YSU ASDS, Statistics, Fall 2019

Lecture 10-11

Michael Poghosyan

25 Sep 2019

Descriptive Statistics / Probability Reminder

Contents

Last Lecture ReCap

- Give the definition of the Sample Covariance

Last Lecture ReCap

- Give the definition of the Sample Covariance
- Give the definition of the Sample Correlation Coefficient

Last Lecture ReCap

- Give the definition of the Sample Covariance
- Give the definition of the Sample Correlation Coefficient
- Why(when) is Covariance preferrable?

Last Lecture ReCap

- Give the definition of the Sample Covariance
- Give the definition of the Sample Correlation Coefficient
- Why(when) is Covariance preferrable?
- Why(when) is Correlation Coefficient preferrable?

Last Lecture ReCap

- Give the definition of the Sample Covariance
- Give the definition of the Sample Correlation Coefficient
- Why(when) is Covariance preferrable?
- Why(when) is Correlation Coefficient preferrable?
- What can be inferred from the Covariance?

Last Lecture ReCap

- Give the definition of the Sample Covariance
- Give the definition of the Sample Correlation Coefficient
- Why(when) is Covariance preferrable?
- Why(when) is Correlation Coefficient preferrable?
- What can be inferred from the Covariance?
- What can be inferred from the Correlation Coefficient?

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

There exist different measures to measure the relationship between variables. We will talk about two of them: Spearman's ρ and Kentall's τ.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

There exist different measures to measure the relationship between variables. We will talk about two of them: Spearman's ρ and Kentall's τ. These are examples of *Rank Correlation* measures.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

There exist different measures to measure the relationship between variables. We will talk about two of them: Spearman's ρ and Kentall's τ. These are examples of *Rank Correlation* measures.

To define them, first we define **Ranks**: assume $x : x_1, x_2, ..., x_n$ is a 1D Dataset.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

There exist different measures to measure the relationship between variables. We will talk about two of them: Spearman's ρ and Kentall's τ. These are examples of *Rank Correlation* measures.

To define them, first we define **Ranks**: assume $x : x_1, x_2, ..., x_n$ is a 1D Dataset. Then the Rank of x_1 is its place in the sorted array $sort(x)$, the Rank of x_2 is its place in that sorted array, etc.

Rank Correlations

Well, we have learned that the Sample Correlation Coefficient is measuring the *strength of linear relationship* between the variables. But, unfortunately, it is not always capturing nonlinear relationships.

There exist different measures to measure the relationship between variables. We will talk about two of them: Spearman's ρ and Kentall's τ. These are examples of *Rank Correlation* measures.

To define them, first we define **Ranks**: assume $x : x_1, x_2, ..., x_n$ is a 1D Dataset. Then the Rank of x_1 is its place in the sorted array $sort(x)$, the Rank of x_2 is its place in that sorted array, etc. and $Rank(x)$ is the Dataset of the elements $Rank(x_1), ..., Rank(x_n)$.

Example:

Example: Construct the Ranks Dataset for

$$x: \ 4, -3, 1, 55, 6, 2$$

Solution: OTB

Example:

Example: Construct the Ranks Dataset for

$$x: \ 4, -3, 1, 55, 6, 2$$

Solution: OTB

Spearman's ρ

Now assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

Spearman's ρ

Now assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

First we calculate the ranks:

$$Rank(x): \; r_1, r_2, ..., r_n \qquad \text{and} \qquad Rank(y): \; i_1, ..., i_n.$$

Spearman's ρ

Now assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

First we calculate the ranks:

$$Rank(x): \; r_1, r_2, ..., r_n \qquad \text{and} \qquad Rank(y): \; i_1, ..., i_n.$$

Definition: The **Spearman's rank correlation coefficient** ρ is defined as:

$$\rho = cor(Rank(x), Rank(y)).$$

Spearman's ρ

Now assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

First we calculate the ranks:

$$Rank(x): \; r_1, r_2, ..., r_n \qquad \text{and} \qquad Rank(y): \; i_1, ..., i_n.$$

Definition: The **Spearman's rank correlation coefficient** ρ is defined as:

$$\rho = cor(Rank(x), Rank(y)).$$

So, basically, Spearman's ρ between x and y is the Correlation between the ranks of x and y.

