Statistics

MATH2089

Semester 1, 2018 – Lecture 7

This lecture

7. Inference concerning a mean

Additional reading: Sections 5.6, 7.1 and 7.2 in the textbook

Introduction

The purpose of most **statistical inference** procedures is to generalise the information contained in an observed random sample to the population from which the sample were obtained.

This can be divided into two major areas:

- **estimation**, including point estimation and interval estimation
- **tests of hypotheses**

In this chapter we will present some theory and largely illustrate it with some methods which pertain to the estimation of means.

Point estimation

Recall that we wish to estimate an unknown parameter θ of a population. For instance, the population mean μ, from a random sample of size n, say $X_1, X_2, \ldots, X_n$.

To do so, we select an estimator, which must be a statistic (i.e. a value computable from the sample), say $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$

$\rightarrow$ **An estimator is a random variable**, which has its mean, its variance and its probability distribution, known as the sampling distribution

For instance, to estimate the population mean μ, we suggested in the previous chapter to use the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

We derived that

$$\mathbb{E}(\bar{X}) = \mu \qquad \text{and } \mathbb{V}\mathrm{ar}(\bar{X}) = \frac{\sigma^2}{n},$$

where σ^2 is the population variance

Specifically, if $X_i \sim \mathcal{N}(\mu, \sigma)$ for all i, we showed $\boxed{\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$

Properties of estimators

The choice of the sample mean to estimate the population mean seems quite natural. However, there are many other estimators that can be used to calculate an estimate. Why not:

- $\hat{\Theta}_1 = X_1$, the first observed value;
- $\hat{\Theta}_2 = (X_1 + X_n)/2$;
- $\hat{\Theta}_3 = (aX_1 + bX_n)/(a + b)$, for two constants a, b $(a + b \neq 0)$

$\rightarrow$ criteria for selecting the 'best' estimator are needed

What do we expect from an estimator for θ ?

$\rightarrow$ certainly that it should give estimates reasonably close to θ, the parameter it is supposed to estimate

However, this 'closeness' is not easy to comprehend: first, θ is unknown, and second, the estimator is random.

$\rightarrow$ we have to properly define what "close" means in this situation

Properties of estimators: unbiasedness

The first desirable property that a good estimator should possess is that it is **unbiased**.

Definition

An estimator $\hat{\Theta}$ of θ is said to be unbiased if and only if its mean is equal to θ, whatever the value of θ, i.e.

$$\mathbb{E}(\hat{\Theta}) = \theta$$

$\rightarrow$ an estimator is unbiased if "on the average" its values will equal the parameter it is supposed to estimate

If an estimator is not unbiased, then the difference

$$\mathbb{E}(\hat{\Theta}) - \theta$$

is called the **bias** of the estimator $\quad \rightarrow$ systematic error

For instance, we showed that $\mathbb{E}(\bar{X}) = \mu$

$\rightarrow$ the sample mean $\bar{X}$ is an unbiased estimator for μ

Properties of estimators: unbiasedness

The property of unbiasedness is one of the most desirable properties of an estimator, although it is sometimes outweighted by other factors.

One shortcoming is that it will generally not provide a unique estimator for a given estimation problem.

For instance, for the above defined estimators for μ,

$$\mathbb{E}(\hat{\Theta}_1) = \mathbb{E}(X_1) = \mu$$

$$\mathbb{E}(\hat{\Theta}_2) = \mathbb{E}\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2}(\mathbb{E}(X_1) + \mathbb{E}(X_n)) = \frac{1}{2}(\mu + \mu) = \mu$$

$$\mathbb{E}(\hat{\Theta}_3) = \mathbb{E}\left(\frac{aX_1 + bX_n}{a+b}\right) = \frac{1}{a+b}(a\mathbb{E}(X_1) + b\mathbb{E}(X_n)) = \frac{1}{a+b}(a\mu + b\mu) = \mu$$

$\rightarrow$ $\hat{\Theta}_1$, $\hat{\Theta}_2$ and $\hat{\Theta}_3$ are also unbiased estimators for μ

$\rightarrow$ we need a further criterion for deciding which of several **unbiased** estimators is best for estimating a given parameter

Properties of estimators: efficiency

That further criterion becomes evident when we compare the variances of $\bar{X}$ and $\hat{\Theta}_1$.

