Section:

Subsection: Introduction:
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function.
   \[ p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}. \]

2. Continuous case: get joint probability density function divided by marginal probability density function:
   \[ f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \]

3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.

4. Example: Two instruments produce measurements \( X, Y \).
Conditional distributions:

Discrete case: Just joint probability function divided by marginal probability function. 
\[ p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_{Y}(y)}. \]
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \).

2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \).

In both cases, for independent variables, conditional probability function or probability density function is same as marginal.

Example: Two instruments produce measurements \( X_1, X_2 \).
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \).

2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \).

3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.

Example: Two instruments produce measurements \( X, 2 \)
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = p_{X,Y}(x,y)/p_Y(y) \).
2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y) \).
3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
4. Example: Two instruments produce measurements \( X \),
1 Conditional distributions:
   1 Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \).
   2 Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \).
   3 In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
   4 Example: Two instruments produce measurements \( X \),

2 Moment Generating Function Summary
1. Conditional distributions:

   1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = p_{X,Y}(x,y)/p_Y(y) \).
   2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y) \).
   3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
   4. Example: Two instruments produce measurements \( X \),

2. Moment Generating Function Summary

   1. General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \):
      \[
      m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \quad \text{for} \quad t \in \mathcal{R}.
      \]
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \).
2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \).
3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
4. Example: Two instruments produce measurements \( X \),

Moment Generating Function Summary

1. General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \): \( m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \) for \( t \in \mathbb{R} \).
2. Binomial \( X \sim \text{Bin}(\pi, m) \): \( m_X(t) = (\exp(t)\pi + 1 - \pi)^m \) for \( t \in \mathbb{R} \).
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. 
   \[ p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}. \]

2. Continuous case: get joint probability density function divided by marginal probability density function: 
   \[ f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \]

3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.

4. Example: Two instruments produce measurements \( X \),

Moment Generating Function Summary

1. General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \): 
   \[ m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \text{ for } t \in \mathbb{R}. \]

2. Binomial \( X \sim \text{Bin}(\pi, m) \): 
   \[ m_X(t) = (\exp(t)\pi + 1 - \pi)^m \text{ for } t \in \mathbb{R}. \]

3. Poisson \( X \sim \text{Pois}(\lambda) \): 
   \[ m_X(t) = \exp([\exp(t) - 1]\lambda) \text{ for } t \in \mathbb{R}. \]
1 Conditional distributions:
   1 Discrete case: Just joint probability function divided by marginal probability function. \( p_{X \mid Y}(x \mid y) = p_{X, Y}(x, y) / p_{Y}(y) \).
   2 Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X \mid Y}(x \mid y) = f_{X, Y}(x, y) / f_{Y}(y) \).
   3 In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
   4 Example: Two instruments produce measurements \( X, Y \).

2 Moment Generating Function Summary
   1 General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \): \( m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \) for \( t \in \mathbb{R} \).
   2 Binomial \( X \sim \text{Bin}(\pi, m) \): \( m_X(t) = (\exp(t)\pi + 1 - \pi)^m \) for \( t \in \mathbb{R} \).
   3 Poisson \( X \sim \text{Pois}(\lambda) \): \( m_X(t) = \exp([\exp(t) - 1] \lambda) \) for \( t \in \mathbb{R} \).
   4 Negative Binomial \( N \sim \text{NBin}(\pi, k) \): \( m_N(t) = \exp(kt)[1 - \exp(t)(1 - \pi)]^{-k}\pi^k \) if \( t < -\ln(1 - \pi) \).
1. **Conditional distributions:**
   1. Discrete case: Just joint probability function divided by marginal probability function. $p_{X \mid Y}(x \mid y) = p_{X, Y}(x, y)/p_Y(y)$.
   2. Continuous case: get joint probability density function divided by marginal probability density function: $f_{X \mid Y}(x \mid y) = f_{X, Y}(x, y)/f_Y(y)$.
   3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
   4. Example: Two instruments produce measurements $X$, $Y$.

2. **Moment Generating Function Summary**
   1. General discrete distribution for $X$ with probabilities $\pi_i$ on points $\zeta_i$: $m_X(t) = \sum_i \pi_i \exp(\zeta_i t)$ for $t \in \mathbb{R}$.
   2. Binomial $X \sim \text{Bin}(\pi, m)$: $m_X(t) = (\exp(t)\pi + 1 - \pi)^m$ for $t \in \mathbb{R}$.
   3. Poisson $X \sim \text{Pois}(\lambda)$: $m_X(t) = \exp([\exp(t) - 1]\lambda)$ for $t \in \mathbb{R}$.
   4. Negative Binomial $N \sim \text{NBin}(\pi, k)$: $m_N(t) = \exp(kt)[1 - \exp(t)(1 - \pi)]^{-k}\pi^k$ if $t < -\ln(1 - \pi)$.
   5. Gamma distribution $X \sim \Gamma(k, \beta)$: $m_X(t) = (1 - \beta t)^{-k}$ if $t < 1/\beta$.

---

: Introduction: Lecture 16
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \).
2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \).
3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
4. Example: Two instruments produce measurements \( X \),

Moment Generating Function Summary

1. General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \): \( m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \) for \( t \in \mathbb{R} \).
2. Binomial \( X \sim \text{Bin}(\pi, m) \): \( m_X(t) = (\exp(t)\pi + 1 - \pi)^m \) for \( t \in \mathbb{R} \).
3. Poisson \( X \sim \text{Pois}(\lambda) \): \( m_X(t) = \exp([\exp(t) - 1]\lambda) \) for \( t \in \mathbb{R} \).
4. Negative Binomial \( N \sim \text{NBin}(\pi, k) \):
   \[
m_N(t) = \exp(kt)[1 - \exp(t)(1 - \pi)]^{-k}\pi^k \text{ if } t < -\ln(1 - \pi).
\]
5. Gamma distribution \( X \sim \Gamma(k, \beta) \): \( m_X(t) = (1 - \beta t)^{-k} \) if \( t < 1/\beta \).
6. Laplace \( X \sim \text{Lap} \): \( m_X(t) = \frac{(1+t)^{-1}}{2} + \frac{(1-t)^{-1}}{2} \) for \(|t| < 1\).
Conditional distributions:

1. Discrete case: Just joint probability function divided by marginal probability function. \( p_{X\mid Y}(x\mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \).
2. Continuous case: get joint probability density function divided by marginal probability density function: \( f_{X\mid Y}(x\mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \).
3. In both cases, for independent variables, conditional probability function or probability density function is same as marginal.
4. Example: Two instruments produce measurements \( X, Y \).

Moment Generating Function Summary

1. General discrete distribution for \( X \) with probabilities \( \pi_i \) on points \( \zeta_i \): \( m_X(t) = \sum_i \pi_i \exp(\zeta_i t) \) for \( t \in \mathbb{R} \).
2. Binomial \( X \sim \text{Bin}(\pi, m) \): \( m_X(t) = (\exp(t)\pi + 1 - \pi)^m \) for \( t \in \mathbb{R} \).
3. Poisson \( X \sim \text{Pois}(\lambda) \): \( m_X(t) = \exp([\exp(t) - 1]\lambda) \) for \( t \in \mathbb{R} \).
4. Negative Binomial \( N \sim \text{NBin}(\pi, k) \):
   \[
m_N(t) = \exp(kt)[1 - \exp(t)(1 - \pi)]^{-k}\pi^k \text{ if } t < -\ln(1 - \pi).
   \]
5. Gamma distribution \( X \sim \Gamma(k, \beta) \): \( m_X(t) = (1 - \beta t)^{-k} \) if \( t < 1/\beta \).
6. Laplace \( X \sim \text{Lap} \): \( m_X(t) = \frac{(1+t)^{-1}}{2} + \frac{(1-t)^{-1}}{2} \) for \( |t| < 1 \).
7. Normal \( Y \sim \mathcal{N}(\mu, \sigma^2) \): \( m_Y(t) = \exp\left(\frac{1}{2}t^2\sigma^2 + \mu t\right) \) for \( t \in \mathbb{R} \).
Higher-Dimensional Distributions

Sample space consists of a region in $\mathbb{R}^k$.

distribution function:

$$F_{X}(x_1, \ldots, x_k) = P (X_1 \leq x_1, \ldots, X_k \leq x_k)$$

probability function:

$$p_{X}(x_1, \ldots, x_k) = P (X_1 = x_1, \ldots, X_k = x_k)$$

probability density function:

$$f_{X}(x_1, \ldots, x_k) \text{ such that } \int_{A} f_{X}(x_1, \ldots, x_k) \, dx_1 \cdots dx_k = P ((X_1, \ldots, X_k) \in A)$$

Marginal and Conditional Quantities

Marginal distribution function: If $m < k$,

$$F_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P (X_1 \leq x_1, \ldots, X_k \leq x_k) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} F_{X_1, \ldots, X_k}(x_1, \ldots, x_k)$$

Marginal probability density function: If $m < k$,

$$f_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = \int_{\infty}^{-\infty} \cdots \int_{\infty}^{-\infty} f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) \, dx_{m+1} \cdots dx_k$$

Conditional distributions: If $X$ continuous, then

$$f_{X_1, \ldots, X_m | X_{m+1}, \ldots, X_k}(x_1, \ldots, x_m | x_{m+1}, \ldots, x_k) = \frac{f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)}{f_{X_{m+1}, \ldots, X_k}(x_{m+1}, \ldots, x_k)}$$
Higher-Dimensional Distributions

Sample space consists of a region in $\mathbb{R}^k$. 
Higher-Dimensional Distributions

- Sample space consists of a region in $\mathbb{R}^k$.
- Distribution function: $F_X(x_1, \cdots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k)$
- Probability function: $p_X(x_1, \cdots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$
- Probability density function: $f_X(x_1, \cdots, x_k)$ such that
  \[
  \int_{A} f_X(x_1, \cdots, x_k) \, dx_1 \cdots dx_k = P((X_1, \ldots, X_k) \in A)
  \]

Marginal and Conditional Quantities

- Marginal distribution function: If $m < k$,
  \[
  F_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P(X_1 \leq x_1, \ldots, X_k \leq x_k)
  \]
- Marginal probability density function: If $m < k$,
  \[
  f_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \ldots, X_k}(x_1, \cdots, x_k) \, dx_{m+1} \cdots dx_k
  \]
- Conditional distributions: If $X$ continuous, then
  \[
  f_{X_1, \ldots, X_m | X_{m+1}, \ldots, X_k}(x_1, \cdots, x_m | x_{m+1}, \ldots, x_k) = \frac{f_{X_1, \ldots, X_k}(x_1, \cdots, x_k)}{f_{X_{m+1}, \ldots, X_k}(x_{m+1}, \ldots, x_k)}
  \]
Higher-Dimensional Distributions

1. Sample space consists of a region in \( \mathbb{R}^k \).
2. Distribution function: \( F_X(x_1, \ldots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k) \)
3. Probability function: \( p_X(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k) \)

Marginal and Conditional Quantities

1. Marginal distribution function: If \( m < k \),
   \[
   F_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P(X_1 \leq x_1, \ldots, X_k \leq x_k)
   \]
2. Marginal probability density function: If \( m < k \),
   \[
   f_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) \, dx_{m+1} \cdots dx_k
   \]
3. Conditional distributions: If \( X \) continuous, then
   \[
   f_{X_1, \ldots, X_m|X_{m+1}, \ldots, X_k}(x_1, \ldots, x_m|x_{m+1}, \ldots, x_k) = \frac{f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)}{f_{X_{m+1}, \ldots, X_k}(x_{m+1}, \ldots, x_k)}
   \]
Higher-Dimensional Distributions

1. Sample space consists of a region in $\mathbb{R}^k$.
2. Distribution function: $F_X(x_1, \ldots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k)$
3. Probability function: $p_X(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$
4. Probability density function: $f_X(x_1, \ldots, x_k)$ such that

$$\int_A f_X(x_1, \ldots, x_k) \, dx_1 \cdots \, dx_k = P((X_1, \ldots, X_k) \in A)$$
Higher-Dimensional Distributions

1. Sample space consists of a region in \( \mathbb{R}^k \).

2. Distribution function: \( F_X(x_1, \ldots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k) \)

3. Probability function: \( p_X(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k) \)

4. Probability density function: \( f_X(x_1, \ldots, x_k) \) such that

\[
\int_A f_X(x_1, \ldots, x_k) \, dx_1 \cdots dx_k = P((X_1, \ldots, X_k) \in A)
\]

5. Probability density function is derivative of distribution function:
Higher-Dimensional Distributions

1. Sample space consists of a region in $\mathbb{R}^k$.
2. Distribution function: $F_X(x_1, \cdots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k)$
3. Probability function: $p_X(x_1, \cdots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$
4. Probability density function: $f_X(x_1, \cdots, x_k)$ such that

$$\int_A f_X(x_1, \cdots, x_k) \, dx_1 \cdots dx_k = P((X_1, \ldots, X_k) \in A)$$

5. Probability density function is derivative of distribution function:

Marginal and Conditional Quantities

2. Marginal distribution function: If $m < k$,

$$F_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P(X_1 \leq x_1, \ldots, X_k \leq x_k) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} F_{X_1, \ldots, X_k}(x_1, \cdots, x_k)$$

2. Marginal probability density function: If $m < k$,

$$f_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \int_{\infty}^{-\infty} \cdots \int_{\infty}^{-\infty} f_{X_1, \ldots, X_k}(x_1, \cdots, x_k) \, dx_{m+1} \cdots dx_k$$

2. Conditional distributions: If $X$ continuous, then

$$f_{X_1, \ldots, X_m | X_{m+1}, \ldots, X_k}(x_1, \cdots, x_m | x_{m+1}, \ldots, x_k) = \frac{f_{X_1, \ldots, X_k}(x_1, \cdots, x_k)}{f_{X_{m+1}, \ldots, X_k}(x_{m+1}, \ldots, x_k)}$$
Higher-Dimensional Distributions

1. Sample space consists of a region in $\mathbb{R}^k$.
2. Distribution function: $F_{\mathbf{X}}(x_1, \cdots, x_k) = P (X_1 \leq x_1, \ldots, X_k \leq x_k)$
3. Probability function: $p_{\mathbf{X}}(x_1, \cdots, x_k) = P (X_1 = x_1, \ldots, X_k = x_k)$
4. Probability density function: $f_{\mathbf{X}}(x_1, \cdots, x_k)$ such that
   $$\int_A f_{\mathbf{X}}(x_1, \cdots, x_k) \, dx_1 \cdots dx_k = P ((X_1, \ldots, X_k) \in A)$$
5. Probability density function is derivative of distribution function:

Marginal and Conditional Quantities

1. Marginal distribution function: If $m < k$,
   $$F_{X_1,\ldots,X_m}(x_1, \cdots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P (X_1 \leq x_1, \ldots, X_k \leq x_k)$$
   $$= \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} F_{X_1,\ldots,X_k}(x_1, \cdots, x_k)$$
Higher-Dimensional Distributions

1. Sample space consists of a region in $\mathbb{R}^k$.
2. Distribution function: $F_X(x_1, \cdots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k)$
3. Probability function: $p_X(x_1, \cdots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$
4. Probability density function: $f_X(x_1, \cdots, x_k)$ such that
   \[
   \int_A f_X(x_1, \cdots, x_k) \, dx_1 \cdots dx_k = P((X_1, \ldots, X_k) \in A)
   \]
5. Probability density function is derivative of distribution function:

Marginal and Conditional Quantities

1. Marginal distribution function: If $m < k$,
   \[
   F_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P(X_1 \leq x_1, \ldots, X_k \leq x_k) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} F_{X_{m+1}, \ldots, X_k}(x_1, \cdots, x_k)
   \]
2. Marginal probability density function: If $m < k$,
   \[
   f_{X_1, \ldots, X_m}(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_{m+1}, \ldots, X_k}(x_1, \cdots, x_k) \, dx_{m+1} \cdots dx_k
   \]
Higher-Dimensional Distributions

1. Sample space consists of a region in $\mathbb{R}^k$.
2. Distribution function: $F_X(x_1, \cdots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k)$
3. Probability function: $p_X(x_1, \cdots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$
4. Probability density function: $f_X(x_1, \cdots, x_k)$ such that
   \[ \int_A f_X(x_1, \cdots, x_k) \, dx_1 \cdots dx_k = P((X_1, \ldots, X_k) \in A) \]
5. Probability density function is derivative of distribution function:

Marginal and Conditional Quantities

1. Marginal distribution function: If $m < k$,
   \[ F_{X_1,\ldots,X_m}(x_1, \cdots, x_m) = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} P(X_1 \leq x_1, \ldots, X_k \leq x_k) \]
   \[ = \lim_{x_{m+1} \to \infty} \cdots \lim_{x_k \to \infty} F_{X_1,\ldots,X_k}(x_1, \cdots, x_k) \]
2. Marginal probability density function: If $m < k$,
   \[ f_{X_1,\ldots,X_m}(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(x_1, \cdots, x_k) \, dx_{m+1} \cdots dx_k \]
3. Conditional distributions: If $X$ continuous, then
   \[ f_{X_1,\ldots,X_m|X_{m+1},\ldots,X_k}(x_1, \cdots, x_m|x_{m+1}, \ldots, x_k) = \frac{f_{X_1,\ldots,X_k}(x_1, \cdots, x_k)}{f_{X_{m+1},\ldots,X_k}(x_{m+1}, \ldots, x_k)} \]
Transformations of multiple random variables

1. The setup: \( U = u(X, Y), \ V = v(X, Y). \)
Transformations of multiple random variables

1. The setup: $U = u(X, Y), \ V = v(X, Y)$.
   - Inverses $X = x(U, V), \ Y = y(U, V)$;
Transformations of multiple random variables

1. The setup: $U = u(X, Y), \ V = v(X, Y)$.
   1. Inverses $X = x(U, V), \ Y = y(U, V)$;
   2. The objective:
Transformations of multiple random variables

1. The setup: \( U = u(X, Y), \ V = v(X, Y) \).
   1. Inverses \( X = x(U, V), \ Y = y(U, V); \)
   2. The objective:

2. Transformations in the Discrete case:

   Let \( J^{-1} \) be the Jacobian of the inverse transformation.