Spearman's ρ

There is another formula to calculate the Spearman's *rho*: if we will denote $d_k = rank(x_k) - rank(y_k)$, then

$$\rho = 1 - \frac{6 \cdot \sum_{k=1}^{n} d_k^2}{n(n^2-1)}.$$

Spearman's ρ

There is another formula to calculate the Spearman's *rho*: if we will denote $d_k = rank(x_k) - rank(y_k)$, then

$$\rho = 1 - \frac{6 \cdot \sum_{k=1}^{n} d_k^2}{n(n^2 - 1)}.$$

Idea: ρ is measuring monotonic relationship between the Datasets. $\rho = \pm 1$ means (in case we do not have repeatitionsin the Datasets) there exists a monotonic relationship between the Datasets.

Spearman's ρ

There is another formula to calculate the Spearman's *rho*: if we will denote $d_k = rank(x_k) - rank(y_k)$, then

$$\rho = 1 - \frac{6 \cdot \sum_{k=1}^{n} d_k^2}{n(n^2-1)}.$$

Idea: ρ is measuring monotonic relationship between the Datasets. $\rho = \pm 1$ means (in case we do not have repeatitionsin the Datasets) there exists a monotonic relationship between the Datasets.
Note: Pearson's Correlation Coefficient is sensitive to outliers, but Spearman's ρ is robust wrt outliers.

Example:

Example: Calculate the Spearman's ρ for

$$x: \ 0, 3, 1 \qquad \text{and} \qquad y: \ -3, -2, -1$$

Solution: OTB

Example:

Example: Calculate the Spearman's ρ for

$$x:\ 0, 3, 1 \qquad \text{and} \qquad y:\ -3, -2, -1$$

Solution: OTB

```
x <- c(0,3,1); y <- -3:-1
cor(x,y)
```

```
## [1] 0.3273268
```

```
cor(x,y,method = "spearman")
```

```
## [1] 0.5
```

Examples:

```r
x <- runif(100,-1,1)
y <- x^2 + 0.3* rnorm(100)
cor(x,y)
```

```
## [1] -0.05774362
```

```r
cor(x,y,method = "spearman")
```

```
## [1] -0.005640564
```

Examples:

```r
x <- runif(100,-1,1)
y <- x^3 + 0.3* rnorm(100)
cor(x,y)
```

```
## [1] 0.7267431
```

```r
cor(x,y,method = "spearman")
```

```
## [1] 0.7118032
```

Examples:

```r
x <- runif(100,0,1)
y <- x^4 + rnorm(100)
z <- y; z[1] = 100 ## Introducing an outlier
cor(x,y)
```

```
## [1] 0.1829122
```

```r
cor(x,y,method = "spearman")
```

```
## [1] 0.1264086
```

```r
cor(x,z)
```

```
## [1] -0.164603
```

```r
cor(x,z,method = "spearman")
```

```
## [1] 0.1224002
```

Kendall's τ

Another measure of the Rank Correlation is the **Kendal's** τ: again assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

Kendall's τ

Another measure of the Rank Correlation is the **Kendal's** τ: again assume we have two numerical 1D Datasets of the same size:

$$x: \ x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \ y_1, y_2, ..., y_n.$$

We say that the pair (x_i, y_i) and (x_j, y_j) $(i \neq j)$ are *concordant*, if either

$$x_i < x_j \qquad \text{and} \qquad y_i < y_j$$

or

$$x_i > x_j \qquad \text{and} \qquad y_i > y_j.$$

Kendall's τ

Another measure of the Rank Correlation is the **Kendal's** τ: again assume we have two numerical 1D Datasets of the same size:

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

We say that the pair (x_i, y_i) and (x_j, y_j) $(i \neq j)$ are *concordant*, if either

$$x_i < x_j \qquad \text{and} \qquad y_i < y_j$$

or

$$x_i > x_j \qquad \text{and} \qquad y_i > y_j.$$

We say that the pair (x_i, y_i) and (x_j, y_j) $(i \neq j)$ are *discordant*, if either

$$x_i < x_j \qquad \text{and} \qquad y_i > y_j$$

or

$$x_i > x_j \qquad \text{and} \qquad y_i < y_j.$$

Kendall's τ

Now, let $con(x, y)$ be the number of concordant pairs in x, y, and $dis(x, y)$ be the number of discordant pairs.

Kendall's τ

Now, let $con(x, y)$ be the number of concordant pairs in x, y, and $dis(x, y)$ be the number of discordant pairs.

