf_{x} \left[F(x) = u \right]$$

$$\Pr\left[F(X) \leq u\right] = \Pr\left[X \leq F^{-1}(u)\right] = F\left[F^{-1}(u)\right] = u \qquad 0 \leq u \leq 1$$

One Sample Problem can be put in the following form by use of the probability transformation:

Given $X_1, X_2, \ldots, X_n$; let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the X's ordered increasingly and define $U_i = F_o(X_{(i)})$ $\quad i = 1, 2, \ldots n$

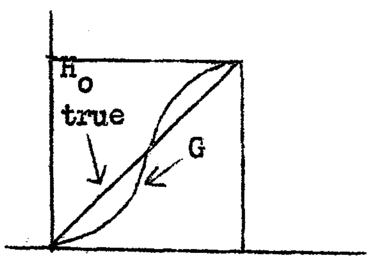
we have a set of ordered observations $U_1, U_2, \ldots, U_n$ on the interval $(0, 1)$ which under

H_o: has uniform distribution (with density $n!$)

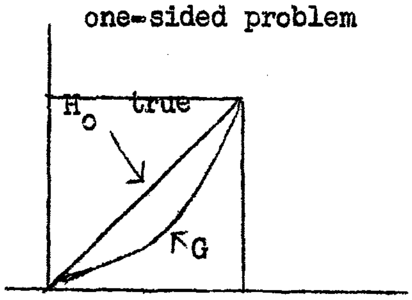
[or equivalently, that the original X's had d.f. $F_o(x)$; i.e., that the correct probability transformation was used]

and under H_1: has distribution $F_1\left[F_o^{-1}\right] = G$.

two-sided problem



one-sided problem



$G(u) \leq u$ -- with the strict inequality for some interval

Some of the Tests for One Sample Problems

1. $\bar{U}$ which is $A.N.(\frac{1}{2}, \frac{1}{12n})$

-- for the one-sided situation
-- consistent
-- reject H if $\bar{U}$ is "too large", i.e. if $\bar{U} > \frac{1}{2} + z_{1-\alpha} \left(\frac{1}{12n}\right)$

(for the above pictured situation)

or if $\bar{U}$ is "too small" in the converse situation [i.e., $G(u) > u$].

2. $-2 \sum_{i=1}^{n} \ln U_i$ which is χ^2 with 2n d.f. (ref. problem 6)

-- χ^2 is exact, not asymptotic
-- consistent
-- for the one-sided problem
-- used in combination problems
-- reject H if χ^2 is "too large"

3. $-2 \sum_{i=1}^{n} \ln (1 - U_i)$ -- also is χ^2 with 2n d.f.

-- for the one-sided problem
-- Pearson's counter to Fisher's advocating No. 2
-- consistent
-- reject H if χ^2 is "too small"

4. Distance Type Problem

Kolmogorov Statistic defined as follows

$$\left.\begin{aligned} D^+ &= \sup_{0 \le u \le 1} \left[F_n(u) - u\right] \\ D^- &= \sup_{0 \le u \le 1} \left[u - F_n(u)\right] \end{aligned}\right\} \text{ for the one-sided problem}$$

$$D = \sup_{0 \le u \le 1} \left|F_n(u) - u\right| \quad \text{for the two-sided problem}$$

-- is consistent

5. Related to the Kolmogorov statistic is

$$R^+ = \sup_{a \le u \le 1} \frac{Fn(u)-u}{u} \quad \text{one-sided problem}$$

$$R = \sup_{a \le u \le 1} \frac{|Fn(u)-u|}{u} \quad \text{two-sided problem}$$

-- not necessarily consistent
-- "a" is arbitrary but positive
-- derived by a Hungarian, Renyi
-- not of very great merit

6. $\omega_n^2 = \int_0^1 \left[F_n(u) - u\right]^2 \, du$

-- two-sided or one-sided problem
-- sort of a continuous analogue of χ^2
-- consistent
-- due to Von Mises and Smirnov

7. $\omega_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| S_i - \frac{1}{n+1} \right|$ $\qquad S_i = u_i - u_{i-1}$

-- ref: Sherman, Annals, 1950
-- one-sided or two-sided problem
-- consistent
-- is A.N. $\left(\frac{1}{e}, \frac{2e-5}{10e^2} \right)$

8. $\beta = \sum_{i=1}^{n+1} S_i^2$

-- one-sided or two-sided problem
-- due to Moran
-- is consistent
-- $\frac{n\beta}{2} - 1$ is A.N.(0, 1)

9. U_1 (Wilkinson Combination Procedure)

-- one-sided problem
-- generally not consistent

Problem 58:

a) Find the test based on U_1 for the set of alternatives $G(u) > u$ with $g(u) = G'(u)$

b) Find the power of the test for the alternative $G(u) = u^k$ $\quad 0 < k < 1$

c) Find the limiting power as $n \longrightarrow \infty$ (if the limit is 1, the test is consistent)

10. χ^2

11. Neyman's smooth tests
-- discovered by Neyman about 1937, but never generally used
-- ref: Neyman, Skandinavisk Aktuarietidskrift, 1937
Pearson - Biometrika, 1938
David - Biometrika, 1938

Problem 59: $X_1, X_2, \ldots, X_n$ are independent with d.f. $F(x)$, density $f(x)$

let $R = \max_{1 \leq i \leq n} X_i - \min_{1 \leq i \leq n} X_i$

a) Find the distribution of R.

b) Suppose $F(x)$ has density of the form $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

where $\int f(x)\,dx = 1 \qquad \int x^2 f(x)\,dx = 1$

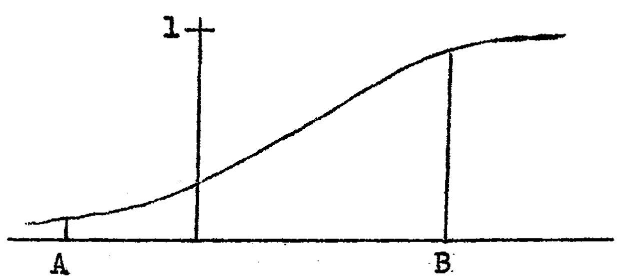
show that $E[R] = k(f, n)\sigma$.

<u>Theorem 31</u>:

If F is continuous, $F_n \xrightarrow[p]{} F$ for all x.

<u>Proof</u>: We want to show that given ε, δ we can find N such that for $n > N$

$$\Pr\left[\,|F_n(x) - F(x)| < \varepsilon \text{ for all } x\right] > 1 - \delta$$



Let A, B be such that

$F(A) < \varepsilon/2$

$1-F(B) < \varepsilon/2$

Pick out points $X_{(1)}, X_{(2)}, \ldots, X_{(k-1)}$ in (A, B) such that $F[X_{(i)}] - F[X_{(i-1)}] <$ which we can do because of uniform continuity.

Now set $A = X_{(o)} \qquad B = X_{(k)}$

Consider a sample of size n. Let $n_i = \#\,X_j$ that fall in the i^{th} interval $(x_{(i-1)}, x_{(i)})$.

$n_o, n_1, \ldots, n_{k+1}$ are multinomial with probabilities $p_i = F[X_{(i)}] - F[X_{(i-1)}]$

Now, $\displaystyle\sum_{i=o}^{k+1} \frac{(n_i - np_i)^2}{np_i}$ has a χ^2 distribution with $k+1$ d.f.

We can find an M such that $\Pr\left[\chi^2_{k+1} > M\right] < \delta$

or $$\Pr\left[\chi^2_{k+1} < M\right] > 1 - \delta$$

Since the above sum is less than M, each term and also its square root are certainly less than M, therefore:

$$\Pr\left[\frac{|n_i - np_i|}{\sqrt{np_i}} < M \text{ for each } i = 0, 1, \ldots, k+1\right] > 1 - \delta$$

which could be written $$\Pr\left[\frac{\left|\frac{n_i}{n} - p_i\right|}{\sqrt{p_i}} < \frac{M}{\sqrt{n}} \text{ for each } i = 0, 1, \ldots, k+1\right] > 1 - \delta$$

Recall that $$F_n\left[x_{(i)}\right] = \frac{\# x_j \leq x_{(i)}}{n} = \frac{\sum_{j=o}^{i} n_j}{n}$$

Now $$\left| F_n\left[x_{(i)}\right] - F\left[x_{(i)}\right]\right| = \left|\sum_{j=o}^{i}\frac{n_j}{n} - \sum_{j=o}^{i} p_j\right| = \left|\sum_{j=o}^{i}\left(\frac{n_j}{n} - p_j\right)\right| \leq \sum_{j=o}^{i}\left|\frac{n_j}{n} - p_j\right|$$

Choose n so large that $\frac{(k+2)M\sqrt{p_i}}{\sqrt{n}} < \frac{\varepsilon}{2}$ for all i

Hence, if n is chosen this large, with probability $1 - \delta$ the following relationship will hold:

$$\left| F_n\left[x_{(i)}\right] - F\left[x_{(i)}\right]\right| < \frac{(i+1)M\sqrt{p_i}}{\sqrt{n}} < \frac{\varepsilon}{2}$$

Consider x lying between $x_{(i-1)}$ and $x_{(i)}$

$$-\varepsilon = -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} \quad \overset{*}{\leq}$$

$$F\left[x_{(i)}\right] - F\left[x_{(i)}\right] + F\left[x_{(i-1)}\right] - F_n\left[x_{(i)}\right] =$$

$$F\left[x_{(i-1)}\right] - F_n\left[x_{(i)}\right] \leq F(x) - F_n(x) \leq F\left[x_{(i)}\right] - F_n\left[x_{(i-1)}\right]$$

$$\overset{**}{\leq} F\left[x_{(i-1)}\right] + \frac{\varepsilon}{2} - F_n\left[x_{(i-1)}\right]$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

with probability $1 - \delta$

~~* utilizing the following relationships:~~

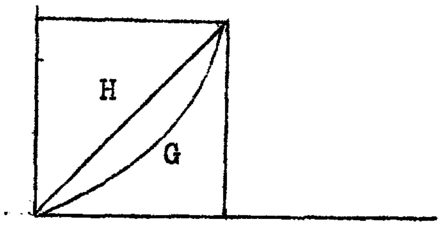
** making use of the following relationship:

$-\frac{\varepsilon}{2} < F[x_{(i)}] - F_n[x_{(i)}] < \frac{\varepsilon}{2}$ $\qquad$ $F[x_{(i)}] - F[x_{(i-1)}] \leq \frac{\varepsilon}{2}$

$-\frac{\varepsilon}{2} < F[x_{(i-1)}] - F[x_{(i)}] < 0$

Therefore the theorem holds.

Example: Referring to test No. 4 given on page 122.



$F_n \xrightarrow{p} F$

If H is true, $F_n \longrightarrow$ uniform, thus for n sufficiently large $|F_n(u) - u| < \varepsilon$.

If G is true, $|F_n(u) - G(u)| < \varepsilon$ and thus $|F_n(u) - u| > \varepsilon$.

Therefore, the test based on $D = \sup |F_n(u) - u|$ will reject H with probability tending to 1;

hence the test based on D is consistent.

<u>Problem 59 (Addendum)</u> (see p. 124):

c) Find the distribution of R for

1) $X_1, X_2, \ldots, X_n$ uniform on $(0, 1)$

2) $X_1, X_2, \ldots, X_n$ with exponential density $f(x) = a\,e^{-ax}$

note: in the uniform case U_1 and R are dependent, in the exponential case they are independent since the upper limit of the observations is ∞

<u>Remark</u>: The Kolmogorov statistics, D_n, D^+, and D^- are in fact invariant under the probability integral transform.

<u>Proof</u>: we have to show that $\sup_{-\infty<x<\infty} |F_n(x) - F(x)| = \sup_{0 \leq u \leq 1} |F_n(u) - u|$

$$\sup_{0 \leq u \leq 1} |F_n(u) - u| = \sup_{0 \leq F(x) \leq 1} |F_n[F(x)] - F(x)|$$

Now: $F_n(x) = \frac{i}{n}$ $\quad Y_i \leq x \leq Y_{i+1}$ where the Y's are the ordered observations $X_1, X_2, \ldots, X_n$

$$F_n[F(x)] = \frac{i}{n} \quad \text{where } F(Y_i) \leq F(x) \leq F(Y_{i+1})$$

Therefore:

$$\sup_{0 \leq u \leq 1} |F_n(u) - u| = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

Distributions of D:

When H is true, and $U_1, U_2, \ldots, U_n$ are uniform, then

a) $\lim_{n \to \infty} \Pr[\sqrt{n}\, D_n^{\mp} < z] = 1 - e^{-2z^2}$ $\quad 0 \leq z < \infty$ (due to Smirnov)

-- Finite distribution of $D^{\mp}$ given by Z. Birnbaum + Tingey in the Annals of Math. Stat.
-- Tabled by Miller in JASA, 1956, pp. 113-115.

b) $L(z) = \lim_{n \to \infty} \Pr[\sqrt{n}\, D_n^{\mp} < z] = 2 \sum_{m=1}^{\infty} (-1)^{m+1} e^{-2m^2 z^2}$ $\quad 0 \leq z < \infty$ (due to Kolmogorov)

-- Tabled by Smirnov in the Annals of Math. Stat., 1948

The simplest proof of both results is due to Doob in the Annals of Math. Stat., 1949

Some tables on the finite distribution of D_n are given by Massey in JASA, 1949, pp. 68-77.

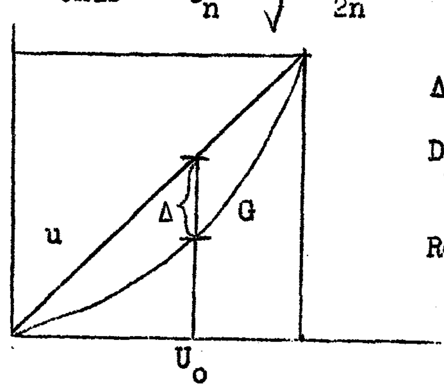
Consider now that H_1 is true, i.e., that U has d.f. G(u):

D^- test is to reject H if $D_n^- > \varepsilon_n$ where

$$\Pr[D_n^- > \varepsilon_n] = \alpha$$

$$\Pr[\sqrt{n}\, D_n^- > \sqrt{n}\, \varepsilon_n] = e^{-2n\varepsilon_n^2} = \alpha$$

thus $\varepsilon_n = \sqrt{\frac{|\ln \alpha|}{2n}}$



Δ = max. difference between u, G(u)

$$D_n^- = \sup_{0 \leq u \leq 1} [u - F_n(u)]$$

Reject H if $D_n^- > \varepsilon_n$

Suppose H_1 is ture so that F_n is the sample d.f. from $G(u)$ with maximum $|u - G(u)| = \Delta$ at $u = u_o$.

Then:
$$\Pr\left[D_n^- > \varepsilon_n\right] \geqslant \Pr\left[u_o - F_n(u_o) > \varepsilon_n\right]$$
$$\geqslant \Pr\left[F_n(u_o) - u_o < -\varepsilon_n\right]$$
$$\geqslant \Pr\left[n\, F_n(u_o) < n(u_o - \varepsilon_n)\right]$$

But $F_n(u_o)$ is a binomial random variable with expectation $(u_o - \Delta)$ and

$$\text{Binomial probability} = \sum_{k=o}^{n(u_o-\varepsilon_n)} F(k;\ n,\ u_o - \Delta)$$

<u>Problem 60:</u> Find the bound on the power of the D_n^- test where $G(u) = u^2$

Using the normal approximation given an explicit form for this bound in terms of

$$n,\ \alpha,\ \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

<u>Test No. 5:</u> Renyi Statistic

$$R^+ = \sup_{a \leq u \leq 1} \frac{F_n(u)-u}{u}$$

$$R = \sup_{a \leq u \leq 1} \frac{|F_n(u)-u|}{u}$$

Limiting distribution:

$$\lim_{n \to \infty} \Pr\left[\sqrt{n}\, R^+ < z\right] = \sqrt{\frac{2}{\pi}} \int_0^{z\sqrt{\frac{a}{1-a}}} e^{-t^2/2}\, dt$$

For the distribution of R and further discussion see an article by Renyi in Acta Mathematica (Magyar), 1953.

<u>Test No. 6:</u> $$\omega_n^2 = \int_0^1 \left[F_n(u) - u\right]^2 du = \int_{-\infty}^{\infty} \left[F_n(x) - F(x)\right]^2 dF(x)$$

$$= \int_0^1 \left[F_n^2(u) - 2uF_n(u) + u^2\right] du$$

$$F_n(u) = \frac{i}{n} \qquad u_i \leq u \leq u_{i+1}$$

e.g., $F_n(u)$ is flat for an interval

$$= \sum_{i=1}^{n} \left\{ \int_{u_i}^{u_{i+1}} \left[\left(\frac{i}{n} \right)^2 - 2u \frac{i}{n} \right] du \right\} + \int_0^1 u^2 \, du$$

recall: $u_{n+1} = 1 \qquad u_o = 0$

$$= \sum_{i=1}^{n+1} \left(\frac{i}{n} \right)^2 (u_{i+1} - u_i) - \frac{2}{2} \sum_{i=1}^{n} \frac{i}{n} (u_{i+1}^2 - u_i^2) + \frac{1}{3}$$

$$= 1 + \frac{1}{n^2} \sum_{i=1}^{n} (1 - 2_i) u_i - \frac{1}{n} \left(-\sum_{i=1}^{n} u_i^2 \right) - 1 + \frac{1}{3}$$

$$\omega_n^2 = \sum_{i=1}^{n} \frac{u_i^2}{n} - 2 \sum_{i=1}^{n} \frac{i u_i}{n^2} + \sum_{i=1}^{n} \frac{u_i}{n^2} + \frac{1}{3}$$

Cramer shows that: $\omega_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left[u_i - \frac{2i-1}{2n} \right]^2$

Tables of $\lim_{n \to \infty} \Pr\left[n\omega^2 < z \right]$ have been given by T. W. Anderson and Darling in in the Annals of Math. Stat., 1952, p. 206.

<u>Problem 61</u>: Find $E\left[\omega_n^2\right]$

<u>Approaches to Combining Probabilities</u>:

<u>Example</u>: The following probabilities are from a one-sided t-test:

p	F	p	F
.026	.1	.76	.6
.115	.2	.81	.7
.27	.3	.89	.8
.36	.4	.92	.9
.75	.5	.98	1.0

Under the basic hypothesis the p's are uniform on (0, 1)

Alternative tests:

1. $\chi^2_{(20)} = -2 \sum \ln U_i = 18.274$ P = 0.57 accept H

2. To test alternatives of the form $G(u) > u$ the Kolmogorov-Smirnov test statistic is $D_n^+ = F_n - u$.

$$\sup D_{10}^+ = 0.085$$

From Miller's tables: $\Pr\left[D_{10}^+ > .342\right] = .15$

3. $\bar{U} = .588$

$$z = \frac{\bar{U} - \frac{1}{2}}{\sqrt{\frac{1}{12n}}} = \frac{.088}{\sqrt{\frac{1}{120}}} = 0.964$$

$\Pr\left[z < .964\right] = .67$

Since $E\left[\bar{U}\right]$ under $H_1 < E\left[\bar{U}\right]$ under H_0 we will reject H if $z < z_\alpha$

therefore we cannot reject H_0.

4. Against two-sided alternatives

$D = .35$ (.75 - .40)

See Massey's tables for small sample size probabilities.

Using the large sample approximation:

$$\Pr\left[\sqrt{n}\, D > z\right] = 2 \sum_{m=1}^{\infty} (-1)^m e^{-2m^2 z^2} = L(z)$$

$$\Pr\left[D > \frac{.35}{\sqrt{10}}\right] = L(.35)$$

Example:

In the following table:

-- the X_i are taken from a table of N(0, 1) normal deviates.

