Mathematics for Machine Learning

Vazgen Mikayelyan

September 15, 2020

FAST DISCOVERING THE FUTURE

The Weak Law of Large Numbers

The Weak Law of Large Numbers

Markov's theorem

Let X_n be a sequence of random variables, each having a finite mean $(\mathbb{E}[X_k] = a_k)$ and variance. If

$$\lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{k=1}^{n} X_k\right) = 0, \text{ then } \frac{1}{n}\sum_{k=1}^{n}(X_k - a_k) \xrightarrow{\mathbb{P}} 0.$$

The Weak Law of Large Numbers

Markov's theorem

Let X_n be a sequence of random variables, each having a finite mean $(\mathbb{E}[X_k] = a_k)$ and variance. If

$$\lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{k=1}^{n} X_k\right) = 0, \text{ then } \frac{1}{n}\sum_{k=1}^{n}(X_k - a_k) \xrightarrow{\mathbb{P}} 0.$$

Corollary

If X_n is a sequence of independent and identically distributed random variables, each having a finite mean $(\mathbb{E}[X_1] = a)$ and variance, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} a.$$

The Strong Law of Large Numbers

Borel's theorem

If X_n is a sequence of independent and identically distributed random variables, each having a finite moment of order 4 and $\mathbb{E}\left[X_1\right] = a$, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} a.$$

The Strong Law of Large Numbers

Borel's theorem

If X_n is a sequence of independent and identically distributed random variables, each having a finite moment of order 4 and $\mathbb{E}\left[X_1\right] = a$, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} a.$$

Kolmogorov's theorem

If X_n is a sequence of independent and identically distributed random variables, each having a finite mean $(\mathbb{E}\left[X_1\right] = a)$, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} a.$$

The Central Limit Theorem

Theorem

If X_n is a sequence of independent and identically distributed random variables, each having a finite mean μ and variance σ^2, then the CDFs of

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal distribution as $n \to \infty$.

Double Integrals

Double Integrals

Let $f : A \to \mathbb{R}$, where $A = [a, b] \times [c, d]$. Also let

$$P_1 = \{x_0, x_1, \ldots, x_n\}, P_2 = \{y_0, y_1, \ldots, y_n\}$$

are partitions of the segments $[a, b]$ and $[c, d]$ respectively.

Double Integrals

Let $f : A \to \mathbb{R}$, where $A = [a, b] \times [c, d]$. Also let

$$P_1 = \{x_0, x_1, \ldots, x_n\}, P_2 = \{y_0, y_1, \ldots, y_n\}$$

are partitions of the segments $[a, b]$ and $[c, d]$ respectively. Denote by λ the maximal length of the diagonals of the rectangles obtained by doing partitions.

Double Integrals

Let $f : A \to \mathbb{R}$, where $A = [a, b] \times [c, d]$. Also let

$$P_1 = \{x_0, x_1, \ldots, x_n\}, P_2 = \{y_0, y_1, \ldots, y_n\}$$

are partitions of the segments $[a, b]$ and $[c, d]$ respectively. Denote by λ the maximal length of the diagonals of the rectangles obtained by doing partitions.

Definition

The Riemann sum of a function $f(x, y)$ over this partition of A is

$$\sigma = \sum_{i=1}^{n} \sum_{j=1}^{n} f(u_i, v_j) \Delta x_i \Delta y_j,$$

where (u_i, v_j) is some point in the rectangle $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$ and $\Delta x_i = x_i - x_{i-1}$, $\Delta y_i = y_i - y_{i-1}$.

