\[ x = \{ 2, 3, 3, 4, 4, 4, 5, 5, 6 \} \]
\[
\bar{x} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}
\]
An Alternative to Calculate the Mean?
An Alternative to Calculate the Mean?

\[ X = \left\{ 2, 3, 3, 4, 4, 4, 5, 5, 6 \right\} \]

\[ \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|} \]
An Alternative to Calculate the Mean?

\[ X = \left\{ 2, 3, 3, 4, 4, 4, 5, 5, 6 \right\} \]

\[ \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|} \]

\[ 2 \times \frac{1}{9} + 3 \times \frac{2}{9} + 4 \times \frac{3}{9} + 5 \times \frac{2}{9} + 6 \times \frac{1}{9} \]
An Alternative to Calculate the Mean?

\[ X = \left\{ 2, 3, 3, 4, 4, 4, 5, 5, 6 \right\} \]

\[ \frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|} \]

\[
2 \times \frac{1}{9} + 3 \times \frac{2}{9} + 4 \times \frac{3}{9} + 5 \times \frac{2}{9} + 6 \times \frac{1}{9}
\]

\[ = 2 \times \frac{|x = 2|}{|X|} + 3 \times \frac{|x = 3|}{|X|} + \ldots \]
An Alternative to Calculate the Mean?

\[
x = \left\{ 2, 3, 3, 4, 4, 4, 5, 5, 6 \right\}
\]

\[
\frac{\text{SUM}}{\text{COUNT}} = \frac{\sum_{i \in \text{range}(X)} x_i}{|X|}
\]

\[
2 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} + 4 \cdot \frac{3}{9} + 5 \cdot \frac{2}{9} + 6 \cdot \frac{1}{9}
\]

\[
= 2 \cdot \frac{|x = 2|}{|X|} + 3 \cdot \frac{|x = 3|}{|X|} + \ldots
\]

\[
= \sum_{v \in X} v \cdot \frac{|x = v|}{|X|}
\]
Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?
Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The *expected value* of $X$ is:

$$E(X) = \int x \, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) \, dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $E(X) = EX = (x) = \mu = \mu x$
Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The *expected value* of $X$ is:

$$E(X) = \int x \, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) \, dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $E(X) = E_X = (x) = \mu = \mu x$

“expectation”  “mean”  “first moment”
Expectation

Conceptually: Approximately: Just given the distribution and no other information: what value should I expect?

Formally: The expected value of $X$ is:

$$E(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $E(X) = EX = (x) = \mu = \mu x$

“expectation” “mean” “first moment”

Alternative Conceptualization: If I had to summarize a distribution with only one number, what would do that best? (the average of a large number of randomly generated numbers from the distribution)
Expectation

the PDF, \( f(x) \), is the derivative of the CDF, \( F(x) \).

Conceptually: Approximate the area under the distribution curve with no other information: what value should I expect?

Formally: The expected value of \( X \) is:

\[
E(X) = \int x \, dF(x) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
\]

denoted:

\[
E(X) = \mathbb{E}X = (x) = \mu = \mu x
\]

"expectation"  "mean"  "first moment"

Alternative Conceptualization: If I had to summarize a distribution with only one number, what would do that best?
(the average of a large number of randomly generated numbers from the distribution)
Expectation

Examples:

$X \sim \text{Bernoulli}(p)$:

$X \sim \text{Uniform}(-3,1)$:

The **expected value** of $X$ is:

$$E(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted:

$$E(X) = EX = (x) = \mu = \mu x$$
Expectation

Examples:

X ~ Bernoulli(p):

X ~ Uniform(-3,1):

The expected value of X is:

\[ E(X) = \int x \, dF(x) = \begin{cases} \sum_x x \, f(x) & \text{if } X \text{ is discrete} \\ \int x \, f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \]

denoted: \[ E(X) = EX = (x) = \mu = \mu x \]
Expectation

Examples:

$X \sim \text{Bernoulli}(p)$:

$X \sim \text{Uniform}(-3,1)$:

The expected value of $X$ is:

$$
E(X) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
$$

denoted: $E(X) = EX = (x) = \mu = \mu x$
Expectation

Examples:

\[ X \sim \text{Bernoulli}(p) : \]

\[ X \sim \text{Uniform}(-3,1) : \]

The expected value of \( X \) is:

\[
\mathbb{E}(X) = \int x \, df(x) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
\]

denoted: \( \mathbb{E}(X) = \text{EX} = (x) = \mu = \mu x \)
Expectation

Examples:

$X \sim \text{Bernoulli}(p)$:

$X \sim \text{Uniform}(-3,1)$:

\[
f(x) = \binom{n}{x} p^x (1 - p)^{n-x} = \binom{1}{x} p^x (1 - p)^{1-x}
= p^x (1 - p)^{1-x}
\]

\[
\sum_{x \in \{0,1\}} x f(x) = 0(p^0 (1 - p)^{(1-0)}) + 1(p^1 (1 - p)^{(1-1)})
= 1(p^1 (1)) = p
\]

