Mathematics for Machine Learning

Vazgen Mikayelyan

September 8, 2020

FAST DISCOVERING THE FUTURE

Geometric Distribution

Definition

Let X be the number of independent Bernoulli trials $B(1,p), 0 < p < 1$ needed to obtain a successful outcome.

Geometric Distribution

Definition

Let X be the number of independent Bernoulli trials $B(1,p), 0 < p < 1$ needed to obtain a successful outcome. We will call such X **Geometric** *random variable. Its PMF is*

$$P(X = k) =$$

Geometric Distribution

Definition

Let X *be the number of independent Bernoulli trials* $B(1, p), 0 < p < 1$ *needed to obtain a successful outcome. We will call such* X **Geometric** *random variable. Its PMF is*

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots.$$

The fact that X *is a geometric random variable is denoted as* $X \sim Geo(p)$.

Geometric Distribution

Definition

Let X be the number of independent Bernoulli trials $B(1,p), 0 < p < 1$ needed to obtain a successful outcome. We will call such X **Geometric** *random variable. Its PMF is*

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \dots.$$

The fact that X is a geometric random variable is denoted as $X \sim Geo(p)$.

Example

Let X be a geometric random variable with parameter p. Prove that

$$\mathbb{E}\left[X\right] = \frac{1}{p}, Var\left(X\right) = \frac{1-p}{p^2}.$$

Uniform Distribution

Definition

A random variable is said to be Uniformly distributed on the interval (a, b) *if its PDF is given by*

$$f(x) = \begin{cases} 0, & \textit{if } x \notin (a, b), \\ \dfrac{1}{b-a}, & \textit{if } x \in (a, b). \end{cases}$$

Uniform Distribution

Definition

A random variable is said to be Uniformly distributed on the interval (a, b) *if its PDF is given by*

$$f(x) = \begin{cases} 0, & \text{if } x \notin (a, b), \\ \dfrac{1}{b-a}, & \text{if } x \in (a, b). \end{cases}$$

Proposition

Prove that CDF of uniform random variable on the interval (a, b) *is the following function*

$$F(x) = \begin{cases} 0, & \text{if } x \leq a, \\ \dfrac{x-a}{b-a}, & \text{if } x \in (a, b), \\ 1, & \text{if } x \geq b. \end{cases}$$

Continuous random variables

Normal Random Variable

A continuous random variable X is said to be **Normally Distributed** (or simply, **Normal Random Variable**) with parameters μ and σ^2 if the PDF of X has the following form:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

The parameter μ is called **mean.** The fact that X is a normal RV with parameters μ and σ^2 will be denoted as $X \sim N(\mu, \sigma^2)$.

Continuous random variables

Normal Random Variable

A continuous random variable X is said to be **Normally Distributed** (or simply, **Normal Random Variable**) with parameters μ and σ^2 if the PDF of X has the following form:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

The parameter μ is called **mean.** The fact that X is a normal RV with parameters μ and σ^2 will be denoted as $X \sim N(\mu, \sigma^2)$.

Example

Prove that if X is normally distributed with parameters μ and σ^2, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a^2\sigma^2$.

Continuous random variables

Standardization of Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

Continuous random variables

Standardization of Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.
Such a RV Z is said to be a **standard**, or a **unit**, normal RV.

Continuous random variables

Standardization of Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.
Such a RV Z is said to be a **standard**, or a **unit**, normal RV.

Example

Show that the parameters μ *and* σ^2 *of a normal random variable represent, respectively, its expected value and variance.*

Continuous random variables

Exponential random variable

A continuous RV whose PDF is given, for some $\lambda > 0$, by

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

is said to be an **exponential** RV (or, more simply, is said to be exponentially distributed) with parameter λ.

Continuous random variables

Exponential random variable

A continuous RV whose PDF is given, for some $\lambda > 0$, by

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

is said to be an **exponential** RV (or, more simply, is said to be exponentially distributed) with parameter λ.

Example

Let X be an exponential random variable with parameter λ. Find the CDF $F(a)$ of the RV X.

Definition

We say that a nonnegative random variable X is **memoryless** *if*

$$\mathbb{P}\{X > s + t | X > t\} = \mathbb{P}\{X > s\} \textit{ for all } s, t \geq 0.$$

Continuous random variables

Definition

We say that a nonnegative random variable X is **memoryless** *if*

$$\mathbb{P}\{X > s + t | X > t\} = \mathbb{P}\{X > s\} \textit{ for all } s, t \geq 0.$$

Proposition

Let X be an exponential random variable. Prove that it is memoryless.

Continuous random variables

Definition

We say that a nonnegative random variable X is **memoryless** *if*

$$\mathbb{P}\{X > s + t | X > t\} = \mathbb{P}\{X > s\} \text{ for all } s, t \geq 0.$$

Proposition

Let X be an exponential random variable. Prove that it is memoryless.

Proposition

Let X be an exponential random variable with parameter λ. Prove that

$$\mathbb{E}[X] = \frac{1}{\lambda} \text{ and } Var(X) = \frac{1}{\lambda^2}.$$

In practice, the exponential distribution often arises as the distribution of the amount of time until some specific event occurs.

In practice, the exponential distribution often arises as the distribution of the amount of time until some specific event occurs. For instance, the amount of time

- until an earthquake occurs,
- until a new war breaks out,
- until a telephone call you receive turns out to be a wrong number

are all random variables that tend in practice to have exponential distributions.

