Central Limit Theorem

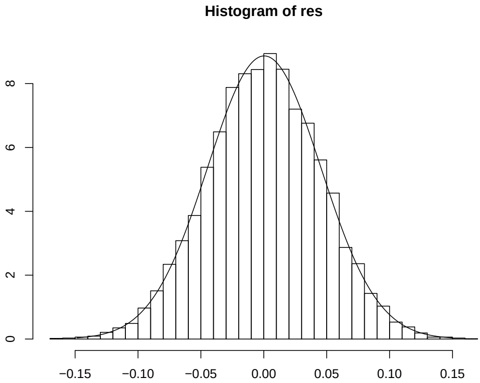
- We know that if $X_1, ..., X_n$ are independent and $X_i \sim N(\mu, \sigma^2),\ i = 1, \ldots, n$, then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n}).$$

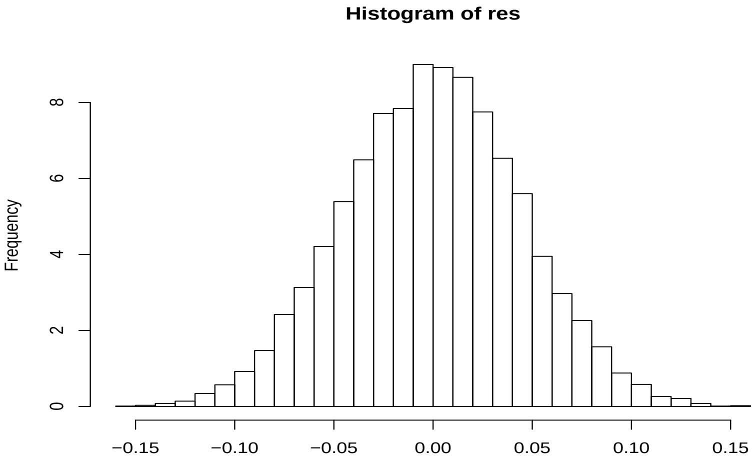
But what if $X_1, ..., X_n$ are not normally distributed? What is the distribution of the sample mean?

- Consider the following simulation experiment:
  (i) Think of a number $n \geq 1$, and then generate n values (observations) from the random variables $X_1, ..., X_n$ with a common distribution (e.g. binomial, uniform, normal, exponential etc.)
  (ii) Produce the sample mean $\bar{X}$ from the sample.
  (iii) Repeat this process 1,000 times to produce $\bar{X}_1, ..., \bar{X}_{1,000}$
  (iv) Plot a histogram of the $\bar{X}'s$.

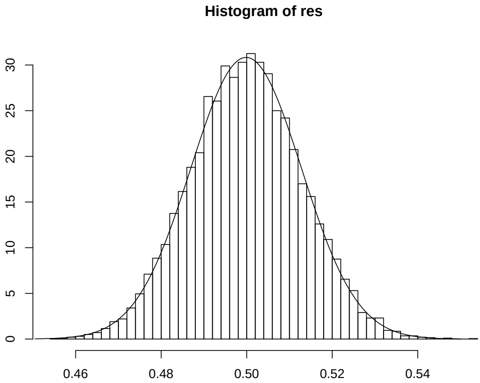
  - When $X_i \sim N(\mu, \sigma^2)$ are normal, then $\bar{X} \sim N(\mu, \sigma^2/n)$, and so the histograms from the $\bar{X}'s$ should look normal in that case.