Definition: Kendall's Rank Correlation Coefficient τ is defined by:

$$\tau = \frac{con(x, y) - dis(x, y)}{con(x, y) + dis(x, y)} = \frac{con(x, y) - dis(x, y)}{\binom{n}{2}}.$$

Kendall's τ

Now, let $con(x, y)$ be the number of concordant pairs in x, y, and $dis(x, y)$ be the number of discordant pairs.

Definition: Kendall's Rank Correlation Coefficient τ is defined by:

$$\tau = \frac{con(x, y) - dis(x, y)}{con(x, y) + dis(x, y)} = \frac{con(x, y) - dis(x, y)}{\binom{n}{2}}.$$

Another, explicit way, of writing this is:

$$\tau = \frac{\sum_{i<j} sign(x_i - x_j) \cdot sign(y_i - y_j)}{\frac{n(n-1)}{2}}.$$

Kendall's τ

Facts:

- If the rankings of x and y are the same (so x and y are in increasing relationship), then $\tau = 1$

Kendall's τ

Facts:

- If the rankings of x and y are the same (so x and y are in increasing relationship), then $\tau = 1$
- If the rankings of x and y are inverse of each other (so x and y are in decreasing relationship), then $\tau = -1$

Example:

```
x <- c(1,3,4); y <- c(-5, 10, 1000)
cor(x,y)
```

```
## [1] 0.7643896
```

```
cor(x,y,method = "spearman")
```

```
## [1] 1
```

```
cor(x,y,method = "kendall")
```

```
## [1] 1
```

Notes and Additions: Rank Correlations

- There can be ties (when $x_i = x_j$ or $y_i = y_j$), and there are different methods to deal with ties. See Wiki.

Notes and Additions: Rank Correlations

- There can be ties (when $x_i = x_j$ or $y_i = y_j$), and there are different methods to deal with ties. See Wiki.
- Rank correlations are more robust than the Pearson's Correlation Coefficient;

Notes and Additions: Rank Correlations

- There can be ties (when $x_i = x_j$ or $y_i = y_j$), and there are different methods to deal with ties. See Wiki.
- Rank correlations are more robust than the Pearson's Correlation Coefficient;
- Rank correlations can be calculated even for Ordinal Variable Datasets

Notes and Additions: Rank Correlations

- There can be ties (when $x_i = x_j$ or $y_i = y_j$), and there are different methods to deal with ties. See Wiki.
- Rank correlations are more robust than the Pearson's Correlation Coefficient;
- Rank correlations can be calculated even for Ordinal Variable Datasets
- There are other Rank Correlation measures, see Wiki

Correlation Coefficient as a Cosine

Recall from Algebra/Calculus, that if

$$\mathbf{a} = [a_1, ..., a_n]^T \qquad \text{and} \qquad \mathbf{b} = [b_1, ..., b_n]^T,$$

then we define the angle α between **a** and **b** in the following way:

Correlation Coefficient as a Cosine

Recall from Algebra/Calculus, that if

$$\mathbf{a} = [a_1, ..., a_n]^T \qquad \text{and} \qquad \mathbf{b} = [b_1, ..., b_n]^T,$$

then we define the angle α between $\mathbf{a}$ and $\mathbf{b}$ in the following way:

$$\cos \alpha = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|},$$

Correlation Coefficient as a Cosine

Recall from Algebra/Calculus, that if

$$\mathbf{a} = [a_1, ..., a_n]^T \qquad \text{and} \qquad \mathbf{b} = [b_1, ..., b_n]^T,$$

then we define the angle α between $\mathbf{a}$ and $\mathbf{b}$ in the following way:

$$\cos\alpha = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|},$$

where the numerator is the dot product of $\mathbf{a}$ and $\mathbf{b}$ and

$$\|\mathbf{a}\| = \sqrt{(a_1)^2 + (a_2)^2 + ... + (a_n)^2}$$

is the length of the vector $\mathbf{a}$.