We have shown that $\mathbb{V}\mathrm{ar}(\bar{X}) = \frac{\sigma^2}{n}$, while we have $\mathbb{V}\mathrm{ar}(\hat{\Theta}_1) = \mathbb{V}\mathrm{ar}(X_1) = \sigma^2$

$\rightarrow$ the variance of $\bar{X}$ is n times smaller than the variance of $\hat{\Theta}_1$!

$\rightarrow$ it is far more likely that $\bar{X}$ will be closer to its mean, μ, than $\hat{\Theta}_1$ is to μ

Fact

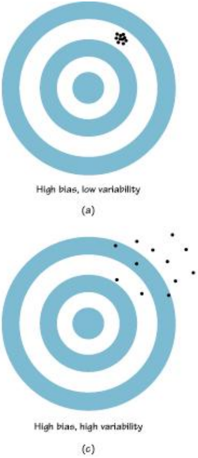
Estimators with smaller variances are more likely to produce estimates close to the true value θ.

$\rightarrow$ a logical principle of estimation is to choose the unbiased estimator that has minimum variance

Such an estimator is said to be **efficient** among the unbiased estimators.

Properties of estimators

A useful analogy is to think of each value taken by an estimator as a shot at a target, the target being the population parameter of interest



Properties of estimators: consistency

Consider the 'minimum variance' argument as the sample size increases:

We desire an estimator that is more and more likely to be close to θ as the number of observations increases.

Namely, we require that the probability that the estimator is 'close' to θ increases to one as the sample size increases.

Such estimators are called **consistent**.

An easy way to check that an unbiased estimator is consistent is to show that its variance decreases to 0 as n increases to ∞.

For instance, $\mathbb{V}\mathrm{ar}(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty \to$ **$\bar{X}$ is consistent for μ**

On the other hand, it can be verified that

$$\mathbb{V}\mathrm{ar}(\hat{\Theta}_1) = \sigma^2 \nrightarrow 0, \quad \mathbb{V}\mathrm{ar}(\hat{\Theta}_2) = \frac{\sigma^2}{2} \nrightarrow 0, \quad \mathbb{V}\mathrm{ar}(\hat{\Theta}_3) = \sigma^2 \frac{a^2 + b^2}{(a+b)^2} \nrightarrow 0$$

$\to$ none of them are consistent

Sample mean

We have seen thus far that the sample mean $\bar{X}$ is unbiased and consistent as an estimator of the population mean μ.

It can be also shown that in most practical situations where we estimate the population mean μ, the variance of no other estimator is less than the variance of the sample mean.

Fact

In most practical situations, the sample mean is a very good estimator for the population mean μ.

Note: there exist several other criteria for assessing the goodness of point estimation methods, but we shall not discuss them in this course

$\Rightarrow$ Here, we will always use the sample mean $\bar{X}$ when we will have to estimate the population mean μ.

Standard error of a point estimate

Although we estimate the population parameter θ with an estimator that we know to have certain desirable properties (unbiasedness, consistency), the chances are slim, virtually non existent, that the estimate will actually equal θ.

$\rightarrow$ **an estimate remains an approximation of the true value!**

$\rightarrow$it is unappealing to report your estimate only, as there is nothing inherent in $\hat{\theta}$ that provides any information about how close it is to θ

Hence, it is usually desirable to give some idea of the precision of the estimation $\rightarrow$ the measure of precision usually employed is the standard error of the estimator.

Definition

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation $\text{sd}(\hat{\Theta})$.

Note: If the standard error involves some unknown parameters that can be estimated, substitution of those values into $\text{sd}(\hat{\Theta})$ produces an estimated standard error, denoted $\widehat{\text{sd}(\hat{\Theta})}$

Standard error of the sample mean

Suppose again that we estimate the mean μ of a population with the sample mean $\bar{X}$ calculated from a random sample of size n.

We know that $\mathbb{E}(\bar{X}) = \mu$ and $\mathbb{V}\mathrm{ar}(\bar{X}) = \frac{\sigma^2}{n}$, so the standard error of $\bar{X}$ as an estimator of μ is

$$\mathrm{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}},$$

However, we cannot report a numerical value for this standard error as it depends in turn on the population standard deviation σ, which is usually unknown.