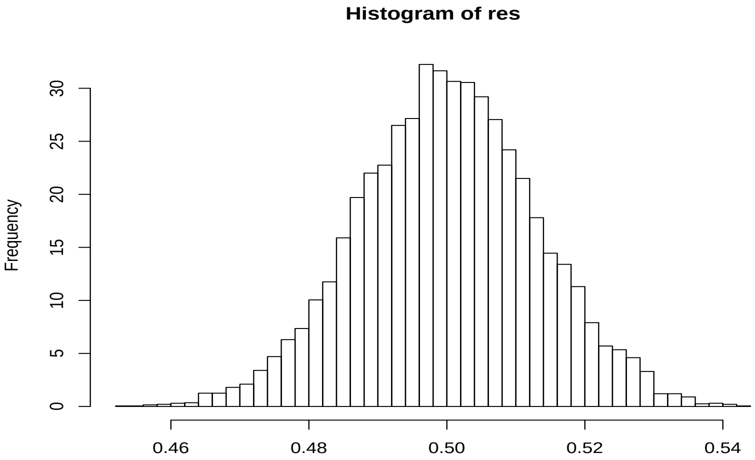
Histogram of 1,000 values of the sample mean based on $n = 500$ observations from $N(0, 1)$, with the imposed graph of a normal density function.


Histogram of res
Frequency
0
2
4
6
8
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15

- Now let us repeat the experiment but this time we generate $1,000$ values of the sample mean where each mean is based on observations of the random variables $X_1, ..., X_n$ that have the same uniform distribution on $(0, 1)$.



Histogram of 1,000 values of the sample mean based on $n = 500$ observations from $U(0, 1)$, with the imposed graph of a normal density function.


Histogram of res
Frequency
0
5
10
15
20
25
30
0.46
0.48
0.50
0.52
0.54

Theorem (The Central Limit Theorem). Suppose that $X_1, ..., X_n$ are independent and identically distributed r.v.'s with a common mean μ and variance σ^2.

Then we have

$$P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq x\right) \stackrel{n\to\infty}{\longrightarrow} \Phi(x), \quad \text{for all } x \in \mathbb{R}.$$

- Thus, if n is large,

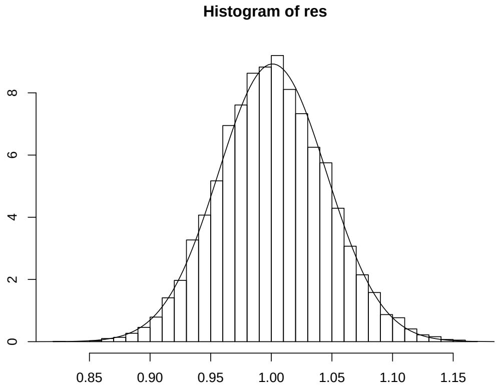
$$\frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}} \stackrel{approx}{\approx} N(0,1)$$

or

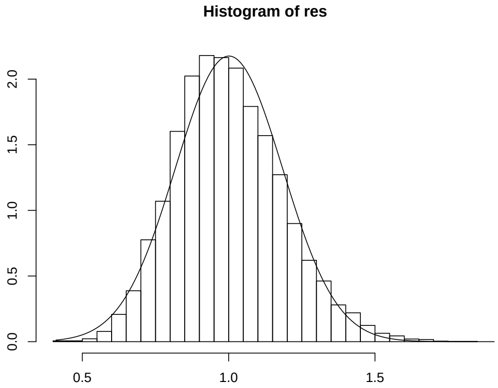
$$\bar{X} \stackrel{approx}{\approx} N(\mu, \frac{\sigma^2}{n}) \quad and \quad \sum_{i=1}^{n} X_i \stackrel{approx}{\approx} N(n\mu, n\sigma^2).$$

- Note that in the formulation of this theorem we do not specify the common distribution, say F, of the variables $X_1, X_2, \ldots$, only its moments. This common distribution can be continuous or discrete!

- Rules of thumb for using the Central Limit Theorem (C.L.T.):
    - In general if the number of observations exceeds 30, then the C.L.T. often provides a reasonable approximation.
    - If the distribution F of the observations is "close" to being unimodal, not too skewed, and is "close" to being continuous, then the C.L.T. provides approximations that are acceptable for even smaller values of n (5-15).
    - If the distribution F is highly skewed, or discrete with a small number of possible values, then a larger value of n might be necessary ($n > 50$).



Histogram of 1,000 values of the sample mean based on $n = 500$ observations from *Exp*(1), with the imposed graph of a normal density function.