The expected value of $X$ is:

\[
\mathbb{E}(X) = \int x dF(x) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
\]

denoted: $\mathbb{E}(X) = \mathbb{E}X = (x) = \mu = \mu x$
Expectation

Examples:

\( X \sim \text{Bernoulli}(p) \):

\( X \sim \text{Uniform}(-3,1) \):

\[
\begin{align*}
(\text{practice}) \\
\frac{a+b}{2}
\end{align*}
\]

The expected value of \( X \) is:

\[
\mathbb{E}(X) = \int x \, dF(x) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
\]

denoted: \( \mathbb{E}(X) = \mathbb{E}X = (x) = \mu = \mu_x \)
Variance, Second Moment

**Conceptually:** The expected difference from the mean.

The **variance** of $X$ is:

\[ V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 dF(x) \]
Conceptually: The expected difference from the mean.

\[ V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 \, dF(x) \]

\[ \sum_{x \in \{0,1\}} p^x (1-p)^{1-x} \]

**Variance, Second Moment**

\[ \int x \, dF(x) = \begin{cases} \sum x \, f(x) & \text{if } X \text{ is discrete} \\ \int x \, f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \]

\( X \sim \text{Bernoulli}(p) \)
Variance, Second Moment

\[
\int x dF(x) = \begin{cases} 
\sum_x x f(x) & \text{if } X \text{ is discrete} \\
\int x f(x) dx & \text{if } X \text{ is continuous}
\end{cases}
\]

Conceptually: The expected difference from the mean.

The \textbf{variance} of \( X \) is:

\[
V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 dF(x)
\]

\( X \sim \text{Bernoulli}(p): \)

\[
\sum_{x \in \{0,1\}} (x - \mu)^2 (p^x (1-p)^{1-x}) = (0 - \mu)^2 (p^0 (1-p)^{1-0}) + (1 - \mu)^2 (p^1 (1-p)^{1-1})
\]
Variance, Second Moment

\[ \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases} \]

Conceptually: The expected difference from the mean.

The **variance** of \( X \) is:

\[ V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 dF(x) \]

\( X \sim \text{Bernoulli}(p) \):

\[ p^x (1 - p)^{1-x} \]

\[ \sum_{x \in \{0,1\}} (x - \mu)^2 (p^x (1 - p)^{1-x}) = (0 - \mu)^2 (p^0 (1 - p)^{1-0}) + (1 - \mu)^2 (p^1 (1 - p)^{1-1}) \]

\[ = \mu^2 (1 - p) + (1 - \mu)^2 (p) = p^2 (1 - p) + (1 - p)^2 (p) \]
Variance, Second Moment

\[ \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \]

Conceptually: The expected difference from the mean.

The \textbf{variance} of X is:

\[ V(X) = \sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 \, dF(x) \]

\( X \sim \text{Bernoulli}(p) \):

\[ p^x (1-p)^{1-x} \]

\[ \sum_{x \in \{0, 1\}} (x - \mu)^2 (p^x (1-p)^{1-x}) = (0 - \mu)^2 (p^0 (1-p)^{1-0}) + (1 - \mu)^2 (p^1 (1-p)^{1-1}) = \mu^2 (1-p) + (1 - \mu)^2 (p) = p^2 (1-p) + (1 - p)^2 (p) = p(1 - p)(p + 1 - p) = p(1 - p) \]
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point mass at $a$</td>
<td>$a$</td>
<td>0</td>
</tr>
<tr>
<td>Bernoulli($p$)</td>
<td>$p$</td>
<td>$p(1 - p)$</td>
</tr>
<tr>
<td>Binomial($n, p$)</td>
<td>$np$</td>
<td>$np(1 - p)$</td>
</tr>
<tr>
<td>Geometric($p$)</td>
<td>$1/p$</td>
<td>$(1 - p)/p^2$</td>
</tr>
<tr>
<td>Poisson($\lambda$)</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>Uniform($a, b$)</td>
<td>$(a + b)/2$</td>
<td>$(b - a)^2/12$</td>
</tr>
<tr>
<td>Normal($\mu, \sigma^2$)</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Exponential($\beta$)</td>
<td>$\beta$</td>
<td>$\beta^2$</td>
</tr>
<tr>
<td>Gamma($\alpha, \beta$)</td>
<td>$\alpha\beta$</td>
<td>$\alpha\beta^2$</td>
</tr>
<tr>
<td>Beta($\alpha, \beta$)</td>
<td>$\alpha/(\alpha + \beta)$</td>
<td>$\alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$</td>
</tr>
<tr>
<td>$t_\nu$</td>
<td>0 (if $\nu &gt; 1$)</td>
<td>$\nu/(\nu - 2)$ (if $\nu &gt; 2$)</td>
</tr>
<tr>
<td>$\chi^2_p$</td>
<td>$p$</td>
<td>$2p$</td>
</tr>
<tr>
<td>Multinomial($n, p$)</td>
<td>$np$</td>
<td>see below</td>
</tr>
<tr>
<td>Multivariate Normal($\mu, \Sigma$)</td>
<td>$\mu$</td>
<td>$\Sigma$</td>
</tr>
</tbody>
</table>