$\rightarrow$ we have a natural estimate of the population standard deviation given by the observed sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$\rightarrow$ estimate the standard error $\mathrm{sd}(\bar{X})$ with $\widehat{\mathrm{sd}(\bar{X})} = \frac{s}{\sqrt{n}}$

Note: we will study the sample standard deviation S as an estimator of σ in more detail later (Chapter 9)

Interval estimation: introduction

When the estimator is normally distributed, we can be 'reasonably confident' that the true value of the parameter lies within two standard errors of the estimate (recall the 68-95-99-rule)

$\rightarrow$ it is often easy to determine an interval of plausible values for a parameter

$\rightarrow$ such observations are the basis of **interval estimation**

$\rightarrow$instead of giving a point estimate $\hat{\theta}$, that is a single value that we know not to be equal to θ anyway, we give an interval in which we are very confident to find the true value,

and we specify what "very confident" means

Basic interval estimation: example

Example

An article in the *Journal of Heat Transfer* described a new method of measuring thermal conductivity of Armco iron. At 100°F and a power input of 550 Watts, the following measurements of thermal conductivity (in BTU/hr-ft-°F) were obtained:

`41.60, 41.48, 42.34, 41.95, 41.86, 42.18, 41.72, 42.26, 41.81, 42.04`

A point estimate of the mean thermal conductivity μ (at 100°F and 550 Watts) is the sample mean

$$\bar{x} = \frac{1}{10}(41.60 + 41.48 + \ldots + 42.04) = 41.924 \text{ BTU/hr-ft-°F}$$

We know that the standard error of the sample mean as an estimator for μ is $\text{sd}(\bar{X}) = \sigma/\sqrt{n}$, and since σ is unknown, we may replace it by the sample standard deviation $s = \ldots = 0.284$ (BTU/hr-ft-°F) to obtain the estimated standard error

$$\widehat{\text{sd}(\bar{X})} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898 \text{ BTU/hr-ft-°F}$$

Basic interval estimation: example

$\rightarrow$ the standard error is about 0.2 percent of the sample mean

$\rightarrow$ we have a relatively precise point estimate of the 'true' mean thermal conductivity μ under those conditions

If we can assume that thermal conductivity is normally distributed (can we?), then 2 times the (estimated) standard error is

$$2\widehat{\text{sd}(\bar{X})} = 2 \times 0.0898 = 0.1796,$$

and we are thus 'highly confident' that the true mean thermal conductivity is within the interval

$$[41.924 \pm 0.1796] = [41.744, 42.104]$$

$\rightarrow$ the term 'highly confident' obviously needs to be quantified.

Confidence intervals

The preceding interval is called a confidence interval.

Definition

A **confidence interval** is an interval for which we can assert with a reasonable degree of certainty (or confidence) that it will contain the true value of the population parameter under consideration.

A confidence interval is always calculated by first selecting a confidence level, which measures its degree of reliability
$\rightarrow$ a confidence interval of level $100 \times (1 - \alpha)\%$ means that we are $100 \times (1 - \alpha)\%$ confident that the true value of the parameter is included into the interval (α is a real number in $[0, 1]$)

The most frequently used confidence levels are 90%, 95% and 99%

$\rightarrow$the higher the confidence level, the more strongly we believe that the value of the parameter being estimated lies within the interval

Confidence intervals: remarks

Remark 1: information about the precision of estimation is conveyed by the length of the interval: a short interval implies precise estimation, a wide interval, however, gives the message that there is a great deal of uncertainty concerning the parameter that we are estimating.

Note that the higher the level of the interval, the wider it must be!

Remark 2 : it is sometimes tempting to interpret a $100 \times (1 - \alpha)\%$ confidence interval for θ as saying that there is a $100 \times (1 - \alpha)\%$ probability that θ belongs to it

$\rightarrow$ this is *not really true*!

Fact

The $100 \times (1 - \alpha)\%$ refers to the percentage of all samples of the same size possibly drawn from the same population which would produce an interval containing the true θ.

Confidence intervals: remarks

Remark 2: (ctd.)

