Introduction to Probability Theory

Daria Lavrentev

University of Freiburg

October 16, 2014

Main reference:
Mathematical Statistics with Applications by John E. Freund.
Available in the seminar library **S1/1427**

Basics

We would often be interested in studying a phenomenon, where there is an uncertainty about the outcome. A random phenomenon is also refered to as an experiment.

- A set of all possible outcomes of an experiment is called the **sample space** and is denoted by Ω (or sometimes S).

One can be interested in results which are not given directly by a specific element of a sample spase.

- An **event** is a subset of the sumple space.

Example

Rolling a dice.
The sample space is: $\{1, 2, 3, 4, 5, 6\}$.
$\{1, 2\}$ represents an event that rolling a dice resulted in 1 or 2.

Basics

We denote events by capital letters.
The probability of an event A is denoted by $Prob(A)$ (or often by $P(A)$).

Some useful properties:

- $P(\Omega) = 1$
- $P(A') = 1 - P(A)$
- $P(\emptyset) = 0$
- Given A and B two events, such that $A \subseteq B$, it holds that $P(A) \leq P(B)$.
- $P(A) \in [0, 1]$
- If A and B are mutualy exclusive events ($A \cap B = \emptyset$), then $Prob(A \cup B) = P(A) + P(B)$

Basics

The **conditional probability** of event A given event B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Two events A and B are said to be **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Random Variable

A **random variable** is a function that associates a unique numerical value with outcome of a random phenomenon/an experiment.
One distinguishes between two types of random variables: discrete and continuous.

- **Discrete** random variable is a random variable that can obtain a finite, or countable infinite number of values. Note that since the number of values is countable, we can denote them by :$x_1, x_2, ..., x_n, ...$
- **Continuous** random variable is a random variable that can obtain any value in $\mathbb{R}$.

Remark: the definition of the continious r.v. is often different.

Random Variable

Examples

- Rolling a dice.
- Number of mistakes on a page of a book.
- Waiting time in a queue at a service point.
- Weight of a child of a certain age.
- Stock price.

p.d.f and c.d.f

- The **probability distribution function** of a discrete random variable, $f(x)$ describes the probability of the variable to obtain a certain value x,

$$f(x) = Prob(X = x)$$

- **Cumulative destribution function** is defined as the probability of the given variable X to obtain a value equal or smaller than x,

$$F(x) = Prob(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

p.d.f and c.d.f

Properties:

- $f(x) \geq 0$
- $\sum_{i=1}^{\infty} f(x_i) = 1$

Example

Consieder a random variable that describes the outcome when rolling a fair dice once. The possible values are $\{1, 2, 3, 4, 5, 6\}$. Since the dice is fair, there is the same probability of all the outcomes, that is

$$f(x) = \begin{cases} \frac{1}{6} & x = 1, 2, 3, 4, 5, 6 \\ 0 & otherwise. \end{cases}$$

p.d.f and c.d.f

Example

The cumulative distribution finction in this case is:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6}, & x \in [1, 2) \\ \frac{2}{6}, & x \in [2, 3) \\ \frac{3}{6}, & x \in [3, 4) \\ \frac{4}{6}, & x \in [4, 5) \\ \frac{5}{6}, & x \in [5, 6) \\ 1 & x \geq 6 \end{cases}$$

Expected Value

For a discrete random variable X with density function $f(x)$ the expected value is defined as:

$$\mathbb{E}(X) = \mu_X = \sum_{i=1}^{\infty} x_i \cdot f(x_i)$$

The following terms often refer to the same thing!

Expected Value = Mean = Expectations = Average

Example

Continuing the previous example,

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \cdot f(x_i) =$$

$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} + 7 \cdot 0 + \ldots$$

$$= \frac{21}{6} = 3.5$$

Variation and Standard Deviation

The **variance** of a discrete random variable X is defined as:

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{\infty}(x_i - \mu_X)^2 \cdot f(x_i).$$

Standard deviation is: $\sigma_X = \sqrt{Var(X)}$

Example

For the example discussed above we have:

$Var(X) = \sum_{i=1}^{\infty}(x_i - \mu_X)^2 \cdot f(x_i) =$

$(1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} +$

$+(5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} + (7-3.5)^2 \cdot 0 + ... = \frac{17.5}{6}$

$\Longrightarrow \sigma \approx 1.7078$

Moment of Order r

For a random variable X, it's moment of order r is defined as:

$$M_X^r = \mathbb{E}(X^r) = \sum_{i=1}^{\infty} x_i^r \cdot f(x_i),$$

sometimes one talks about the central moment, also refered to as the moment about the mean, that is:

$$M_X^r = \mathbb{E}[(X - \mu_X)^r] = \sum_{i=1}^{\infty} (x_i - \mu_X)^r \cdot f(x_i)$$