Double Integrals

Definition

The double integral of a function $f(x, y)$ in the rectangular region A is the following limit

$$\iint\limits_{A} f(x, y)\, dx dy = \lim_{\lambda \to 0} \sum_{i=1}^{n} \sum_{j=1}^{n} f(u_i, v_j)\, \Delta x_i \Delta y_j$$

Double Integrals

Definition

The double integral of a function $f(x,y)$ in the rectangular region A is the following limit

$$\iint\limits_{A} f(x,y)\,dxdy = \lim_{\lambda\to 0}\sum_{i=1}^{n}\sum_{j=1}^{n} f(u_i,v_j)\,\Delta x_i \Delta y_j$$

To define the double integral over a bounded region A other than a rectangle, we choose a rectangle B such that $A \subset B$ and define the function $g(x,y)$ so that

$$g(x,y) = \begin{cases} f(x,y), & \text{if } (x,y) \in A, \\ 0, & \text{if } (x,y) \in B \setminus A. \end{cases}$$

Double Integrals

Definition

The double integral of a function $f(x,y)$ in the rectangular region A is the following limit

$$\iint\limits_{A} f(x,y)\,dxdy = \lim_{\lambda \to 0} \sum_{i=1}^{n} \sum_{j=1}^{n} f(u_i, v_j)\,\Delta x_i \Delta y_j$$

To define the double integral over a bounded region A other than a rectangle, we choose a rectangle B such that $A \subset B$ and define the function $g(x,y)$ so that

$$g(x,y) = \begin{cases} f(x,y), & \text{if } (x,y) \in A, \\ 0, & \text{if } (x,y) \in B \setminus A. \end{cases}$$

We define $\iint\limits_{A} f(x,y)\,dxdy = \iint\limits_{B} g(x,y)\,dxdy.$

Double Integrals

Fubini's Theorem

If $A = [a, b] \times [c, d]$ and $f \in C(A)$ then

$$\iint\limits_A f(x, y)\, dx dy = \int\limits_a^b \left(\int\limits_c^d f(x, y)\, dy \right) dx = \int\limits_c^d \left(\int\limits_a^b f(x, y)\, dx \right) dy.$$

Double Integrals

Fubini's Theorem

If $A = [a, b] \times [c, d]$ and $f \in C(A)$ then

$$\iint\limits_A f(x, y)\, dxdy = \int\limits_a^b \left(\int\limits_c^d f(x, y)\, dy \right) dx = \int\limits_c^d \left(\int\limits_a^b f(x, y)\, dx \right) dy.$$

Change of Variable

Assume that the mapping $x = X(u, v), y = Y(u, v)$ from the domain T in the uv-plane to the domain S in the xy-plane is bijective, the functions X and Y are continuous and have continuous first order partial derivatives and the Jacobian $J(u, v)$ is never zero. Then

$$\iint\limits_S f(x, y)\, dxdy = \iint\limits_T f(X(u, v), Y(u, v))\, |J(u, v)|\, dudv.$$

Jointly Distributed Random Variables.

Jointly Distributed Random Variables.

Definition of joint CDF of 2 RVs

Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be two random variables. The function

$$F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2$$

is called **joint cumulative probability distribution function** (CDF) of X and Y

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x, y) \leq 1$ for any $(x, y) \in \mathbb{R}^2$.

Properties of Joint CDF

1. $0 \le F_{(X,Y)}(x, y) \le 1$ for any $(x, y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are increasing functions.

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x, y) \leq 1$ for any $(x, y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are increasing functions.
3. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are right continuous functions.

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x, y) \leq 1$ for any $(x, y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are increasing functions.
3. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are right continuous functions.
4. $F_{(X,Y)}(x, +\infty) = F_X(x), x \in \mathbb{R}$, where F_X is the CDF of X. Similarly, $F_{(X,Y)}(+\infty, y) = F_Y(y)$ for any $y \in \mathbb{R}$.

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x,y) \leq 1$ for any $(x,y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x,\bullet)$ and $F_{(X,Y)}(\bullet,y)$ are increasing functions.
3. $F_{(X,Y)}(x,\bullet)$ and $F_{(X,Y)}(\bullet,y)$ are right continuous functions.
4. $F_{(X,Y)}(x,+\infty) = F_X(x), x \in \mathbb{R}$, where F_X is the CDF of X. Similarly, $F_{(X,Y)}(+\infty,y) = F_Y(y)$ for any $y \in \mathbb{R}$.
5. $F_{(X,Y)}(+\infty,+\infty) = 1$,

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x, y) \leq 1$ for any $(x, y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are increasing functions.
3. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are right continuous functions.
4. $F_{(X,Y)}(x, +\infty) = F_X(x), x \in \mathbb{R}$, where F_X is the CDF of X. Similarly, $F_{(X,Y)}(+\infty, y) = F_Y(y)$ for any $y \in \mathbb{R}$.
5. $F_{(X,Y)}(+\infty, +\infty) = 1$,
6. $F_{(X,Y)}(x, -\infty) = F_{(X,Y)}(-\infty, y) = 0$ for any $x, y \in \mathbb{R}$.