   Hence \( f_{U, V}(u, v) = f_{X, Y}(x(u, v), y(u, v)) \cdot |J^{-1}| \)

   Multiple inverses: Sum over them.
Transformations of multiple random variables

1. The setup: $U = u(X, Y), \ V = v(X, Y)$.
   1. Inverses $X = x(U, V), \ Y = y(U, V)$;
   2. The objective:

2. Transformations in the Discrete case:
   1. $f_{U, V}(u, v) = f_{X, Y}(x(u, v), y(u, v))$
Transformations of multiple random variables

1. The setup: \( U = u(X, Y), \ V = v(X, Y) \).
   - Inverses \( X = x(U, V), \ Y = y(U, V) \);
   - The objective:

2. Transformations in the Discrete case:
   - \( f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \)

3. Transformations in the Continuous case:

: Introduction: Lecture 16
Transformations of multiple random variables

The setup: \( U = u(X, Y), \ V = v(X, Y) \).

- Inverses \( X = x(U, V), \ Y = y(U, V) \);
- The objective:

Transformations in the Discrete case:

- \( f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \)

Transformations in the Continuous case:

- Let \( J^- \) be the Jacobian of the inverse transformation.
Transformations of multiple random variables

The setup: \( U = u(X, Y), \ V = v(X, Y) \).

1. Inverses \( X = x(U, V), \ Y = y(U, V) \);
2. The objective:

Transformations in the Discrete case:

1. \( f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \)

Transformations in the Continuous case:

1. Let \( J^- \) be the Jacobian of the inverse transformation.
2. Hence \( f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) |J^-| \)
Transformations of multiple random variables

The setup: $U = u(X, Y), \ V = v(X, Y)$.
1. Inverses $X = x(U, V), \ Y = y(U, V)$;
2. The objective:

Transformations in the Discrete case:
1. $f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))$

Transformations in the Continuous case:
1. Let $J^-$ be the Jacobian of the inverse transformation.
2. Hence $f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) |J^-|$  
3. Multiple inverses: Sum over them.
Distribution of a lower-dimensional transformation
Distribution of a lower-dimensional transformation

Ex., \((X, Y)\) have a joint distribution.
Distribution of a lower-dimensional transformation

1. Ex., \((X, Y)\) have a joint distribution.
2. Want distribution of \(U = u(X, Y)\)
Distribution of a lower-dimensional transformation

1. Ex., $(X, Y)$ have a joint distribution.
2. Want distribution of $U = u(X, Y)$
3. Change of variables formula in multiple dimensions doesn’t work,
Distribution of a lower-dimensional transformation

1 Ex., \((X, Y)\) have a joint distribution.
2 Want distribution of \(U = u(X, Y)\)
3 Change of variables formula in multiple dimensions doesn’t work,
4 Fix via transforming to a space of the same dimension and marginalizing.
Distribution of a lower-dimensional transformation

1. Ex., \((X, Y)\) have a joint distribution.
2. Want distribution of \(U = u(X, Y)\).
3. Change of variables formula in multiple dimensions doesn’t work,
4. Fix via transforming to a space of the same dimension and marginalizing.

The mgf method: \(U = u(X, Y)\)
Distribution of a lower-dimensional transformation

1. Ex., \((X, Y)\) have a joint distribution.
2. Want distribution of \(U = u(X, Y)\)
3. Change of variables formula in multiple dimensions doesn’t work,
4. Fix via transforming to a space of the same dimension and marginalizing.

The mgf method: \(U = u(X, Y)\)

1. \(m_U(\nu) = E(\exp(\nu u(X, Y)))\)
Distribution of a lower-dimensional transformation

1. Ex., \((X, Y)\) have a joint distribution.
2. Want distribution of \(U = u(X, Y)\)
3. Change of variables formula in multiple dimensions doesn’t work,
4. Fix via transforming to a space of the same dimension and marginalizing.

The mgf method: \(U = u(X, Y)\)

1. \(m_U(\nu) = E(\exp(\nu u(X, Y)))\)
2. Try to recognize this.
Expectations in Multiple Dimensions

Set up:

1. Consider jointly-distributed random variables \((X, Y)\).
2. Consider some summary function \(g(x, y)\).
3. Want \(E(g(X, Y))\).
Expectations in Multiple Dimensions

1 Setup:
   1 Consider jointly-distributed random variables \((X, Y)\)
Expectations in Multiple Dimensions

Setup:

1. Consider jointly-distributed random variables \((X, Y)\).
2. Consider some summary function \(g_1(x, y)\).
Expectations in Multiple Dimensions

Setup:

1. Consider jointly-distributed random variables \((X, Y)\).
2. Consider some summary function \(g_1(x, y)\).
3. Want \(E(g_1(X, Y))\).
Expectations in Multiple Dimensions

1. Setup:
   1. Consider jointly-distributed random variables \((X, Y)\).
   2. Consider some summary function \(g_1(x, y)\).
   3. Want \(E(g_1(X, Y))\).

2. Discrete distributions:
Expectations in Multiple Dimensions

1. Setup:
   1. Consider jointly-distributed random variables $(X, Y)$
   2. Consider some summary function $g_1(x, y)$.
   3. Want $E(g_1(X, Y))$.

2. Discrete distributions:
   1. Equivalent to
   \[
   \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} g_1(x, y) p_{X, Y}(x, y).
   \]
Continuous case:

Result is

$$E(Z) = \int \int g(x, y) f_{X,Y}(x, y) \, dy \, dx.$$ 

Summability:

Suppose $X$ and $Y$ have a joint probability function $f_{X,Y}(x, y)$.

Then $E(X+Y) = E(X) + E(Y)$.

By extension this holds for any number of summands.

Ex.: a $\text{Bin}(m, \pi)$ variable has expectation $m \pi$.

Hence the expectation has the advantage of transforming easily.
Continuous case:

Result is

\[ E(Z) = \int_{\mathcal{X}} \int_{\mathcal{Y}} g_1(x, y) f_{X,Y}(x, y) \, dy \, dx. \]
Continuous case:

Result is

\[ E(Z) = \int_X \int_Y g_1(x, y) f_{X,Y}(x, y) \, dy \, dx. \]

Summability:

Suppose \( X \) and \( Y \) have a joint probability function \( f_{X,Y}(x, y) \).

Then \( E(X + Y) = E(X) + E(Y) \).

By extension this holds for any number of summands.

Ex.: a Bin\( (m, \pi) \) variable has expectation \( m\pi \).

Hence the expectation has the advantage of transforming easily.
Continuous case:

Result is

$$E(Z) = \int_X \int_Y g_1(x, y)f_{X,Y}(x, y) \, dy \, dx.$$ 

Summability:

Suppose $X$ and $Y$ have a joint probability function $f_{X,Y}(x, y)$.

By extension this holds for any number of summands.

Ex.: a Bin($m, \pi$) variable has expectation $m\pi$. Hence the expectation has the advantage of transforming easily.
Continuous case:

Result is

$$E(Z) = \int_{X} \int_{Y} g_1(x, y) f_{X,Y}(x, y) \, dy \, dx.$$ 

Summability:

Suppose $X$ and $Y$ have a joint probability function $f_{X,Y}(x, y)$.

Then $E(X + Y) = E(X) + E(Y)$.
Continuous case:

Result is

$$E(Z) = \int_{X} \int_{Y} g_1(x, y) f_{X,Y}(x, y) \, dy \, dx.$$ 

Summability:

Suppose $X$ and $Y$ have a joint probability function $f_{X,Y}(x, y)$.

Then $E(X + Y) = E(X) + E(Y)$

By extension this holds for any number of summands.
Continuous case:

Result is

$$E(Z) = \int_X \int_Y g_1(x, y) f_{X,Y}(x, y) \, dy \, dx.$$ 

Summability:

Suppose $X$ and $Y$ have a joint probability function $f_{X,Y}(x, y)$.

Then $E(X + Y) = E(X) + E(Y)$

By extension this holds for any number of summands.

Ex.: a Bin($m, \pi$) variable has expectation $m\pi$. 

Hence the expectation has the advantage of transforming easily.
Continuous case:

Result is

\[ E(Z) = \int_{X} \int_{Y} g_1(x, y) f_{X,Y}(x, y) \, dy \, dx. \]

Summability:

1. Suppose \( X \) and \( Y \) have a joint probability function \( f_{X,Y}(x, y) \).
2. Then \( E(X + Y) = E(X) + E(Y) \).
3. By extension this holds for any number of summands.
4. Ex.: a Bin\((m, \pi)\) variable has expectation \( m\pi \).
5. Hence the expectation has the advantage of transforming easily.
Multiplicative property under independence:
Multiplicative property under independence:

If $X$ and $Y$ are independent,
Multiplicative property under independence:

If $X$ and $Y$ are independent,

g and $h$ are functions
Multiplicative property under independence:

1. If $X$ and $Y$ are independent,
2. $g$ and $h$ are functions
3. Then $E(g(X)h(Y)) = E(g(X))E(h(Y))$. 
Distributions of Sums of independent random variables:

1. In general

\[ W \text{ and } X \] are independent random variables

\[ Y = W + X \]

The distribution of \( Y \) is called the convolution of the distributions of \( X \) and \( Y \).

Carry along \( Z = X \) to make the transformation invertible.

Examine the function \((y, z) = r(w, x) = (w + x, x)\).

Remove \( Z \) by marginalizing.
Distributions of Sums of independent random variables:

1. In general
   1. $W$ and $X$ are independent random variables
Distributions of Sums of independent random variables:

1. In general
   1. $W$ and $X$ are independent random variables
   2. Produce the distribution of $Y = W + X$. 
Distributions of Sums of independent random variables:

In general

1. $W$ and $X$ are independent random variables
2. Produce the distribution of $Y = W + X$.
3. The distribution of $Y$ is called the *convolution* of the distributions of $X$ and $Y$. 
Distributions of Sums of independent random variables:

1. In general
   1. \( W \) and \( X \) are independent random variables
   2. Produce the distribution of \( Y = W + X \).
   3. The distribution of \( Y \) is called the \textit{convolution} of the distributions of \( X \) and \( Y \).
   4. Carry along \( Z = X \) to make the transformation invertible.
Distributions of Sums of independent random variables:

1. In general
   1. $W$ and $X$ are independent random variables
   2. Produce the distribution of $Y = W + X$.
   3. The distribution of $Y$ is called the *convolution* of the distributions of $X$ and $Y$.
   4. Carry along $Z = X$ to make the transformation invertible.
   5. Examine the function $(y, z) = r(w, x) = (w + x, x)$. 
Distributions of Sums of independent random variables:

1. In general
   1. $W$ and $X$ are independent random variables
   2. Produce the distribution of $Y = W + X$.
   3. The distribution of $Y$ is called the \textit{convolution} of the distributions of $X$ and $Y$.
   4. Carry along $Z = X$ to make the transformation invertible.
   5. Examine the function $(y, z) = r(w, x) = (w + x, x)$.
   6. Remove $Z$ by marginalizing.
The mgf of a collection of independent random variables is the product of the mgfs.
The mgf of a collection of independent random variables is the product of the mgfs.

One can sometimes perform convolution by recognizing the distribution from the product of the mgfs.
The mgf of a collection of independent random variables is the product of the mgfs.

One can sometimes perform convolution by recognizing the distribution from the product of the mgfs.

Sums of general discrete independent random variables
The mgf of a collection of independent random variables is the product of the mgfs.

One can sometimes perform convolution by recognizing the distribution from the product of the mgfs.

Sums of general discrete independent random variables

To obtain the probability function of the sum alone, marginalize by summing out $Z$ to get $p_Y(y) = \sum_{z \in X} p_W(y - z)p_X(z)$. 

Poisson Example:

\[ X \sim \text{Pois}(\gamma), \quad W \sim \text{Pois}(\lambda), \quad X \perp W, \quad Y = X + W. \]

Result:

\[ Y \sim \text{Pois}(\lambda + \gamma). \]

Binomial Example:

\[ X \sim \text{Bin}(m, \pi), \quad W \sim \text{Bin}(m, \pi), \quad X \perp W, \quad Y = X + W. \]

Result:

\[ Y \sim \text{Bin}(2m, \pi). \]

Extension: If \( X_j \sim \text{Bin}(m_j, \pi) \), independent, then

\[ \sum_j X_j \sim \text{Bin}(\sum_j m_j, \pi). \]
Poisson Example:

\[ X \sim \text{Pois}(\gamma), \ W \sim \text{Pois}(\lambda), \ X \perp W, \ Y = X + W. \]
Poisson Example:
1. $X \sim \text{Pois}(\gamma)$, $W \sim \text{Pois}(\lambda)$, $X \perp W$, $Y = X + W$.
2. Result: $Y \sim \text{Pois}(\lambda + \gamma)$
1 Poisson Example:
   1. $X \sim \text{Pois}(\gamma)$, $W \sim \text{Pois}(\lambda)$, $X \perp W$, $Y = X + W$.
   2. Result: $Y \sim \text{Pois}(\lambda + \gamma)$

2 Binomial Example:
Poisson Example:

1. $X \sim \text{Pois}(\gamma)$, $W \sim \text{Pois}(\lambda)$, $X \perp W$, $Y = X + W$.
2. Result: $Y \sim \text{Pois}(\lambda + \gamma)$

Binomial Example:

1. $X \sim \text{Bin}(m, \pi)$, $W \sim \text{Bin}(m, \pi)$, $X \perp W$, $Y = X + W$

Extension: If $X_j \sim \text{Bin}(m_j, \pi)$, independent, then $\sum_j X_j \sim \text{Bin}(\sum_j m_j, \pi)$.
1 Poisson Example:
   1. $X \sim \text{Pois}(\gamma)$, $W \sim \text{Pois}(\lambda)$, $X \perp W$, $Y = X + W$.
   2. Result: $Y \sim \text{Pois}(\lambda + \gamma)$

2 Binomial Example:
   1. $X \sim \text{Bin}(m, \pi)$, $W \sim \text{Bin}(m, \pi)$, $X \perp W$, $Y = X + W$
   2. Result: $Y \sim \text{Bin}(m + m, \pi)$
Poisson Example:

1. $X \sim \text{Pois} (\gamma), \ W \sim \text{Pois} (\lambda), \ X \perp W, \ Y = X + W.$
2. Result: $Y \sim \text{Pois} (\lambda + \gamma)$

Binomial Example:

1. $X \sim \text{Bin} (m, \pi), \ W \sim \text{Bin} (m, \pi), \ X \perp W, \ Y = X + W$
2. Result: $Y \sim \text{Bin} (m + m, \pi)$
3. Extension: If $X_j \sim \text{Bin} (m_j, \pi)$, independent, then $\sum_j X_j \sim \text{Bin} (\sum_j m_j, \pi)$. 

: Introduction: Lecture 17
Negative Binomial Example:
1 Negative Binomial Example:

1. \( X \sim \text{NBin}(k, \pi), \ W \sim \text{NBin}(m, \pi), \ X \perp W, \ Y = X + W \)
Negative Binomial Example:
1. $X \sim \text{NBin}(k, \pi)$, $W \sim \text{NBin}(m, \pi)$, $X \perp W$, $Y = X + W$
2. Result: $Y \sim \text{NBin}(m + k, \pi)$
Negative Binomial Example:

1. $X \sim \text{NBin}(k, \pi)$, $W \sim \text{NBin}(m, \pi)$, $X \perp W$, $Y = X + W$
2. Result: $Y \sim \text{NBin}(m + k, \pi)$
3. Convolution directly as hard as for binomial.
Negative Binomial Example:

1. $X \sim \text{NBin}(k, \pi), \ W \sim \text{NBin}(m, \pi), \ X \perp W, \ Y = X + W$
2. Result: $Y \sim \text{NBin}(m + k, \pi)$
3. Convolution directly as hard as for binomial.

A cautionary case
Negative Binomial Example:

1. $X \sim \text{NBin}(k, \pi)$, $W \sim \text{NBin}(m, \pi)$, $X \perp W$, $Y = X + W$
2. Result: $Y \sim \text{NBin}(m + k, \pi)$
3. Convolution directly as hard as for binomial.

A cautionary case

1. $W \sim \sqrt{2} \text{Bin}(m, \pi)$, $X \sim \text{Bin}(n, \pi)$, $W \perp X$, $Y = X + W$. 

Negative Binomial Example:

1. \( X \sim \text{NBin}(k, \pi) \), \( W \sim \text{NBin}(m, \pi) \), \( X \perp W \), \( Y = X + W \)
2. Result: \( Y \sim \text{NBin}(m + k, \pi) \)
3. Convolution directly as hard as for binomial.

A cautionary case

1. \( W \sim \sqrt{2}\text{Bin}(m, \pi) \), \( X \sim \text{Bin}(n, \pi) \), \( W \perp X \), \( Y = X + W \).
2. Every possible \( y \) is associated with only one possible \((x, w)\) pair, because:
Sums of general continuous independent random variables

Same setup as for discrete case: $X, W$ independent, $Y = X + W, Z = X$.

The determinant of matrix of derivatives is then
$$\begin{vmatrix} dw & dx \\ dy & dz \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}.$$

The determinant of this matrix is then 1, and the probability density function is
$$f_{Y, Z}(y, z) = f_W(y - z)f_X(z).$$

To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get
$$f_Y(y) = \int_X f_W(y - z)f_X(z) \, dz.$$

Uniform Example:
$X \sim \text{Unif}(0, 1), W \sim \text{Unif}(0, 1), X \perp W, Y = X + W.$

By general result,
$$f_Y(y) = \int_0^1 f_W(y - z)f_X(z) \, dz.$$

$f_X(z) = 1$ for $x \in [0, 1]$.

Hence
$$f_Y(y) = \int_0^1 f_W(y - z) \, dz.$$

The probability density function for the sum $Y$ is defined in pieces:

Repeating this with more copies of $W$ gives densities for sums of arbitrary numbers of summands,
Sums of general continuous independent random variables

Same setup as for discrete case: $X, W$ independent, $Y = X + W$, $Z = X$. 

The determinant of matrix of derivatives is then
$$\begin{vmatrix}
\frac{dw}{dx}, \frac{dy}{dx} \\
\frac{dz}{dx}, \frac{dz}{dz}
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-1 & 1
\end{vmatrix}.$$ 

The determinant of this matrix is then 1, and the probability density function is
$$f_{Y, Z}(y, z) = f_W(y - z)f_X(z).$$ 

To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get
$$f_Y(y) = \int f_W(y - z)f_X(z) \, dz.$$ 

Uniform Example:

$X \sim \text{Unif}(0, 1)$, $W \sim \text{Unif}(0, 1)$, $X \perp W$, $Y = X + W$.