Histogram of 1,000 values of the sample mean based on $n = 30$ observations from $Exp(1)$, with the imposed graph of a normal density function.

Example. The life time of battery is random with mean 40 hours and standard deviation 20 hours. After a battery is used, it is replaced by a new one.

Suppose we have a stockpile of 30 such batteries the lifetimes of which are independent. Approximate the probability that over 1000 hours of use can be obtained.

Exercise. Customers arrive at the automatic teller machine in accordance with a Poisson process with rate 20 per hour. Assuming that the machine is in use for 10 hours daily, and the arrival times are independent, approximate the probability that the 200'th customer will arrive before the machine is closed for the day[1].

Clicker Question(s).

[1] see also Example on page 249.

- The Central Limit Theorem can be used to obtain Normal approximations to some standard distributions. Examples include:

  (i) If $X_n \sim Binomial(n, p)$, and n is large, then

  $$\frac{X_n - np}{\sqrt{np(1-p)}} \overset{approx}{\sim} N(0, 1).$$

  Thus, for large n, $X_n \overset{approx}{\sim} N(np, np(1-p))$.

  (ii) If $X_\mu \sim Poi(\mu)$, and μ is large, then

  $$\frac{X_\mu - \mu}{\sqrt{\mu}} \overset{approx}{\sim} N(0, 1)$$

  Thus, for large μ, $X_\mu \overset{approx}{\sim} N(\mu, \mu)$.

Clicker Question(s).

Example. Suppose that Billy flips a fair coin 30 times. Approximate the probability that the number of heads is between 5 and 20. Compare the answer with the exact value.

Let X be the number of heads. Then $X \sim \mathit{Binomial}(30, 0.5)$, and by the Normal approximation

$$X \sim N(\mu = np = 15, \sigma^2 = np(1-p) = 7.5).$$

Therefore, for $Z \sim N(0,1)$, we have

$$\begin{aligned}
P(5 \leq X \leq 20) &= P\left(\frac{5-15}{\sqrt{7.5}} \leq \frac{X-15}{\sqrt{7.5}} \leq \frac{20-15}{\sqrt{7.5}}\right) \\
&= P(-3.651 \leq Z \leq 1.826) \\
&= \Phi(1.826) - \Phi(-3.651) = 0.966 - 0.0001 \\
&= 0.9659.
\end{aligned}$$

The exact value is[2]

$$\sum_{x=5}^{20} \binom{30}{x} (0.5)^x (0.5)^{30-x} = 0.9786$$

[2]calculated using R: *pbinom*$(20, 30, 0.5)$ − *pbinom*$(4, 30, 0.5)$

- **The continuity correction:** if we are using a Normal distribution to approximate

$$P(a \leq X \leq b)$$

where a and b are integers and X follows a **discrete** distribution, then we can obtain a more accurate approximation by replacing

$$a \leq X \leq b \quad \text{with} \quad a - 0.5 \leq X \leq b + 0.5.$$

For example, for $X \sim Binomial(n, p)$ the continuity correction method yields:

$$\begin{aligned} P(a \leq X \leq b) &= P(a - 0.5 \leq X \leq b + 0.5) \\ &\approx P\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

We do not use the correction when n is large.

Example. Suppose that Billy flips a fair coin 30 times. Using the continuity correction, approximate the probability that the number of heads is between 5 and 20. Compare the answer with the exact value.

Let X be the number of heads. As before, by the Normal approximation we have

$$X \sim N(\mu = np = 15, \sigma^2 = np(1-p) = 7.5).$$

Using the continuity correction, we get

$$\begin{aligned} P(5 \leq X \leq 20) &= P\left(\frac{4.5-15}{\sqrt{7.5}} \leq \frac{X-15}{\sqrt{7.5}} \leq \frac{20.5-15}{\sqrt{7.5}}\right) \\ &= P(-3.834 \leq Z \leq 2.008) \\ &= \Phi(2.008) - \Phi(-3.834) = 0.9776. \end{aligned}$$

The exact value was 0.9786.