-- the $X_{(i)}$ are the ordered X_i

-- the $U_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X_{(i)}} e^{-t^2/2}\, dt = \Pr\left[z < X_i\right]$

-- the S_i are the spacings between the U_i

X_i	$X_{(i)}$	U_i	S_i	$\frac{1}{n+1}$
			.35	.09
.42	-0.38	.35		
			.14	.09
-0.02	-0.02	.49		
			.04	.09
.88	.08	.53		
			.01	.09
.40	.09	.54		
			.10	.09
1.76	.37	.64		
			.02	.09
.09	.40	.66		
			.00	.09
.08	.42	.66		
			.05	.09
1.12	.88	.71		
			.16	.09
-0.38	1.12	.87		
			.09	.09
.37	1.76	.96		
			.04	.09

Under H_0: the X's are N(0, 1).

The test statistic is:

$$\omega_n = \frac{1}{2} \sum_{i=0}^{10} \left| S_i - \frac{1}{n+1} \right| = 0.38$$

Ignoring the slight negative correlation between the S_i, one could use a normal approximation with:

$$E\left[\omega_n\right] = \frac{1}{e} = 0.37 \qquad V(\omega_n) = \frac{2e-5}{10e^2} = (0.07)^2$$

Example:

If you want to test H_1: the X's are $\chi^2_{(5)}$ then you should use the probability transformation:

$$U = \frac{1}{2^{5/2}\Gamma(\frac{5}{2})} \int_0^X e^{-t/2}\, t^{\frac{5}{2}-1}\, dt$$

Combination of Test Probabilities:

H_o: U is uniform

H_1: U has distribution $G(u) > u$ (i.e., the observations tend to be smaller)

ref: A. Birnbaum, JASA, September 1954

He considers a comparison of the following tests:

1- Fisher: $-2 \sum \ln U_i$

2- Pearson: $-2 \sum \ln (1 - U_i)$

-- found unsatisfactory for most applications.

3- U_1 -- not consistent, but is this important?

Birnbaum concluded that $-2 \sum \ln U_i$ was better for the one particular case of testing normal means.

Other possible tests: D, ω_n, $\bar{\bar{U}}$; have not been studied in this light.

Goodness of Fit Tests:

H_o: $F = F_o$ completely specified

-- The test usually involves estimation of parameters.

-- The only completely worked out theory is for the χ^2-test.

For other suggestions see:

-- F. David, Biometrika, 1938-9

-- Kac, Kiefer, and Wolfowitz, Annals of Math. Stat., June 1955. They present a Monte Carlo derived distribution of ω_n^2 and D_n for the case of testing H: X's are $N(\mu, \sigma^2)$ where μ, σ^2 are estimated by $\bar{x}$, s^2 working with n=25, n=100.

For consideration of one-sided tests see: Chapman, Annals of Math. Stat., 1958

-- The D_n^+ test is a "minimax" test (among Fisher, Pearson, D_n^+, ω_n^2, $\bar{\bar{U}}$) of the one-sided hypothesis H_o: $F = F_o$ versus H_1: $F = F_1 < F_o$.

i.e., "minimax" in the sense that it has minimum power to pick up easy-to-detect alternatives, maximum power to pick up hard-to-detect alternatives.

IV. Two Sample Problems:

$X_1, X_2, \ldots, X_m$ have d.f. $F(x)$

$Y_1, Y_2, \ldots, Y_n$ have d.f. $G(y)$

H_o: $F = G$ $\qquad$ H_1: $F < G$ or $F > G$

H_1': $F \neq G$

In the parametric case one could use the normal approximation and the two-sample t-test on the means, but these hypotheses are somewhat wider.

Partial list of tests:

1. Median Test
2. Runs Test
3. Wilcoxen's Test (also called the Mann-Whitney test)
4. Kolmogorov-Smirnov D-test
5. Ven der Waerden's X-test (or Terry's C-test)

All we need for these tests is to be able to order the observations; magnitudes are not important; e.g. X X X X Y Y Y X Y Y

Test 1: For the median test set up a 2x2 table classifying the observations as above or below the median of the combined sample. For example:

	Below	Above	Total
X:	4	1	m
Y:	1	4	n
Total:	$\frac{m+n}{2}$	$\frac{m+n}{2}$	m + n

Test Statistics is the usual χ^2 for 2x2 contingency tables with one d.f.

Test 2: For the runs test, set

r = number of runs of X's and of Y's in the combined sample (in our example r = 4)

H_o is rejected if $r < r_o$.

If H_o is true, then the X's and Y's are intermingled and the value of r will be "large"; if H_1 is true, then r will be "small".

Test 3: For the Wilcoxen (Mann-Whitney) test, we define:

$$u_{ij} = 1 \quad \text{if} \quad X_i < Y_j$$
$$= 0 \quad \text{otherwise}$$

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij} \quad \text{in our example } U = 22 \ (4+4+4+5+5)$$

Wilcoxen's original test was based on R_i^x

where R_i^x = the sum of the ranks of X_i in the combined sample ordering from the smallest.

Similarly R_j^y = the sum of the ranks of Y_j.

Problem 62: Prove: $U = mn + \frac{m(m+1)}{2} - \sum_{i=1}^{m} R_i^x = \sum_{j=1}^{n} R_j^y - \frac{n(n+1)}{2}$

Test 4: Kolmogorov-Smirnov define D for the two-sample problem as:

$$D_{mn} = \sup_{-\infty < x < \infty} \left| F_m(x) - G_n(x) \right|$$

Test 5: For Van der Waerden's X-test we

let $$\phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} \, dt$$

and define: $$\psi(u) = \phi^{-1}(u)$$

then the test statistic is: $$X = \sum_{i=1}^{m} \psi\left(\frac{R_i^x}{m+n-1} \right)$$

Problem 63: Let Q = number of Y's which exceed $\max(X_1, X_2, \ldots, X_m)$.

a) Find the distribution of Q in the general case.

b) Specialize the result in (a) when H_0 is true.

c) Find the limiting distribution of Q for case (b) as $m \to \infty$, $n \to \infty$, $\frac{m}{n} \to \lambda$.

Runs Test (Test No. 2):

Ref.: Mood Chap. 16

r = no. of runs of X's or Y's in the ordered combined sample.

There are two cases to be considered—that of an even and that of an odd number of runs.

Even case: The number of arrangements of m X's and of n Y's that have the property of giving rise to $2k$ runs can be found by a simple generating function device.

Again, there are two possibilities, i.e. starting with a run of X's or of Y's, so starting with either, we must multiply the end result by 2 since the two starts are symmetric.

Starting with the X's, they are divided into k groups, all non-zero. To find the number of ways of doing this, consider the coefficient of t^m in the expansion of $(t + t^2 + t^3 + \dots)^k$.

$$(t + t^2 + t^3 + \dots)^k = \left(\frac{t}{1-t}\right)^k = t^k\,(1 - t)^{-k}$$

$$= t^k\left(1 + kt + \frac{(-k)(-k-1)}{2}t^2 + \dots\right)$$

$$= t^k \sum_{j=0}^{\infty} t^j \frac{(k+j-1)!}{j!\,(k-1)!} = \sum_{j=0}^{\infty} \binom{k+j-1}{j} t^j$$

The coefficient of t^m is found when $j = m - k$, and

$$= \frac{(k + m - k - 1)!}{(m - k)!\,(k - 1)!} = \binom{m - 1}{k - 1}$$

Similarly the number of arrangements of the n Y's into k non-zero groups is

$$\binom{n - 1}{k - 1}$$

Therefore, the number of arrangements of X's and Y's with $2k$ runs is

$$2\binom{m - 1}{k - 1}\binom{n - 1}{k - 1}$$

The total number of arrangements possible with m X's and n Y's is $\frac{(m + n)!}{m!\,n!}$.

Therefore:

$$Pr\left[r = 2k\right] = \frac{2\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{m}}$$

Odd Case: (r = 2k + 1) For the case when the number of runs is odd, i.e., to determine $Pr\left[r = 2k + 1\right]$, the argument is similar, but we start and end with either X or Y (instead of starting with one end and ending with the other), therefore

$$Pr\left[r = 2k + 1\right] = \frac{\binom{m-1}{k}\binom{n-1}{k-1} + \binom{m-1}{k-1}\binom{n-1}{k}}{\binom{m+n}{m}}$$

$$E(r) = \frac{2mn}{m+n} + 1 \approx 2N\alpha\beta$$

$$V(r) = \frac{2mn(2mn - m - n)}{(m+n)(m+n-1)} \approx 4N\alpha^2\beta^2$$

Where: $N = m + n$ $m = N\alpha$ $n = N\beta$ $\alpha + \beta = 1$

By Sterling's approximation methods, we can show that

$$\lim_{N \to \infty} \frac{r - 2N\alpha\beta}{2\alpha\beta\sqrt{N}} \text{ is } N(0,1)$$

Test: H_o is rejected if $r < r_o$

r_o can be determined from the left hand tail of the normal approximation or from tables in

1. Swed + Eisenhart, 1943 Annals of Math. Stat.
2. Siegel, Table F
3. Dixon + Massey, Table 11

Wilcoxen (Mann-Whitney) Test (Test No. 3):

$$U = \sum_i \sum_j u_{ij} \qquad \text{where} \qquad u_{ij} = 1 \text{ if } X_i < Y_j$$

$$= 0 \text{ otherwise}$$

If H is true, $E(U) = \sum_i \sum_j E(u_{ij}) = \frac{mn}{2}$

For the general case, assume $Pr[Y > X] = p$

$$= \int_{-\infty}^{\infty} Pr[Y > x \mid X = x] \, dF(x)$$

$$= \int_{-\infty}^{\infty} [1 - G(x)] \, dF(x)$$

$$= E[\{1 - G(x)\} \mid F(x)]$$

In this general case $E(U) = m\, n\, p$

$$E(U^2) = \sum_i \sum_j E(u_{ij}^2) \quad (1)$$

$$+ \sum_i \sum_{j \neq b} \sum E(u_{ij}, u_{ib}) \quad (2)$$

$$+ \sum_{i \neq a} \sum \sum_j E(u_{ij}, u_{aj}) \quad (3)$$

$$+ \sum_{i \neq a} \sum \sum_{j \neq b} \sum E(u_{ij}, u_{ab}) \quad (4)$$

To evaluate this, we can examine each part separately, e.g.:

(1) $E(u_{ij}^2) = p$ mn such terms

(4) $E(u_{ij}, u_{ab}) = E(u_{ij}) \, E(u_{ab}) = p^2$ m(m-1)n(n-1) such terms

(2) $E(u_{ij}, u_{ib}) = Pr[Y_j \quad X_i, Y_b \quad X_i]$

$$= \int_{-\infty}^{\infty} Pr[Y_j > x, Y_b > x \mid X_i = x] \, dF(x)$$

$$= \int_{-\infty}^{\infty} [1 - G(x)]^2 \, dF(x)$$

$$= E[\{1 - G(x)\}^2 \mid F]$$ mn(n-1) such terms

(3) by similar argument $E(u_{ij}, u_{aj}) = \int_{-\infty}^{\infty} F^2(y) \, dG(y)$

$$= E[F^2(Y) \mid G]$$ mn(m-1) such terms

Thus:

$$V(U) = mnp + mn(m-1)(n-1)p^2 + mn(n-1)\,E\left[(1-G)^2 \,\middle|\, F\right] + mn(m-1)\,E\left[F^2 \,\middle|\, G\right] - m^2n^2p^2$$

Exercise: If H is true, verify that $E(F^2) = E(1 - G)^2 = \frac{1}{3}$

and then $Var(U) = \frac{mn}{12}(m + n + 1)$

Mann-Whitney in their further work on Wilcoxen's test proved that

$$\frac{U - \frac{mn}{2}}{\sqrt{\frac{mn}{12}(m+n+1)}}$$

is $AN(0,1)$. This they found by discovering a recursion formula to get all the moments of the distribution of U, then observing that the limits of the moments, as $m \to \infty$, $n \to \infty$ were the moments of the normal distribution.

U -- may be used for one-sided or two-sided tests.

-- the most complete tables are given by Hodges + Fix, Annals of Math. Stat., 1955.

Problem 64:

Take 10 observations of (a) X which is $N(0,1)$ and

(b) Y which is $N(1,1)$

Apply each of the five tests to the data to obtain tests for

H_o: $F = G$ against H_1: $F \neq G$.

Problem 65:

Let $a_1, a_2, \ldots, a_{k-1}$ be fixed points.

Let $X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ be independent observations from $F(x)$, $G(y)$.

Define: $z_i = F_m(a_i) - G_n(a_i)$

$$Z = \sum_{i=1}^{k-1} z_i$$

(a) Find $E(Z)$, $Var(Z)$ in general and for the case $F = G$.

denote: $F(a_i) = f_i$ $G(a_i) = g_i$

(b) Find E(Z), Var(Z) for the case when

$$F(u) = u \qquad G(u) = 0 \qquad 0 \leq u < b \qquad 0 \leq b \leq \frac{1}{2}$$

$$= 1 + \frac{u - 1 + b}{1 - 2b} \qquad b \leq u \leq 1-b$$

$$= 1 \qquad 1-b < u \leq 1$$

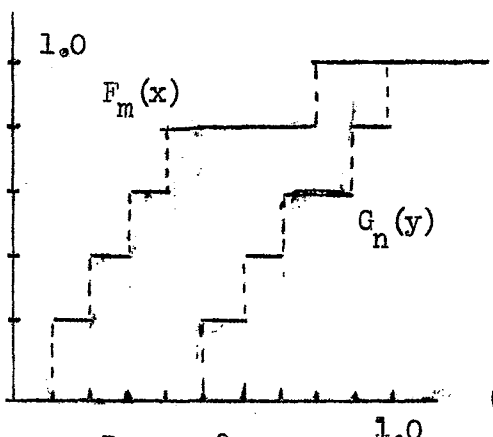
In this problem let $a_i = \frac{i}{k}$

(c) Assuming Z is AN, is it consistent for alternatives of the form of G in (b).

Kolmogorov-Smirnov Test (Test No. 4):

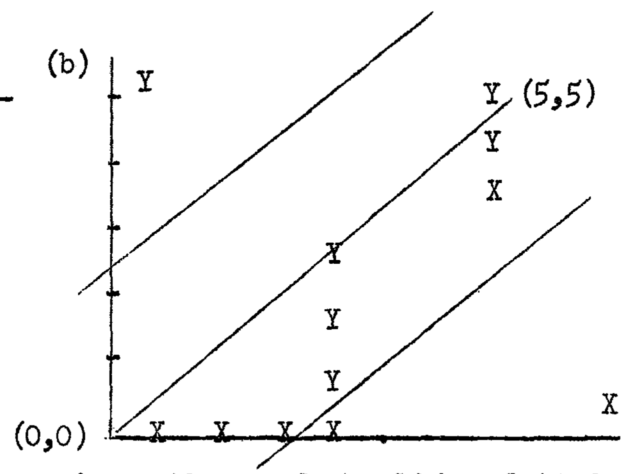
Our basic sample (X X X X Y Y Y X Y Y) could be expressed graphically in two ways:

(a)



$D_{mn} = .8$

(b)



i.e., the sample could be plotted as a two-dimensional random walk reaching the point (5,5)--reject if the walk strays beyond a line parallel to the 45^o line.

Asymptotic distributions of D_{mn}, if H is true:

$$\lim_{m,n \to \infty} \Pr\left[\sqrt{\frac{mn}{m+n}}\, D_{mn} \leq z\right] = L(z) = 2\sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 z^2}$$

L(z) has been tabled in the Annals of Math. Stat., 1948.

Also, $D^{\pm}_{mn}$ have the same limiting distribution as $D^{\pm}_{n}$ (one-sample statistic) with the normalization $\sqrt{\frac{mn}{m+n}}$ (see p. 127)

Test: Reject H if $D_{mn} > d$.

Van der Waerden's X-test (Test No. 5):

$$X = \sum_{i=1}^{m} \Psi\left(\frac{R_i^x}{N+1}\right) \qquad \text{where} \quad \Psi(u) = \phi^{-1}(u)$$

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2}\, dt$$

$$X \text{ is } AN\left(0, \frac{mn}{N-1} Q\right) \qquad \text{where } Q = \frac{1}{N} \sum_{i=1}^{N} \Psi^2\left(\frac{i}{N+1}\right) \qquad N = m + n$$

Example: X X X X Y Y Y X Y Y

$$R_i^x = 1 \quad 2 \quad 3 \quad 4 \quad 8$$

$$\frac{R_i^x}{N+1} = \frac{1}{11} \quad \frac{2}{11} \quad \frac{3}{11} \quad \frac{4}{11} \quad \frac{8}{11} = .09, .18, .27, .36, .73$$

using normal deviate tables

$$\Psi\left(\frac{R_i^x}{N+1}\right) = z_{.09} \quad z_{.18} \quad z_{.27} \quad z_{.36} \quad z_{.73}$$

$$= -1.34, \quad -.91, \quad -.60, \quad -.35, \quad +.60$$

$$X = \sum_i \Psi\left(\frac{R_i^x}{N+1}\right) = -2.60$$

For determining the variance of X, tables of Q have been given by Van der Waerden in an appendix to his text.

- -

Theorem 32 (Pitman's Theorem on A.R.E.):

Assume: T_n, T_n^* are A. N. Statistics

Ω' is a subset of Ω indexed by ρ such that

when H is true $\rho = 0$.

Let the sequence of ρ's tends to θ, i.e., $\rho_1\ \rho_2\ \dots\ \rho_n \longrightarrow 0$.

Test: T -- reject H if $T_n > t_{n\alpha}$

T -- reject H if $T^* > t^*_{n\alpha}$

Assumptions:

(1) $\frac{d}{d\rho} E\rho\,(T_n) > 0$

(2) $\lim \frac{d}{d\rho} \frac{E\rho\,(T_n)\,\big|\,\rho=0}{\sqrt{n}\ \sigma_o} = c$ if $\rho_n = \frac{k}{\sqrt{n}}$

(3) $\lim\limits_{n \to \infty} \frac{\frac{d}{d\rho}\ E\rho\,(T_n)\,\big|\,\rho = \rho_n}{\frac{d}{d\rho}\ E\rho\,(T_n)\,\big|\,\rho = 0} = 1$

(4) $\lim\limits_{n \to \infty} \frac{\sigma_{\rho_n}\,(T_n)}{\sigma_0\,(T_n)} = 1$

<u>Theorem:</u> Under these regularity conditions, the limiting power of the T_n test for alternatives $\rho_n = \frac{k}{\sqrt{n}}$ as $n \to \infty$ is $1 - \phi\,(z_\alpha - kc)$.

The A.R.E. of T^*_n to T^* is

$$= \left(\frac{c}{c^*}\right)^2 = \lim_{n \to \infty} \left[\frac{\frac{d}{d\rho}\,E\rho\,(T_n)\,\big|\,\rho = 0}{\frac{d}{d\rho}\,E\rho\,(T^*_n)\,\big|\,\rho = 0}\right]^2 \quad \frac{\sigma_0^2\,(T^*_n)}{\sigma_0^2\,(T_n)}$$

Proof:

$$\Pr\left[\frac{T_n - E_0(T_n)}{\sigma_0} > \frac{t_{n\alpha} - E_0(T_n)}{\sigma_0}\right] = \alpha$$

For large n, since T is asymptotically normal, $t_{n\alpha} = z_\alpha \sigma_0 + E_0(T_n)$

$$\lim_{n \to \infty} \Pr[\text{Reject } H \mid \rho_n] = \lim_{n \to \infty} \Pr\left[\frac{T_n - E\rho_n(T_n)}{\sigma_{\rho_n}} > \frac{z_\alpha \sigma_0 + E_0(T_n) - E\rho_n(T_n)}{\sigma_{\rho_n}}\right]$$

$$= \lim_{n \to \infty} \Pr\left[z > z_\alpha \frac{\sigma_0}{\sigma_{\rho_n}} + \frac{E_0(T_n) - E\rho_n(T_n)}{\sigma_0} \frac{\sigma_0}{\sigma_{\rho_n}}\right]$$

$$\lim_{n \to \infty} z_\alpha \left(\frac{\sigma_0}{\sigma_{\rho_n}}\right) + \frac{E_0(T_n) - E\rho_n(T_n)}{\sigma_0} \left(\frac{\sigma_0}{\sigma_{\rho_n}}\right) = z_\alpha - kc$$

from assumption 4, as $n \longrightarrow \infty$, $\left(\frac{\sigma_0}{\sigma_{\rho_n}}\right) \longrightarrow 1$

using a Taylor series expansion:

$$E\rho_n(T_n) = E_o(T_n) + \rho_n\left[\frac{d}{d\rho} E\rho(T_n)\right]_{\rho_n'} \qquad 0 < \rho_n' < \rho_n$$

Putting $\rho_n = \frac{k}{\sqrt{n}}$

$$\lim_{n \to \infty}\left[\frac{E_o(T_n) - E\rho_n(T_n)}{\sigma_o}\right] = \lim_{n \to \infty}\left[-\frac{k}{\sqrt{n}} \frac{\left[\frac{d}{d\rho} E_\rho(T_n)\right]\Big|_{\rho = \rho_n'}}{\sigma_o}\right]$$

$$= -kc$$

Therefore: $\lim_{n \to \infty} \Pr\left[\text{Reject } H \mid \rho_n\right] = 1 - \phi(z_{1-\alpha} - kc)$

By a similar argument, the limiting power of the T* test is

$1 - \phi(z_{1-\alpha} - k^* c^*)$

$$\text{where } c^* = \lim_{n \to \infty} \frac{\frac{d}{d\rho} E_\rho(T_n^*)\Big|_{\rho = 0}}{\sqrt{n}\ \sigma_o^*}$$

We want to determine sequences $\{n_1\}$, $\{n_1^*\}$ such that $1 - \emptyset\,(Z_{1-\alpha} - kc) = 1 - \emptyset\,(Z_{1-\alpha} - k^*c^*)$ which means that $kc = k^*c^*$. Also, for the two sequences to be the same $\rho_n = \rho_n^*$ or $\frac{k}{\sqrt{n_1}} = \frac{k^*}{\sqrt{n_1^*}}$.