$\rightarrow$ if we consider taking sample after sample from the population and use each sample separately to compute $100 \times (1-\alpha)$% confidence intervals, then in the long-run roughly $100 \times (1-\alpha)$% of these intervals will capture θ

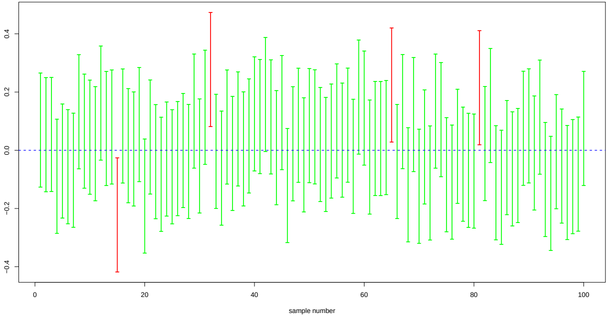
A correct probabilistic interpretation lies in the realisation that a confidence interval is a random interval, because its end-points are calculated from a random sample and are therefore random variables.

However, once the confidence interval has been computed, the true value either belongs to it or does not belong to it, and any probability statement is pointless.

$\rightarrow$ that is why we use the term “confidence level” instead of “probability”

Confidence intervals: remarks

As an illustration, we successively computed 95%-confidence intervals for μ for 100 random samples of size 100 independently drawn from a $\mathcal{N}(0, 1)$ population



$\rightarrow$ 96 intervals out of 100 ($\simeq$ 95%) contain the true value $\mu = 0$

Of course in practice we do not know the true value of μ, and we cannot tell whether the interval we have computed is one of the 'good' 95% intervals or one of the 'bad' 5% intervals.

Confidence interval on the mean of a normal distribution, variance known

The basic ideas for building confidence intervals are most easily understood by first considering a simple situation:

Suppose we have a normal population with unknown mean μ and known variance σ^2.

Note that this is somewhat unrealistic, as typically both the mean and the variance are unknown.

$\rightarrow$ we will address more general situations later

We have thus a random sample $X_1, X_2, \ldots, X_n$, such that, for all i,

$$X_i \sim \mathcal{N}(\mu, \sigma),$$

with μ unknown and σ a known constant

$\rightarrow$ we would like a confidence interval for μ

Confidence interval on the mean of a normal distribution, variance known

In that situation, we know that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

We may standardise this normally distributed random variable:

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

Suppose we desire a confidence interval for μ of level $100 \times (1 - \alpha)$%.

From our random sample, this can be regarded as a 'random interval', say $[L, U]$, where L and U are statistics (i.e. computable from the sample) such that

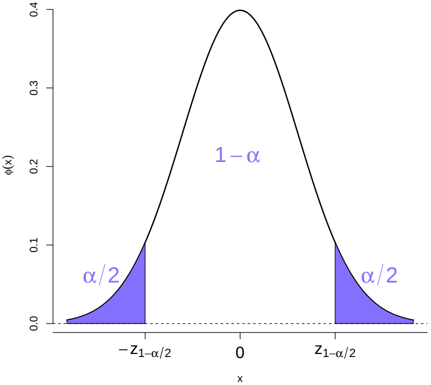
$$\mathbb{P}(L \leq \mu \leq U) = 1 - \alpha$$

Confidence interval on the mean of a normal distribution, variance known

In our situation, because $Z \sim \mathcal{N}(0,1)$, we may write

$$\mathbb{P}(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$

where $z_{1-\alpha/2}$ is the quantile of level $1 - \alpha/2$ of the standard normal distribution



Hence it is the case that

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

Confidence interval on the mean of a normal distribution, variance known

Isolating μ, it follows

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

$\rightarrow$ here are L and U, two statistics such that

$$\mathbb{P}(L \leq \mu \leq U) = 1 - \alpha$$

$\rightarrow$ L and U will yield the bounds of the confidence interval!

$\rightarrow$if $\bar{x}$ is the sample mean of an observed random sample of size n from a normal distribution with known variance σ^2, a confidence interval of level $100 \times (1 - \alpha)\%$ for μ is given by

$$\boxed{\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Confidence interval on the mean of a normal distribution, variance known

From Slide 19, Lecture 6, $z_{0.95} = 1.645$, $z_{0.975} = 1.96$ and $z_{0.995} = 2.575$

$\rightarrow$a confidence interval for μ of level 90% is

$$\left[\bar{x} - 1.645 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.645 \times \frac{\sigma}{\sqrt{n}}\right]$$

$\rightarrow$a confidence interval for μ of level 95% is

$$\left[\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right]$$

$\rightarrow$a confidence interval for μ of level 99% is

$$\left[\bar{x} - 2.575 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 2.575 \times \frac{\sigma}{\sqrt{n}}\right]$$

