Stat 410 Random Variable and Confidence Interval Review

Dr. D. Scott

August 23, 2005

Fundamental object is the density function:

$$X \sim f(x) = f(x_1, x_2, \ldots, x_p),$$

which encodes "structure."

Sometimes one of the variables is labeled differently:

$$(X, Y) \sim f(x, y) = f(x_1, \ldots, x_p, y).$$

Here, Y is the *dependent* or *response* variable, while $X_1, \ldots, X_p$ are the *independent* or *predictor* variables.

Regression is the estimation of the *conditional mean*

$$E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p].$$

Thus regression is an extension of the *unconditional mean*, μ, reviewed in Chapter 1:

$$\mu = \mu_x = E[X] \quad \text{where } X \sim f(x)\,.$$

The variance is $\sigma^2 = \sigma_x^2 = E[X - \mu]^2$.

Random sample $\{X_1, X_2, \ldots, X_n\}$ *i.i.d.* means..?)
(Note: Different X_i's... How can you tell?)

Properties of estimator

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

$$E[\hat{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{n\mu_x}{n} = \mu_x$$

so unbiased.

Also $\bar{X} \to \mu_x$ (consistency) since

$$var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{1}{n}\sigma^2_X \to 0 \quad \text{as } n \to \infty$$

by Chebyshev's inequality.

$$
\begin{aligned}
\sigma_{\bar{x}}^2 &= E(\bar{X} - \mu_{\bar{x}})^2 \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right]^2 \\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i \neq j}(X_i - \mu)(X_j - \mu)\right] \\
&= \frac{1}{n^2} \cdot n\,\sigma_x^2 + \frac{1}{n^2}n(n-1) \cdot 0 \cdot 0 \quad \text{Why?} \\
&= \frac{\sigma_x^2}{n}\,.
\end{aligned}
$$

Normal Results $\quad Z \sim N(0,1) \quad X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Facts:

$$\begin{aligned}
\frac{X-\mu}{\sigma} &\sim N(0,1) \qquad \text{(exactly)} \\
\bar{X} &\sim N(\mu, \frac{\sigma^2}{n}) \\
\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} &\sim N(0,1) \\
Z^2 &\sim \chi^2(1) \qquad \text{where } Z \sim N(0,1) \\
\sum_{i=1}^{n} Z_i^2 &\sim \chi^2(n)\,.
\end{aligned}$$

CLT: $\sum$ of *i.i.d. r.v.*'s $\approx$ Normal. (RVLS)

Confidence Intervals for Parameters: *pivots*

Rearrange

$$Prob(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96) = 95\%$$

to get

$$Prob(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$$

($\pm$2.576 for a 99% confidence interval)

Pivot for σ^2 : $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$

$$\sum (X_i - \bar{X})^2 = \sum \left[(X_i - \mu) - (\bar{X} - \mu)\right]^2 =$$

$$\sum (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum (X_i - \mu) + n(\bar{X} - \mu)^2$$

Now, $\sum (X_i - \mu) = n(\bar{X} - \mu)$, so

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

or, dividing by σ^2 and rearranging,

$$\sum \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 = \sum \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Again, the identity is:

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

or

$$\frac{(n-1)S^2}{\sigma^2} \quad + \quad \chi^2(1) = \quad \chi^2(n)$$

From this we can conclude that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

by counting degrees of freedom!

Since $E[\chi^2(p)] = p$ and $Var[\chi^2(p)] = 2p$,

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

or

$$E[S^2] = \sigma^2 \qquad\qquad (unbiased)$$

Now have a pivot for a C.I. for σ^2:

$$Prob\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 1 - \alpha$$

iff

$$Prob\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right),$$

where $Pr(\chi^2_{n-1} < a) = Pr(\chi^2_{n-1} > b) = \frac{\alpha}{2}$.

(note: show R code to plot $\chi^2(6)$)

Need a better pivot for μ, since do not really know σ^2 for the normal random variable.

Obvious idea is to use S^2 in place of σ^2!

$$
\begin{aligned}
S^2 &= \frac{1}{n-1}\sum(X_i - \bar{X})^2 \\
S^2(\bar{X}) &= \frac{1}{n}S^2 = \frac{1}{n(n-1)}\sum(X_i - \bar{X})^2
\end{aligned}
$$

is unbiased for $\sigma^2_{\bar{x}}$.

A century ago, "Student" (Gossett) showed

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

and

$$\frac{\bar{X}-\mu}{S(\bar{X})} \sim T_{n-1}$$

called Student's t-distribution. Can show

$$T_{n-1} = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}}$$

Thus a C.I. for μ follows:

$$Prob(-a < T_{n-1} < a) = 1 - \alpha$$

becomes

$$Prob(\mu \in \bar{X} \pm a \cdot S(\bar{X})) = 1 - \alpha$$

which should be compared to

$$Prob(\mu \in \bar{X} \pm 1.96\sigma(\bar{X})) = 95\%$$

.

Relationship to hypothesis testing.

$$H_0 : \mu = \mu_0 \qquad \text{vs.} \qquad H_1 : \mu \neq \mu_0$$

Compute $\widehat{T} = (\bar{X} - \mu_0)/S(\bar{X})$ and reject if $|\widehat{T}| > a$.

Same as checking if μ_0 is in C.I. (symmetry)!

p-value $= Prob(|T_{n-1}| > |\widehat{T}|)$ (learn)

Finally, the F distribution is a pivot for comparing variances.

$$F = \frac{\chi^2(\nu_1)/\nu_1}{\chi^2(\nu_2)/\nu_2} \sim F(\nu_1, \nu_2) = F_{\nu_1,\nu_2}$$

Recall: $$T_{n-1} = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}}$$

Thus

$$T^2_{n-1} = \frac{Z^2}{\chi^2_{n-1}/(n-1)} = \frac{\chi^2(1)}{\chi^2_{n-1}/(n-1)} = F_{1,n-1}$$