$$\operatorname{Var} \bar{X} = \operatorname{Var}\left[\frac{1}{n}\sum^{n} X_i\right] = \frac{1}{n^2} \operatorname{Var} \sum^{n} X_i =$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$\sqrt{\operatorname{Var} \bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Confidence intervals

An interval I_θ that covers θ with probability $1-\alpha$ is called a confidence interval for θ with conf. level $1-\alpha$

Let $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ be a sample from distribution $X \sim P_X$ that depends on the parameter θ and $\mathbf{X} = (X_1 \dots X_n)$ corresponding theoretical sample

Given definition means that there are functions of theoretical sample $a_1(\mathbf{X})$ and $a_2(\mathbf{X})$ so that

$$P(a_1(\mathbf{X}) \le \theta \le a_2(\mathbf{X})) = 1-\alpha$$

Then values of these RVs $a_1(\boldsymbol{x})$ and $a_2(\boldsymbol{x})$ are called lower and upper conf. limits

$$I_\theta = (a_1(\boldsymbol{x}), a_2(\boldsymbol{x}))$$

is conf. interval at the conf. limit $1-\alpha$

Requirements for conf. interval:

- as narrow as possible
- conf. level as high as possible (closer to 1)

A very wide conf. interval gives the message that there is a great deal of uncertainty concerning the value of what we are estimating

if conf. limits $a_1(\boldsymbol{x})$ and $a_2(\boldsymbol{x})$ are finite, the intervals is then called two sided, otherwise one-sided $(-\infty, a_2)$, (a_2, ∞)

Classical conf. intervals are 90%, 95%, 99%

Example

If $X \sim N(\mu, \sigma^2)$ then from our prev. RVs we are interested in such λ_α

$$P(X < \mu + \lambda_\alpha \sigma) = 1 - \alpha$$

$$P(X < \mu + \lambda_\alpha \sigma) = P\left(\frac{X-\mu}{\sigma} < \lambda_\alpha\right) = \Phi(\lambda_\alpha) =$$
$$= 1 - \alpha$$

So λ_α is the $(1-\alpha)$-quantile of $N(0,1)$ and is also called the complement α-quantile

Same logic applies to:

$$P(\mu - \lambda_{\alpha/2} \sigma < X < \mu + \lambda_{\alpha/2} \sigma) = 1 - \alpha$$

Example 2

$X \sim N(\mu, \sigma^2)$

$P(\mu - \lambda_{\alpha/2}\sigma < X < \mu + \lambda_{\alpha/2}\sigma) = 1 - \alpha$

where $\lambda_{\alpha/2}$ is $N(0,1)$ complement $\alpha/2$ quantile

Remarks

Example 2

Single Random Sample. Conf. interval for the Mean

Let $x_1, \dots, x_n$ be a random sample from $N(\mu, \sigma^2)$. We want to construct a conf. interval for the mean μ

Theorem

Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where μ is unknown. Then:

$$I_\mu = \bar{X} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$

$$I_\mu = \bar{X} \pm t_{\alpha/2, f} \frac{S}{\sqrt{n}} \quad \text{if } \sigma \text{ unknown}$$

where S is stand. deviation of the sample, $\lambda_{\alpha/2}$ and $t_{\alpha/2,f}$ are $\alpha/2$ complement quantiles of $N(0,1)$ and $t(f)$, $f = n-1$

Chi-square and t-distribution

A random variable X with a dencity function of the form

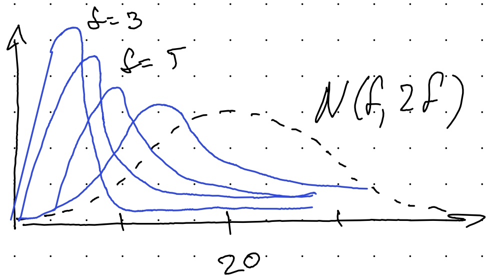
$$k\, x^{f/2 - 1} e^{-x/2} \qquad (x \geq 0)$$

is said to have a χ^2 distr. with f degrees of freedom and k being a normalizing constant

