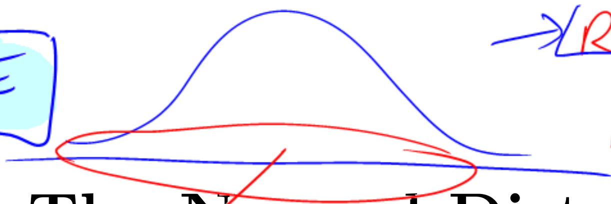
Exam prep: PRACTISE
Start NOW!



→ Revision List
Info Exam.
Past exams on Website

Chapter 5: The Normal Distribution and the Central Limit Theorem

The Normal distribution is the familiar bell-shaped distribution. It is probably the most important distribution in statistics, mainly because of its link with the Central Limit Theorem, which states that any large sum of independent, identically distributed r.v.s (random variables) is approximately Normal:

$$\boxed{\begin{array}{l}\text{if } X_1, X_2, \ldots, X_n \text{ are I.I.D. and } n \text{ is large,} \\ \text{then } X_1 + X_2 + \ldots + X_n \sim \text{approx Normal.}\end{array}}$$

Before studying the Central Limit Theorem, we look at the Normal distribution and some of its general properties.

5.1 The Normal Distribution

The Normal distribution has two parameters, the mean, μ, and the variance, σ^2. (or sometimes we use μ and $\sigma = s.d.$ instead.)

μ and σ^2 satisfy $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

We write $X \sim \text{Normal}(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma^2)$.

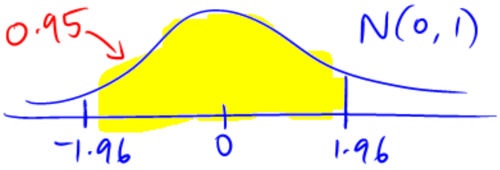
Probability density function, $f_X(x)$

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\{-(x-\mu)^2/2\sigma^2\}} \quad \text{for } -\infty < x < \infty.}$$

