01.03.2018.

Limits of Dist^n. f^n's.

It can be useful to approximate one prob. dist^n. by another, which is more convenient to handle.

- e.g. computing Binomial probs. for large n.

Can we use another dist^n. to appro. this?

Thus we are interested in the prob. dist^n. of a R.V. X_n as n increases.

Note: we express the approximation of prob. dist^n. of X_n in terms of dist^n. f^n., rather than pdf.

Thus let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of R.Vs with corresponding dist^n. f^n's.

$$F_1(X), \ F_2(X), \ \ldots \ F_n(X).$$

Now the limit $\lim_{n\to\infty} F_n(X)$ may or may not exist.

Even it does, $\lim_{n\to\infty} F_n(X)$ may not be a dist^n. f^n.

If the sequence of D.F's does converge to a f^n. $F(X)$ at each pt. of continuity of $F(X)$, where $F(X)$ is a dist^n. f^n., then we say that the corresponding sequence of Random var. $X_1, \ldots, X_n \ldots$

converges in dist^n. or in Law to the R.V. X whose dist^n. f^n. is $F(X)$.

and we write

$$\lim X_n = X$$

or $$X_n \xrightarrow{L} X$$

Note: There are other types of convergence that may be defined for R.Vs — but these would be appropriate in a more advanced course.

why we demand convergence of $F_n(x)$ to $F(x)$ only at p.t.s of continuity of $F(x)$:

– The following example will clarify this:

Let $X_n \sim N(0, \sigma_n)$, $\sigma_n \to 0$.

$$F_n(x) = P[X_n \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-\frac{x^2}{2\sigma_n^2}} dx$$

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x<0 \\ \frac{1}{2} & x=0 \\ 1 & x>0 \end{cases} \quad \text{as } \sigma_n \to 0.$$

Thus $F(x)$ is ~~dis-cont.~~ dis-cts at $x=0$.

At p.t.s of cts of $F(x)$, we have $F(x)=0$ $x<0$
$F(x)=1$ $x>0$

And we use the right ~~ct~~ continuity of $F(x)$ to say that we must have $F(0)=1$

i.e. $F(x)=0$ for $x<0$
$F(x)=1$ for $x\geq 0$.

Thus $F_n(x) \to F(x)$ only at p.t.s of ~~cts~~ continuity of $F(x)$. and the fact $F_n(x) \to \frac{1}{2}$ at $x=0$ does not matter.

Note: If $X_n \xrightarrow{L} X$

this does not imply that the pdf of X_n converges to the pdf of X.

E.g. above example, $f_n(X) \to 0$ for $X \neq 0$
$\to \infty$ for $X = 0$.

How do we find the limiting forms of distⁿ. fⁿ's ?

One way is to find the D.F's themselves and attempting to find the limiting form. Can be very difficult.

or through the use of chara. ~~Fn~~ fⁿ. relying on:

Continuity Theorem:
If the sequence of distⁿ. fⁿ. $F_n(X)$ converges to a distⁿ. fⁿ. $F(x)$, the corresponding sequence of char. fⁿ's

$\phi_n(t)$ converges to the chara. fⁿ. $\phi(t)$ of $F(x)$.

Conversely: If a sequence of chara. fⁿ's $\phi_n(t) \to \phi(t)$

which is cts at $t=0$, then $\phi(t)$ is a chara. fⁿs and the corresponding sequence of distⁿ. fⁿ's converges to the distⁿ fⁿ. determined by $\phi(t)$.

Application of the Theorem

We now illustrate the use of this theorem to show that if X_n is $Bin(n,p)$, then the limiting dist^n. of

$$\frac{X_n - np}{\sqrt{npq}} \quad \text{is} \quad N(0,1).$$

Proof: The chara. f^n. of X_n is $(pe^{it}+q)^n$

Thus, the chara. f^n. of $Z = \frac{X_n - np}{\sqrt{npq}}$ is

$$\begin{aligned}
\varphi_Z(t) &= E[e^{itZ}] \\
&= E[e^{it\left(\frac{X_n - np}{\sqrt{npq}}\right)}] \\
&= E[e^{-it\cdot\frac{np}{\sqrt{npq}}} \cdot e^{it\cdot\frac{X_n}{\sqrt{npq}}}] \\
&= e^{-it\frac{np}{\sqrt{npq}}}\left(p\cdot e^{\frac{it}{\sqrt{npq}}} + q\right)^n \\
&= \left(p\cdot e^{\frac{it(1-p)}{\sqrt{npq}}} + q\, e^{-\frac{itp}{\sqrt{npq}}}\right)^n \\
&= \left(p\cdot e^{\frac{itq}{\sqrt{npq}}} + q e^{-\frac{itp}{\sqrt{npq}}}\right)^n
\end{aligned}$$

Taylor expansion at zero.

$$\begin{aligned}
&= \Big[p\Big\{1 + \frac{itq}{\sqrt{npq}} - \frac{t^2q^2}{2npq} + o\left(\frac{t^2}{n}\right)\Big\} \\
&\quad + q\Big\{1 - \frac{itp}{\sqrt{npq}} - \frac{t^2p^2}{2npq} + o\left(\frac{t^2}{n}\right)\Big\}\Big]^n. \\
&= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n.
\end{aligned}$$

As $n \to \infty$, this char. f^{n}. $\to e^{-\frac{t^2}{2}}$, which is a chara. f^{n}. for $N(0,1)$.

Recall: $\lim_{n\to\infty} (1+\frac{1}{n})^n = e$.

$\lim_{n\to\infty} [1 - \frac{t^2}{2n} + o(\frac{t^2}{n})]^n$, the third item $\to 0$ as $n \to \infty$.

$$= \lim_{n\to\infty} \left(1 - \frac{t^2}{2n}\right)^n = \lim_{n\to\infty} \left(1 + -\frac{t^2}{2n}\right)^{n \times -\frac{2n}{t^2} \times \frac{-t^2}{2n}}$$

$$= \lim_{n\to\infty} \left(1 + -\frac{t^2}{2n}\right)^{-\frac{2n}{t^2} \times \frac{-t^2}{2}} = e^{-\frac{t^2}{2}}$$

Thus, using the continuity Theorem, the limiting prob. dist^{n}. of $Z \sim N(0,1)$, and the limiting prob. dist^{n}. of X_n is $N(np, npq)$.

Therefore, $n \to \infty$ X_n is $\doteq N(np, npq)$.

Notice that the R.V. in the above may be regarded as the sum of n indep. R.Vs.:

$$X_n = \sum_{i=1}^{n} Y_i$$

where each Y_i is a Bernoulli R.Vs, with mean p, variance $p(1-p)$.

Thus, we have shown that $\sum_{i=1}^{n} Y_i$ for large n is $\doteq N(np, npq)$.

This result is a special case of the most important theorem — Central Limit Theorem.

Examples on Normal approximation to Binomial.

Let $n = 100$ $\quad p = \frac{1}{2}$.

Now $E(X) = np = 50$

$\quad Var(X) = npq = 25$.

$$P[X \leq 46] = \sum_{k=0}^{46} \binom{100}{k} \left(\tfrac{1}{2}\right)^{100}$$

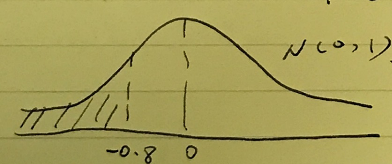
$\hookrightarrow$ mission impossible.

Use $X \sim N(np, npq)$.

Thus $\dfrac{X - np}{\sqrt{npq}} \sim N(0, 1)$

Now $P[X \leq 46] = P\left[\dfrac{X - np}{\sqrt{npq}} \leq \dfrac{46 - 50}{5}\right]$

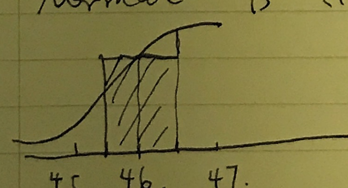
$$= P[Z \leq -0.8] = 0.2119$$



Exact answer from Tables of Binomial $= 0.2421$

The approx. does not seem very good. Can improve it by using <u>A continuity Correction</u>:

Remember that the Bin. Distn. is discrete, while normal is cts.



Regard ~~ing~~ prob. associated with 46 as running from 45.5 to 46.5.

Thus

$$P[X_B \le 46] = P[X_N \le 46.5]$$

$$= P[Z \le -0.7] = 0.2420.$$

Extremely close.

<u>2^{nd} Example:</u> Much smaller $n=10$ $p=\frac{1}{2}$.

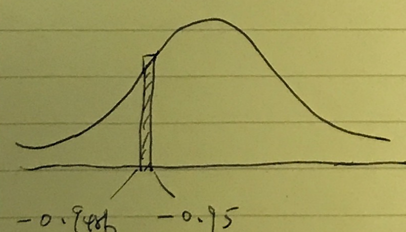
$$P[X \le 3] = \sum_{0}^{3} \binom{10}{k} \left(\frac{1}{2}\right)^{10} = 0.1719.$$

Normal Approx.

$$P[X_B \le 3] \doteq P[X_N \le 3.5] = P[Z \le -0.9486]$$

check table $P[Z \le -0.95] = 0.1711.$

≈ 0.1715 by interpoln.



regard area $\approx (0.95 - 0.9486)\overset{f(-0.95)}{\cancel{f(0.95)}}$

Suppose $p = 0.1$, Would the approx. for $n=10$ be as good?

Central Limit Theorem

We state the theorem in its simplest form

Let $X_1, \ldots, X_n$ be a sequence of indep. and identically distributed R.Vs. such that

$E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, both exist & finite.
Then the limiting distⁿ of

$$\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right) \text{ is } N(0,1) \text{ as } n \to \infty.$$

where $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

proof: Let $\phi(t)$ be the chara. fⁿ. $Z_i = \frac{X_i - \mu}{\sigma}$

$\Rightarrow E(Z_i) = 0,$
$Var(Z_i) = 1 = \frac{E[(X_i - \mu)^2]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1.$

$\because$ 1st & 2nd moments of Z_i exist, ~~then~~
$\therefore$ 1st & 2nd derivatives of $\phi(t)$ exist and at $t=0$, we have

$$\phi'(t)\big|_{t=0} = 0; \quad \phi''(t)\big|_{t=0} = -1$$

We ~~can expand~~ can obtain taylor expansion around $t=0$ up to (t^2).

i.e. $\phi(t) = 1 - \frac{t^2}{2} + o(t^2)$.

Now we consider

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma n}$$

$$= \frac{1}{\sqrt{n}} \sum \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum Z_i = \sum \frac{Z_i}{\sqrt{n}}$$

Since the X_i are indep, so also are the Z_i and the $\frac{Z_i}{\sqrt{n}}$.

Chara. f^{n}. of $\frac{Z_i}{\sqrt{n}} = \phi\left(\frac{t}{\sqrt{n}}\right)$

$\Rightarrow$ Chara. f^{n}. of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n$.

$$= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \quad \cancel{\text{as } n \to \infty},$$

as $n \to \infty$, the above $\to \exp\left(-\frac{t^2}{2}\right)$

Thus by the Levy ~~et~~ continuity Theorem, we conclude that the dist^{n}. f^{n}. of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ tends to the Normal Dist^{n}. f^{n}. $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ as $n \to \infty$.

