Tues. Jan. 21st

Confidence Intervals Part 2

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ Both unknown.

Estimate μ with $\bar{X}_n$, want confidence interval

Have $\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}$ for large n

Based on $Z_n = \dfrac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} Z \sim N(0,1)$

What is the exact (small-sample) distribution of

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

The t distribution, William Gosset (1908) "Student"

Let $Z \sim N(0,1)$ & $Y \sim \chi^2(\nu)$ be independent

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu)$$

Density of T: Jacobian problem

$X_1 \sim N(0,1)$, $X_2 \sim \chi^2(\nu)$

$Y_1 = \frac{X_1}{\sqrt{X_2/\nu_2}}$, $Y_2 = X_2$ get

$$f_T(t \mid \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \nu}\, \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

for $-\infty < t < \infty$

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

Lots of work, directed tow. obtaining Z, & Y from a normal random sample

Background

- Functions of independent RVs are independent
- For normally distributed RVs, zero covariance implies independence. (STA302)
- If $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- If $Z \sim N(0,1)$, $Z^2 \sim \chi^2(\nu=1)$
- Linear combinations of normals are normal (STA302)
- If $Y_1, \ldots, Y_n$ ind, $Y_i \sim \chi^2(\nu_i)$, $\sum_{i=1}^{n} Y_i \sim \chi^2(\sum_{i=1}^{n} \nu_i)$
- $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$
- So $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1)$

Recall

$$Cov(X,X) = Var(X)$$

$$Cov(aX, bY) = ab\, Cov(X,Y)$$

$$Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$$

Preparing for χ^2 part

We know that if $Y_1 \sim \chi^2(r_1)$, $Y_2 \sim \chi^2(r_2)$ are independent, then $Y = Y_1 + Y_2 \sim \chi^2(r_1 + r_2)$ using MGFs

<u>Thm</u> Let Y_1 & Y_2 be independent RVs,

$Y = Y_1 + Y_2$, $Y \sim \chi^2(r_1 + r_2)$, $Y_2 \sim \chi^2(r_2)$

Then $Y_1 \sim \chi^2(r_1)$

If $X \sim \chi^2(r)$, $M_X(t) = (1-2t)^{-r/2}$

<u>Proof</u> By independence

$$M_Y(t) = M_{Y_1}(t)\, M_{Y_2}(t)$$

$$\Rightarrow (1-2t)^{-\frac{r_1+r_2}{2}} = M_{Y_1}(t)\,(1-2t)^{-r_2/2}$$

$$\Longrightarrow \frac{(1-2t)^{-\frac{r_1}{2}}\cancel{(1-2t)^{-\frac{r_2}{2}}}}{\cancel{(1-2t)^{-r_2/2}}} = (1-2t)^{-r_1/2}$$

MGF of $\chi^2(r_1)$

Working toward $Y = Y_1 + Y_2$ (5)

$X_1 \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

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

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

$$= \sum_{i=1}^{n}(X_i-\bar{X})^2 + 2(\bar{X}-\mu)\underbrace{\left[\sum_{i=1}^{n}X_i - n\bar{X}\right]}_{=0} + n(\bar{X}-\mu)^2$$

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

$$\Longrightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

$$\Longrightarrow \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

$$Y \sim \chi^2(n) = Y_1 + Y_2 \sim \chi^2(1)$$

Y_1 & Y_2 ind bec S^2 & $\bar{X}$ are ind.

$Z_i = \frac{X_i-\mu}{\sigma} \sim N(0,1)$

$Z_i^2 \sim \chi^2(1)$

$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$

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

$\cancel{Z} = \frac{X-\mu}{\sigma} \sim N(0,1)$

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

$Z^2 \sim \chi^2(1)$

So $Y_1 \sim \chi^2(n-1)$

Have $Y = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Again if $Z \sim N(0,1)$, $Y \sim \chi^2(r)$ ind

$$T = \frac{Z}{\sqrt{Y/r}} \sim t(r)$$

Set $Z = \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$

$$T = \frac{\dfrac{\sqrt{n}(\bar{X}-\mu)}{\cancel{\sigma}}}{\sqrt{\dfrac{\cancel{(n-1)}S^2}{\cancel{\sigma^2}}\Big/\cancel{(n-1)}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

$$\sim t(n-1)$$

- $T_n = \dfrac{\sqrt{n}(\bar{X}_n-\mu)}{S_n}$ special case of $\dfrac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}$

$$\xrightarrow{d} Z \sim N(0,1)$$

- Can't integrate $f_T(t/r)$, but good numerical approximations are available

Filling a hole: Why are S^2 & $\bar{X}$ independent?