By setting $f = 1, 2 \ldots$ we obtain a whole family of distributions.
All distributions in the family are skew, but the larger is f the more symmetr. is the distribution
if $X \sim \chi^2(f)$ then $EX = f$ and $\text{Var}\, X = 2f$

Remark

if f is large, then $\chi^2(f) \approx N(f, 2f)$



Chi-square distribution

if the random variables $X_1, .. X_n$ are independent and have distributions $\chi^2(f_1), \chi^2(f_2), .., \chi^2(f_n)$ respectively then

$$Y = \sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} f_i\right)$$

Theorem

If X_i $i=1 ... n$ are independent random var-s and $X_i \sim N(0,1) \Rightarrow$

$$\sum^{n} X_i^2 \sim \chi^2(n)$$

and

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

$$\bar{X} = \frac{1}{n} \sum^{n} X_i$$

Student's t-distribution

Def a random var X with dencity function

$$k\left(1+\frac{x^2}{f}\right)^{-(f+1)/2}, \quad (-\infty < x < \infty)$$

where

$$k = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\sqrt{f\pi}\,\Gamma\left(\frac{f}{2}\right)}, \quad f \in \{1, 2 \dots\}$$

is said to have a t-distrib. with f degree of freedom

Property

If $X \sim N(0,1)$ and $Y \sim \chi^2(f)$

Where X and Y are independent

$$Z = \sqrt{f}\,\frac{X}{\sqrt{Y}} \sim t(f)$$

Theorem

if $X_i \sim N(\mu, \sigma^2)$ are independent random var-s $i = 1, 2, \dots, n$ then

$$\sqrt{n}\,\frac{\bar{X} - \mu}{S} \sim t(n-1)$$

where $\bar{X} = \frac{1}{n}\sum^{n} x_i$

$$S = \left[\frac{1}{n-1}\sum^{n}(X_i - \bar{X})^2\right]^{1/2}$$

Conf. Interval for the Mean

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

in other words

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

Two samples. Conf. Interval for difference between mean

The following model is often employed in practice: Two indep. samples have been collected

$x_1, \dots x_n$ from $N(\mu_1, \sigma_1^2)$

$y_1, \dots, y_n$ from $N(\mu_2, \sigma_2^2)$

Theorem

Let $x_1, \dots, x_{n_1}$ and $y_1, \dots, y_{n_2}$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ $\Rightarrow$

$$I_{\mu_1-\mu_2} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

if σ_1, σ_2 are known

$$I_{\mu_1-\mu_2} = \bar{x} - \bar{y} \pm t_{\alpha/2, f}\, S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

if $\sigma_1 = \sigma_2 = \sigma$ are unknown

are two-sided conf. intervals for $\mu_1 - \mu_2$ with conf. level $1-\alpha$ where $S^2 = \dfrac{\sum^{n_1} (x_i - \bar{x})^2 + \sum^{n_2} (y_i - \bar{y})^2}{n_1 - n_2 - 2}$

and $f = n_1 + n_2 - 2$

Paired Samples

Dependent (paired) samples - repeated measurements on the same object (for example before and after using some treatment medicament etc)

Before: $X_1, \dots, X_n$: $x_i \leftarrow X_i \sim N(\mu_i, \sigma_1^2)$

After: $Y_1, \dots, y_n$; $y_i \leftarrow Y_i \sim N(\mu_i + \Delta, \sigma_2^2)$

derive a new variable $W_i =$

$= Y_i - X_i \sim N(\Delta, \sigma_w^2)$ where

$\sigma_w^2 = \sigma_1^2 + \sigma_2^2$

Theorem (from previous section) is used (where both of the parameters are unknown):

$$I_\mu = \overline{X} \pm t_{\alpha/2, f} \frac{S}{\sqrt{n}} \quad \text{if } \sigma \text{ unknown}$$

Theorem