<u>Note:</u> We should stress that in any event, the R.V. $\bar{X}_n = \frac{1}{n} \sum X_i$ will have mean μ (whether or not indep) and variance σ^2/n, due to indep.

For large n, the CLT tells us something about the shape of the dist^{n}. of $\bar{X}_n$, — namely that it is approx normal in shape.

Prob Dist^{n} of Functions of a R.V.

Here we are concerned with the investigation of the prob. dist^{n} of some f^{n} of a R.V. X.

Assume X's prob. dist^{n} is known and that the f^{n} $y = u(x)$ has been specified. The dist^{n} of ~~X~~ $Y = u(X)$ is of interest.

If X is a discrete R.V. then Y is also discrete, ~~(irrespond~~ regardless the $u(X)$.

$$X: \quad x_1, \quad x_2, \quad -- \quad x_k, \quad ---$$
$$\text{Probs}: \quad f(x_1), \quad f(x_2), \quad -- \quad f(x_k), \quad ---$$

$$\begin{array}{lllll} & u_1 & u_2 & -- & u_k, \; --- \\ \text{Then } Y \text{ takes values} & u(x_1), & u(x_2), & -- & u(x_k) \; --- \\ \quad \text{with prob}: & f(x_1), & f(x_2), & -- & f(x_k) \; --- \end{array}$$

Now it may happen that several ~~value~~ x_i give rise to same value of y.

Thus in general the prob. f^{n} of Y is given by

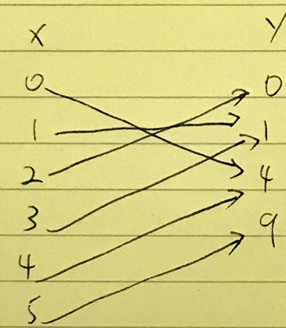
$$g(y_i) = \sum_{\substack{x_j \text{ for} \\ \text{which} \\ u(x_j) = y_i}} P[X = x_j]$$

Example:

X:	0	1	2	3	4	5
Probs:	p_0	p_1	p_2	p_3	p_4	p_5

Let the f^{n} be $u(X) = (X-2)^2$, determine prob. f^{n} of $Y = u(X)$

Possible values:	0	1	4	9
probs:	p_2	p_1+p_3	p_0+p_4	p_5



If X is cts, then $Y = \cancel{X} u(X)$ may be discrete or cts, depends on $u(X)$.

- If X is cts, $Y = u(X)$ is discrete.

<u>eg.</u> $u(X) =$ integer part of X.

or $u(X) = 1$ for $X > 0$

$= 0$ for $X \le 0$.

there it is again a simple matter to find the prob. distⁿ. of $Y = u(X)$.

Suppose Y may take on values $y_1, y_2, \cdots y_k$, then

$$P[Y = y_i] = \int_{\substack{x \\ \text{s.t.} \\ u(x) = y_i}} f(x)\,dx.$$

pdf of X.

- Both X & $Y = u(X)$ are cts.

Here there are ~~the~~ 3 techniques that may be used in finding the PDF of Y:

1) Cumulative distⁿ. fⁿ.
2) MGF
3) Transformation Technique.

We will present examples to illustrate the application of each approach:

Suppose $X \sim N(0,1)$ and $u(X) = x^2$

1) Now use CDF approach.

$$G(y) = P[Y \le y] = P[X^2 \le y]$$
$$= P[-\sqrt{y} \le X \le \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$
$$\cancel{= F(\sqrt{y}}$$
$$= 2F(\sqrt{y}) - 1, \text{ since symmetry}$$

Differentiate w.r.t. y.

$$g(y) = 2\frac{d}{dy}F(\sqrt{y}) = 2f(\sqrt{y}) \cdot \frac{1}{2} y^{-\frac{1}{2}}$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}}$$
$$= \frac{1}{\sqrt{\pi}} \frac{y^{-\frac{1}{2}}}{2^{\frac{1}{2}}} e^{-\frac{y}{2}}$$

Can check $g(y)$ is a pdf by

07.03.2018. $\int_0^\infty \frac{1}{\sqrt{\pi}} \cdot \frac{y^{-\frac{1}{2}}}{2^{\frac{1}{2}}} e^{-\frac{y}{2}} dy$

$$= \frac{1}{\sqrt{\pi}} \int_0^\infty \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \left(\frac{1}{2}\right)^{-\frac{1}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} dy$$
$$= \frac{1}{2\sqrt{\pi}} \int_0^\infty \left(\frac{y}{2}\right)^{-\frac{1}{2}} e^{-\frac{y}{2}} dy, \quad \text{let } \frac{y}{2} = t, \ dy = 2dt.$$
$$= \frac{1}{2\sqrt{\pi}} \int_0^\infty t^{-\frac{1}{2}} e^{-t} 2\,dt = \frac{1}{\sqrt{\pi}} \int_0^\infty t^{\frac{1}{2}-1} e^{-t} dt$$
$$= \frac{1}{\sqrt{\pi}} \cdot \Gamma\left(\frac{1}{2}\right), \quad \text{recall } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$
$$= 1$$

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

Recall: Gamma pdf:

$$f(x) = \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{\beta^{\alpha}} e^{-\frac{x}{\beta}}$$

Thus above, we actually have a pdf of Gamma dist^{n} with $\alpha = \frac{1}{2}$, $\beta = 2$.

This particular form of the Gamma dist^{n} is also known as

CHi square dist^{n} with 1 degree of freedom.

χ^2_1

Note: General form of PDF for χ^2 is ~~formed~~ found from Gamma by taking

$\beta = 2$ and $\alpha = \frac{n}{2}$

Thus $f(x) = \frac{1}{\Gamma(\frac{n}{2})} \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}} e^{-\frac{x}{2}}$ for $x>0$,

parameter of this dist^{n} family is n, termed the # of degree of freedom. shall see the reason later.

2) Using MGF: again $X \sim N(0,1)$, $Y = u(X) = X^2$

$$E(e^{sY}) = E[e^{sX^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx^2} e^{-\frac{1}{2}x^2} dx$$

$$= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2s)x^2} dx, \quad \text{let } t = \sqrt{1-2s}\, x$$

$$dt = \sqrt{1-2s}\, dx$$

$$= \underbrace{\int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt}_{=1} \times \frac{1}{\sqrt{1-2s}} = \frac{1}{\sqrt{1-2s}}$$

MGF for Gamma is $= \frac{1}{(1-\beta s)^{\alpha}}$

$\Rightarrow$ Y has a gamma distⁿ with $\alpha = \frac{1}{2}$, $\beta = 2$, as before.

Note: At this point it is opportunity ~~it~~ to mention that if $Z = \sum_{i=1}^{n} X_i^2$ where $X_i \sim N(0,1)$, and X_i are indep.

then $MGF(Z) = \prod_{i=1}^{n} MGF(X_i^2)$

$$= \prod_{i=1}^{n} \frac{1}{(1-2s)^{\frac{1}{2}}} = \frac{1}{(1-2s)^{\frac{n}{2}}},$$

which is a Gamma with $\alpha = \frac{n}{2}$, $\beta = 2$., i.e. the χ_n^2 distⁿ with n d.o.f.

Note: Mean and Variance for χ^2.

We will quote these results here, since we have derived twice. for Γ distⁿ.

If $U \sim \chi_n^2$, then $E(U) = n$, ~~Var~~
$Var(U) = 2n$.

since $E(\Gamma) = \alpha\beta$
$Var(\Gamma) = \alpha\beta^2$

We ~~say~~ saw that $X_i \sim N(0,1)$, then $U = \sum_{i=1}^{n} X_i$ (for X_i indep) has a χ^2 distⁿ with n d.f.

Since ~~$U = Z$~~ U is a sum of i.i.d. R.Vs, whose 2nd moments exists, based on CLT, we claim:

$$\frac{U}{n} \sim N\left(1, \frac{2}{n}\right) \text{ for large } n.,$$

$\Rightarrow U \sim N(n, 2n)$.

(3) Transformation Technique.
We need the following result before we start:

Let X be a cts. R.V. with pdf $f(x)$. Let $Y=h(X)$ be a transformation s.t. $y=h(x)$ is strictly monotonic so that $x=h^{-1}(y)$ exists.

strictly increase or decrease.

We further assume that $x=h^{-1}(y)$ has a derivative which is cts for all y in the range of h.

Then the pdf of Y exists and is

$$g\cancel{(y)=} \cancel{f(x)\left|\frac{dx}{dy}\right|}$$

$$g(y)=f(x)\left|\frac{dx}{dy}\right|$$

08.03.2018

$$=f[h^{-1}(y)]\left|\frac{d}{dy}h^{-1}(y)\right|$$

Proof:

Assume to begin with that $y=h(x)$ is monotonically increase i.e. $\frac{dy}{dx}>0$, $\frac{dx}{dy}>0$.

Now $G(y)=$ Dist^{n}. f^{n}. of Y

$$=P[Y\le y] = P[h(X)\le y]$$

$$=P[X\le h^{-1}(y)] = \int_{-\infty}^{h^{-1}(y)} f(x)\,dx,$$

Now $x=h^{-1}(y)$, $dx=\frac{dh^{-1}(y)}{dy}\,dy$

$$=\int_{-\infty}^{y} f[h^{-1}(y)]\frac{dh^{-1}(y)}{dy}\,dy$$

Thus $g(y) = f[h^{-1}(y)] \dfrac{dh^{-1}(y)}{dy}$

$$= f(x)\frac{dx}{dy} \quad \text{where } x = h^{-1}(y)$$

Suppose now that $y = h(x)$ is monotonically $\downarrow$
i.e. $\dfrac{dx}{dy} < 0$

Then $G(y) = P[Y \le y] = P[h(x) \le y] = P[X \ge h^{-1}(y)]$

$$= \int_{h^{-1}(y)}^{\infty} f(x)\,dx, \quad \text{Let } y = h(x),\ x = h^{-1}(y)$$

$$dx = \frac{dh^{-1}(y)}{dy}\,dy.$$

$$= \int_y^{-\infty} f[h^{-1}(y)] \frac{d}{dy} h^{-1}(y)\,dy$$

when $x = \infty$, $y = -\infty$

$x = h^{-1}(y)$, $y = y$.

$$= \int_{-\infty}^{y} f[h^{-1}(y)] \underbrace{\left[-\frac{d}{dy} h^{-1}(y)\right]}_{>0} dy$$

$$= \int_{-\infty}^{y} f[h^{-1}(y)] \left|\frac{dx}{dy}\right| dy.$$

Thus for monotonic $y = h(x)$
The PDF of $Y = h(X)$ is given by

$$g(y) = f[h^{-1}(y)] \left|\frac{dx}{dy}\right|$$

Frequently denoted by J and termed the Jacobian of the (inverse) transformation.

<u>Example</u>: Suppose $X \sim R(0,1)$, uniform distⁿ on $0, 1$,

and $Y = -\ln X$, wish to find pdf of Y.

