Chapter 4
Jointly distributed Random variables

Multivariate distributions

Joint probability function

• For a Discrete RV, the joint probability function:
  \[ p(x, y) = P[X = x, Y = y] \]

• Marginal distributions
  \[ p_X(x) = \sum_y p(x, y) \quad p_Y(y) = \sum_x p(x, y) \]

• Conditional distributions
  \[ p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)} \quad p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)} \]
Joint probability function

- For a Continuous RV, the joint probability function:
  \[ f(x,y) = \text{Pf}[X = x, Y = y] \]

- Marginal distributions
  \[ f_X(x) = \int_{-\infty}^{\infty} f(x,y) \, dy \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) \, dx \]

- Conditional distributions
  \[ f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \]

Independence

Definition: Independence
Two random variables \( X \) and \( Y \) are defined to be independent if

- \( p(x,y) = p_X(x) \cdot p_Y(y) \) if \( X \) and \( Y \) are discrete
- \( f(x,y) = f_X(x) \cdot f_Y(y) \) if \( X \) and \( Y \) are continuous

Note: \[ p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p_X(x)p_Y(y)}{p_X(x)} = p_Y(y) \]
\[ p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x) \]

Thus, in the case of independence

**marginal distributions \( \equiv \) conditional distributions**
The Multiplicative Rule for densities

if $X$ and $Y$ are discrete

\[ p(x, y) = \begin{cases} p_X(x) p_{Y|X}(y|x) \\ p_Y(y) p_{X|Y}(x|y) \end{cases} \]

\[ = p_X(x) p_Y(y) \quad \text{if } X \text{ and } Y \text{ are independent} \]

if $X$ and $Y$ are continuous

\[ f(x, y) = \begin{cases} f_X(x) f_{Y|X}(y|x) \\ f_Y(y) f_{X|Y}(x|y) \end{cases} \]

\[ = f_X(x) f_Y(y) \quad \text{if } X \text{ and } Y \text{ are independent} \]

The Multiplicative Rule for densities

**Proof:**

This follows from the definition for conditional densities:

\[ p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)} \quad p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)} \]

Hence

\[ p(x, y) = p_X(x) p_{Y|X}(y|x) \]

and

\[ p(x, y) = p_Y(y) p_{X|Y}(x|y) \]

The same is true for continuous random variables.
Suppose that a rectangle is constructed by first choosing its length, $X$ and then choosing its width $Y$.

Its length $X$ is selected from an exponential distribution with mean $\mu = 1/\lambda = 5$. Once the length has been chosen its width, $Y$, is selected from a uniform distribution from 0 to half its length.

Find the probability that its area $A = XY$ is less than 4.

Solution: 

$$f_X(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \text{ for } x \geq 0$$

$$f_{Y|X}(y|x) = \frac{1}{x/2} \text{ if } 0 \leq y \leq x/2$$

$$f(x, y) = f_X(x) f_{Y|X}(y|x)$$

$$= \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \frac{1}{x/2} = \frac{2}{\lambda x} e^{-\frac{x}{\lambda}} \text{ if } 0 \leq y \leq x/2, x \geq 0$$
RS – 4 - Jointly distributed RV (b)

Finding CDFs (or pdfs) is difficult - Example

\[ P[X Y \leq 4] = \int_0^{2\sqrt{2}} \int_0^{\frac{4-x^2}{4}} f(x, y) \, dy \, dx + \int_{2\sqrt{2}}^{\infty} \int_0^{\frac{4-x^2}{4}} f(x, y) \, dy \, dx \]

This part can be evaluated

This part may require Numerical evaluation

\[ P[X Y \leq 4] = \int_0^{2\sqrt{2}} \int_0^{\frac{2}{\sqrt{3}}} e^{-\frac{x+y}{3}} \, dy \, dx + \int_{2\sqrt{2}}^{\infty} \int_0^{\frac{2}{\sqrt{3}}} e^{-\frac{x+y}{3}} \, dy \, dx \]

\[ = \frac{2}{\sqrt{3}} \int_0^{\frac{2}{\sqrt{3}}} e^{-\frac{x}{3}} \, \frac{1}{2} \, dx + \frac{2}{\sqrt{3}} \int_{2\sqrt{2}}^{\infty} e^{-\frac{x}{3}} \, \frac{1}{x} \, dx \]

\[ = \frac{1}{3} \int_0^{2\sqrt{2}} e^{-\frac{x}{3}} \, dx + \frac{8}{3} \int_{2\sqrt{2}}^{\infty} x^{-2} e^{-\frac{x}{3}} \, dx \]
Functions of Random Variables

Methods for determining the distribution of functions of Random Variables

Given some random variable $X$, we want to study some function $h(X)$.

$X$ is a transformation from $(Ω, Σ, P)$ into the probability space $(χ, A_X, P_X)$. Now, $Y$ is a transformation from $(χ, A_X, P_X)$ into the probability space $(Υ, A_Y, P_Y)$.

That is, we need to discover how the probability measure $P_Y$ relates to the measure $P_X$.

For some $F \in \mathcal{F}$, we have that

$$P_Y[F] = P_X\{x \in χ : Y(x) \in F\} = P\{ω \in Ω : g(X(ω)) \in F\} = P\{ω \in Ω : X(ω) \in h^{-1}(F)\} = P_X[h^{-1}(F)]$$
Methods for determining the distribution of functions of Random Variables

With non-transformed variables, we step "backwards" from the values of $X$ to the set of events in $\Omega$. In the transformed case, we take two steps backwards: 1) once from the range of the transformation back to the values of the $X$, and 2) again back to the set of events in $\Omega$.

Potential problem: The transformation $h(x)$ --we need to work with $h^{-1}(x)$-- may not yield unique results if $h(X)$ is not monotonic.

Methods to get $P_Y$:
1. Distribution function method
2. Transformation method
3. Moment generating function method

Method 1: Distribution function method

Let $X, Y, Z, \ldots$ have joint density $f(x,y,z,\ldots)$
Let $W = h(X, Y, Z, \ldots)$

First step
Find the distribution function of $W$
\[ G(w) = P[W \leq w] = P[h(X, Y, Z, \ldots) \leq w] \]

Second step
Find the density function of $W$
\[ g(w) = G'(w), \]
Distribution function method: Example 1 - $\chi^2_1$

Let $X$ have a normal distribution with mean 0, and variance 1 -i.e., a standard normal distribution. That is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Let $W = X^2$. Find the distribution of $W$.