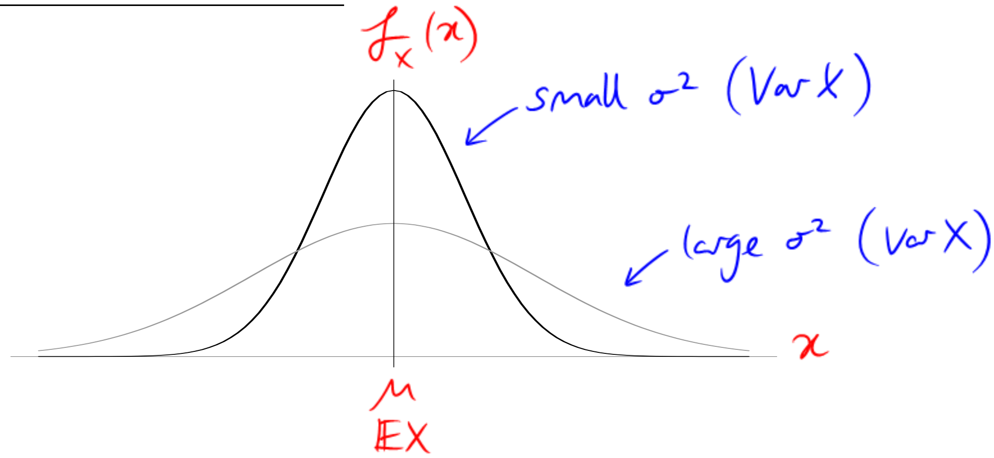
don't need to learn

Distribution function, $F_X(x)$

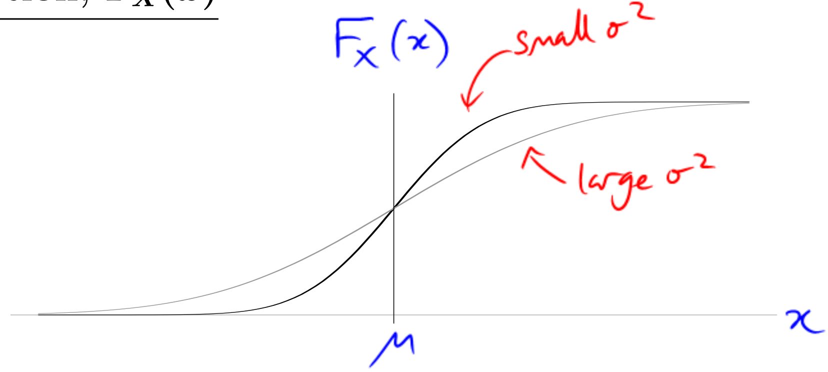
There is no closed form for the distribution function of the Normal distribution. If $X \sim \text{Normal}(\mu, \sigma^2)$, then $F_X(x)$ can can only be calculated by computer;
R command: $F_X(x)$ = pnorm(x, mean = μ, sd = sqrt(σ^2))



Probability density function, $f_X(x)$



Distribution function, $F_X(x)$



Mean and Variance

For $X \sim \text{Normal}(\mu, \sigma^2)$, $\boxed{\mathbb{E}X = \mu \text{ and } Var(X) = \sigma^2}$

Linear transformations

If $X \sim \text{Normal}(\mu, \sigma^2)$, then for any constants a and b,

$Y = aX + b$

$$\boxed{aX + b \sim \text{Normal}\Big(a\mu + b,\ a^2\sigma^2\Big).}$$

$Var(aX+b) = a^2 Var X$
$= a^2\sigma^2$

$\mathbb{E}(aX+b) = a\,\mathbb{E}X + b = a\mu + b$

In particular, put $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$, then $Y = aX + b = \frac{X-\mu}{\sigma}$

and $\mathbb{E}Y = 0$, $Var Y = 1$

$$\boxed{\text{If } X \sim \text{Normal}(\mu, \sigma^2), \text{ then } Y = \left(\frac{X-\mu}{\sigma}\right) \sim \text{Normal}(0,1).}$$

$Z \sim \text{Normal}(0, 1)$ is called the Standard Normal random variable.

hugely important for Confidence intervals

Proof that $aX + b \sim$ Normal$\left(a\mu + b,\ a^2\sigma^2\right)$:

Let $X \sim$ Normal(μ, σ^2), and let $Y = aX + b$. We wish to find the distribution of Y. Use *the change of variable technique.*

1) $y(x) = ax + b$ is monotone, so we can apply the Change of Variable technique.

2) Let $y = y(x) = ax + b$ for $-\infty < x < \infty$.

3) Then $x = x(y) = \frac{y-b}{a}$ for $-\infty < y < \infty$.

4) $\left|\dfrac{dx}{dy}\right| = \left|\dfrac{1}{a}\right| = \dfrac{1}{|a|}$.

5) So $\quad f_Y(y) = f_X(x(y))\left|\dfrac{dx}{dy}\right| = f_X\left(\dfrac{y-b}{a}\right)\dfrac{1}{|a|}. \qquad (\star)$

But $X \sim$ Normal(μ, σ^2), so $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$

$$\text{Thus } f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\frac{y-b}{a}-\mu)^2/2\sigma^2}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-(a\mu+b))^2/2a^2\sigma^2}.$$

Returning to $(\star)$,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\cdot\frac{1}{|a|} = \frac{1}{\sqrt{2\pi a^2\sigma^2}}e^{-(y-(a\mu+b))^2/2a^2\sigma^2} \quad \text{for } -\infty < y < \infty.$$

But this is the p.d.f. of a Normal$(a\mu + b,\ a^2\sigma^2)$ random variable.

So, if $X \sim$ Normal(μ, σ^2), then $aX + b \sim$ Normal$\left(a\mu + b,\ a^2\sigma^2\right)$.