Correlation Coefficient as a Cosine

Now, assume we have two Datasets

$$x: \ x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \ y_1, y_2, ..., y_n.$$

Correlation Coefficient as a Cosine

Now, assume we have two Datasets

$$x: \; x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_n.$$

We center our Datasets: calculate

$$x - \bar{x}: \; x_1 - \bar{x}, ..., x_n - \bar{x}, \qquad \text{and} \qquad y - \bar{y}: \; y_1 - \bar{y}, ..., y_n - \bar{y}.$$

Correlation Coefficient as a Cosine

Now, assume we have two Datasets

$$x:\ x_1, x_2, ..., x_n, \qquad \text{and} \qquad y:\ y_1, y_2, ..., y_n.$$

We center our Datasets: calculate

$$x-\bar{x}:\ x_1-\bar{x}, ..., x_n-\bar{x}, \qquad \text{and} \qquad y-\bar{y}:\ y_1-\bar{y}, ..., y_n-\bar{y}.$$

If we will denote

$$\mathbf{a} = [x_1-\bar{x}, ..., x_n-\bar{x}]^T, \qquad \mathbf{b} = [y_1-\bar{y}, ..., y_n-\bar{y}]^T,$$

Correlation Coefficient as a Cosine

Now, assume we have two Datasets

$$x: \ x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \ y_1, y_2, ..., y_n.$$

We center our Datasets: calculate

$$x - \bar{x}: \ x_1 - \bar{x}, ..., x_n - \bar{x}, \qquad \text{and} \qquad y - \bar{y}: \ y_1 - \bar{y}, ..., y_n - \bar{y}.$$

If we will denote

$$\mathbf{a} = [x_1 - \bar{x}, ..., x_n - \bar{x}]^T, \qquad \mathbf{b} = [y_1 - \bar{y}, ..., y_n - \bar{y}]^T,$$

then it is easy to see that the Correlation Coefficient $cor(x, y)$ is the Cosine of the angle between **a** and **b**.

Correlation Coefficient as a Cosine

Now, assume we have two Datasets

$$x: \ x_1, x_2, ..., x_n, \qquad \text{and} \qquad y: \ y_1, y_2, ..., y_n.$$

We center our Datasets: calculate

$$x - \bar{x}: \ x_1 - \bar{x}, ..., x_n - \bar{x}, \qquad \text{and} \qquad y - \bar{y}: \ y_1 - \bar{y}, ..., y_n - \bar{y}.$$

If we will denote

$$\mathbf{a} = [x_1 - \bar{x}, ..., x_n - \bar{x}]^T, \qquad \mathbf{b} = [y_1 - \bar{y}, ..., y_n - \bar{y}]^T,$$

then it is easy to see that the Correlation Coefficient $cor(x, y)$ is the Cosine of the angle between **a** and **b**.

See also Wiki

Correlation is not Causation

- Some Examples: Spurious Correlations

Anscombe Quartet

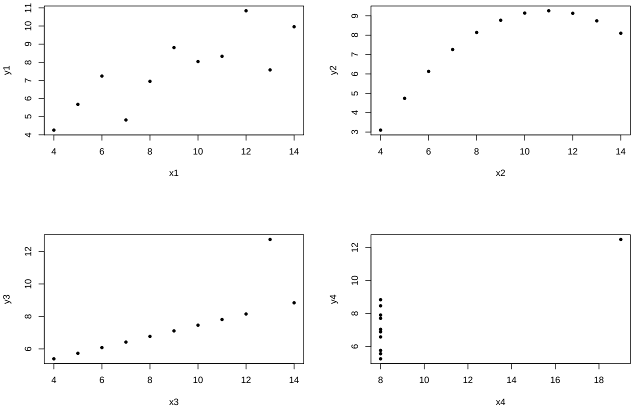
See Wiki

```
anscombe
```

```
##    x1 x2 x3 x4    y1   y2    y3    y4
## 1  10 10 10  8  8.04 9.14  7.46  6.58
## 2   8  8  8  8  6.95 8.14  6.77  5.76
## 3  13 13 13  8  7.58 8.74 12.74  7.71
## 4   9  9  9  8  8.81 8.77  7.11  8.84
## 5  11 11 11  8  8.33 9.26  7.81  8.47
## 6  14 14 14  8  9.96 8.10  8.84  7.04
## 7   6  6  6  8  7.24 6.13  6.08  5.25
## 8   4  4  4 19  4.26 3.10  5.39 12.50
## 9  12 12 12  8 10.84 9.13  8.15  5.56
## 10  7  7  7  8  4.82 7.26  6.42  7.91
## 11  5  5  5  8  5.68 4.74  5.73  6.89
```