$y=h(x)$: $\underbrace{y=-\ln x}_{\text{monotonic}} \Rightarrow x=e^{-y}=h^{-1}(y)$

$f(x)=1$ $\qquad dx=-e^{-y}\,dy.$

Thus $g(y)=1\cdot\left|\frac{dx}{dy}\right|=e^{-y}$

$\Rightarrow Y \sim$ exp distn. on $(0,\infty)$.

Note: Change of variable in an integral:

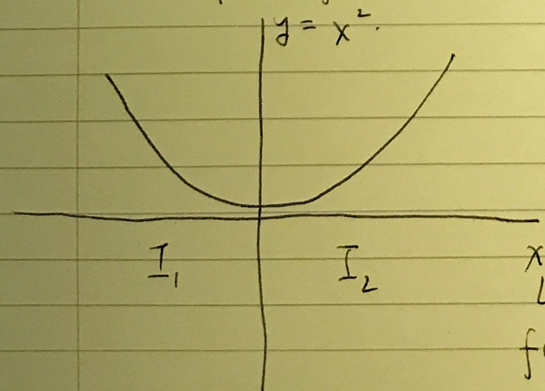
$\int_a^b f(x)\,dx$ suppose $y=h(x) \Rightarrow x=h^{-1}(y)$

$dx=\frac{d}{dy}h^{-1}(y).dy.$

$=\int\left[f[h^{-1}(y)]\frac{dh^{-1}(y)}{dy}\right]dy.$

Function of a R.V., when the transformation $Y=h(X)$ is not one-one., s.t. the inverse $X=h^{-1}(y)$ is multi-valued.

Procedure: Partition S into intervals s.t. $y=h(x)$ is strictly monotonic in each interval $\Rightarrow$ in each interval, the inverse is unique, each such interval contributes to the PDF of Y. — we add to contributions.



$Y=X^2$

$g(y)=\sum_{I_i} f(h^{-1}(y))\left|\frac{dh^{-1}(y)}{dy}\right|$

Let $X\sim N(0,1)$.

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

$$\left.\begin{aligned} &\text{For } I_1: \ g_1(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot \left[\frac{1}{2} y^{-\frac{1}{2}}\right] \\ &\text{For } I_2: \ g_2(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \left[\frac{1}{2} y^{-\frac{1}{2}}\right] \end{aligned}\right\} \ g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} y^{-\frac{1}{2}}.$$

Multi-variate Transformation.

There are corresponding results for the transformation of p R.V.s $X_1, \dots, X_p$ into R.V.s. $Y_1, \dots, Y_p$.

We shall state (without proof) the result for $p=2$:
Let X_1, X_2 be a pair of cts. R.V.s with joint pdf. $f(x_1, x_2)$ and let

$$y_1 = h_1(x_1, x_2)$$
$$y_2 = h_2(x_1, x_2)$$

be a 1-1 transformation of $(x_1, x_2) \in S$ onto $(y_1, y_2) \in T$ (which is one-one) such that the partial derivatives of $x_1 = h_1^{-1}(y_1, y_2)$ and $x_2 = h_2^{-1}(y_1, y_2)$ exist and are cts for all $(y_1, y_2) \in T$.

Then the joint pdf of (Y_1, Y_2) exists and is

$$g(y_1, y_2) = f[h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)] \, |J|,$$

where

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

is the ~~Jacb~~ Jacobian of the inverse transformation.

12.03.2018

(or the Jacobian of (x_1, x_2) w.r.t. (y_1, y_2))
provided $J \neq 0$ on T.

Example :

Let X_1, X_2 be 2 R.Vs having Gamma dist^n's.

i.e. $X_1 \sim \Gamma(\alpha_1, \beta)$

$X_2 \sim \Gamma(\alpha_2, \beta)$

and let us assume that X_1 & X_2 are indep.

Suppose wish to find the joint dist^n. of the R.Vs

$Y_1 = X_1 + X_2$,

$Y_2 = \frac{X_1}{X_2}$, what is dist^n of Y_1?

Need to find X_1, X_2 in terms of y_1, y_2 (for finding J)

sol^n: $X_1 = \frac{y_1 y_2}{1+y_2}$

$X_2 = \frac{y_1}{1+y_2}$

Thus $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{y_2}{1+y_2} & \frac{y_1}{1+y_2} - \frac{y_1 y_2}{(1+y_2)^2} \\ \frac{1}{y_2+1} & \frac{-(y_1)}{(1+y_2)^2} \end{vmatrix}$

$= \frac{-y_1 y_2}{(1+y_2)^3} - \frac{y_1}{(1+y_2)^2} + \frac{y_1 y_2}{(1+y_2)^3}$

$= -\frac{y_1}{(1+y_2)^2}$ notice $J \neq 0$, for any (y_1, y_2) in T.

and $|J| = \frac{y_1}{(1+y_2)^2}$

Thus $g(y_1, y_2) = f(x_1, x_2)\,|J|$ +

$f(x_1, x_2) = f_1(x_1)\, f_2(x_2)$, since x_1, x_2 indep.

Thus

$$g(y_1, y_2) = \left[\Gamma(\alpha_1)\, e^{-\frac{x_1}{\beta}} \frac{(x_1)^{\alpha_1 - 1}}{\beta^{\alpha_1}}\right]\left[\frac{1}{\Gamma(\alpha_2)}\, e^{-\frac{x_2}{\beta}} \frac{(x_2)^{\alpha_2 - 1}}{\beta^{\alpha_2}}\right]\frac{y_1}{(1+y_2)^2}$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\, e^{-\frac{x_1+x_2}{\beta}} \left(\frac{y_1 y_2}{1+y_2}\right)^{\alpha_1-1}\left(\frac{y_1}{1+y_2}\right)^{\alpha_2-1}\frac{y_1}{(1+y_2)^2}\,\frac{1}{\beta^{\alpha_1+\alpha_2}}.$$

$$= \underbrace{\frac{1}{\Gamma(\alpha_1+\alpha_2)}\, e^{-\frac{y_1}{\beta}} \frac{y_1^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2}}}\ \underbrace{\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\, \frac{y_2^{\alpha_1-1}}{(1+y_2)^{\alpha_1+\alpha_2}}}$$

$$= g_1(y_1)\, g_2(y_2)$$

i.e. we can factor the joint pdf of y_1, y_2 into the form $g_1(y_1) \times g_2(y_2)$

$\Rightarrow$ y_1, y_2 are indep.

$\Rightarrow$ $g_1(y_1)$ is the marginal pdf of Y_1,
$g_2(y_2)$ " " " " " Y_2.

We know $g_1(y_1)$ is a Gamma pdf ~~of~~, ~~we~~ can tell since $Y_1 = X_1 + X_2$,

$$G_{X_1} = \frac{1}{(1-s\beta)^{\alpha_1}}, \quad G_{X_2} = \frac{1}{(1-s\beta)^{\alpha_2}}$$

s.t. $$G_{X_1+X_2} = \frac{1}{(1-s\beta)^{\alpha_1+\alpha_2}}$$

Special Case of the result.

we have shown that for $\alpha = \frac{n}{2}$, $\beta = 2$,
the Gamma dist^n ~~is termed~~ $\Gamma(\frac{n}{2}, 2)$ is termed as
~~Sh~~ χ^2_n dist^n with n degree of freedom.

Thus if $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ } indep.

then for $Y_1 = X_1 + X_2$ and $U_2 = \frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}} = \frac{n_2}{n_1} Y_2$

we can derive the joint pdf of (Y_1, Y_2).
Further, that $Y_1 \sim \chi^2_{n_1+n_2}$

easily show that U_2 has pdf $\frac{\Gamma(\frac{n_1+n_2}{2}) (\frac{n_1}{n_2} U_2)^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(1+\frac{n_1}{n_2}U_2)^{\frac{n_1+n_2}{2}}} \cdot \frac{n_1}{n_2}$

and U_2 is said to have an F dist^n, with n_1 and
(Fisher).
n_2 degree of freedom. i.e. F dist^n is specifed by
2 para. n_1, n_2.

Thus for the F dist^n.

$$g(u) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{u^{\frac{n_1}{2}-1}}{(1+\frac{n_1}{n_2}u)^{\frac{n_1+n_2}{2}}} \quad \text{for } u > 0$$

Note: $E(U) = \frac{n_2}{n_2 - 2}$ for $n_2 > 2$.

$$Var(U) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} \quad \text{for } n_2 > 4$$

The t dist^{n}. (Student's t Dist^{n}.)

Suppose the R.V. $X \sim N(0,1)$ and
$U \sim \chi_n^2$, indep of X.

Suppose we form the new R.V.

$$T = \frac{X}{\sqrt{\frac{U}{n}}}$$

Find the dist^{n} of T.

Consider the transformation from X, U to T, S, where

$$t = \frac{x}{\sqrt{\frac{u}{n}}}, \quad s = u,$$

Inverse of transformation : $x = t \cdot \sqrt{\frac{s}{n}}$, $u = s$.

Jacobian of x, u w.r.t. t, s.

$$J = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial u}{\partial s} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{s}{n}} & \frac{1}{2}t\left(\frac{s}{n}\right)^{-\frac{1}{2}} \\ 0 & 1 \end{vmatrix}$$

$$= \sqrt{\frac{s}{n}}.$$

Thus the joint pdf of T and S is
$g(t,s) = f(x,u)\,|J|$ with $x = t \cdot \sqrt{\frac{s}{n}}$, $u = s$.

indep.
$$= \frac{1}{\sqrt{2\pi}} e^{-t^2\frac{s}{2n}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} s^{\frac{n}{2}-1} e^{-\frac{s}{2}} |J|$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} s^{\frac{n}{2}-1} e^{-\frac{s}{2}(1+\frac{t^2}{n})} \sqrt{\frac{s}{n}}.$$

Marginal PDF of $T = \int_0^\infty g(t,s)\,ds$

$$\propto \int_0^\infty s^{\frac{n-1}{2}} e^{-\frac{s}{2}(1+\frac{t^2}{n})} ds.$$

proportional to

let $\delta = \frac{s}{2}(1+\frac{t^2}{n})$, $d\delta = \frac{1}{2}(1+\frac{t^2}{n})ds$.

$$\Rightarrow T \propto \int_0^\infty \left(\frac{2\delta}{1+\frac{t^2}{n}}\right)^{\frac{n-1}{2}} e^{-\delta} \cdot \frac{2}{(1+\frac{t^2}{n})} d\delta.$$

$$\propto \int_0 \frac{2^{\frac{n+1}{2}}}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \cdot \delta^{\frac{n-1}{2}} e^{-\delta} d\delta.$$

<u>Recall:</u> $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx.$

$\cancel{T \propto}$

$$f_T(t) = \frac{1}{\sqrt{2\pi n}\,\Gamma(\frac{n}{2})\, 2^{\frac{n}{2}}} \cdot \frac{2^{\frac{n+1}{2}}}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \Gamma\left(\frac{n+1}{2}\right).$$

$$= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \cdot \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \qquad -\infty < t < \infty$$

— known as the t dist^{n}. with one para. n as degree of freedom.