$\left\{ \begin{array}{l} \mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y \text{ always} \\ \text{Var}(X+Y) = \text{Var}X + \text{Var}Y \text{ if indept.} \end{array} \right.$

Sums of Normal random variables

If X and Y are independent and $X \sim \text{Normal}(\mu_1, \sigma_1^2)$, $Y \sim \text{Normal}(\mu_2, \sigma_2^2)$, then

$$X + Y \sim \text{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Proof: needs more advanced methods

More generally, if $X_1, X_2, \ldots, X_n$ are independent, and $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$, then

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n \sim \text{Normal}\Big((a_1\mu_1 + \ldots + a_n\mu_n), \ (a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2) \Big).$$

For mathematicians: properties of the Normal distribution

1. Proof that $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

The full proof that $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\{-(x-\mu)^2/(2\sigma^2)\}}\,dx = 1$

relies on the following result:

$$\text{FACT:} \qquad \int_{-\infty}^{\infty} e^{-y^2}\,dy \;=\; \sqrt{\pi}.$$

This result is non-trivial to prove. See Calculus courses for details.

Using this result, the proof that $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ follows by using the change of variable $y = \dfrac{(x-\mu)}{\sqrt{2}\sigma}$ in the integral.

2. Proof that $\mathbb{E}(X) = \mu$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Change variable of integration: let $z = \frac{x-\mu}{\sigma}$: then $x = \sigma z + \mu$ and $\frac{dx}{dz} = \sigma$.

$$\text{Then} \quad \mathbb{E}(X) \;=\; \int_{-\infty}^{\infty} (\sigma z + \mu) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-z^2/2} \cdot \sigma\,dz$$

$$= \underbrace{\int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \cdot e^{-z^2/2}\, dz}_{\substack{\text{this is an odd function of } z \\ \text{(i.e. } g(-z) = -g(z)\text{), so it} \\ \text{integrates to 0 over range} \\ -\infty \text{ to } \infty.}} \quad + \quad \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz}_{\text{p.d.f. of } N(0,1) \text{ integrates to 1.}}$$

$$\begin{aligned} \text{Thus} \quad \mathbb{E}(X) &= 0 + \mu \times 1 \\ &= \mu. \end{aligned}$$

3. Proof that $\mathrm{Var}(X) = \sigma^2$.

$$\begin{aligned} \mathrm{Var}(X) &= E\left\{(X-\mu)^2\right\} \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}\, dx \\ &= \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z^2\, e^{-z^2/2}\, dz && \left(\text{putting } z = \frac{x-\mu}{\sigma}\right) \\ &= \sigma^2 \left\{ \frac{1}{\sqrt{2\pi}} \left[-ze^{-z^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \right\} && \text{(integration by parts)} \\ &= \sigma^2 \{0 + 1\} \\ &= \sigma^2. \end{aligned}$$

$\square$

5.2 The Central Limit Theorem (CLT)

also known as. . . the Piece of Cake Theorem

The Central Limit Theorem (CLT) is one of the most fundamental results in statistics. In its simplest form, it states that if a large number of independent random variables are drawn from ***any*** distribution, then the distribution of their sum (or alternatively their sample average) always converges to the Normal distribution.

'CLT converts std errors into confidence intervals.'

Theorem (The Central Limit Theorem):

Let $X_1, \ldots, X_n$ be independent r.v.s with mean μ and variance σ^2 from ANY distribution.

For example: let $X_i \sim \text{Bin}(m, p)$ for all i, so $\mu = mp$ and $\sigma^2 = mp(1-p)$.

Then the sum, $S_n = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$ has a distribution that tends to Normal as $n \to \infty$.

The ***mean*** of the Normal distribution is $\mathbb{E}(S_n) = \mathbb{E}(X_1 + \cdots + X_n) = n\mu$.

The ***variance*** of the Normal distribution is

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}(X_1 + \cdots + X_n) \\ &= \text{Var}(X_1) + \cdots + \text{Var}(X_n) \quad \text{by indep.} \\ &= n\sigma^2. \end{aligned}$$

So $\boxed{S_n = X_1 + \cdots + X_n \sim \text{approx Normal}(n\mu, n\sigma^2) \text{ as } n \to \infty.}$

Notes:

1. This is a remarkable theorem, because the limit holds for ***any*** distribution of $X_1, \ldots, X_n$.

2. A sufficient condition on X for the Central Limit Theorem to apply is that $\text{Var}(X)$ is finite. Other versions of the Central Limit Theorem relax the conditions that $X_1, \ldots, X_n$ are independent and have the same distribution.