Example

σ_x^2 is the central moment of X.

p.d.f and c.d.f

- A function $f(x)$ defined on the set of real numbers is called a **probability density function** of a continious random variable X if and only if

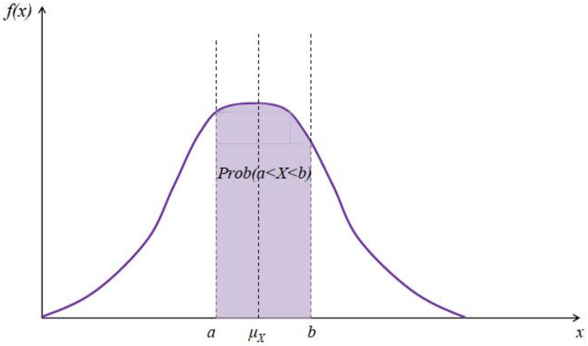
$$Prob(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

for any $a, b \in \mathbb{R}$, such that $a \leq b$.

- **Cumulative destribution functio**n is defined as the probability of the given variable X to obtain a value equal or smaller than x,

$$F(x) = Prob(X \leq x) = \int_{-\infty}^{x} f(z)dz$$

p.d.f and c.d.f



p.d.f and c.d.f

- Note the difference to the discrete case!
- A function $f(x)$ can serve as a probability density function only if it has the following properties:
  1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}$,
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
- $P(a \leq X \leq b) = F(b) - F(a)$
- If c.d.f. of a random variable X is differentiable, then

$$f(x) = \frac{dF(x)}{dx}.$$

p.d.f and c.d.f

Example

A random variable X follows the exponential distribution with parameter λ if it has the folowing c.d.f. and p.d.f.

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

and

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

We denote $X \sim exp(\lambda)$

Expected Value

- For a continuous random variable X with density function $f(x)$ the expected value is defined as:

$$\mathbb{E}(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- Let Y be a randon variable defined by $Y = g(X)$, where $g(*)$ is a real function of single variable. The expected value of Y is:

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$$

Variation and Standard Deviation

The variance of a continuous random variable is:

$$Var(X) = \sigma^2 = \mathbb{E}((X - \mu_X)^2)$$

Note that

$$Var(X) = \mathbb{E}((X - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)dx =$$

$$= \int_{-\infty}^{\infty} (x^2 - 2x\mu_X + \mu_X^2) \cdot f(x)dx =$$

$$= \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \int_{-\infty}^{\infty} 2x\mu_X \cdot f(x)dx + \int_{-\infty}^{\infty} \mu_X^2 \cdot f(x)dx =$$

$$= \mathbb{E}(X^2) - \mu_X^2$$

Example

Consieder $X \sim exp(\lambda)$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{0}^{\infty} x\lambda e^{-\lambda x}dx = \frac{1}{\lambda}\int_{0}^{\infty} ye^{-y}dy =$$

$$= \frac{1}{\lambda}(-ye^{-y}|_0^\infty + \int_0^\infty e^{-y}dy) = \frac{1}{\lambda}(0 - e^{-y}|_0^\infty) = \frac{1}{\lambda}(0-0+1) = \frac{1}{\lambda}$$

where a substitution of variable $y = \lambda x$ is performed first, and the resulting integral is calculated using integration by parts.

Exercise

Show that $Var(X) = \frac{1}{\lambda^2}$

Moment of Order r

Similar to the discrete case, a moment of order r is defined as:

$$M_X^r = \mathbb{E}(X^r)$$

The moment about the mean is:

$$M_X^r = \mathbb{E}[(X - \mu_X)^r]$$

Example

σ_x^2 is the second moment around the mean.

Important Discrete Distributions

- **Bernoulli** distribution
  A single toss of a coin with probability p of obtaining heads. Let X be a random variable obtaining 1 if the result is heads and 0 if the result is tails.

$$f(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

$$\mathbb{E}(X) = p,$$

$$Var(X) = p(1 - p)$$

Important Discrete Distributions

- **Binomial** distribution
  Represents the number of successes in a sequence of n independent yes/no experiments, each of which yields success(yes) with probability p.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, ...$$

$$\mathbb{E}(X) = np,$$

$$Var(X) = np(1-p)$$

Important Continuous Distributions

- **Uniform** distribution
  X is uniform distributed on interval (a, b) it it's density function is:

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

$$\mathbb{E}(X) = \frac{a+b}{2},$$

$$Var(x) = \frac{(b-a)^2}{12}$$

Important Continuous Distributions

- **Normal** distribution with mean μ and variance σ^2

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

If $\sigma^2 = 1$ and $\mu = 0$ we say that the distribution is a **standard normal** dositribution.

