Sampling and Statistical Inference

Up to now, much of the course material has been concerned with the theory and methods of Probability (i.e. models for randomness)

We now move toward the study of some of the concepts and methods of STATISTICS.
We began the course with the concept of a random experiment and its sample space, and went on to consider random variables.
Suppose a random experiment were to be repeated n times, resulting in observations:

$$x_1, x_2, x_3 \dots, x_n$$

These observations can provide information on the underlying random experiment.

SAMPLING: The process of gathering data from several performances of a random experiment is called Sampling.

The individual items of data are termed observations

The collection of observations is termed a SAMPLE

Prior to getting the numerical values of the observations, each individual "OBSERVATION-TO-BE" has the potential for variability, and we regard each "OBSERVATION-TO-BE" as a random variable.

We shall be mainly concerned with observations on a SINGLE VARIABLE (I.E. UNIVARIATE STATISTICS.)

RANDOM SAMPLING AND STATISTICAL INFERENCE

Typical Sampling Processes that arise in practice are as follows

1) We have a finite collection of objects — a Population. Objects are drawn one at a time, all those in the pop. having equal chance of selection.
When an object is drawn, some characteristic assoc. with the object may be observed.
— object is then replaced — and pop mixed before next drawing.

2) Finite pop. again, but unlike 1), the object is not replaced after each drawing

3) The observations are obtained as the result of repeated indept performances of an experiment, where the conditions of the expt (that can be controlled) are kept const.

e.g. measuring $\begin{cases} g \\ \text{the speed of light} \end{cases}$

Testing a new alloy under stress conditions.
Observing the yield from an (agric) crop.

We have already used X to denote the random var. associated with the random expt.
~~—~~ Thus can regard a sample as a set of observations of the random variable X.

The prob. distrib of X is determined by the pattern of distrib of values in the population.
— We refer to the prob. distrib of X as the

POPULATION DISTRIB.

Before the observations in a sample are made, we have in mind to observe X, n times
— or alternately to observe $X_1, X_2, \ldots X_n$

r. Sample

When these n random vars are indept and identically distrib. (i.i.d.) the term RANDOM SAMPLE is used to refer to the n observations on the random var. X.

Thus if the distrib of X in the pop. is characterized by PDF $f(x)$, we have the following DEFN:

If the r.vars $X_1, X_2, \ldots, X_n$ have a joint PDF

$$f(x_1, x_2, \ldots, x_n) = \prod_1^n f(x_i)$$

where $f(\cdot)$ is the common PDF of X_i, then the r.vars $\{X_i\}$ are said to constitute a random sample of size n from the population characterized by $f(\cdot)$.

One of the central problems in Statistics is to determine the form of the density $f(\cdot)$ from the observations $x_1, x_2, \ldots, x_n$.

We shall see that in many instances, $f(\cdot)$ is a known function of certain unknown parameters θ and our task then becomes that of finding good estimates for θ.

e.g. If we could assume ~~that~~ the blood-pressures of a pop. of students were Normally Distrib (but μ, σ unknown) then we should attempt to estimate μ & σ which are the Mean & Std Devns characterizing the population.

STATISTICAL INF. is concerned with making statements concerning the population on the basis of SAMPLE OBSERVATIONS.

It is customary to subdivide the main study areas of Stat^l Inf. as follows:

I Estimation

How should we provide an estimate for a certain char. of a populn.

— e.g. mean

How do we choose among alternative estimates.

I2 Hypothesis Testing

How do we determine whether our observations are consistent with certain assumptions — or HYPOTHESES about the population.

— e.g the assumption that blood pressures are Normally distrib.

I3 Decision Theory

We have some process which we wish to control. Observations are taken and resulting from an analysis of these observations, an action follows to keep the process in control

e.g. Machine filling commodity into containers

The ~~containers~~ filled packages are examined periodically and action may follow.

NOTE: In this course, no time for in depth study of these areas — in remaining time shall conc. on the theory & application of the more common & useful statistical procedures.