3. The ***speed*** of convergence of S_n to the Normal distribution depends upon the distribution of X. Skewed distributions converge more slowly than symmetric Normal-like distributions. It is usually safe to assume that the Central Limit Theorem applies whenever $n \geq 30$. It might apply for as little as $n = 4$.

If $X \sim$ Normal
then $aX+b \sim$ Normal.
$\mathbb{E}(aX+b) = a\mathbb{E}X + b$

Distribution of the sample mean, $\overline{X}$, using the CLT

Let $X_1, \ldots, X_n$ be independent, identically distributed with mean $\mathbb{E}(X_i) = \mu$, and variance $Var(X_i) = \sigma^2$ for all i.

The sample mean, $\overline{X}$, is defined as:

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}.$$

So $\overline{X} = \frac{S_n}{n}$, where $S_n = X_1 + \cdots + X_n \sim \text{approx Normal}(n\mu, n\sigma^2)$ by CLT.

(So $\overline{X} = aS_n$ where $a = \frac{1}{n}$ and $S_n \sim$ Normal.)

Because $\overline{X}$ is a scalar multiple of a Normal r.v. as n grows large, $\overline{X}$ *itself is approximately Normal for large* n:

$$\frac{X_1 + \cdots + X_n}{n} \sim \text{approx Normal}\left(\mu, \frac{\sigma^2}{n}\right) \text{ as } n \to \infty$$

The following three statements of the Central Limit Theorem are equivalent:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim \text{approx Normal}\left(\mu,\ \frac{\sigma^2}{n}\right) \quad \text{as } n \to \infty.$$

$$S_n = X_1 + X_2 + \ldots + X_n \sim \text{approx Normal}\left(n\mu,\ n\sigma^2\right) \quad \text{as } n \to \infty.$$

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim \text{approx Normal}(0,\ 1) \quad \text{as } n \to \infty.$$

The essential point to remember about the Central Limit Theorem is that large sums or sample means of independent random variables converge to a Normal distribution, WHATEVER the distribution of the original r.v.s.

More general version of the CLT

A more general form of CLT states that, if $X_1, \ldots, X_n$ are independent, and $\mathbb{E}(X_i) = \mu_i$, $\text{Var}(X_i) = \sigma_i^2$ (not necessarily all equal), then

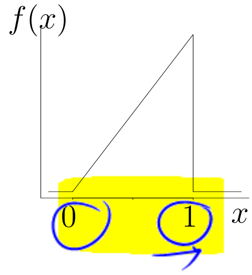
$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \to \text{Normal}(0, 1) \quad \text{as } n \to \infty.$$

Other versions of the CLT relax the condition that $X_1, \ldots, X_n$ are independent.

The Central Limit Theorem in action : simulation studies

The following simulation study illustrates the Central Limit Theorem, making use of several of the techniques learnt in STATS 210. We will look particularly at *how fast the distribution of S_n converges to the Normal distribution.*

Example 1: Triangular distribution: $f_X(x) = 2x$ for $0 < x < 1$.



Find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$:

$$\mu = \mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$= \int_0^1 x \; 2x \, dx$$ ← skip straight here if you like.

$$= \left[\frac{2x^3}{3} \right]_0^1$$

$$\mu = \frac{2}{3}$$

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \int_0^1 x^2 * 2x\,dx - \left(\frac{2}{3}\right)^2$$

$$= \left[\frac{2x^4}{4} \right]_0^1 - \frac{4}{9}$$

$$\therefore \sigma^2 = \frac{1}{18}$$

Let $S_n = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent.

Then $\mathbb{E}(S_n) = \mathbb{E}(X_1 + \cdots + X_n) = n\mu = \frac{2n}{3}$.

1 mark.

$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = n\sigma^2$ by independence

$\Rightarrow \mathrm{Var}(S_n) = \frac{n}{18}$.