Example. Suppose 80% of people who buy new car say they are satisfied with the car when surveyed one year after the purchase. Let X be the number of people in a group of n randomly chosen new car buyers who report satisfaction with their car.

(i) Determine the number n to ensure that there is a 95% chance that the proportion of satisfied with car in the sample is between 79% and 81%.

(ii) Let Y be the number of satisfied owners in a second (independent) survey of n randomly chosen new car buyers. Assuming that $n = 60$, find $P(|X - Y| \geq 3)$ using a suitable approximation. (A continuity correction is expected)

(i) The proportion of satisfied in the sample is X/n. By the Normal approximation

$$\frac{X}{n} \sim N(p, \frac{p(1-p)}{n}), \quad \text{with} \quad p = 0.8.$$

We want to find n that satisfies

$$\begin{aligned} 0.95 &= P(0.79 \leq \frac{X}{n} \leq 0.81) \\ &= P(\frac{0.79 - p}{\sqrt{p(1-p)/n}} \leq \frac{X/n - p}{\sqrt{p(1-p)/n}} \leq \frac{0.81 - p}{\sqrt{p(1-p)/n}}) \\ &= P(\frac{0.79 - 0.8}{\sqrt{0.8(1-0.8)/n}} \leq Z \leq \frac{0.81 - 0.8}{\sqrt{0.8(1-0.8)/n}}) \\ &= P(-0.025 \cdot \sqrt{n} \leq Z \leq 0.025 \cdot \sqrt{n}) = 2\Phi(0.025 \cdot \sqrt{n}) - 1 \end{aligned}$$

Thus

$$2\Phi(0.025 \cdot \sqrt{n}) - 1 = 0.95 \quad \Rightarrow \quad 0.025 \cdot \sqrt{n} = \Phi^{-1}(0.975)$$

Since

$$(\Phi^{-1}(0.975)/0.025)^2 = 6146.3,$$

n should be at least equal to 6147.

(ii) By the Normal approximation

$$X \sim N(np, np(1-p)) \quad \text{and} \quad Y \sim N(np, np(1-p))$$

Thus $X - Y \sim N(0, 2np(1-p))$, and

$$\begin{aligned} P(|X-Y| \geq 3) &= P(X-Y \geq 3 \;\text{ or }\; X-Y \leq -3) \\ &= P(X-Y \geq 3) + P(X-Y \leq -3) \\ &= P(X-Y \geq 2.5) + P(X-Y \leq -2.5) \\ &= P(\frac{X-Y}{\sqrt{2np(1-p)}} \geq \frac{2.5}{\sqrt{2np(1-p)}}) \\ &\quad +P(\frac{X-Y}{\sqrt{2np(1-p)}} \leq \frac{-2.5}{\sqrt{2np(1-p)}})) \\ &= P(Z \geq \frac{2.5}{\sqrt{2np(1-p)}}) + P(Z \leq \frac{-2.5}{\sqrt{2np(1-p)}})) \\ &= P(Z \geq 0.5705) + P(Z \leq -0.5705) \\ &= 0.2842 + 0.2842 = 0.5684. \end{aligned}$$

Clicker Question(s).

Moment Generating Functions

Definition. The **moment generating function** (m.g.f. or MGF) of a random variable X is given by

$$M_X(t) := E(e^{tX})$$

for arguments t for which the right-hand side is finite.

- If X is discrete with p.f. $f(x)$ then
$$M_X(t) = \sum_{x \in range(X)} e^{tx} f(x).$$
- If X is continuous with density $f(x)$, then
$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$
- We shall assume that the m.g.f. is finite in an interval around 0. Note that we always have $M_X(0) = 1$.