By general result,
$$f_Y(y) = \int_0^1 f_W(y - z) \, dz.$$ 

$f_X(z) = 1$ for $x \in [0, 1]$.

Hence
$$f_Y(y) = \int_0^1 f_W(y - z) \, dz.$$ 

The probability density function for the sum $Y$ is defined in pieces:

Repeating this with more copies of $W$ gives densities for sums of arbitrary numbers of summands,
Sums of general continuous independent random variables

1. Same setup as for discrete case: \( X, W \) independent, \( Y = X + W \), \( Z = X \).

2. The determinant of matrix of derivatives is then

\[
\begin{vmatrix}
(dw, dx) \\
(dy, dz)
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-1 & 1
\end{vmatrix}.
\]

3. The probability density function is

\[
f_Y(y, z) = f_W(y - z) f_X(z).
\]

4. To obtain the probability density function of the sum alone, marginalize by integrating out \( Z \) to get

\[
f_Y(y) = \int_X f_W(y - z) f_X(z) \, dz.
\]

5. Uniform Example:

\[
X \sim \text{Unif}(0, 1), \quad W \sim \text{Unif}(0, 1), \quad X \perp W, \quad Y = X + W.
\]

6. By general result,

\[
f_Y(y) = \int_0^1 f_W(y - z) \, dz.
\]

7. \( f_X(z) = 1 \) for \( z \in [0, 1] \).

8. Hence

\[
f_Y(y) = \int_0^1 f_W(y - z) \, dz.
\]

9. The probability density function for the sum \( Y \) is defined in pieces:

10. Repeating this with more copies of \( W \) gives densities for sums of arbitrary numbers of summands,
Sums of general continuous independent random variables


2. The determinant of matrix of derivatives is then

$$\begin{vmatrix} (dw, dx) \\ (dy, dz) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}.$$ 

3. The determinant of this matrix is then 1, and the probability density function is $f_{Y,Z}(y, z) = f_W(y - z)f_X(z)$.
Sums of general continuous independent random variables


2. The determinant of matrix of derivatives is then

$$\begin{vmatrix} (dw, dx) \\ (dy, dz) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}.$$ 

3. The determinant of this matrix is then 1, and the probability density function is $f_{Y,Z}(y, z) = f_W(y - z)f_X(z)$.

4. To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get $f_Y(y) = \int_X f_W(y - z)f_X(z) \, dz$. 

Uniform Example:
$X \sim \text{Unif}(0, 1)$, $W \sim \text{Unif}(0, 1)$, $X \perp W$, $Y = X + W$.

By general result,

$$f_Y(y) = \int_0^1 f_W(y - z)f_X(z) \, dz.$$ 

$f_X(z) = 1$ for $z \in [0, 1]$.

Hence

$$f_Y(y) = \int_0^1 f_W(y - z) \, dz.$$ 

The probability density function for the sum is defined in pieces:

Repeating this with more copies of $W$ gives densities for sums of arbitrary numbers of summands.
Sums of general continuous independent random variables

Same setup as for discrete case: \( X, W \) independent, \( Y = X + W \), \( Z = X \).

The determinant of matrix of derivatives is then

\[
\begin{vmatrix}
(dw, dx) \\
(dy, dz)
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-1 & 1
\end{vmatrix}.
\]

The determinant of this matrix is then 1, and the probability density function is \( f_{Y,Z}(y, z) = f_W(y - z)f_X(z) \).

To obtain the probability density function of the sum alone, marginalize by integrating out \( Z \) to get \( f_Y(y) = \int f_W(y - z)f_X(z) \, dz \).

Uniform Example:
Sums of general continuous independent random variables

Same setup as for discrete case: \( X, W \) independent, \( Y = X + W, Z = X \).

The determinant of matrix of derivatives is then

\[
\begin{vmatrix}
(dw, dx) \\
(dy, dz)
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-1 & 1
\end{vmatrix}.
\]

The determinant of this matrix is then 1, and the probability density function is

\[ f_{Y,Z}(y, z) = f_W(y - z)f_X(z). \]

To obtain the probability density function of the sum alone, marginalize by integrating out \( Z \) to get

\[ f_Y(y) = \int_X f_W(y - z)f_X(z) \, dz. \]

Uniform Example:

\( X \sim \text{Unif}(0, 1), W \sim \text{Unif}(0, 1), X \perp W, Y = X + W. \)
Sums of general continuous independent random variables


2. The determinant of matrix of derivatives is then

$$
\left| \begin{pmatrix} dw & dx \\ dy & dz \end{pmatrix} \right| = \left| \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right|.
$$

3. The determinant of this matrix is then 1, and the probability density function is $f_{Y,Z}(y,z) = f_W(y-z)f_X(z)$.

4. To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get $f_Y(y) = \int_X f_W(y-z)f_X(z) \, dz$.

Uniform Example:

1. $X \sim \text{Unif}(0,1)$, $W \sim \text{Unif}(0,1)$, $X \perp W$, $Y = X + W$.

2. By general result, $f_Y(y) = \int_0^1 f_W(y-z)f_X(z) \, dz$. 
Sums of general continuous independent random variables

2. The determinant of matrix of derivatives is then

$$\left| \begin{pmatrix} dw & dx \\ dy & dz \end{pmatrix} \right| = \left| \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right|.$$

3. The determinant of this matrix is then 1, and the probability density function is $f_{Y,Z}(y, z) = f_W(y - z)f_X(z)$.
4. To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get $f_Y(y) = \int_X f_W(y - z)f_X(z) \, dz$.

Uniform Example:

1. $X \sim \text{Unif}(0, 1)$, $W \sim \text{Unif}(0, 1)$, $X \perp W$, $Y = X + W$.
2. By general result, $f_Y(y) = \int_0^1 f_W(y - z)f_X(z) \, dz$.
3. $f_X(z) = 1$ for $x \in [0, 1]$. 


Sums of general continuous independent random variables

Same setup as for discrete case: $X, W$ independent, $Y = X + W, Z = X$.

The determinant of matrix of derivatives is then

$$\begin{vmatrix} \frac{dw}{dx} & \frac{dx}{dz} \\ \frac{dy}{dz} & \frac{dz}{dx} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}.$$  

The determinant of this matrix is then 1, and the probability density function is $f_{Y,Z}(y,z) = f_W(y-z)f_X(z)$.

To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get $f_Y(y) = \int f_W(y-z)f_X(z) \, dz$.

Uniform Example:

$X \sim \text{Unif}(0,1), W \sim \text{Unif}(0,1), X \perp W, Y = X + W$.

By general result, $f_Y(y) = \int_0^1 f_W(y-z)f_X(z) \, dz$.

$f_X(z) = 1$ for $x \in [0,1]$.

Hence $f_Y(y) = \int_0^1 f_W(y-z)1 \, dz$. 
Sums of general continuous independent random variables

1. Same setup as for discrete case: \( X, W \) independent, \( Y = X + W \), \( Z = X \).

2. The determinant of matrix of derivatives is then

\[
\begin{vmatrix}
(dw, dx) \\
(dy, dz)
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-1 & 1
\end{vmatrix}.
\]

3. The determinant of this matrix is then 1, and the probability density function is \( f_{Y,Z}(y, z) = f_W(y - z)f_X(z) \).

4. To obtain the probability density function of the sum alone, marginalize by integrating out \( Z \) to get \( f_Y(y) = \int_X f_W(y - z)f_X(z) \, dz \).

Uniform Example:

1. \( X \sim \text{Unif}(0, 1), W \sim \text{Unif}(0, 1), X \perp W, Y = X + W \).

2. By general result, \( f_Y(y) = \int_0^1 f_W(y - z)f_X(z) \, dz \).

3. \( f_X(z) = 1 \) for \( x \in [0, 1] \).

4. Hence \( f_Y(y) = \int_0^1 f_W(y - z)1 \, dz \).

5. The probability density function for the sum \( Y \) is defined in pieces:
Sums of general continuous independent random variables

2. The determinant of matrix of derivatives is then
   \[
   \begin{vmatrix}
   (dw, dx) \\ (dy, dz)
   \end{vmatrix} = \begin{vmatrix}
   1 & 0 \\ -1 & 1
   \end{vmatrix}.
   \]
3. The determinant of this matrix is then 1, and the probability density function is $f_{Y, Z}(y, z) = f_W(y - z)f_X(z)$.
4. To obtain the probability density function of the sum alone, marginalize by integrating out $Z$ to get $f_Y(y) = \int_X f_W(y - z)f_X(z)\,dz$.

Uniform Example:

1. $X \sim \text{Unif}(0, 1), W \sim \text{Unif}(0, 1), X \perp W, Y = X + W$.
2. By general result, $f_Y(y) = \int_0^1 f_W(y - z)f_X(z)\,dz$.
3. $f_X(z) = 1$ for $x \in [0, 1]$.
4. Hence $f_Y(y) = \int_0^1 f_W(y - z)1\,dz$.
5. The probability density function for the sum $Y$ is defined in pieces:
6. Repeating this with more copies of $W$ gives densities for sums of arbitrary numbers of summands,
Normal Example:

\[ X \sim N(\mu, \sigma^2), \quad W \sim N(\nu, \tau^2), \quad X \perp W, \quad Y = X + W. \]

Direct convolution shows \( X + W \sim N(\mu + \nu, \tau^2 + \sigma^2) \), because

\[ \sum_j X_j \sim N\left(\sum_j \nu_j, \sum_j \sigma_j^2\right). \]

Cauchy Example:

\[ X \text{ and } W \text{ Cauchy, } X \perp W, \text{ and } Y = X + W. \]

\[ Y/2 \text{ is Cauchy.} \]
Normal Example:

\[ X \sim \mathcal{N}(\mu, \sigma^2), \ W \sim \mathcal{N}(\nu, \tau^2), \ X \perp W, \ Y = X + W. \]
Normal Example:

1. $X \sim \mathcal{N}(\mu, \sigma^2)$, $W \sim \mathcal{N}(\nu, \tau^2)$, $X \perp W$, $Y = X + W$.

2. Direct convolution shows $X + W \sim \mathcal{N}(\mu + \nu, \tau^2 + \sigma^2)$, because
Normal Example:

1. $X \sim \mathcal{N}(\mu, \sigma^2)$, $W \sim \mathcal{N}(\nu, \tau^2)$, $X \perp W$, $Y = X + W$.
2. Direct convolution shows $X + W \sim \mathcal{N}(\mu + \nu, \tau^2 + \sigma^2)$, because
3. Corollary.: If $X_j$ are independent $\mathcal{N}(\nu_j, \sigma^2_j)$, then
   $\sum_j X_j \sim \mathcal{N}(\sum_j \nu_j, \sum_j \sigma^2_j)$. 

Cauchy Example:

$X$ and $W$ Cauchy, $X \perp W$, and $Y = X + W$.

$Y/2$ is Cauchy.
Normal Example:

1. $X \sim \mathcal{N}(\mu, \sigma^2)$, $W \sim \mathcal{N}(\nu, \tau^2)$, $X \perp W$, $Y = X + W$.
2. Direct convolution shows $X + W \sim \mathcal{N}(\mu + \nu, \tau^2 + \sigma^2)$, because
3. Corollary: If $X_j$ are independent $\mathcal{N}(\nu_j, \sigma_j^2)$, then $\sum_j X_j \sim \mathcal{N}(\sum_j \nu_j, \sum_j \sigma_j^2)$.

Cauchy Example:

1. $X$ and $W$ Cauchy, $X \perp W$, and $Y = X + W$.
2. $\frac{Y}{2}$ is Cauchy.
1 Normal Example:
   1 $X \sim \mathcal{N}(\mu, \sigma^2)$, $W \sim \mathcal{N}(\nu, \tau^2)$, $X \perp W$, $Y = X + W$.
   2 Direct convolution shows $X + W \sim \mathcal{N}(\mu + \nu, \tau^2 + \sigma^2)$, because
   3 Corollary.: If $X_j$ are independent $\mathcal{N}(\nu_j, \sigma_j^2)$, then
     $\sum_j X_j \sim \mathcal{N}(\sum_j \nu_j, \sum_j \sigma_j^2)$.

2 Cauchy Example:
   1 $X$ and $W$ Cauchy, $X \perp W$, and $Y = X + W$. 
1 Normal Example:
   1 \( X \sim \mathcal{N}(\mu, \sigma^2), \ W \sim \mathcal{N}(\nu, \tau^2), \ X \perp W, \ Y = X + W. \)
   2 Direct convolution shows \( X + W \sim \mathcal{N}(\mu + \nu, \tau^2 + \sigma^2) \), because
   3 Corollary.: If \( X_j \) are independent \( \mathcal{N}(\nu_j, \sigma^2_j) \), then
      \( \sum_j X_j \sim \mathcal{N}(\sum_j \nu_j, \sum_j \sigma^2_j) \).

2 Cauchy Example:
   1 \( X \) and \( W \) Cauchy, \( X \perp W \), and \( Y = X + W. \)
   2 \( Y/2 \) is Cauchy.
Gamma Example:

\[ X \sim \Gamma(k, \lambda) \quad \text{and} \quad W \sim \Gamma(m, \lambda), \quad X \perp W. \]

Result:

\[ Y = X + W \sim \Gamma(k + m, \lambda). \]

Direct convolution:

By extension, \( X_j \sim \Gamma(k_j, \lambda), \) independent, then

\[ Y = \sum_{j=1}^{n} X_j \sim \Gamma\left( \sum_{j} k_j, \lambda \right). \]

Special case with shape parameter an integer divided by two, scale parameter \( \frac{1}{2} \) is chi-square.
Gamma Example:

1. $X \sim \Gamma(k, \lambda)$ and $W \sim \Gamma(m, \lambda)$, and $X \perp W$. 

Result:

$Y = X + W \sim \Gamma(k + m, \lambda)$. 

Direct convolution:

By extension, $X_j \sim \Gamma(k_j, \lambda)$, independent, then $Y = \sum_{n} X_j \sim \Gamma(\sum_{j} k_j, \lambda)$. 

Special case with shape parameter an integer divided by two, scale parameter $1/2$ is chi-square.
Gamma Example:

1. $X \sim \Gamma(k, \lambda)$ and $W \sim \Gamma(m, \lambda)$, and $X \perp W$.
2. Result $Y = X + W \sim \Gamma(k + m, \lambda)$. 
Gamma Example:

1. $X \sim \Gamma(k, \lambda)$ and $W \sim \Gamma(m, \lambda)$, and $X \perp W$.
2. Result $Y = X + W \sim \Gamma(k + m, \lambda)$.
3. Direct convolution:
Gamma Example:

1. $X \sim \Gamma(k, \lambda)$ and $W \sim \Gamma(m, \lambda)$, and $X \perp W$.
2. Result $Y = X + W \sim \Gamma(k + m, \lambda)$.
3. Direct convolution:
4. By extension, $X_j \sim \Gamma(k_j, \lambda)$, independent, then $Y = \sum_{j=1}^{n} X_j \sim \Gamma(\sum_j k_j, \lambda)$. 

Special case with shape parameter an integer divided by two, scale parameter $1/2$ is chi-square.
Gamma Example:

1. $X \sim \Gamma(k, \lambda)$ and $W \sim \Gamma(m, \lambda)$, and $X \perp W$.
2. Result $Y = X + W \sim \Gamma(k + m, \lambda)$.
3. Direct convolution:
4. By extension, $X_j \sim \Gamma(k_j, \lambda)$, independent, then $Y = \sum_{j=1}^{n} X_j \sim \Gamma(\sum k_j, \lambda)$.
5. Special case with shape parameter an integer divided by two, scale parameter $1/2$ is \textit{chi-square}.
Counts of Events

The Exponential Expon($\lambda$) Distribution

Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent Expon($\lambda$) variables.

$N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\}$

The $X$'s are called inter-arrival times. See Fig. 1.

Number of events before a fixed time is Poisson.

Claim $N(t) \sim \text{Pois}(t \lambda)$.

By definition, $P(N(t) \geq k) = P(X_1 + \ldots + X_k \leq t)$.
Counts of Events

The Exponential Expon(1/\(\lambda\)) Distribution

Let \(X_1, \ldots, X_n, \ldots\) be a sequence of independent Expon(\(\lambda\)) variables.

\[ N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\} \]

The \(X\)'s are called inter-arrival times. See Fig. 1.

Number of events before a fixed time is Poisson.

Claim \(N(t) \sim \text{Pois}(t\lambda)\).
1 Counts of Events
   1 The Exponential $\text{Expon}(1/\lambda)$ Distribution
   2 Poisson distribution with expectation $\lambda$
Counts of Events

1. The Exponential $\text{Expon}(1/\lambda)$ Distribution
2. Poisson distribution with expectation $\lambda$
3. Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent $\text{Expon}(\lambda)$ variables
Counts of Events

1. The Exponential Expon\((1/\lambda)\) Distribution
2. Poisson distribution with expectation \(\lambda\)
3. Let \(X_1, \ldots, X_n, \ldots\) be a sequence of independent Expon\((\lambda)\) variables
4. \(N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\}\)
Counts of Events

1. The Exponential Expon(1/λ) Distribution
2. Poisson distribution with expectation λ
3. Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent Expon(λ) variables
4. $N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\}$
5. The X’s are called *inter-arrival times*. See Fig. 1.
Counts of Events

1. The Exponential Expon\((1/\lambda)\) Distribution
2. Poisson distribution with expectation \(\lambda\)
3. Let \(X_1, \ldots, X_n, \ldots\) be a sequence of independent Expon\((\lambda)\) variables
4. \(N(t) = \max\{n \geq 0| \sum_{j=1}^{n} X_j \leq t\}\)
5. The \(X\)'s are called *inter-arrival times*. See Fig. 1.

Number of events before a fixed time is Poisson
Counts of Events

1. The Exponential $\text{Expon}(1/\lambda)$ Distribution
2. Poisson distribution with expectation $\lambda$
3. Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent $\text{Expon}(\lambda)$ variables
4. $N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\}$
5. The $X$'s are called *inter-arrival times*. See Fig. 1.