So

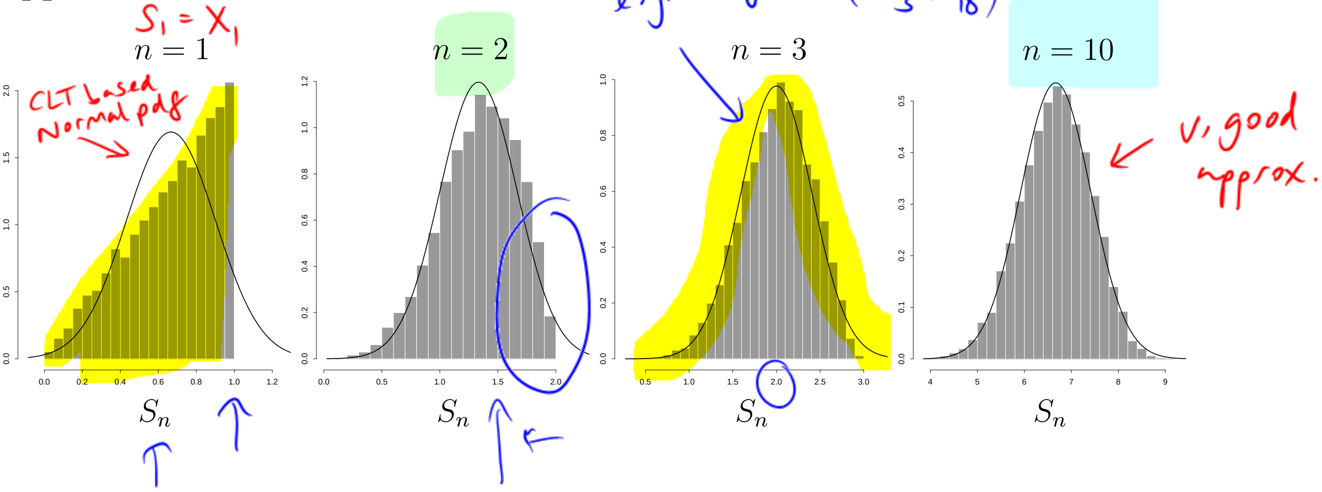
$S_n \sim$ approx Normal$\left(\frac{2n}{3}, \frac{n}{18}\right)$ for large n, by CLT.

Central limit Thm.

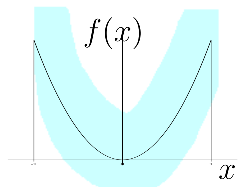
$S_n = X_1 + \dots + X_n$

The graph shows histograms of 10 000 values of $S_n = X_1 + \ldots + X_n$ for $n = 1, 2, 3$, and 10. The Normal p.d.f. $\text{Normal}(n\mu, n\sigma^2) = \text{Normal}(\frac{2n}{3}, \frac{n}{18})$ is superimposed across the top. Even for n as low as 10, the Normal curve is a very good approximation.

e.g. PDF for $N(\frac{2*3}{3}, \frac{3}{18})$



Example 2: U-shaped distribution: $f_X(x) = \frac{3}{2}x^2$ for $-1 < x < 1$.



We find that $\mathbb{E}X = \mu = 0$, $Var(X) = \sigma^2 = \frac{3}{5}$ (Exercise).

Let $S_n = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent.