Some terminology:

STATISTIC

A statistic is any known function of observable random variables — does not involve any unknown parameters.

e.g. random sample: $X_1, X_2, ---, X_n$ from distrib with PDF $f(x)$

$$\bar{X} = \frac{\Sigma X_i}{n} \quad ; \quad \frac{\sum^n (x_i - \bar{X})^2}{n-1}$$

$$\frac{X_1 + X_2}{2}$$

$$X_{(n)} - X_{(1)}$$

are all STATISTICS

But $\bar{X} - \mu$ or $\frac{X_{(n)} - X_{(1)}}{\sigma}$

are not statistics unless μ, σ are known.

SAMPLE MOMENTS

Common examples of statistics are the sample moments about origin

$$m_1' = \frac{\sum x_i}{n} = \bar{x}$$

$$m_2' = \frac{\sum x_i^2}{n}$$

$$m_r' = \frac{\sum x_i^r}{n}$$

Corresponding sample moments about sample mean are

$$m_1 = 0 = \frac{\sum^n (x_i - \bar{x})}{n}$$

$$m_2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$\vdots \qquad \vdots$$

$$m_r = \frac{1}{n} \sum^n (x_i - \bar{x})^r$$

It is useful to find the Expected value of these Sample Moments — these expected values can be found providing the corresponding population moments exist.

e.g. $$E(m_1') = E(\bar{x})$$

$$= E\left[\frac{\sum^n x_i}{n}\right] = \frac{1}{n} \sum E X_i$$

$$= \frac{1}{n} n\mu = \mu$$

$$E(m_2') = E\left[\frac{\sum^n x_i^2}{n}\right] = \frac{1}{n} \sum^n E X_i^2$$

$$= \mu_2'$$

$$E(m_r') = \mu_r'$$

The sample moments about the origin are said to be <u>UNBIASED</u>

Sample Moments about Sample mean

$$E(m_2) = E\left[\frac{1}{n}\sum^n (x_i - \bar{x})^2\right]$$
$$= \frac{1}{n}E\left[\sum^n x_i^2 - n(\bar{x})^2\right]$$
$$= \frac{1}{n}\left\{\sum^n E X_i^2 - nE(\bar{X})^2\right\}$$
$$= \frac{1}{n}\left\{n(\sigma^2+\mu^2) - nE(\bar{X})^2\right\}$$

*

$$E(\bar{X})^2 = \frac{1}{n^2}E\left(\sum^n X_i\right)^2$$
$$= \frac{1}{n^2}E\left[\sum^n X_i^2 + 2\sum_{i>j}\sum X_i X_j\right]$$
$$= \frac{1}{n^2}\left\{\sum^n E X_i^2 + 2\sum_{i>j}\sum (EX_i)(EX_j)\right]$$
$$= \frac{1}{n^2}\left\{n(\sigma^2+\mu^2) + n(n-1)\mu^2\right\}$$
$$= \frac{1}{n}\left\{\sigma^2 + \cancel{\mu^2} + n\mu^2 - \cancel{\mu^2}\right\}$$
$$= \mu^2 + \frac{\sigma^2}{n}$$

$$E(m_2) = \frac{1}{n}\left\{n(\sigma^2+\mu^2) - n\mu^2 - \sigma^2\right\}$$
$$= \frac{n-1}{n}\sigma^2$$
$$= \frac{n-1}{n}\mu_2$$

ALT. TO *: $E(\bar{X})^2 = Var(\bar{X})^2 + (\mu)^2$
$$= \frac{\sigma^2}{n} + \mu^2$$

earlier work
$Var(\sum a_i X_i)$
$= \sum a_i^2 \, Var X_i + Cov$

ESTIMATOR

An estimator for a {distib / pop. parameter} is a statistic that may be used to provide an estimate for that parameter.

e.g. it seems sensible (intuitive) to use

$\bar{X}$ as an estimate of μ
and m_2 as " " " μ_2

But $E(m_2) = \frac{n-1}{n}\mu_2$
and because of this m_2 is said to be a
BIASED estimator of μ_2. (But $\bar{X}$ is unbiased estimator of μ)

UNBIASED ESTIMATOR OF μ_2
would be $\dfrac{\sum^n (x_i - \bar{x})^2}{n-1}$

— shall refer to this as the SAMPLE VARIANCE
and denote it by s^2

—— || ——

It is important that you understand the difference between
the moments of the prob distib : The Pop. Moments
& " " " " sample obs : The Sample "