It is symmetric : $E(T) = 0$.

$$Var(T) = \frac{n}{n-2} \text{ for } n>2.$$

Sampling and Statistical Inference.

Up to know, much of the course material has been concerned with the theory and method of probability.

Now come to the concepts of statistics.

Recall: We began the course with random experiments

and its sample space, then to consider R.Vs,

Suppose a Random Exp. were to be repeated n times, resulting in observations:

$$x_1, x_2, \text{---}, x_n.$$

These obs. can provide information on underlying random exp.

Sampling: The process of gathering data from several performance of a R.E. is called sampling.

The individual items of data are termed observations

The collection of observations is called a sample.

Prior to getting the numerial values of the observations, each individual "observation - to - be" has the potential for variability, and we regard each "observation - to - be" as a R.V.

We'll ~~most~~ mainly focus on obs^{n}. on a single variable. i.e, univariate statistics.

Random sampling and statistical Inference.

Typical sampling process 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 population having equal chance of selection.
When an object is drawn, some characteristic associate with the object may be observed.
– object is then replaced – and population mixed before next drawing.

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

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

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

Testing a new alloy under stress conditions.

We have already used X to denote the R.Vs. associated with the random experiment.

$\Rightarrow$ can regard a sample as a set of obs. of the R.V. X.

The prob. of X is determined by the pattern of dist^{n} of values in the population.

We refer to the prob. dist^{n} of X as the <u>population dist^{n}</u>.

Before the observations in a sample are made, we have in mind to observe X n times, or alternatively to observe $X_1, X_2, \cdots, X_n$.

[When these n R.V.s are indep. and identically dist^{n} (i.i.d), the term <u>Random sample</u> is used to refer to the n obs. on the R.V. X.

Thus if the dist^{n}. of X is in the population is characterised by pdf $f(x)$, we have the following def^{n}.:

If the R.V.s $X_1, X_2, ---, X_n$ have a joint pdf $f(x_1, x_2, ---, x_n) = \prod_{i=1}^{n} f(x_i)$, where

$f(\cdot)$ is the common pdf of X_i, then the R.Vs X_i are said to constitute a random sample of size n from the population characteristized by $f(\cdot)$.

One of the central problems in statistics is to determine the form of the density $f(\cdot)$ from the obs^{n}. $x_1, x_2, ----, x_n$.

We shall see that in many instances, $f(\cdot)$ is a known f^{n}. of certain unknown parameter $\underline{\theta}$ and our task is to find good estimates for $\underline{\theta}$.

eg. If we could assume that the blood-pressures of a pop^{n}. of students were normally dist^{d}. (but μ, σ unknown) then we shall attempt to estimate μ & σ, which are the mean and SD characterising the pop^{n}.

Statistical Inference. is concerned with making statements concerning the pop^{n}. on the basis of sample observations.

It is customary to subdivide the main study area of stat. inference as follows:

(1) Estimation

How should we provide an estimate for a certain char. of a pop^{n}.
How do we chose among alternative estimates.

(2) Hypothesis Testing.
How do we determine whether our obs^{n}'s are consistent with certain assumptions — or hypothesis about the pop^{n}.

(3) Decision Theory.
We have some process we wish to control. Obs^{n}. are taken and resulting from an analysis of the obs^{n}'s, an action follows to keep the process in control.

eg. Machine filling commodity into containers.
The filled ~~package~~ are examined peordically and action may follow.

Note: In this course, no time for depth study of these areas — in remaining time shall ~~we~~ cont. on the theory + application of more common and useful statistical procedure.

Some Terminologies.

Statistics:

A statistic is any known f^{n}. of observable random variables — does not involve any unknown parameters.

eg. Random sample: $X_1, X_2, \cdots, X_n$ from $dist^n$. with pdf. $f(x)$

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

$$\frac{X_1 + X_2}{2}, \quad X_n - X_1$$ are all statistics.

But $\bar{X} - \mu$, $\frac{X_n - X_1}{\sigma}$ are not stat. unless μ, σ are known.

Sample moments

Sample moments about origin.

$$m_1' = \frac{\Sigma X_i}{n} = \bar{X},$$

$$m_2' = \frac{\Sigma X_i^2}{n}$$

$$\vdots$$

$$m_r' = \frac{\Sigma X_i^r}{n}$$

Corresponding sample moments ~~are~~ about sample mean are:

$$m_1 = 0 = \frac{\sum^{n} (X_i - \bar{X})}{n}$$

$$m_2 = \frac{1}{n} \Sigma (X_i - \bar{X})^2$$

$$\vdots$$

$$m_r = \frac{1}{n} \Sigma (X_i - \bar{X})^r.$$

[It is useful to find the expected value of these samples moments — these expected values can be found providing the corresponding pop^n. moments exist.

eg. ~~m_i~~' $E[m_1'] = E[\bar{X}] = E[\frac{\sum X_i}{n}]$

$= \frac{1}{n} \times \sum E[X_i] = \frac{1}{n} \cdot n\mu$

$= \mu.$

$E[m_2'] = E[\frac{1}{n} \sum X_i^2] = \frac{1}{n} \sum^{n} E[X_i^2] = \mu_2'$

$E[m_r'] = \mu_r'$

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

<u>Sample moments about sample mean</u>

$E[m_2] = E[\frac{1}{n} \sum^{n} (X_i - \bar{X})^2] = \frac{1}{n} E[\sum X_i^2 - n(\bar{X})^2]$

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

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

20.03.2018.

$E(\bar{X}^2) = \frac{1}{n^2} E[(\sum^{n} X_i)^2] = \frac{1}{n^2} E[\sum X_i^2 + 2 \sum_{i>j} \sum X_i X_j]$

$= \frac{1}{n^2} \{ \sum^{n} E(X_i^2) + 2 \sum_{i>j} \sum E(X_i) E(X_j) \}$

$\downarrow 2 \times {}^nC_2$

$= \frac{1}{n^2} \{ n(\sigma^2 + \mu^2) + n(n-1)\mu^2 \}$

$= \frac{1}{n} \{ \sigma^2 + \mu^2 + n\mu^2 - \mu^2 \} = \mu^2 + \frac{\sigma^2}{n}.$

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

$= \frac{n-1}{n}\mu_2.$

or $E[\bar{X}^2] = Var(\bar{X}) + (\mu)^2 = \frac{\sigma^2}{n} + \mu^2.$

Estimator

An ~~E~~estimator for a $\left\{ \begin{array}{l} \text{dist}^{n} \\ \text{pop}^{n}\text{. para.} \end{array} \right\}$ is a statistic that may be used to provide an estimate for the parameter.

eg. it seems intuitive to use $\bar{X}$ as an estimate of μ and m_2 as . . . . . μ_2.

But $E(m_2) = \frac{n-1}{n} \cdot \mu_2$ 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 is $\frac{\sum^{n}(X_i - \bar{X})^2}{n-1}$

— refer to this as the sample variance and

denote as S^2.

—————————— // ——————————

You need to distinguish.

the moment of $\cancel{\text{pop}}$ pop^n. dist^n. : The ~~prob~~ pop^n. moment.

" " " sample dist^n. : The sample moment

Pop^n. Moments.	Sample Moment.
$\mu_1' = \mu$.	$m_1' = \bar{X}$
$\mu_2 = \sigma^2$	$m_2 = \frac{1}{n}\sum(X_i - \bar{X})^2$
$\vdots$	$\vdots$
$\mu_r = \sim$	$m_r = \frac{1}{n}\sum(X_i - \bar{X})^r$.

Sampling from a Normal Dist^n.

The R.V. $X \sim N(\mu, \sigma^2)$

$E(X) = \mu$, $Var(X) = \sigma^2$.

Suppose a random sample of n obs^n's from this dist^n: Intuitive estimators for μ & σ^2 are $\bar{X}$ & S^2

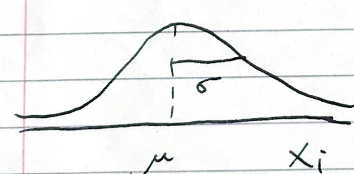
Now $\bar{X} = \frac{\Sigma X_i}{n}$ is a statistic and thus a R.V. $\Rightarrow$ It has a dist^n:

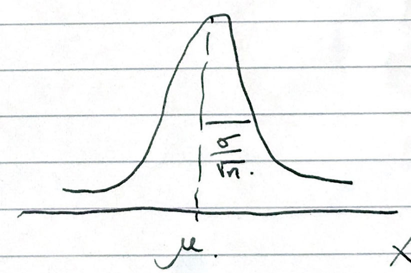
In the context of sampling, the prob. dist^n of $\bar{X}$ is termed the sampling dist^n of $\bar{X}$

or " " " " the statistic $\bar{X}$.

What is the sampling dist^n of $\bar{X}$?

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$





Interval Estimate for μ

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

$P[|\bar{X} - \mu| \leq 1.96 \cdot \frac{\sigma}{\sqrt{n}}] = 0.95$.

$\Rightarrow P[-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}] = 0.95$.

$\Rightarrow P[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}] = 0.95$

i.e. The prob. that μ lies in the Random Interval is 95%.

— The ends of the interval are R.V.s.

Once the obs^{n}s are made, $\bar{X}$ assumes a numerical value (lets assume σ known).
$\Rightarrow$ Some interval such as $100.9 \to 110.1$.

But the predominant school of thought in STATS would avoid the statement that μ lies in this fixed interval with prob. 0.95.

Why? Because μ is either in the fixed interval or not.

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

Confidence Interval
" Coefficient — rather than prob.

Thus if σ is known, ~~the~~ the expression for the 95% CI for μ is the case of sampling from a normal dist^{n}, is

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}.$$

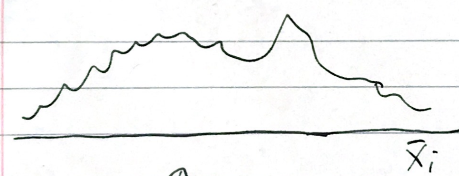
Sampling from a General prob. Dist^n.

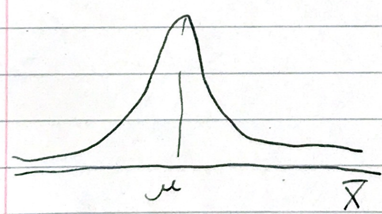
Suppose the mean + variance of the prob. dist^n. is μ + σ^2 respectively.

Random sample of size n selected, and $n \to \infty$.

By the CLT, the sampling dist^n. of $\bar{X}$ will be normal with mean μ
var σ^2/n.

i.e. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ for large n.





Thus taking σ to be known, the approx. 95% CI for μ is given by

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}.$$

Independence of $\bar{X}$ and s^2 (Give handout)

→ Sampling dist^n of s^2.

Can we find a CI for σ^2 just as we did for μ?

Find the sampling dist^n of s^2.

Let $X_1, X_2, \ldots X_n$ be a Random sample from $N(\mu, \sigma^2)$.
Introduce new R.Vs $Y_1, Y_2, \ldots, Y_n$ by means of a linear transformation of $X_1, \ldots, X_n$.