Then

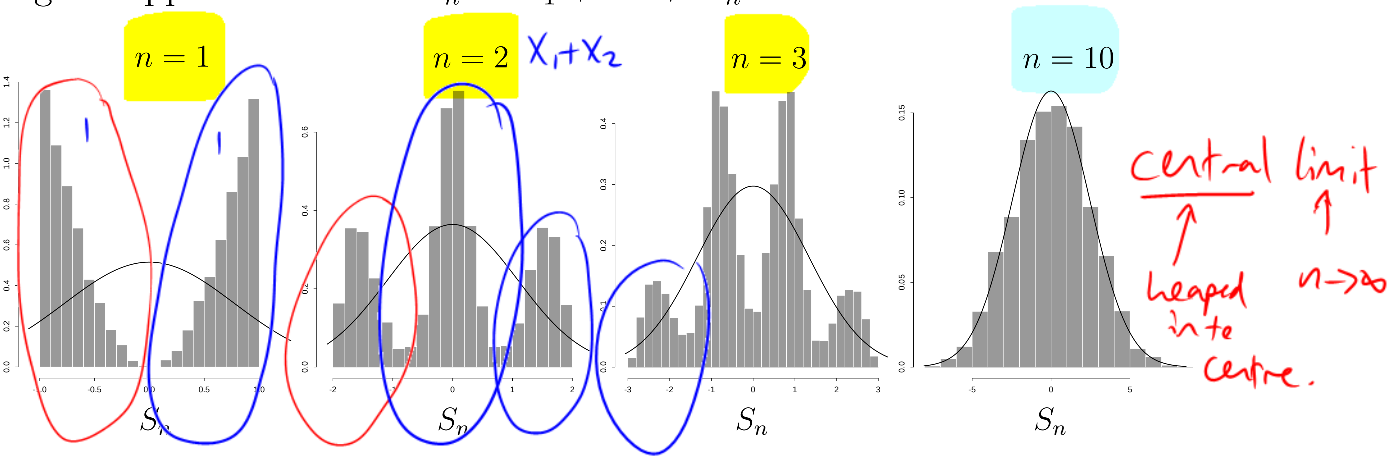
$$\mathbb{E}(S_n) = \mathbb{E}(X_1 + \dots + X_n) = n\mu = 0$$

$$Var(S_n) = Var(X_1 + \dots + X_n) = n\sigma^2 \text{ by indep.}$$

$$\Rightarrow Var(S_n) = \frac{3n}{5}.$$

So $S_n \sim$ approx Normal $(0, \frac{3n}{5})$ for large n, by CLT.

Even with this highly non-Normal distribution for X, the Normal curve provides a good approximation to $S_n = X_1 + \ldots + X_n$ for n as small as 10.



Normal approximation to the Binomial distribution, using the CLT

Let $Y \sim \text{Binomial}(n, p)$.

We can think of Y as the *sum of n Bernoulli random variables:*

prob p

$Y = X_1 + X_2 + \ldots + X_n$, where $X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise (prob } 1-p). \end{cases}$

So $Y = X_1 + \cdots + X_n$ and each X_i has $\mu = \mathbb{E}(X_i) = p$
and $\sigma^2 = \text{Var}(X_i) = p(1-p)$.

Thus by the CLT,

$Y = X_1 + \cdots + X_n \to \text{Normal}(n\mu, n\sigma^2)$ large n

$\sim \text{Normal}(np, np(1-p))$.

Thus,

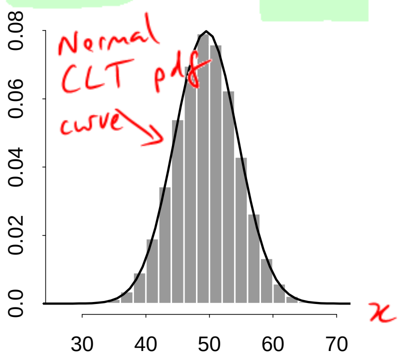
$$\textit{Bin}(n,p) \to \textit{Normal}\Big(\underbrace{np}_{\text{mean of Bin}(n,p)}, \underbrace{np(1-p)}_{\text{var of Bin}(n,p)}\Big) \text{ as } n \to \infty \text{ with } p \text{ fixed.}$$

The Binomial distribution is therefore well approximated by the Normal distribution when n is large, for any fixed value of p.

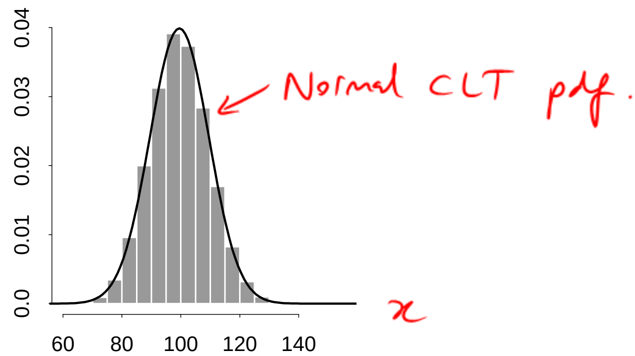
The Normal distribution is also a good approximation to the Poisson(λ) distribution when λ is large:

$$\textit{Poisson}(\lambda) \to \textit{Normal}(\lambda, \lambda) \text{ when } \lambda \text{ is large.}$$

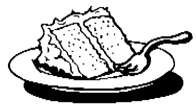
$P(X=x)$ Binomial($n = 100$, $p = 0.5$)



$P(X=x)$ Poisson($\lambda = 100$)



Why the Piece of Cake Theorem? . . .



