Advanced statistics and modelling

2020. február 19.

Bivariate distribution

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ and $\lim_{x,y\to\infty} F(x,y) = 1$.

Bivariate distribution

Sample space, Probability, Variable and Distribution

Random variable, CDF and PDF
Bivariate distrubtion
Conditional distribution
Multivariate distribution
IID variables
Transformations

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x \to -\infty} F(x,y) = \lim_{y \to -\infty} F(x,y) = 0$ and $\lim_{x,y \to \infty} F(x,y) = 1$.

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x \to -\infty} F(x,y) = \lim_{y \to -\infty} F(x,y) = 0$ and $\lim_{x,y \to \infty} F(x,y) = 1$.

Bivariate distribution

Cumulative distribution function

The joint CDF of random variables X and Y is defined as

$$F_{X,Y}(x,y) := P(X < x, Y < y).$$

Properties:

- Monotonous, non-decreasing function of its variables.
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ and $\lim_{x,y\to\infty} F(x,y) = 1$.

Joint mass function

For discrete random variables X and Y the joint mass function is defined as

$$f_{X,Y}(x,y) = P(X = x, Y = y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1, x_2]$ and $[y_1, y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1, x_2]$ and $[y_1, y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Bivariate distribution

Joint probability density function

For continuous random variables X and Y the joint probability density function is given by the function $\rho(x,y)$ connected to the joint CDF $F(x,y)$ as

$$F(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \rho(x',y') \qquad \rho(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Properties:

- Normalised:

$$\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \rho(x',y') = 1.$$

- The probability that X and Y fall into given $[x_1,x_2]$ and $[y_1,y_2]$ intervals can be written as

$$P(x_1 < X < x_2, y_1 < Y < y_2) = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy \rho(x,y).$$

Marginal distributions

Marginal CDF and probability densities

The joint distribution of X and Y uniquely determines the distribution of either X or Y, and these are called the marginals of the joint CDF:

$$F_X(x) = P(X < x, Y < \infty) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} dx' \int_{-\infty}^{\infty} dy' \rho(x', y'),$$

$$F_Y(y) = P(X < \infty, Y < y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{y} dy' \rho(x', y').$$

The marginal probability densities:

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x, y), \qquad \rho_Y(y) = \int_{\infty}^{\infty} dx \rho(x, y).$$

Marginal distributions

Marginal CDF and probability densities

The joint distribution of X and Y uniquely determines the distribution of either X or Y, and these are called the marginals of the joint CDF:

$$F_X(x) = P(X < x, Y < \infty) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} dx' \int_{-\infty}^{\infty} dy' \rho(x', y'),$$

$$F_Y(y) = P(X < \infty, Y < y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{y} dy' \rho(x', y').$$

The marginal probability densities:

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x, y), \qquad \rho_Y(y) = \int_{\infty}^{\infty} dx \rho(x, y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad \rho(x, y) = \rho_X(x)\rho_Y(y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad \rho(x, y) = \rho_X(x)\rho_Y(y).$$

Independent random variables

Independence of random variables

The random variables X and Y are independent if for any $x_1 \leq x_2$ and $y_1 \leq y_2$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2)P(y_1 \leq Y \leq y_2).$$

- For discrete variables this means that for any x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- For continuous variables this means that for any x and y

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \qquad \rho(x, y) = \rho_X(x)\rho_Y(y).$$

Conditional distributions

Conditional distribution

- The conditional CDF of a random variable X given that the variable Y takes the value $Y = y$ can be defined as

$$F_X(x \mid Y = y) := \lim_{\Delta y \to 0} P(X < x \mid y \le Y < y + \Delta y).$$

- The joint CDF of X and Y fully determines this distribution:

$$\begin{aligned} F_X(x \mid y) &= \lim_{\Delta y \to 0} P(X < x \mid y \le Y < y + \Delta y) = \\ & \lim_{\Delta y \to 0} \frac{P(X < x, y \le Y < y + \Delta y)}{P(y \le Y < y + \Delta y)} = \\ & \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \\ & \lim_{\Delta y \to 0} \frac{\frac{F(x,y+\Delta y) - F(x,y)}{\Delta y}}{\frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}} = \frac{\frac{\partial}{\partial y} F(x, y)}{\rho_Y(y)} \end{aligned}$$