Anscombe Quartet

```
attach(anscombe)
par(mfrow=c(2,2))
plot(y1~x1, pch = 20); plot(y2~x2, pch = 20);
plot(y3~x3, pch = 20); plot(y4~x4, pch = 20);
```



Anscombe Quartet

```r
c(mean(x1), mean(x2), mean(x3), mean(x4))
```

```
## [1] 9 9 9 9
```

```r
c(mean(y1), mean(y2), mean(y3), mean(y4))
```

```
## [1] 7.500909 7.500909 7.500000 7.500909
```

```r
c(var(x1), var(x2), var(x3), var(x4))
```

```
## [1] 11 11 11 11
```

```r
c(var(y1), var(y2), var(y3), var(y4))
```

```
## [1] 4.127269 4.127629 4.122620 4.123249
```

```r
c(cor(x1,y1), cor(x2,y2), cor(x3,y3), cor(x4,y4))
```

```
## [1] 0.8164205 0.8162365 0.8162867 0.8165214
```

Anscombe Quartet

Moral: Just calculating numbers (summary statistics) is not enough, visualize your Data if possible.

Reminder on Random Variables and Distributions

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

- Ω is the Sample Space

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

- Ω is the Sample Space
- $\mathcal{F}$ is the set of all Events

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

- Ω is the Sample Space
- $\mathcal{F}$ is the set of all Events
- $\mathbb{P}$ is a Probability Measure

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

- Ω is the Sample Space
- $\mathcal{F}$ is the set of all Events
- $\mathbb{P}$ is a Probability Measure

Definition: Any (measurable) function $X : \Omega \to \mathbb{R}$ is called a r.v. on the Probability Space $(\Omega, \mathbb{P})$.

Random Variables

Everything starts at the Probability Space (Experiment, Model): we are given

$$(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{or, we usually use} \qquad (\Omega, \mathbb{P}),$$

where

- Ω is the Sample Space
- $\mathcal{F}$ is the set of all Events
- $\mathbb{P}$ is a Probability Measure

Definition: Any (measurable) function $X : \Omega \to \mathbb{R}$ is called a r.v. on the Probability Space $(\Omega, \mathbb{P})$.

So $X = X(\omega)$, but usually we forget about ω, and use X.

Main Complete Characteristics of a r.v.

If X is a r.v., then we get the **complete information** (everything we can get) about X from either its CDF or PDF/PMF.

Main Complete Characteristics of a r.v.

If X is a r.v., then we get the **complete information** (everything we can get) about X from either its CDF or PDF/PMF.

Definition: The CDF of X is defined as

$$F(x) = F_X(x) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R}.$$

Main Complete Characteristics of a r.v.

If X is a r.v., then we get the **complete information** (everything we can get) about X from either its CDF or PDF/PMF.

Definition: The CDF of X is defined as

$$F(x) = F_X(x) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R}.$$

Definition: We say that X is a *Continuous r.v.*, if it has a PDF: a function $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt, \qquad \forall x \in \mathbb{R}.$$

Main Complete Characteristics of a r.v.

If X is a r.v., then we get the **complete information** (everything we can get) about X from either its CDF or PDF/PMF.

Definition: The CDF of X is defined as

$$F(x) = F_X(x) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R}.$$

Definition: We say that X is a *Continuous r.v.*, if it has a PDF: a function $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt, \qquad \forall x \in \mathbb{R}.$$

So for a Continuous r.v., another complete characteristic, besides the CDF, is its PDF.

Discrete r.v.s

Definition: We say that X is a *Discrete r.v.*, if the set of values of X is finite or countably infinite.

Discrete r.v.s

Definition: We say that X is a *Discrete r.v.*, if the set of values of X is finite or countably infinite. And if the possible values are x_k, $k = 1, 2, ...$, then we define the PMF of X as

$$f(x_k) = \mathbb{P}(X = x_k), \qquad k = 1, 2, ...,$$

Discrete r.v.s

Definition: We say that X is a *Discrete r.v.*, if the set of values of X is finite or countably infinite. And if the possible values are x_k, $k = 1, 2, ...$, then we define the PMF of X as

$$f(x_k) = \mathbb{P}(X = x_k), \qquad k = 1, 2, ...,$$

or, in a table form,

Values of X	x_1	x_2	...
$\mathbb{P}(X = x)$	p_1	p_2	...

Main Partial Characteristics of a r.v.