Example 1: Chi-square - $\chi^2_1$

**First step**
Find the distribution function of $W$

$$G(w) = P[W \leq w] = P[X^2 \leq w]$$

$$= P[-\sqrt{w} \leq X \leq \sqrt{w}] \text{ if } w \geq 0$$

$$= \int_{-\sqrt{w}}^{\sqrt{w}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = F(\sqrt{w}) - F(-\sqrt{w})$$

where $$F'(x) = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$
Example 1: Chi-square - $\chi^2$

**Second step**

Find the density function of $W$

$$g(w) = G'(w).$$

$$= F'\left(\sqrt{w}\right) \frac{d}{dw}\sqrt{w} - F'\left(-\sqrt{w}\right) \frac{d}{dw}\left(-\sqrt{w}\right)$$

$$= f\left(\sqrt{w}\right) \frac{1}{2} w^{-\frac{1}{2}} + f\left(-\sqrt{w}\right) \frac{1}{2} w^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{w}{2}} \frac{1}{2} w^{-\frac{1}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{w}{2}} \frac{1}{2} w^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2\pi}} w^{-\frac{1}{2}} e^{-\frac{w}{2}} \text{ if } w \geq 0.$$

---

Example 1: Chi-square - $\chi^2$

Thus, if $X$ has a standard Normal distribution $\Rightarrow W = X^2$ follows

$$g(w) = \frac{1}{\sqrt{2\pi}} w^{-\frac{1}{2}} e^{-\frac{w}{2}} \text{ if } w \geq 0.$$

This distribution is the Gamma distribution with $\alpha = \frac{1}{2}$ and $\lambda = \frac{1}{2}$.

This distribution is the $\chi^2$ distribution with 1 degree of freedom.

- Using the additive properties of a gamma distribution, the sum of $T$ independent $\chi^2$ RVs produces a $\chi^2$ distributed RV. Alternatively, the sum of $T$ independent $N(0, 1)^2$ RVs produces a $\chi^2$ distributed RV.
- **Note:** If we add $T$ independent $N(\mu, \sigma^2)$ RVs, the $\sum_i (X_i/\sigma_i)^2$ follows a non-central $\chi^2$ distribution, with non-centrality parameter $\sum_i (\mu_i/\sigma_i)^2$. This distribution is common in the power analysis of statistical tests in which the null distribution is (asymptotically) a $\chi^2$. 
Suppose that $X$ and $Y$ are independent random variables each having an exponential distribution with parameter $\lambda$ \((=) E(X) = 1/\lambda\)

\[
f_1(x) = \lambda e^{-\lambda x} \quad \text{for} \quad x \geq 0
\]

\[
f_2(y) = \lambda e^{-\lambda y} \quad \text{for} \quad y \geq 0
\]

\[
f(x, y) = f_1(x)f_2(y) = \lambda^2 e^{-\lambda(x+y)} \quad \text{for} \quad x \geq 0, \quad y \geq 0
\]

Find the distribution of $W = X + Y$.
Example 2: Sum of Exponentials

\[ P[X + Y \leq w] = \int_0^w \int_0^{w-x} f_1(x) f_2(y) dy dx \]
\[ = \int_0^w \int_0^{w-x} \lambda^2 e^{-\lambda(x+y)} dy dx \]
Example 2: Sum of Exponentials

\[ P \{ X + Y \leq w \} = \int_0^w \int_0^{w-x} f_1(x) f_2(y) dy \, dx \]

\[ = \int_0^w \int_0^{w-x} \lambda^2 e^{-\lambda(x+y)} \, dy \, dx \]

\[ = \lambda^2 \int_0^w e^{-\lambda x} \left[ \int_0^{w-x} e^{-\lambda y} \, dy \right] \, dx \]

\[ = \lambda^2 \int_0^w e^{-\lambda x} \left[ \frac{e^{-\lambda y}}{-\lambda} \right]_{y=0}^{y=w-x} \, dx \]

\[ = \lambda^2 \int_0^w e^{-\lambda x} \left[ \frac{e^{-\lambda (w-x)} - e^0}{-\lambda} \right] \, dx \]

Example 2: Sum of Exponentials

\[ P \{ X + Y \leq w \} = \lambda^2 \int_0^w e^{-\lambda x} \left[ \frac{e^{-\lambda (w-x)} - e^0}{-\lambda} \right] \, dx \]

\[ = \lambda \int_0^w \left[ e^{-\lambda x} - e^{-\lambda w} \right] \, dx \]

\[ = \lambda \left[ \frac{e^{-\lambda x}}{-\lambda} - xe^{-\lambda w} \right]_0^w \]

\[ = \lambda \left[ \left( \frac{e^{-\lambda w}}{-\lambda} - we^{-\lambda w} \right) - \left( \frac{e^0}{-\lambda} \right) \right] \]

\[ = \lambda \left[ 1 - e^{-\lambda w} - \lambda we^{-\lambda w} \right] \]
Example 2: Sum of Exponentials

Second step
Find the density function of $W$

$$g(w) = G'(w).$$

$$= \frac{d}{dw} \left[ 1 - e^{-\lambda w} - \lambda we^{-\lambda w} \right]$$

$$= \left[ \lambda e^{-\lambda w} - \lambda \left( \frac{d}{dw} e^{-\lambda w} + w \frac{d}{dw} e^{-\lambda w} \right) \right]$$

$$= \left[ \lambda e^{-\lambda w} - \lambda e^{-\lambda w} + \lambda^2 we^{-\lambda w} \right]$$

$$= \lambda^2 we^{-\lambda w} \quad \text{for} \quad w \geq 0$$

Example 2: Sum of Exponentials

Hence if $X$ and $Y$ are independent random variables each having an exponential distribution with parameter $\lambda$ then $W$ has density

$$g(w) = \lambda^2 we^{-\lambda w} \quad \text{for} \quad w \geq 0$$

This distribution can be recognized to be the Gamma distribution with parameters $\alpha = 2$ and $\lambda$. 
Distribution function method: Example 3
Student’s t distribution