Conditional distributions

Conditional distribution

- The conditional CDF of a random variable X given that the variable Y takes the value $Y = y$ can be defined as

$$F_X(x \mid Y = y) := \lim_{\Delta y \to 0} P(X < x \mid y \leq Y < y + \Delta y).$$

- The joint CDF of X and Y fully determines this distribution:

$$\begin{aligned} F_X(x \mid y) \quad = \quad & \lim_{\Delta y \to 0} P(X < x \mid y \leq Y < y + \Delta y) = \\ & \lim_{\Delta y \to 0} \frac{P(X < x, y \leq Y < y + \Delta y)}{P(y \leq Y < y + \Delta y)} = \\ & \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \\ & \lim_{\Delta y \to 0} \frac{\frac{F(x,y+\Delta y) - F(x,y)}{\Delta y}}{\frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}} = \frac{\frac{\partial}{\partial y} F(x, y)}{\rho_Y(y)} \end{aligned}$$

Conditional distribution

Conditional probability density

- By taking the partial derivative of $F_X(x \mid y)$-t with respect to x we obtain the conditional probability density of X as

$$\rho_X(x \mid y) = \frac{\partial}{\partial x} F_X(x \mid y) = \frac{\partial}{\partial x} \frac{\frac{\partial}{\partial y} F(x,y)}{\rho_Y(y)} = \frac{\rho(x,y)}{\rho_Y(y)}.$$

Similarly, the conditional CDF and density of Y given that $X = x$ can be written as

$$F_Y(y \mid X = x) = \frac{\frac{\partial}{\partial x} F(x,y)}{\rho_X(x)}, \qquad \rho_Y(y \mid x) = \frac{\rho(x,y)}{\rho_X(x)}.$$

Conditional distribution

Conditional probability density

- By taking the partial derivative of $F_X(x \mid y)$-t with respect to x we obtain the conditional probability density of X as

$$\rho_X(x \mid y) = \frac{\partial}{\partial x} F_X(x \mid y) = \frac{\partial}{\partial x} \frac{\frac{\partial}{\partial y} F(x,y)}{\rho_Y(y)} = \frac{\rho(x,y)}{\rho_Y(y)}.$$

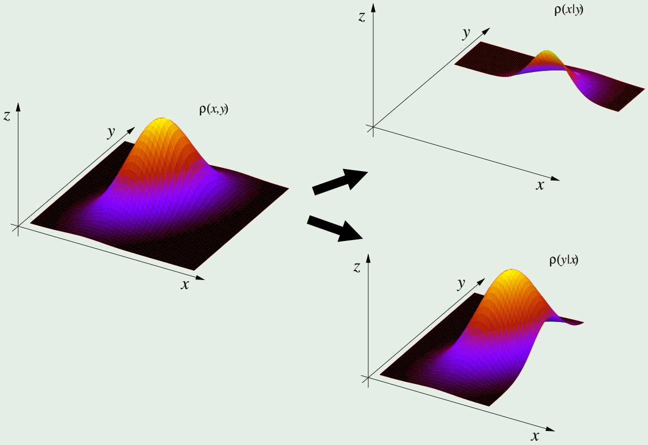
Similarly, the conditional CDF and density of Y given that $X = x$ can be written as

$$F_Y(y \mid X = x) = \frac{\frac{\partial}{\partial x} F(x,y)}{\rho_X(x)}, \qquad \rho_Y(y \mid x) = \frac{\rho(x,y)}{\rho_X(x)}.$$

Conditional probability

Illustration

Marginal density distribution $\leftrightarrow$ projection
Conditional density distribution $\leftrightarrow$ intersection



Conditional density and marginal density

- Since

$$\rho(x,y) = \rho_X(x \mid y)\rho_Y(y) = \rho_Y(y \mid x)\rho_X(x),$$

the marginals can be expressed with the help of the conditional densities as

$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x,y) = \int_{-\infty}^{\infty} dy \rho_X(x \mid y)\rho_Y(y),$$

$$\rho_Y(y) = \int_{-\infty}^{\infty} dx \rho(x,y) = \int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x).$$

(These are analogous to the law of total probability).