Main partial characteristics of a r.v. X are:

Main Partial Characteristics of a r.v.

Main partial characteristics of a r.v. X are:

- the Expected Value (Mean):

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \, (cont.) \quad | \quad \mathbb{E}(X) = \sum_{k} x_k \cdot \mathbb{P}(X = x_k) \, (disc.).$$

Main Partial Characteristics of a r.v.

Main partial characteristics of a r.v. X are:

- the Expected Value (Mean):

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \, (cont.) \quad | \quad \mathbb{E}(X) = \sum_k x_k \cdot \mathbb{P}(X = x_k) \, (disc.).$$

Note:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx \, (cont.) \quad | \quad \mathbb{E}(g(X)) = \sum_k g(x_k) \cdot \mathbb{P}(X = x_k) \, ($$

Main Partial Characteristics of a r.v.

Main partial characteristics of a r.v. X are:

- the Expected Value (Mean):

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \, (cont.) \quad | \quad \mathbb{E}(X) = \sum_k x_k \cdot \mathbb{P}(X = x_k) \, (disc.).$$

Note:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx \, (cont.) \quad | \quad \mathbb{E}(g(X)) = \sum_k g(x_k) \cdot \mathbb{P}(X = x_k) \, ($$

- The Variance

$$Var(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}(X^2) - \Big[\mathbb{E}(X)\Big]^2.$$

Important Discrete Distributions

Bernoulli Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)

Bernoulli Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim \mathit{Bernoulli}(p)$;

Bernoulli Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Bernoulli(p)$;
- Support: $\{0, 1\}$

Bernoulli Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Bernoulli(p)$;
- Support: $\{0,1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1 - p$	p

Bernoulli Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim \mathit{Bernoulli}(p)$;
- Support: $\{0,1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X=x)$	$1-p$	p

Note: This can be written in the form:

$$f(x) = f(x;p) = f(x|p) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0,1\}.$$

Bernoulli Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim \mathit{Bernoulli}(p)$;
- Support: $\{0,1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1-p$	p

Note: This can be written in the form:

$$f(x) = f(x;p) = f(x|p) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0,1\}.$$

- Mean and Variance: $\mathbb{E}(X) = p$, $\mathit{Var}(X) = p(1-p)$.

Bernoulli Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Bernoulli(p)$;
- Support: $\{0, 1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1 - p$	p

Note: This can be written in the form:

$$f(x) = f(x; p) = f(x|p) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0, 1\}.$$

- Mean and Variance: $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$.
- Models: Models binary output, "success-failure" type Experiments, a lot of examples.

Bernoulli Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Bernoulli(p)$;
- Support: $\{0,1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1-p$	p

Note: This can be written in the form:

$$f(x) = f(x;p) = f(x|p) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0,1\}.$$

- Mean and Variance: $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$.
- Models: Models binary output, "success-failure" type Experiments, a lot of examples.
- **R** name: `binom` with the parameters `size=1` and `prob`

Bernoulli Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Bernoulli(p)$;
- Support: $\{0,1\}$
- PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1-p$	p

Note: This can be written in the form:

$$f(x) = f(x;p) = f(x|p) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0,1\}.$$

- Mean and Variance: $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$.
- Models: Models binary output, "success-failure" type Experiments, a lot of examples.
- **R** name: `binom` with the parameters `size=1` and `prob`

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;
- Support: $\{0, 1, 2, ..., n\}$

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Binom(n,p)$;
- Support: $\{0, 1, 2, ..., n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\binom{n}{1} p^1 (1-p)^{n-1}$	...	$\binom{n}{k} p^k (1-p)^{n-k}$	...	$\binom{n}{n} p^n (1-p$

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;
- Support: $\{0, 1, 2, ..., n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\binom{n}{1} p^1 (1-p)^{n-1}$	...	$\binom{n}{k} p^k (1-p)^{n-k}$	...	$\binom{n}{n} p^n (1-p$

- Mean and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1 - p)$.

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;
- Support: $\{0, 1, 2, ..., n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\binom{n}{1} p^1 (1-p)^{n-1}$	...	$\binom{n}{k} p^k (1-p)^{n-k}$	...	$\binom{n}{n} p^n (1 - p$

- Mean and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1-p)$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment.