Let $Z$ and $U$ be two independent random variables with:
1. $Z$ having a Standard Normal distribution - i.e., $Z \sim N(0,1)$
   \[ f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \]
2. $U$ having a $\chi^2$ distribution with $\nu$ degrees of freedom
   \[ h(u) = \left(\frac{1}{2}\right)^{\frac{\nu}{2}} u^{\frac{\nu-2}{2}} e^{-\frac{u}{2}} \]

Find the distribution of $T = \frac{Z}{\sqrt{\frac{U}{\nu}}}$

Example 3: Student-t Distribution

Therefore the joint density of $Z$ and $U$ is:
\[
f(z, u) = f(z) h(u) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}} u^{\frac{\nu-2}{2}} e^{-\frac{z^2+u}{2}}}{\sqrt{2\pi} \Gamma\left(\frac{\nu}{2}\right)}
\]

The distribution function of $T$ is:
\[
G(t) = P[T \leq t] = P\left[\frac{Z}{\sqrt{\frac{U}{\nu}}} \leq t\right] = P\left[Z \leq \frac{t}{\sqrt{\nu}} \sqrt{U}\right] = \int_0^\infty \int_{-\infty}^{t \sqrt{\nu}} \frac{\left(1/2\right)^{\nu/2} u^{\nu-2/2} e^{-z^2+u/2}}{\sqrt{2\pi} \Gamma\left(\nu/2\right)} dz du
\]
Example 3: Student-t Distribution

\[ G(t) = \int_{0}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{\nu}{\Gamma(\nu/2)} u^{\nu/2-1} e^{-\frac{1}{2} \frac{z^2 + u}{u}} du dz \]

Then:

\[ g(t) = G'(t) = \int_{0}^{\infty} \frac{d}{dt} \left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{\nu}{\Gamma(\nu/2)} u^{\nu/2-1} e^{-\frac{1}{2} \frac{z^2 + u}{u}} du \right] dz du \]

Example 3: Student-t Distribution

Using:

\[ \frac{d}{dt} \int_{a}^{b} F(x, t) dx = \int_{a}^{b} \frac{d}{dt} F(x, t) dx \]

Using the FTC:

\[ F(x) = \int_{a}^{x} f(t) dt \Rightarrow F'(x) = f(x) \]

Then:

\[ g(t) = \int_{0}^{\infty} \frac{d}{dt} \left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{\nu}{\Gamma(\nu/2)} u^{\nu/2-1} e^{-\frac{1}{2} \frac{z^2 + u}{u}} du \right] dz du \]

\[ = \frac{1}{\sqrt{2\pi}} \frac{\nu}{\Gamma(\nu/2)} \int_{0}^{\infty} \frac{u^{\nu/2-1}}{2} e^{-\frac{u}{2} - \frac{z^2 + u}{2\nu}} \sqrt{u} du \]
Example 3: Student-t Distribution

Hence

\[ g(t) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\sqrt{2\pi} \sqrt{\nu} \Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty u^{\frac{\nu-1}{2}} e^{-\frac{\left(t^2\nu+1\right)u}{2}} du \]

Using

\[ 1 = \int_0^\infty \alpha^\alpha x^{\alpha-1} e^{-\lambda x} dx \Rightarrow \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha} \]

Thus,

\[ \int_0^\infty u^{\frac{\nu-1}{2}} e^{-\frac{\left(t^2\nu+1\right)u}{2}} du = \frac{\Gamma\left(\frac{\nu+1}{2}\right)2^{\frac{\nu+1}{2}}}{\left(\frac{t^2}{\nu}+1\right)^{\frac{\nu+1}{2}}} \]

or

\[ g(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi} \nu \Gamma\left(\frac{\nu}{2}\right)} \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}} = K \left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}} \]

where

\[ K = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi} \nu \Gamma\left(\frac{\nu}{2}\right)} \]

This is the Student's t distribution. Its denoted as \( t \sim t_\nu \)
where \( \nu \) is referred as degrees of freedom (df).

William Gosset (1876 - 1937)
Example 3: Student-t Distribution

Let

\[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \sqrt{n} \frac{\bar{x} - \mu}{\sigma} \sim N(0,1) \]

Let

\[ U = \frac{(n - 1)s^2}{\sigma^2} \sim \chi^2_{n-1} \]

Assume that Z and U are independent (later, we will learn how to check this assumption). Then,

\[ t = \frac{\sqrt{n} \frac{\bar{x} - \mu}{\sigma}}{\sqrt{(n - 1)s^2 / \sigma^2} / (n - 1)} = \frac{\sqrt{n} (\bar{x} - \mu)}{s} = \frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t_{n-1} \]
Let \( x_1, x_2, \ldots, x_n \) denote a sample of size \( n \) from the density \( f(x) \).
Find the distribution of \( M = \max(x_i) \).
Repeat this computation for \( m = \min(x_i) \)
Assume that the density is the uniform density from 0 to \( \theta \). Thus,
\[
f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}
\]
\[
F(x) = P[X \leq x] = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & 0 \leq x \leq \theta \\ 1 & x > \theta \end{cases}
\]

Example 4: Distribution of Max & Min

Finding the distribution function of \( M \).
\[
G(t) = P[M \leq t] = P[\max(x_i) \leq t]
= P[x_1 \leq t, \ldots, x_n \leq t]
= P[x_1 \leq t] \cdots P[x_n \leq t]
= \begin{cases} 0 & t < 0 \\ \left(\frac{t}{\theta}\right)^n & 0 \leq t \leq \theta \\ 1 & t > \theta \end{cases}
\]
Example 4: Distribution of Max & Min

Differentiating we find the density function of \( M \).

\[
g(t) = G'(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & 0 \leq t \leq \theta \\ 0 & \text{otherwise} \end{cases}
\]

Example 4: Distribution of Max & Min

Finding the distribution function of \( m \).

\[
G(t) = P[m \leq t] = P\left[ \min(x_i) \leq t \right]
\]

\[
= 1 - P[x_1 > t, \ldots, x_n > t]
\]

\[
= 1 - P[x_1 > t] \cdots P[x_n > t]
\]

\[
= \begin{cases} 
0 & t < 0 \\
1 - \left(1 - \frac{t}{\theta}\right)^n & 0 \leq t \leq \theta \\
1 & t > \theta
\end{cases}
\]
Example 4: Distribution of Max & Min

Differentiating we find the density function of \( m \).

\[
g(t) = G'(t) = \begin{cases} \frac{n}{\theta} \left(1 - \frac{t}{\theta}\right)^{n-1} & 0 \leq t \leq \theta \\ 0 & \text{otherwise} \end{cases}
\]

The Probability Integral Transformation

Theorem: Let \( U \sim \text{Uniform}(0,1) \) and \( F \) is a CDF, then a RV with CDF \( F \) can be generated as \( F^{-1}(U) \), where \( F^{-1} \) is the inverse of \( F \), if it exists (more generally, it is the quantile function for \( F \)).

