$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\,\mathrm{Var}(X_1)}{n^2} = \frac{\sigma^2}{n}$$

Recall $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$

This is called Markov's inequality. Now consider a random variable X with finite mean μ and variance σ^2. Setting $Y = (X - \mu)^2$ and $c = \varepsilon^2$ in Markov's inequality yields

$$\mathbb{P}(Y \geqslant c) \leqslant \mathbb{E}(Y)/c$$

$$\mathbb{P}(|X - \mu| \geqslant \varepsilon) = \mathbb{P}\left((X - \mu)^2 \geqslant \varepsilon^2\right) \leqslant \sigma^2/\varepsilon^2$$

This is called Chebyshev's inequality. We can use Chebyshev's inequality to show consistency of estimators. For example, let $\bar{X}$ be the estimator corresponding to the sample mean from a simple random sample of size n. Then

$\mathbb{E}[\bar{X}] = \mu$

$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$

$$\mathbb{P}(|\bar{X} - \mu| \geqslant \varepsilon)) \leqslant \frac{\mathrm{Var}(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

From this inequality we can see that $\bar{X}$ is a $\boxed{\text{Consistent}}$ estimator of μ. In other words, the average of a large number of independent and identically distributed random variables tends to the expected value as the sample size goes to infinity. This result is known as the **Law of Large Numbers**.

Central limit theorem

If $X_1, \ldots, X_n$ is a simple random sample from a $\mathsf{Normal}(\mu, \sigma^2)$ distribution, then

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

$$\mathbb{E}\left(\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}\right) = \frac{\mathbb{E}(\bar{X}) - \mu}{\sqrt{\sigma^2/n}} = \frac{\mu - \mu}{\sqrt{\sigma^2/n}} = 0$$

$$\mathrm{Var}\left(\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}\right) = \frac{1}{\sigma^2/n}\cdot\mathrm{Var}(\bar{X}) = \frac{1}{\sigma^2/n}\cdot\frac{\sigma^2}{n} = 1$$

Suppose now we have a simple random sample $X_1, \ldots, X_n$ of size n, where $\mathbb{E}\,X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, but the distribution of the X_i is not necessarily normal. It is one of the remarkable results of probability and statistics that $\bar{X}$ has approximately a normal distribution with mean μ and variance σ^2/n. More precisely, for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leqslant x\right) = \Phi(x),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This result is called the **Central Limit Theorem**.

As a sketch of the ideas involved, suppose that $X_1, \ldots, X_n$ form a simple random sample with $\mathbb{E}\,X_i = 0$ and $\mathrm{Var}(X_i) = 1$. The general result follows by considering random variables $Y_i := \mu + \sigma X_i$. If we can show that the moment generating function $M_{Z_n}(t)$ of $Z_n := n^{-1/2}\sum_{i=1}^{n} X_i$ converges to $\exp(t^2/2)$ (the MGF of the standard normal distribution) for all t in some neighbourhood of 0, then it follows (*from Lévy's continuity theorem — a result we will not study*) that the distribution of Z_n converges to a standard normal distribution.

Let $M_X(t)$ denote the moment generating function of the X_i. The moment generating function of Z_n is if X and Y independent, then $\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$

$$
\begin{aligned}
M_{Z_n}(t) &= \mathbb{E}[e^{tZ_n}] = \mathbb{E}[e^{tn^{-1/2}\sum_{i=1}^n X_i}] \\
&= \mathbb{E}[e^{tn^{-1/2}X_1 + tn^{-1/2}X_2 + \cdots + tn^{-1/2}X_n}] \\
&= \mathbb{E}[e^{tn^{-1/2}X_1}] \times \cdots \times \mathbb{E}[e^{tn^{-1/2}X_n}] \quad \text{[by independence]} \\
&= M_{X_1}(tn^{-1/2}) \times \cdots \times M_{X_n}(tn^{-1/2}) = \left(M_X(tn^{-1/2})\right)^n
\end{aligned}
$$

As the moment generating function of the standard normal distribution is $\exp(t^2/2)$, we would like to show that

ln natural logarithm — log.