- An identity analogous to Bayes' theorem:

$$\rho_X(x \mid y) = \frac{\rho(x,y)}{\rho_Y(y)} = \frac{\rho_Y(y \mid x)\rho_X(x)}{\int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x)}$$

Conditional density and marginal density

- Since
$$\rho(x,y) = \rho_X(x \mid y)\rho_Y(y) = \rho_Y(y \mid x)\rho_X(x),$$
the marginals can be expressed with the help of the conditional densities as
$$\rho_X(x) = \int_{-\infty}^{\infty} dy \rho(x,y) = \int_{-\infty}^{\infty} dy \rho_X(x \mid y)\rho_Y(y),$$
$$\rho_Y(y) = \int_{-\infty}^{\infty} dx \rho(x,y) = \int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x).$$
(These are analogous to the law of total probability).
- An identity analogous to Bayes' theorem:
$$\rho_X(x \mid y) = \frac{\rho(x,y)}{\rho_Y(y)} = \frac{\rho_Y(y \mid x)\rho_X(x)}{\int_{-\infty}^{\infty} dx \rho_Y(y \mid x)\rho_X(x)}$$

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \ldots, x_n) := P(X_1 < x_1, X_1 < x_2, \ldots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \quad \lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ and $\lim_{x_1, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1.$

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \ldots, x_n) := P(X_1 < x_1, X_1 < x_2, \ldots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \quad \lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ and $\lim_{x_1, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1$.

Multivariate distributions

Multivariate CDF

The joint CDF of random variables $X_1, X_2, \dots, X_n$ is defined similarly to the bivariate case as

$$F(x_1, x_2, \dots, x_n) := P(X_1 < x_1, X_1 < x_2, \dots, X_n < x_n).$$

Properties:

- It is a monotonous non-decreasing function of its variables.
- $\forall i \in [1, n] \quad \lim_{x_i \to -\infty} F(x_1, \dots, x_i, \dots, x_n) = 0$ and $\lim_{x_1, \dots, x_n \to \infty} F(x_1, x_2, \dots, x_n) = 1.$

Multivariate distributions

Multivariate PDF

The joint PDF of continuous random variables $X_1, X_2, ..., X_n$ denoted by $\rho(x_1, x_2, ..., x_n)$ fulfils

$$F(x_1, x_2, ..., x_n) = \int_{-\infty}^{x_1} dx_1' \int_{-\infty}^{x_2} dx_2' \cdots \int_{-\infty}^{x_n} dx_n' \rho(x_1', x_2', ..., x_n')$$

$$\rho(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

Properties:

- Normalised

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \rho(x_1, x_2, ..., x_n) = 1.$$

- The probability that $(X_1, X_2, ..., X_n)$ falls in some region E is

$$P((X_1, X_2, ..., X_n) \in E) = \iint \cdots \int_E dx_1 dx_2 ... dx_n \rho(x_1, x_2, ..., x_n).$$

Multivariate distributions

Multivariate PDF

The joint PDF of continuous random variables $X_1, X_2, \ldots, X_n$ denoted by $\rho(x_1, x_2, \ldots, x_n)$ fulfils

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} dx_1' \int_{-\infty}^{x_2} dx_2' \cdots \int_{-\infty}^{x_n} dx_n' \rho(x_1', x_2', \ldots, x_n')$$

$$\rho(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

Properties:

- Normalised

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \rho(x_1, x_2, \ldots, x_n) = 1.$$

- The probability that $(X_1, X_2, \ldots, X_n)$ falls in some region E is

$$P\left((X_1, X_2, \ldots, X_n) \in E\right) = \iint \cdots \int_E dx_1 dx_2 \ldots dx_n \rho(x_1, x_2, \ldots, x_n).$$

IID variables

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \ldots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \ldots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

- This is equivalent to

$$\rho(x_1, x_2, \ldots, x_n) = \rho_{X_1}(x_1)\rho_{X_2}(x_2)\cdots\rho_{X_n}(x_n)$$

IID variables

Independent and Identically distributed variables

- The independence of random variables $X_1, X_2, \ldots, X_n$ is defined similarly to the bivariate case, i.e., they are independent if for any $A_1, A_2, \ldots, A_n$ where $A_i \subset \mathbb{R}$,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

- This is equivalent to

$$\rho(x_1, x_2, \ldots, x_n) = \rho_{X_1}(x_1)\rho_{X_2}(x_2)\cdots\rho_{X_n}(x_n)$$

- If in addition all X_i have the same marginal distribution, then we call them independent and identically distributed (IID). This means that they correspond to independent draws from the same distribution.