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x, y) \leq 1$ for any $(x, y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are increasing functions.
3. $F_{(X,Y)}(x, \bullet)$ and $F_{(X,Y)}(\bullet, y)$ are right continuous functions.
4. $F_{(X,Y)}(x, +\infty) = F_X(x), x \in \mathbb{R}$, where F_X is the CDF of X. Similarly, $F_{(X,Y)}(+\infty, y) = F_Y(y)$ for any $y \in \mathbb{R}$.
5. $F_{(X,Y)}(+\infty, +\infty) = 1$,
6. $F_{(X,Y)}(x, -\infty) = F_{(X,Y)}(-\infty, y) = 0$ for any $x, y \in \mathbb{R}$.

Remark: The distribution functions F_X and F_Y are sometimes referred to as the **marginal distribution functions** of X and Y respectively.

Properties of Joint CDF

1. $0 \leq F_{(X,Y)}(x,y) \leq 1$ for any $(x,y) \in \mathbb{R}^2$.
2. $F_{(X,Y)}(x,\bullet)$ and $F_{(X,Y)}(\bullet,y)$ are increasing functions.
3. $F_{(X,Y)}(x,\bullet)$ and $F_{(X,Y)}(\bullet,y)$ are right continuous functions.
4. $F_{(X,Y)}(x,+\infty) = F_X(x), x \in \mathbb{R}$, where F_X is the CDF of X. Similarly, $F_{(X,Y)}(+\infty,y) = F_Y(y)$ for any $y \in \mathbb{R}$.
5. $F_{(X,Y)}(+\infty,+\infty) = 1$,
6. $F_{(X,Y)}(x,-\infty) = F_{(X,Y)}(-\infty,y) = 0$ for any $x,y \in \mathbb{R}$.

Remark: The distribution functions F_X and F_Y are sometimes referred to as the **marginal distribution functions** of X and Y respectively.

Theorem

$$P(a < X \leq b, c < Y \leq d) =$$

$$F_{(X,Y)}(a,c) + F_{(X,Y)}(b,d) - F_{(X,Y)}(a,d) - F_{(X,Y)}(b,c).$$

Joint PMF of 2 Discrete RVs

In the case when X and Y are both discrete random variables defined on the same sample space, it is convenient to define the **joint probability mass function** (joint PMF) of X and Y by the formula

$$p_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y).$$

Joint PMF of 2 Discrete RVs

In the case when X and Y are both discrete random variables defined on the same sample space, it is convenient to define the **joint probability mass function** (joint PMF) of X and Y by the formula

$$p_{(X,Y)}(x,y) = \mathbb{P}(X = x, Y = y).$$

The probability mass function of X can be obtained from the joint PMF $p_{(X,Y)}(x,y)$ by

$$p_X(x) = \sum_{y:p_{(X,Y)}(x,y)>0} p_{(X,Y)}(x,y)$$

Joint PMF of 2 Discrete RVs

In the case when X and Y are both discrete random variables defined on the same sample space, it is convenient to define the **joint probability mass function** (joint PMF) of X and Y by the formula

$$p_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y).$$

The probability mass function of X can be obtained from the joint PMF $p_{(X,Y)}(x, y)$ by

$$p_X(x) = \sum_{y: p_{(X,Y)}(x,y) > 0} p_{(X,Y)}(x, y)$$

Similarly,

$$p_Y(y) = \sum_{x: p_{(X,Y)}(x,y) > 0} p_{(X,Y)}(x, y)$$

Joint PMF of 2 Discrete RVs

In the case when X and Y are both discrete random variables defined on the same sample space, it is convenient to define the **joint probability mass function** (joint PMF) of X and Y by the formula

$$p_{(X,Y)}(x,y) = \mathbb{P}(X = x, Y = y).$$

The probability mass function of X can be obtained from the joint PMF $p_{(X,Y)}(x,y)$ by

$$p_X(x) = \sum_{y : p_{(X,Y)}(x,y) > 0} p_{(X,Y)}(x,y)$$

Similarly,

$$p_Y(y) = \sum_{x : p_{(X,Y)}(x,y) > 0} p_{(X,Y)}(x,y)$$

Remark: The functions p_X and p_Y are sometimes called the marginal PMFs of X and Y respectively.