Pop Moments	Sample Moments
$\mu_1' = \mu$	$m_1' = \bar{x}$
$\mu_2 = \sigma^2$	$m_2 = \frac{1}{n}\sum^n (x_i - \bar{x})^2$
$\vdots$	$\vdots$
μ_r	$m_r = \frac{1}{n}\sum_1^n (x_i - \bar{x})^r$

SAMPLING FROM A NORMAL DISTRIB

The r.v. $X \sim N(\mu, \sigma^2)$

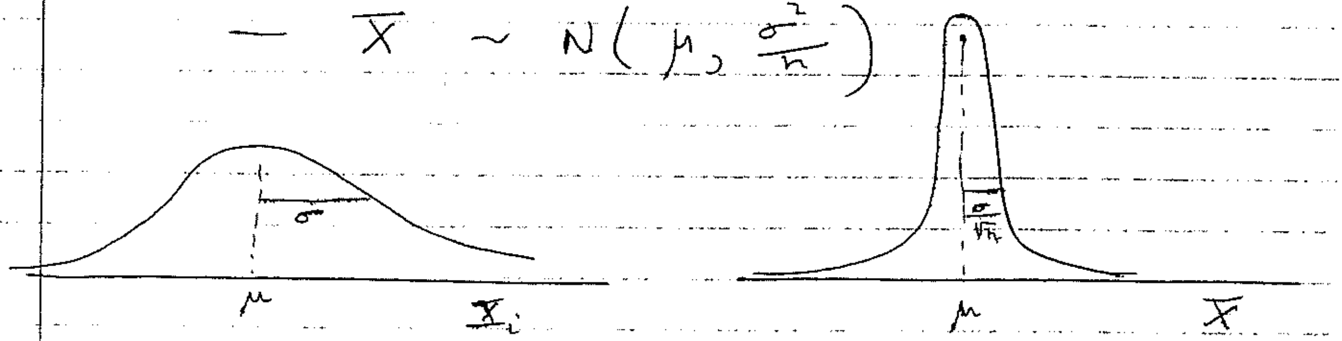
i.e. $E(X) = \mu$
$V(X) = \sigma^2$

Suppose we have a random sample of n obs from this distrib

Intuitive estimators for μ and σ^2
are $\bar{X}$ for μ
and S^2 for σ^2

Now $\bar{X} = \frac{\sum^n X_i}{n}$ is a STATISTIC
and thus is a random variable.
— Thus it has a probability distrib. In this context of sampling, the prob. distrib of $\bar{X}$ is termed
the SAMPLING DISTRIB OF $\bar{X}$
or " " " OF THE STATISTIC $\bar{X}$

What is the sampling distrib of $\bar{X}$?
— $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$



INTERVAL ESTIMATE FOR μ

Thus $P\left[|\bar{X} - \mu| > 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.05$

or $P\left[|\bar{X} - \mu| \leqslant 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.95$

We can write this as

$P\left[-1.96 \frac{\sigma}{\sqrt{n}} \leqslant \bar{X} - \mu \leqslant 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.95$

i.e. $P\left[-\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leqslant -\mu \leqslant -\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.95$

i.e. $P\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right] = 0.95$

i.e. The prob that μ lies in this <u>RANDOM INTERVAL</u> is 95%

— the ends of the interval are random vars.

Once the observations are made, $\overline{X}$ assumes a numerical value (lets assume σ known)

$\Rightarrow$ some interval such as $100.9 \rightarrow 110.1$

But the predominant school of ~~thought~~ in STATS would avoid the statement that μ lies in this FIXED INTERVAL with prob 0.95

<u>WHY</u>?

because μ is either in the fixed interval or not.

To avoid any such conclusion being drawn, the following terminology is used:

CONFIDENCE INTERVAL

" COEFFICENT — rather than <u>prob</u>.