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \lim_{n\to\infty} n \ln M_X(tn^{-1/2}) = \frac{t^2}{2}$$

for all t is some neighbourhood of 0. Let $y = n^{-1/2}$. Then we can write the limit as

$$\lim_{y\to 0} \frac{\ln M_X(yt)}{y^2}.$$

$\lim_{y\to 0} \log M_X(yt) = 0$

We need to use L'Hopital's rule to evaluate this limit as $\lim_{y\to 0} M_X(yt) = 1$ for any t in a neighbourhood of 0. Applying L'Hoptial's rule once gives

$\frac{d}{dy} \log M_X(yt) = \frac{1}{M_X(yt)} \cdot t M_X'(yt)$

$$\lim_{y\to 0} \frac{\ln M_X(yt)}{y^2} = \lim_{y\to 0}\left(\frac{t M_X'(yt)}{M_X(yt)} \Big/ 2y\right) = \lim_{y\to 0} \frac{t M_X'(yt)}{2y M_X(yt)}$$

This is still of indeterminate form since $M_X'(0) = \mathbb{E}X = 0$. Applying L'Hopital's rule again gives

$$\lim_{y\to 0} \frac{\ln M_X(yt)}{y^2} = \lim_{y\to 0} \frac{t^2 M_X''(yt)}{2(1\cdot M_X(yt) + y\cdot t M_X'(yt))} = \frac{t^2}{2}$$

as $M_X''(0) = \mathbb{E}(X^2) = 1$.

$=$ (variance since $\mathbb{E}X = 0$)

Example. Let $X \sim \mathsf{Binomial}(n, p)$. As $X = X_1 + X_2 + \cdots X_n$, where the X_i are independent $\mathsf{Bernoulli}(p)$ random variables, we have

$\mathbb{E}(X_i) = p$

$Var(X_i) = p(1-p)$

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{X/n - p}{\sqrt{p(1-p)/n}} \sim \mathsf{Normal}(0,1) \quad \text{(approximately)}.$$

Example. Let $X_1, \ldots, X_n$ be a simple random sample from a $\mathsf{Poisson}(\lambda)$ distribution. Define $Y = \sum_{i=1}^n X_i$. We have seen that $Y \sim \mathsf{Poisson}(n\lambda)$. The central limit theorem implies that

$\mathbb{E}(Y) = n\lambda$

$Var(Y) = n\lambda$

$$\frac{Y/n - \lambda}{\sqrt{\lambda/n}} \sim \mathsf{Normal}(0,1) \quad \text{(approximately)}.$$

$$= \frac{Y - n\lambda}{\sqrt{n\lambda}}$$

Confidence intervals

How can we gauge the accuracy of an estimator of θ? *Confidence intervals* (sometimes called *interval estimates*) provide a precise way of describing the uncertainty in an estimator.

generally θ

Ou aim is to construct random variables T_1 and T_2 so that the probability of μ being in the interval (T_1, T_2) is sufficiently high. For example, we might want to construct T_1 and T_2 (with $T_1 < T_2$) that ensure that the probability of the mean μ being in (T_1, T_2) is 95%.

Formally, given random variables $X_1, \ldots, X_n$ whose joint distribution depends on some unknown $\theta \in \Theta$, a **$(1-\alpha)$ stochastic confidence interval** is a pair of statistics

$$T_1(X_1, \ldots, X_n) \quad \text{and} \quad T_2(X_1, \ldots, X_n)$$

functions of the sample

with the property that

usually =

$$\mathbb{P}(T_1 < \theta < T_2) \geqslant 1 - \alpha, \quad \text{for all } \theta \in \Theta,$$

where the number $1 - \alpha \in [0, 1]$ is the *coverage probability.*

That is, (T_1, T_2) is a *random* interval, based only on the (as yet to be observed) outcomes $X_1, \ldots, X_n$, that contains the unknown θ with probability at least $1 - \alpha$.

A realisation of the random interval, say (t_1, t_2), is called a **$(1-\alpha)$ numerical confidence interval** for θ.