Example

We are rolling a fair six-sided die. Define two random variables on $\Omega = \{1, 2, 3, 4, 5, 6\}$ *as*

$$X(\omega) = \begin{cases} 0, & \text{if } w = 1, \\ 1, & \text{if } w \text{ is composite} \\ 2, & \text{if } w \text{ is prime.} \end{cases}$$

$$Y(\omega) = \begin{cases} -5, & \text{if } w \leq 3, \\ 10, & \text{if } w \geq 4 \end{cases}$$

Construct the joint PMF of X *and* Y *and compute* $F_{(X,Y)}(\sqrt{2}, 7)$ *and* $\mathbb{P}(Y > 5X)$.

Jointly Distributed Random Variables.

Definition

Two random variables X and Y , defined on the same sample space, are called **jointly (absolutely) continuous** *if there exists a non-negative function f, defined on the plane, such that the joint CDF of X and Y is representable as*

$$F_{(X,Y)}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \mathrm{d}u \mathrm{d}v, \quad (x, y) \in \mathbb{R}^2.$$

The function f is called the **joint probability density function** *(PDF) of X and Y.*

Computing Marginal PDFs from the Joint PDF

Theorem

If X and Y are jointly continuous random variables then they are also individually continuous. Moreover, their (marginal) PDFs, f_X and f_Y , can be computed from the joint PDF, $f_{(X,Y)}$, as follows:

- $f_X(u) = \int\limits_{-\infty}^{+\infty} f(u,v)\mathrm{d}v$
- $f_Y(v) = \int\limits_{-\infty}^{+\infty} f(u,v)\mathrm{d}u$

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \textit{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \textit{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α*;*

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \textit{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α*;*

b. Find the marginal density functions (PDFs);

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \text{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α;

b. Find the marginal density functions (PDFs);

c. Find the joint and the marginal distribution functions (CDFs);

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \textit{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α*;*

b. Find the marginal density functions (PDFs);

c. Find the joint and the marginal distribution functions (CDFs);

d. Compute the probability $\mathbb{P}(0 < X < 1; 1 < Y < 2)$*;*

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \text{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α*;*

b. Find the marginal density functions (PDFs);

c. Find the joint and the marginal distribution functions (CDFs);

d. Compute the probability $\mathbb{P}(0 < X < 1; 1 < Y < 2)$*;*

e. Compute $\mathbb{P}(0 < Y/X < 2)$*;*

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \textit{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. Find the constant α*;*

b. Find the marginal density functions (PDFs);

c. Find the joint and the marginal distribution functions (CDFs);

d. Compute the probability $\mathbb{P}(0 < X < 1; 1 < Y < 2)$*;*

e. Compute $\mathbb{P}(0 < Y/X < 2)$*;*

f. Find the PDF of the new random variable $Z = Y/X$ *;*

Example

Assume the piecewise defined function

$$f(x,y) = \begin{cases} \alpha \cdot xy \cdot e^{-\frac{x^2+y^2}{2}}, & \text{if } x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

is a joint PDF of some random vector (X, Y).

a. *Find the constant* α;

b. *Find the marginal density functions (PDFs);*

c. *Find the joint and the marginal distribution functions (CDFs);*

d. *Compute the probability* $\mathbb{P}(0 < X < 1; 1 < Y < 2)$;

e. *Compute* $\mathbb{P}(0 < Y/X < 2)$;

f. *Find the PDF of the new random variable* $Z = Y/X$;

g. *Calculate* $\mathbb{P}(X^2 + Y^2 \leq 4)$.

Signal Processing