- **Exponential** distribution

Joint Probability and Density

- Given two discrete r.v. X and Y, the function given by

$$f(x,y) = Prob(X = x, Y = y)$$

is called the **joint probability distribution function** of X and Y.

- A bivariate function $f(x,y)$ is called a **joint probability density function** of the continuous r.v. X and Y if and only if

$$Prob[(X,Y) \in A] = \int\int_A f(x,y)dxdy$$

for any region A in the xy plane.

Joint Distribution

- Given two discrete r.v. X and Y, the function

$$F(x, y) = Prob(X \leq x, Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} f(s, t)$$

where $(x, y) \in \mathbb{R}^2$ and $f(s, t)$ is the **joint probability density**, is the **joint distribution function**, or the **joint cumulative distribution**, of X and Y.

- Given two continuous r.v. X and Y, the function

$$F(x, y) = Prob(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) ds dt$$

where $(x, y) \in \mathbb{R}^2$ and $f(s, t)$ is the joint probability density, is called the **joint distribution function** of X and Y.

Marginal Distribution and Density.

- Given two discrete r.v. X and Y with joint probability distribution $f(x, y)$, the function

$$g(x) = \sum_{y} f(x, y)$$

is called the **marginal distribution** of X.
In a similar way, $h(y) = \sum_{x} f(x, y)$ is the marginal distribution of Y.

- Given two continuous r.v. X and Y,with joint dinsity $f(x, y)$, the function

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

is called the **marginal density** of X.
In a similar way, $h(y) = \int_{-\infty}^{\infty} f(x, y)dy$ is the marginal density of Y.

Independent Random Variables

Given two random variables X_1 and X_2 with joint distribution/density function $f(x_1, x_2)$ and distribution/density functions $f_1(x_1)$ and $f_2(x_2)$, we say that the variables X_1 and X_2 are **independent** if and only if

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2).$$

If the above does not call, we say that the two random variablles are **dependent** or **correlated**.

Dependent Random Variables

- **Covariance** between two random variables X and Y measures how much the two random variables change together. It is defined as

$$\mathbb{C}ov(X, Y) = \sigma_{X,Y} = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]$$

- **Correlation coefficient** between two random variables X and Y, denoted by $\rho_{X,Y}$, is defined as:

$$\rho_{X,Y} = \frac{\mathbb{C}ov(X, Y)}{\sigma_X \sigma_Y}$$

Note that this definition applies to both discrete and continuous variables.

Useful Properties

- $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[X^2] = Var[X] + \mathbb{E}^2[X]$
- If X and Y are independent r.v. then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
- $Var[\alpha X] = \alpha^2 Var[X]$
- If X and Y are independent r.v. then $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y]$
- If X and Y are dependent r.v. then $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2ab\mathbb{C}ov(X, Y)$

where $\alpha, a, b \in \mathbb{R}$, X and Y are random variables.

Sample Mean, Variance and Covariance

Given a random sample $x_1, x_2, x_3, ..., x_N$ we can find the mean as

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

The sample variance is

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N}$$

The sample covariance of two random samples $x_1, x_2, x_3, ..., x_N$ and $y_1, y_2, y_3, ..., y_N$ is

$$\mathbb{C}ov(X, Y) = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{N}$$

Remark: for variance and covariance, sometimes N will be replaced with $N - 1$.

Var-Covariance Matrix

Given N observations for each of M random varibles,which are given in a matrix form X, where each column of the matrix gives the N observations of a single variable, the variance-covariance matrix is the following matrix,

$$\Sigma = \left(\sigma_{ij}\right)_{i,j=1\ldots M}, \text{ where } \sigma_{ij} = \mathbb{C}ov(X_i, X_j)$$

We note that this matrix is symetric and that the diagonal elements are the sample variance of the variables.

Var-Covariance Matrix

Important Example

Assuming that the variavles represented by the matrix X have a sample variance of zero, the var-covar. matrix can be calculated as

$$\frac{1}{M} X^\top X$$

Important Application

Consider an economy with two periods, today ($t = 0$) and tomorrow (t=1). There is uncertainty about the economy state tomorrow. There are s possible states. The probabilities of each state to occur is given by $(\pi_1, \pi_2, ..., \pi_s)$. In addition we know that the state contingent consumption of an agent in this economy is given by $(c_1, c_2, ..., c_s)$ and that the agent's utility of consuming c units is equal to $u(c) = \sqrt{c}$.
The expected utility of the agent in this economy is:

$$E(u(c)) = \sum_{i=1}^{s} \pi_i u(c_i) = \sum_{i=1}^{s} \pi_i \sqrt{c_i}$$