Thus we can determine the equality of the following ratios:

$$\frac{n_1^*}{n_1} = \left(\frac{k^*}{k}\right)^2 = \left(\frac{c}{c^*}\right)^2 = \lim_{n \to \infty} \left(\frac{n\ \sigma_0^{*2}}{n\ \sigma_0^2}\right)\left(\frac{\frac{d}{d\rho} E\rho\ (T_n)\ \Big|\ \rho=0}{\frac{d}{d\rho} E\rho\ (T_n^*)\ \Big|\ \rho=0}\right)^2$$

The A.R.E. of T to T* is given by any of these ratios, or as stated in the theorem:

$$\text{A.R.E.} = \left(\frac{c}{c^*}\right)^2 = \lim_{n \to \infty} \left(\frac{\sigma_0^{*2}}{\sigma_0^2}\right)\left(\frac{\frac{d}{d\rho} E\rho\ (T_n)\ \Big|\ \rho = 0}{\frac{d}{d\rho} E\rho\ (T_n^*)\ \Big|\ \rho = 0}\right)^2$$

<u>Example</u>: Obtaining the A.R.E. for the Wilcoxen test versus the normal mean test, i.e.

U versus $Z = \dfrac{\bar{Y} - \bar{X}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}$ which is asymptotically equivalent to the two-sample t-test.

X and Y have d.f. F(x) under H_0.
X has d.f. F(x), Y has d.f. $F(x - \mu)$ under H_1.
In both cases the variance = σ^2.

For Wilcoxen's test:

$$p = \Pr\left[Y > X\right] = \int_{-\infty}^{\infty} \left[1 - G(x)\right] dF(x) = \int_{-\infty}^{\infty} \left[1 - F(x - \mu)\right] f(x)\, dx$$

$$\frac{dp}{d\mu}\Bigg|_{\mu=0} = \int_{-\infty}^{\infty} f(x - \mu)\, f(x)\, dx = \int_{-\infty}^{\infty} f^2(x)\, dx$$

$$E(U) = mnp \qquad \frac{d}{d\mu} E(u) \Big|_{\mu=0} = mn \int_{-\infty}^{\infty} f^2(x)\, dx$$

$$\text{Var } (U) = \frac{mn(m + n + 1)}{12}$$

$$E(Z) = \frac{\int x\, dF(x - \mu) - \int x\, dF(x)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\mu}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\frac{dE(Z)}{d\mu} = \frac{1}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1}{\sigma\sqrt{\frac{m+n}{mn}}} \qquad \text{Var } (Z) = 1$$

Therefore, the A.R.E. of U to Z is

$$\text{A.R.E.} = \lim_{n \to \infty} \left(\frac{1}{\frac{mn(m + n + 1)}{12}} \right) \left(\frac{mn \int_{-\infty}^{\infty} f^2(x)\, dx}{\frac{1}{\sigma}\sqrt{\frac{mn}{m+n}}} \right)^2$$

which reduces to:

$$\text{A.R.E. (U to Z)} = 12\, \sigma^2 \left[\int_{-\infty}^{\infty} f^2(x)\, dx \right]^2$$

Thus we can compare U and Z for any $f(x)$ whatsoever.

For instance, if $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2}$ then $\int f^2(x)\, dx = \frac{1}{2\pi\sigma^2} \int e^{-x^2/\sigma^2}\, dx$

using the transformation $\frac{x}{\sigma} = \frac{y}{\sqrt{2}}$

$$\int f^2(x)\, dx = \frac{1}{\sqrt{2\pi\sigma}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$$= \frac{1}{2\sigma\sqrt{\pi}}$$

and $$12\, \sigma^2 \left[\int_{-\infty}^{\infty} f^2(x)\, dx \right]^2 = 12\, \sigma^2 \left(\frac{1}{2\sigma\sqrt{\pi}} \right)^2$$

$$= \frac{12\, \sigma^2}{4\pi\sigma^2} = \frac{3}{\pi} = .955$$

All of which says that the A.R.E. of the Wilcoxen test to the Normal (two-sample) test, if the underlying populations are normal, and we are testing "slippage of the mean" is $3/\pi$.

Problem 66:

(a) Evaluate the A.R.E. of U and Z when

(1) $f(x) = 1 \quad 0 \leq x \leq 1$, i.e.

$F(x) = x \qquad 0 \leq x \leq 1$

$F(x - \mu) = x - \mu \qquad \mu < x \leq 1 + \mu$

(2) $f(x) = e^{-x} \qquad 0 \leq x < \infty$

(b) Find an $f(x)$ such that the A.R.E. of U to Z is $+\infty$.

Remark: It has been shown that the A.R.E. of U to Z in this case, i.e. testing slippage, is always $\geq .864$. - Hodges and Lehmen, Annals of Math. Stat., 1955.

Test	A.R.E. of test to Z (testing for slippage)	Consistency
1. Median	$2/\pi$	yes, if the median of $F_0 \neq$ median of F_1
2. Runs	0	consistent for all $F_0 = F_1$
3. U	$3/\pi$	yes, if the median of $F_0 \neq$ median of F_1
4. K-S	???	consistent for all $F_0 = F_1$
5. X	1	yes, if the median of $F_0 \neq$ median of F_1

Robustness of a test (as propounded by Box) refers to the behavior of a test when the various assumptions made for the validity of the test are not fulfilled.

Type 1 error - Pr [reject H when true under the assumptions]

Power - Pr [reject H when false under the assumptions]

A test is said to be robust if Pr [reject H when true if assumptions are not fulfilled] remains close to α regardless of the assumptions.

Note: the Z-test, or two-tailed t-test, is robust.

The proponents of distribution-free statistics argue that the disadvantage of the Z-test is that the power may slip if the assumptions (of normalcy, etc) are not satisfied.

V. k-Sample Tests:

1. Median Tests
2. χ^2-Test with arbitrary groupings
3. Kruskal-Wallis Rank Order Test
4. Kolmogorov-Smirnov Type Tests

samples:

1. X_{11} X_{12} . . . X_{1n_1}
2. X_{21} X_{22} . . . X_{2n_2}

$$N = \sum_{j=1}^{k} n_j$$

.
.
.

k. X_{k1} X_{k2} . . . X_{kn_k}

1. Median Test is made by setting up a 2xk table:

sample	1	2 ...	i	...	k	
Above			m_i			N/2
Below			$n_i - m_i$			N/2
	n_1	n_2 ...	n_i	...	n_k	N

where m_i = the number of observations in sample i above the median of the combined sample.

Use a χ^2-test of the null hypothesis with the expected values = $n_i/2$ when N is even, and $\chi^2 = \sum_{i=1}^{k} \frac{2(m_i - n_i/2)^2}{n_i/2}$ with k-1 d.f. under the null hypothesis that the k samples all came from the same distribution.

2. χ^2-Test:

Being given or arbitrarily choosing groups A_j ($a_j \leq X \leq a_{j+1}$) define n_{ij} = the number of X's in sample i that fall in A_j.

Under the null hypothesis Pr[X falls in A_j] = p_j independently of i.

This can be tested by χ^2 in the usual manner -- as an rxk test of homogeniety, where χ^2 has (r-1)(k-1) d.f.

3. Kruskal-Wallis Test:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left[\bar{R}_i - \frac{N+1}{2} \right]^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

where R_i = sum of the ranks of sample i taken within the combined sample

$\bar{R}_i = R_i/n_i$

H is asymptotically distributed as χ^2 with k-1 d.f.

ref: March 1959 JASA for small sample approximations.

Problem 67: What does H reduce to when k = 2?

Prove your answer.

4. K-S Type Tests:

Would involve drawing a step-function for each sample on the same graph. Unfortunately nothing is now known about the distributions.

ref: Kefer, Annals of Math. Stat., article to be published probably in 1959.

Consistency:

The Median and H are consistent against all alternatives if at least one of the sub-group medians differs from the others.

A.R.E.:

A.R.E. for slippage alternatives:

median test	against the standard ANOVA test	$2/\pi$
H test	against the standard ANOVA test	$3/\pi$

where the underlying distribution is normal.

If the underlying distribution is rectangular, then the A.R.E.'s become:

median	against ANOVA	1/3
H	against ANOVA	1

CHAPTER VI

TESTING OF HYPOTHESES -- PARAMETRIC THEORY -- POWER

Refs: Cramer, ch. 35
Kendall, vol. 2, chs. 26-27
Lehmann, "Theory of Testing Hypotheses", U. of Cal. Bookstore (notes by C. Blyth)
Fraser, "Nonparametric Methods in Statistics", ch. 5

1. Generalities

$X_1, X_2, \ldots, X_n$ have d.f. $F(\underline{x}, \underline{\theta})$

-- usually the X's are independent with density $f(\underline{x}, \underline{\theta})$, the density having a specified parametric form with one or more unknown parameters.

For the parameter space Ω H_0: $\theta \in \omega_0$ H_1: $\underline{\theta} \in \omega_1$

Recall that $\phi(\underline{x})$ is a test function of size α such that

$$\phi(\underline{x}) = 1 \qquad \text{reject H with probability 1}$$
$$= k \qquad \text{reject H with probability k}$$
$$= 0 \qquad \text{do not reject H (accept H)}$$

$$\text{where } E\left[(\phi) \mid \underline{\theta} \in \omega_0\right] \leq \alpha$$

Power function: $\beta_\phi(\theta) = E\left[\phi \mid \theta\right]$

Ref: Defs. 33-38 in chapter 5.

Def. 41: ϕ^* is a uniformly most powerful (u.m.p.) test of size α, if ϕ being any other size α test

$$\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta) \text{ for all } \theta \in \omega_1$$

2. Probability Ratio Tests

Neyman-Pearson Theorem:

X is a continuous random variable with density $f(x, \theta)$.

H_0: $\theta = \theta_0$ (simple hypothesis) H_1: $\theta = \theta_1$ (simple alternative)

Assume: $f(x, \theta_0) > 0$; $f(x, \theta_1) > 0$ for the same set S.

Assume: $\dfrac{f(x, \theta_1)}{f(x, \theta_0)}$ is a continuous random variable.

Theorem 33 (Neyman-Pearson):

The most powerful test of H_0 against H_1 is given by $\phi^*(x)$ defined as follows:

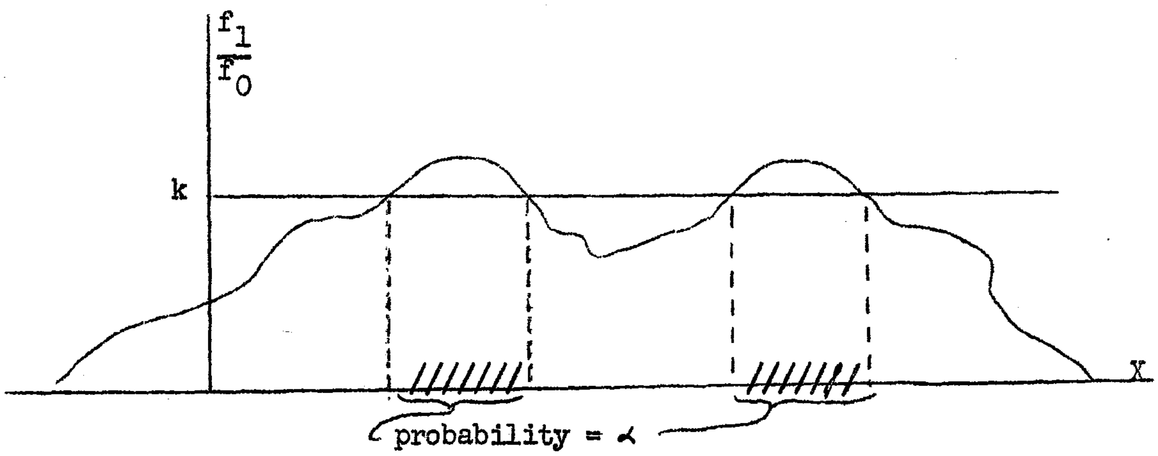
$$\phi^*(x) = 1 \quad \text{when } f(x, \theta_1) > k\, f(x, \theta_0) \text{ or } f_1 > k\, f_0$$
$$= 0 \quad \text{elsewhere}$$

where k can be chosen so that ϕ^* is of size α.

If the test ϕ^* is independent of θ_1 for $\theta_1 \varepsilon \omega_1$ then ϕ^* is the u.m.p. test of H_0 against H_1: $\theta_1 \varepsilon \omega_1$.

Remark: This is what has been called the probability ratio test, since Neyman-Pearson originally expressed the theorem that

$$\phi^* = 1 \quad \text{if } \frac{f_1}{f_0} > k$$



Proof:

1. To show there is a required k that makes ϕ^* of size α

define: $$\alpha(k) = \Pr\left[\frac{f_1(x)}{f_0(x)} > k \;\middle|\; X \text{ has density } f_0\right]$$

$$\alpha(0) = 1 \qquad \alpha(\infty) = 0$$

since $\frac{f_1}{f_0}$ is a continuous random variable, $1 - \alpha(k)$, which is the d.f. of this random variable, is a continuous function and is monotone non-decreasing, hence for some k' we must have $\alpha(k') = \alpha$.

2. To show that ϕ^* is u.m.p.

Let ϕ be any other test of size α.

We want to show that:

$$\int \phi^*(x)\, f(x, \theta_1)\, dx \geq \int \phi(x)\, f(x, \theta_1)\, dx$$

Consider: $$\int \left[\phi^*(x) - \phi(x)\right] \left[f(x, \theta_1) - k\; f(x, \theta_0)\right] dx \geq 0$$

That this integral is ≥ 0 follows since

when $f_1 - k\, f_0 > 0$, then $\phi^* = 1$, and $\phi^* - \phi \geq 0$

when $f\ \ - k\, f \leq 0$, then $\phi^* = 0$, and $\phi^* - \phi \leq 0$

Expanding this integral we get:

$$\int \phi^* f_1\, dx - \int \phi\, f_1\, dx - k\left[\int \phi^* f_0\, dx - \int \phi\, f_0\, dx\right] \geq 0$$

but $\int \phi^* f_0\, dx = \int \phi\, f_0\, dx = \alpha$ = the size condition

Therefore: $$\int \phi^* f_1\, dx \geq \int \phi\, f_1\, dx$$

and ϕ^* is u.m.p.

q.e.d.

<u>Example 1</u>: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ $\qquad \sigma^2$ known

H_0: $\mu = 0$ $\qquad H_1$: $\mu = \mu_1 > 0$

$$f_1 = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{\sum (X_i - \mu_1)^2}{2\sigma^2}} \qquad f_0 = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{\sum X_i^{\,2}}{2\sigma^2}}$$

Reject H_o when $\frac{f_1}{f_0} > k$

$$e^{-\frac{\sum X_i - 2\mu_1 \sum X_i + n\mu_1^{\,2} - \sum X_i^{\,2}}{2\sigma^2}} > k$$

or $$\frac{2\mu_1 \sum X_i - n\mu_1^{\,2}}{2\sigma^2} > \ln k$$

or $\quad \sum x_i > \dfrac{(\ln k)\, 2\sigma^2 + n\mu_1^{\,2}}{2\mu_1}$

i.e., if $\bar{X} > K$

To satisfy the size condition, $K = z_{1-\alpha}\left(\sigma/\sqrt{n}\right)$ (which is independent of μ_1)

If H_0 is true, $\bar{X}$ is $N(0, \frac{\sigma^2}{n})$ $\qquad Pr\left[\dfrac{\bar{X}}{\sigma/\sqrt{n}} > z_{1-\alpha}\right] = \alpha$

Remarks: If the most powerful test of H_0 against H_1: $\theta = \theta_1$ is independent of θ_1 for some family ω_1, then the probability ratio test, ϕ^*, is u.m.p. for H_0: $\theta = \theta_0$ against H_1: $\theta \in \omega$.

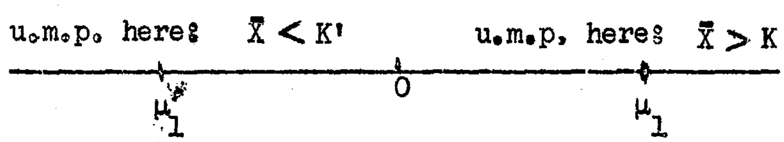
Therefore, $\bar{X} > z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$ is u.m.p. for H_0 against H_1: $\mu > 0$.

For H_0: $\mu = 0$ against H_1: $\mu < 0$ the u.m.p. test is $\bar{X} < z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$

Example 2: $X_1, X_2, \ldots, X_n$ are NID (μ, σ^2)

H_0: $\mu = 0 \qquad H_1$: $\mu_1 \neq 0$

There can be no u.m.p. test for this problem since the u.m.p. tests for $\mu < 0$ and $\mu > 0$ differ.



Problem 68: $X_1, X_2, \ldots, X_n$ are independently distributed with density

$f(x) = \theta e^{-\theta x} \qquad x > 0 \qquad \theta > 0$

H_0: $\theta = \theta_0 \qquad H_1$: $\theta = \theta_1 > \theta_0$

(1) Find the u.m.p. test explicitly (i.e., find the distribution of the test statistic).

(2) Write down the power function in terms of a familiar tabulation.

Remark: $\frac{f_1(x)}{f_0(x)}$ is required to be a continuous random variable.

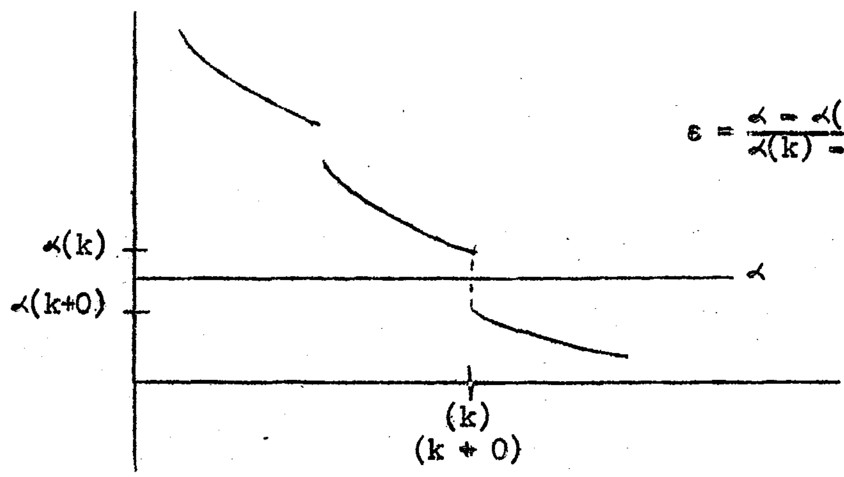
If this ratio is not a continuous random variable, then $\alpha(k) = \Pr\left[\frac{f_1}{f_0} > k\right]$ may be discontinuous, so that $\alpha(k) = \alpha$ has no solution. (e.g. small binomial situations where one can't get $\alpha = .05$ exactly.)