$$Cov(\bar{X}_n, X_j - \bar{X}_n) = Cov(\bar{X}, X_j) - Cov(\bar{X}_n, \bar{X}_n)$$

$$= Cov\left(X_j, \frac{1}{n}\sum_{i=1}^{n} X_i\right) - Var(\bar{X}_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} Cov(X_i, X_j) - \frac{\sigma^2}{n}$$

$$= \frac{1}{n}\Big(Cov(X_j, X_j) + \sum_{i \neq j} \underbrace{Cov(X_i, X_j)}_{0}\Big) - \frac{\sigma^2}{n}$$

$$= \frac{1}{n}(\sigma^2 + 0) - \frac{\sigma^2}{n} = 0$$ Any random sample

If normal random sample

Bec linear combos of normals are normal $\bar{X}_n$ & $X_j - \bar{X}$ are independent, for $j = 1, \dots, n$

$\bar{X}_n$ is independent of all functions of $(X_1 - \bar{X}), (X_2 - \bar{X}), \dots (X_n - \bar{X})$. In particular

$$S^2 = \frac{\sum_{j=1}^{n}(X_j - \bar{X})^2}{n-1}$$ is ind of $\bar{X}_n$ ∎

Have $T = \dfrac{\sqrt{n}\,(\bar{X}-\mu)}{S} \sim t(n-1)$

a) Derive a $(1-\alpha)100\%$ confidence interval for μ.

$$1-\alpha = P\left(\cancel{t_{\alpha/2}(n-1) < T \le t_{1-}}\right)$$

$$= P\left(-t_{1-\alpha/2}(n-1) < T < t_{1-\alpha/2}(n-1)\right)$$

Pivotal Quantity

$$= P\left(-t_{1-\alpha/2} < \frac{\sqrt{n}(\bar{X}-\mu)}{S} < t_{1-\alpha/2}\right)$$

$$= P\left(-t_{1-\alpha/2}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

$$= P\left(-\bar{X}-t_{1-\alpha/2}\frac{S}{\sqrt{n}} < -\mu < -\bar{X}+t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}+t_{1-\alpha/2}\frac{S}{\sqrt{n}} > \mu > \bar{X}-t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}-t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \bar{X}+t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right)$$

$$= P\left(L < \mu < U\right)$$

or $\bar{X} \pm t_{1-\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}$

b) what is the capacity of the human stomach?
Stomachs from a sample of $n=12$ cadavers were measured, yielding $\bar{X} = 994$ ml, $S^2 = 407$

Give a 95% CI for μ.

$$\bar{X} \pm t_{1-\alpha/2}\frac{S}{\sqrt{n}} = 994 \pm 2.201\sqrt{\frac{407}{12}}$$

$$= 994 \pm 12.82 = (981.18, 1006.82)$$

c) Derive a $(1-\alpha)100\%$ CI for σ^2

Have $Y = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$$1-\alpha = P\left(\chi^2_{\alpha/2} < Y < \chi^2_{1-\alpha/2}\right)$$

$$= P\left(\chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right)$$

$$= P\left(\frac{\chi^2_{\alpha/2}}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{1-\alpha/2}}{(n-1)S^2}\right)$$

$$= P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} > \sigma^2 > \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right)$$

$$= P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right) = P(L < \sigma^2 < U)$$

Using $\left(\dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \dfrac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \right)$,

$$= \left(\frac{11 * 407}{21.92}, \frac{11*407}{3.82} \right)$$

$$= (204.24, 1171.99)$$

(Recall $s^2 = 407$)