This transformation allows one to convert observations that come from a uniform distribution from 0 to 1 (easy to generate using a standard function) to observations that come from an arbitrary pdf.

Let \( U \) denote an observation having a \( U(0,1) \):

\[
g(u) = \begin{cases} 1 & 0 \leq u \leq 1 \\ 0 & \text{elsewhere} \end{cases}
\]

Let \( f(x) \) denote an arbitrary pdf and \( F(x) \) its corresponding CDF. Let \( X = F^{-1}(U) \); we want to find the distribution of \( X \).
The Probability Integral Transformation

Now, we go from the $u$'s – i.e., from $(0,1)$- to the $x$'s.

Find the distribution of $X$.

$G(x) = P[X \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = F(x)$ (because $U$ is uniform.)

Hence: $g(x) = G'(x) = F'(x) = f(x)$

Thus if $U \sim \text{Uniform}(0,1)$, then,

$X = F^{-1}(U)$ has density $f(x)$. 

X=F^{-1}(U)
The Probability Integral Transformation

- The goal of some estimation methods is to simulate an expectation, say \( E[h(Z)] \). To do this, we need to simulate \( Z \) from its distribution. The probability integral transformation is very handy for this task.

**Example:** Exponential distribution (\( X \sim \text{Exponential}(\lambda) \)).

Let \( U \sim \text{Uniform}(0,1) \) and \( F(x) = 1 - \exp(-\lambda x) \).

\[
X = -\log(1 - U)/\lambda \sim F \text{ (exponential distribution)}
\]

Code in R

```r
> set.seed(90)
> lambda <- 3
> U <- runif(500,0,1)
> XX <- -log(1-U)/lambda
> mean(XX)
[1] 0.3127794
```

**Example** (continuation):

hist(XX, main="Histogram for Simulated Draws",xlab="Freq", breaks=20)
The Probability Integral Transformation

**Example:** If $F$ is the standard normal, $F^{-1}$ has no closed form solution. Most computer programs have a routine to approximate $F^{-1}$ for the standard normal distribution. We can use this to simulate other distributions—for example, truncated normals.

Say, we want to sample from $X \sim N(\mu, \sigma^2)$ with truncation points $a$ & $b$. Suppose we can evaluate points from the untruncated inverse normal cdf $F^{-1}$. Then,

1. Calculate endpoints $p_a = F(a)$ & $p_b = F(b)$.
2. Sample $U \sim \text{Unif}(p_a; p_b)$.
3. Set $X = F^{-1}(U)$.

Steps (1) and (2) avoid calculating the normalizing constant for the pdf. -i.e., $1/[F^{-1}(b) - F^{-1}(a)]$.

The Probability Integral Transformation

**Example:** Code in R ($a=-1; b=5, \mu=0, \sigma=1$)

```r
> set.seed(90)
> pa <- pnorm(-1)
> pb <- pnorm(5)
> U <- runif(500,pa,pb)
> XX <- qnorm(U)  # inverse normal from mu=0, sd=1
> mean(XX)
[1] 0.2416027
> hist(XX, main="Histogram for Simulated Draws", xlab="Observations", breaks=20)
```

![Histogram for Simulated Draws](image-url)
Method 2: The Transformation Method

Theorem

Let $X$ denote a RV with pdf $f(x)$ and $U = h(X)$.

Assume that $h(x)$ is either strictly increasing (or decreasing) then the probability density of $U$ is:

$$g(u) = f(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right| = f(x) \left| \frac{dx}{du} \right|$$

Proof

Use the distribution function method.

Step 1 Find the distribution function, $G(u)$

Step 2 Differentiate $G(u)$ to find the probability density function $g(u)$

Method 2: The Transformation Method

$$G(u) = P[U \leq u] = P[h(X) \leq u]$$

$$= \begin{cases} P[X \leq h^{-1}(u)] & h \text{ strictly increasing} \\ P[X \geq h^{-1}(u)] & h \text{ strictly decreasing} \end{cases}$$

$$= \begin{cases} F(h^{-1}(u)) & h \text{ strictly increasing} \\ 1 - F(h^{-1}(u)) & h \text{ strictly decreasing} \end{cases}$$

Thus, $g(u) = G'(u)$

$$= \begin{cases} F'(h^{-1}(u)) \frac{dh^{-1}(u)}{du} & h \text{ strictly increasing} \\ -F'(h^{-1}(u)) \frac{dh^{-1}(u)}{du} & h \text{ strictly decreasing} \end{cases}$$
Method 2: The Transformation Method

\[ g(u) = G'(u) \]

\[
= \begin{cases} 
F'(h^{-1}(u)) \frac{dh^{-1}(u)}{du} & \text{if } h \text{ strictly increasing} \\
-F'(h^{-1}(u)) \frac{dh^{-1}(u)}{du} & \text{if } h \text{ strictly decreasing}
\end{cases}
\]

Or,

\[ g(u) = f(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right| = f(x) \left| \frac{dx}{du} \right| \]

Example 1: Log Normal Distribution

Suppose that \( X \sim \text{N}(\mu, \sigma^2) \). Find the distribution of \( U = h(x) = e^x \).

**Solution:** Recall transformation formula:

\[ g(u) = f(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right| = f(x) \left| \frac{dx}{du} \right| \]

\[ f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ h^{-1}(u) = \ln(u) \quad \text{and} \quad \frac{dh^{-1}(u)}{du} = \frac{d\ln(u)}{du} = \frac{1}{u} \]
Example 1: Log Normal Distribution

Replacing in transformation formula:

\[
g(u) = f \left(h^{-1}(u)\right) \left|\frac{dh^{-1}(u)}{du}\right| = f \left(x\right) \left|\frac{dx}{du}\right|
\]

\[
= \frac{1}{\sqrt{2\pi} \sigma} \frac{1}{u} e^{-\frac{(\ln(u) - \mu)^2}{2\sigma^2}} \text{ for } u > 0
\]

Since the distribution of \(\log(U)\) is normal, the distribution of \(U\) is called the log-normal distribution.