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

i.e. $Y = PX$.
Can check P is orthogonal, i.e. $PP^T = P^TP = I$
$\mathrm{Det}(P) = 1$

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

$$\left.\begin{array}{ll} E(Y_i) = 0 & \text{for } i > 1. \\ Var(Y_i) = \sigma^2 & \forall i \\ cov(Y_i, Y_j) = 0 & \forall i \neq j \end{array}\right\} \text{check !!}$$

The joint pdf of $Y_1, Y_2, \ldots Y_n$ is.
$g(y_1, y_2, \ldots y_n) = f(x_1, \ldots, x_n)\,|J|$

Now $J = \frac{\partial(x_1, \ldots x_n)}{\partial(y_1, \ldots y_n)}$ the ~~Jach~~ Jacobian of X w.r.t Y.

since P is orthogonal,
$Y = PX \Rightarrow P^TY = P^TPX = X$.

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

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

when we must replace X_i's in terms of y_i's.

Now $\sum_1^n (X_i-\mu)^2 = (X-\mu)^T(X-\mu)$

where $x = \begin{pmatrix} x_1 \\ | \\ | \\ x_n \end{pmatrix}$, $\mu = \begin{pmatrix} \mu \\ | \\ | \\ \mu \end{pmatrix}$

$$= (x-\mu)^T P^T P (x-\mu)$$
$$= [P(x-\mu)]^T P(x-\mu)$$

Now $P(X-\mu) = PX - P\mu$.

$$= \begin{pmatrix} y_1 \\ | \\ | \\ y_n \end{pmatrix} - \begin{bmatrix} \sqrt{n}\,\mu \\ 0 \\ | \\ 0 \end{bmatrix}$$

Thus $\sum_{i=1}^{n} (X_i-\mu)^2 = (y_1 - \sqrt{n}\,\mu)^2 + \sum_{i=2}^{n} y_i^2$

Note: $Y_1, \text{---}\, Y_n$ are all indep.

$$\begin{cases} Y_1 \sim N(\sqrt{n}\mu, \sigma) \\ Y_i \sim N(0, \sigma);\ i > 1. \end{cases}$$

Thus

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

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

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

and using $y_1 = \sqrt{n}\bar{X}$, we get

$$\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2 = \sum_{i=2}^{n} y_i^2$$

i.e. $\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=2}^{n} y_i^2$

$$\Rightarrow \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} = \sum_{i=2}^{n} \left(\frac{y_i}{\sigma}\right)^2$$

But $y_i/\sigma \sim N(0,1)$ and y_i are indep.

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

$$\boxed{\text{Thus } \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \equiv \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}}$$

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

Usually expressed as $\bar{X}$ and s^2 are indep.

21.03.2018.

Expected Value of Sampling Standard Deviation when Sampling from a Normal Dist^{n}.

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}

Denote this as X,

$$E[\sqrt{X}] = E\left[\sqrt{\frac{n-1}{\sigma}}\, s\right] = \sqrt{2} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$ 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.

eg. $n=4$: $E(s) = \sigma\sqrt{\frac{2}{3}} \frac{\Gamma(2)}{\Gamma(\frac{3}{2})} = \sigma\sqrt{\frac{2}{3}} \frac{1}{\frac{1}{2}\sqrt{\pi}} = 0.9213\sigma.$

$n=25$, $E(s) = 0.99\sigma.$

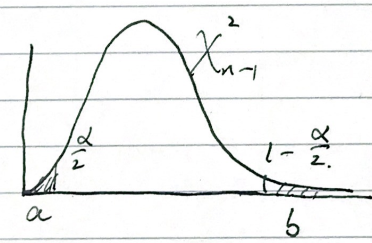
Confidence Interval for σ^2

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

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

let $a = \frac{\alpha}{2}$ for χ^2_{n-1}

$b = 1 - \frac{\alpha}{2}$ for χ^2_{n-1}



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

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

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

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

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

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

Typical Single Sample Problem.

~~Given~~ 10 students are drawn as a Random sample with the following Blood pressure:

105, 112, 130, 128, 100, 98, 122, 118, 106, 110.

Assume B.P $\sim N(\mu, \sigma^2)$, here B.P is a population with para. μ, σ^2. ~~We~~ Need to estimate μ, σ^2.

Point estimator for μ & σ^2 $\Big\}$: $\bar{X} = 112.9$

$$s^2 = \frac{1}{\underbrace{\boxed{9}}_{=n-1}}\left[128{,}601 - \frac{(1129)^2}{10}\right] = 126.32.$$

If σ^2 is known, the 95% CI for μ is

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}} \quad \text{since} \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

If σ^2 is unknown, can we use $\bar{X} \pm 1.96 \frac{s}{\sqrt{n}}$ as 95% CI? No!!

Derive the dist^n. of $\frac{\bar{X}-\mu}{s/\sqrt{n}}$.

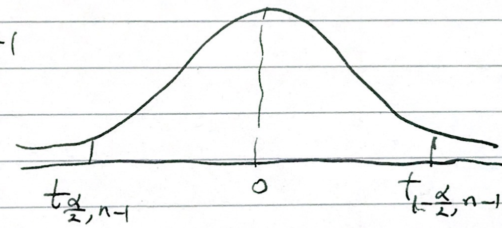
Given $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$,

and $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ indep of $\bar{X}$ and thus $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.

Recall, if $X \sim N(0,1)$, $U \sim \chi^2_n$, Then $\frac{X}{\sqrt{\frac{U}{n}}} \sim t_n$ dist^n.

Therefore, here

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2(n-1)}}} \sim t_{n-1}$$



i.e. $\dfrac{\bar{X}-\mu}{\cancel{\sigma}\, s/\sqrt{n}} \sim t_{n-1}$

Thus $P\left[-t_{\frac{\alpha}{2},n-1} < \dfrac{\bar{X}-\mu}{s/\sqrt{n}} < t_{\frac{\alpha}{2},n-1}\right] = 100(1-\alpha).$

$\Rightarrow P\left[\bar{X} - t_{\frac{\alpha}{2},n-1}\cdot\dfrac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},n-1}\cdot\dfrac{s}{\sqrt{n}}\right] = 100(1-\alpha).$

i.e. $\bar{x} \pm t_{\frac{\alpha}{2},n-1}\dfrac{s}{\sqrt{n}}$ is the $100(1-\alpha)\%$ CI for μ, if σ^2 is unknown.

Apply to our data,

95% CI : $t_{\frac{\alpha}{2},9} = 2.26$, check the table.

$\Rightarrow 112.9 \pm 2.26\cdot\dfrac{\sqrt{126.3}}{\sqrt{10}} = 112.9 \pm 2.26 \times 3.55$

Summary: CI for μ. $100(1-\alpha)\%$

σ known : $\bar{X} \pm Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$.

σ unknown: $\bar{X} \pm t_{\frac{\alpha}{2},n-1}\cdot\dfrac{s}{\sqrt{n}}$.

$\alpha = 0.05$ for 95% CI.

If $n\to\infty$, even σ is unknow, we can use

$$\bar{X} \pm Z_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}}.$$

CI for σ^2 : $\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \rightarrow \dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}$

$n-1=9$, $\alpha = 5\%$

$\Rightarrow$ 95% CI for σ^2 : $\dfrac{1136.9}{17.535} \rightarrow \dfrac{1136.9}{2.18}$

Check table.

For σ^2 $(64.8, 521.5)$

" σ $(8, 22.8)$

22. 03. 2018

Method of Moment

Here, we equate the popⁿ. moment to corresponding sample moments, then solve for the para.

One para. case.

Usually, equate popⁿ. mean to sample mean.

eg. $X_i \sim Poi(m)$, sample of size n.

$E(X_i) = m, \; = \bar{X}$

$\Rightarrow \hat{m} = \bar{X}$

$X_i \sim$ Neg. ~~B~~ $Exp(\lambda)$

$E(X_i) = \frac{1}{\lambda}, \; = \bar{X}, \Rightarrow \hat{\lambda} = \frac{1}{\bar{X}}$

Geometric: $P[X=r] = q^{r-1} p$, $E(X) = \frac{1}{p} = \bar{X}$,

$\hat{p} = \frac{1}{\bar{X}}$

Two para. Case: Need two equations.

Set 1^{st} & 2^{nd} popⁿ. moments $= 1^{st}$ & 2^{nd} sample mean.

1^{st} moment: $E(X_i) = \bar{X}$ as before.

2^{nd} moment: $E(X_i^2) = \frac{\Sigma X_i^2}{n}$

or $E[(X_i - \mu)^2] = \frac{\Sigma(X_i - \bar{X})^2}{n}$

E.g. Binomial: $X_i \sim Bin(n, p)$: observe $x_1, \cdots, x_N$.

$$E(X_i) = np, \overset{set}{=} m_1$$

$$E(X_i^2) = \underbrace{np(1-p)}_{Var(X_i)} + (np)^2 \overset{set}{=} m_2 = \frac{\sum X_i^2}{N}$$

$$\Rightarrow \hat{n} = \frac{m_1^2}{m_1 - m_2 + m_1^2}, \quad \hat{p} = \frac{m_1 - m_2 + m_1^2}{m_1}$$

$X_i \sim Gamma(\alpha, \beta)$

$$E(X_i) = \alpha\beta = \bar{X}, \qquad m_1'$$

$$Var(X_i) = \alpha\beta^2 = \frac{\sum (X_i - \bar{X})^2}{n} \qquad m_2'$$

$$\Rightarrow \hat{\beta} = \frac{m_2'}{\bar{X}}, \quad \hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{\bar{X}^2}{m_2'}$$

Bias $= E[\hat{\theta}] - \theta$, $\hat{\theta}$ is called unbiased estimator if $E(\hat{\theta}) = \theta$.

MSE (Mean Square Error)

$$MSE[g(x)] = \cancel{E[g(x)] - \theta} \; E[(g(x) - \theta)^2]$$

Easy to show that

$$MSE = Variance + (Bias)^2.$$

$$MSE = E[(g(x) - E(g) + E(g) - \theta)^2]$$

$$= E[(g(x) - E(g))^2] + E[(E(g) - \theta)^2]$$

$$= Var[g(x)] + (Bias)^2,$$

Note: $E(g) - \theta$ = Bias is a constant.

E.g. Binomial: $X_i \sim Bin(n, p)$: observe $x_1, ---, x_N$.

$E(X_i) = np, \overset{set}{=} m_1$

$E(X_i^2) = \underbrace{np(1-p)}_{Var(X_i)} + (np)^2 . \overset{set}{=} m_2 = \frac{\Sigma X_i^2}{N}$

$\Rightarrow \hat{n} = \frac{m_1^2}{m_1 - m_2 + m_1^2}$, $\hat{p} = \frac{m_1 - m_2 + m_1^2}{m_1}$

$X_i \sim Gamma(\alpha, \beta)$

$E(X_i) = \alpha\beta = \bar{X}$, m_1'

$Var(X_i) = \alpha\beta^2 = \frac{\Sigma(X_i - \bar{X})^2}{n}$ m_2'

$\Rightarrow \hat{\beta} = \frac{m_2'}{\bar{X}}$, $\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{\bar{X}^2}{m_2'}$

Bias $= E[\hat{\theta}] - \theta$, $\hat{\theta}$ is called unbiased estimator if $E(\hat{\theta}) = \theta$.