Number of events before a fixed time is Poisson

1. Claim $N(t) \sim \text{Pois}(t\lambda)$. 

Counts of Events

1. The Exponential Expon(1/\(\lambda\)) Distribution
2. Poisson distribution with expectation \(\lambda\)
3. Let \(X_1, \ldots, X_n, \ldots\) be a sequence of independent Expon(\(\lambda\)) variables
4. \(N(t) = \max\{n \geq 0 | \sum_{j=1}^{n} X_j \leq t\}\)
5. The \(X\)'s are called *inter-arrival times*. See Fig. 1.

Number of events before a fixed time is Poisson

1. Claim \(N(t) \sim \text{Pois}(t\lambda)\).
2. By definition, \(P(N(t) \geq k) = P(X_1 + \ldots + X_k \leq t)\)
Describing movements of two variables:

1. Measure association using covariance

\[ \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \]

Alternate formulation:
\[ \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \], because

Definition is symmetric:
\[ \text{Cov}(X, Y) = \text{Cov}(Y, X) \].

Intuition:
\[ \text{Cov}(aX + b, Y) = a \text{Cov}(X, Y) \] because

Unrelated (independent) random variables have covariance 0:

Think of covariance as inner product of random variables; by this measure, the independence sign \( \perp \) represents perpendicular.
Describing movements of two variables:

1. Measure association using covariance
   - Definition: \( \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \)
Describing movements of two variables:

1 Measure association using covariance
   1 Definition: \( \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \)
   2 Alternate formulation: \( \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \), because...
Describing movements of two variables:

1. Measure association using covariance
   - Definition: \( \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \)
   - Alternate formulation: \( \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \), because
   - Definition is symmetric: \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \).
Describing movements of two variables:

1. Measure association using covariance
   - Definition: $\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$
   - Alternate formulation: $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$, because $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
   - Definition is symmetric: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

4. Intuition:
Describing movements of two variables:

1. Measure association using covariance
   - Definition: \( \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \)
   - Alternate formulation: \( \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \), because
   - Definition is symmetric: \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \).
   - Intuition:
     - \( \text{Cov}(aX + b, Y) = a\text{Cov}(X, Y) \) because

Unrelated (independent) random variables have covariance 0: Think of covariance as inner product of random variables; by this measure, the independence sign \( \perp \) represents perpendicular.
Describing movements of two variables:

1. Measure association using covariance
   - Definition: \( \text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \)
   - Alternate formulation: \( \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \), because
   - Definition is symmetric: \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \).
   - Intuition:
     - \( \text{Cov}(aX + b, Y) = a\text{Cov}(X, Y) \) because
     - Unrelated (independent) random variables have covariance 0:
Describing movements of two variables:

1 Measure association using covariance
   1 Definition: Cov(X, Y) = E((X - E(X))(Y - E(Y)))
   2 Alternate formulation: Cov(X, Y) = E(XY) - E(X)E(Y), because
   3 Definition is symmetric: Cov(X, Y) = Cov(Y, X).
   4 Intuition:
   5 Cov(aX + b, Y) = aCov(X, Y) because
   6 Unrelated (independent) random variables have covariance 0:
   7 Think of covariance as inner product of random variables; by this
      measure, the independence sign ⊥ represents perpendicular
Converse does not hold

Dependent random variables may have zero covariance: Ex., $(X, Y)$ taking values on $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ with equal probability.

Ex., variables uniform in unit circle.

Ex., variables taking values on unit circle with probabilities proportional to arc length.

Preliminaries: Schwartz's inequality: $\sqrt{\mathbb{E}(U^2)} \sqrt{\mathbb{E}(V^2)} \leq \mathbb{E}(UV) \leq \sqrt{\mathbb{E}(U^2)} \sqrt{\mathbb{E}(V^2)}$, holds $\iff$ one of these variables is a linear function of the other.

Proof: 

: Introduction: Lecture 19
Converse does not hold

Dependent random variables may have zero covariance:

Ex., \((X, Y)\) taking values on \(\{(0, 1), (0, -1), (1, 0), (-1, 0)\}\) with equal probability.

Ex., variables uniform in unit circle.

Ex., variables taking values on unit circle with probabilities proportional to arc length.

Preliminaries: Schwartz's inequality:

\[
\sqrt{\mathbb{E}(U^2)} \leq \mathbb{E}(UV) \leq \sqrt{\mathbb{E}(V^2)},
\]

holds \(\iff\) one of these variables is a linear function of the other.

Proof:
Converse does not hold

1. Dependent random variables may have zero covariance:
2. Ex., $(X, Y)$ taking values on $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ with equal probability. 2.
Converse does not hold

1. Dependent random variables may have zero covariance:
2. Ex., $(X, Y)$ taking values on \{$(0, 1), (0, -1), (1, 0), (-1, 0)$\} with equal probability. 2.
3. Ex., variables uniform in unit circle.
Converse does not hold

1. Dependent random variables may have zero covariance:
   Ex., $(X, Y)$ taking values on $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ with equal probability. 2.

2. Ex., variables uniform in unit circle.

3. Ex., variables taking values on unit circle with probabilities proportional to arc length.
Converse does not hold

1. Dependent random variables may have zero covariance:
   - Ex., \((X, Y)\) taking values on \{\((0, 1), (0, -1), (1, 0), (-1, 0)\)\} with equal probability.
   - Ex., variables uniform in unit circle.
   - Ex., variables taking values on unit circle with probabilities proportional to arc length.

2. Preliminaries: Schwartz’s inequality:
Converse does not hold

1. Dependent random variables may have zero covariance:
   1. Ex., \((X, Y)\) taking values on \(\{(0, 1), (0, -1), (1, 0), (-1, 0)\}\) with equal probability.
   2. Ex., variables uniform in unit circle.
   3. Ex., variables taking values on unit circle with probabilities proportional to arc length.

Preliminaries: Schwartz’s inequality:

1. \(\sqrt{E(U^2)} \cdot \sqrt{E(V^2)} \leq E(UV) \leq \sqrt{E(U^2)} \cdot \sqrt{E(V^2)},\)
Converse does not hold

Dependent random variables may have zero covariance:

Ex., \((X, Y)\) taking values on \(\{(0, 1), (0, -1), (1, 0), (-1, 0)\}\) with equal probability.

Ex., variables uniform in unit circle.

Ex., variables taking values on unit circle with probabilities proportional to arc length.

Preliminaries: Schwartz’s inequality:

\[-\sqrt{\mathbb{E}(U^2)}\sqrt{\mathbb{E}(V^2)} \leq \mathbb{E}(UV) \leq \sqrt{\mathbb{E}(U^2)}\sqrt{\mathbb{E}(V^2)},\]

= holds \iff one of these variables is a linear function of the other.
Converse does not hold

1. Dependent random variables may have zero covariance:
   1. Ex., \((X, Y)\) taking values on \(\{(0, 1), (0, -1), (1, 0), (-1, 0)\}\) with equal probability.
   2. Ex., variables uniform in unit circle.
   3. Ex., variables taking values on unit circle with probabilities proportional to arc length.

2. Preliminaries: Schwartz's inequality:
   1. \(-\sqrt{E(U^2)}\sqrt{E(V^2)} \leq E(UV) \leq \sqrt{E(U^2)}\sqrt{E(V^2)},\)
   2. \(=\) holds \(\iff\) one of these variables is a linear function of the other.
   3. Proof:
How big can covariance be?

Schwartz’s inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives

$$|\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y);$$

equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$.

Correlation measures association independent of scale.

Define correlation as $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$.

Properties:

Cauchy-Schwartz shows $|\rho(X, Y)| \leq 1$. Introduction: Lecture 19 19/42
How big can covariance be?

Schwartz's inequality applied to $X - E(X)$ and $Y - E(Y)$ gives
$$|Cov(X, Y)| \leq SD(X)SD(Y);$$

Correlation measures association independent of scale.

Define correlation as $ho(X, Y) = \frac{Cov(X, Y)}{SD(X)SD(Y)}$.

Properties:

1. Cauchy-Schwartz shows $|\rho(X, Y)| \leq 1$.
How big can covariance be?

1. Schwartz's inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives $|\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y)$;
2. equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$. 

Correlation measures association independent of scale.

Define correlation as $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$.

Properties:
1. Cauchy-Schwartz shows $|\rho(X, Y)| \leq 1$: Introduction: Lecture 19 19/42
1 How big can covariance be?
   1 Schwartz’s inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives
     $|\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y)$;
   2 equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$.

2 Correlation measures association independent of scale.
How big can covariance be?

Schwartz’s inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives

$$|\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y);$$

equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$.

Correlation measures association independent of scale.

Define correlation as $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$. 
How big can covariance be?

1. Schwartz’s inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives $|\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y)$;
2. equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$.

Correlation measures association independent of scale.

1. Define correlation as $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$.
2. Properties:
How big can covariance be?

1. Schwartz's inequality applied to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ gives
   \[ |\text{Cov}(X, Y)| \leq \text{SD}(X)\text{SD}(Y); \]
2. equality holds $\iff Y = aX + b$ for some $a \neq 0$ and $b$.

Correlation measures association independent of scale.

1. Define correlation as $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$.
2. Properties:
3. Cauchy-Schwartz shows $|\rho(X, Y)| \leq 1$
Moments of Sums of Two Variables

Variance of Sum is the sum of variances plus twice the covariance, because:

Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

If $X_1, \ldots, X_n$ are random variables, let $Y = \sum_{i=1}^{n} a_i X_i$ for constants $a_i$.

$E(Y) = \sum_{i=1}^{n} a_i \mu_i$.

Also $V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \text{Cov}(X_i, X_j)$.
Moments of Sums of Two Variables

Variance of Sum is the sum of variances plus twice the covariance, because:

Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

If $X_1, \ldots, X_n$ are random variables,

Let $Y = \sum_{i=1}^{n} a_i X_i$ for constants $a_i$.

$E(Y) = \sum_{i=1}^{n} a_i \mu_i$.

Also $V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i=1}^{n} a_i a_j \text{Cov}(X_i, X_j)$. 
Moments of Sums of Two Variables

1. Variance of Sum is the sum of variances plus twice the covariance, because:
2. Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$. 
Moments of Sums of Two Variables

Variance of Sum is the sum of variances plus twice the covariance, because:

Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.
Moments of Sums of Two Variables

- Variance of Sum is the sum of variances plus twice the covariance, because:
  - Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

- If $X_1, \ldots, X_n$ are random variables,
Moments of Sums of Two Variables

1. Variance of Sum is the sum of variances plus twice the covariance, because:

2. Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

1. If $X_1, \ldots, X_n$ are random variables,

2. Let $Y = \sum_{i=1}^{n} a_i X_i$ for constants $a_i$. 
Moments of Sums of Two Variables

1. Variance of Sum is the sum of variances plus twice the covariance, because:
2. Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

1. If $X_1, \ldots, X_n$ are random variables,
2. Let $Y = \sum_{i=1}^{n} a_i X_i$ for constants $a_i$.
3. $\mathbb{E}(Y) = \sum_{i=1}^{n} a_i \mu_i$. 

Also $\text{Var}(Y) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \text{Cov}(X_i, X_j)$. 
Moments of Sums of Two Variables

1. Variance of Sum is the sum of variances plus twice the covariance, because:
2. Suppose $X \perp Y$, implying $\text{Cov}(X, Y) = 0$.

Additivity extends easily to larger sums.

1. If $X_1, \ldots, X_n$ are random variables,
2. Let $Y = \sum_{i=1}^{n} a_i X_i$ for constants $a_i$.
3. $\mathbb{E}(Y) = \sum_{i=1}^{n} a_i \mu_i$.
4. Also

$$\text{V}(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \text{Cov}(X_i, X_j).$$
Variances add for sums of independent random variables.
Variance add for sums of independent random variables.

Take $X_1, \ldots, X_n$ are independent, $a_i$ constants.
Variances add for sums of independent random variables.

1. Take $X_1, \ldots, X_n$ are independent, $a_i$ constants.
2. Then $V \left( \sum_{i=1}^{n} a_i X_i \right) = \sum_{i=1}^{n} a_i^2 V(X_i)$
Two Special Multivariate Distributions

1 Multinomial distribution

A Multinomial distribution is an extension of the binomial distribution. It arises when you categorize $n$ independent items into one of $K$ categories. Let $X_i$ be the count in cell $i$, for $i \leq K$. The probability for category $i$, $\pi_i$, is such that $\sum_{i=1}^{K} \pi_i = 1$. The probability mass function is given by:

$$P (X_j = x_j \forall j) = \left( x_1 x_2 \cdots x_K \right) \prod_{i=1}^{K} \pi_i x_i$$

This stays multinomial after rearranging categories and after collapsing categories. Conditional and marginal distributions are also multinomial (possibly binomial). The expectation and variance of $X_i$ are:

$$E (X_i) = n \pi_i$$
$$V (X_i) = n \pi_i (1 - \pi_i)$$

The binomial distribution is a special case of the multinomial distribution. The covariance of $X_i$ and $X_j$ is:

$$\text{Cov} (X_i, X_j) = \begin{cases} n \pi_i \pi_j, & \text{if } i = j \\ -n \pi_i \pi_j, & \text{if } i \neq j \end{cases}$$

Independent Poissons conditioned on their sum are multinomial.
Two Special Multivariate Distributions

1. Multinomial distribution
   1. Extension of binomial

Result of categorizing $n$ independent items into one of $K$ categories.

$X_i$ is count in cell $i$, for $i \leq K$.

$\pi_i = \text{probability for category } i, i \leq K$.

$P (X_j = x_j \forall j) = \left(\frac{n}{x_1 x_2 \cdots x_K}\right) \prod_{i=1}^{K} \pi_i x_i$, because:

1. Stays multinomial after rearranging categories
2. Stays multinomial after collapsing categories
3. Conditional and marginal distributions: multinomial (possibly binomial)
4. Expectation and Variance:
   $E (X_i) = n \pi_i$ and $V (X_i) = n \pi_i (1 - \pi_i)$,
5. Binomial as special case of multinomial:
6. Trick: Use variance-of-sum formula to show $Cov (X_i, X_j) = -n \pi_i \pi_j$ if $i \neq j$.
7. Independent Poissons conditioned on sum are multinomial:
8. Distribution function in multiple dimensions
Two Special Multivariate Distributions

1. Multinomial distribution
   1. Extension of binomial
   2. Result of categorizing $n$ independent items into one of $K$ categories.
Two Special Multivariate Distributions

1. Multinomial distribution
   1. Extension of binomial
   2. Result of categorizing $n$ independent items into one of $K$ categories.
   3. $X_i$ is count in cell $i$, for $i \leq K$

$\pi_i = \text{probability for category } i, i \leq K.$

$P (X_j = x_j \forall j) = \left(\frac{n}{\prod_{i=1}^{K} \pi_i} \right).$

Stays multinomial after rearranging categories

Stays multinomial after collapsing categories

Conditional and marginal distributions: multinomial (possibly binomial)

Expectation and Variance: $E (X_i) = n \pi_i$ and $V (X_i) = n \pi_i (1 - \pi_i)$,

Binomial as special case of multinomial:

Trick: Use variance-of-sum formula to show $\text{Cov} (X_i, X_j) = -n \pi_i \pi_j$ if $i \neq j$.

Independent Poissons conditioned on sum are multinomial:

Distribution function in multiple dimensions
Two Special Multivariate Distributions

1. Multinomial distribution
   1. Extension of binomial
   2. Result of categorizing $n$ independent items into one of $K$ categories.
   3. $X_i$ is count in cell $i$, for $i \leq K$
   4. $\pi_i = \text{probability for category } i, \ i \leq K$. 

\[
P(X_j = x_j \forall j) = \left( \frac{n x_1 x_2 \cdots x_K}{\prod_{i=1}^K \pi_i x_i} \right)
\]

1. Stays multinomial after rearranging categories
2. Stays multinomial after collapsing categories
3. Conditional and marginal distributions: multinomial (possibly binomial)
4. Expectation and Variance: $E(X_i) = n \pi_i$ and $V(X_i) = n \pi_i (1 - \pi_i)$,
5. Binomial as special case of multinomial:
6. Trick: Use variance-of-sum formula to show $\text{Cov}(X_i, X_j) = -n \pi_i \pi_j$ if $i \neq j$.
7. Independent Poissons conditioned on sum are multinomial:
8. Distribution function in multiple dimensions
9. Introduction: Lecture 20 22/42
Two Special Multivariate Distributions

1. Multinomial distribution

   1. Extension of binomial
   2. Result of categorizing \( n \) independent items into one of \( K \) categories.
   3. \( X_i \) is count in cell \( i \), for \( i \leq K \)
   4. \( \pi_i \) = probability for category \( i \), \( i \leq K \).
   5. \( P(X_j = x_j \forall j) = \binom{n}{x_1 \ldots x_K} \prod_{i=1}^{K} \pi_i^{x_i} \), because:
## Two Special Multivariate Distributions

1. **Multinomial distribution**
   1. Extension of binomial
   2. Result of categorizing $n$ independent items into one of $K$ categories.
   3. $X_i$ is count in cell $i$, for $i \leq K$
   4. $\pi_i = \text{probability for category } i, i \leq K.$
   5. $P(X_j = x_j \forall j) = \left(\begin{array}{c} n \\ x_1 x_2 \cdots x_K \end{array}\right) \prod_{i=1}^{K} \pi_i^{x_i}$, because:
   6. Stays multinomial after rearranging categories

---

1. **Conditional and marginal distributions:** multinomial (possibly binomial)
2. **Expectation and Variance:**
   - $E(X_i) = n \pi_i$ and $V(X_i) = n \pi_i (1 - \pi_i)$,
3. **Binomial as special case of multinomial:**
4. **Trick:** Use variance-of-sum formula to show $\text{Cov}(X_i, X_j) = -n \pi_i \pi_j$ if $i \neq j$.
5. **Independent Poissons conditioned on sum are multinomial:**
6. **Distribution function in multiple dimensions**
Two Special Multivariate Distributions

1 Multinomial distribution
   1 Extension of binomial
   2 Result of categorizing $n$ independent items into one of $K$ categories.
   3 $X_i$ is count in cell $i$, for $i \leq K$
   4 $\pi_i = \text{probability for category } i, i \leq K$.
   5 $P(X_j = x_j \forall j) = \binom{n}{x_1x_2...x_K} \prod_{i=1}^{K} \pi_i^{x_i}$, because:
   6 Stays multinomial after rearranging categories
   7 Stays multinomial after collapsing categories
Two Special Multivariate Distributions