ϕ^* in this case is defined to be:

$$= 1 \quad \text{when } \frac{f_1(x)}{f_0(x)} > k$$

$$= 0 \quad \text{when } \frac{f_1(x)}{f_0(x)} < k$$

$$= \varepsilon \quad \text{when } \frac{f_1(x)}{f_0(x)} = k$$ where ε is chosen so that the size of the test comes out as α



$$\varepsilon = \frac{\alpha - \alpha(k+0)}{\alpha(k) - \alpha(k+0)}$$

Problem 69: For f_1 let S_1 be the set where $f_1 > 0$.

For f_0 let S_0 be the set where $f_0 > 0$.

(1) If S_0 and S_1 are not coincident then there may be no test of size α.

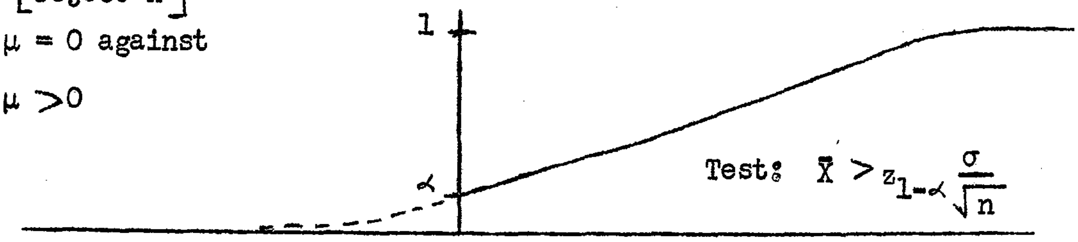
(2) If there is no test of size α given by the Neyman-Pearson Lemma (Theorem 33) there is a test of size $< \alpha$ with power = 1.

Hypotheses noted thus far have been simple hypotheses. If ω has more than one point, the hypothesis is called a composite hypothesis, e.g.:

$$X_1, X_2, \ldots, X_n \text{ are NID } (\mu, \sigma^2)$$

H_0: $\mu \leq 0$ $\qquad$ H_1: $\mu > 0$

$\beta(\theta) = \Pr[\text{reject } H]$
for H_0: $\mu = 0$ against
H_1: $\mu > 0$



Extension of the u.m.p. test to composite hypotheses by means of a "most unfavorable distribution of θ".

$$H_0: \theta \in \omega \qquad H_1: \theta = \theta_1$$

Let $\lambda(\theta)$ be a distribution of θ over ω.

Then X has a density $f(x, \theta)$.

$$h_\lambda(x) = \int_\omega f(x, \theta)\, d\lambda(\theta)$$ = density of X under H_0 plus the additional information.

Let $\tilde{H}_0$ be: X has density $h_\lambda(x)$ $\qquad$ H_1: $\theta = \theta_1$
i.e., the density of X is $f(x, \theta_1)$

Let ϕ_λ^* be the most powerful test of $\tilde{H}_0$ against H_1.

Theorem 34: If ϕ^* is of size α for the original test, it is m.p. for this test.

Proof: Let ϕ be any other test of H_0.

$$\phi(x)\, f(x, \theta)\, dx \leq \alpha \quad \text{for all } \theta \in \omega$$

$$\text{Then} \quad \alpha \geq \int_\omega \left[\int_x \phi(x)\, f(x, \theta)\, dx\right] d\lambda(\theta)$$

$$= \int_x \phi(x) \left\{ \int_\omega f(x, \theta)\, d\lambda(\theta) \right\} dx$$

$$= \int \phi(x)\, h_\lambda(\theta)\, dx$$

Thus ϕ is of required size for $\tilde{H}_0$.

and: $$\int \phi^*(x)\, f(x, \theta)\, dx \geq \int \phi(x)\, f(x, \theta)\, dx \qquad \text{q.e.d.}$$

Example: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ σ^2 known

H_0: $\mu \leq 0$ H_1: $\mu > 0$

ϕ^* test: reject if $\bar{X} > z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ which was found to be u.m.p. for H_0': $\mu = 0$ against H_1: $\mu > 0$. We know that Pr $\left[\text{rejecting H with } \phi^* \middle| \mu\right] \leq \alpha$ for $\mu \leq 0$.

Put $\Lambda(\theta) = 0$ $\theta < 0$

$= 1$ $\theta \geq 0$

(i.e., concentrate the distribution at 0 which is the worst spot from the standpoint of testing or distinguishing)

This makes our composite distribution:

$$h_\lambda(x) = \int f(x, \theta)\, d\lambda(\theta) = f(x, 0)$$

so that our problem is back to H_0' and we have for H_0' a u.m.p. test which is also a test of the original H_0.

One way to get an optimum test in the absense of a u.m.p. test is to restrict the class of tests to be considered and look for probability ratio test for restricted class. Such restrictions are:

1. Unbiasedness

Def. 42: ϕ is an unbiased test if $\beta_\phi(\theta) \geq \alpha$ for all $\theta \in \omega_1$

2. Similarity

Def. 43: ϕ is a similar test if $E_{\underline{\theta}}(\phi) = \alpha$ for all $\underline{\theta} \in \omega_0$

3. Invariance

X has density $f(x, \theta)$

G = a family of transformations of the sample space of X onto itself.

e.g.: $g(x) = cx$ change of scale

$= x + d$ translation of axes

Let $g(X)$ have density $f\left[x, \bar{\bar{g}}(\theta)\right]$ $\bar{\bar{g}}(\theta) \in \Omega$

$\bar{\bar{g}}$ is a transformation of the parameter space induced by the transformation of the sample space.

e.g. : if X is $N(\mu, \sigma^2)$ $\quad$ cX is $N(c\mu, c^2\sigma^2)$

$$\bar{g}(\mu, \sigma^2) = c\mu, c^2\sigma^2$$

if X is $N(\mu, \sigma^2)$ $\quad$ $X + d$ is $N(\mu + d, \sigma^2)$

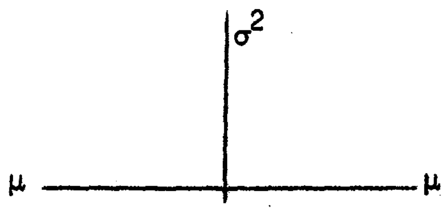
$$\bar{g}(\mu, \sigma^2) = \mu + d, \sigma^2$$

If we have H_0: $\theta \varepsilon \omega$ and H_1: $\theta \varepsilon \Omega - \omega$ then a transformation induced by g leaves the problem <u>invariant</u> if

$$\bar{g}(\theta) \varepsilon \omega \text{ when } \theta \varepsilon \omega$$

$$\bar{g}(\theta) \varepsilon \Omega - \omega \text{ when } \theta \varepsilon \Omega - \omega$$

Example: X is $N(\mu, \sigma^2)$



H_0: $\mu = 0$

$g(X) = cX \qquad c > 0$

$\bar{g}(\mu, \sigma^2) = c\mu, c^2\sigma^2$

ω is the σ^2 axis

under H_0 $\quad g(X)$ is $N(0, c^2\sigma^2)$

H_1 $\quad g(X)$ is $N(c\mu, c^2\sigma^2)$ $\qquad$ which doesn't change ω

<u>Def. 44</u>: A test, $\phi(x)$, is invariant under a transformation g (for which the corresponding $\bar{g}$ leaves the problem invariant) if $\phi\left[g(x)\right] = \phi(x)$

Example: $$t = \frac{\bar{X}}{s/\sqrt{n}} = \frac{\frac{1}{n}\sum X_i}{\left(\frac{\sum (X_i - \bar{X})^2}{(n-1)\,n}\right)^{1/2}} = (\text{under } g = cX)\ \frac{\frac{1}{n}\sum cX_i}{\left(\frac{\sum (cX_i - \bar{X}c)^2}{(n-1)\,n}\right)^{1/2}}$$

$$= \frac{\frac{1}{n}\sum X_i}{\left(\frac{\sum (X_i - \bar{X})^2}{(n-1)\,n}\right)^{1/2}} = t$$

<u>Def. 45</u>: If among all unbiased (or similar or invariant) tests there is a ϕ^* which is u.m.p., then ϕ^* is u.m.p.u. (or u.m.p.s. or u.m.p.i.).

Uniformly Most Powerful Unbiased (u.m.p.u.) Tests:

Single parameter θ H_0: $\theta = \theta_0$ which is inside an open interval of the parameter space.

"f" is differentiable with respect to θ.

Remark: If ϕ is unbiased, then

$$\left.\frac{\partial \beta_\theta(\theta)}{\partial \theta}\right|_{\theta=\theta_0} = 0$$

Proof: $\beta_\phi(\theta_0) \leq \alpha$

$\beta_\phi(\theta) > \alpha$ for $\theta \neq \theta_0$

Hence $\beta_\phi(\theta)$ has a minimum at $\theta = \theta_0$

$\beta_\phi(\theta) = \int \phi(x)\, f(x, \theta)\, dx$

$\beta_\phi(\theta)$ is differentiable with respect to θ

and hence $\left.\frac{\partial \beta_\phi(\theta)}{\partial \theta}\right|_{\theta=\theta_0} = 0$

Note: For unbiasedness, alternatives must be two-sided (otherwise the power curve has no minimum).

Assuming that $\int f(x, \theta)\, dx$ is differentiable under the integral sign, and with H_0: $\theta = \theta_0$ H_1: $\theta = \theta_1$ (two-sided) we can get:

Theorem 35: If there exist k_1, k_2 so that

$$\phi_u^* = 1 \text{ when } f(x, \theta_1) > k_1 f(x, \theta_0) + k_2 \left.\frac{\partial f}{\partial \theta}\right|_{\theta=\theta_0}$$

$= 0$ elsewhere

is of size α for H_0 and unbiased, then it is m.p. for alternatives θ_1. If the test does not depend on θ_1 it is u.m.p.u. for H_1: $\theta \in \omega$.

Proof: (ref: Cramer p.532)

Let ϕ be any other unbiased test.

Then $\int \phi_u^* f_0\, dx = \alpha = \int \phi f_0\, dx$ — - size conditions

$$\int \phi_u^* \left.\frac{\partial f}{\partial \theta}\right|_{\theta=\theta_0} dx = 0 = \int \phi \left.\frac{\partial f}{\partial \theta}\right|_{\theta=\theta_0} dx$$ - unbiasedness conditions

We want to show that: $\int \phi_u^* f_1 \, dx \geq \int \phi f_1 \, dx$

$$\int (\phi^* - \phi) f_1 dx = \int (\phi^* - \phi) \left[f_1 - k_1 f_0 - k_2 \frac{\partial f}{\partial \theta} \Big|_{\theta = \theta_0} \right] dx \geq 0$$

This integrand must always be ≥ 0

$$\text{since if } f_1 - k_1 f_0 - k_2 \frac{\partial f}{\partial \theta} \Big|_{\theta = \theta_0} \begin{cases} \geq 0 & 1 = \phi^* \geq \phi \\ \leq 0 & 0 = \phi^* \leq \phi \end{cases}$$

Thus the desired relationship always holds.

Comment: If you are trying to find a bounded function, $a \leq \phi \leq b$ which maximizes $\int \phi f \, dx$ subject to side conditions; $\int \phi f_i \, dx = c_i$ $i = 1, 2, \ldots, n$; then this maximum will be given by choosing:

$\phi = b$ where $f > \sum k_i f_i(x)$ where k_i are chosen to satisfy the n side conditi[...]

$= a$ otherwise

Example: $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2) σ^2 known

H_0: $\mu = 0$ H_1: $\mu \neq 0$

Consider a particular alternative μ_1.

$$f(\underline{x}, \mu_1) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{\sum (X_i - \mu_1)^2}{2\sigma^2}}$$

$$\frac{\partial f}{\partial \mu_1} = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \left(e^{-\frac{\sum (X_i - \mu_1)^2}{2\sigma^2}} \right) \left(\frac{\sum (X_i - \mu)}{\sigma^2} \right)$$

If we can find proper k's, the u.m.p.u. test is:

Reject H if

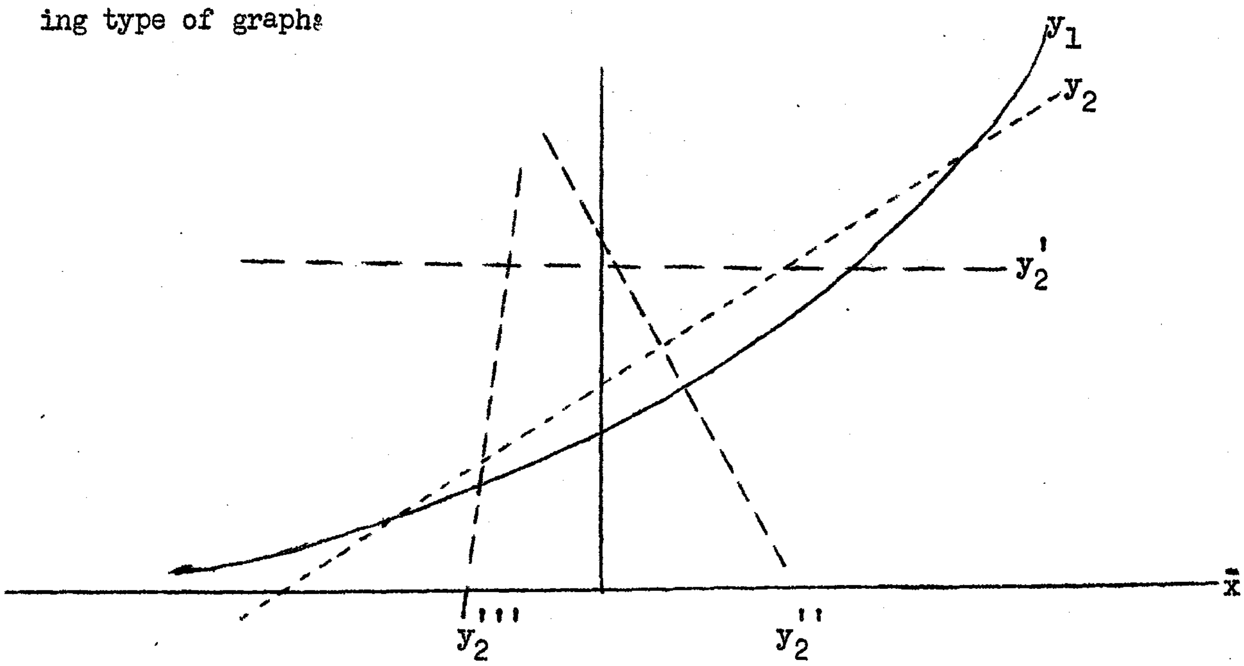
$$\frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{\sum (X_i-\mu_1)^2}{2\sigma^2}} \geq k_1 \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{\sum X_i^2}{2\sigma^2}} + k_2 \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{\sum X_i^2}{2\sigma^2}} \left(\frac{\sum X_i}{\sigma^2}\right)$$

(derivative evaluated at $\mu = 0$)

$$e^{\mu_1 \frac{\sum X_i}{\sigma^2}}\, e^{-\frac{n\mu_1^2}{2\sigma^2}} \geq k_1 + \frac{k_2}{\sigma^2}\sum X_i$$

$$e^{\frac{n\mu_1}{\sigma^2}\bar{\bar{x}}} \geq k_1' + k_2'\,\bar{\bar{x}} \qquad \text{where: } k_1' = e^{\frac{n\mu_1^2}{2\sigma^2}} \qquad k_2' = \frac{n}{\sigma^2}\, e^{\frac{n\mu_1^2}{2\sigma^2}}$$

If we set y_1 = the left hand side of this inequality, y_2 = the right hand side, and restrict ourselves to the cases where $\mu_1 \geq 0$; then we could get the following type of graph:



(y_2, y_2', y_2'', y_2''' are possible y_2 lines)

The test says to reject H if $\bar{\bar{x}} < a$ or $\bar{\bar{x}} > b$ where a may be $-\infty$

b may be $+\infty$

[y_2 gives a two-tailed test (finite a, b) -- y_2', y_2'', y_2''' all give one-tailed tests (only one intersection with y_1)]

Requiring that the test be of size α specifies that:

$$\Pr\left[a < \bar{x} < b\right] = 1 - \alpha$$

$$\text{i.e.,} \quad \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \int_a^b e^{-\frac{n\bar{x}^2}{2\sigma^2}} \, d\bar{x} = 1 - \alpha \qquad [1]$$

Also, the unbiasedness condition requires that:

$$\frac{\partial}{\partial \mu_1} \left[\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{a} e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} + \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \int_b^{\infty} e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} \right]_{\mu_1 = 0} = 0$$

$$\text{or} \quad \frac{\partial}{\partial \mu_1} \left[1 - \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} \right]_{\mu_1 = 0} = 0$$

$$\text{or} \quad -\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{n\bar{x}^2}{2\sigma^2}} \cdot \frac{n\bar{x}}{\sigma^2} \, d\bar{x} = 0 \qquad [2]$$

The function under the integral in [2] is an odd function, so that the integral is zero only if $a = -b$ (i.e., if a, b are symetrical).

[1] thus becomes:

$$\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-a}^{a} e^{-\frac{n\bar{x}^2}{2\sigma^2}} \, d\bar{x} = 1 - \alpha$$

so that we can determine a as: $a = z_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

Thus the test is:

Reject H if $\left| \bar{x} \right| > z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$

and since it is independent of μ_1 it is u.m.p.u.

Problem 70: $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2) μ known

H_0: $\sigma = \sigma_0$ H_1: $\sigma \neq \sigma_0$

Find the u.m.p.u. test for H_0.

note: $\sum (X_i - \mu)^2$ is sufficient for σ^2.

Theorem 36: If $\underline{T}$ is sufficient for $\underline{\theta}$ then given any test $\phi(\underline{x})$ there exists a test $\psi(\underline{T})$ with the same power function. Hence in looking for optimum tests, only functions of $\underline{T}$ need be considered.

Proof: define $\psi(\underline{T}) = E\left[\phi(\underline{x}) \,\middle|\, T\right]$ which is independent of θ by definition

Thus:

$$E\left[\psi(\underline{T})\right] = E_T E_X \left[\phi(\underline{x}) \,\middle|\, T\right] = E\left[\phi(\underline{x})\right] \text{ q.e.d.}$$

Invariant Tests

Refs: Lehmann, "Notes on Testing Hypotheses", ch. 4
Fraser, "Non Parametric Methods in Statistics", ch. 2

We have: observations: $\underline{X}$ parameter: θ density: $f(\underline{x}, \theta)$

H_0: $\theta \varepsilon \omega_0$ H_1: $\theta \varepsilon \omega_1$

Transformations: $x' = g(x)$ -- x' has density $f\left[\underline{x}, g(\theta)\right]$

i.e., there is an induced transformation on the parameter space:

$$\tilde{\theta} = \bar{g}(\theta)$$

G: the group of transformations that leave the problem invariant, i.e.

$$\bar{g}(\theta) \ \varepsilon \ \omega_0 \quad \text{if } \theta \ \varepsilon \ \omega_0$$

$$\bar{g}(\theta) \ \varepsilon \ \omega_1 \quad \text{if } \theta \ \varepsilon \ \omega_1$$

Example: $X_1, X_2, \ldots, X_n$ are NID (μ, σ^2)

$x' = cx + d$ x' is NID $(c\mu + d, c^2\sigma^2)$

$g(x) = cx + d$ $\bar{g}(\mu, \sigma^2) = (c\mu + d, c^2\sigma^2)$

H_0: $\mu = 0$ H_1: $\mu > 0$

If we set $d = 0$ and $c > 0$, then the problem is invariant, and we have:

$x' = g(x) = cx$ $\bar{g}(\mu, \sigma^2) = (c\mu, c^2\sigma^2)$

Sufficient statistics for μ, σ^2 are $\bar{X}$ and $S = \sum (X_i - \bar{X})^2$

Under the transformation:

$t = \frac{\bar{X}}{\sqrt{S}}$ is invariant (the inclusion of constants does not affect invariance).