Note: It is easy to obtain the mean of a log-normal variable.

\[
E[u] = \int_{0}^{\infty} u g(u) du = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\ln(u) - \mu)^2}{2\sigma^2}} du
\]

Example 1: Log Normal Distribution

\[
E[u] = \int_{0}^{\infty} u g(u) du = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\ln(u) - \mu)^2}{2\sigma^2}} du
\]

Substitution: \(y = (\ln(u) - \mu)/\sigma \Rightarrow u = e^{y+\mu}\)

\[
dy \sigma = (1/\sigma)du \Rightarrow du = \sigma e^{y+\mu} dy.
\]

Then,

\[
E[u] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(y+\mu)^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{y^2}{2\sigma^2} + y\sigma} dy
\]

\[
= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{y^2}{2} + y\sigma} dy
\]

\[
= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{y^2}{2} + y\sigma} e^{-\frac{\sigma^2}{2}} dy = e^{\mu + \frac{\sigma^2}{2}}
\]
Example 1: Log Normal Distribution - Graph

In finance, a popular assumption is that stock prices follow a log-normal distribution. We have the following inks between normal and log-normal variables.

<table>
<thead>
<tr>
<th>Asset (Price)</th>
<th>$S_T$</th>
<th>$\ln(S_T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Return)</td>
<td>$(S_T - S_{T-1})/1$</td>
<td>$\ln(S_T) - \ln(S_{T-1})$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Lognormal</th>
<th>Normal</th>
</tr>
</thead>
</table>

- First moment: $\exp(\mu + 0.5 \sigma)$, $\mu$
- Second moment: $\exp(2\mu + 2 \sigma)$, $\sigma$
Example 1: Log Normal Distribution

**Example:** We value a call option, \( c \), by discounting the expected option payoff at expiration, using \( r \) as the discount rate, where the expectation is calculated in a *risk-neutral* world (under “Q measure”):

\[
c = e^{-rT} E^Q \left[ \max (S_T - K, 0) \right]
\]

- Under the Black-Scholes model, \( S \) is lognormally distributed, then \( \ln(S) \) follows a normal distribution. Then,

\[
\ln(S_T) \sim N(\mu, \sigma^2) \quad \Rightarrow \quad E[S_T] = e^{\mu + \frac{\sigma^2}{2}}
\]

\[
\Rightarrow \quad \text{Var}[S_T] = e^{2\mu + \sigma^2}
\]

where \( \mu \) is expected return and \( \sigma \) is volatility. Rates of return—i.e., \( \ln(S_T) - \ln(S_t) \)—are also normally distributed!

We want to value a call option, \( c \), with a strike price \( K \):

\[
c = e^{-rT} E^Q \left[ \max (S_T - K, 0) \right]
\]

Divided into two terms:

\[
c = e^{-rT} E^Q[S_T | S_T > K] - e^{-rT} E^Q[K | S_T > K]
\]

Looking at the second part:

\[
E^Q[K | S_T > K] = K \text{ Prob } [\ln S_T > \ln K]
\]

\[
= K \text{ Prob } [(\ln S_T - \mu)/\sigma > (\ln K - \mu)/\sigma]
\]

\[
= K \text{ Prob } [Z_T > (\ln K - \mu)/\sigma]
\]

\[
= K N(d_2)
\]

The first part is more complicated and relies on the following result:

If \( \ln(S_T) \sim N(\mu, \sigma) \), then \( E[S_T | S_T > K] = E[S_T] N(\sigma - (\ln(K) - \mu)/\sigma) \).

Then,

\[
E^Q[S_T | S_T > K] = E^Q[S_T] N(\sigma - (\ln(K) - \mu)/\sigma)
\]

\[
= E^Q[S_T] N(d_2 + \sigma)
\]

\[
= E^Q[S_T] N(d_1) = \exp(\mu + 0.5 \sigma) N(d_1)
\]
Example 2: Inverse Gamma Distribution

Suppose that $X$ has a Gamma distribution with $\alpha$ and $\lambda$ parameters. Find the distribution of $U = h(x) = 1/X$.

Solution: Recall transformation formula:

$$
    g (u) = f \left( h^{-1}(u) \right) \left| \frac{d h^{-1}(u)}{du} \right| = f \left( x \right) \left| \frac{dx}{du} \right|
$$

$$
    f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x > 0.
$$

$h^{-1}(u) = 1/u$ and \( \frac{dh^{-1}(u)}{du} = -\frac{1}{u^2} \quad x > 0.$

Example 2: Inverse Gamma Distribution

Replacing in transformation formula:

$$
    g (u) = f \left( h^{-1}(u) \right) \left| \frac{d h^{-1}(u)}{du} \right| = f \left( x \right) \left| \frac{dx}{du} \right|
$$

$$
    = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left( \frac{1}{u} \right)^{\alpha-1} e^{-\lambda \frac{1}{u}} \frac{1}{u^2} = \frac{\lambda^\alpha}{\Gamma(\alpha)} u^{-\alpha-1} e^{-\frac{\lambda}{u}} \quad u > 0.
$$

Since the distribution of $X$ is gamma, the distribution of $U$ is called the inverse-gamma distribution.

It’s straightforward to calculate its mean:

$$
    E[u] = \int_0^\infty u g(u) du = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} u^{-\alpha-1} e^{-\frac{\lambda}{u}} du
$$

$$
    = \frac{\lambda}{(\alpha-1) \Gamma(\alpha-1)} \int_0^\infty u^{-\alpha} e^{-\frac{\lambda}{u}} du = \frac{\lambda}{(\alpha-1)} \quad (\alpha > 1).
$$
In Bayesian econometrics, it is common to assume that $\sigma^2$ follows an inverse gamma distribution ($\sigma^{-2}$ follows a gamma!). Usual values for $(\alpha, \lambda)$: $\alpha = T/2$ and $\lambda = 1/(2\eta^2) = \Phi/2$, where $\eta^2$ is related to the variance of the $T N(0, \eta^2)$ variables we are implicitly adding.