1 Multinomial distribution
   1 Extension of binomial
   2 Result of categorizing $n$ independent items into one of $K$ categories.
   3 $X_i$ is count in cell $i$, for $i \leq K$
   4 $\pi_i = $ probability for category $i$, $i \leq K$.
   5 $P (X_j = x_j \forall j) = \binom{n}{x_1 x_2 \ldots x_K} \prod_{i=1}^{K} \pi_i^{x_i}$, because:
   6 Stays multinomial after rearranging categories
   7 Stays multinomial after collapsing categories
   8 Conditional and marginal distributions: multinomial (possibly binomial)
Multinomial distribution

- Extension of binomial
- Result of categorizing \( n \) independent items into one of \( K \) categories.
- \( X_i \) is count in cell \( i \), for \( i \leq K \)
- \( \pi_i = \) probability for category \( i \), \( i \leq K \).
- \( P (X_j = x_j \forall j) = \left( \binom{n}{x_1x_2...x_K} \right) \prod_{i=1}^{K} \pi_i^{x_i} \), because:
- Stays multinomial after rearranging categories
- Stays multinomial after collapsing categories
- Conditional and marginal distributions: multinomial (possibly binomial)
- Expectation and Variance: \( E(X_i) = n\pi_i \) and \( V(X_i) = n\pi_i(1 - \pi_i) \),
Two Special Multivariate Distributions

1 Multinomial distribution

1 Extension of binomial
2 Result of categorizing \( n \) independent items into one of \( K \) categories.
3 \( X_i \) is count in cell \( i \), for \( i \leq K \)
4 \( \pi_i \) = probability for category \( i \), \( i \leq K \).
5 \( P (X_j = x_j \forall j) = \left( \frac{n}{x_1 x_2 \ldots x_K} \right) \prod_{i=1}^{K} \pi_i^{x_i} \), because:
6 Stays multinomial after rearranging categories
7 Stays multinomial after collapsing categories
8 Conditional and marginal distributions: multinomial (possibly binomial)
9 Expectation and Variance: \( E (X_i) = n\pi_i \) and \( V (X_i) = n\pi_i(1 - \pi_i) \),
10 Binomial as special case of multinomial:
Multinomial distribution

1. Extension of binomial
2. Result of categorizing \( n \) independent items into one of \( K \) categories.
3. \( X_i \) is count in cell \( i \), for \( i \leq K \)
4. \( \pi_i \) = probability for category \( i \), \( i \leq K \).
5. \( P (X_j = x_j \forall j) = \binom{n}{x_1, x_2, \ldots, x_K} \prod_{i=1}^{K} \pi_i^{x_i}, \) because:
6. Stays multinomial after rearranging categories
7. Stays multinomial after collapsing categories
8. Conditional and marginal distributions: multinomial (possibly binomial)
9. Expectation and Variance: \( E (X_i) = n\pi_i \) and \( V (X_i) = n\pi_i(1 - \pi_i) \),
10. Binomial as special case of multinomial:
11. Trick: Use variance-of-sum formula to show \( \text{Cov} (X_i, X_j) = -n\pi_i\pi_j \) if \( i \neq j \).
Two Special Multivariate Distributions

1 Multinomial distribution
   1 Extension of binomial
   2 Result of categorizing $n$ independent items into one of $K$ categories.
   3 $X_i$ is count in cell $i$, for $i \leq K$
   4 $\pi_i = \text{probability for category } i, \ i \leq K$
   5 $P(\mathbf{X} = \mathbf{x}) = \left(\frac{n}{x_1x_2...x_K}\right) \prod_{i=1}^{K} \pi_i^{x_i}$, because:
   6 Stays multinomial after rearranging categories
   7 Stays multinomial after collapsing categories
   8 Conditional and marginal distributions: multinomial (possibly binomial)
   9 Expectation and Variance: $E(X_i) = n\pi_i$ and $V(X_i) = n\pi_i(1 - \pi_i)$
   10 Binomial as special case of multinomial:
   11 Trick: Use variance-of-sum formula to show $\text{Cov}(X_i, X_j) = -n\pi_i\pi_j$ if $i \neq j$.
   12 Independent Poissons conditioned on sum are multinomial:
Two Special Multivariate Distributions

1. Multinomial distribution
   1. Extension of binomial
   2. Result of categorizing $n$ independent items into one of $K$ categories.
   3. $X_i$ is count in cell $i$, for $i \leq K$
   4. $\pi_i = \text{probability for category } i$, $i \leq K$.
   5. $P(X_j = x_j \forall j) = \binom{n}{x_1,x_2,...,x_K} \prod_{i=1}^{K} \pi_i^{x_i}$, because:
   6. Stays multinomial after rearranging categories
   7. Stays multinomial after collapsing categories
   8. Conditional and marginal distributions: multinomial (possibly binomial)
   9. Expectation and Variance: $E(X_i) = n\pi_i$ and $V(X_i) = n\pi_i(1 - \pi_i)$,
   10. Binomial as special case of multinomial:
   11. Trick: Use variance-of-sum formula to show $\text{Cov}(X_i, X_j) = -n\pi_i\pi_j$ if $i \neq j$.
   12. Independent Poissons conditioned on sum are multinomial:
   13. Distribution function in multiple dimensions
Bivariate Normal Distribution: Special case with standard normal marginals.

Joint probability density function:

\[ f_{Y_1, Y_2}(y_1, y_2) = \exp\left( -\frac{(y_2 - 2\rho y_1 y_2)^2}{2(1 - \rho^2)} \right) \sqrt{\frac{2}{\pi}} \sqrt{1 - \rho^2}. \]

Also \( \text{Cov}(Y_1, Y_2) = \rho \).

\( \rho \) is also the correlation.

General Form of Bivariate Normal:

Here \( Y_1 \sim N(\mu_1, \sigma_1^2) \), \( Y_2 | Y_1 = N(\mu_2 + \sigma_2^2 \rho (Y_1 - \mu_1) / \sigma_1, \sigma_2^2 (1 - \rho^2)) \).

Marginal distribution from bivariate normal is still normal.
Bivariate Normal Distribution: Special case with standard normal marginals.

1. \( Y_1 \sim \mathcal{N}(0, 1), \ Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)
Bivariate Normal Distribution: Special case with standard normal marginals.

1. \( Y_1 \sim \mathcal{N}(0, 1) \), \( Y_2|Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)

2. Joint probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2(1 - \rho^2)}\right)}{2\pi \sqrt{1 - \rho^2}}.
\]
Bivariate Normal Distribution: Special case with standard normal marginals.

1. \( Y_1 \sim \mathcal{N}(0, 1) \), \( Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)

2. Joint probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2(1 - \rho^2)}\right)}{2\pi \sqrt{1 - \rho^2}}.
\]

3. Argument is symmetric if you change direction of conditioning:

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\frac{y_2^2}{2}\right) \exp\left(-\frac{(y_1 - \rho y_2)^2}{2(1 - \rho^2)}\right)}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi \sqrt{1 - \rho^2}}}.
\]
Bivariate Normal Distribution: Special case with standard normal marginals.

1. $Y_1 \sim \mathcal{N}(0, 1)$, $Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2)$

2. Joint probability density function

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(- (y_1^2 - 2\rho y_1 y_2 + y_2^2)/(2(1 - \rho^2)))}{2\pi \sqrt{1 - \rho^2}}.$$

3. Argument is symmetric if you change direction of conditioning:

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(-y_2^2/2) \exp(- (y_1 - \rho y_2)^2/(2(1 - \rho^2)))}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi \sqrt{1 - \rho^2}}}.$$

4. Also $\text{Cov}(Y_1, Y_2) = \rho$. 
Bivariate Normal Distribution: Special case with standard normal marginals.

\( Y_1 \sim \mathcal{N}(0, 1), \ Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)

2 Joint probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2(1 - \rho^2)}\right)}{2\pi \sqrt{1 - \rho^2}}.
\]

3 Argument is symmetric if you change direction of conditioning:

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\frac{y_2^2/2}{\sqrt{2\pi}}\right) \exp\left(-\frac{(y_1 - \rho y_2)^2}{2(1 - \rho^2)}\right)}{\sqrt{2\pi} \sqrt{2\pi} \sqrt{1 - \rho^2}}.
\]

4 Also \( \text{Cov}(Y_1, Y_2) = \rho \).

5 \( \rho \) is also the correlation.
Bivariate Normal Distribution: Special case with standard normal marginals.

1. $Y_1 \sim \mathcal{N}(0, 1)$, $Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2)$

2. Joint probability density function

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(-(y_1^2 - 2\rho y_1 y_2 + y_2^2)/(2(1 - \rho^2)))}{2\pi \sqrt{1 - \rho^2}}.$$ 

3. Argument is symmetric if you change direction of conditioning:

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(-y_2^2/2) \exp(-(y_1 - \rho y_2)^2/(2(1 - \rho^2)))}{\sqrt{2\pi}} \frac{\exp(-y_1^2/2)}{\sqrt{2\pi} \sqrt{1 - \rho^2}}$$

4. Also $\text{Cov}(Y_1, Y_2) = \rho$.

5. $\rho$ is also the correlation.

General Form of Bivariate Normal
Bivariate Normal Distribution: Special case with standard normal marginals.

1. $Y_1 \sim \mathcal{N}(0, 1), \ Y_2|Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2)$
2. Joint probability density function
   \[
   f_{Y_1,Y_2}(y_1, y_2) = \frac{\exp\left(-\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right)/(2(1 - \rho^2))\right)}{2\pi \sqrt{1 - \rho^2}}.
   \]
3. Argument is symmetric if you change direction of conditioning:
   \[
   f_{Y_1,Y_2}(y_1, y_2) = \frac{\exp\left(-y_2^2/2\right) \exp\left(-\left(y_1 - \rho y_2\right)^2/(2(1 - \rho^2))\right)}{\sqrt{2\pi} \sqrt{2\pi} \sqrt{1 - \rho^2}}.
   \]
4. Also $\text{Cov}(Y_1, Y_2) = \rho$.
5. $\rho$ is also the correlation.

General Form of Bivariate Normal

1. Here $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \ Y_2|Y_1 = \mathcal{N}(\mu_2 + \sigma_2 \rho (Y_1 - \mu_1)/\sigma_1, \sigma_2^2(1 - \rho^2))$. 


Bivariate Normal Distribution: Special case with standard normal marginals.

1. \( Y_1 \sim \mathcal{N}(0, 1), \ Y_2 | Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)

2. Joint probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right)/(2(1 - \rho^2))\right)}{2\pi \sqrt{1 - \rho^2}}.
\]

3. Argument is symmetric if you change direction of conditioning:

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-y_2^2/2\right) \exp\left(-\left(y_1 - \rho y_2\right)^2/(2(1 - \rho^2))\right)}{\sqrt{2\pi} \sqrt{2\pi} \sqrt{1 - \rho^2}}.
\]

4. Also \( \text{Cov}(Y_1, Y_2) = \rho \).

5. \( \rho \) is also the correlation.

General Form of Bivariate Normal

1. Here \( Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \ Y_2 | Y_1 = \mathcal{N}(\mu_2 + \sigma_2 \rho (Y_1 - \mu_1)/\sigma_1, \sigma_2^2(1 - \rho^2)) \).

2. Probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(y_1 - \mu_1)}{\sigma_1} \frac{(y_2 - \mu_2)}{\sigma_2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2}\right)/(2(1 - \rho^2))\right)}{2\sigma_1 \sigma_2 \pi \sqrt{1 - \rho^2}}
\]
Bivariate Normal Distribution: Special case with standard normal marginals.

1. \( Y_1 \sim \mathcal{N}(0, 1), \ Y_2|Y_1 = \mathcal{N}(\rho Y_1, 1 - \rho^2) \)

2. Joint probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(-\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right)/(2(1 - \rho^2)))}{2\pi \sqrt{1 - \rho^2}}.
\]

3. Argument is symmetric if you change direction of conditioning:

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp(-y_2^2/2) \exp(-\left(y_1 - \rho y_2\right)^2/(2(1 - \rho^2)))}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi \sqrt{1 - \rho^2}}}.
\]

4. Also \( \text{Cov}(Y_1, Y_2) = \rho \).

5. \( \rho \) is also the correlation.

2. General Form of Bivariate Normal

1. Here \( Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \ Y_2|Y_1 = \mathcal{N}(\mu_2 + \sigma_2 \rho(\ Y_1 - \mu_1)/\sigma_1, \sigma_2^2(1 - \rho^2)). \)

2. Probability density function

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{\exp\left(-\left(\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)}{\sigma_1}\frac{(y_2-\mu_2)}{\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right)/(2(1 - \rho^2))\right)}{2\sigma_1\sigma_2\pi \sqrt{1 - \rho^2}}.
\]

3. Marginal distribution from bivariate normal is still normal.
More on the Multivariate Normal

Sums of components of bivariate normals are still normal.

Correlation 0 implies independence for multivariate normals. Always, independence implies zero correlation. Bivariate normality is needed for the opposite direction.
More on the Multivariate Normal

Sums of components of bivariate normals are still normal.
More on the Multivariate Normal

1. Sums of components of bivariate normals are still normal.

2. Correlation 0 implies independence for multivariate normals.
More on the Multivariate Normal

1. Sums of components of bivariate normals are still normal.

2. Correlation 0 implies independence for multivariate normals
   - Always, independence implies zero correlation
More on the Multivariate Normal

1. Sums of components of bivariate normals are still normal.
2. Correlation 0 implies independence for multivariate normals
   - Always, independence implies zero correlation
   - Bivariate normality is needed for the opposite direction.
More on the Multivariate Normal

1. Sums of components of bivariate normals are still normal.

2. Correlation 0 implies independence for multivariate normals
   - Always, independence implies zero correlation
   - Bivariate normality is needed for the opposite direction.
   - Multivariate normal not just a joint distribution with normal marginals.
Measure association using conditional distributions:

1. Measure association using conditional expectation:

\[
E \left( g(X, Y) \mid X = x \right) = \int_Y g(x, y) f_Y \left( y \mid x \right) \, dy \\
\text{or} \sum_{y \in Y} g(x, y) P \left( y \mid x \right)
\]

2. Use notation \( E(Y \mid X) \) when no specific \( x \) is in mind.

3. Example: Suppose that \( X \) and \( Y \) are uniform on triangle \( 0 \leq X \leq Y \leq 1 \).

4. Conditional on \( X \), functions of \( X \) behave like constants.
Measure association using conditional distributions:

1. Measure association using conditional expectation:

   Denote conditional expectation as before:
   
   $\mathbb{E}(g(X, Y) | X = x) = \int_y g(x, y) f_{Y|X}(y | x) \, dy$ or $\sum_{y \in Y} g(x, y) P(y | x)$

2. Use notation $\mathbb{E}(Y | X)$ when no specific $x$ is in mind.

3. Example: Suppose that $X$ and $Y$ are uniform on triangle $0 \leq X \leq Y \leq 1$.


5. Iterated Expectation

   Expectation is average of conditional expectations:
   
   $\mathbb{E}(g(Y)) = \mathbb{E}(\mathbb{E}(g(Y) | X))$, because ...

   : Introduction: Lecture 21
Measure association using conditional distributions:

1. Measure association using conditional expectation:
   - Denote conditional expectation as before:
     \[ E(g(X, Y)|X = x) = \int_Y g(x, y)f_{Y|X}(y|x) \, dy \text{ or } \sum_{y \in Y} g(x, y)P(y|x) \]
   - Use notation \( E(Y|X) \) when no specific \( x \) is in mind.

2. Example: Suppose that \( X \) and \( Y \) are uniform on triangle \( 0 \leq X \leq Y \leq 1 \).
   - Conditional on \( X \), functions of \( X \) behave like constants.
Measure association using conditional distributions:

1. Measure association using conditional expectation:
   - Denote conditional expectation as before:
     \[ \mathbb{E}(g(X, Y)|X = x) = \int_Y g(x, y)f_{Y|X}(y|x) \, dy \text{ or } \sum_{y \in Y} g(x, y)P(y|x) \]
   - Use notation \( \mathbb{E}(Y|X) \) when no specific \( x \) is in mind.
   - Example: Suppose that \( X \) and \( Y \) are uniform on triangle \( 0 \leq X \leq Y \leq 1 \).

2. Iterated Expectation
   - Expectation is average of conditional expectations:
     \[ \mathbb{E}(g(Y)) = \mathbb{E}(\mathbb{E}(g(Y)|X)) \]
     because

3. Introduction: Lecture 21
Measure association using conditional distributions:

1. Measure association using conditional expectation:
   - Denote conditional expectation as before:
     \[ E (g(X, Y)|X = x) = \int_y g(x, y)f_Y|X(y|x) \, dy \text{ or } \sum_{y \in Y} g(x, y)P(y|x) \]
   - Use notation \( E(Y|X) \) when no specific \( x \) is in mind.
   - Example: Suppose that \( X \) and \( Y \) are uniform on triangle \( 0 \leq X \leq Y \leq 1 \).
   - Conditional on \( X \), functions of \( X \) behave like constants.
Measure association using conditional distributions:

1. Measure association using conditional expectation:
   1. Denote conditional expectation as before:
      \[
      E(g(X, Y)|X = x) = \int_Y g(x, y)f_{Y|X}(y|x) \, dy \quad \text{or} \quad \sum_{y \in Y} g(x, y)P(y|x)
      \]
   2. Use notation \(E(Y|X)\) when no specific \(x\) is in mind.
   3. Example: Suppose that \(X\) and \(Y\) are uniform on triangle \(0 \leq X \leq Y \leq 1\).
   4. Conditional on \(X\), functions of \(X\) behave like constants.

2. Iterated Expectation
Measure association using conditional distributions:

1. Measure association using conditional expectation:
   - Denote conditional expectation as before:
     \[ E(g(X, Y) | X = x) = \int_Y g(x, y) f_{Y|X}(y|x) \, dy \quad \text{or} \quad \sum_{y \in Y} g(x, y) P(y|x) \]
   - Use notation \( E(Y|X) \) when no specific \( x \) is in mind.
   - Example: Suppose that \( X \) and \( Y \) are uniform on triangle \( 0 \leq X \leq Y \leq 1 \).
   - Conditional on \( X \), functions of \( X \) behave like constants.