MSE (Mean Square Error)

$MSE[g(x)] = \cancel{E[g(x) - \theta} \; E[(g(x) - \theta)^2]$

Easy to show that

$MSE = Variance + (Bias)^2$.

$MSE = E[(g(x) - E(g) + E(g) - \theta)^2]$

$= E[(g(x) - E(g))^2] + E[(E(g) - \theta)^2]$

$= Var[g(x)] + (Bias)^2$,

Note: $E(g) - \theta$ $= Bias$ is a constant.

The Method of Maximum Likelihood.

~~The~~

Consider R.V.s. $X_1, X_2, \dots, X_n$ having joint pdf $f(x_1, x_2, \dots, x_n)$ — or, if discrete, having joint prob. f^n. $f(x_1, x_2, \dots, x_n)$ — which contains one or more parameters θ.

The likelihood of θ, as a f^n. of $x_1, \dots, x_n$ is defined as

$$L(\theta) = L(\theta | x_1, \dots x_n) = f(x_1, \dots, x_n | \theta)$$

Here, we shall confine our consideration to random samples $X_1, \dots, X_n$ s.t.

$$L(\theta) = \prod_{i=1}^{n} f(x_i | \theta),$$

where X_i are i.i.d. from a pop^n. with pdf (or pmf) $f(x|\theta)$.

Note: The likelihood f^n. gives the prob. of observing the given values $(X_1, \dots, X_n)$, as a f^n. of para. θ, in the ~~pr~~ discrete case — and in the cts case the likelihood f^n. is proportional to the prob. of observing values in the neighbourhood of the given values.

The Maximum likelihood ~~of~~ estimator (MLE) of θ is ~~that~~ the value that maximizes the likelihood i.e. make the observed data most likely to be observed.

More nature to maximise the natural log of $L(\theta)$, known as log-likelihood, denoted as $l(\theta)$.

The process of maximizing is often done by differentiation, with r.t. θ, set derivative to zero, solve for θ.

Some Examples.

A Random sample from $Poi(m)$.

$$L(m) = \prod_{i=1}^{n} \frac{m^{X_i} e^{-m}}{(X_i)!}$$

$$\ln L(m) = \sum_{i=1}^{n} [X_i \ln m - m - \ln(X_i!)]$$

$$= \sum_{i=1}^{n} [X_i \ln m - m] + C.$$

$$\frac{d l(m)}{dm} = \frac{1}{m} \sum_{i=1}^{n} X_i - n = 0.$$

$$\Rightarrow \hat{m} = \bar{X} \quad \text{(same as for Method of Moments)}$$

To check that $\hat{m}$ yields a max to the likelihood f^{n}, diff. again, - and check the sign for $m = \bar{X}$, which is negative, indicating that this is a max.

Cts case: $X \sim exp(\lambda)$

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$$

$$\ln L(\lambda) = \sum_{i=1}^{n} [\ln \lambda - \lambda X_i] = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i$$

$$\frac{d l(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0 \qquad \hat{\lambda} = \frac{1}{\bar{X}}$$

ML Estimation for Two Parameters.

Typical examples : $N(\mu, \sigma^2)$, $Gamma(\alpha, \beta)$

Consider a sample $(X_1, \dots, X_n)$ from $N(\mu, \sigma^2)$.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum (X_i - \mu)^2.$$

$$\cancel{\partial \ln} \quad \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^3} = 0.$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}.$$

To check that $\hat{\mu}, \hat{\sigma}$ maxmizes $L(\mu, \sigma^2)$ one could consider the Hessian matrix, which will not be coveredin this module.

ML estimation of (α, β) from the Gamma Dist^{n}.

~~$\ln L(\alpha, \lambda)$~~ Suppose $X_i \sim Gamma(\alpha, \lambda)$.

$$\ln L(\alpha, \lambda) = \ln\left[\prod_{i=1}^{n}\left\{\frac{1}{\Gamma(\alpha)}(\lambda X_i)^{\alpha-1} e^{-\lambda X_i}\lambda\right\}\right]$$

$$= \sum\left[\alpha\ln\lambda - \ln\Gamma(\alpha) + (\alpha-1)\ln(X_i) - \lambda X_i\right]$$

$$= n\alpha\ln\lambda - n\ln\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\ln X_i - \lambda\sum_{i=1}^{n} X_i$$

$$\frac{d\ln L}{d\alpha} = n\ln\lambda + \sum_{i=1}^{n}\ln X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\frac{d\ln L}{d\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i$$

$$\Rightarrow \hat{\lambda} = \frac{n\hat{\alpha}}{\Sigma X_i} = \frac{\hat{\alpha}}{\bar{X}}$$

$\Rightarrow n\ln\left(\frac{\hat{\alpha}}{\bar{X}}\right) + \sum_{i=1}^{n}\ln X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0$, which is a non-linear eq^{n}. for $\hat{\alpha}$, and cannot be solved explicitly.

Need to use numerical methods.

<u>Invariance Property for ML estimators:</u>
If $\hat{\theta}$ is the MLE for θ, then $g(\hat{\theta})$ is the MLE for $g(\theta)$.

eg. For ~~N~~ normal data ~~N~~ from $N(\mu, \sigma^2)$,

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$
$\Rightarrow$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{N}(X_i-\bar{x})^2.$$

<u>Applicibility to Truncated or Censored Data.</u>

This is a very useful aspect of ML estimation, which will be used in a later course on Survival Analysis. The data may consist of n_1 observations of a R.V. X, and a further n_2 obs^{n}'s where we know only that the obs^{n}. is greater than a value t.
Then the likelihood f^{n}. is

$$L(\theta) = \left[\prod_{i=1}^{n_1} f(x_i, \theta)\right]\left[P[X>t]\right]^{n_2}.$$

and the MLE for θ could be found by max $L(\theta)$.

Biased Estimator + Mean Square Error.

$\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$. A measure of performance for a biased estimator is the Mean square error (MSE)

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$, which is not the variance of $\hat{\theta}$

It is the second moment of $\hat{\theta}$ about θ.

$$\cancel{MSE(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2]}$$

The Dist^n. of an MLE for large n.

This is also known as the asymptotic dist^n. of a MLE.

Suppose that $\hat{\theta}$ is the MLE for θ. It can be shown that, as $n \to \infty$

(i) $\hat{\theta}$ is unbiased.

(ii) $\hat{\theta}$ is normally distributed.

(iii).

$$Var(\hat{\theta}) = -\frac{1}{E\left[\frac{d^2}{d\theta^2} \ln L(\theta)\right]} \quad \text{or} \quad \cancel{\frac{1}{E\left[\frac{d}{d\theta} \ln L(\theta)\right]^2}}$$

$$\frac{1}{E\left[\left(\frac{d}{d\theta} \ln L(\theta)\right)^2\right]}$$

The expression for $Var(\hat{\theta})$ is known as the Cramer-Rao Lower Bound (CRLB)

No unbiased estimator for θ can have a smaller variance than the CRLB. (The CRLB does not apply when the range of values for the dist^{n}. is related to θ, eg. $U(0, \theta)$).

Examples of CRLB Evaluation

For $X_1, \ldots, X_n \sim Poi(m)$,

$\hat{m} = \bar{X}$. $\quad E(\hat{m}) = E(\bar{X}) = m$

$+ \; Var(\hat{m}) = Var(\bar{X}) = \frac{m}{n}$. $\leftarrow$ Check this !!

$$\ln L(m) = \sum_{i=1}^{n} (X_i \ln m) - mn - (\ln X_i!)$$

$$\frac{d^2 \ln L(m)}{dm^2} = -\frac{1}{m^2} \sum_{i=1}^{n} X_i$$

So that $E\left[\frac{d^2 \ln L(m)}{dm^2}\right] = -\frac{1}{m^2} \sum_{i=1}^{n} E(X_i) = -\frac{n}{m}$.

$\Rightarrow$ CRLB $= \frac{m}{n}$.

We can say that the sample mean $\bar{X}$ attains the CR lower bound for $Var(\hat{m})$ in this case.

For $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, let $\theta = \sigma^2$.

$$\ln L(\mu, \theta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta - \frac{1}{2\theta} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{d^2 L(\mu, \theta)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^2} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2.$$

check this !!

$$E\left[\frac{d^2}{d\theta^2} \ln L(\mu, \theta)\right] = \frac{n}{2\theta^2} - \frac{n}{\theta^2}$$

~~Another example: The CRLB~~

Thus the CRLB for estimation of σ^2 from a random sample from $N(\mu, \sigma^2)$:

$$= \frac{2\theta^2}{n} = \frac{2\sigma^4}{n}.$$

Thus for any unbiased estimator $\hat{\theta}$ for ~~θ~~ σ^2,

$$Var(\hat{\theta}) \geq \frac{2\sigma^4}{n}.$$

Example: A random sample $\{X_i\}_{i=1,\dots,n}$ from pdf $f(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}$ for $x > 0$

$$\ln L(x_1, \dots, x_n) = \sum_{i=1}^{n} \ln x_i - n \ln \theta - \frac{\sum x_i^2}{2\theta}$$

$$\frac{d}{d\theta} \ln L = \frac{-n}{\theta} + \frac{\sum x_i^2}{2\theta^2} = 0$$

$$\Rightarrow \hat{\theta} = \frac{\sum x_i^2}{2n}$$

$$E(\hat{\theta}) = \frac{\sum_{i=1}^{n}}{2n} \int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} dx.$$

let $u = \frac{x^2}{2\theta}$, $\Rightarrow du = \frac{x}{\theta} dx$ $\quad x^2 = 2\theta u.$

$$E(\hat{\theta}) = \frac{\sum_{i=1}^{n} \int_0^\infty 2\theta u \, e^{-u} du}{2n} = \frac{\sum_{i=1}^{n} 2\theta \Gamma(2)}{2n} = \theta.$$

$$CRLB = \frac{-1}{E\left[\frac{d^2}{d\theta^2} \ln L\right]}$$

$$\frac{d^2 \ln L(\theta)}{d\theta^2} = \frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} x_i^2}{\theta^3}$$

$$E\left[\frac{d^2}{d\theta^2} \ln L\right) = \frac{n}{\theta^2} - \frac{\sum}{\theta^3} \int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} dx.$$

$$= \frac{n}{\theta^2} - \frac{\sum_{1}^{n} 2\theta}{\theta^3} = -\frac{n}{\theta^2}$$

$$\Rightarrow CRLB = \frac{\theta^2}{n}.$$

Hypothesis Testing: Single sample problems.

We have a random sample $X_1, \ldots, X_n$ from a prob. dist^n. $f(x)$ and a hypothesis concerning $f(x)$ is to be tested using the sample data.

i.e. is the data consistent with the hypothesis?
— how consistent or inconsistent?

The hypothesis under test is termed the Null hypothesis and is defined by H_0. In this course, the

hypothesis ~~test~~ to be tested will only be concerned with parameters of the prob. dist^n. $f(x)$.
e.g. $\mu = \mu_0$ — denoted by $H_0: \mu = \mu_0$.