Example 2: Inverse Gamma Distribution - Graph
Method 3: Use of moment generating functions

Review: MGF - Definition

Let $X$ denote a random variable with probability density function $f(x)$ if continuous (probability mass function $p(x)$ if discrete). Then

$$m_X(t) = \text{the moment generating function of } X$$

$$= E \left( e^{tX} \right)$$

$$= \begin{cases} 
\int_{-\infty}^{\infty} e^{tx} f(x) \, dx & \text{if } X \text{ is continuous} \\
\sum_{x} e^{tx} p(x) & \text{if } X \text{ is discrete}
\end{cases}$$

The distribution of a random variable $X$ is described by either

1. The density function $f(x)$ if $X$ continuous (probability mass function $p(x)$ if $X$ discrete), or
2. The cumulative distribution function $F(x)$, or
3. The moment generating function $m_X(t)$
Review: MGF - Properties

1. \( m_X(0) = 1 \)

2. \( m_X^{(k)}(0) = k^{th} \) derivative of \( m_X(t) \) at \( t = 0 \).
   \[
   \mu_k = E(X^k) = \begin{cases} 
   \int x^k f(x) \, dx & \text{X continuous} \\
   \sum x^k p(x) & \text{X discrete}
   \end{cases}
   \]

3. \( m_X(t) = 1 + \mu_1 t + \frac{\mu_2}{2!} t^2 + \frac{\mu_3}{3!} t^3 + \cdots + \frac{\mu_k}{k!} t^k + \cdots \).

4. Let \( X \) be a RV with MGF \( m_X(t) \). Let \( Y = bX + a \). Then, \( m_Y(t) = m_{bX + a}(t) = E(e^{(bX + a)t}) = e^{at} m_X(bt) \)

5. Let \( X \) and \( Y \) be two independent random variables with moment generating function \( m_X(t) \) and \( m_Y(t) \). Then \( m_{X+Y}(t) = m_X(t) m_Y(t) \).

6. Let \( X \) and \( Y \) be two random variables with moment generating function \( m_X(t) \) and \( m_Y(t) \) and two distribution functions \( F_X(x) \) and \( F_Y(y) \) respectively. If \( m_X(t) = m_Y(t) \), then \( F_X(x) = F_Y(x) \). This ensures that the distribution of a random variable can be identified by its moment generating function.

7. \( m_{X+b}(t) = e^{bt} m_X(t) \)
Using of moment generating functions to find the distribution of functions of Random Variables

**Example: Sum of 2 independent exponentials**

Suppose that $X$ and $Y$ are two independent distributed exponential random variables with pdf’s given by

\[
f(x) = \lambda e^{-\lambda x}
\]

\[
f(y) = \lambda e^{-\lambda y}
\]

**Solution:**

\[
m_X(t) = \left(\frac{\lambda}{\lambda - t}\right) = (1 - \frac{t}{\lambda})^{-1}
\]

\[
m_Y(t) = (1 - \frac{t}{\lambda})^{-1}
\]

\[
m_{X+Y}(t) = m_X(t)m_Y(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}\left(1 - \frac{t}{\lambda}\right)^{-1} = \left(1 - \frac{t}{\lambda}\right)^{-2}
\]

= MGF of the gamma distribution with $\alpha=2$. 
Example: Affine Transformation of a Normal

• Suppose that $X \sim N(\mu, \sigma^2)$. What is the distribution of $Y = aX + b$?

• Solution:

$$m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$$m_{aX+b}(t) = e^{bt} m_X(at) = e^{bt} e^{\mu(at) + \frac{\sigma^2(at)^2}{2}}$$

$$= e^{(a\mu+b)t + \frac{a^2\sigma^2 t^2}{2}}$$

= MGF of the normal distribution with mean $a\mu + b$ and variance $a^2\sigma^2$.

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

Example: Affine Transformation of a Normal

Special Case: the $z$ transformation

$$Z = \frac{X - \mu}{\sigma} = \left(\frac{1}{\sigma}\right) X + \left(-\frac{\mu}{\sigma}\right) = aX + b$$

$$\mu_Z = a\mu + b = \left(\frac{1}{\sigma}\right) \mu + \left(-\frac{\mu}{\sigma}\right) = 0$$

$$\sigma^2_Z = a^2\sigma^2 = \left(\frac{1}{\sigma}\right)^2 \sigma^2 = 1$$

Thus $Z$ has a standard normal distribution.
Example 1: Distribution of $X+Y$

Suppose that $X$ and $Y$ are independent. Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Find the distribution of $S = X + Y$.

Solution:

$$m_X(t) = e^{\mu_X t + \frac{\sigma_X^2 t^2}{2}} \quad m_Y(t) = e^{\mu_Y t + \frac{\sigma_Y^2 t^2}{2}}$$

Now,

$$m_{X+Y}(t) = m_X(t) m_Y(t) = e^{\mu_X t + \frac{\sigma_X^2 t^2}{2}} e^{\mu_Y t + \frac{\sigma_Y^2 t^2}{2}}$$

$$m_{X+Y}(t) = e^{(\mu_X + \mu_Y)t + \frac{(\sigma_X^2 + \sigma_Y^2)t^2}{2}}$$

Thus, $Y = X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Example 2: Distribution of $aX + bY$

Suppose that $X$ and $Y$ are independent. Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Find the distribution of $L = aX + bY$.

Solution:

$$m_X(t) = e^{\mu_X t + \frac{\sigma_X^2 t^2}{2}} \quad m_Y(t) = e^{\mu_Y t + \frac{\sigma_Y^2 t^2}{2}}$$

Now,

$$m_{aX+bY}(t) = m_{aX}(t) m_{bY}(t) = m_X(at) m_Y(bt)$$

$$= e^{\mu_X(at) + \frac{\sigma_X^2 (at)^2}{2}} e^{\mu_Y(bt) + \frac{\sigma_Y^2 (bt)^2}{2}}$$

$$m_{aX+bY}(t) = e^{(a\mu_X + b\mu_Y)t + \frac{(a^2 \sigma_X^2 + b^2 \sigma_Y^2)t^2}{2}}$$

Thus, $Y = aX + bY \sim \mathcal{N}(a\mu_X + b\mu_Y, a^2 \sigma_X^2 + b^2 \sigma_Y^2)$. 

Example 2: Distribution of $aX + bY$

Special Case:
$a = +1$ and $b = -1$.