(discussion to be continued after def. 46 and theorem 37.)

Def. 46: $m(x)$ is a maximal invariant function under a group of transformations if

1) $m\left[g(x)\right] = m(x)$

2) $m(x_1) = m(x_2) \Longrightarrow$ there is a g such that $g(x_2) = x_1$ of vice vers

Theorem 37: The necessary and sufficient condition that the test function $\phi(\underline{x})$ be invariant under G is that it depend only on $m(\underline{x})$.

Proof: 1) $\phi(\underline{x}) = \Psi[m(\underline{x})]$

$$\phi[g(\underline{x})] = \Psi\{m[g(\underline{x})]\} = \Psi[m(\underline{x})] = \phi(\underline{x})$$

2) Suppose that $\phi(\underline{x})$ is invariant. We have to show that $\phi(\underline{x})$ is a function of $m(\underline{x})$.

Given $\phi[g(\underline{x})] = \phi(\underline{x})$

show $m(x_1) = m(x_2) \Longrightarrow \phi(x_1) = \phi(x_2)$

$m(x_1) = m(x_2) \Longrightarrow$ for some g, call it g', $g'(x_2) = x_1$

thus: $\phi(x_1) = \phi[g'(x_2)] = \phi(x_2)$ which is what we set out to prove.

Returning to the "Student" problem:

It remains to show that $t = \frac{\bar{X}}{\sqrt{S}}$ is maximal invariant.

Setting $t = \frac{\bar{X}}{\sqrt{S}}$; $t' = \frac{\bar{X}'}{\sqrt{S'}}$ we have to show that given $t = t'$, we can find g such that $\bar{X}', S' = g(\bar{X}, S)$.

Consider $\frac{\bar{X}'}{\bar{X}}$ and call this ratio "a".

$$\frac{\bar{X}}{\sqrt{S}} = \frac{\bar{X}'}{\sqrt{S'}} \Longrightarrow \frac{\bar{X}'}{\bar{X}} = \frac{\sqrt{S'}}{\sqrt{S}} = a \qquad \text{or } S' = a^2 S$$

But this is just one of the members of the original family of transformations so t is maximal invariant.

Hence the problem is reduced to finding the u.m.p. test based on t.

In summary: $X_1, X_2, \ldots, X_n$ are NID (μ, σ^2)

H_0: $\mu = 0$ $\qquad$ H_1: $\mu > 0$

1) Reduce by using sufficient statistics: $\bar{X}$, S.

2) Impose the invariance conditions under the transformation $x' = cx$, $c > 0$. This reduces the problem to tests based on t.

3) Find the distribution of t under H_0 and H_1 and apply Theorem 33 (Probability Ratio Test).

Distribution of Non-Central t (t-distribution under H):

Refs: Neyman + Tokarska, JASA 1936, pp. 318-326 (tables for the one-sided case)
Welch + Johnson, Biometrika, 1940, pp. 362-389
Resnikoff + Lieberman, "Tables of the Non-Central t-distribution", Stanford University Press, 1957

τ, the non-central t-variable, is defined by:

$$\tau = \tau(\delta, f) = \frac{z + \delta}{\sqrt{w/f}}$$

where: z is N(0, 1)

w is χ^2 with f d.f.

δ is a constant > 0

The usual t-variable is

$$t = \tau(0, f) = \frac{\sqrt{n}\,\bar{X}}{\sqrt{\frac{S^2}{n-1}}} = \frac{\sqrt{n}\,\frac{\bar{X}}{\sigma}}{\sqrt{\frac{S^2}{\sigma^2(n-1)}}}$$

when H_0 is true

If H_1 is true, $\mu = \mu_1$ and $\frac{\sqrt{n}(\bar{X}-\mu_1)}{\sigma}$ is N(0, 1)

$$t = \frac{\sqrt{n}\left(\frac{\bar{X}-\mu_1}{\sigma}\right) + \sqrt{n}\,\frac{\mu_1}{\sigma}}{\sqrt{\frac{S^2}{\sigma^2(n-1)}}}$$

The joint density of z and w is:

$$f(z, w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{2^{f/2}\Gamma\frac{f}{2}} w^{\frac{f}{2}-1} e^{-\frac{w}{2}} \qquad \begin{matrix} -\infty < z < \infty \\ w \geq 0 \end{matrix}$$

To get the density of τ (the non-central t) we make the transformation:

$$\tau = \frac{z + \delta}{\sqrt{\frac{w}{f}}} \qquad u = w \qquad |J| = \frac{\sqrt{u}}{\sqrt{f}}$$

Thus:

$$f(\tau) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{f/2}\Gamma\frac{f}{2}\sqrt{f}} \int_0^\infty e^{\frac{1}{2}\left(\frac{\sqrt{u}\,\tau}{\sqrt{f}} - \delta\right)^2} u^{\frac{f-1}{2}} e^{-\frac{u}{2}}\, du$$

To get the distribution of the usual t, put $\delta = 0$.

The integral in $f(\tau)$ becomes

$$\int_0^\infty u^{\frac{f+1}{2}-1} e^{-\frac{u}{2}\left(\frac{\tau^2}{f}+1\right)} du$$

which can be readily evaluated recalling that

$$\int_0^\infty e^{-ax} x^{b-1}\, dx = \frac{\Gamma(b)}{a^b}$$

Thus,

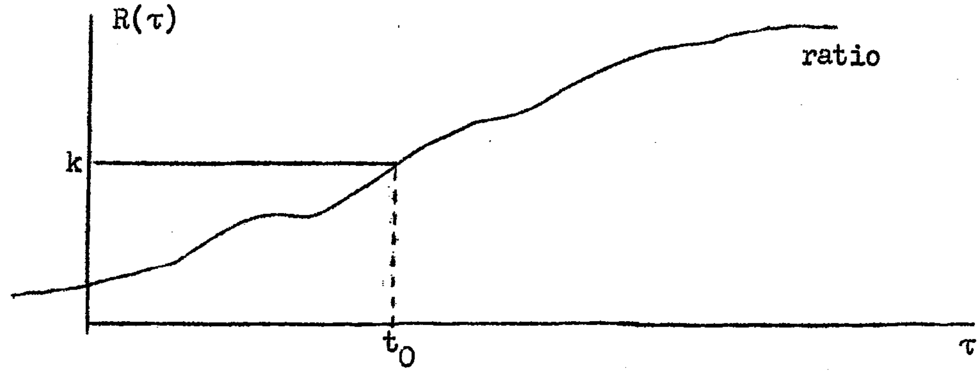
$$f(t) = \frac{1}{\sqrt{\pi f}} \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{f}\right)^{\frac{f+1}{2}}}$$

ref: Cramer, p.238

The u.m.p.i. test is to reject H_0 if :

$$R(\tau) = \frac{f(\tau, \delta)}{f(\tau, 0)} = C \left[\int_0^{\infty} e^{-\frac{1}{2}\left(\frac{\sqrt{u}\,\tau}{\sqrt{f}} - \delta\right)^2} e^{-\frac{u}{2}} u^{\frac{f+1}{2} - 1} \, du \right] \left(1 + \frac{\tau^2}{f}\right)^{\frac{f+1}{2}} > k$$

Note: This ratio is a monotone increasing function of τ. (The simplest proof of the required monotonisity is given by Kruskal, Annals of Math Stat, 1954, pp. 162-3.)



Since $t > t_o$ when the above ratio $> k$, the probability ratio test thus reduces to:

Reject H_o when $t > t_o$

Final result is that the m.p.i. test for H_1: $\mu = \mu_1 > \mu_o$ is:

Reject H_o when $t = \frac{\sqrt{n}\,\bar{X}}{\sqrt{S}} > t_{1-\alpha}(n-1)$

This is independent of μ_1 and hence is u.m.p.i. for H_0 against H_1.

Example: $X_1, X_2, \ldots, X_m$ are NID (μ_1, σ^2)

$Y_1, Y_2, \ldots, Y_n$ are NID (μ_2, σ^2) σ^2 ~~known~~ unknown

H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 > \mu_2$

Sufficient statistics for the three parameters are:

$$\bar{X}, \bar{Y}, \quad s_p = \sqrt{\frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{m+n-2}}$$

Consider: $X' = aX + b$ X' is $N(a\mu_1 + b, a^2\sigma^2)$

$Y' = cY + d$ Y' is $N(c\mu_2 + d, c^2\sigma^2)$

Invariance requires that:

$a\mu_1 + b = c\mu_2 + d$ when $\mu_1 = \mu_2$ for all μ

$a\mu_1 + b > c\mu_2 + d$ when $\mu_1 > \mu_2$

The first line requires that: $b = d$; $a = c$

The second line adds the requirement that: $a = c > 0$

Therefore: $X' = aX + b$, $Y' = aY + b$, $a > 0$ leaves the problem invariant.

To be proven:

$$t = \frac{\bar{X} - \bar{Y}}{s_p}$$ is a maximal invariant statistic.

$t = t' \Longrightarrow$ there exists an a, b such that $X' = aX + b$, $Y' = aY + b$

$$t = \frac{\bar{X} - \bar{Y}}{s_p} = \frac{\bar{X}' - \bar{Y}'}{s'_p} = t'$$

define: $\frac{s'_p}{s_p} = c > 0$

$$\frac{\bar{X} - \bar{Y}}{s_p} = \frac{\bar{X}' - \bar{Y}'}{c\, s_p}$$

$$c(\bar{X} - \bar{Y}) = \bar{X}' - \bar{Y}'$$

now let $\bar{X}' - c\bar{X} = d$

$$c(\bar{X} - \bar{Y}) = c\bar{X} + d - \bar{Y}'$$

$$-c\bar{Y} = d - \bar{Y}'$$

or: $\bar{Y}' = c\bar{Y} + d$

$\bar{X}' = c\bar{X} + d$ so that the same transformation has been applied to $\bar{X}$ and $\bar{Y}$.

The u.m.p. test invariant under the family of transformations,

$$X' = aX + b, \quad Y' = aY + b, \quad a > 0$$

is:

Reject H if $$t = \frac{\bar{X} - \bar{Y}}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{1-\alpha}(m+n-2)$$

Under the alternative $\mu_1 - \mu_2 = d$, thus

$$\frac{\frac{(\bar{X} - \bar{Y} - d)}{\sigma} + \frac{d}{\sigma}}{\frac{s_p}{\sigma}\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

has a non-central t distribution with parameters $\left[\frac{d}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}},\ m+n-2\right]$

Exercise: $\frac{d}{\sigma} = 0.8$ $\alpha = 0.05$

Find m, n required to have the power $\geq .90$ (put m = n).

Try to find m, n also by normal approximation.

Answer: d.f. = 2(n-1) non-centrality parameter $= \rho = \frac{d\sqrt{n}}{\sigma\sqrt{2}} = \frac{.8\sqrt{n}}{\sqrt{2}}$

$= .5657$ n

n	ρ
30	3.1
28	2.99
27	2.94
25	2.8

For the power to exceed .90, from the Neyman-Tokarska tables we need:

$\rho = 2.99$ at d.f. = 30

$\rho = 2.93$ at d.f. = ∞

Thus, by rough interpolation, the minimum size required is n = m = 28.

From the normal approximation: n = 26.7 or 27

Problem 71: $X_1, X_2, \ldots, X_m$ are NID (μ_1, σ_1^2)

$Y_1, Y_2, \ldots, Y_n$ are NID (μ_2, σ_2^2)

H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 > \sigma_2^2$

a) Find the group of transformations leaving H_0 invariant.

b) Find sufficient statistics for the parameters.

c) Find the maximum invariant function.

d) Find the u.m.p.i. test.

e) Find the u.m.p.u.i. test (i.e., set down conditions to get the rejection region as in problem 70) for: H_0: $\sigma_1^2 = \sigma_2^2$ against H_1': $\sigma_1^2 \neq \sigma_2^2$.

f) Show that the usual test which is to reject if $F > F_{1-\alpha}$ (m-1, n-1) where:

$$s_x^2 = \frac{\sum (X_i - \bar{X})^2}{m-1} \quad s_y^2 = \frac{\sum (Y_i - \bar{Y})}{n-1} \quad F = \max\left(\frac{s_x^2}{s_y^2};\ \frac{s_y^2}{s_x^2}\right)$$

is a test of size 2α with "equal tail probabilities".

g) Plot the power of the test of (f) for $m = n = 10$ with $2\alpha = 0.05$. (Include the points $\sigma_1^2/\sigma_2^2 = 0.5, 0.8, 1.25, 1.5, 2.0$).

Tests for Variances:

$X_1, X_2, \ldots, X_n$ are NID (μ, σ_2)

H_0: $\sigma^2 = \sigma_0^2$ $\qquad$ H_1: $\sigma^2 > \sigma_0^2$

$\qquad\qquad\qquad\qquad$ H_1': $\sigma^2 \neq \sigma_0^2$

1) μ known $\qquad \sum (X_i - \mu)^2$ is sufficient for σ^2.

Reject H_0 if $\sum (X_i - \mu)^2 > K$ -- u.m.p. for H_0 against H_1.

Reject H_0 if $\sum (X_i - \mu)^2 < K_1$ or $> K_2$ -- u.m.p.u. for H_0 against H_1'.

ref: problem 70

2) μ unknown $\qquad \bar{X}, \ \dfrac{\sum (X_i - \bar{X})^2}{n-1}$ are sufficient for μ, σ^2

Problem is invariant under translation, i.e.:

$$X' = X + a \qquad \bar{X}' = \bar{X} + a \qquad (s'^2) = s^2$$

To find a maximum invariant function

$$f(\bar{X}', s'^2) = f(\bar{X} + a, s^2) = f(\bar{X}, s^2) \text{ must hold for all a.}$$

This says that $f(\bar{X}, s^2)$ is independent of $\bar{X}$. Thus invariant functions are functions of s^2 only.

Hence, since $\dfrac{(n-1)\ s^2}{\sigma^2}$ is χ^2 with (n-1) d.f. when H_0 is true, the problem is exactly that of (1) with the d.f. reduced by 1 (i.e., n-1 vice n).

Summary of Normal Tests:

$X_1, X_2, \ldots, X_m$ are NID (μ_x, σ_x^2)
$Y_1, Y_2m \ldots, Y_n$ are NID (μ_y, σ_y^2)

X's, Y's are independent

Alternative Hypothesis	Test: Reject H if	Classification (for H against the given alternative)	Invariant under transformations of the form
1. $H_0: \mu_x = 0$	$(\sigma_x^2$ known$)$		
$H_1: \mu_x > 0$	$\bar{x} > z_{1-\alpha} \frac{\sigma_x}{\sqrt{m}}$	u.m.p.	
$H_1': \mu_x \neq 0$	$\lvert\bar{x}\rvert > z_{1-\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{m}}$	u.m.p.u.	
2. $H_0: \mu_x = \mu_0$	$(\sigma_x^2$ unknown$)$		
$H_1: \mu_x > \mu_0$	$t > t_{1-\alpha}$	u.m.p.i.	$X' = cX \quad c > 0$
$H_1': \mu_x \neq \mu_0$	$\lvert t\rvert > t_{1-\frac{\alpha}{2}}$	u.m.p.u.i.	$X' = cX$ c arbitrary
3. $H_0: \sigma_x^2 = \sigma_0^2$	$(\mu$ known$)$		
$H_1: \sigma_x^2 > \sigma_0^2$	$\sum (X_i - \mu)^2 > \chi^2_{1-\alpha}(m)$	u.m.p.	
$H_1': \sigma_x^2 \neq \sigma_0^2$	$\sum (X_i - \mu)^2 < K_1$ or $> K_2$ (for equations for K_1, K_2 see problem 70)	u.m.p.u.	
4. $H_0: \sigma_x^2 = \sigma_0^2$	$(\mu$ unknown$)$		
$H_1: \sigma_x^2 > \sigma_0^2$	$\sum (X_i - \bar{X})^2 > \chi^2_{1-\alpha}(m-1)$	u.m.p.i.	$X' = X + a$
$H_1': \sigma_x^2 \neq \sigma_0^2$	$\sum (X_i - \bar{X})^2 < K_1$ or $> K_2$ (for K_1, K_2 see problem 70 - use m-1 d.f.)	u.m.p.u.i.	$X' = X + a$

Alternative Hypothesis	Test: Reject H if	Classification (for H against the given alternative)	Invariant under transformations of the form
5. $H_0: \mu_x = \mu_y$	$\left(\sigma_x^2, \sigma_y^2 \text{ known}\right)$		
$H_1: \mu_x > \mu_y$	$\dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sigma_x^2}{m} + \dfrac{\sigma_y^2}{n}}} > z_{1-\alpha}$	u.m.p.i.	$X' = X + a$ $Y' = Y + a$
$H_1': \mu_x \neq \mu_y$	$\dfrac{\lvert\bar{X} - \bar{Y}\rvert}{\sqrt{\dfrac{\sigma_x^2}{m} + \dfrac{\sigma_y^2}{n}}} > z_{1-\frac{\alpha}{2}}$	u.m.p.u.i.	$X' = X + a$ $Y' = Y + a$
6. $H_0: \mu_x = \mu_y$	$\left(\sigma_x^2, \sigma_y^2 \text{ unknown}\right)$		
a. $\sigma_x^2 = \sigma_y^2$	$\bar{X}, \bar{Y}, s_p^2 = \dfrac{\sum (X_i-\bar{X})^2 + \sum (Y_i-\bar{Y})^2}{m+n-2}$ are sufficient for μ_x, μ_y, σ^2		
$H_1: \mu_x > \mu_y$	$\dfrac{\bar{X} - \bar{Y}}{s_p\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} > t_{1-\alpha}(m+n-2)$	u.m.p.i.	$X' = aX + b$ $Y' = aY + b$
$H_1': \mu_x \neq \mu_y$	$\dfrac{\lvert\bar{X} - \bar{Y}\rvert}{s_p\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} > t_{1-\frac{\alpha}{2}}(m+n-2)$	u.m.p.u.i.	$X' = aX + b$ $Y' = aY + b$
b. $\sigma_x^2 \neq \sigma_y^2$	$\bar{X}, \bar{Y}, s_x^2, s_y^2$ are sufficient statistics for $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$.		

This is the classical Fisher-Behrens Problem--no exact test is known. The approximations thus far have tried to keep the size of the test under control, and very little attention has been paid to the power. The approximation used is:

$$t' = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}}}$$

t' is approximately distributed as t with modified d.f. (i.e., modifications by: Smith-Satterthwaite, Cochran-Cox, and Dixon-Massey)

Ref: Anderson and Bancroft, p. 80.

Tables for t' in the one-sided case ($\alpha = 0.05, 0.01$) have been given by Aspen in Biometrika, 1949.

If $m = n$ $$t' = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_x^2 + s_y^2}{n}}}$$

$$t = \frac{\bar{X} - \bar{Y}}{\left(\dfrac{(n-1)\ s_x^2 + (n-1)\ s_y^2}{2(n-1)}\right)^{\frac{1}{2}} \left(\dfrac{1}{n} + \dfrac{1}{n}\right)^{\frac{1}{2}}} = t'$$

Test proposed, when $m = n$, is to reject H if $t > t_{1-\alpha}(n-1)$ for the one-sided case.

Empirical results:

based on 1000 samples, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$:

$m = n = 15$	significance level	5%	1%
	actual rejections	4.9%	1.1%

rejections based on 28 d.f.

$m = n = 5$	significance level	5%	1%
	actual rejections	6.4%	1.8%

rejections based on 8 d.f.

In re power in the Fisher-Behrens problem:

Ref: Gronow; Biometrika, 1951, pp. 252-256

He gives a note on the power of the U and the t tests for the 2 sample problems with unequal variances.

With $n_1 \neq n_2$, $\sigma_1^2 \neq \sigma_2^2$, the U test stays fairly close to α in size, whereas the t-test jumps wildly and a comparison of the power becomes very difficult.