In practice, the exponential distribution often arises as the distribution of the amount of time until some specific event occurs. For instance, the amount of time

- until an earthquake occurs,
- until a new war breaks out,
- until a telephone call you receive turns out to be a wrong number

are all random variables that tend in practice to have exponential distributions.

Example

The time (in hours) required to repair a machine is an exponentially distributed random variable with parameter $\lambda = \frac{1}{2}$. What is
(a) the probability that a repair time exceeds 2 hours?
(b) the conditional probability that a repair takes at least 10 hours, given that its duration exceeds 9 hours?

Continuous random variables

Example

The number of years a radio functions is exponentially distributed with parameter $\lambda = 1/8$. *If Jones buys a used radio, what is the probability that it will be working after an additional 8 years?*

Continuous random variables

Example

The number of years a radio functions is exponentially distributed with parameter $\lambda = 1/8$. *If Jones buys a used radio, what is the probability that it will be working after an additional 8 years?*

Example

Suppose that the number of km that a car can run before its battery wears out is exponentially distributed with an average value of 20,000 km. If a person desires to take a 1000-km trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery? What can be said when the distribution is not exponential?

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right)=Cov\left(Y,X\right),$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right)=Cov\left(Y,X\right),$
2. $Cov\left(X,X\right)=Var\left(X\right),$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right) = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)\left(Y - \mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right) = Cov\left(Y,X\right),$
2. $Cov\left(X,X\right) = Var\left(X\right),$
3. $Cov\left(a \cdot X,Y\right) = a \cdot Cov\left(X,Y\right),$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right)=Cov\left(Y,X\right),$
2. $Cov\left(X,X\right)=Var\left(X\right),$
3. $Cov\left(a\cdot X,Y\right)=a\cdot Cov\left(X,Y\right),$
4. $Cov\left(X,Y\right)=\mathbb{E}\left[XY\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right)=Cov\left(Y,X\right),$
2. $Cov\left(X,X\right)=Var\left(X\right),$
3. $Cov\left(a\cdot X,Y\right)=a\cdot Cov\left(X,Y\right),$
4. $Cov\left(X,Y\right)=\mathbb{E}\left[XY\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$
5. $Cov\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{k}Y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{k}Cov\left(X_i,Y_j\right),$

Covariance

Definition

The covariance between random variables X and Y is defined by

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

Properties

1. $Cov\left(X,Y\right)=Cov\left(Y,X\right),$
2. $Cov\left(X,X\right)=Var\left(X\right),$
3. $Cov\left(a\cdot X,Y\right)=a\cdot Cov\left(X,Y\right),$
4. $Cov\left(X,Y\right)=\mathbb{E}\left[XY\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$
5. $Cov\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{k}Y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{k}Cov\left(X_i,Y_j\right),$
6. $Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}Var\left(X_i\right)+2\sum_{i<j}Cov\left(X_iX_j\right),$

Correlation

Definition

The correlation between random variables X *and* Y*, as long as* X *and* Y *are not constant, is defined by*

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Correlation

Definition

The correlation between random variables X and Y, as long as X and Y are not constant, is defined by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Properties

1. $\rho(X,Y) = \rho(Y,X)$,

Correlation

Definition

The correlation between random variables X and Y, as long as X and Y are not constant, is defined by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

Properties

1. $\rho(X,Y) = \rho(Y,X)$,
2. $|\rho(X,Y)| \leq 1$,

Correlation

Definition

The correlation between random variables X *and* Y*, as long as* X *and* Y *are not constant, is defined by*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)\, Var(Y)}}.$$

Properties

1. $\rho(X, Y) = \rho(Y, X)$,
2. $|\rho(X, Y)| \leq 1$,
3. $Y = aX + b$, for some constants a, b iff $\rho(X, Y) = \pm 1$.

Sequence of Functions

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim\limits_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Definition

We will say that f_n converges to f uniformly on X, if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ holds $|f_n(x) - f(x)| < \varepsilon$ for all $x \in X$. We will write $f_n(x) \rightrightarrows f(x), x \in X$.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Definition

We will say that f_n converges to f uniformly on X, if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ holds $|f_n(x) - f(x)| < \varepsilon$ for all $x \in X$. We will write $f_n(x) \rightrightarrows f(x), x \in X$.

Theorem

f_n converges uniformly on X if and only if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $m \in \mathbb{N}$ holds $|f_{n+m}(x) - f_n(x)| < \varepsilon$ for all $x \in X$.

Series of Functions

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *everywhere on* X*, if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *everywhere on* X*.*

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *everywhere on* X, *if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *everywhere on* X.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *uniformly on* X, *if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *uniformly on* X.

Theorem

If there exists a sequence of real numbers λ_n *such that* $|u_n(x)| \leq \lambda_n$ *for all* $x \in X$ *and* $n \in \mathbb{N}$ *and* $\sum\limits_{n=1}^{\infty} \lambda_n < +\infty$, *then* $\sum\limits_{n=1}^{\infty} u_n(x)$ *is uniformly convergent.*

Theorem

If there exists a sequence of real numbers λ_n *such that* $|u_n(x)| \leq \lambda_n$ *for all* $x \in X$ *and* $n \in \mathbb{N}$ *and* $\sum_{n=1}^{\infty} \lambda_n < +\infty$, *then* $\sum_{n=1}^{\infty} u_n(x)$ *is uniformly convergent.*

Example

Prove that the series $\sum_{n=1}^{\infty} \frac{\sin nx}{1+n^2}$ *is uniformly convergent on* $\mathbb{R}$.