Thus $Y = X - Y$ has a normal distribution with mean $\mu_X - \mu_Y$ and variance:
\[
(+1)^2 \sigma_X^2 + (-1)^2 \sigma_Y^2 = \sigma_X^2 + \sigma_Y^2
\]

Example 3: (Extension to $n$ independent RV’s)

Suppose that $X_1, X_2, \ldots, X_n$ are independent each having a normal distribution with means $\mu_i$, standard deviations $\sigma_i$ (for $i = 1, 2, \ldots, n$)

Find the distribution of $L = a_1X_1 + a_1X_2 + \ldots + a_nX_n$

Solution:
\[
m_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}} \quad (\text{for } i = 1, 2, \ldots, n)
\]

Now
\[
m_{a_1X_1 + \ldots + a_nX_n}(t) = m_{a_1X_1}(t) \cdots m_{a_nX_n}(t)
\]
\[
= m_{X_1}(a_1 t) \cdots m_{X_n}(a_n t)
\]
\[
= e^{\mu_1(a_1 t) + \frac{\sigma_1^2 (a_1 t)^2}{2}} \cdots e^{\mu_n(a_n t) + \frac{\sigma_n^2 (a_n t)^2}{2}}
\]
Example 3: (Extension to \( n \) independent RV's)

or

\[
m_{a_1X_1 + \cdots + a_nX_n}(t) = e^{(a_1\mu_1 + \cdots + a_n\mu_n)t + \frac{(a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2)t^2}{2}}
\]

= the MGF of the normal distribution

with mean \( a_1\mu_1 + \cdots + a_n\mu_n \)

and variance \( a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \)

Thus \( Y = a_1X_1 + \cdots + a_nX_n \) has a normal distribution with mean \( a_1\mu_1 + \cdots + a_n\mu_n \) and variance \( a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \).

---

Example 3: (Extension to \( n \) independent RV's)

Special case:

\[
a_1 = a_2 = \cdots = a_n = \frac{1}{n}
\]

\[
\mu_1 = \mu_2 = \cdots = \mu_n = \mu
\]

\[
\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2
\]

In this case \( X_1, X_2, \ldots, X_n \) is a sample from a normal distribution with mean \( \mu \) and standard deviations \( \sigma \), and

\[
L = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \bar{X} = \text{the sample mean}
\]
Thus \[ Y = \bar{x} = a_1 x_1 + \ldots + a_n x_n = \left( \frac{1}{n} \right) x_1 + \ldots + \left( \frac{1}{n} \right) x_n \]
has a normal distribution with mean
\[ \mu_Y = a_1 \mu_1 + \ldots + a_n \mu_n = \left( \frac{1}{n} \right) \mu + \ldots + \left( \frac{1}{n} \right) \mu = \mu \]
and variance
\[ \sigma^2_Y = a_1^2 \sigma^2_1 + \ldots + a_n^2 \sigma^2_n \]
\[ = \left( \frac{1}{n} \right)^2 \sigma^2 + \ldots + \left( \frac{1}{n} \right)^2 \sigma^2 = n \left( \frac{1}{n} \right)^2 \sigma^2 = \frac{\sigma^2}{n} \]

**Summary**
Let \( x_1, x_2, \ldots, x_n \) be a sample from a normal distribution with mean \( \mu \),
and standard deviations \( \sigma \). Then,
\[ \bar{X} \sim \text{N}(\mu, \sigma^2 / n) \]
Application: Simple Regression Framework

Let the random variable $Y_i$ be determined by:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, 2, ..., n,$$

where the $X_i$'s are (exogenous or pre-determined) numbers, $\alpha$ and $\beta$ constants. Let the random variable $\varepsilon_i$ follow a normal distribution with zero mean and constant variance. That is,

$$\varepsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, ..., n.$$

Then,

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i,$$

$$Var(Y_i) = Var(\alpha + \beta X_i + \varepsilon_i) = Var(\varepsilon_i) = \sigma^2.$$

Since $Y$ is a linear function of a normal RV $\Rightarrow Y_i \sim N(\alpha + \beta X_i, \sigma^2)$.

Central and Non-central Distributions

- **Noncentrality parameters** are parameters of families of probability distributions which are related to other central families of distributions. If the noncentrality parameter of a distribution is zero, the distribution is identical to a distribution in the central family.

For example, the standard Student's $t$-distribution is the central family of distributions for the Noncentral $t$-distribution family.

- Noncentrality parameters are used in the following distributions:
  - $t$-distribution
  - $F$-distribution
  - $\chi^2$ distribution
  - $\chi$ distribution
  - Beta distribution
Central and Non-central Distributions

• In general, noncentrality parameters occur in distributions that are transformations of a normal distribution. The central versions are derived from zero-mean normal distributions; the noncentral versions generalize to arbitrary means.

Example: The standard (central) $\chi^2$ distribution is the distribution of a sum of squared independent $N(0, 1)$. The noncentral $\chi^2$ distribution generalizes this to $N(\mu, \sigma^2)$.

• There are extended versions of these distributions with two noncentrality parameters: the doubly noncentral beta distribution, the doubly noncentral $F$-distribution and the doubly noncentral $t$-distribution.

Central and Non-central Distributions

• There are extended versions of these distributions with two noncentrality parameters. This happens for distributions that are defined as the quotient of two independent distributions.

• When both source distributions are central, the result is a central distribution. When one is noncentral, a (singly) noncentral distribution results, while if both are noncentral, the result is a doubly noncentral distribution.

Example: A $t$-distribution is defined (ignoring constants) as the ratio of a $N(0,1)$ RV and the square root of an independent $\chi^2$ RV. Both RV can have noncentral parameters. This produces a doubly noncentral $t$-distribution.
The Central Limit Theorem (CLT): Preliminaries

The proof of the CLT is very simple using moment generating functions. We will rely on the following result:

Let \( m_1(t), m_2(t), \ldots \), be a sequence of moment generating functions corresponding to the sequence of distribution functions: \( F_1(x), F_2(x), \ldots \)

Let \( m(t) \) be a moment generating function corresponding to the distribution function \( F(x) \).

Then, if  \[ \lim_{i \to \infty} m_i(t) = m(t) \quad \text{for all } t \text{ in an interval about } 0. \]

then  \[ \lim_{i \to \infty} F_i(x) = F(x) \quad \text{for all } x. \]

The Central Limit Theorem (CLT)

Let \( x_1, x_2, \ldots, x_n \) be a sequence of independent and identically distributed RVs with finite mean \( \mu \), and finite variance \( \sigma^2 \). Then as \( n \) increases, \( \bar{x} \), the sample mean, approaches the normal distribution with mean \( \mu \) and variance \( \sigma^2/n \).