2. Iterated Expectation
   - Expectation is average of conditional expectations:
     \[ E(g(Y)) = E(E(g(Y)|X)), \text{ because} \]
Conditional Expectation and Prediction

We want to predict $Y$ from $X$. Call prediction $d(X)$. Error measured by mean squared error $E((Y - d(X))^2)$. Minimize for $X$ held fixed: $d^*(X) = E(Y | X)$. Minimum mean squared error is conditional variance $E((Y - E(Y | X))^2 | X) = E(Y^2 | X) - E(Y | X)^2 = V(Y | X)$. Similarly, the conditional median minimizes absolute deviation. Degenerate case: Prediction for one random variable without using any other information is the expectation.
Conditional Expectation and Prediction

We want to predict \( Y \) from \( X \).
Conditional Expectation and Prediction

1. We want to predict \( Y \) from \( X \).
2. Call prediction \( d(X) \).

Error measured by mean squared error

\[
E \left( (Y - d(X))^2 \right)
\]

Minimize for \( X \) held fixed:

\[
d^*(X) = E(Y | X)
\]

Minimum mean squared error is conditional variance

\[
E \left( (Y - E(Y | X))^2 | X \right) = E(Y^2 | X) - E(Y | X)^2 = \text{Var}(Y | X)
\]

Similarly, the conditional median minimizes absolute deviation.

Degenerate case: Prediction for one random variable without using any other information is the expectation.
Conditional Expectation and Prediction

1. We want to predict $Y$ from $X$.
2. Call prediction $d(X)$.
3. Error measured by mean squared error $E((Y - d(X))^2)$.

Minimum mean squared error is conditional variance $E((Y - E(Y|X))^2) = E(Y^2|X) - E(Y|X)^2 = \text{V}(Y|X)$.

Similarly, the conditional median minimizes absolute deviation.

Degenerate case: Prediction for one random variable without using any other information is the expectation.
Conditional Expectation and Prediction

1. We want to predict $Y$ from $X$.
2. Call prediction $d(X)$.
3. Error measured by mean squared error $E((Y - d(X))^2)$.
4. Minimize for $X$ held fixed: $d^*(X) = E(Y|X)$. 

Minimum mean squared error is conditional variance $E(Y^2|X) - E(Y|X)^2 = V(Y|X)$. Similarly, the conditional median minimizes absolute deviation.

Degenerate case: Prediction for one random variable without using any other information is the expectation.
Conditional Expectation and Prediction

1. We want to predict $Y$ from $X$.
2. Call prediction $d(X)$.
3. Error measured by mean squared error $E((Y - d(X))^2)$.
4. Minimize for $X$ held fixed: $d^*(X) = E(Y|X)$
5. Minimum mean squared error is conditional variance $E((Y - E(Y|X))^2|X) = E(Y^2|X) - E(Y|X)^2 = V(Y|X)$.
Conditional Expectation and Prediction

1. We want to predict $Y$ from $X$.
2. Call prediction $d(X)$.
3. Error measured by mean squared error $E \left( (Y - d(X))^2 \right)$.
4. Minimize for $X$ held fixed: $d^*(X) = E(Y|X)$
5. Minimum mean squared error is conditional variance
   $E \left( (Y - E(Y|X))^2 | X \right) = E(Y^2|X) - E(Y|X)^2 = V(Y|X)$.
6. Similarly, the conditional median minimizes absolute deviation.
Conditional Expectation and Prediction

1. We want to predict $Y$ from $X$.
2. Call prediction $d(X)$.
3. Error measured by mean squared error $E((Y - d(X))^2)$.
4. Minimize for $X$ held fixed: $d^*(X) = E(Y|X)$.
5. Minimum mean squared error is conditional variance $E((Y - E(Y|X))^2|X) = E(Y^2|X) - E(Y|X)^2 = V(Y|X)$.
6. Similarly, the conditional median minimizes absolute deviation.
7. Degenerate case: Prediction for one random variable without using any other information is the expectation.
Conditional expectation example continued.

Recall iterated expectation formula: 
\[
E(Y) = E(E(Y|X))
\]

Counterpart for variance: 
\[
V(Y) = V(E(Y|X)) + E(V(Y|X)) 
\geq E(V(Y|X))
\]
Conditional expectation example continued.

From example above, prediction is \( \frac{1}{2}(1 + X) \),
Conditional expectation example continued.

From example above, prediction is $\frac{1}{2}(1 + X)$,

Marginal and Conditional Variance
Conditional expectation example continued.

From example above, prediction is $\frac{1}{2}(1 + X)$,

Marginal and Conditional Variance

Recall iterated expectation formula: $E(Y) = E(E(Y|X))$
Conditional expectation example continued.

From example above, prediction is $\frac{1}{2}(1 + X),$

Marginal and Conditional Variance

Recall iterated expectation formula: $E(Y) = E(E(Y|X))$

Counterpart for variance:

$V(Y) = V(E(Y|X)) + E(V(Y|X)) \geq E(V(Y|X))$
Order Statistics

Definition of Order Statistics
Order Statistics

1 Definition of Order Statistics

Suppose you have a set of random variables $X_1, \ldots, X_n$. 

Order Statistic Preliminaries

Notation: Let $X(1), X(2), \ldots, X(n)$ be the ordered values.

Assumptions:

Easy Cases: First and last order statistics.

Maximum:

Minimum:

Assumptions in both cases:
Definition of Order Statistics

1. Suppose you have a set of random variables $X_1, \ldots, X_n$
2. You are interested in a summary that calculated from their order.
Order Statistics

1. Definition of Order Statistics
   1. Suppose you have a set of random variables $X_1, \ldots, X_n$
   2. You are interested in a summary that calculated from their order.

2. Order Statistic Preliminaries
Order Statistics

1. Definition of Order Statistics
   1. Suppose you have a set of random variables $X_1, \ldots, X_n$.
   2. You are interested in a summary that calculated from their order.

2. Order Statistic Preliminaries
   1. Notation: Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered values.
Order Statistics

1 Definition of Order Statistics
   1 Suppose you have a set of random variables $X_1, \ldots, X_n$
   2 You are interested in a summary that calculated from their order.

2 Order Statistic Preliminaries
   1 Notation: Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered values.
   2 Assumptions:
Definition of Order Statistics

1. Suppose you have a set of random variables $X_1, \ldots, X_n$.
2. You are interested in a summary that calculated from their order.

Order Statistic Preliminaries

1. Notation: Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered values.
2. Assumptions:

Easy Cases: First and last order statistics.
Order Statistics

1 Definition of Order Statistics
   1 Suppose you have a set of random variables $X_1, \ldots, X_n$
   2 You are interested in a summary that calculated from their order.

2 Order Statistic Preliminaries
   1 Notation: Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered values.
   2 Assumptions:

3 Easy Cases: First and last order statistics.
   1 Maximum:
Definition of Order Statistics

Suppose you have a set of random variables \( X_1, \ldots, X_n \).

You are interested in a summary that calculated from their order.

Order Statistic Preliminaries

Notation: Let \( X_{(1)}, X_{(2)}, \ldots, X_{(n)} \) be the ordered values.

Assumptions:

Easy Cases: First and last order statistics.

Maximum:

Minimum:
# Order Statistics

1. **Definition of Order Statistics**
   1. Suppose you have a set of random variables $X_1, \ldots, X_n$.
   2. You are interested in a summary that calculated from their order.

2. **Order Statistic Preliminaries**
   1. Notation: Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered values.
   2. Assumptions:

3. **Easy Cases: First and last order statistics.**
   1. Maximum:
   2. Minimum:
   3. Assumptions in both cases:
Densities in general:

Given a random variable $X$, we want to evaluate the probability density function $f_X(x)$ at $x$. For the special case of odd sample size, we can use the following examples:

1. $X_j \sim \text{Exponential} \text{ with rate } \lambda$.
2. $X_j \sim \text{Beta}(1, 1) = \text{Uniform}(0, 1)$. 

Introduction: Lecture 22
Densities in general:

- General in sense of $j$ not necessarily either 1 or $n$
Densities in general:

1. General in sense of $j$ not necessarily either 1 or $n$
2. Pick $j$ and value $x$ where you want to evaluate the probability density function of $X_{(j)}$. 

\[
 f_{X_{(j)}}(x) = \binom{n}{j-1} \frac{1}{n-j+1} F_{X}(x)^{j-1} (1-F_{X}(x))^{n-j}, \text{ because:}
\] 

Special case of median for odd sample size:

1. $X_{j} \sim \text{Expon with rate } \lambda$.
2. $X_{j} \sim \text{Beta}(1, 1) = \text{Unif}(0, 1)$. 

---

: Introduction: Lecture 22
Densities in general:

1. General in sense of $j$ not necessarily either 1 or $n$
2. Pick $j$ and value $x$ where you want to evaluate the probability density function of $X_{(j)}$.
3. $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1 - F_X(x))^{n-j}$, because:

Special case of median for odd sample size:

Examples:
1. $X_{(j)} \sim \text{Expon}$ with rate $\lambda$.
2. $X_{(j)} \sim \text{Beta}(1, 1) = \text{Unif}(0, 1)$. 

: Introduction: Lecture 22
1 Densities in general:

1. General in sense of \( j \) not necessarily either 1 or \( n \)
2. Pick \( j \) and value \( x \) where you want to evaluate the probability density function of \( X(j) \).
3. \[ f_{X(j)}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1 - F_X(x))^{n-j}, \text{ because:} \]
4. Special case of median for odd sample size:
Densities in general:

1. General in sense of $j$ not necessarily either 1 or $n$
2. Pick $j$ and value $x$ where you want to evaluate the probability density function of $X(j)$.
3. $f_{X(j)}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1 - F_X(x))^{n-j}$, because:
4. Special case of median for odd sample size:

Examples:
Densities in general:

1. General in sense of \( j \) not necessarily either 1 or \( n \)
2. Pick \( j \) and value \( x \) where you want to evaluate the probability density function of \( X_j \).
3. \[
f_{X_{(j)}}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x) (1 - F_X(x))^{n-j},\]
because:
4. Special case of median for odd sample size:

Examples:

1. \( X_j \sim \text{Expon} \) with rate \( \lambda \).
Densities in general:

1. General in sense of $j$ not necessarily either 1 or $n$
2. Pick $j$ and value $x$ where you want to evaluate the probability density function of $X_{(j)}$.
3. $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1 - F_X(x))^{n-j}$, because:
4. Special case of median for odd sample size:

Examples:

1. $X_j \sim \text{Expon}$ with rate $\lambda$.
2. $X_j \sim \text{Beta}(1, 1) = \text{Unif}(0, 1)$. 
Distributions Derived from the Normal

1 Distribution of the Sample Mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

Moments Without Assuming Independence:

Moments Assuming Independence:

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

Shape (assuming independence, normality):

Distribution of Sum of Squares

Sum of Squares from mean can be written as the sum of squared independent random variables.

Let $Q_n = \sum_{j=1}^{n}(X_j - \bar{X}_n)^2$, $\bar{X}_n = \frac{\sum_{j=1}^{n} X_j}{n}$.

Express in terms of quantities with last omitted:

$Q_n = \sum_{j=1}^{n}(X_j - \bar{X}_n)^2 = \sum_{j=2}^{n-1}(X_j - \bar{X}_n - 1)^2 + (X_n - \bar{X}_n)^2$.

Claim:

$Q_n = \sum_{j=2}^{n}(X_j - \bar{X}_j - 1)^2 \left( 1 - \frac{1}{j} \right)$.

Before squaring, each summand has expectation 0.

Summands are independent:

Before squaring, each summand has variance $\sigma^2$.

Hence $Q_n / \sigma^2$ has same distribution as sum of $n - 1$ independent squared normals: $\chi^2_{n-1}$.

$n - 1$ is called the degrees of freedom.

Furthermore, $\bar{X}$ and $Q_n$ are independent.
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
2. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.
Distributions Derived from the Normal

1. Distribution of the Sample Mean \( \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \).
   - Let \( \mu = \mathbb{E}(X_i) \), \( \sigma^2 = \text{Var}(X_i) \).
   - Moments Without Assuming Independence:

2. Moments Assuming Independence:
   \[ \text{Var}(\bar{X}) = \text{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sigma^2}{n} \]
   because

3. Shape (assuming independence, normality):
   - Distribution of Sum of Squares
     - Sum of Squares from mean can be written as the sum of squared independent random variables.
     - Let \( Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 \), \( \bar{X}_n = \frac{\sum_{j=1}^{n} X_j}{n} \)

4. Express in terms of quantities with last omitted:
   \[ Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 + (X_n - \bar{X}_n)^2 \]

5. Claim:
   \[ Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j) \]

6. Before squaring, each summand has expectation 0.

7. Summands are independent:
   - Before squaring, each summand has variance \( \sigma^2 \).

8. Hence \( Q_n/\sigma^2 \) has same distribution as sum of \( n-1 \) independent squared normals: \( \chi^2_{n-1} \).

9. \( n-1 \) is called the degrees of freedom.

10. Furthermore, \( \bar{X} \) and \( Q_n \) are independent.
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i/n$.
   1. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2. Moments Without Assuming Independence:

3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \sigma^2/n$, because
Distributions Derived from the Normal

1 Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.

1 Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2 Moments Without Assuming Independence:

3 Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \sigma^2 / n$, because

4 Shape (assuming independence, normality):
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   - Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2. Moments Without Assuming Independence:

3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left( \frac{1}{n} \right)^2 = \sigma^2 / n$, because

4. Shape (assuming independence, normality):

2. Distribution of Sum of Squares

Before squaring, each summand has expectation 0.

Summands are independent:

Before squaring, each summand has variance $\sigma^2$.

Hence $Q_n / \sigma^2$ has same distribution as sum of $n-1$ independent squared normals: $\chi^2_{n-1}$.

$n-1$ is called the degrees of freedom.

Furthermore, $\bar{X}$ and $Q_n$ are independent.
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   - Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.
   - Moments Without Assuming Independence:
   - Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left( \frac{1}{n} \right)^2 = \sigma^2 / n$, because
   - Shape (assuming independence, normality):

2. Distribution of Sum of Squares
   - Sum of Squares from mean can be written as the sum of squared independent random variables.
Distributions Derived from the Normal

1 Distribution of the Sample Mean \( \bar{X} = \sum_{i=1}^{n} X_i / n \).
   1 Let \( \mu = E(X_i) \), \( \sigma^2 = V(X_i) \).
2 Moments Without Assuming Independence:
3 Moments Assuming Independence: \( V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left( \frac{1}{n} \right)^2 = \sigma^2 / n \), because
4 Shape (assuming independence, normality):

2 Distribution of Sum of Squares
1 Sum of Squares from mean can be written as the sum of squared independent random variables.
2 Let \( Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 \), \( \bar{X}_n = \sum_{j=1}^{n} X_j / n \).
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   - Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2. Moments Without Assuming Independence:
   - Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i)(\frac{1}{n})^2 = \sigma^2 / n$, because

3. Shape (assuming independence, normality):

4. Distribution of Sum of Squares
   - Sum of Squares from mean can be written as the sum of squared independent random variables.
   - Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$
   - Express in terms of quantities with last omitted:
     $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.

Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_j - 1)^2(1 - 1/j)$.

1. Before squaring, each summand has expectation 0.
2. Summands are independent:
3. Before squaring, each summand has variance $\sigma^2$.
4. Hence $Q_n/\sigma^2$ has same distribution as sum of $n-1$ independent squared normals: $\chi^2_{n-1}$.
5. $n-1$ is called the degrees of freedom.

Furthermore, $\bar{X}$ and $Q_n$ are independent.
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.

   - Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2. Moments Without Assuming Independence:

3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \sigma^2 / n$.

   - because

4. Shape (assuming independence, normality):

   - Distribution of Sum of Squares

   - Sum of Squares from mean can be written as the sum of squared independent random variables.

   - Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$

   - Express in terms of quantities with last omitted:

     $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.

   - Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j)$. 
Distributions Derived from the Normal

1. Distribution of the Sample Mean \( \bar{X} = \sum_{i=1}^{n} X_i / n \).
   - Let \( \mu = \mathbb{E}(X_i) \), \( \sigma^2 = \mathbb{V}(X_i) \).

2. Moments Without Assuming Independence:

3. Moments Assuming Independence: \( \mathbb{V}(\bar{X}) = \mathbb{V}(\sum_{i=1}^{n} X_i) \left( \frac{1}{n} \right)^2 = \sigma^2 / n \), because

4. Shape (assuming independence, normality):

2. Distribution of Sum of Squares

1. Sum of Squares from mean can be written as the sum of squared independent random variables.

2. Let \( Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 \), \( \bar{X}_n = \sum_{j=1}^{n} X_j / n \)

3. Express in terms of quantities with last omitted:
   \( Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2 \).

4. Claim: \( Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2(1 - 1/j) \).

5. Before squaring, each summand has expectation 0.
Distributions Derived from the Normal

Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.

1. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

Moments Without Assuming Independence:

Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \sigma^2 / n$, because

Shape (assuming independence, normality):

Distribution of Sum of Squares

Sum of Squares from mean can be written as the sum of squared independent random variables.

1. Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$

2. Express in terms of quantities with last omitted:

$Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.

3. Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2(1 - 1/j)$.

4. Before squaring, each summand has expectation 0.

5. Summands are independent:
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.
   
   Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.

2. Moments Without Assuming Independence:

3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \frac{\sigma^2}{n}$, because

4. Shape (assuming independence, normality):

2. Distribution of Sum of Squares

1. Sum of Squares from mean can be written as the sum of squared independent random variables.

2. Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$

3. Express in terms of quantities with last omitted:
   
   $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.

4. Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j)$.

5. Before squaring, each summand has expectation 0.

6. Summands are independent:

7. Before squaring, each summand has variance $\sigma^2$. 
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   1. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.
   2. Moments Without Assuming Independence:
   3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) (\frac{1}{n})^2 = \sigma^2 / n$, because
   4. Shape (assuming independence, normality):

2. Distribution of Sum of Squares
   1. Sum of Squares from mean can be written as the sum of squared independent random variables.
   2. Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$
   3. Express in terms of quantities with last omitted:
      $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.
   4. Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j)$.
   5. Before squaring, each summand has expectation 0.
   6. Summands are independent:
   7. Before squaring, each summand has variance $\sigma^2$.
   8. Hence $Q_n/\sigma^2$ has same distribution as sum of $n - 1$ independent squared normals: $\chi^2_{n-1}$. 

: Introduction: Lecture 23
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   1. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.
   2. Moments Without Assuming Independence:
   3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) \left(\frac{1}{n}\right)^2 = \sigma^2 / n$.