- The Central Limit Theorem makes whole realms of statistics into a piece of cake.
- After seeing a theorem this good, you deserve a piece of cake!

5.3 Confidence intervals

Example: Remember the margin of error for an opinion poll?

An opinion pollster wishes to estimate the level of support for Labour in an upcoming election. She interviews n people about their voting preferences. Let p be the true, unknown level of support for the Labour party in New Zealand. Let X be the number of of the n people interviewed by the opinion pollster who plan to vote Labour. Then $X \sim \text{Binomial}(n, p)$.

At the end of Chapter 2, we said that the maximum likelihood estimator for p is $\hat{p} = \dfrac{X}{n}$.

In a large sample (large n), we now know that

$$X \sim \text{approx Normal}(np, npq) \text{ for large } n \text{ (CLT)}$$

where $q = 1 - p$.

So $\hat{p} = aX + b$ where $X \overset{\cdot}{\sim} \text{Normal}$, $a = \frac{1}{n}$, $b = 0$.

So

$$\hat{p} \sim \text{approx Normal}\left(p, \frac{pq}{n}\right)$$

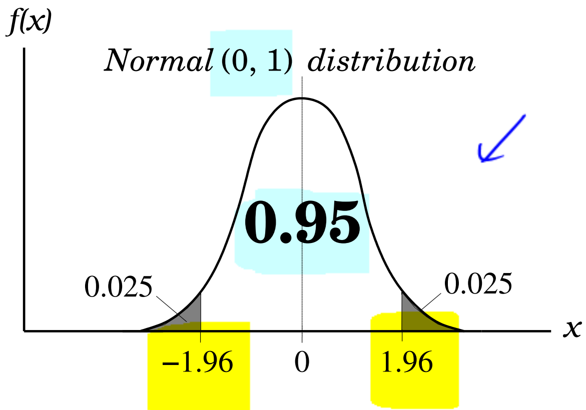
So

$$\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim \text{approx Normal}(0, 1).$$

Now if $Z \sim \text{Normal}(0, 1)$, we find (using a computer) that the 95% central probability region of Z is from -1.96 to $+1.96$:

$$\mathbb{P}(-1.96 < Z < 1.96) = 0.95.$$

Check in R: `pnorm(1.96, mean=0, sd=1) - pnorm(-1.96, mean=0, sd=1)`



Putting $Z = \frac{\widehat{p}-p}{\sqrt{\frac{pq}{n}}}$, we obtain

$$\mathbb{P}\left(-1.96 < \frac{\widehat{p}-p}{\sqrt{\frac{pq}{n}}} < 1.96\right) \simeq 0.95.$$

Rearranging to put the unknown p in the middle:

$$\mathbb{P}\left(\underbrace{\widehat{p} - 1.96\sqrt{\frac{pq}{n}}}_{\text{random}} < \underbrace{p}_{\text{const but unknown}} < \underbrace{\widehat{p} + 1.96\sqrt{\frac{pq}{n}}}_{\text{random}}\right) \simeq 0.95.$$

This enables us to form an estimated 95% confidence interval for the unknown parameter p: estimated 95% C.I. is:

$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{to} \quad \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The 95% confidence interval *has RANDOM end-points, which depend on* $\widehat{p}$. *About 95% of the time, these random end-points will enclose the true unknown value,* p.

Confidence intervals are extremely important for helping us to assess how useful our estimate is.

A narrow confidence interval suggests a useful estimate (low variance).
A wide confidence interval suggests a poor estimate (high variance).