This theorem is sometimes stated as  \[ \frac{\sqrt{n} (\bar{x} - \mu)}{\sigma} \xrightarrow{d} N(0,1) \]

where \( \xrightarrow{d} \) means “the limiting distribution (asymptotic distribution) is” (or convergence in distribution).

Note: This CLT is sometimes referred as the Lindeberg-Lévy CLT.
The Central Limit Theorem (CLT)

**Proof:** (use moment generating functions)
Let $X_1, X_2, \ldots$ be a sequence of independent random variables from a distribution with moment generating function $m(t)$ and CDF $F(x)$.

Let $S_n = X_1 + X_2 + \ldots + X_n$ then
\[
m_{S_n}(t) = m_{X_1 + X_2 + \ldots + X_n}(t) = m_{X_1}(t)m_{X_2}(t)\ldots m_{X_n}(t) = [m(t)]^n
\]

now $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{S_n}{n}$

or $m_{\overline{X}}(t) = m_{\left(\frac{1}{n}\right)S_n}(t) = m_{S_n}(t^*n) = m\left(\frac{t}{n}\right)^n$

The Central Limit Theorem (CLT)

Let $z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}\overline{X} - \sqrt{n}\mu}{\sigma}$

then $m_z(t) = e^{-\frac{\sqrt{n}\mu}{\sigma}t} m_{\overline{X}}\left(\frac{\sqrt{n}t}{\sigma}\right) = e^{-\frac{\sqrt{n}\mu}{\sigma}t} \left[ m\left(\frac{\sqrt{n}t}{\sigma}\right) \right]^n$

and $\ln \left[ m_z(t) \right] = -\frac{\sqrt{n}\mu}{\sigma}t + n \ln \left[ m\left(\frac{t}{\sigma/\sqrt{n}}\right) \right]$

Let $u = \frac{t}{\sigma/\sqrt{n}}$ or $\sqrt{n} = \frac{t}{u}$ and $n = \frac{t^2}{u^2}$

\[
\ln \left[ m_u(t) \right] = -\frac{t^2\mu}{\sigma^2u} + \frac{t^2}{\sigma^2u^2} \ln \left[ m(u) \right] = \frac{t^2}{u^2} \ln \left[ m(u) \right] - \mu u
\]
The Central Limit Theorem (CLT)

\[
\ln \left[ m_z(t) \right] = \frac{t^2}{\sigma^2} \ln \left[ m(u) \right] - \mu u
\]

Now \( \lim_{n \to \infty} \left( \ln \left[ m_z(t) \right] \right) = \lim_{n \to 0} \left( \ln \left[ m_z(t) \right] \right) \)

\[
= \frac{t^2}{\sigma^2} \lim_{u \to 0} \frac{\ln \left[ m(u) \right] - \mu u}{u^2}
\]

\[
= \frac{t^2}{\sigma^2} \lim_{u \to 0} \frac{m'(u)}{m(u)} - \frac{\mu}{2u}
\]

using L'Hopital's rule

\[
= \frac{t^2}{\sigma^2} \lim_{u \to 0} \frac{m''(u)m(u) - \left[ m'(u) \right]^2}{2} \]

using L'Hopital's rule again

\[
= \frac{t^2}{\sigma^2} \lim_{u \to 0} \frac{\left[ m(u) \right]^2}{2}
\]

The Central Limit Theorem (CLT)

\[
= \frac{t^2}{\sigma^2} \frac{m''(0) - \left[ m'(0) \right]^2}{2} = \frac{t^2}{\sigma^2} \frac{E(x_i^2) - \left[ E(x_i) \right]^2}{2} = \frac{t^2}{2}
\]

thus \( \lim_{n \to \infty} \left( \ln \left[ m_z(t) \right] \right) = \frac{t^2}{2} \) and \( \lim_{n \to \infty} \left( m_z(t) \right) = e^{\frac{t^2}{2}} \)

Now \( m(t) = e^{\frac{t^2}{2}} \)

• This is the moment generating function of the standard normal distribution.

• Thus, the limiting distribution of \( z \) is the standard normal distribution:

\[ \text{i.e.} \quad \lim_{n \to \infty} F_z(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du \quad \text{Q.E.D.} \]
**CLT: Asymptotic Distribution**

The CLT states that the *asymptotic distribution* of the sample mean is a normal distribution.

Q: What is meant by “asymptotic distribution”? In general, the “asymptotic distribution of $X_n$” means a non-constant RV $X$, along with real-number sequences $\{a_n\}$ and $\{b_n\}$, such that $a_n(X_n - b_n)$ has as limiting distribution the distribution of $X$.

In this case, the distribution of $X$ might be referred to as the *asymptotic* or *limiting distribution* of either $X_n$ or of $a_n(X_n - b_n)$, depending on the context.

The real number sequence $\{a_n\}$ plays the role of a stabilizing transformation to make sure the transformed RV –i.e., $a_n(X_n - b_n)$– does not have zero variance as $n \to \infty$.

Example: $\bar{X}$ has zero variance as $n \to \infty$. But, $n^{1/2} \bar{X}$ has a finite variance, $\sigma^2$.

**CLT: Remarks**

- The CLT gives only an asymptotic distribution. As an approximation for a finite number of observations, it provides a reasonable approximation only when the observations are close to the mean; it requires a very large number of observations to stretch it into the tails.

- The CLT also applies in the case of sequences that are not identically distributed. Extra conditions need to be imposed on the RVs.

- Lindeberg found a condition on the sequence $\{X_n\}$, which guarantees that the distribution of $S_n$ is asymptotically normally distributed. W. Feller showed that Lindeberg’s condition is necessary as well (if the condition does not hold, then the sum $S_n$ is not asymptotically normally distributed).

- A sufficient condition that is stronger (but easier to state) than Lindeberg’s condition is that there exists a constant $A$, such that $|X_n| < A$ for all $n$. 


RS – 4 - Jointly distributed RV (b)

Jarl W. Lindeberg (1876 – 1932)
Vilibald S. (Willy) Feller (1906-1970)
Paul Pierre Lévy (1886–1971)