Important multivariate distributions

Multinomial distribution

- This is the **multivariate** version of the **binomial distribution**.

  Let us consider N draws with replacement from an urn with balls of q different colours, where the probability of drawing colour i is p_i. Let $X = (X_1, X_2, \dots, X_q)$ count the number of drawn balls with the different colours. The probability mass function can be written as

$$f_X(k_1, k_2, \dots, k_q) = \frac{N!}{k_1!k_2!\cdots k_q!} p_1^{k_1} p_2^{k_2} \cdots p_q^{k_q}.$$

- The marginal distributions of $X = (X_1, X_2, \dots, X_q)$ ~Multinomial(N, p) where $p = (p_1, p_2, \dots, p_q)$ are given by X_i ~Binom(N, p_i).

Important multivariate distributions

Multinomial distribution

- This is the **multivariate** version of the **binomial distribution**.

  Let us consider N draws with replacement from an urn with balls of q different colours, where the probability of drawing colour i is p_i. Let $X = (X_1, X_2, \ldots, X_q)$ count the number of drawn balls with the different colours. The probability mass function can be written as

$$f_X(k_1, k_2, \ldots, k_q) = \frac{N!}{k_1!k_2!\cdots k_q!} p_1^{k_1} p_2^{k_2} \cdots p_q^{k_q}.$$

- The marginal distributions of $X = (X_1, X_2, \ldots, X_q)$ ~Multinomial(N, p) where $p = (p_1, p_2, \ldots, p_q)$ are given by X_i ~Binom(N, p_i).

Important multivariate distributions

Multivariate normal distribution

The **multivariate Normal distribution** is parametrised by the vector of expected values $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and the covariance matrix Σ. A random vector $X = (X_1, X_2, \ldots, X_n)$ has multivariate Normal distribution if the joint density distribution can be written as

$$\rho(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})[\Sigma]^{-1}(\vec{x}-\vec{\mu})} = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}\sum_{ij}\left[\Sigma^{-1}\right]_{ij}(x_i-\mu_i)(x_j-\mu_j)}$$

Transformation of random variables

Transformation of random variables

Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$

$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$

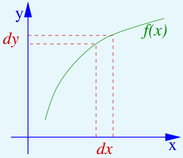
Transformation of random variables

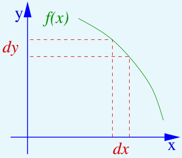
Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$





$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

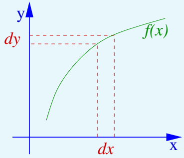
$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$

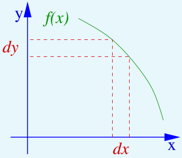
Transformation of variables

- If $Y = f(X)$, then how can we formulate the PDF of Y based on $\rho_X(x)$?

$$y = f(x) \;\rightarrow\; dy = f'(x)dx$$

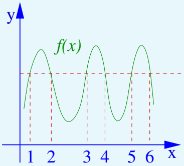
$$\rho_Y(y)\,|dy| = \rho_X(x)\,|dx|$$





$$\rho_Y(y) = \rho_X(x)\frac{|dx|}{|dy|} = \rho_X(f^{-1}(y))\frac{1}{|f'(f^{-1}(y))|}$$

$$\rho_Y(y) = \sum_n \rho_X(f_n^{-1}(y))\frac{1}{|f'(f_n^{-1}(y))|}$$



Transformations of random variables

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \dots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \dots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \dots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \ f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \dots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$\begin{aligned} F_{\vec{Y}}(\vec{y}) &= P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \\ &\int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_n. \end{aligned}$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformations of random variables