- MGF's of some standard distributions:

(i) If $X \sim B(n, p)$, then

$$M_X(t) = (pe^t + 1 - p)^n, \quad t \in \mathbb{R}.$$

(ii) If $X \sim Poi(\lambda)$, then

$$M_X(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}.$$

(iii) If $X \sim Exponential(\theta)$, then

$$M_X(t) = \frac{1}{1 - \theta t}, \quad t < \frac{1}{\theta}.$$

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

$$M_X(t) = e^{t\mu + t^2\sigma^2/2}, \quad t \in \mathbb{R}.$$

Exercise. Show that the m.g.f. of a r.v. X that has uniform discrete distribution on $\{1, 2, \ldots, 5\}$ is given by

$$M_X(t) = \frac{1}{5}[e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t}], \quad t \in \mathbb{R}.$$

Exercise. Show that the m.g.f. of a r.v. X that has uniform distribution on $(0, 5)$ is given by

$$M_X(t) = \frac{1}{5t}(e^{5t} - 1), \quad t \neq 0,$$

with $M_X(0) = 1$.

Clicker Question(s).

- Some of the reasons m.g.f.'s (and other transforms) are great tools to work with:
  - We can compute the following expected values[3]

    $$E(X^k), \quad k = 1, 2, \ldots,$$

    directly from the m.g.f. of X.
  - Moment generating functions uniquely identify the corresponding distributions.

    We can use this property to find, for example, the limiting distribution of a sequence of distributions.

    Or we can identify the distribution of some transformations of random variables.

[3]we call them the moments of the r.v. X

Theorem. Let the random variable X have moment generating function $M(t)$. Then

$$E(X^k) = M^{(k)}(0) \quad \text{for} \quad k = 1, 2, \ldots,$$

where

$$M^{(k)}(0) = \frac{d^k}{dt^k} M(t)|_{t=0}.$$

Proof. Assuming that the series converges, we have

$$\begin{aligned} M^{(k)}(t) &= \frac{d^k}{dt^k} \sum_x e^{tx} f(x) \\ &= \sum_x \frac{d^k}{dt^k} e^{tx} f(x) \\ &= \sum_x x^k e^{tx} f(x). \end{aligned}$$

If we evaluate $M^{(k)}(t)$ at zero, we get

$$M^{(k)}(0) = \sum_x x^k \cdot 1 \cdot f(x) = E(X^k).$$

Example. Find $E(X)$ and $Var(X)$ of a random variable

$$X \sim Exponential(\theta).$$

Exercise. Find $E(X)$ and $Var(X)$ of a r.v. X that has uniform discrete distribution on $\{1, 2, \ldots, 5\}$.

Uniqueness Theorem for Moment Generating Functions. If X and Y have MGF's M_X and M_Y respectively, and

$$M_X(t) = M_Y(t)$$

for all t for which both MGF's are finite, then X and Y have the same distribution.

Exercise. Find the distribution that corresponds to the following MGF:

$$M_X(t) = e^{-0.5t+t^2}, \quad t \in \mathbb{R}.$$

Exercise. Suppose that $X \sim N(\mu, \sigma^2)$. Show that for any constants a and b

$$Z := aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Clicker Question(s).

One more property of MGF's

Very useful property of MGF's:[4]

Theorem. Suppose that X and Y are independent r.v's with moment generating functions M_X and M_Y, respectively. Then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t),$$

for all t for which both sides are well defined.

Example. Suppose that $X \sim Poi(\lambda)$ and $Y \sim Poi(\mu)$, and X and Y are independent. Then

$$X + Y \sim Poi(\lambda + \mu).$$

[4] It is described in Section 10.3 and will not be covered on the final; but certainly you can use it if you want to.

Since

$$M_X(t) = e^{\lambda(e^t-1)} \quad \text{and} \quad M_Y(t) = e^{\mu(e^t-1)}, \qquad t \in \mathbb{R},$$

we have

$$M_{X+Y}(t) = e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}, \qquad t \in \mathbb{R},$$

which, by the uniqueness theorem, shows $X + Y \sim Poi(\lambda + \mu)$.

- Thus we have three methods of finding the distribution of a sum of two independent r.v.'s:
  - by using p.f.:

$$P(X+Y=t) = \sum_{x,y:x+y=t} f_X(x) f_Y(y),$$

  - by general reasoning,
  - by using MGF's.