Statistics

2023 Lectures
Part 2 - Random Variables

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

A numerical function of the outcomes

- In most cases, an experimenter focuses on some specific characteristics of the experiment
- E.g., a traffic engineer may focus on the number of vehicles traveling on a certain road or in a certain direction rather then on the brand of vehicles or number of passengers in each vehicle
- Each outcome of the experiment can be associated with a number by specifying a rule of association
- Passing from the experimental outcomes to a numerical function of the outcomes is allowed by the concept of a random variable, a useful tools for describing events (and much more).

Random variable

- A general idea: random variable is a number depending on chance
- Aim: to define random variable associated with the studied phenomenon rather than with a specific sample space

Definition 1: A real-valued function $X[(S, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}^1)]$ is a random variable, provided it is measurable, i.e.

$$B \in \mathcal{B}^1 \Rightarrow X^{-1}(B) \in \mathcal{A}.$$

- Hence, each $\{s \in S | X(s) \leq t\}$ or for brevity just $\{X \leq t\}$ is an event for each $t \in \mathbb{R}$.

Invariance of random variables

Definition 2: X, X' random variables, defined on sample spaces S, S', respectively, describing the same phenomenon, are equivalent if for every t we have $\{X \leq t\}$ occurs if and only if $\{X' \leq t\}$ occurs and these events have the same probability.

- the above equivalence satisfies the requirements for relation of equivalence - reflexivity, symmetry and transitivity

Example 1: In the tossing of three fair coins, let r.v. X be defined as the number of tails. Then X has only values $0, 1, 2, 3$. We can associate with these values probabilities in the following way:

$P(X = 0) = P(\{H, H, H\}) = 1/8;$
$P(X = 1) = P(\{H, H, T\} \cup \{H, T, H\} \cup \{T, H, H\}) = 3/8;$
$P(X = 2) = P(\{H, T, T\} \cup \{T, H, T\} \cup \{T, T, H\}) = 3/8;$
$P(X = 3) = P(\{T, T, T\}) = 1/8.$

Distribution of random variable

Definition 3: By distribution of the random variable X we mean the assignment of probabilities to all events $\{X \in A\} \subset S$, where $A \subset \mathbb{R}$.

- Basic type of events are given by intervals

$$\{a < X < b\}, \{a \leq X < b\}, \{a < X \leq b\}, \{a \leq X \leq b\},$$

for $-\infty \leq a \leq b \leq \infty$.

Discrete distribution

Definition 4: A random variable X is called discrete if there is a finite or countable set of real numbers $U = \{x_1, x_2, \dots\}$ such that

$$P(X \in U) = \sum_n P(X = x_n) = 1.$$

Experiment	Random Variable (x)	Possible Values for the Random Variable
Contact five customers	Number of customers who place an order	0, 1, 2, 3, 4, 5
Inspect a shipment of 50 radios	Number of defective radios	0, 1, 2, $\cdots$, 49, 50
Operate a restaurant for one day	Number of customers	0, 1, 2, 3, $\cdots$
Sell an automobile	Gender of the customer	0 if male; 1 if female

Definition 5: The probability mass function (or discrete density function) of a discrete random variable X is the function

$$p_X(k) = \begin{cases} P(X = k) & \text{if } k \in U; \\ 0 & \text{else.} \end{cases}$$

Uniform distribution

- some distributions with certain characteristics have special names
- there are many classes of (parametric) distributions of random variables

Example 2: $U = \{x_1, \ldots, x_n\}, P(X = x_i) = 1/n, i = 1, \ldots, n.$
E.g., the selection (assumed fair) of the number of the winning lottery ticket, where n is the total number of tickets.

Such distribution of random variable X is called discrete uniform distribution.

Binomial distribution

Example 3:

$U = \{0, 1, \ldots, n\}, P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, \ldots, n$

Random variable with this distribution is often defined as total number of successes in n independent experiments, each with probability p

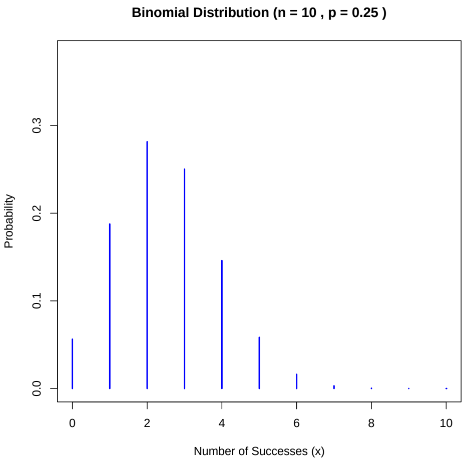
The probability of k successes and $n - k$ failures in any specific order equals $p^k(1-p)^{n-k}$. There is $\binom{n}{k}$ of such different orders. Clearly,

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Such distribution is called binomial
We write $X \sim BIN(n, p)$.

Binomial distribution: Example

Example 4: Consider $X \sim BIN(10, 0.25)$.
The graph of the probability mass function



More examples of discrete distribution

Example 5:
$U = \{0, 1, 2, \dots\}, P(X = k) = p(1-p)^k, k = 0, 1, 2 \dots$
X is the number of trials before the first success in series

$$\sum_{k=0}^{\infty} P(X = k) = p \sum_{k=0}^{\infty} (1-p)^k = p\frac{1}{p} = 1.$$

Such distribution is called geometric and we write $X \sim GEO(p)$.