We see that the respective lengths of these intervals are

$$3.29 \times \frac{\sigma}{\sqrt{n}},\ 3.92 \times \frac{\sigma}{\sqrt{n}} \text{ and } 5.15 \times \frac{\sigma}{\sqrt{n}}$$

Sample Size for CI on the mean (normal, variance known)

The length of a CI is a measure of the precision of the estimation
$\rightarrow$ the precision is inversely related to the confidence level

However, it is often desirable to obtain a confidence interval that is both

- short enough for decision-making purposes
- of an adequate confidence level

$\rightarrow$ one way to reduce the length of a confidence interval with prescribed confidence level is by choosing n large enough

From the above, we know that in using $\bar{x}$ to estimate μ, the error $e = |\bar{x} - \mu|$ is less than $z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ with $100 \times (1-\alpha)\%$ confidence

$\rightarrow$ in other words, we can be $100 \times (1-\alpha)\%$ confident that the error will not exceed a given amount e when the sample size is

$$\boxed{n = \left(\frac{z_{1-\alpha/2}\sigma}{e}\right)^2}$$

Confidence interval on the mean of a normal distribution, variance known: example

Example

The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (in J) on specimens of steel cut at 60°C are as follows:

$$64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$$

Assume that impact energy is normally distributed with $\sigma = 1$ J. a) Find a 95% CI for μ, the mean impact energy for that kind of steel

Confidence interval on the mean of a normal distribution, variance known: example

Example (ctd.)

b) Determine how many specimens we should test to ensure that the 95% CI on the mean impact energy μ has a length of at most 1 J

Confidence interval on the mean of a normal distribution, variance known: remarks

Remark 1: if the population is normal, the confidence interval

$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad (\star)$$

is valid for all sample sizes $n \geq 1$

Remark 2: this interval is not the only $100 \times (1-\alpha)\%$ confidence interval for μ. For instance, starting from $\mathbb{P}(z_{\alpha/4} \leq Z \leq z_{1-3\alpha/4}) = 1-\alpha$ on Slide 23, another $100 \times (1-\alpha)\%$ CI could be

$$\left[\bar{x} - z_{1-3\alpha/4}\frac{\sigma}{\sqrt{n}}, \bar{x} - z_{\alpha/4}\frac{\sigma}{\sqrt{n}}\right]$$

However, interval $(\star)$ is often preferable, as it is symmetric around $\bar{x}$

Also, it can be shown that the symmetric confidence interval $(\star)$

$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

is the shortest one among all confidence intervals of level $1-\alpha$

Confidence interval on the mean of a normal distribution, variance known: remarks

Remark 3: in the same spirit, we have

$$\mathbb{P}(Z \leq z_{1-\alpha}) = \mathbb{P}(-z_{1-\alpha} \leq Z) = 1 - \alpha$$

Hence,

$$\left(-\infty, \bar{x} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right]$$

and

$$\left[\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right)$$

are also $100 \times (1 - \alpha)\%$ CI for μ

These are called one-sided confidence intervals, as opposed to ($\star$) (two-sided CI).

They are also sometimes called (upper and lower) confidence bounds.

What if the distribution is not normal?

So far, we have assumed that the population distribution is normal. In that situation, we have

$$Z = \sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

for any sample size n.

This sampling distribution is the cornerstone when deriving confidence intervals for μ, and directly follows from $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma)$.

A natural question is now:

What if the population is not normal? (X_i not $\mathcal{N}(\mu, \sigma)$)

$\rightarrow$ surprisingly enough, the above results still hold most of the time, *at least approximately*, due to the so-called

Central Limit Theorem

The Central Limit Theorem

The Central Limit Theorem (CLT) is certainly one of the most remarkable results in probability ("*the unofficial sovereign of probability theory*"). Loosely speaking, it asserts that

the sum of a large number of independent random variables has a distribution that is approximately normal

It was first postulated by Abraham de Moivre who used the bell-shaped curve to approximate the distribution of the number of heads resulting from many tosses of a fair coin.

However, this received little attention until the French mathematician Pierre-Simon Laplace (1749-1827) rescued it from obscurity in his monumental work "*Théorie Analytique des Probabilités*", which was published in 1812.