— || —

Thus if <u>σ is known</u>, the expression for the 95% conf. interval for μ in the case of Sampling from a Normal Prob Distrib is

$$\overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

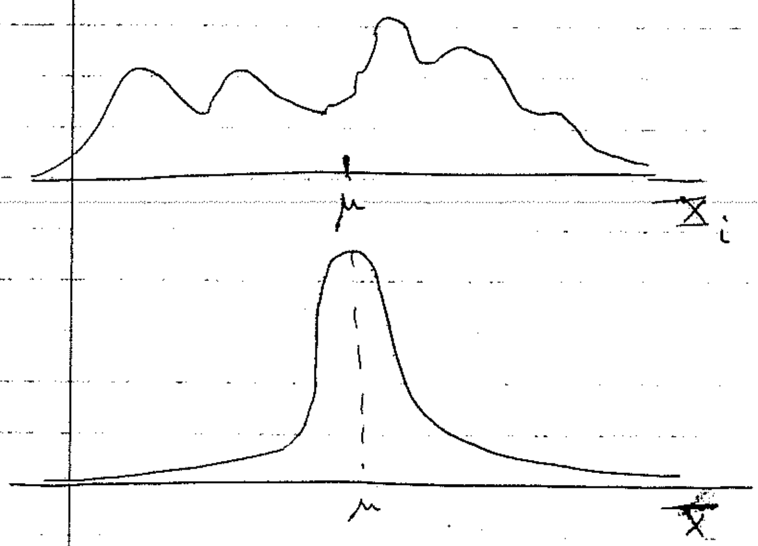
SAMPLING FROM A GENERAL PROB DISTRIB

Suppose the mean of the prob. distrib is μ
var " " " is σ^2

Random sample of size n selected
Suppose n is quite large

By the C. L Theorem, the Sampling Distrib of $\overline{X}$ will be approx Normal with mean μ
and " var $\frac{\sigma^2}{n}$

I.E $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ for large n



Thus taking σ to be known, the approx 95% Conf Int is given by

$$\overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

(SAMPLING FROM A NORMAL DISTRIB.)

INDEPENDENCE OF $\overline{X}$ AND S^2

$\longrightarrow$ SAMPLING DISTRIBUTION OF S^2

Can we find a conf. interval for σ^2 just as we did for μ?

In the case of μ, we used our knowledge of the Sampling Distrib of $\overline{X}$, the estimator for μ.

Thus we should try to find the sampling distrib of S^2

Let $X_1, X_2 \; --- \; X_n$ be a random sample from $N(\mu, \sigma^2)$

We introduce new random vars $Y_1, Y_2, ---, Y_n$ by means of a linear transformation of $X_1, X_2 --- X_n$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & ---- & & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & --- & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 \; --- & 0 \\ & & & & 0 \\ ---- & ---- & ---- & ---- & \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & ---- & ---- & -\frac{(n-1)}{\sqrt{n(n-1)}} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{bmatrix}$$

I.E. $\underline{Y} = \underline{P}\,\underline{X}$

$$\left.\begin{array}{l} \text{The Matrix is an ORTHOGONAL MATRIX} \\ \text{I.E.} \quad PP^T \equiv P^T P = I \\ \qquad \operatorname{Det} P = 1 \end{array}\right\}$$

Notice that $E(Y_1) = \frac{1}{\sqrt{n}} \sum E X_i = \sqrt{n}\, E X_i = \sqrt{n}\,\mu$

$$\left.\begin{array}{ll} E(Y_i) = 0 & \text{for } \forall i > 1 \\ V(Y_i) = \sigma^2 & \forall i \\ \operatorname{Cov}(Y_i, Y_j) = 0 & \forall i \neq j \end{array}\right\} \underline{\text{CHECK}}$$

A NOTE ON ORTHOGONAL MATRICES

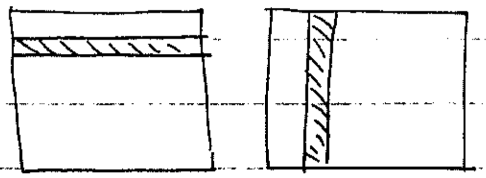
DEFN A, an $n \times n$ matrix is ORTHOGONAL if