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \; f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$\begin{aligned} F_{\vec{Y}}(\vec{y}) &= P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \\ & \int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n. \end{aligned}$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformation of multivariate distributions

How to obtain the distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g., $\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \; f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$\begin{aligned} F_{\vec{Y}}(\vec{y}) \quad &= \quad P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \\ & \int\int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n. \end{aligned}$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformation of multivariate distributions

How to obtain the distribution of
$\vec{Y} = (Y_1, Y_2, \ldots, Y_q) = \vec{f}(\vec{X}) = \vec{f}(X_1, X_2, \ldots, X_k)$ based on e.g.,
$\rho_{\vec{X}}(x_1, x_2, \ldots, x_k)$?

The general recipe:

- For each $\vec{y}$ find the set

$$A_{\vec{y}} = \{\vec{x} : \ f_1(\vec{x}) < y_1, f_2(\vec{x}) < y_2, \ldots, f_n(\vec{x}) < y_n\}.$$

- Based on that, the CDF can be written as

$$\begin{aligned} F_{\vec{Y}}(\vec{y}) &= P(\vec{Y} < \vec{y}) = P(\vec{f}(\vec{x}) < \vec{y}) = P\left(\{\vec{x} : \vec{f}(\vec{x}) < \vec{y}\}\right) = \\ & \int \int \cdots \int_{A_{\vec{y}}} \rho_{\vec{X}}(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_n. \end{aligned}$$

- The PDF is

$$\rho_{\vec{Y}}(\vec{y}) = F'_{\vec{Y}}(\vec{y}).$$

Transformations of random variables

- In the special case of $\vec{Y} = \vec{f}(\vec{X})$ where both Y and X have dimension n and f is invertible:

$$\vec{y} = \vec{f}(\vec{x}), \quad \rightarrow \quad d\vec{y} = D\vec{f}(\vec{x})d\vec{x}, \quad \rho_Y(\vec{y})\,|d\vec{y}| = \rho_X(\vec{x})\,|d\vec{x}|\,,$$

$$\rho_Y(\vec{y}) = \rho_X(\vec{x})\frac{|d\vec{x}|}{|d\vec{y}|} = \rho_X(\vec{f}^{-1}(\vec{y}))\left|\det \boldsymbol{J}\left[\vec{f}^{-1}(\vec{y})\right]\right|,$$

where the determinant of the Jacobi matrix is

$$\det \boldsymbol{J}\left[\vec{g}(\vec{x})\right] = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{pmatrix}$$

EXPECTATION, INEQUALITIES, CONVERGENCE

EXPECTATION AND VARIANCE

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as

$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$

- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then

$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^{n} a_i Z_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$

- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as

$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$

- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then

$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^{n} a_i Z_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$

- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as
$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$
- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then
$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^n a_i Z_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$
- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then
$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as
$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$
- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then
$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^{n} a_i Z_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$
- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then
$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i-\mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\text{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x) dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX + b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Mean and variance of important distributions

Distribution	Mean	Variance
Bernoulli(p)	p	$p(1-p)$
Binomial(N,p)	Np	$Np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a,b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ,σ^2)	μ	σ^2
Exponential(λ)	λ	λ^2
Gamma(q,λ)	$q\lambda$	$q\lambda^2$
χ^2_n	n	$2n$
t_n	0 if $n>1$	$n/(n-2)$ if $n>2$

Sample mean and sample variance

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned}\mathbb{E}(\overline{X}_n) &= \mu,\\ \mathbb{V}(\overline{X}_n) &= \frac{\sigma^2}{n},\\ \mathbb{E}(S_n^2) &= \sigma^2.\end{aligned}$$

Sample mean and sample variance

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned} \mathbb{E}(\overline{X_n}) &= \mu, \\ \mathbb{V}(\overline{X_n}) &= \frac{\sigma^2}{n}, \\ \mathbb{E}(S_n^2) &= \sigma^2. \end{aligned}$$

Sample mean and sample variance

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned} \mathbb{E}(\overline{X_n}) &= \mu, \\ \mathbb{V}(\overline{X_n}) &= \frac{\sigma^2}{n}, \\ \mathbb{E}(S_n^2) &= \sigma^2. \end{aligned}$$