But it was not before 1901 that it was defined in general terms and formally proved by the Russian mathematician Aleksandr Lyapunov (1857-1918).

The Central Limit Theorem

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample taken from a population with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, then the limiting distribution of

$$\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$$

as $n \to \infty$, is the **standard normal distribution**

Proof: no proof provided

The Central Limit Theorem

When $X_i \sim \mathcal{N}(\mu, \sigma)$, $\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$ for all n.

What the CLT states is that, when the X_i's are not normal (whatever they are!), $\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$ when n is infinitely large

$\rightarrow$ the standard normal distribution provides a reasonable approximation to the distribution of $\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$ when "n is large"

This is usually denoted

$$\boxed{\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \overset{a}{\sim} \mathcal{N}(0,1)}$$

with $\overset{a}{\sim}$ for 'approximately follows' (or 'asymptotically ($n \to \infty$) follows')

The Central Limit Theorem

The power of the CLT is that it holds true **for any population distribution**, discrete or continuous! For instance,

$$X_i \sim \text{Exp}(\lambda) \quad \left(\mu = \frac{1}{\lambda}, \sigma = \frac{1}{\lambda}\right) \implies \sqrt{n}\,\frac{\bar{X} - 1/\lambda}{1/\lambda} \overset{a}{\sim} \mathcal{N}(0,1)$$

$$X_i \sim U_{[a,b]} \quad \left(\mu = \frac{a+b}{2}, \sigma = \frac{b-a}{\sqrt{12}}\right) \implies \sqrt{n}\,\frac{\bar{X} - \frac{a+b}{2}}{\frac{b-a}{\sqrt{12}}} \overset{a}{\sim} \mathcal{N}(0,1)$$

$$X_i \sim \text{Bern}(\pi) \quad \left(\mu = \pi, \sigma = \sqrt{\pi(1-\pi)}\right) \implies \sqrt{n}\,\frac{\bar{X} - \pi}{\sqrt{\pi(1-\pi)}} \overset{a}{\sim} \mathcal{N}(0,1)$$

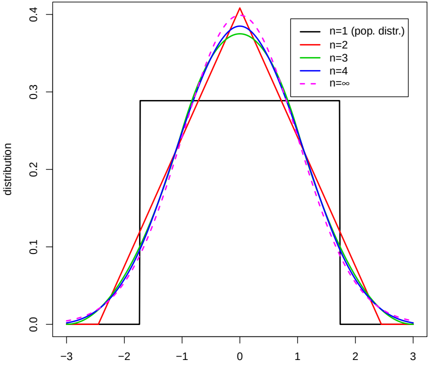
<u>Facts:</u>

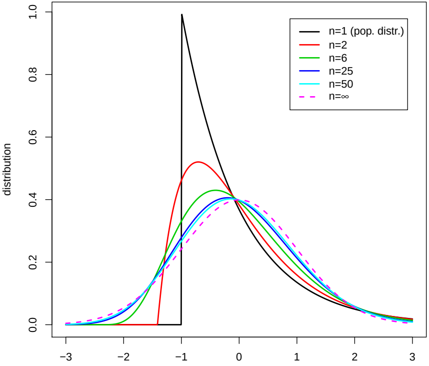
- the larger n, the better the normal approximation
- the closer the population distribution is to being normal, the more rapidly the distribution of $\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$ approaches normality as n gets large

The Central Limit Theorem: illustration

Probability density functions for $\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$

$\boxed{X_i \sim U_{[-\sqrt{3},\sqrt{3}]}}$ $(\mu = 0, \sigma = 1)$ $\boxed{X_i \sim \text{Exp}(1) - 1}$ $(\mu = 0, \sigma = 1)$