Alternative Hypothesis	Test: Reject H if	Classification (for H against the given alternative)	Invariant under transformations of the form
7. $H_0: \sigma_x^2 = \sigma_y^2$ $(\mu_x, \mu_y \text{ known})$			
$H_1: \sigma_x^2 > \sigma_y^2$	$\dfrac{\sum (X_i - \mu_x)^2 / m}{\sum (Y_i - \mu_y)^2 / n} > F_{1-\alpha}(m,n)$	u.m.p.i.	$X' = aX + b$ $Y' = (\pm a)Y + c$
$H_1': \sigma_x^2 \neq \sigma_y^2$	$\dfrac{\sum (X_i - \mu_x)^2 / m}{\sum (Y_i - \mu_y)^2 / n} < F_1$ or $> F_2$	u.m.p.u.i.	$X' = aX + b$ $Y' = (\pm a)Y + c$

Where F_1, F_2 are chosen to satisfy the unbiasedness conditions as in problem 71.

Since F_1, F_2 depend on complicated equations, we usually use the test:

Reject H_0 if $\max\left(\dfrac{s_x^2}{s_y^2}, \dfrac{s_y^2}{s_x^2}\right) > F_{1-\frac{\alpha}{2}}$ (m,n or n,m)

d.f. depending on which term is in the numerator.

8. $H_0: \sigma_x^2 = \sigma_y^2$ $(\mu_x, \mu_y \text{ unknown})$

maximal invariant statistic is $\dfrac{\sum (X_i - \bar{x})^2}{\sum (Y_i - \bar{y})^2}$

Alternative Hypothesis	Test: Reject H if	Classification	Invariant under
$H_1: \sigma_x^2 > \sigma_y^2$	$\dfrac{\sum (X_i - \bar{x})^2 / m-1}{\sum (Y_i - \bar{y})^2 / n-1} > F_{1-\alpha}(m-1, n-1)$	u.m.p.i.	$X' = aX + b$ $Y' = (\pm a)Y + c$
$H_1': \sigma_x^2 \neq \sigma_y^2$	$\dfrac{\sum (X_i - \bar{x})^2 / m-1}{\sum (Y_i - \bar{y})^2 / n-1} < F_1$ or $> F_2$	u.m.p.u.i.	$X' = aX + b$ $Y' = (\pm a)Y + c$

Same comments on the determination of F_1 and F_2 apply as in No. 7, and the same alternative approach is usually taken (with the appropriate modification of d.f.)

3. Maximum Likelihood Ratio Tests:

H_0: $\theta \in \omega_0$ $\qquad\qquad$ H_1: $\theta \in \Omega - \omega_0$

observations: $X_1, X_2, \ldots, X_n$

$$\lambda = \frac{\max\limits_{\theta \in \omega_0} f(\underline{x}, \theta)}{\max\limits_{\theta \in \Omega} f(\underline{x}, \theta)}$$

Maximum Likelihood Ratio Test is to reject H_0 if $\lambda < \lambda_0$ where λ_0 is chosen to satisfy the size conditions.

For small samples in general this procedure may not give reasonable results. For large samples, as with the m.l.e., results are fairly good.

Remark: Under suitable regularity conditions $-2 \ln \lambda$ is asymptotically distributed as χ^2. The degrees of freedom depend upon the number of parameters specified by the hypothesis; i.e., if $\underline{\theta}$ has m components in Ω and k component in ω_0 then the d.f. in the asymptotic distribution of $-2 \ln \lambda$ (under H_0) are $(m - k)$.

Proof: See S. S. Wilks: Annals of Math Stat, 1938, or "Mathematical Statistics", Princeton University Press

Wald has proven that the m.l.r. test is asymptotically most powerful or asymptoticall most powerful unbiased test.

Ref: A. Wald: Annals of Math Stat, 1941
Transactions of American Math Society, 1943
(both papers in his collected papers)

Example of m.l.r. test:

$X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ $\qquad$ σ^2 unknown

H_0: $\mu = 0$ $\qquad\qquad$ H_1: $\mu \neq 0$

$$f(x; \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}}$$

When H_0 is true, $\dfrac{\sum X_i^2}{n}$ is the m.l.e. of σ^2.

And in general, $\bar{X}$, $\dfrac{\sum (X_i - \bar{X})^2}{n}$ are the m.l.e. of μ, σ^2.

$$\lambda = \frac{\dfrac{1}{\left[2\pi \dfrac{\sum X_i^2}{n}\right]^{\frac{n}{2}}} e^{-\dfrac{n\sum X_i^2}{2\sum X_i^2}}}{\dfrac{1}{\left[2\pi \dfrac{\sum (X_i - \bar{X})^2}{n}\right]^{\frac{n}{2}}} e^{-\dfrac{n\sum (X_i - \bar{X})^2}{2\sum (X_i - \bar{X})^2}}}$$

$$\lambda = \left[\frac{\sum (X_i - \bar{X})^2}{\sum X_i^2}\right]^{\frac{n}{2}}$$

$$\lambda^{\frac{2}{n}} = \frac{\sum (X_i - \bar{X})^2}{\sum X_i^2} = \frac{(n-1)\, s^2}{(n-1)\, s^2 + n\bar{x}^2}$$

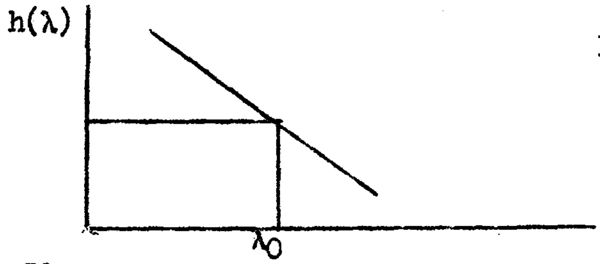
$$\lambda^{-\frac{2}{n}} = 1 + \left(\frac{n}{n-1}\right)\frac{\bar{X}^2}{s^2}$$

$$\sqrt{n-1}\ (\lambda^{-2/n} - 1)^{1/2} = \frac{\sqrt{n}\ |\bar{X}|}{s} = t$$

$h(\lambda) = \sqrt{n-1}\ (\lambda^{-2/n} - 1)^{1/2}$ can be used if it is a monotone function of λ.

$$h'(\lambda) = -\sqrt{n-1}\,(\tfrac{1}{2})\,(\lambda^{-2/n}-1)^{-1/2}\,\lambda^{-(2/n)-1}\,(\tfrac{2}{n}) < 0$$

Therefore $h(\lambda)$ is a monotone decreasing function of λ.



Reject H_0 when: $\lambda < \lambda_0$

$h(\lambda) > h_0$

$|t| > t_0$

Problem 72: Find the m.l.r. test for H_0: $\mu = 0$ against H_1: $\mu > 0$. Does it reduce to the u.m.p.i. test?

Problem 73: $X_1, X_2, \ldots, X_n$ are NID(μ, σ^2) $\qquad$ μ, σ^2 unknown

H_0: $\frac{\sigma}{\mu} = c_o$ (specified) $\qquad$ H_1: $\frac{\sigma}{\mu} = c_1 > c_o$

Find a u.m.p.i. test for H_0 against H_1.

Perform the test where $\bar{X} = 5$, $s = 12$, $c_o = 1$, $n = 20$

---

4. General Linear Hypothesis:

assumptions: X_i are NID(μ_i, σ^2) $\qquad$ $i = 1, 2, \ldots, N$

$$\mu_i = \sum_{j=1}^{p} a_{ij}\,\beta_j$$

$\beta_1, \ldots, \beta_p$ are unknown parameters

$p \leq N$

rank $A = (a_{ij})$ is p

$$H_0: \quad \sum_{j=1}^{p} b_{kj}\,\beta_j = 0$$

$k = 1, 2, \ldots, s \qquad s \leq p$

i.e., s linearly independent equations in the parameters.

a's, b's known

Alternative formulation:

We may solve for $\beta_1, \beta_2, \ldots, \beta_p$ in terms of p of the μ's and then get as the assumptions:

$$\sum_{i=1}^{N} \lambda_{\ell i}\,\mu_i = 0 \qquad \ell = 1, 2, \ldots, N-p$$

H_0 can be written as an additional set of s equations in the μ's:

$$H_0. \quad \sum_{i=1}^{N} \rho_{ki}\,\mu_i = 0 \qquad k = 1, 2, \ldots, s$$

Example: X_{ij} are $N(\mu_{ij}, \sigma^2)$

$\mu_{ij} = \mu + \alpha_i + \beta_j$ i.e., the 2-factor, no interaction model

$\sum \alpha_i = \sum \beta_j = 0$ $i = 1, 2, \ldots, a$ $j = 1, 2, \ldots, b$

ab observations

Parameters are $\mu, \alpha_1, \ldots, \alpha_{a-1}, \beta_1, \ldots, \beta_{b-1}$

$a + b - 1$ parameters in all

$$H_0: \begin{aligned} \alpha_1 &= 0 \\ \alpha_2 &= 0 \\ &\ldots \\ \alpha_{a-1} &= 0 \end{aligned}$$

are the a-1 equations in the parameters

Lemma: If $\sum_{j=1}^{n} a_{ij} y_j$ $(i = 1, 2, \ldots, m)$ are m linearly independent equations in n unknowns $(m \leq n)$ then there exists an equivalent set of equations with matrix C which is orthogonal, i.e.

$$\sum_{j=1}^{n} c_{ij}^2 = 1 \qquad \text{for } i = 1, 2, \ldots, m$$

$$\sum_{j=1}^{n} c_{ij} c_{\ell j} = 0 \qquad i \neq \ell$$

Proof: Ref: Mann, Analysis and Design of Experiments.

given m equations in n unknowns:

$$\begin{aligned} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n &= 0 \\ a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n &= 0 \\ \ldots \quad \ldots \quad & \ldots \\ a_{m1}y_1 + a_{m2}y_2 + \ldots + a_{mn}y_n &= 0 \end{aligned}$$

The first row in the equivalent set is determined by setting:

$$c_{1j} = \frac{a_{1j}}{\sum_{\ell=1}^{n} a_{1\ell}^2} \qquad , \qquad \sum_{j=1}^{n} c_{1j}^2 = 1$$

To get the second row:

$$c'_{2j} = a_{2j} - \lambda c_{1j}$$

$$\sum_j c_{1j} c'_{2j} = \sum_j c_{1j} a_{2j} - \lambda \sum_j c_{1j}^2$$

This = 0 (as required) if $\lambda = \sum_j c_{1j} a_{2j}$

Thus, we set $c_{2j} = \frac{c'_{2j}}{\sum_{\ell=1}^{n} (c'_{2\ell})^2}$ which will hold if the denominator $\neq 0$ -- but by virtue of the independence of the original equations the equality can not hold.

For the third row:

$$c'_{3j} = a_{3j} - \lambda_1 c_{1j} - \lambda_2 c_{2j}$$

$$\sum_j c_{1j} c'_{3j} = \sum_j a_{3j} c_{1j} - \lambda_1 (1) - \lambda_2 (0)$$

This = 0 if $\lambda_1 = \sum_j a_{3j} c_{1j}$

Similarly $\lambda_2 = \sum_j a_{3j} c_{2j}$

Finally we set:

$$c_{3j} = \frac{c'_{3j}}{\sum_{\ell=1}^{n} (c'_{3j})^2}$$

The completion of the proof follows readily by induction.

In the alternative formulation of the general linear hypothesis, we had the followin N-p+s ($\leq$ N) linearly independent equations:

$$\sum_{i=1}^{N} \lambda_{\ell i} \mu_i = 0 \qquad \ell = 1, 2, \ldots, N-p$$

$$\sum_{i=1}^{N} \rho_{ki} \mu_i = 0 \qquad k = 1, 2, \ldots, s$$

From the lemma we may assume these form the first N-p+s rows of an orthogonal matrix. These can be extended to form a complete orthogonal matrix (NxN) -- this is a well known matrix algebra lemma, proof is in Cramer or Mann. If we call the complete orthogonal matrix Λ, we have

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1N} \\ \vdots & & \vdots \\ \lambda_{N-p,\,1} & \cdots & \lambda_{N-p,\,N} \\ \rho_{11} & \cdots & \rho_{1N} \\ \vdots & & \vdots \\ \rho_{s1} & \cdots & \rho_{sN} \\ \tau_{11} & \cdots & \tau_{1N} \\ \vdots & & \vdots \\ \tau_{p-s,\,1} & \cdots & \tau_{p-s,\,N} \end{pmatrix}$$

λ's, ρ's orthogonalize

τ's added to complete the matrix.

$$\underline{Y} = \Lambda \underline{X}$$

$$E(Y_\ell) = \sum_{i=1}^{N} \lambda_{\ell i} \mu_i \qquad \ell = 1, 2, \ldots, N-p$$

$$= 0 \qquad \text{by the assumptions}$$

$$E(Y_{N-p+k}) = \sum_{i=1}^{N} \rho_{ki} \mu_i \qquad k = 1, 2, \ldots, s$$

$$= 0 \qquad \text{if } H_0 \text{ is true}$$

Y's are independent normal variables with means as shown and variances σ^2 (an orthogonal transformation changes NID variables into other NID variables with the same variances).

This is the canonical form of the general linear hypothesis.

Y_i are NID (μ_i, σ^2) where $\mu_i = 0$ $i = 1, 2, \ldots, N-p$

H_0: $\mu_i = 0$ $i = N-p+1, \ldots, N-p+s$

H_1: one or more of these s μ_i's $\neq 0$

M.L.R. Test for the canonical form:

$$f(\underline{y}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^N} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N-p} Y_i^2 + \sum_{i=N-p+1}^{N-p+s} (Y_i - \mu_i)^2 + \sum_{i=N-p+s+1}^{N} (Y_i - \mu_i)^2\right]}$$

m.l.e. of the μ's in Ω are obtained by setting $\mu_i = Y_i$ $(i = N-p+1, \ldots, N)$

and $\hat{\sigma}^2 = \sum_{i=1}^{N-p} Y_i^2 / N$

m.l.e. of the μ's in ω are found by setting $\mu_i = Y_i$ $(i = N-p+s+1, \ldots, N)$

and $\hat{\sigma}^2 = \sum_{i=1}^{N-p+s} Y_i^2 / N$

$$\lambda = \frac{\left(\frac{1}{\hat{\sigma}_\omega}\right)^N e^{-N/2}}{\left(\frac{1}{\hat{\sigma}_\Omega}\right)^N e^{-N/2}} = \left(\frac{\hat{\sigma}_\Omega^2}{\hat{\sigma}_\omega^2}\right)^{\frac{N}{2}}$$

$$= \left(\frac{\sum_{i=1}^{N-p} Y_i^2}{\sum_{i=1}^{N-p+s} Y_i^2}\right)^{\frac{N}{2}} = \left(\frac{Q_a}{Q_r}\right)^{N/2}$$

a = absolute minimum (nothing specified about the hypothes
r = relative minimum

$$Q_r = Q_a + \sum_{i=N-p+1}^{N-p+s} Y_i^2$$

$$\lambda^{\frac{2}{N}} = \frac{Q_a}{Q_r}$$

$$\lambda^{-\frac{2}{N}} = \frac{Q_r}{Q_a} = 1 + \frac{Q_r - Q_a}{Q_a}$$

$$\lambda^{-\frac{2}{N}} - 1 = \frac{Q_r - Q_a}{Q_a}$$

We can use any monotone function of λ for test purposes, therefore the test is:

Reject H_0 if $\frac{Q_r - Q_a}{Q_a} > K$

which is equivalent to rejecting H_0 if $\lambda < \lambda_0$.

$$Q_r - Q_a = \sum_{i=N-p+1}^{N-p+s} Y_i^2 \quad \text{is} \quad \chi^2_{(s)}$$

$$Q_a = \sum_{i=1}^{N-p} Y_i^2 \quad \text{is} \quad \chi^2_{(N-p)}$$

Therefore $\frac{(Q_r - Q_a)/s}{Q_a/N-p}$ has the F-distribution with parameters (s, N-p)

If H_1 is true, then $E(Y_i) = d_i$ $\quad i = N-p+1, \ldots, N-p+s$

$$\sum d_i^2 > 0$$

Recall that $\sum_{i=N-p+1}^{N-p+s} Y_i^2$ is the sum of squares of variables with non-zero means, and hence has a non-central χ^2-distribution with parameters $(s, \sum_{i=1}^{s} d_i^2)$.

The ratio of a non-central χ^2 to a central χ^2 is a non-central F, thus $\frac{(Q_r - Q_a)/s}{Q_a/N-p}$ under H_1 has a non-central F-distribution with parameters $(s, N-p, \sum_{i=1}^{s} d_i^2)$.

Thus the problem is solved in terms of the Y's (the orthogonalized X's).

The original problem was:

X's are NID (μ_i, σ^2)

1) $\sum_{i=1}^{N} \lambda_{\ell i}\, \mu_i = 0 \qquad \ell = 1, 2, \ldots, N-p$

2) $\sum_{i=1}^{n} \rho_{ki}\, \mu_i = 0 \qquad k = 1, 2, \ldots, s$

M.L.R. test:

define: $Q_a' = \min \sum (X_i - \mu_i)^2$ under restrictions (1)

$Q_r' = \min \sum (X_i - \mu_i)^2$ under restrictions (1) and (2)

$\hat{\sigma}$ (in Ω) $= (Q_a')^{1/2}$

$\hat{\sigma}$ (in ω) $= (Q_r')^{1/2}$

$$\lambda = \left(\frac{Q_a'}{Q_r'}\right)^{N/2} \frac{e^{-N/2}}{e^{-N/2}}$$

As in the other case, this is a monotone decreasing function of $\frac{Q_r' - Q_a'}{Q_a'}$

Recall that under an orthogonal transformation, sums of squares are preserved.

Thus: Q_a' is carried into Q_a

Q_r' is carried into Q_r

Hence: $\frac{Q_r' - Q_a'}{Q_a'} = \frac{Q_r - Q_a}{Q_a}$ and has the same distributions under H_0 (F-distributio[n] and under H_1 (non-central F).

<u>Problem 74</u>: X_{ijk} is $N(\mu_{ij}, \sigma^2)$ $\quad i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$\sum_i \alpha_i = \sum_j \beta_j = 0$$

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$

H_0: all $\alpha_i = 0$ $\qquad$ H_0': all $\alpha_i = 0$; all $(\alpha\beta)_{ij} = 0$

Find the usual F-test for 1) H_0

2) H_0'

Power of the ANOVA test:

Power of the ANOVA test has been tabled by Tang (Statistical Research Memoirs, Vol. 2 for $\alpha = 0.05$, $\alpha = 0.01$ for various values of:

$$\phi = \left[\frac{2\lambda}{s+1}\right]^{1/2} \qquad \lambda = \frac{\sum d_i^2}{2\sigma^2}$$

Tables in: Mann
Kempthorne

$$2\sigma^2\lambda = \sum_{k=1}^{s} d_i^2 = \sum_k \left(\sum_{i=1}^{N} \rho_{ki}\, \mu_i\right)^2$$

$$Y_{N-p+k} = \sum_{i=1}^{N} \rho_{ki}\, X_i$$

$$Q_r - Q_a = \sum_{k=1}^{s} Y^2_{N-p+k} = \sum_{k=1}^{s} \left(\sum_{i=1}^{N} \rho_{ki}\, X_i\right)^2$$

hence $2\sigma^2\lambda$ is $Q_r - Q_a$ with X_i replaced by $E(X_i) = \mu_i$

Example: X_{ijk} is $N(\mu_{ij}, \sigma^2)$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

or $X_{ijk} = \mu + \alpha_i + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$

$$\sum_i \alpha_i = 0 \qquad \sum_j \beta_j = 0 \qquad \sum_i (\alpha\beta)_{ij} = 0 \qquad \sum_j (\alpha\beta)_{ij} = 0$$

H_0: $\alpha_i = 0$ $\qquad$ H_1: α_i not all zero

$$Q_r - Q_a = \sum\sum\sum (\bar{X}_{i..} - \bar{X}_{...})^2$$

under H_1: $E(\bar{X}_{i..}) = \mu + \alpha_i + 0 + 0$

$$E(\bar{X}_{...}) = \mu$$

$$E(\bar{X}_{i..} - \bar{X}_{...}) = \alpha_i$$

hence $2\sigma^2\lambda = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \alpha_i^2 = nb \sum_{i=1}^{a} \alpha_i^2$

thus $\phi = \left[\frac{2nb \sum \alpha_i^2}{2\sigma^2 \; a}\right]^{1/2} = \left[\frac{nb}{\sigma^2} \frac{\sum \alpha_i^2}{a}\right]^{1/2}$

<u>Exercise</u>: If $a = 3$, $b = 5$ how large should n be so that we can detect

$\alpha_1 = -0.25\sigma$; $\alpha_2 = -0.25\sigma$; $\alpha_3 = 0.50\sigma$

with probability .75 ($\alpha = 0.05$).