$$A^{-1} = A'$$

$$\Rightarrow \quad AA' = A'A = I$$

$\Rightarrow$ A orthogonal iff sum of squares of elts of each row $= 1$

and the sum of products of corresponding elts in different rows $= 0$



Some Properties of Orthogonal A

1) $|\det A| = 1$

PROOF: $\det A = \det A'$ for any A

Now $\det AA' = (\det A)(\det A') = (\det A)^2$

$1 \leftarrow \det(I) = (\det A)^2$

$\Rightarrow \quad |\det A| = 1$

2) Suppose $Y = AX$ where X is $(n \times 1)$ vector

then $\sum X_i^2 = \sum Y_i^2$

Proof: $\sum Y_i^2 = Y'Y = (AX)'AX$

$= X'A'AX$

$= X'IX$

$= \sum X_i^2$

$Y = AX$ is termed an orthogonal linear transformation of X

If X is a random vector, the Jacobian of the inverse transfn

$X = A'Y$

is $\dfrac{\partial(x_1\, x_2 - - x_n)}{\partial(y_1\, y_2 - - - y_n)} = \begin{bmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_2}{\partial y_1} & - - - & \dfrac{\partial x_n}{\partial y_1} \\ & & & \\ & & & \end{bmatrix} = A'$

$|J| = (\text{Det } A') = 1$

The joint PDF of $Y_1, Y_2 \cdots Y_n$ is

$$g(y_1, y_2 \cdots, y_n) = f(x_1, x_2 \cdots, x_n)\,|J|$$

(extension of the earlier result that we quoted — where the x_i's must be replaced in terms of y_i's)

Now $$J = \frac{\partial(x_1, x_2 \cdots x_n)}{\partial(y_1, y_2 \cdots y_n)}$$

— the Jacobian of $\underline{X}$ wrt $\underline{Y}$

Since P is orthogonal

$$\underline{Y} = \underline{P}\,\underline{X} \implies \underline{P}^T\underline{Y} = \underline{P}^T\underline{P}\,\underline{X} = \underline{X}$$

Thus $$J = \text{Det } P^T$$
$$= \text{Det } P = 1$$

Thus $$g(y_1, y_2 \cdots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu)^2\right]} \qquad 1$$

where we must replace x_i's in terms of y_i's.

Now $$\sum_1^n (x_i-\mu)^2 = (\underline{x}-\underline{\mu})^T(\underline{x}-\underline{\mu})$$

where $$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \underline{\mu} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix}$$

$$= (\underline{x}-\underline{\mu})^T P^T P (\underline{x}-\underline{\mu})$$
$$= \left[\underline{P}(\underline{x}-\underline{\mu})\right]^T \underline{P}(\underline{x}-\underline{\mu})$$

Now $$\underline{P}(\underline{x}-\underline{\mu}) = \underline{P}\,\underline{x} - \underline{P}\,\underline{\mu}$$
$$= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \sqrt{n}\mu \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus $$\underbrace{\sum_1^n (x_i-\mu)^2}_{\wedge} = (y_1-\sqrt{n}\mu)^2 + \sum_2^n y_i^2 \implies \begin{cases} Y_1 \sim N(\sqrt{n}\mu, \sigma) \\ Y_i \sim N(0, \sigma);\ i>1 \end{cases}$$

NOTE $Y_1, Y_2 \cdots Y_n$ are all indept

Thus
$$g(y_1, y_2, \dots y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\left[(y_1 - \sqrt{n}\mu)^2 + \sum_{i=2}^{n} y_i^2\right]}$$
$$= \frac{1}{\sqrt{2\pi}\,\frac{\sigma}{\sqrt{n}}} e^{-\frac{1}{2\left(\frac{\sigma}{\sqrt{n}}\right)^2}\left[\overline{X}-\mu\right]^2} \frac{1}{(\sqrt{2\pi})^{n-1}\sigma^{n-1}} e^{-\frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2}$$

Now
$$\sum_{i=1}^{n} x_i^2 = \underline{x}^T \underline{x} = \underline{x}^T \underline{P}^T \underline{P}\,\underline{x}$$
$$= (\underline{P}\,\underline{x})^T \underline{P}\,\underline{x}$$
$$= \underline{y}^T \underline{y}$$
$$= \sum_{i=1}^{n} y_i^2$$

Thus $\displaystyle\sum_{i=1}^{n} x_i^2 - y_1^2 = \sum_{2}^{n} y_i^2$

and using $y_1 = \sqrt{n}\,\overline{X}$, we get
$$\sum_{i=1}^{n} x_i^2 - n(\overline{X})^2 = \sum_{2}^{n} y_i^2$$
i.e. $\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{2}^{n} y_i^2$