(Wasserman, 2003)
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point mass at $a$</td>
<td>$a$</td>
<td>0</td>
</tr>
<tr>
<td>Bernoulli($p$)</td>
<td>$p$</td>
<td>$p(1 - p)$</td>
</tr>
<tr>
<td>Binomial($n, p$)</td>
<td>$np$</td>
<td>$np(1 - p)$</td>
</tr>
<tr>
<td>Geometric($p$)</td>
<td>$1/p$</td>
<td>$(1 - p)/p^2$</td>
</tr>
<tr>
<td>Poisson($\lambda$)</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>Uniform($a, b$)</td>
<td>$(a + b)/2$</td>
<td>$(b - a)^2/12$</td>
</tr>
<tr>
<td>Normal($\mu, \sigma^2$)</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Exponential($\beta$)</td>
<td>$\beta$</td>
<td>$\beta^2$</td>
</tr>
<tr>
<td>Gamma($\alpha, \beta$)</td>
<td>$\alpha\beta$</td>
<td>$\alpha\beta^2$</td>
</tr>
<tr>
<td>Beta($\alpha, \beta$)</td>
<td>$\alpha/ (\alpha + \beta)$</td>
<td>$\alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$</td>
</tr>
<tr>
<td>$t_\nu$</td>
<td>0 (if $\nu &gt; 1$)</td>
<td>$\nu/(\nu - 2)$ (if $\nu &gt; 2$)</td>
</tr>
<tr>
<td>$\chi^2_p$</td>
<td>$p$</td>
<td>$2p$</td>
</tr>
<tr>
<td>Multinomial($n, p$)</td>
<td>$np$</td>
<td>see below</td>
</tr>
<tr>
<td>Multivariate Normal($\mu, \Sigma$)</td>
<td>$\mu$</td>
<td>$\Sigma$</td>
</tr>
</tbody>
</table>

(Wasserman, 2003)
CDF to PDF trick
Monty Hall Problem
Population and Samples

Population

- complete: described by pdf of RV
- expectation: $E(X)$ or $\mu$
- variance: $Var(X)$ or $\sigma^2$

Sample

- presumed random subset of population
- sample mean: $\bar{X}$
- sample variance: $s^2$
Population and Samples

Population

complete: described by pdf of RV
expectation: $E(X)$ or $\mu$
variance: $Var(X)$ or $\sigma^2$

Sample

presumed random subset of population

sample mean: $\bar{X}$
sample variance: $s^2$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
Population and Samples

Population

complete: described by pdf of RV
expectation: \( E(X) \) or \( \mu \)
variance: \( Var(X) \) or \( \sigma^2 \)

Sample

presumed random subset of population
sample mean: \( \bar{X} \)
sample variance: \( s^2 \)

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2
\]
Population and Samples

Population

Sample 1

Sample 2

complete: described by pdf of RV
expectation: $E(X)$ or $\mu$
variance: $Var(X)$ or $\sigma^2$

presumed random subset of population
sample mean: $\bar{X}$
sample variance: $s^2$
Population and Samples

Population

- complete: described by pdf of RV
- expectation: $E(X)$ or $\mu$
- variance: $Var(X)$ or $\sigma^2$

Sample 1

- presumed random subset of population
- sample mean: $\bar{X}_1$
- sample variance: $s_1^2$

Sample 2

- sample mean: $\bar{X}_2$
- sample variance: $s_2^2$
Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

if \(X_1, \ldots, X_n\) are iid then \(\bar{X}_n \xrightarrow{p} \mu\)
Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

If $X_1, \ldots, X_n$ are iid then $\overline{X}_n \xrightarrow{p} \mu$

means $P(|\overline{X} - \mu| > \epsilon) \rightarrow 0$

independent and identically distributed
Law of Large Numbers

"Weak Law of Large Numbers" (WLLN)

\[ \text{if } X_1, \ldots, X_n \text{ are iid then } \bar{X}_n \xrightarrow{p} \mu \]

means \( P(|\bar{X} - \mu| > \epsilon) \to 0 \)

The sample mean converges with the population mean in probability for every \( \epsilon > 0 \)

(\( \bar{X} \) is close to \( \mu \) with high probability)
Can we describe the distribution of \( \bar{X} \) converges to a normal distribution.

\[ \sqrt{n}(\bar{X}_n - \mu) \]