Example: Jars of coffee are being filled by a machine. The machine is set to put 225g in each jar. A random sample of 16 jars is taken from the output from the machine and $\bar{X}$ found to be 223.7 g.

Suppose the std Dev. of the scatter in fill of the filling machine is 2g.
The packer would like to know if the machine is operating satisfactorily. I.E. he asks if

$H_0: \mu = \mu_0 = 225$ is true.

We should like to study how to test whether the data are consistent with this hypothesis.

Simple & Composite Hypothesis.

When a hypothesis fully specifies a particular prob. dist^n, it is called simple.

When " " specifies a range of prob. dist^{n}s, it is called composite.

In testing the hypothesis H_0 we are essentially trying to decide whether H_0 is consistent with the data or not.

It is usual to use the term Alternative Hypoth. in referring to the situation when H_0 is not true, denoted as H_1 or H_A.

The terminology used is that we are testing H_0 against H_1.

2 Possible Errors.

~~1~~ · Type 1 error : Reject H_0 when H_0 is true.

" 2 error : Accept H_0 " " " False.

We can evaluate a test procedure by determine the prob. of the 2 types of errors that might be incurred using that procedure.

Seriousness of Errors.

In some situations one type of error may be more serious than the other:

e.g. Testing a new drug

Drug is either harmful or not.

status of Nature.

	Drug is harmful	Not harmful
Reject Drug	✓	Error
Accept Drug	Error	✓

Convention is that ~~Type~~ the term Type I error is applied to the more serious error.

— Just choose the hypothesis names accordingly.

In our example, this leads to the choice

$H_0: \mu = 225$

$H_1: \mu \neq 225$.

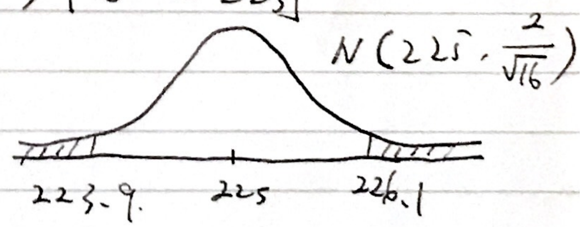
P-value : A very useful way of quantifying the strength of the evidence against H_0 is to find the P-value — which is the prob. of observing a value for the sample statistic that is at least as extreme as the observed value, assuming H_0 true

Thus, we would have to find

$$P[(\bar{X} \leq 223.9) \text{ or } (\bar{X} \geq 226.1) \mid \mu = 225]$$

$$= 2P[\bar{X} \geq 226.1 \mid \mu = 225]$$

$$= 2P[\bar{X} \geq 226.1 \mid \mu = 225]$$



$$= 2P\left[\frac{\bar{X} - 225}{2/\sqrt{16}} \geq \frac{226.1 - 225}{2/\sqrt{16}}\right]$$

$$= 2P[Z \geq 2.2] = 0.028 \leftarrow \text{P-value}.$$

Terminology: 0.028 has also been termed the level of statistical significance of the obs^n. $\bar{x} = 223.9$ in relation to $H_0: \mu = 225$.

→ Test statistic: The quantity $\frac{\bar{X} - 225}{\sigma/\sqrt{n}}$ is an example of a test statistic — which is some f^n. of the observations that we can use to ~~disc~~ discriminate between H_0 & H_1.

The critical region is the set of values of the test statistic for which H_0 is rejected.
– extent of ~~cor~~ critical region is determined by α (significant level).
Here we want $P\left[\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \in \text{Crit. Regin} \mid H_0 \text{ true}\right] = 0.05$.

⇒ critical regin extends from $-\infty$ to -1.96
$+\infty$ to $+1.96$.

Test procedure:

Evaluate the test statistic for given data ~~if in~~
if in critical Region: Reject H_0

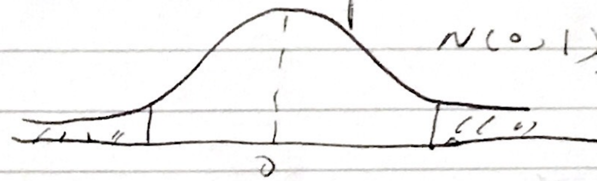
else : Accept.

Here $\frac{\bar{X} - 225}{1/2} = \frac{223.9 - 225}{\frac{1}{2}} = -2.2$ ⇒ Reject.

~~Another common procedure is as follows:~~
The ~~acceptance~~ acceptable level of Type I error, (α) ~~B~~ is specified, e.g. 5% and 1%.

The dist^n. of the test statistic under the assumption that H_0 is true is determined, in our example,

Test statistic : $\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$



Test of Hypothesis $\mu = \mu_0$ when σ unknown.

Need to know the dist^{n}. of test statistic under the assumption that H_0 is true.

$\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ (for σ is known) is $N(0, 1)$ under $H_0 : \mu = \mu_0$.

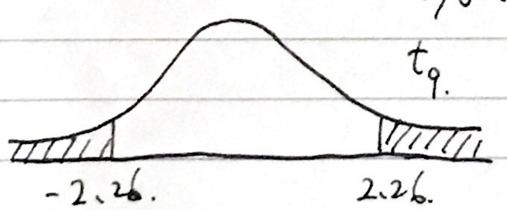
When σ is unknown, use $\dfrac{\bar{X}-\mu_0}{s/\sqrt{n}}$ as test statistic and under H_0, it has a t_{n-1} dist^{n}.

Example: Use the data on student Blood pressure.

Test $H_0 : \mu = 120$

$H_1 : \mu \neq 120$.

Test statistic : $\dfrac{\bar{X}-120}{s/\sqrt{n}}$, 5% signif. level.



Value of test statistic $= \dfrac{112.9-120}{\sqrt{\dfrac{126.32}{10}}} = \dfrac{-7.1}{3.55} = -2.$

$\Rightarrow$ Accept H_0, p-value = 0.038 ?

Note: 95% CI for μ : $112.9 \pm 2.26 \times 3.55 \Rightarrow 104.9 \rightarrow 120.9$.

One tail & two tail Tests

The two tests we have done are ~~two~~ two tail tests.
– because the critical region is made up of 2 areas in the tails of the sampling distⁿ of our test statistic.
→ This follows from the form of

$$H_1: \mu \neq 225 \quad \& \quad H_1: \mu \neq 120$$

In other testing problems, the appropriate form of the alternative hypothesis may be of the following forms:

$$H_1: \mu > \mu_0 \quad \text{or} \quad H_1: \mu < \mu_0$$

eg. Rope manufacturer : → traditional strength = 8000 lbs.
New process introduced ~~wit~~ which it is hoped will increase the strength
sample mean of 25 ropes = 8120

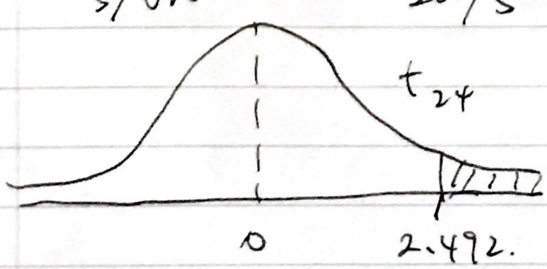
" variance " " " $s^2 = 40{,}000 \Rightarrow s = 200$

~~the~~ Use 1% level of signif., should we conclude increased strength?

$H_0: \mu = 8000$
$H_1: \mu > 8000$

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{8120 - 8000}{200/5} = \frac{120}{40} = 3 > 2.492.$$



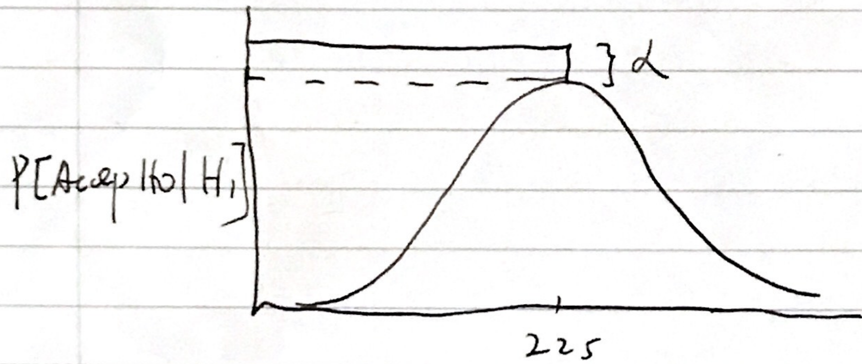
P-value = 0.003 < 0.001
⇒ Reject Ho

CI for μ : $8120 \pm 2.492 \times 40$ 8020.3 → 8219.7.

Operating Char. Curve

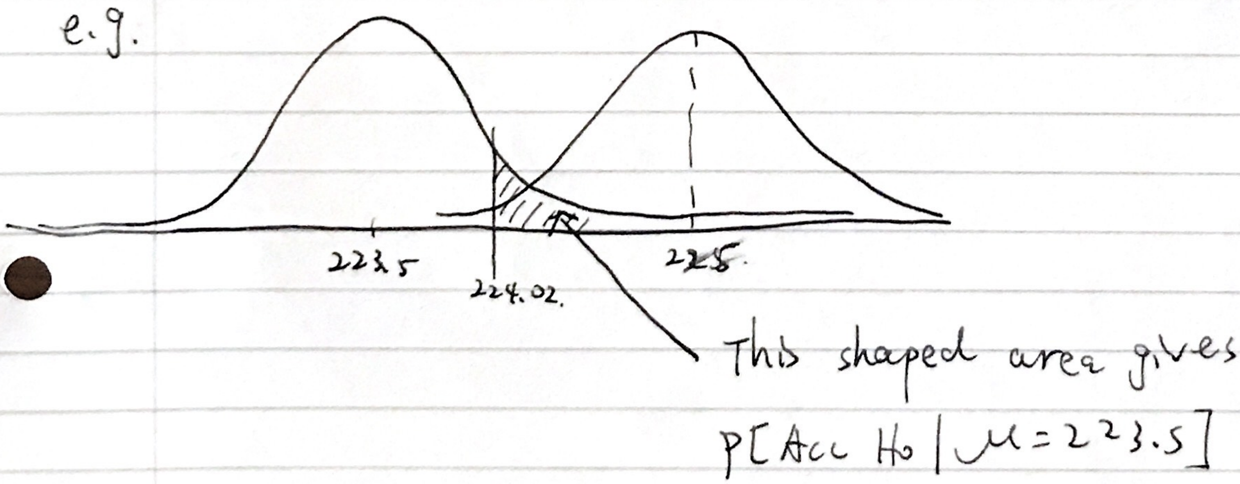
We have said little of the Type 2 error prob.
In the case of hypo: $H_0: \mu = 225$
$H_1: \mu \neq 225$.

- Suppose wish to compute P[Accept H_0 | H_1 is true]



- must be done for range of μ values within H_1.
- Usually plotted against μ as shown and referred to as OC curve.

e.g.

