18.05 Lecture 26
April 13, 2005

Confidence intervals for parameters of Normal distribution.

Confidence intervals for μ_0, σ_0^2 in $N(\mu_0, \sigma_0^2)$
$\hat{\mu} = \overline{x}, \hat{\sigma}^2 = \overline{x^2} - (\overline{x})^2$
$\hat{\mu} \to \mu_0, \hat{\sigma}^2 \to \sigma_0^2$ with large n, but how close exactly?

You can guarantee that the mean or variance are in a particular interval with some probability:
Definition: Take $\alpha \in [0,1], \alpha-$ confidence level
If $\mathbb{P}(S_1(X_1, ..., X_n) \leq \mu_0 \leq S_2(X_1, ..., X_n)) = \alpha$,
then interval $[S_1, S_2]$ is the confidence interval for μ_0 with confidence level α.

Consider $Z_0, ..., Z_n$ - i.i.d., N(0, 1)
Definition: The distribution of $Z_1^2 + Z_2^2 + ... + Z_n^2$ is called a chi-square (χ^2) distribution, with n degrees of freedom.
As shown in §7.2, the chi-square distribution is a Gamma distribution $\to \Gamma(\frac{n}{2}, \frac{1}{2})$

Definition:

$$\text{The distribution of } \frac{Z_0}{\sqrt{\frac{1}{n}(Z_1^2 + ... + Z_n^2)}} \text{is called a t-distribution with n d.o.f.}$$

The t-distribution is also called Student's distribution, see §7.4 for detail.

To find the confidence interval for $N(\mu_0, \sigma_0^2)$, need the following:
Fact: $Z_1, ..., Z_n \sim i.i.d. N(0,1)$

$$\overline{z} = \frac{1}{n}(Z_1 + ... + Z_n), \overline{z^2} - (\overline{z})^2 = \frac{1}{n}\sum z_i^2 - (\frac{1}{n}\sum z_i)^2$$

Then, $A = \sqrt{n}\overline{z} \sim N(0,1), B = n(\overline{z^2} - (\overline{z})^2) \sim \chi^2_{n-1}$,
and A and B are independent.

Take $X_1, ..., X_n \sim N(\mu_0, \sigma_0^2), \mu_0, \sigma_0^2$ unknown.

$$Z_1 = \frac{x_1 - \mu_0}{\sigma_0}, ..., Z_n = \frac{x_n - \mu_0}{\sigma_0} \sim N(0,1)$$

$$A = \sqrt{n}\overline{z} = \sqrt{n}(\frac{\overline{x} - \mu_0}{\sigma_0})$$

$$B = n(\overline{z^2} - (\overline{z})^2) = \sqrt{n}(\frac{1}{n}\sum \frac{(x_i - \mu_0)^2}{\sigma_0^2} - (\frac{\overline{x} - \mu_0}{\sigma_0})^2) = \frac{\sqrt{n}}{\sigma_0^2}(\frac{1}{n}\sum (x_i - \mu_0)^2 - (\overline{x} - \mu_0)^2) =$$

$$= \frac{\sqrt{n}}{\sigma_0^2}(\overline{x^2} - 2\mu_0\overline{x} + \mu_0^2 - \overline{x}^2 + 2\mu_0\overline{x} - \mu_0^2) = \frac{\sqrt{n}}{\sigma_0^2}(\overline{x^2} - (\overline{x})^2)$$

To summarize:

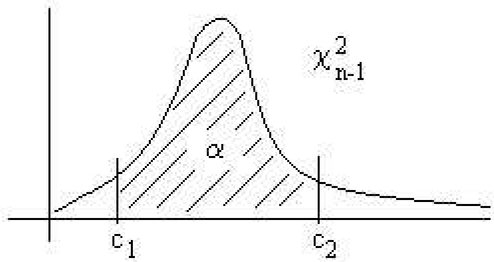
$$A = \sqrt{n}(\frac{\overline{x} - \mu_0}{\sigma_0}) \sim N(0,1); B = \frac{\sqrt{n}}{\sigma_0^2}(\overline{x^2} - (\overline{x})^2) \sim \chi^2_{n-1}$$

and, A and B are independent.

You can't compute B, because you don't know σ_0, but you know the distribution:

$$B = \frac{\sqrt{n}(\overline{x^2} - (\overline{x})^2)}{\sigma_0^2} \sim \chi^2_{n-1}$$

Choose the most likely values for B, between c_1 and c_2.



Choose the c values from the chi-square tabled values, such that area $= \alpha$ confidence.

With probability = confidence (α), $c_1 \leq B \leq c_2$

$$c_1 \leq \frac{\sqrt{n}(\overline{x^2} - (\overline{x})^2)}{\sigma_0^2} \leq c_2$$

Solve for σ_0:

$$\frac{\sqrt{n}(\overline{x^2} - (\overline{x})^2)}{c_2} \leq \sigma_0^2 \leq \frac{\sqrt{n}(\overline{x^2} - (\overline{x})^2)}{c_1}$$

Choose c_1 and c_2 such that the right tail has probability $\frac{1-\alpha}{2}$, same as left tail.
This results in throwing away the possibilities outside c_1 and c_2
Or, you could choose to make the interval as small as possible, minimize: $\frac{1}{c_1} - \frac{1}{c_2}$ given α

Why wouldn't you throw away a small interval in between c_1 and c_2, with area $1 - \alpha$?
Though it's the same area, you are throwing away very likely values for the parameter!

** End of Lecture 26