Thus $\displaystyle\sum_{i=1}^{n} \frac{(x_i - \overline{X})^2}{\sigma^2} = \sum_{2}^{n} \left(\frac{y_i}{\sigma}\right)^2$

But $y_i/\sigma \sim N(0,1)$ and the y_i are all indept

So that $\displaystyle\sum_{i=2}^{n} \left(\frac{y_i}{\sigma}\right)^2 \sim \chi^2_{n-1}$

Thus $\displaystyle\sum_{i=1}^{n} \frac{(x_i - \overline{X})^2}{\sigma^2} \equiv \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Furthermore, because all the y_i's are indept, this quantity is indept of $\overline{X}$ $\left(= \frac{y_1}{\sqrt{n}}\right)$

(Usually expressed as $\overline{X}$ and S^2 are <u>indept</u>.)

EXPECTED VALUE OF SAMPLE STANDARD DEVIATION WHEN SAMPLING FROM A NORMAL DISTRIBUTION.

FIRST, SUPPOSE X IS χ^2_n ; WE WILL EVALUATE $E[\sqrt{X}]$

$$E[\sqrt{X}] = \int_0^\infty x^{\frac{1}{2}} \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \frac{1}{2}\, dx$$

$$= \sqrt{2}\, \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{\Gamma(\frac{n+1}{2})} \left(\frac{x}{2}\right)^{\frac{n+1}{2}-1} e^{-\frac{x}{2}} \frac{1}{2}\, dx$$

$$= \sqrt{2}\, \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$$

THIS IS RELEVANT TO THE EVALUATION OF $E(S)$, BECAUSE $\frac{(n-1)S^2}{\sigma^2}$ IS χ^2_{n-1}

LET'S DENOTE THIS BY X

THUS $E[\sqrt{X}] = E\left[\sqrt{\frac{n-1}{\sigma^2}}\, S\right] = \sqrt{2}\, \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$

AND SO, $E[S] = \sigma \sqrt{\frac{2}{n-1}}\, \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$

FOR LARGE VALUES OF n, $E(S)$ IS VERY CLOSE TO σ, BUT EVEN FOR SMALL VALUES OF n, IT'S QUITE CLOSE.

e.g. $n = 4$:

$$E(S) = \sigma \sqrt{\frac{2}{3}}\, \frac{\Gamma(2)}{\Gamma(\frac{3}{2})}$$

$$= \sigma \sqrt{\frac{2}{3}}\, \frac{1}{(\frac{1}{2})\sqrt{\pi}} = 0.9213\,\sigma$$

FOR $n = 25$ $\quad E(S) = 0.99\,\sigma$

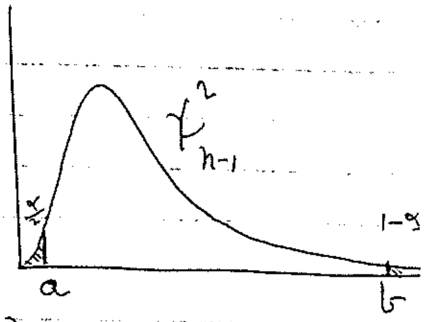
RETURN TO FIND CONF. INT for σ^2

We now know that $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

Choose some conf. level e.g. 95% $= (1-\alpha)100\%$

Suppose a denotes $\frac{\alpha}{2}$ pt for χ^2_{n-1}

b " $(1-\frac{\alpha}{2})$ " "



$$P\left[a \leq \frac{(n-1)s^2}{\sigma^2} \leq b\right] = (1-\alpha)$$

Thus $P\left[\frac{1}{a} \geq \frac{\sigma^2}{(n-1)s^2} \geq \frac{1}{b}\right] = (1-\alpha)$

or $P\left[\frac{(n-1)s^2}{b} \leq \sigma^2 \leq \frac{(n-1)s^2}{a}\right] = 1-\alpha$

$\chi^2_{(1-\frac{\alpha}{2}, n-1)}$ $\chi^2_{(\frac{\alpha}{2}, n-1)}$

The interval $\left(\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}, \dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}}\right)$

is termed the $100(1-\alpha)\%$ conf. int for σ^2

e.g. (Numerical LATER (P.17))