2. Distribution of Sum of Squares
   1. Sum of Squares from mean can be written as the sum of squared independent random variables.
   2. Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$.
   3. Express in terms of quantities with last omitted:
      $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.
   4. Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j)$.
   5. Before squaring, each summand has expectation 0.
   6. Summands are independent:
   7. Before squaring, each summand has variance $\sigma^2$.
   8. Hence $Q_n / \sigma^2$ has same distribution as sum of $n - 1$ independent squared normals: $\chi^2_{n-1}$.
   9. $n - 1$ is called the degrees of freedom.
Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i / n$.
   1. Let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$.
   2. Moments Without Assuming Independence:
   3. Moments Assuming Independence: $V(\bar{X}) = V(\sum_{i=1}^{n} X_i) (\frac{1}{n})^2 = \sigma^2 / n$, because
   4. Shape (assuming independence, normality):

2. Distribution of Sum of Squares

1. Sum of Squares from mean can be written as the sum of squared independent random variables.
2. Let $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j / n$
3. Express in terms of quantities with last omitted:
   $Q_n = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1} (X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2$.
4. Claim: $Q_n = \sum_{j=2}^{n} (X_j - \bar{X}_{j-1})^2 (1 - 1/j)$.
5. Before squaring, each summand has expectation 0.
6. Summands are independent:
7. Before squaring, each summand has variance $\sigma^2$.
8. Hence $Q_n / \sigma^2$ has same distribution as sum of $n - 1$ independent squared normals: $\chi^2_{n-1}$.
9. $n - 1$ is called the degrees of freedom.
10. Furthermore, $\bar{X}$ and $Q_n$ are independent.
1 Distribution of the studentized sample mean

Again suppose

\[ X_j \sim N(\mu, \sigma^2), \text{ independent.} \]

Then

\[ \sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0,1), \] because:

When \( \sigma \) is unknown, it is often approximated by

\[ \hat{\sigma} = \sqrt{\frac{Q_n}{n-1}}. \]

Unfortunately,

\[ T = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma} \] is not \( \sim N(0,1) \),

Distribution \( T \) of a standard normal divided by the square root of an independent \( \chi^2 \) divided by degrees of freedom is called \( t \) distribution on \( n-1 \) degrees of freedom.

\( T \) probability density function \( \propto \left(1 + \frac{t^2}{k}\right)^{-k/2 - 1/2} \), for \( k \) degrees of freedom of denominator, because:
Distribution of the studentized sample mean

Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.
Distribution of the studentized sample mean

1. Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.
2. Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:
Distribution of the studentized sample mean

1. Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.
2. Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:
3. When $\sigma$ is unknown, it is often approximated by $\hat{\sigma} = \sqrt{Q_n/(n-1)}$. 
Distribution of the studentized sample mean

1. Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.
2. Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:
3. When $\sigma$ is unknown, it is often approximated by $\hat{\sigma} = \sqrt{Q_n/(n - 1)}$.
4. Unfortunately, $T = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma}$ is not $\sim \mathcal{N}(0, 1)$,
Distribution of the studentized sample mean

1. Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.
2. Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:
3. When $\sigma$ is unknown, it is often approximated by $\hat{\sigma} = \sqrt{Q_n/(n - 1)}$.
4. Unfortunately, $T = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma}$ is not $\sim \mathcal{N}(0, 1)$,
5. Distribution $T$ of a standard normal divided by the square root of an independent $\chi^2$ divided by degrees of freedom is called $t$ distribution on $n - 1$ degrees of freedom.
Distribution of the studentized sample mean

Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.

Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:

When $\sigma$ is unknown, it is often approximated by $\hat{\sigma} = \sqrt{Q_n/(n-1)}$.

Unfortunately, $T = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma}$ is not $\sim \mathcal{N}(0, 1)$.

Distribution $T$ of a standard normal divided by the square root of an independent $\chi^2$ divided by degrees of freedom is called $t$ distribution on $n - 1$ degrees of freedom.

$T$ probability density function $\propto (1 + t^2/k)^{-k/2-1/2}$, for $k$ degrees of freedom of denominator, because:
Distribution of the ratio of sums of squares.

Suppose that $F = \frac{Q_a}{a} / \frac{Q_b}{b}$, then the distribution of variable $F$ is called the $F$-distribution with $a$ and $b$ degrees of freedom.
1 Distribution of the ratio of sums of squares.
   1 Suppose that

\[
F = \frac{Q_{a}}{Q_{b}}
\]

Then the distribution of variable \(F\) is called the \(F\)-distribution with \(a\) and \(b\) degrees of freedom.

Distribution depends on numerator, denominator df.
Distribution of the ratio of sums of squares.

Suppose that

Then the distribution of variable $F = (Q_a/a)/(Q_b/b)$ is called the $F$-distribution with $a$ and $b$ degrees of freedom.
Distribution of the ratio of sums of squares.

1. Suppose that
2. Then the distribution of variable $F = (Q_a/a)/(Q_b/b)$ is called the $F$-distribution with $a$ and $b$ degrees of freedom.
3. Distribution depends on numerator, denominator df.
Section:

Subsection: Limit Theorems
Central Limit Theorem Statement:

Remember Φ is the distribution function for a standard normal.

Let $S_n = \sum_{j=1}^{n} Y_j$, $S^*_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}}$.

Expectations and variances add.

Conclusion: $\lim_{n \to \infty} P(S^*_n \leq z) = \Phi(z)$.

Elementary Central Limit Theorem Proof

Proof technique is called the Lindeberg Bridge.

Avoids characteristic functions.

w.o.l.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$.

Let $T_n = \sum_{i=1}^{n} Z_i$ for $Z_i$ any other set of iid random variables with the same expectation, variance.

I will prove $\lim_{n \to \infty} P(S_n \leq s\sqrt{n}) = \lim_{n \to \infty} P(T_n \leq s\sqrt{n}) \forall s$.

Common limit is $\Phi(s)$, as can be seen by taking $T_j \sim N(0, 1)$.
Statements about a sample average.

Central Limit Theorem Statement:

- Conditions and Definitions: independence, equal distribution, finite variance.
Central Limit Theorem Statement:

1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember $\Phi$ is the distribution function for a standard normal.
Central Limit Theorem Statement:

1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember $\Phi$ is the distribution function for a standard normal.
3. Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$. 

Expectations and variances add.

Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$. 

Elementary Central Limit Theorem Proof

Proof technique is called the Lindeberg Bridge.

Avoids characteristic functions.

w.o.l.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$.

Let $T_n = \sum_{i=1}^{n} Z_i$ for $Z_i$ any other set of iid random variables with the same expectation, variance.

I will prove $\lim_{n \to \infty} P(S_n \leq s\sqrt{n}) = \lim_{n \to \infty} P(T_n \leq s\sqrt{n}) \forall s$.

Common limit is $\Phi(s)$, as can be seen by taking $T_j \sim N(0,1)$. 

Statements about a sample average.
Central Limit Theorem Statement:

1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember $\Phi$ is the distribution function for a standard normal.
3. Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$.
4. Expectations and variances add.

Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$. 

Elementary Central Limit Theorem Proof

1. Proof technique is called the Lindeberg Bridge.
2. Avoids characteristic functions.
3. w.o.l.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$.
4. Let $T_n = \sum_{i=1}^{n} Z_i$ for $Z_i$ any other set of iid random variables with the same expectation, variance.
5. I will prove $\lim_{n \to \infty} P(S_n^* \leq s\sqrt{n}) = \lim_{n \to \infty} P(T_n^* \leq s\sqrt{n})$ $\forall s$.
6. Common limit is $\Phi(s)$, as can be seen by taking $T_j \sim N(0,1)$. 

Limit Theorems Lecture 24
Central Limit Theorem Statement:

1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember $\Phi$ is the distribution function for a standard normal.
3. Let $S_n = \sum_{j=1}^{n} Y_j$, $S^*_n = (S_n - E(S_n))/\sqrt{V(S_n)}$.
4. Expectations and variances add.
5. Conclusion: $\lim_{n \to \infty} P(S^*_n \leq z) = \Phi(z)$.
Statements about a sample average.

1 Central Limit Theorem Statement:
   1 Conditions and Definitions: independence, equal distribution, finite variance.
   2 Remember $\Phi$ is the distribution function for a standard normal.
   3 Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - \text{E}(S_n))/\sqrt{\text{V}(S_n)}$.
   4 Expectations and variances add.
   5 Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$.

2 Elementary Central Limit Theorem Proof
Statements about a sample average.

1. Central Limit Theorem Statement:
   1. Conditions and Definitions: independence, equal distribution, finite variance.
   2. Remember $\Phi$ is the distribution function for a standard normal.
   3. Let $S_n = \sum_{j=1}^n Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$.
   4. Expectations and variances add.
   5. Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$.

2. Elementary Central Limit Theorem Proof
   1. Proof technique is called the Lindeberg Bridge.
Statements about a sample average.

1 Central Limit Theorem Statement:
   1 Conditions and Definitions: independence, equal distribution, finite variance.
   2 Remember Φ is the distribution function for a standard normal.
   3 Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$.
   4 Expectations and variances add.
   5 Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$.

2 Elementary Central Limit Theorem Proof
   1 Proof technique is called the Lindeberg Bridge.
   2 Avoids characteristic functions.
Statements about a sample average.

1 Central Limit Theorem Statement:
   1 Conditions and Definitions: independence, equal distribution, finite variance.
   2 Remember $\Phi$ is the distribution function for a standard normal.
   3 Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$.
   4 Expectations and variances add.
   5 Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$.

2 Elementary Central Limit Theorem Proof
   1 Proof technique is called the Lindeberg Bridge.
   2 Avoids characteristic functions.
   3 w.o.l.o.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$. 
Statements about a sample average.

Central Limit Theorem Statement:
1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember $\Phi$ is the distribution function for a standard normal.
3. Let $S_n = \sum_{j=1}^{n} Y_j$, $S^*_n = (S_n - E(S_n))/\sqrt{V(S_n)}$.
4. Expectations and variances add.
5. Conclusion: $\lim_{n \to \infty} P(S^*_n \leq z) = \Phi(z)$.

Elementary Central Limit Theorem Proof
1. Proof technique is called the Lindeberg Bridge.
2. Avoids characteristic functions.
3. w.o.l.o.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$.
4. Let $T_n = \sum_{i=1}^{n} Z_i$ for $Z_i$ any other set of iid random variables with the same expectation, variance.
Central Limit Theorem Statement:
1. Conditions and Definitions: independence, equal distribution, finite variance.
2. Remember \( \Phi \) is the distribution function for a standard normal.
3. Let \( S_n = \sum_{j=1}^{n} Y_j \), \( S_n^* = (S_n - \text{E}(S_n))/\sqrt{\text{V}(S_n)} \).
4. Expectations and variances add.
5. Conclusion: \( \lim_{n \to \infty} P( S_n^* \leq z ) = \Phi(z) \).

Elementary Central Limit Theorem Proof
1. Proof technique is called the Lindeberg Bridge.
2. Avoids characteristic functions.
3. w.o.l.o.g. take \( \text{E}(Y_j) = 0 \) and \( \text{V}(Y_j) = 1 \).
4. Let \( T_n = \sum_{i=1}^{n} Z_i \) for \( Z_i \) any other set of iid random variables with the same expectation, variance.
5. I will prove \( \lim_{n \to \infty} P( S_n \leq s \sqrt{n} ) = \lim_{n \to \infty} P( T_n \leq s \sqrt{n} ) \forall s \).
1 Central Limit Theorem Statement:
   1 Conditions and Definitions: independence, equal distribution, finite variance.
   2 Remember $\Phi$ is the distribution function for a standard normal.
   3 Let $S_n = \sum_{j=1}^{n} Y_j$, $S_n^* = (S_n - E(S_n))/\sqrt{V(S_n)}$.
   4 Expectations and variances add.
   5 Conclusion: $\lim_{n \to \infty} P(S_n^* \leq z) = \Phi(z)$.

2 Elementary Central Limit Theorem Proof
   1 Proof technique is called the Lindeberg Bridge.
   2 Avoids characteristic functions.
   3 w.o.l.o.g. take $E(Y_j) = 0$ and $V(Y_j) = 1$.
   4 Let $T_n = \sum_{i=1}^{n} Z_i$ for $Z_i$ any other set of iid random variables with the same expectation, variance.
   5 I will prove $\lim_{n \to \infty} P(S_n \leq s\sqrt{n}) = \lim_{n \to \infty} P(T_n \leq s\sqrt{n}) \forall s$.
   6 Common limit is $\Phi(s)$, as can be seen by taking $T_j \sim \mathcal{N}(0, 1)$. 
Limits of this proof:

1. Only for identically distributed case
2. Only for independent case
3. Moving between $Y_i$ to $Z_i$ one variable at a time.
4. After adding a small random variable $V$

Define distribution function with summands swapped sequentially:

Some facts about absolute values and maximum values:
Limits of this proof:

1. Only for identically distributed case
1. Limits of this proof:
   1. Only for identically distributed case
   2. Only for independent case
1. Limits of this proof:
   1. Only for identically distributed case
   2. Only for independent case
2. Moving between $Y_i$ to $Z_i$ one variable at a time.
1 Limits of this proof:
   1 Only for identically distributed case
   2 Only for independent case

2 Moving between $Y_i$ to $Z_i$ one variable at a time.
   1 After adding a small random variable $V$
Limits of this proof:
1. Only for identically distributed case
2. Only for independent case

Moving between $Y_i$ to $Z_i$ one variable at a time.
1. After adding a small random variable $V$
2. Express distribution function of sum as the expected value of a conditional distribution:
1 **Limits of this proof:**
   1. Only for identically distributed case
   2. Only for independent case

2 **Moving between \( Y_i \) to \( Z_i \) one variable at a time.**
   1. After adding a small random variable \( V \)
   2. Express distribution function of sum as the expected value of a conditional distribution:
   3. Define distribution function with summands swapped sequentially:
Limits of this proof:
1. Only for identically distributed case
2. Only for independent case

Moving between $Y_i$ to $Z_i$ one variable at a time.
1. After adding a small random variable $V$
2. Express distribution function of sum as the expected value of a conditional distribution:
3. Define distribution function with summands swapped sequentially:
4. Some facts about absolute values and maximum values:
Bounding the largest CDF difference

The bound on each difference is the same Taylor series:

\[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F'_V(s)(-Z_j/\sqrt{n}) + F''_V(s) Z_j^2/n \] for some \( s_z \), and similarly for \( Y_j \).

The factor \( n \) multiplying \( E(n|P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n})|c) \) cancels with \( n \) in denominator, and result goes to zero as \( n \to \infty \), because Central Limit Theorem Proof via Moment Generating Function

More enlightening, but with more swept under the rug.

Suppose that w.o.l.g. \( E(Y_j) = 0 \) and \( V(Y_j) = 1 \).

Suppose that \( Y_j \) have a common mgf \( m_1(t) \).

The mgf of the sum \( S_n \) is \( m_1(t)n \).

The mgf of the standardized sum \( S^*_n = S_n/\sqrt{n} \) is \( m_1(t/\sqrt{n})n \).

Then \( \lim_{n \to \infty} m_n(t) = \exp(t^2/2) \), because...
Bounding the largest CDF difference

First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:

The bound on each difference is the same Taylor series:

$$F_V(s - Z_j / \sqrt{n}) = F_V(s) + F_V'(s)(-Z_j / \sqrt{n}) + F_V''(s)(Z_j^2 / n)$$ for some $s_z$, and similarly for $Y_j$.

The factor $n$ multiplying $E(n | P(Y_j + V \sqrt{n} \geq s \sqrt{n}) - P(Z_j + V \sqrt{n} \geq s \sqrt{n})$ cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because Central Limit Theorem Proof via Moment Generating Function

More enlightening, but with more swept under the rug.

Suppose that w.o.l.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.

Suppose that $Y_j$ have a common mgf $m_1(t)$.

The mgf of the sum $S_n$ is $m_1(t)n$.

The mgf of the standardized sum $S^*_n = S_n / \sqrt{n}$ is $m_1(t/\sqrt{n})n$.

Then $\lim_{n \to \infty} m_n(t) = \exp(t^2/2)$, because Limit Theorems Lecture 24 36/42
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2. The bound on each difference is the same
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2. The bound on each difference is the same
3. Taylor series:
   \[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F_V'(s)(-Z_j/\sqrt{n}) + F_V''(s_z)(Z_j^2/n) \]
   for some $s_z$, and similarly for $Y_j$. 

