Math 3070, Applied Statistics

Section 1

November 6, 2019

Lecture Outline, 11/6

Section 7.4

- Confidence Intervals for unknown σ

Problem

If Z is a standard normal random variable, what is the pdf of Z^2?

Letting $F(x)$ be the cdf of $X = Z^2$, for $t \geq 0$,

$$F(t) = P(Z^2 \leq t) = P(|Z| \leq t^{1/2}) = 1 - 2\Phi(-t^{1/2})$$

Differentiating, we find the pdf:

$$\begin{aligned} f(t) = F'(t) &= \frac{d}{dt}(1 - 2\Phi(-t^{1/2})) = -2\phi(-t^{1/2})\frac{d}{dt}(-t^{1/2}) \\ &= t^{-1/2}\phi(t^{-1/2}) = \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2} \end{aligned}$$

Up to a constant factor, we recognize this as the pdf of a gamma random variable, $\frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}$, with $k = 1/2$ and $\lambda = 1/2$. Since both are valid pdfs, the constant factors must agree; therefore,

$$\frac{1}{\sqrt{2\pi}} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} = \frac{1}{\sqrt{2}\Gamma(1/2)} \implies \Gamma(1/2) = \sqrt{\pi}$$

Chi-squared Distribution

If Z is a standard normal random variable, then Z^2 has a so-called *chi-squared* distribution with one *degree of freedom*. It turns out that this is simply a gamma distribution with $k = \lambda = 1/2$.

The following generalization has important applications in statistical inference,:

A *chi-square* random variable with ν degrees of freedom is a gamma random variable with $k = \nu/2$ and $\lambda = 1/2$ and has pdf

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Idea

Assume $X_1, \ldots, X_n$ are i.i.d. with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$.

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

Recall: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$.
We expect $\overline{X}$ and μ to be close when $n \to \infty$.

$$\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \approx \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_1^2$$

More careful analysis shows that

$$\sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma^2}\right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

https://en.wikipedia.org/wiki/Cochran%27s_theorem#Estimation_of_variance

Confidence Interval for Variance of Normal

Suppose we want to find a confidence interval for the variance σ^2 of a normal distribution based on a random sample $X_1, \ldots, X_n$.

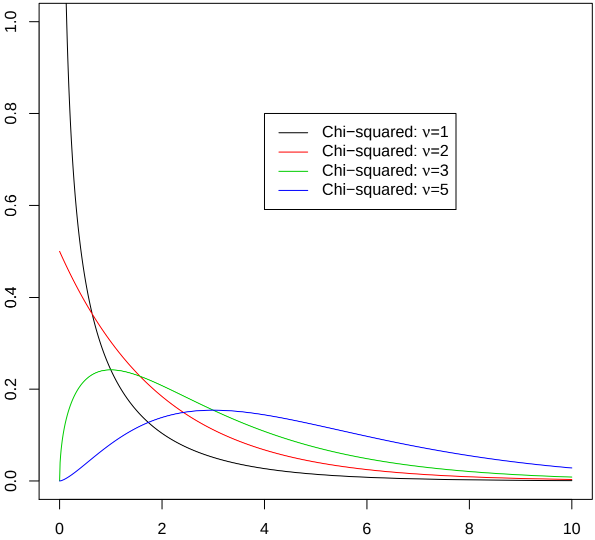
The statistic $(n-1)S^2/\sigma^2$ has a so-called χ^2 distribution with $\nu = n-1$ degrees of freedom.

Given a random sample $X_1, \ldots, X_n$ from a normal distribution with unknown mean μ and variance σ^2, A $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right]$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are critical values from a χ^2 *distribution* with $\nu = n-1$ degrees of freedom.

Chi-squared distribution



Example

Recall the breakdown voltage data, assumed to be normal:

$$1470, 1510, 1690, 1740, 1900, 2000, 2030, 2100, 2200,$$
$$2290, 2380, 2390, 2480, 2500, 2580, 2190, 2700$$

Find a 95% confidence interval for the standard deviation σ.

The observed sample mean and sample variance are $\overline{X} = 2126.5$ and $S^2 = 137324.3$. The critical values are $\chi^2_{.025,16} = 28.845$ and $\chi^2_{.975,16} = 6.908$, which gives a 95% confidence interval of

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right] = [76172.3, 318064.4]$$

for σ^2. The corresponding confidence interval for σ is

$$[\sqrt{76172.3}, \sqrt{318064.4}] = [276.0, 564.0]$$

Estimating a Proportion

Suppose we have a sequence of n Bernoulli trials, where the probability p of success is unknown. If we observe X successes, we know that the maximum likelihood estimator of p is the sample proportion $\hat{p} = X/n$.

How do we construct a confidence interval for p based on $\hat{p}$?

Suppose X is a binomial random variable counting the number of successes in n trials where each trial has probability p of success. If n is sufficiently large an approximate $100(1-\alpha)\%$ confidence interval is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p} = X/n$ is the sample proportion, and $z_{\alpha/2}$ is a critical value from the standard normal distribution.

Rule of thumb: This may be used if the number of successes X and the number of failures $n - X$ are both at least 10.

Example

A quality control team for a manufacturer tests 200 randomly selected devices, out of which 15 are defective. Assume that defective devices occur independently of one another. Find an approximate 95% confidence interval for the proportion defective.

Here the sample proportion is $\hat{p} = 15/200 = .075$, and the critical value is $z_{.025} = 1.96$, so the approximate 95% confidence interval is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .075 \pm 1.96\sqrt{\frac{.075(1-.075)}{200}}$$
$$= .075 \pm .037$$

Necessary Sample Size for Estimating Proportion

In the previous example, a 95% confidence interval for the proportion was $.075 \pm .037$. Estimate the required sample size to achieve a margin of error of .01.

Setting the margin of error equal to .01 gives an equation

$$.01 = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Solving for n gives

$$n = \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{.01^2}$$

Unfortunately, this depends on $\hat{p}$, which is unknown until *after* the new sample is taken. However, we can estimate the required n by using our previous sample proportion $\hat{p} = .075$:

$$n \approx \frac{1.96^2(.075)(1-.075)}{.01^2} = 2665.11 \approx 2666$$

Summary

Confidence interval for mean μ of normal, σ known	$\overline{X} \pm \dfrac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$
Large-sample approximate confidence interval for mean μ	$\overline{X} \pm \dfrac{z_{\alpha/2} \cdot S}{\sqrt{n}}$
Confidence interval for mean μ of normal, σ unknown	$\overline{X} \pm \dfrac{t_{\alpha/2,n-1} \cdot S}{\sqrt{n}}$
Prediction interval for normal observation	$\overline{X} \pm t_{\alpha/2,n-1} \cdot S\sqrt{1 + \dfrac{1}{n}}$
Confidence interval for variance σ^2 of normal	$\left[\dfrac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right]$
Approximate confidence interval for proportion p	$\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$