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;
- Support: $\{0, 1, 2, ..., n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\binom{n}{1} p^1 (1-p)^{n-1}$	...	$\binom{n}{k} p^k (1-p)^{n-k}$	...	$\binom{n}{n} p^n (1-p$

- Mean and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1-p)$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment.
- **R** name: `binom` with the parameters `size` and `prob`

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Binom(n,p)$;
- Support: $\{0,1,2,...,n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X=x)$	$\binom{n}{0}p^0(1-p)^{n-0}$	$\binom{n}{1}p^1(1-p)^{n-1}$	...	$\binom{n}{k}p^k(1-p)^{n-k}$	...	$\binom{n}{n}p^n(1-p$

- Mean and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1-p)$.
- Models: Models the independent repetition of the $Bernoulli(p)$ Experiment.
- **R** name: `binom` with the parameters `size` and `prob`
- Additional: If $X_1, X_2, ..., X_n \sim Bernoulli(p)$ are independent, then $X_1 + X_2 + ... + X_n \sim Binom(n,p)$.

Binomial Distribution

- Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Binom(n, p)$;
- Support: $\{0, 1, 2, ..., n\}$
- PMF:

Values of X	0	1	...	k	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0}p^0(1-p)^{n-0}$	$\binom{n}{1}p^1(1-p)^{n-1}$	...	$\binom{n}{k}p^k(1-p)^{n-k}$	...	$\binom{n}{n}p^n(1-p$

- Mean and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1-p)$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment.
- **R** name: `binom` with the parameters `size` and `prob`
- Additional: If $X_1, X_2, ..., X_n \sim Bernoulli(p)$ are independent, then $X_1 + X_2 + ... + X_n \sim Binom(n, p)$.
- Example:

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, ...\}$ (or, sometimes, $\{0, 1, 2, 3, ...\}$)

Geometric Distribution

- Parameter: $p \in [0,1]$ (usually, $p \in (0,1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, \ldots\}$ (or, sometimes, $\{0, 1, 2, 3, \ldots\}$)
- PMF:

Values of X	1	2	3	...
$\mathbb{P}(X = x)$	p	$p(1-p)$	$p(1-p)^2$	...

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, ...\}$ (or, sometimes, $\{0, 1, 2, 3, ...\}$)
- PMF:

Values of X	1	2	3	...
$\mathbb{P}(X = x)$	p	$p(1-p)$	$p(1-p)^2$	...

- Mean and Variance: $\mathbb{E}(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$.

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, ...\}$ (or, sometimes, $\{0, 1, 2, 3, ...\}$)
- PMF:

Values of X	1	2	3	...
$\mathbb{P}(X = x)$	p	$p(1-p)$	$p(1-p)^2$	...

- Mean and Variance: $\mathbb{E}(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment until the *First Success*.

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, ...\}$ (or, sometimes, $\{0, 1, 2, 3, ...\}$)
- PMF:

Values of X	1	2	3	...
$\mathbb{P}(X = x)$	p	$p(1-p)$	$p(1-p)^2$	...

- Mean and Variance: $\mathbb{E}(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment until the *First Success*.
- **R** name: `geom` with the parameter `prob`

Geometric Distribution

- Parameter: $p \in [0, 1]$ (usually, $p \in (0, 1)$)
- Notation: $X \sim Geom(p)$;
- Support: $\{1, 2, 3, ...\}$ (or, sometimes, $\{0, 1, 2, 3, ...\}$)
- PMF:

Values of X	1	2	3	...
$\mathbb{P}(X = x)$	p	$p(1-p)$	$p(1-p)^2$	...

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$.
- Models: Models the independent repetition of the *Bernoulli*(p) Experiment until the *First Success*.
- **R** name: `geom` with the parameter `prob`
- Example:

```
rgeom(10,prob = 0.3)
```

```
##  [1] 11  2  0  0  6  0  2  3  5  0
```

Poisson Distribution

- Parameter: $\lambda > 0$

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

- Mean and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$.

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

- Mean and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$.
- Models: Models the number of success in a fixed interval. Say, number of calls, number of customers, number of clicks, number of page visits, ...

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

- Mean and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$.
- Models: Models the number of success in a fixed interval. Say, number of calls, number of customers, number of clicks, number of page visits, ... λ is the average number of calls, customers, clicks, page visits,...