When you see newspapers quoting the ***margin of error*** on an opinion poll:

- Remember: margin of error $= 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Think: Central Limit Theorem!
- Have: a piece of cake. ☺

Confidence intervals for the Poisson λ parameter

We saw in section 3.6 that if $X_1, \ldots, X_n$ are independent, identically distributed with $X_i \sim \text{Poisson}(\lambda)$, then the maximum likelihood estimator of λ is

$$\widehat{\lambda} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Now $\mathbb{E}(X_i) = \mu = \lambda$ and $\text{Var}(X_i) = \sigma^2 = \lambda$ for $i = 1, \ldots, n$.

Thus, when n is large,

$$\widehat{\lambda} = \overline{X} \sim \text{approx Normal}(\mu, \frac{\sigma^2}{n})$$

by the Central Limit Theorem. In other words,

$$\hat{\lambda} = \text{approx Normal}\left(\lambda, \frac{\lambda}{n}\right). \quad \text{as } n \to \infty.$$

We use the same transformation as before to find approximate 95% confidence intervals for λ as n grows large:

Let $Z = \dfrac{\widehat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}}$. We have $Z \sim \text{approx } N(0,1)$ for large n.

Thus:

$$\mathbb{P}\left(-1.96 < \frac{\widehat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} < 1.96\right) \simeq 0.95.$$

Rearranging

$$\mathbb{P}\left(\widehat{\lambda} - 1.96\sqrt{\frac{\lambda}{n}} \; < \; \lambda \; < \; \widehat{\lambda} + 1.96\sqrt{\frac{\lambda}{n}}\right) \simeq 0.95.$$

So our ***estimated*** 95% confidence interval for the unknown parameter λ is:

$$\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}}{n}} \quad \text{to} \quad \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}}{n}}$$

Why is this so good?

It's clear that it's important to measure precision, or reliability, of an estimate, otherwise the estimate is almost worthless. However, we have already seen various measures of precision: variance, standard error, coefficient of variation, and now confidence intervals. Why do we need so many?

- The true variance of an estimator, e.g. $\mathrm{Var}(\widehat{\lambda})$, is the most convenient quantity to work with mathematically. However, it is on a non-intuitive scale (squared deviation from the mean), and it usually depends upon the unknown parameter, e.g. λ.

- The ***standard error*** is $\mathrm{se}(\widehat{\lambda}) = \sqrt{\widehat{\mathrm{Var}}\left(\widehat{\lambda}\right)}$. It is an ***estimate*** of the square root of the true variance, $\mathrm{Var}(\widehat{\lambda})$. Because of the square root, the standard error is a direct measure of deviation from the mean, rather than squared deviation from the mean. This means it is measured in more intuitive units. However, it is still unclear how we should comprehend the information that the standard error gives us.

- The beauty of the Central Limit Theorem is that it gives us an incredibly easy way of understanding what the standard error is telling us, using ***Normal-based asymptotic confidence intervals*** as computed in the previous two examples.

  Although it is beyond the scope of this course to see why, the Central Limit Theorem guarantees that almost ***any*** maximum likelihood estimator will be Normally distributed as long as the sample size n is large enough, subject only to fairly mild conditions.

  Thus, if we can find an estimate of the variance, e.g. $\widehat{\mathrm{Var}}(\widehat{\lambda})$ we can immediately convert it to an estimated 95% confidence interval using the CLT / Normal formulation:

  $$\widehat{\lambda} - 1.96\sqrt{\widehat{\mathrm{Var}}(\widehat{\lambda})} \quad \text{to} \quad \widehat{\lambda} + 1.96\sqrt{\widehat{\mathrm{Var}}(\widehat{\lambda})},$$

  or equivalently,

  $$\widehat{\lambda} - 1.96\,\mathrm{se}(\widehat{\lambda}) \quad \text{to} \quad \widehat{\lambda} + 1.96\,\mathrm{se}(\widehat{\lambda}).$$

  The confidence interval has an easily-understood interpretation: on 95% of occasions we conduct a random expt and build a confidence interval, the interval will contain the true parameter.

  So the Central Limit Theorem has given us an incredibly simple and powerful way of converting from a hard-to-understand measure of precision, $\mathrm{se}(\widehat{\lambda})$, to a measure that is easily understood and relevant to the problem at hand. Brilliant!