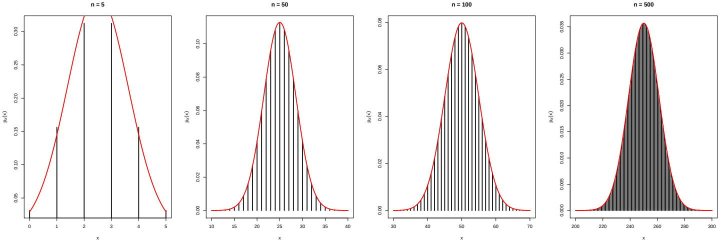




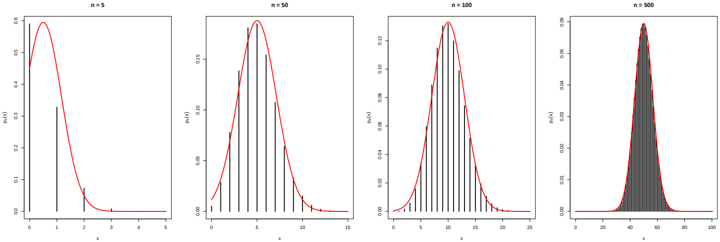
The Central Limit Theorem: illustration

Probability mass functions for $\sum_{i=1}^{n} X_i$, $X_i \sim \text{Bern}(\pi)$

$\pi = 0.5$:



$\pi = 0.1$:



The Central Limit Theorem: further illustration

Go to

http://onlinestatbook.com/stat_sim/sampling_dist/index.html

The Central Limit Theorem: remarks

Remark 1:

The Central Limit Theorem not only provides a simple method for computing approximate probabilities for sums or averages of independent random variables.

It also helps explain why so many natural populations exhibit a bell-shaped (i.e., normal) distribution curve:

Indeed, as long as the behaviour of the variable of interest is dictated by a large number of independent contributions, it should be (at least approximately) normally distributed.

For instance, a person's height is the result of many independent factors, both genetic and environmental. Each of these factors can increase or decrease a person's height, just as each ball in Galton's board can bounce to the right or the left

$\rightarrow$ the Central Limit Theorem guarantees that the sum of these contributions has approximately a normal distribution

The Central Limit Theorem: remarks

Remark 2:

a natural question is '**how large n needs to be**' for the normal approximation to be valid

$\rightarrow$ that depends on the population distribution!

A general rule-of-thumb is that one can be fairly confident of the normal approximation whenever the sample size n is at least 30

$$\boxed{n \geq 30}$$

Note that, in favourable cases (population distribution is not skewed and not long-tailed), the normal approximation will be satisfactory for much smaller sample sizes (like $n = 5$ in the uniform case, for instance)

If $n \geq 30$, the normal distribution will provide a good approximation to the sampling distribution of $\bar{X}$ irrespective of the shape of the population (well, except for variables with extremely right-skewed or very long-tailed distributions).

Confidence interval on the mean of an arbitrary distribution

The Central Limit Theorem allows us to use the procedures described earlier to derive confidence intervals for μ in an arbitrary population, bearing in mind that these will be **approximate confidence intervals** (whereas they were exact in a normal population)

Indeed, if n is large enough,

$$Z = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \overset{a}{\sim} \mathcal{N}(0,1),$$

hence

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

where $z_{1-\alpha/2}$ is the quantile of level $1-\alpha/2$ of the standard normal distribution.

Confidence interval on the mean of an arbitrary distribution

It follows

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

so that if $\bar{x}$ is the sample mean of an observed random sample of size n from any distribution with known variance σ^2, an approximate confidence interval of level $100 \times (1-\alpha)\%$ for μ is given by

$$\boxed{\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Note: because this result requires "n large enough" to be reliable, this type of interval, based on the CLT, is often called large-sample confidence interval.

One could also define large-sample one-sided confidence intervals of level $100 \times (1-\alpha)\%$: $(-\infty, \bar{x} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}]$ and $[\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, +\infty)$.

Objectives

Now you should be able to:

- Explain important properties of point estimators, including bias, variance, efficiency and consistency □
- Know how to compute and explain the precision with which a parameter is estimated □
- Understand the basics of interval estimation and explain what a confidence interval of level $100 \times (1 - \alpha)\%$ for a parameter is □
- Construct exact confidence intervals on the mean of a normal distribution with known variance □
- Understand the Central Limit Theorem and explain the important role of the normal distribution in inference □
- Construct large sample confidence intervals on a mean of an arbitrary distribution □

Recommended exercises:

$\rightarrow$ Q46+Q49, Q50 p.237, Q69 p.239, Q1, Q3 p.293, Q5, Q6 p.294, Q10, Q11 p.301 (2nd edition)

$\rightarrow$ Q48+Q51, Q52 p.241, Q72 p.244, Q1, Q3 p.297, Q5, Q6 p.298, Q10 p.305, Q11 p.306 (3rd edition)