Remark: Whilst *stochastic* confidence intervals contain the unknown θ with probability at least $1 - \alpha$, their *numerical* counterparts either contain θ or they do not. It may be helpful to think of a Bernoulli analogy, where "success" occurs with probability (at least) $1 - \alpha$ — then outcomes are either "successes" or "failures".

Consider a simple random sample of size n from a $\mathsf{Normal}(\mu, \sigma^2)$ distribution. Suppose we know σ^2 and we would like to construct a confidence interval for the unknown parameter μ. We know that

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathsf{Normal}(0, 1).$$

The quantity on the left is often called a *pivot*, because we know its distribution (which does not depend on the unknown parameter of interest) and it contains both a statistic and the unknown parameter of interest.

Hence, e.g. $\mathbb{P}\left(-1.96 \leq \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \leq 1.96\right) = 0.95$

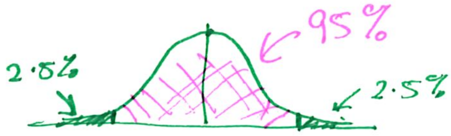
$$\boxed{\mathbb{P}\left(z_{\alpha/2} \leqslant \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leqslant z_{1-\alpha/2}\right) = 1 - \alpha,}$$

where z_γ is the γ-quantile of the standard normal distribution. For example, a standard normal random variable is contained in the interval $(-1.96, 1.96)$ with probability 0.95.

Rearranging, we have

$$\boxed{\mathbb{P}\left(\bar{X} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha.}$$

$$1 - \alpha = \mathbb{P}\left(z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$
$$= \mathbb{P}\left(-\bar{X} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq -\mu \leq -\bar{X} + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$
$$= \mathbb{P}\left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \geq \mu \geq \bar{X} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$



As the standard normal distribution is symmetric about 0, the quantiles satisfy $-z_{\alpha/2} = z_{1-\alpha/2}$.

Hence a stochastic $1-\alpha$ confidence interval for μ in this case is

$$\boxed{\left(\bar{X} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\ , \bar{X} + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\ \right),}$$

which is often *abbreviated* to

$$\boxed{\bar{X} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}}$$

So for example, in 95% of the simple random samples from $\mathsf{Normal}(\mu, \sigma^2)$, μ will be within $1.96 \times \sigma/\sqrt{n}$ of $\bar{X}$.

Exercise: Suppose that we wish to determine the average time it takes to write a 2 Gb file to a hard-drive we are testing, as well as to quantify the uncertainty inherent in the estimate. We will assume that write times are Normally distributed, with unknown mean μ but known standard deviation $\sigma = 1$ s. We have the following data:

$$7.2\ s, 8.3\ s, 7.8\ s, 8.1\ s, 7.5\ s\,.$$

Construct a numerical (numerical) 95% confidence interval for the unknown mean.

We calculate

$$\bar{x} = \frac{1}{5}(7.2 + 8.3 + 7.8 + 8.1 + 7.5) = 7.78\,s$$

From the tabulated values of the standard normal cdf,

MATLAB
norminv
R
qnorm

$$z_{0.975} = 1.96 \quad , \text{ so } \quad z_{0.975} \times \sqrt{\frac{\sigma^2}{n}} = 1.96 \times \sqrt{\frac{1}{5}} \simeq 0.88\,s\,.$$

The (numerical) 95% confidence interval is

$$7.78 \pm 0.88\,s \quad \text{or} \quad (6.90\,s\,, 8.66\,s)$$

This is great. We were able to say something about an unknown parameter μ based on our sample. Unfortunately, this is practically useless since there is no reason why we would know what σ^2 is.

Impact of Unknown Variance

For a random sample from a normal distribution with known variance σ^2, we have seen that the estimator corresponding to the *sample mean* $\bar{X}$ is normally distributed. From this we can construct a confidence interval for the unknown mean μ. How can we proceed when σ^2 is unknown?

It is natural to consider replacing σ^2 by the unbiased estimator of σ^2 given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\,.$$