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

- Mean and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$.
- Models: Models the number of success in a fixed interval. Say, number of calls, number of customers, number of clicks, number of page visits, ... λ is the average number of calls, customers, clicks, page visits,...
- **R** name: `pois` with the parameter `lambda`

Poisson Distribution

- Parameter: $\lambda > 0$
- Notation: $X \sim Pois(\lambda)$;
- Support: $\{0, 1, 2, 3, ...\}$
- PMF:

Values of X	0	1	2	...
$\mathbb{P}(X = x)$	$e^{-\lambda}\frac{\lambda^0}{0!}$	$e^{-\lambda}\frac{\lambda^1}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	...

- Mean and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$.
- Models: Models the number of success in a fixed interval. Say, number of calls, number of customers, number of clicks, number of page visits, ... λ is the average number of calls, customers, clicks, page visits,...
- **R** name: `pois` with the parameter `lambda`
- Example:

```
rpois(10, lambda = 2)
```

Important Continuous Distributions

Uniform Distribution

- Parameters: a, b $(a < b)$

Uniform Distribution

- Parameters: a, b ($a < b$)
- Notation: $X \sim Unif[a, b]$;

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim \mathit{Unif}[a, b]$;
- Support: $[a, b]$

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a,b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*
- **R** name: `unif` with the parameters `min = 0` and `max = 1`

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*
- **R** name: `unif` with the parameters `min = 0` and `max = 1`
- Example:

```r
runif(10, min = 2, max = 5)
```

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim \mathit{Exp}(\lambda)$ (or $\mathit{Exp}(\beta)$);

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim \mathit{Exp}(\lambda)$ (or $\mathit{Exp}(\beta)$);
- Support: $[0, +\infty)$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim \mathit{Exp}(\lambda)$ (or $\mathit{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & \mathit{otherwise} \end{cases}$$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $Exp(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $Exp(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random.

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $Exp(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average "arrival rate", the reciprocal of the average time between the events,

$$\lambda = \frac{1}{\text{average time between events}}.$$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $Exp(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average "arrival rate", the reciprocal of the average time between the events,
  $\lambda = \dfrac{1}{\text{average time between events}}.$
- **R** name: `exp` with the parameter `rate = 1`

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $Exp(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average “arrival rate”, the reciprocal of the average time between the events,

$$\lambda = \frac{1}{\text{average time between events}}.$$

- **R** name: `exp` with the parameter `rate = 1`
- E

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: Different things. Important is because of the CLT

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: Different things. Important is because of the CLT
- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: Different things. Important is because of the CLT
- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`
- Example:

```r
rnorm(10, mean = 2, sd = 3)
```

```
##  [1]  1.0604526  2.6384238  1.8464923 -3.6916389  0.627
##  [7]  6.0648668  0.9021995 -1.0345287  3.2685995
```

Normal (Gaussian) Distribution

Note:

- **R** is using $\mathcal{N}(mean, sd)$ format
- **In Math** we are using the $\mathcal{N}(mean, variance)$ format

Normal (Gaussian) Distribution

Note:

- **R** is using $\mathcal{N}(mean, sd)$ format
- **In Math** we are using the $\mathcal{N}(mean, variance)$ format

So if you want to generate a sample of size 100 from $\mathcal{N}(2, 9)$, use the command `rnorm(100, mean = 2, sd = 3)`.

Normal (Gaussian) Distribution

Note:

- **R** is using $\mathcal{N}(mean, sd)$ format
- **In Math** we are using the $\mathcal{N}(mean, variance)$ format

So if you want to generate a sample of size 100 from $\mathcal{N}(2, 9)$, use the command `rnorm(100, mean = 2, sd = 3)`.

Additional Properties:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ and

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) =$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$
- If $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for $k = 1, ..., n$ are independent, then $X_1 + X_2 + ... + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$.

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$
- If $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for $k = 1, ..., n$ are independent, then $X_1 + X_2 + ... + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$.
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(-\sigma < X - \mu < \sigma) \approx 0.6827,$$

$$\mathbb{P}(-2\sigma < X - \mu < 2\sigma) \approx 0.9545$$

and

$$\mathbb{P}(-3\sigma < X - \mu < 3\sigma) \approx 0.9973.$$

Additions

- See many other Distributions at Wiki or in different textbooks.

Additions

- See many other Distributions at Wiki or in different textbooks.
- Relationships among probability distributions: Wiki

Additions

- See many other Distributions at Wiki or in different textbooks.
- Relationships among probability distributions: Wiki
- Another page for the Relationship: L. Leemis Page