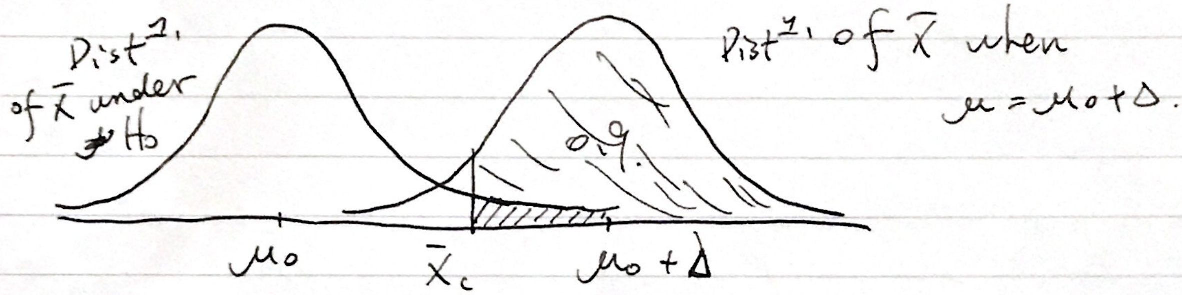


Finding the sample size required in single-sample Hypothesis testing

Testing : $H_0: \mu = \mu_0$

$H_1: \mu > \mu_0$, $\alpha = 1\%$,

suppose we want to have a 90% chance of detecting an increase above μ_0 of size Δ. What size n is needed?



Let $\mu_0 = 3000$, $\Delta = 200$.

n is required s.t. $P[\bar{X} \geq \bar{X}_c \mid \mu = 3000] = 0.01$. ①

and $P[\bar{X} \geq \bar{X}_c \mid \mu = 3200] = 0.9$. ②

From ① $\bar{X}_c = 3000 + 2.33 \frac{\sigma}{\sqrt{n}}$

② $\bar{X}_c = 3200 - 1.28 \frac{\sigma}{\sqrt{n}}$.

$\Rightarrow$ $2.33 \frac{\sigma}{\sqrt{n}} + 1.28 \frac{\sigma}{\sqrt{n}} = 200$

$3.61 \frac{\sigma}{\sqrt{n}} = 200$

Need to estimate σ, say $\sigma \simeq 800$

$\Rightarrow$ $\sqrt{n} = 3.61 \times 4 = 14.44$

$\Rightarrow$ $n = 208.51 \simeq 209$.

Note: The choice of σ, Δ, α and power level (90%) affects the value of n.

Further Single sample Tests.

Testing Population Variance.

Suppose we have a R. sample of size n from $N(\mu, \sigma^2)$
We want to test
$H_0: \sigma^2 = \sigma_0^2$, $H_A: \sigma^2 \neq \sigma_0^2$
The test statistic is $\frac{(n-1)s^2}{\sigma_0^2}$

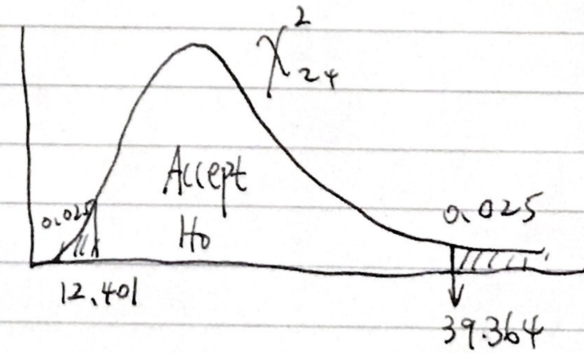
If H_0 is true, this test statistic has a known dist^n. χ^2_{n-1}

For large n, the test works well even when the pop^n. dist^n. is not Normal.

Example: A sample of 25 children (girls, aged 12) was taken, and their heights measured. The sample mean was 130.3, with $s = 3.87$. Test $\sigma = 3$ at $\alpha = 5\%$.

$H_0: \sigma = 3$, $H_1: \sigma \neq 3$.
Under H_0, $\frac{24s^2}{3^2}$ is χ^2_{24}, $n = 25$,

$= 39.94 > 39.364$,
$\Rightarrow$ Reject H_0 at 95% confidence.



Testing the value of the mean of a Poisson Dist^n.

Suppose we have a R. sample of size n from a Po(λ)
Test
$H_0: \lambda = \lambda_0$, $H_A: \lambda \neq \lambda_0$

Test stat is the sample sum $\sum X_i$

If H_0 is true, $\sum X_i \sim Poi(n\lambda_0)$.

Note: If n is large (or when $n\lambda_0$ is large) it is possible to use a normal approx.

$\sum X_i \sim Poi(n\lambda) \Rightarrow N(n\lambda, n\lambda)$

If the normal approx is used, one would use $\bar{X}$ as test stat, with dist^n (under H_0) as follows:

$$\frac{\bar{X} - \lambda_0}{\sqrt{\lambda_0 / n}} \sim N(0,1), \text{ under } H_0$$

or alternatively $\dfrac{\sum X_i - n\lambda_0}{\sqrt{n\lambda_0}} \sim N(0,1)$.

Example: In an investigation of freq. of claims by motorists, it was found that there were 960 claims for 6000 policies. Assuming that the number of claims by individual motorists has a Poisson(λ) dist^n; carry out a test (at $\alpha = 1\%$) of the null hypothesis that $\lambda = 0.17$ against $H_a: \lambda < 0.17$.

Test stat $\sum X_i$ is ~~po~~ Poisson

P-value = Prob[$\sum X_i \leq 960$ | Poisson mean $= 1020$]

where is this from?

It is necessory to use the Normal approx., need ~~cont.~~ continuity correction.

$$\begin{aligned} \text{P-value} &= P[\ \textstyle\sum X_i \leq 960.5 \mid \sum X_i \sim N(1020, 1020)] \\ &= P\left[\frac{\sum X_i - 1020}{\sqrt{1020}} \leq \frac{960.5 - 1020}{\sqrt{1020}}\right] \\ &= P[Z \leq -1.863] = 0.0312. \end{aligned}$$

Thus we could reject H_0, if we are prepared to use a significance level of 4%.

Two - Sample Problems
2 Groups of female rats placed on diets with high & low protein content.

- gain in weight between 28^{th} & 84^{th} days was measured for each rat.

High P		Low P	
134	107	70	94
146	83	118	
104	113	101	
119	129	85	
124	97	107	
161	123	132	

Is the evidence of dietary effect - obtain a 95% CI for the difference in mean weight gain between the 2 diets.

~~Grop~~ Group 1	" 2.
$\bar{X}_1 = 120$	$\bar{X}_2 = 101$
$S_1^2 = 457.45$	$S_2^2 = 425.33$

Need the following assumptions to construct CI and test hypothesis.

(1) Sample 1 & 2 are random sample indep of each other.

(2) for sample 1 $\quad X_i^{(1)} \sim N(\mu_1, \sigma^2)$

" " 2 $\quad X_i^{(2)} \sim N(\mu_2, \sigma^2)$

$\Big\}$ same σ^2.

CI for $\mu_1 - \mu_2$

Earlier assumptions $\Rightarrow$ $\bar{X}_1 \sim N(\mu, \frac{\sigma^2}{n_1})$

$\bar{X}_2 \sim N(\mu, \frac{\sigma^2}{n_2})$

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

If σ^2 were known,

95% CI : $\bar{X}_1 - \bar{X}_2 \pm 1.96\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

Since σ^2 unknown, we are led to follow the same procedure as before : estimate σ^2.

suggest $\hat{\sigma}^2 = \frac{w_1 S_1^2 + w_2 S_2^2}{w_1 + w_2}$, and the best choice of w_1, w_2 is

$$\hat{\sigma}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

Now $\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$ and $\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$

Thus $\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$.

and $\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$

Thus $\frac{[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)] / \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)\sigma}}} \sim t_{n_1+n_2-2}$.

or $\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$.

where $s=\sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$.

95% CI for $\mu_1-\mu_2$

$\bar{X}_1-\bar{X}_2 \pm t_{n_1+n_2-2,\,0.025}\, s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$= 19 \pm 2.11\sqrt{446.12}\sqrt{\frac{1}{12}+\frac{1}{7}}$

$= 19 \pm 2.11 \times \sqrt{100.9} = 19 \pm 21.2.$

$[-2.2 \to 40.2]$

2-Sample Test. 16.04.2018.

May wish to test

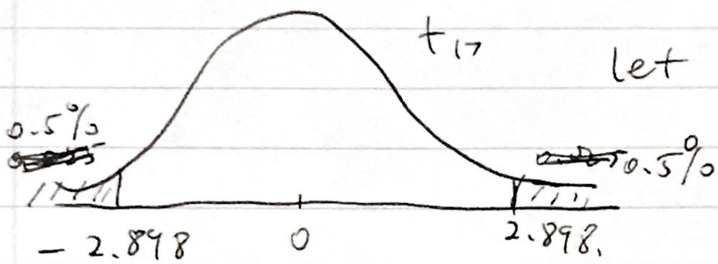
$H_0: \mu_1=\mu_2$
$H_1: \mu_1\neq\mu_2$ } to test for ~~a diff~~ evidence of a dietary effect.

test stat:

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Assuming H_0 true, $t=\dfrac{\bar{X}_1-\bar{X}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ and unusually large +ve or −ve values of t would indicate $\mu_1>\mu_2$ and $\mu_1<\mu_2$.

Dist^{n} of t (under H_0) is known. t_{17}



let $\alpha = 1\%$

$\cancel{t = \frac{19-0}{1.8}}$ $\quad t = \frac{19-0}{10.05} = 1.89.$

Can not reject H_0 at 1% sig. level.

Note: We may accept H_0 as being consistent with the data, but it should also be regarded that there is a range of values of $(\mu_1 - \mu_2)$ apart from $\mu_1 - \mu_2 = 0$ which are also consistent with data.

The extent of this range: construct the 99% CI for $\mu_1 - \mu_2$.

$\Rightarrow 19 \pm 2.898(10.05) = 19 \pm 29.1 = -10.1 \rightarrow 48.1.$

Thus any actual difference between the mean $(\mu_1 - \mu_2)$ in this range could not be detected by our test, using the 1% significance level. I.E. our data is consistent with any actual difference in this range.

Now if some of these differences are considered significant differences to ~~be~~ the experiments, then this experiment has been completely useless in detecting the presence of such differences.

Testing $\sigma_1^2 = \sigma_2^2$

Sample 1: $X_i^{(1)} \sim N(\mu, \sigma_1^2)$ size n_1

" " 2: $X_i^{(2)} \sim N(\mu, \sigma_2^2)$ " n_2.

Wish to test $\sigma_1^2 = \sigma_2^2$

Use $\dfrac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$, $\dfrac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$

Thus $\dfrac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2} \sim F_{n_1-1, n_2-1}$

Under H_0: $\sigma_1^2 = \sigma_2^2$, we see that

$$\frac{s_1^2}{s_2^2} \sim F_{n_1-1, n_2-1}$$

Thus, we can test H_0, using the test statistic $\dfrac{s_1^2}{s_2^2}$

which will be close to 1 when H_0 true.

<u>Test Procedure:</u>

Form $\dfrac{s_1^2}{s_2^2}$ so that ratio > 1

s_1^2: $n_1 - 1$ d.f.

s_2^2: $n_2 - 1$ d.f.

F dist^{n}.