Try n, calculate ϕ, enter the tables and find the power. After successive trials n will be obtained to give the required power. (d.f. for the numerato = 2; for the denominator = ab(n-1) = 15(n-1). Verify that n = 13.

<u>Randomized blocks:</u>

$X_{ij} \mid \mu_{ij}$ are $N(\mu_{ij}, \sigma^2)$ $\qquad i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$

$\mu_{ij} = \mu + \alpha_i + b_j$

b_j are $N(0, \sigma_b^2)$ (this is the additional assumption of randomized blocks)

or $X_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}$ where ε_{ij} is $N(0, \sigma^2)$

b_j is $N(0, \sigma_b^2)$

and they are independent.

H_0: all $\alpha_i = 0$ $\qquad H_1$: some $\alpha_i \neq 0$

Given b_j, $\bar{X}_{i.}$ are $N(\mu + \alpha_i + \frac{\sum_{i=1}^{b} b_j}{b}; \frac{\sigma^2}{b})$

If H_0 is true, $\sqrt{b}\ \bar{X}_{i.}$ is $N(\mu', \sigma^2)$, and $\sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{i.} - \bar{X}_{..})^2 = b \sum_{i=1}^{a} (\bar{X}_{i.} - \bar{X}_{..})^2$ has a χ^2-distribution with a-1 d.f. This conditional (on the b_j) distribution does not involve the b_j, therefore the unconditional distribution of

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{i.} - \bar{X}_{..})^2 \text{ is } \chi^2(a-1).$$

In general, given b_j,

$$\sum_i \sum_j (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 \text{ is } \chi^2 \ (a-1)(b-1).$$

Again we have the same result for the unconditional distribution.

Hence a test of H_0 is based on the statistic:

$$\frac{\sum_i \sum_j (\bar{X}_{i.} - \bar{X}_{..})^2 \big/ a-1}{\sum_i \sum_j (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 \big/ (a-1)(b-1)}$$

which has the F-distribution with $(a-1)$, $(a-1)(b-1)$ d.f.

If H_0 is false, $\sqrt{b}\ \bar{X}_{i.}$ is $N(\mu' + \sqrt{b}\ \alpha_i, \sigma^2)$. $\sum\sum(\bar{X}_{i.} - \bar{X}_{..})^2$ has a non-central χ^2-distribution with parameter $b\sum\alpha_i^2$, d.f. $a-1$.

The Power is also calculated as if the b's were fixed parameters.

Random model: (Heirarchical Classification or Nested Sampling)

$X_{ij} \big| \mu_{ij}$ is $N(\mu_{ij}, \sigma^2)$ $\qquad$ $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, n$

$\mu_{ij} = \mu + a_i$

where the a_i are $N(0, \sigma_a^2)$

or $X_{ij} = \mu + a_i + \varepsilon_{ij}$

ANOVA Table

	d.f.	S.S.	E(MS)
a's	a-1	$\sum\sum(\bar{X}_{i.}-\bar{X}_{..})^2$	$\sigma^2 + n\,\sigma_a^2$
error	a(n-1)	$\sum\sum(X_{ij}-\bar{X}_{i.})^2$	σ^2

Given the a_i, X_{ij} is $N(\mu + a_i, \sigma^2)$

$$\frac{\sum_i \sum_j (\bar{X}_{ij} - \bar{X}_{i.})^2}{\sigma^2} \text{ has a } \chi^2\text{-distribution with } a(n-1) \text{ d.f.}$$

Since this is conditional on the a_i, but independent of them, it is also the unconditional distribution.

$\bar{X}_{i.}$ is $N(\mu, \frac{\sigma^2}{n} + \sigma_a^2)$ -- unconditional distribution.

e.g., in terms of moment generating functions

$$E_{ai}\left[E_{\bar{X}i}(e^{\bar{X}_{i.}t} \mid a_i)\right] = e^{\mu t + \frac{t^2}{2}\left[\frac{\sigma^2}{n} + \sigma_a^2\right]}$$

$$\frac{n \sum_{i=1}^{a} (\bar{X}_{i.} - \bar{X}_{..})^2}{\sigma^2 + n\,\sigma_a^2} \text{ is } \chi^2 \text{ with a-1 d.f.}$$

H_0: $\sigma_a^2 = 0$ leads to the usual F-test against H_1: $\sigma_a^2 > 0$, i.e.,

$$\frac{\sum\sum(\bar{X}_{i.} - \bar{X}_{..})^2}{\sum\sum(X_{ij} - \bar{X}_{i.})^2} \cdot \frac{\sigma^2}{\sigma^2 + n\,\sigma_a^2}$$

has an F-distribution with $\left[(a-1),\ a(n-1)\right]$ d.f.

<u>Problem 75</u>: $X_{ijk} \mid \mu_{ij}$ is $N(\mu_{ij}, \sigma^2)$ $\quad i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$

a) $\mu_{ij} = \mu + \alpha_i + b_{ij}$ where b_{ij} are $N(0, \sigma_b^2)$

b) $\mu_{ij} = \mu + a_i + b_{ij}$ where b_{ij} are $N(0, \sigma_b^2)$

a_i are $N(0, \sigma_a^2)$

and the a's and b's are independent.

For $n = 2$; $b = 2$; $a = 5$; $\alpha = 0.05$, plot the power of the tests:

1) in (a) against $\dfrac{\sum \alpha_i^2}{\sigma^2 + 2\sigma_b^2}$

2) in (b) against $\dfrac{\sigma_a^2}{\sigma^2 + 2\sigma_b^2}$

Multiple Comparisons:

Tests of all contrasts with the general linear hypothesis model following ANOVA.

Tukey's Procedure:

observations: $X_1, X_2, \ldots, X_n$ are NID$(0, \sigma^2)$

let $R = \max_{1 \le i \le n} \{X_i\} - \min_{1 \le i \le n} \{X_i\}$

${s_e}^2$ is an estimate of σ^2 which is independent of $X_1, X_2, \ldots, X_n$ with f d.f.

(i.e., $f\,{s_e}^2/\sigma^2$ has a χ^2-distribution with f d.f.)

The distribution of R/σ can be found in a general form, and in particular when the X's are normal (since the X_i/σ are N(0, 1) -- it is free of any parameters.

The distribution of $R' = \frac{R}{s_e} = \frac{R/\sigma}{s_e/\sigma}$ = Studentized Range has been found by numerical integration and percentage points have been tabulated by Hartley and Pearson in Biometrika Tables, Table 29.

Consider the one-way ANOVA situation:

X_{ij} are NID (μ_i, σ^2) $\qquad i = 1, 2, \ldots, a$; $\quad j = 1, 2, \ldots, n_i = n$

$\bar{X}_{i.}$ are NID $(\mu_i, \frac{\sigma^2}{n})$

$$MSE = \frac{\sum\sum (X_{ij} - \bar{X}_{i.})^2}{a(n-1)}$$ is an independent estimate of σ^2 with a(n-1) d.f.

Consider H_0: $\mu_i = \mu_k$ or $\mu_i - \mu_k = 0$

Given H_0:

$$Pr\left[\frac{\sqrt{n}\,(\bar{X}_{i.} - \bar{X}_{k.})}{\sqrt{\frac{\sum\sum (X_{ij} - X_{i.})^2}{a(n-1)}}} > t\right] \le Pr\left[R' > t\right] = \alpha$$

This t is the statistic Q(a, f) in Snedecor, section 10.6, p. 251 where f = d.f. = a(n-1) in this case.

This inequality will hold for any possible pairwise comparison (i.e., any i, k).

Frame of reference is the total experiment, not the individual pairwise tests -- i.e., for the total experiment and making all possible pairwise tests, the probability of the Type I error is $\leq \alpha$, where α is the level chosen for the Q(a,f) statistic.

Hence we reject H: $\mu_1 - \mu_k = 0$ for any i, k when

$$\frac{\sqrt{n}\,(\bar{X}_{i.} - \bar{X}_{k.})}{\sqrt{MSE}} > Q_\alpha(a,f)$$

Scheffe Test for all contrasts (most general of all).

Ref: Scheffe, Biometrika, June 1952

X_{ij} are NID(μ_i, σ^2) $\qquad$ $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, n_i$

s_e^2 is an estimate of σ^2 which is independent of $\bar{X}_{i.}$ with f d.f.

$$H_0(\underline{c}): \sum_{i=1}^{a} c_i\mu_i = 0 \qquad \sum_{i=1}^{a} c_i = 0$$

Consider the totality of all such $H_0(\underline{c})$.

Theorem 38: If the μ_i are all equal, then

$$\Pr\left[\frac{\left\{\sum_{i=1}^{a} c_i\,(\bar{X}_{i.} - \bar{X}_{..})\right\}^2}{s_e^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}} > t\right] \qquad (a-1).\,c\int_0^t F^{\frac{a-1}{2}-1}\left(1 + \frac{a-1}{f}F\right)^{-\frac{a-1+f}{2}} dF$$

$$\text{where } c = \left(\frac{a-1}{f}\right)^{\frac{a-1}{2}} \frac{\Gamma\left(\frac{a-1+f}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)\Gamma\left(\frac{f}{2}\right)}$$

hence, if we reject any such $H_0(\underline{c})$ when

$$\frac{\left\{\sum_{i=1}^{a} c_i\,(\bar{X}_{i.} - \bar{X}_{..})\right\}^2}{s_e^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}} > (a-1)\,F_{a-1,\,f}(1-\alpha)$$

then the type I error $\leq \alpha$.

Proof: Under H_0 $\sum c_i \bar{X}_{..} = 0$ and $\sum c_i \mu_i = 0$. For all c_i with $\sum c_i = 0$:

$$\Pr\left\{\frac{(\sum c_i \bar{X}_{i.})^2}{s_e^2 \sum \frac{c_i^2}{n_i}} > t\right\} \leq \Pr\left\{\max_{\substack{c_i \\ \sum c_i = 0}} \frac{\left[\sum_i \frac{\sqrt{n_i}\, c_i}{\sqrt{n_i}} (\bar{X}_{i.} - \mu_i)\right]^2}{s_e^2 \sum \frac{c_i^2}{n_i}} > t\right\}$$

define: $$Y_i = \frac{\sqrt{n_i}\,(\bar{X}_{i.} - \mu_i)}{\sigma}$$ Y_i are NID(0, 1)

$i = 1, 2, \ldots, a$

$$d_i = \frac{c_i / \sqrt{n_i}}{\sum_i c_i^2 / n_i} \qquad \sum c_i = \sum \sqrt{n_i}\, d_i = 0$$

$$\sum d_i^2 = 1$$

Expression we wish to maximize is $\dfrac{\left(\sum d_i Y_i\right)^2}{s_e^2 / \sigma^2}$ with respect to d_i

$$\phi = \left[\sum d_i Y_i\right]^2 - \lambda_1 \sum \sqrt{n_i}\, d_i - \lambda_2 \sum d_i^2 \qquad [1]$$

$$\frac{\partial \phi}{\partial d_j} = 2\left[\sum d_i Y_i\right] Y_j - \lambda_1 \sqrt{n_j} - 2\lambda_2 d_j = 0 \qquad j = 1, 2, \ldots, a$$

multiply [1] by $\sqrt{n_j}$ and sum over j:

$$\frac{2\left[\sum \sqrt{n_j}\, Y_j\right]\left[\sum d_i Y_i\right]}{\sum n_j} = \lambda_1 \quad \text{since } \sum \sqrt{n_i} d_i = 0$$

multiply [1] by d_j and sum over j:

$$2\left[\sum d_j Y_j\right]\left[\sum d_i Y_i\right] - 0 - 2\lambda_2(1) = 0 \quad \text{since: i) } \sum \sqrt{n_i} d_i = 0$$

$$\text{ii) } \sum d_i^2 = 1$$

$$\lambda_2 = \left[\sum d_i Y_i\right]^2$$

put λ_1, λ_2 in [1] to obtain the j^{th} equation:

$$2Y_j\left[\sum d_iY_i\right] - 2\frac{\sqrt{n_j}\left[\sum\sqrt{n_i}Y_i\right]\left[\sum d_iY_i\right]}{\sum n_j} - 2d_j\left[\sum d_iY_i\right] = 0$$

$$Y_j - \frac{\sqrt{n_j}\sum\sqrt{n_i}Y_i}{\sum n_i} = d_j\sum d_iY_i$$

multiply this by Y_j and sum over j:

$$\left[\sum d_jY_j\right]^2 = \sum Y_j^{\,2} - \frac{\left[\sum\sqrt{n_j}\,Y_j\right]^2}{\sum n_j}$$

in the n_j are equal, this $= \sum Y_j^{\,2} - \frac{\left[\sum Y_j\right]^2}{a} = \sum_{j=1}^{a}(Y_j - \bar{Y})$

Recall:

$$\Pr\left\{\frac{\left[\sum_{i=1}^{a} c_i\,\bar{X}_{i.}\right]^2}{s_e^{\,2}\sum_i\frac{c_i^{\,2}}{n_i}} > t \quad \text{for all } c_i,\ \sum c_i = 0\right\} \qquad [2]$$

$$\leq \Pr\left\{\max_{\substack{c_i\\ \sum c_i=0}} \frac{\left(\sum c_i\bar{X}_{i.}\right)^2}{s_e^{\,2}\sum\frac{c_i^{\,2}}{n_i}} > t\right\}$$

$$\simeq \Pr\left\{\frac{\sum Y_i^{\,2} - \frac{\left(\sum n_i\,Y_i\right)^2}{\sum n_i}}{s_e^{\,2}\ \sigma^2} > t\right\}$$

$$Y_i = \frac{\sqrt{n_i}\,(\bar{X}_{i.} - \mu_i)}{\sigma_i}$$ and are NID(0, 1).

We can make an orthogonal transformation of the form:

$$Z_1 = \frac{\sum_{i=1}^{a} \sqrt{n_i}\, Y_i}{\left(\sum n_i\right)^{1/2}}$$

$$Z_2 =$$

$$\vdots$$

$$Z_a =$$

Then $$\sum Y_i^2 - \frac{\left(\sum \sqrt{n_i}\, Y_i\right)^2}{\sum n_i} \longrightarrow \sum_{j=1}^{a} Z_j^2 - Z_1^2 = \sum_{j=2}^{a} Z_j^2$$

Hence $$\sum Y_i^2 - \frac{\left(\sum \sqrt{n_i}\, Y_i\right)^2}{\sum n_i}$$ has a χ^2-distribution with (a-1) d.f.

Thus $$\Pr\left[2\right] = \Pr\left[\frac{\sum_{j=2}^{a} Z_j^2 \Big/ a-1}{s_e^2 \big/ \sigma^2} > \frac{t}{a-1}\right]$$

but $$\frac{\sum Z_j^2 \Big/ a-1}{s_e^2 \big/ \sigma^2}$$ has an F-distribution with (a-1, f) d.f.

Therefore the type I error will be less than or equal to α if we reject any $H_0(\underline{c})$ if:

$$\frac{\left[\sum c_i\, \bar{X}_{i.}\right]^2}{s_e^2 \sum \frac{c_i^2}{n_i}} > (a-1)\, F_{1-\alpha}\, (a-1,\, f)$$

5. χ^2-Tests

Single series of trials: Simple hypothesis

classes (or events):	1	2	. . .	k	
probabilities:	p_1	p_2	. . .	p_k	$\sum p_j = 1$
observations:	v_1	v_2	. . .	v_k	$\sum v_j = n$

H_0: $p_j = p_j^0$

If H_0 is true

$$X_j = \frac{v_j - np_j^0}{\sqrt{np_j^0}}$$ is asymptotically normal subject to the restricti[on]

$$\sum \sqrt{p_j^0}\, X_j = 0$$

By an orthogonal transformation it can be shown that

$$\sum_{j=1}^{k} X_j^2 = \sum_{j=1}^{k} \frac{(v_j - np_j^0)^2}{np_j^0}$$ has an asymptotic χ^2-distribution with k-1 d.f. as $n \longrightarrow \infty$.

The power of the test $\longrightarrow 1$ as $n \longrightarrow \infty$.

Power of a simple χ^2-test:

H_1: $p_j = p_j'$ $\qquad p_j' - p_j^0 = \frac{c_j}{\sqrt{n}}$

$$X_j = \frac{v_j - np_j^0}{\sqrt{np_j^0}} = \frac{v_j - np_j'}{\sqrt{np_j'}} \sqrt{\frac{p_j'}{p_j^0}} + \frac{n(p_j' - p_j^0)}{\sqrt{np_j^0}}$$

$$= Y_j e_j + \frac{c_j}{\sqrt{p_j^0}}$$

$$\text{now } e_j = \sqrt{\frac{p_j'}{p_j^0}} = \sqrt{1 + \frac{p_j' - p_j^0}{p_j^0}} = \sqrt{1 + \frac{c_j}{\sqrt{n_j}\, p_j^0}} \longrightarrow 1$$

as $n \longrightarrow \infty$.

Under H_1 the Y_j are asymptotically normal.

With the same type of orthogonal transformation, we have $\sum X_j^2$ under H_1 is a sum of squares of $Y_j + \frac{c_j}{\sqrt{p_j^0}}$ and hence has a non-central χ^2-distribution with parameters $k-1$, $\sum_{j=1}^{k} \frac{c_j^2}{p_j^0}$.

The non-centrality parameter can also be written as $n \sum_{j=1}^{k} \frac{(p_j' - p_j^0)^2}{p_j^0}$

Example: v_i = no. of families with i boys in families of 2 children.

Assume the number of boys is binomially distributed with parameter p.

H_0: $p = 0.5$ H_1: $p = 0.4$

No. boys	p_j^0 under H_0)	p_j' (under H_1)
0	.25	.16
1	.50	.48
2	.25	.36

$$\lambda = n \sum \frac{(p_j' - p_j^0)^2}{p_j^0} = n \left[\frac{(.09)^2}{.25} + \frac{(.02)^2}{.50} + \frac{(.11)^2}{.25} \right]$$

$$= .0816\ n$$

For $n = 100$ $\lambda = 8.16$

The power can be determined from the Tables of the Non-Central χ^2 by E. Fix, University of California Publications in Statistics, Volume 1, No. 2, pp. 15-19.

From the tables, for $\lambda = 8.16$, $k-1 = f = 2$, $\alpha = 0.05$

$$0.7 < \beta \text{ (power)} < 0.8$$

χ^2-test may also be a one-sided test:

H_0: $p_j = p_j^0$ H_1: $p_j > p_j^0$ for $j \leq k'$ (k' exceed p_j^0)

$p_j < p_j^0$ for $j > k'$ ($k - k'$ fall below p_j^0)

χ^2-test is modified as follows:

for $j \leq k'$: if $v_j > E(v_j) = np_j^0$ calculate $\frac{(v_j - np_j^0)^2}{np_j^0}$ as usual

if $v_j < np_j^0$ put the contribution = 0

for $j > k'$: if $v_j > np_j^0$ put the contribution = 0

if $v_j < np_j^0$ calculate χ^2 as usual

χ^2 is rejected if the sample value $> \chi^2_{1-2\alpha}(k-1)$

Example:

$^8(10)$	$^{12}(10)$	20
$^{12}(10)$	$^8(10)$	20
20	20	

H_0: $p = 0.5$

$\chi^2 = 0$ therefore we fail to reject H_0

note: if $v_{11} > 10$ we would then calculate χ^2.