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
2. Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.
3. Suppose that $Y_j$ have a common mgf $m_1(t)$.
4. The mgf of the sum $S_n$ is $m_1(t)^n$.
5. The mgf of the standardized sum $S^*_n = S_n / \sqrt{n}$ is $m_1(t/\sqrt{n})^n$.
6. Then \( \lim_{n \to \infty} m_n(t) = \exp(t^2/2) \), because...
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2. The bound on each difference is the same
3. Taylor series:
   \[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F'_V(s)(-Z_j/\sqrt{n}) + F''_V(s_z)(Z_j^2/n) \]
   for some $s_z$, and similarly for $Y_j$.
4. The factor $n$ multiplying
   \[ E(n|P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n}) \mid) \]
   cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
   - The bound on each difference is the same
   - Taylor series:
     \[
     F_V(s - Z_j / \sqrt{n}) = F_V(s) + F'_V(s)(-Z_j / \sqrt{n}) + F''_V(s_z)(Z_j^2 / n) \text{ for some } s_z, \text{ and similarly for } Y_j.
     \]
   - The factor $n$ multiplying
     \[
     E(n | P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n}) |) \tag{1}
     \]
     cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
   2. The bound on each difference is the same
   3. Taylor series:
      \[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F'_V(s)(-Z_j/\sqrt{n}) + F''_V(s_z)(Z_j^2/n) \text{ for some } s_z, \text{ and similarly for } Y_j. \]
   4. The factor $n$ multiplying
      \[ E(n|P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n}) |) \text{ cancels with } n \]
      in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2. The bound on each difference is the same
3. Taylor series:
   \[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F'_V(s)(-Z_j/\sqrt{n}) + F''_V(s_z)(Z_j^2/n) \] for some $s_z$, and similarly for $Y_j$.
4. The factor $n$ multiplying
   \[ E(n|P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n}) |) \] cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
2. Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$. 
Bounding the largest CDF difference

1 First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2 The bound on each difference is the same
3 Taylor series:
   \[ F_V(s - Z_j / \sqrt{n}) = F_V(s) + F'_V(s)(-Z_j / \sqrt{n}) + F''_V(s_z)(Z_j^2 / n) \]
   for some $s_z$, and similarly for $Y_j$.
4 The factor $n$ multiplying
   \[ E \left( n \mid P(Y_j + V \sqrt{n} \geq s \sqrt{n}) - P(Z_j + V \sqrt{n} \geq s \sqrt{n}) \right) \]
   cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1 More enlightening, but with more swept under the rug.
2 Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.
3 Suppose that $Y_j$ have a common mgf $m_1(t)$. 
Bounding the largest CDF difference

First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:

1. The bound on each difference is the same
2. Taylor series:
   \[ F_V(s - Z_j / \sqrt{n}) = F_V(s) + F'_V(s)(-Z_j / \sqrt{n}) + F''_V(s_z)(Z_j^2 / n) \text{ for some } s_z, \text{ and similarly for } Y_j. \]
3. The factor $n$ multiplying
   \[ E(n|P(Y_j + V \sqrt{n} \geq s \sqrt{n}) - P(Z_j + V \sqrt{n} \geq s \sqrt{n})|) \]
   cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
2. Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.
3. Suppose that $Y_j$ have a common mgf $m_1(t)$.
4. The mgf of the sum $S_n$ is $m_1(t)^n$. 
Bounding the largest CDF difference

1. First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:
2. The bound on each difference is the same
3. Taylor series:
   \[ F_V(s - Z_j/\sqrt{n}) = F_V(s) + F'_V(s)(-Z_j/\sqrt{n}) + F''_V(s_z)(Z_j^2/n) \]
   for some $s_z$, and similarly for $Y_j$.
4. The factor $n$ multiplying
   \[ E(n|P(Y_j + V\sqrt{n} \geq s\sqrt{n}) - P(Z_j + V\sqrt{n} \geq s\sqrt{n})|) \]
   cancels with $n$ in denominator, and result goes to zero as $n \to \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
2. Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.
3. Suppose that $Y_j$ have a common mgf $m_1(t)$.
4. The mgf of the sum $S_n$ is $m_1(t)^n$.
5. The mgf of the standardized sum $S^*_n = S_n/\sqrt{n}$ is $m_1(t/\sqrt{n})^n$. 
Bounding the largest CDF difference

First and last results of sequential swapping are bounded by $n$ times largest CDF difference, because:

1. The bound on each difference is the same
2. Taylor series:
   \[ F_V(s - Z_j / \sqrt{n}) = F_V(s) + F_V'(s)(-Z_j / \sqrt{n}) + F_V''(s_z)(Z_j^2 / n) \]
   for some $s_z$, and similarly for $Y_j$.
3. The factor $n$ multiplying $E(n | P(Y_j + V \sqrt{n} \geq s \sqrt{n}) - P(Z_j + V \sqrt{n} \geq s \sqrt{n}) |)$ cancels with $n$ in denominator, and result goes to zero as $n \rightarrow \infty$, because

Central Limit Theorem Proof via Moment Generating Function

1. More enlightening, but with more swept under the rug.
2. Suppose that w.o.l.o.g. $E(Y_j) = 0$ and $V(Y_j) = 1$.
3. Suppose that $Y_j$ have a common mgf $m_1(t)$.
4. The mgf of the sum $S_n$ is $m_1(t)^n$.
5. The mgf of the standardized sum $S_n^* = S_n / \sqrt{n}$ is $m_1(t/\sqrt{n})^n$.
6. Then $\lim_{n \rightarrow \infty} m_n(t) = \exp(t^2/2)$, because
Completing Proof Requires Concepts Beyond Course
Completing Proof Requires Concepts Beyond Course

- Remember that you trusted me: mgf uniquely defines distribution.
Completing Proof Requires Concepts Beyond Course

1. Remember that you trusted me: mgf uniquely defines distribution.
2. Beyond that, mgfs close together implies probability density functions or probability functions close together.

\[ Y_1 \sim \text{Unif}\left(-\frac{1}{2}, \frac{1}{2}\right), \quad Y_k = Y_1 \text{ for all } k. \]

\[ Y_k \sim \text{Unif}\left(-\frac{1}{2}, \frac{1}{2}\right) / k, \text{ independent, have a distribution that gets more concentrated as } k \text{ increases.} \]

\[ Y_k \sim T(2) \text{ do not have variance, and the distribution of averages does not get more concentrated.} \]
Completing Proof Requires Concepts Beyond Course

1. Remember that you trusted me: mgf uniquely defines distribution.
2. Beyond that, mgfs close together implies probability density functions or probability functions close together.

Each of the assumptions is necessary:
Completing Proof Requires Concepts Beyond Course

1. Remember that you trusted me: mgf uniquely defines distribution.
2. Beyond that, mgfs close together implies probability density functions or probability functions close together.

Each of the assumptions is necessary:

1. $Y_1 \sim \text{Unif}(-1/2, 1/2)$, $Y_k = Y_1$ for all $k$. 

2. $Y_k \sim \text{T}(2)$ do not have variance, and the distribution of averages does not get more concentrated.
Completing Proof Requires Concepts Beyond Course

1. Remember that you trusted me: mgf uniquely defines distribution.
2. Beyond that, mgfs close together implies probability density functions or probability functions close together.

Each of the assumptions is necessary:

1. $Y_1 \sim \text{Unif}(-1/2, 1/2)$, $Y_k = Y_1$ for all $k$.
2. $Y_k \sim \text{Unif}(-1/2, 1/2)/k$, independent, have a distribution that gets more concentrated as $k$ increases.
Completing Proof Requires Concepts Beyond Course

Remember that you trusted me: mgf uniquely defines distribution.

Beyond that, mgfs close together implies probability density functions or probability functions close together.

Each of the assumptions is necessary:

1. $Y_1 \sim \text{Unif}(-1/2,1/2)$, $Y_k = Y_1$ for all $k$.
2. $Y_k \sim \text{Unif}(-1/2,1/2)/k$, independent, have a distribution that gets more concentrated as $k$ increases.
3. $Y_k \sim \mathcal{T}(2)$ do not have variance, and the distribution of averages does not get more concentrated.
When will the Central Limit Theorem be Accurate?

Suppose \( Y_j \) iid expectation zero, variance 1. The most important part of the distribution that the CLT ignores is \( E(S^3) \).

Target normal distribution has \( E(Y^3) = 0 \).

Let \( S_n = \sum_{j=1}^{n} Y_j \), \( S^*_n = \frac{\sum_{j=1}^{n} Y_j}{\sqrt{n}} \).

So \( E(S^*_3) = \frac{n E(Y^3)}{\sqrt{n^3}} = \frac{E(Y^3)}{\sqrt{n}} \).
When will the Central Limit Theorem be Accurate?

Suppose \( Y_j \) iid expectation zero, variance 1.
When will the Central Limit Theorem be Accurate?

1. Suppose $Y_j$ iid expectation zero, variance 1.
2. The most important part of the distribution that the CLT ignores is skewness $E(S^*^3)$.
When will the Central Limit Theorem be Accurate?

1. Suppose $Y_j$ iid expectation zero, variance 1.
2. The most important part of the distribution that the CLT ignores is skewness $E(S^*^3)$.
3. Target normal distribution has $E(Y_1^3) = 0$. 

Let $S_n = \sum_{j=1}^{n} Y_j$, $S^*_n = \sum_{j=1}^{n} Y_j / \sqrt{n}$. 

So $E(S^*^3_n) = n E(Y_1^3) / n^{3/2} = E(Y_1^3) / \sqrt{n}$. 
When will the Central Limit Theorem be Accurate?

1. Suppose $Y_j$ iid expectation zero, variance 1.
2. The most important part of the distribution that the CLT ignores is skewness $E(S^*_3)$.
3. Target normal distribution has $E(Y^3_1) = 0$.
4. Let $S_n = \sum_{j=1}^n Y_j$, $S^*_n = \sum_{j=1}^n Y_j / \sqrt{n}$. 


When will the Central Limit Theorem be Accurate?

1. Suppose $Y_j$ iid expectation zero, variance 1.
2. The most important part of the distribution that the CLT ignores is skewness $E(S^*_3)$.
3. Target normal distribution has $E(Y_1^3) = 0$.
4. Let $S_n = \sum_{j=1}^n Y_j$, $S_n^* = \sum_{j=1}^n Y_j/\sqrt{n}$.
5. So $E(S_n^3) = nE(Y_1^3)/n^{3/2} = E(Y_1^3)/\sqrt{n}$. 
A Further Refinement to the Central Limit Theorem

- Represent discrete distribution as a bar plot
- Adjust area under normal curve to match area of probability to be calculated.

Example:

\[ Y \sim \text{Bin}(m, \pi), \quad m = 20, \quad \pi = \cdots \]
A Further Refinement to the Central Limit Theorem

1. Represent discrete distribution as a bar plot

Example:
\[ Y \sim \text{Bin}(m, \pi), \quad m = 20, \quad \pi = \ldots \]
A Further Refinement to the Central Limit Theorem

1. Represent discrete distribution as a bar plot
2. Adjust area under normal curve to match area of probability to be calculated.

Example:

\[ Y \sim \text{Bin}(m, \pi), \quad m = 20, \quad \pi = \cdots \]
A Further Refinement to the Central Limit Theorem

1. Represent discrete distribution as a bar plot
2. Adjust area under normal curve to match area of probability to be calculated.
3. Example: \( Y_m \sim \text{Bin}(m, \pi), \ m = 20, \ \pi = .4. \)
Continuity correction is less important for large samples
Continuity correction is less important for large samples

Let $\Delta$ represent the space between values of $Y_j$. 

$\frac{\Delta}{\sqrt{n} \sigma}$

Goes to zero as $n \to \infty$

In respect to power of $n$, effect size is comparable to that of skewness.
Continuity correction is less important for large samples

1. Let $\Delta$ represent the space between values of $Y_j$.
2. Let $\sigma$ represent standard deviation of $Y_j$. 

Effect on argument to $\Phi$ is $\Delta / (2 \sqrt{n} \sigma)$.

Goes to zero as $n \to \infty$.

In respect to power of $n$, effect size is comparable to that of skewness.
Continuity correction is less important for large samples

1. Let $\Delta$ represent the space between values of $Y_j$.
2. Let $\sigma$ represent standard deviation of $Y_j$.
3. Effect on argument to $\Phi$ is $\Delta/(2\sqrt{n}\sigma)$.
Continuity correction is less important for large samples

1. Let $\Delta$ represent the space between values of $Y_j$.
2. Let $\sigma$ represent standard deviation of $Y_j$.
3. Effect on argument to $\Phi$ is $\Delta/(2\sqrt{n}\sigma)$.
4. Goes to zero as $n \to \infty$
Continuity correction is less important for large samples

1. Let $\Delta$ represent the space between values of $Y_j$.
2. Let $\sigma$ represent standard deviation of $Y_j$.
3. Effect on argument to $\Phi$ is $\Delta/(2\sqrt{n}\sigma)$.
4. Goes to zero as $n \to \infty$
5. In respect to power of $n$, effect size is comparable to that of skewness.
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation

\[
\text{Suppose that } \bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i.
\]

Then \( E(\bar{Y}_n) = \mu \), \( V(\bar{Y}_n) = \sigma^2 / n \).

Take \( \epsilon > 0 \).

Tchebysheff's Inequality:

\[
P( |\bar{Y}_n - \mu| > \epsilon ) \leq \frac{\sigma^2}{n \epsilon^2}.
\]

Central limit argument:

\[
P( |\bar{Y}_n - \mu| > \epsilon ) = P( |\bar{Y}_n - \mu| / \sigma / \sqrt{n} > \sqrt{n} \epsilon / \sigma ) \approx 2 \Phi(\sqrt{n} \epsilon / \sigma).
\]

CLT bound is smaller than Tchebysheff bound.
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation

   Suppose that

   \[ \bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i, \]

   then

   \[ E(\bar{Y}_n) = \mu, \quad V(\bar{Y}_n) = \frac{\sigma^2}{n}. \]

Take \( \epsilon > 0 \).

Tchebysheff's Inequality:

\[ P(|\bar{Y}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n \epsilon^2}. \]

Central limit argument:

\[ P(|\bar{Y}_n - \mu| > \epsilon) = P \left( \left| \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \right| > \frac{\sqrt{n} \epsilon}{\sigma} \right) \approx 2 \Phi \left( \frac{\sqrt{n} \epsilon}{\sigma} \right). \]

CLT bound is smaller than Tchebysheff bound.
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation
   1. Suppose that
   2. Then $E(\bar{Y}_n) = \mu$, $V(\bar{Y}_n) = \sigma^2/n$
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation
   - Suppose that
   - Then $E(\bar{Y}_n) = \mu$, $V(\bar{Y}_n) = \sigma^2/n$
   - Take $\epsilon > 0$.

   Tchebysheff's Inequality: $P(|\bar{Y}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$.

   Central limit argument: $P(|\bar{Y}_n - \mu| > \epsilon) = P\left(|\bar{Y}_n - \mu| > \frac{\epsilon}{\sigma/\sqrt{n}}\right) \approx 2\Phi\left(\frac{\sqrt{n}\epsilon}{\sigma}\right)$.

   CLT bound is smaller than Tchebysheff bound.
Laws of Large Numbers

1. **Probability Bounds on the Distance between the Average and Expectation**

1. Suppose that
2. Then $E(\bar{Y}_n) = \mu$, $V(\bar{Y}_n) = \sigma^2/n$
3. Take $\epsilon > 0$.
4. Tchebyshhev’s Inequality: $P(|\bar{Y}_n - \mu| > \epsilon) \leq \sigma^2/(n\epsilon^2)$. 
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation
   1. Suppose that
   2. Then \( E(\bar{Y}_n) = \mu, \ V(\bar{Y}_n) = \sigma^2/n \)
   3. Take \( \epsilon > 0 \).
   4. Tchebycheff’s Inequality: \( P(|\bar{Y}_n - \mu| > \epsilon) \leq \sigma^2/(n\epsilon^2) \).
   5. Central limit argument:
      \[
      P(|\bar{Y}_n - \mu| > \epsilon) = P \left( \frac{|\bar{Y}_n - \mu|}{\sigma/\sqrt{n}} > \sqrt{n\epsilon}/\sigma \right) \approx 2\Phi(\sqrt{n\epsilon}/\sigma)
      \]
Laws of Large Numbers

1. Probability Bounds on the Distance between the Average and Expectation

1. Suppose that
2. Then \( E (\bar{Y}_n) = \mu, \, V (\bar{Y}_n) = \sigma^2 / n \)
3. Take \( \varepsilon > 0 \).
4. Tchebycheff’s Inequality: \( P \left( | \bar{Y}_n - \mu | > \varepsilon \right) \leq \frac{\sigma^2}{n \varepsilon^2} \).
5. Central limit argument:
   \( P \left( | \bar{Y}_n - \mu | > \varepsilon \right) = P \left( \frac{| \bar{Y}_n - \mu |}{\sigma/\sqrt{n}} > \sqrt{n\varepsilon/\sigma} \right) \approx 2\Phi \left( \sqrt{n\varepsilon/\sigma} \right) \)
6. CLT bound is smaller than Tchebycheff bound
Statement of Weak Law of Large Numbers

\[ \text{Select independent and identically distributed rvs, with finite expectation and } \epsilon > 0, \]
\[ \text{Then } \lim_{n \to \infty} P(\left| \bar{Y}_n - \mu \right| > \epsilon) = 0: \]

Weak Law of Large Numbers

LLN needs finite expectation

LLN does not need finite variance

Weak and Strong Laws

Preceding Law of Large Numbers refers treats each member of the sequence by itself.

Another version refers to the interplay between the sample points and the conversion parameter \( n \): Limit Theorems Lecture 25
Statement of Weak Law of Large Numbers

Select independent and identically distributed rvs, with finite expectation and $\epsilon > 0$, then

$$\lim_{n \to \infty} P\left(\left| \bar{Y}_n - \mu \right| > \epsilon \right) = 0.$$
Statement of Weak Law of Large Numbers

1. Select independent and identically distributed rvs, with finite expectation and $\epsilon > 0$,
2. Then $\lim_{n \to \infty} P \left( |\bar{Y}_n - \mu| > \epsilon \right) = 0$: Weak Law of Large Numbers
Statement of Weak Law of Large Numbers

1. Select independent and identically distributed rvs, with finite expectation and \( \epsilon > 0 \),
2. Then \( \lim_{n \to \infty} P \left( |\bar{Y}_n - \mu| > \epsilon \right) = 0 \): Weak Law of Large Numbers
3. LLN needs finite expectation
Statement of Weak Law of Large Numbers

1. Select independent and identically distributed rvs, with finite expectation and \( \epsilon > 0 \),
2. Then \( \lim_{n \to \infty} P(\left| \bar{Y}_n - \mu \right| > \epsilon) = 0 \): Weak Law of Large Numbers
3. LLN needs finite expectation
4. LLN does not need finite variance
Statement of Weak Law of Large Numbers

1. Select independent and identically distributed rvs, with finite expectation and $\epsilon > 0$,
2. Then $\lim_{n \to \infty} P \left( \left| \bar{Y}_n - \mu \right| > \epsilon \right) = 0$: Weak Law of Large Numbers

LLN needs finite expectation
LLN does not need finite variance

Weak and Strong Laws
1 **Statement of Weak Law of Large Numbers**
   1. Select independent and identically distributed rvs, with finite expectation and $\epsilon > 0$,
   2. Then $\lim_{n \to \infty} P \left( |\bar{Y}_n - \mu| > \epsilon \right) = 0$: *Weak Law of Large Numbers*

2 **Weak and Strong Laws**
   1. Preceding Law of Large Numbers refers to treats each member of the sequence by itself.
1. Statement of Weak Law of Large Numbers
   1. Select independent and identically distributed rvs, with finite expectation and $\epsilon > 0$,
   2. Then $\lim_{n \to \infty} P \left( |\bar{Y}_n - \mu| > \epsilon \right) = 0$: Weak Law of Large Numbers

2. LLN needs finite expectation
3. LLN does not need finite variance

2. Weak and Strong Laws
   1. Preceding Law of Large Numbers refers treats each member of the sequence by itself.
   2. Another version refers to the interplay between the sample points and the conversion parameter $n$