Refs: (on the χ^2-test):

Cochran: 1952 Annals of Math Stat
1954 Biometrics

χ^2-test of a composite hypothesis:

classes:	1	2	...	k
s series of observations:	v_{11}	v_{12}	...	v_{1k}
	.	.	...	.
	.	.	...	.
	.	.	...	.
	v_{s1}	v_{s2}	...	v_{sk}
probabilities:	p_{11}	p_{12}	...	p_{1k}
	.	.	...	.
	.	.	...	.
	.	.	...	.
	p_{s1}	p_{s2}	...	p_{sk}

$i = 1, 2, \ldots, s$
$j = 1, 2, \ldots, k$

$$\sum_{j=1}^{k} v_{ij} = n_i$$

$p_{ij} = f(\underline{\theta})$

$$X_{ij} = \frac{v_{ij} - n_i p_{ij}(\hat{\underline{\theta}})}{\sqrt{n_i p_{ij}(\hat{\underline{\theta}})}}$$

$$\sum_j \sqrt{p_{ij}}\ X_{ij} = 0 \qquad \text{for } i = 1, 2, \ldots, s$$

H_0: $p_{ij} = p_{ij}^0(\underline{\theta})$

Theorem 39: If $\hat{\underline{\theta}}$ is estimated by m.l.e. or any asymptotically equivalent method then as $n \to \infty$, under the regularity conditions as for estimation:

$$X^2 = \sum_i \sum_j \frac{\left[v_{ij} - n_i p_{ij}(\hat{\underline{\theta}})\right]^2}{n_i p_{ij}(\hat{\underline{\theta}})}$$

has a χ^2-distribution with s(k-1)-t d.f. when H is true.
(t = no. of components in $\underline{\theta}$)

If H_1 is true: $p_{ij} = p'_{ij} = p_{ij}^0 + \dfrac{c_{ij}}{\sqrt{n_i}}$ and $\rho = \dfrac{n_i}{\sum n_i}$

then X^2 has a non-central χ^2-distribution with d.f. as before,

and non-centrality parameter $\underline{\delta}' \left[I - B(B'B)^{-1} B' \right] \underline{\delta}$

where $\underline{\delta}'_{1 \times ks} = \left(\dfrac{c_{ij}\sqrt{\rho_i}}{\sqrt{p_{ij}^0}} \right)$ $\qquad i = 1, 2, \ldots, s$; $j = 1, 2, \ldots, k$; $\ell = 1, 2, \ldots, t$

$$B_{ks \times t} = \left[\frac{\sqrt{\rho_i}}{\sqrt{p_{ij}^0}} \left(\frac{\partial p_{ij}}{\partial \theta_\ell} \right)_{\theta_\ell = \theta_\ell^0} \right]$$

Ref: Cramer: χ^2 section
Mitra: December 1958 Annals of Math Stat
Ph.D. Dissertation, UNC, Institute of Statistics Mimeograph Series No. 142.

Note on the general χ^2 tests of composite hypotheses:

$X_{ij} = \dfrac{v_{ij} - n_i p_{ij}}{\sqrt{n_i\, p_{ij}}}$ are asymptotically normal subject to the linear

restrictions that $\sum_j \sqrt{p_{ij}}\, X_{ij} = 0$ for all $i = 1, 2, \ldots, s$ [1]

For sufficiently large n the m.l.e. reduces to solving the equation:

$$\frac{\partial \ln L}{\partial \theta} = \left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta = \theta_0} + \left(\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta = \theta_0}\right)(\theta - \theta_0) \qquad [2]$$

Recall:

$$L = K \prod_i \prod_j \left[p_{ij}(\theta)\right]^{v_{ij}}$$

$$\ln L = K' + \sum_i \sum_j v_{ij} \ln p_{ij}(\theta)$$

$\ln L$ is linear in the v_{ij}, as are $\dfrac{\partial^h \ln L}{\partial \theta^h}$ $\quad h = 1, 2$

Hence the estimation of θ (single component) means one additional linear restriction on the v_{ij} or on the X_{ij} (since the two are linearly related).

If this restriction [2] is linearly independent of the s restrictions in [1], then we can transform the s+1 restrictions to s+1 orthogonal linear restrictions and then by the usual extension get an orthogonal transformation that takes

$\sum_i \sum_j X_{ij}^2$ into $\sum_i \sum_j Y_{ij}^2$ with $s(k-1) - 1 = sk - (s+1)$ terms and hence we get the result of the general χ^2 theorem (theorem 39).

Example: Power of a χ^2 test of homogeniety, 2 x 2 table.

Probabilities:

1)	p_{11}	$p_{12} = 1 - p_{11}$	1
2)	p_{21}	$p_{22} = 1 - p_{21}$	1

H_0: $p_{11} = p_{21} = \theta$ $\qquad p_{12} = p_{22} = 1 - \theta$

H_1: $p_{11} = \theta + c/\sqrt{n}$ $\qquad p_{12} = 1 - \theta - c/\sqrt{n}$

$p_{21} = \theta - c/\sqrt{n}$ $\qquad p_{22} = 1 - \theta + c/\sqrt{n}$

$$\lambda = \underline{\delta}' \left[I - B(B'B)^{-1} B' \right] \underline{\delta}$$

take $\rho_1 = \rho_2 = \frac{1}{2}$ (sampling fraction)

$$B' = \begin{pmatrix} \frac{1}{\sqrt{2\,\theta}} & \frac{-1}{\sqrt{2(1-\theta)}} & \frac{1}{\sqrt{2\,\theta}} & \frac{-1}{\sqrt{2(1-\theta)}} \end{pmatrix}$$

$$B'B = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

$$(B'B)^{-1} = \theta(1-\theta)$$

$$B(B'B)^{-1}B' = \begin{pmatrix} \frac{1-\theta}{2} & \frac{-\theta(1-\theta)}{2} & \frac{1-\theta}{2} & \frac{-\theta(1-\theta)}{2} \\ \frac{-\theta(1-\theta)}{2} & \frac{\theta}{2} & \frac{-\theta(1-\theta)}{2} & \frac{\theta}{2} \\ \frac{1-\theta}{2} & \frac{-\theta(1-\theta)}{2} & \frac{1-\theta}{2} & \frac{-\theta(1-\theta)}{2} \\ \frac{-\theta(1-\theta)}{2} & \frac{\theta}{2} & \frac{-\theta(1-\theta)}{2} & \frac{\theta}{2} \end{pmatrix}$$

$$\underline{\delta}' = \begin{pmatrix} \frac{c}{\sqrt{2\,\theta}} & \frac{-c}{\sqrt{2(1-\theta)}} & \frac{c}{\sqrt{2\,\theta}} & \frac{-c}{\sqrt{2(1-\theta)}} \end{pmatrix}$$

$$\underline{\delta}' \, I \, \underline{\delta} = \frac{c^2}{\theta} + \frac{c^2}{1-\theta} = \frac{c^2}{\theta(1-\theta)}$$

$$\underline{\delta}' \left[B(B'B)^{-1}B'\right] \underline{\delta} = 0$$

(patterns of + and - signs in $\underline{\delta}$, B are opposite and sufficient to cancel all terms.

Hence the non-centrality parameter, $\lambda = \frac{c^2}{\theta(1-\theta)}$

Which can be expressed in terms of the p's as:

$$\theta = \frac{p_{11} + p_{21}}{2} \qquad 1 - \theta = \frac{p_{12} + p_{22}}{2}$$

$$p_{11} - p_{21} = \frac{2c}{\sqrt{n}} \qquad c^2 = \frac{n}{4}(p_{11} - p_{21})^2$$

$$\lambda = \frac{n(p_{11} - p_{21})^2}{(p_{11} + p_{21})(p_{12} + p_{22})}$$

In the December 1958 Annals of Math Stat Mitra gives the following formula for λ in the 2 x k contingency table:

probabilities:	p_{11}	p_{12}	...	p_{1k}
	p_{21}	p_{22}	...	p_{2k}

H_0: $p_{1j} = p_{2j} = \theta_j$ $\qquad \sum \theta_j = 1$

H_1: $p_{1j} = \theta + c_{1j}/\sqrt{n}$ $\qquad \sum c_{1j} = 0$

$p_{2j} = \theta + c_{2j}/\sqrt{n}$ $\qquad \sum c_{2j} = 0$

$$\lambda = \rho_1 \, \rho_2 \sum_{j=1}^{k} \frac{(c_{1j} - c_{2j})^2}{\theta_j}$$

For the above example: $\rho_1 = \rho_2 = \frac{1}{2}$ $\quad c_{1j} = c$ $\qquad c_{2j} = -c$

$$\lambda = \frac{1}{4}\left(\frac{(2c)^2}{\theta} + \frac{(2c)^2}{1 - \theta}\right) = \frac{c^2}{\theta(1-\theta)}$$

---

6. Other Approaches to Testing Hypotheses and other problems:

1. Most Stringent Tests:

Let the family of tests of size α be $\Phi(\alpha)$.

Let $\bar{\beta}(\theta) = \sup_{\phi \, \varepsilon \, \Phi(\alpha)} \beta_{\phi}(\theta)$

ϕ (x) is most stringent if:

i) it is of size α

ii) for any other size α test ϕ'

$$\sup_\theta \left[\bar{\beta}(\theta) - \beta_\phi(\theta)\right] \leq \sup_\theta \left[\bar{\beta}(\theta) - \beta_{\phi'}(\theta)\right]$$

2. Minimax Tests: Decision Theory Approach

$L\left[D(x), \theta\right]$ = the loss when D(x) is the decision made and θ is the true value of the parameter.

Example: $\underline{X}$ is $N(\mu, 1)$ $\qquad H_0: \mu = 0$

$D(\underline{x}) = 1$

$D(\underline{x}) = 0$ means accept $H_0: \mu = 0$

$L\left[D(\underline{x}), \mu\right] = c\,\mu^2$ when $D(\underline{x}) = 0$

$L\left[D(\underline{x}), \mu\right] = c/\mu$ when $D(\underline{x}) = 1$

Having set up a loss function, we then have a risk function defined as:

$$r_D(\theta) = E(L) = \int_{-\infty}^{\infty} L\left[D(\underline{x}), \theta\right] f(\underline{x}, \theta)\, d\underline{x}$$

It is frequently impossible to minimize this universally with respect to θ, thus:

$D(\underline{x})$ is a minimax decision rule if it minimizes (with respect to all possible decision rules).

$$\sup_\theta r(\theta)$$

or expressed another way we determine: $\inf_D \sup_\theta r_D(\theta)$

ref: Blackwell and Girshick, Theory of Games, Wiley

3. Admissible Decision Rules:

$D(\underline{x})$ is admissible if there is no $D'(x)$ such that

$$r_{D'}(\theta) \leq r_D(\theta) \quad \text{for all } \theta$$

$$r_{D'}(\theta) < r_D(\theta) \quad \text{for some } \theta$$

i.e., you can not improve on $D(\underline{x})$ uniformly.

CHAPTER VII

MISCELLANEOUS

A. Confidence Regions:

$\underline{X}$ is a random variable with d.f. $f(\underline{x}, \theta)$ $\theta \varepsilon \Omega$

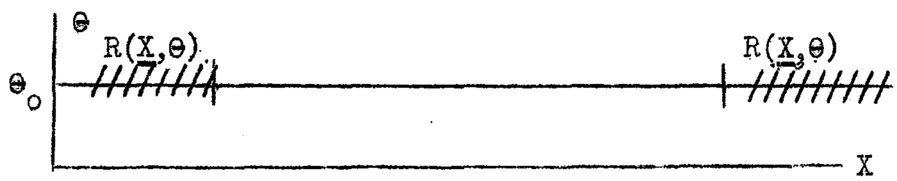
$B(\omega)$ = the totality of subsets of Ω

Let $A(\underline{X})$ be a function from X (sample space) to $B(\omega)$

Def. 47: If $Pr[A(\underline{X}) \supset \underline{\theta}] \geq 1 - \alpha$ then $A(\underline{X})$ is a confidence region with confidence coefficient $1 - \alpha$.

Theorem 40: If a non-randomized test of size α exists for H_o: $\theta = \theta_o$ for every θ_o then there exists a confidence region for θ of size $1 - \alpha$.

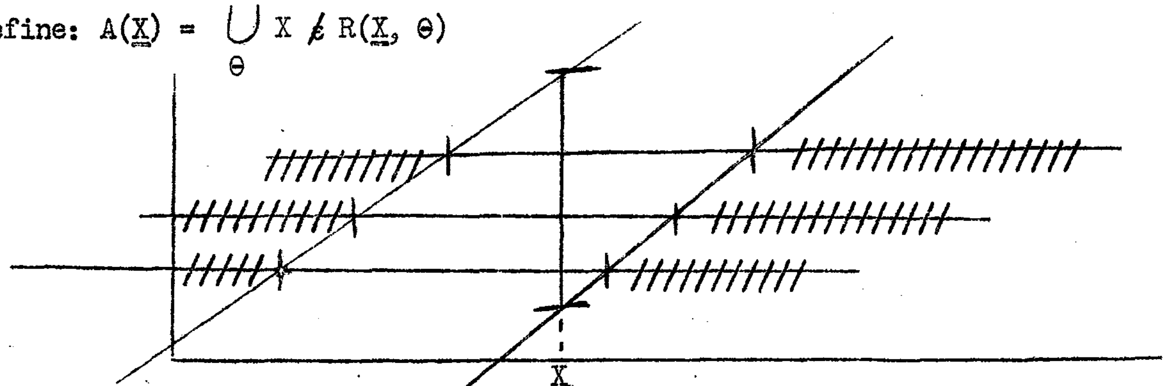
Proof: Let $R(\underline{X}, \theta_o)$ be the set of $\underline{X}$ for which H_o: $\theta = \theta_o$ is rejected, e.g.,



$$\int_{R(\underline{X}, \theta_o)} f(\underline{X}, \theta_o)\, d\underline{X} \leq \alpha \qquad [1]$$

This $R(\underline{X}, \theta_o)$ is defined and satisfies [1] for every θ_o.

Define: $A(\underline{X}) = \bigcup_{\theta} X \not\in R(\underline{X}, \theta)$



$Pr[A(\underline{X}) \supset \theta_o] = Pr[\underline{X} \not\in R(\theta)] = 1 - Pr[\underline{X} \varepsilon R(\theta)] \geq 1 - \alpha$ q.e.d.

Def. 48: A confidence region, $A(\underline{X})$, is uniformly most powerful if $Pr[A(\underline{X}) \supset \theta \mid \theta' \text{ is true}]$ is minimized for all $\theta' \neq \theta$.

note: Kendall uses this same definition with "uniformly most powerful" replaced with "uniformly most selective." Neyman replaces "u.m.p." with "shortest."

Def. 49: A confidence interval is "shortest" if the length of the interval is minimized uniformly in θ.

Confidence Interval for the ratio of two means:

X_i, Y_i are bivariate normal BIV $N(\mu, \alpha\mu, \sigma_x^2, \sigma_y^2, \rho)$ $i = 1, 2, \ldots, n$

where ρ = correlation coefficient and may = 0.

Problem: find a confidence interval for $\alpha = \frac{E(Y)}{E(X)}$

$Z_i = \alpha X_i - Y_i$ is $N(0, \alpha^2\sigma_x^2 - 2\alpha\rho\sigma_x\sigma_y + \sigma_y^2)$

$\sum(Z_i - \bar{Z})^2$ and $\bar{Z}$ are independent and distributed as X^2 and normal respectively.

$$\frac{\sum(Z_i-\bar{Z})^2}{n-1} = \frac{\sum(\alpha X_i - Y_i)^2 - n(\alpha\bar{X} - \bar{Y})^2}{n-1}$$

$$= \frac{\alpha^2\sum(X_i - \bar{X})^2 - 2\alpha\sum(X_i - \bar{X})(Y_i - \bar{Y}) + \sum(Y_i - \bar{Y})^2}{n-1}$$

$$= \alpha^2 s_x^2 - 2\alpha s_{xy} + s_y^2$$

$$\bar{Z} = \alpha\bar{X} - \bar{Y}$$

therefore:

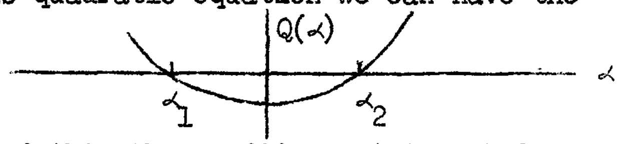
$$\frac{\sqrt{n}(\alpha\bar{X} - \bar{Y})}{(\alpha^2 s_x^2 - 2\alpha s_{xy} + s_y^2)^{1/2}}$$ has a t-distribution with n-1 d.f.

$$\Pr\left[\frac{\sqrt{n}\,|\alpha\bar{X} - \bar{Y}|}{(\alpha^2 s_x^2 - 2\alpha s_{xy} + s_y^2)^{1/2}} \leq t_{1-\frac{\varepsilon}{2}}(n-1)\right] = 1 - \varepsilon$$

We can solve this and get confidence limits for α. This yields the following quadratic equation in α:

$$Q(\alpha) = \alpha^2(n\bar{X}^2 - t^2 s_x^2) - 2\alpha(n\bar{X}\bar{Y} - t^2 s_{xy}) + (n\bar{Y}^2 - t^2 s_y^2) = 0$$

Among the various possible solutions of this quadratic equation we can have the following situations:



A. If $n\bar{X}^2 - t^2 s_x^2 > 0$:

It can be shown that 2 roots always exist and this the confidence interval is:

$$\alpha_1 < \alpha < \alpha_2$$

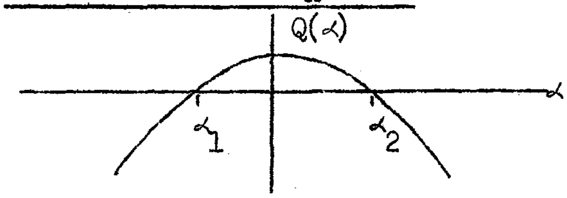
For the above inequality to hold, $\bar{X}$ must be significantly away from the origin;

i.e., we must reject $H_o: \mu = 0$ on the basis of $X_1, X_2, \ldots, X_n$ at the ε level.

e.g., the following condition must hold: $\frac{\sqrt{n}\,|\bar{X}|}{s_x} > t_{1 - \frac{\varepsilon}{2}}$

Note: One could, if desired, change t (and thus ε) to insure that this inequality always held and thus that a "real" confidence interval exists.

B. If $n\bar{X}^2 - t^2 s_x^2 < 0$



α_1, α_2 may be real or complex

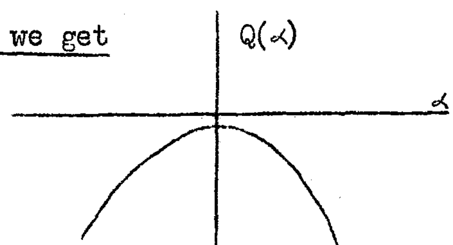
If the roots are real, then the "confidence interval" is $(-\infty, \alpha_1), (\alpha_2, +\infty)$,

i.e., we accept "in the tails."

C. If $4(n\bar{X}\bar{Y} - t^2 s_{xy})^2 < 4(n\bar{X}^2 - t^2 s_x^2)(n\bar{Y}^2 - t^2 s_y^2)$, i.e., $b^2 < 4ac$, then the roots are complex and we have a confidence interval $(-\infty, +\infty)$.

[Thus we can say that $-\infty < \alpha < +\infty$ with confidence $1 - \varepsilon$].

D. If we get



$$Q(\alpha) \leq 0 \iff \frac{\sqrt{n}\,|\alpha\bar{X} - \bar{Y}|}{(\alpha^2 s_x^2 - 2\alpha s_{xy} + s_y^2)^{1/2}} \leq t$$

thus we always accept H_o since the test statistic never exceeds t.

Refs: Fieller, Journal of the Royal Statistical Society, 1940
Paulson, Annals of Math Stat, 1942
Bennett, Sankhya, 1957
Symposium, Journal of the Royal Statistical Society, 1954.

LIST OF THEOREMS

LIST OF DEFINITIONS

LIST OF PROBLEMS