Example 6: $U = \{0, 1, 2, \dots\}, P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots$
Such distribution is called Poisson

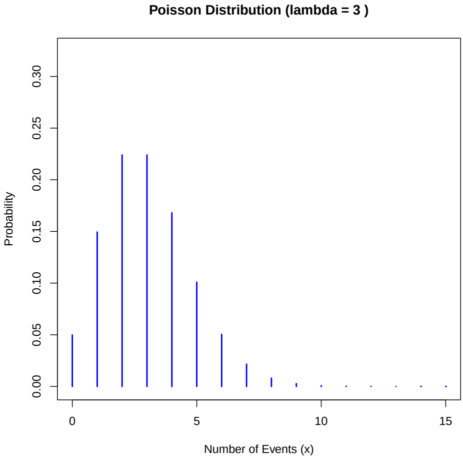
$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

We write $X \sim POI(\lambda)$.

Poisson distribution: Example

Example 7: Consider $X \sim POI(3)$.

The graph of the probability mass function



Random variable and events

Theorem 1: The probabilities of the events of the form $\{a < X \leq b\}$ for all $-\infty \leq a \leq b \leq \infty$ uniquely determine the probabilities of events of the form $\{a < X < b\}$, $\{a \leq X < b\}$ and $\{a \leq X \leq b\}$.

- Moreover $\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$ and so

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a).$$

- Thus probabilities of events $\{a < X \leq b\}$ are determined by probabilities of events $\{X \leq t\}$ for $-\infty < t < \infty$.

Distribution function

Definition 6: For any random variable X, the (cumulative) distribution function (cdf) F_X is defined as

$$F_X(t) = P(X \leq t), \quad t \in \mathbb{R}.$$

Theorem 2: For any random variable X, cdf F_X has the following properties:

a) F_X is nondecreasing;
b) $\lim_{t\to-\infty} F_X(t) = 0,\ \lim_{t\to\infty} F_X(t) = 1;$
c) F_X is continuous on the right.

- As a consequence $P(X = t) = F_X(t) - F_X(t)_-$. So if F_X is continuous at t, then $P(X = t) = 0$.
- For discrete random variables

$$F_X(t) = P(X \leq t) = \sum_{x_i \leq t} P(X = x_i).$$

Properties of cdf: Example

Example 1 (cont.): Recall, X is the number of tails in three tosses of a regular coin. In such simple cases we can represent the distribution by a table

values	0	1	2	3
probabilities	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

We have $0 \leq X \leq 3$ and so $F_X(t) = 0$ for $t < 0$ and $F_X(t) = 1$ for $t \geq 3$.

For $t \in [0, 1)$

$$F_X(t) = P(X = 0) = 1/8$$

For $t \in [1, 2)$

$$F_X(t) = P(X = 0) + P(X = 1) = 1/2$$

For $t \in [2, 3)$

$$F_X(t) = P(X = 0) + P(X = 1) + P(X = 2) = 7/8$$

Remark: Cdf is defined for all real arguments, not only for the values of the random variable. E.g., one can find $F_X(2.27)$.

Continuous random variable

Definition 7: The random variable X is called continuous if there exists a nonnegative function f, called the density of X, such that

$$F_X(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x, \quad -\infty < t < \infty.$$

Theorem 2c: For a continuous random variable X,

a) F_X is nondecreasing;

b) $\lim_{t\to-\infty} F_X(t) = 0,\ \lim_{t\to\infty} F_X(t) = 1;$

c) F_X is continuous.

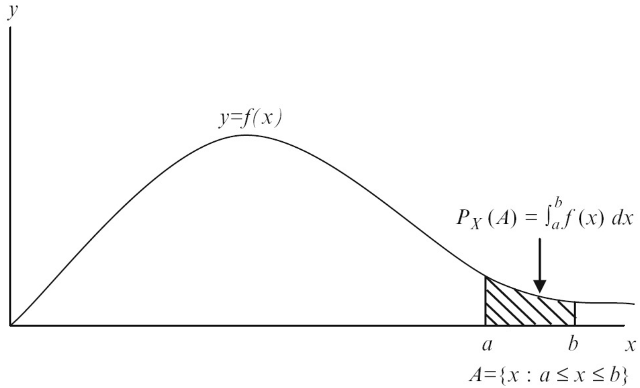
- As a consequence $P(X = t) = 0$ for any $t \in \mathbb{R}$.
- $f(t) = F'(t)$ and $f(t) \geq 0$ almost everywhere

Properties of cdf for continuous distribution

Theorem 3: If X has density f then for all $a < b$ we have

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) =$$

$$= F(b) - F(a) = \int_a^b f(x)\mathrm{d}x.$$



Uniform distribution

Example 8: Let $a < b$ and

$$f(x) = \begin{cases} 0, & \text{if } X < a; \\ c, & \text{if } a \le X \le b; \\ 0, & \text{if } X > b. \end{cases}$$

Then

$$F_X(t) = \begin{cases} 0, & \text{if } t < a; \\ \int_a^t f(x)\mathrm{d}x = c(t-a), & \text{if } a \le t \le b; \\ c(b-a), & \text{if } t > b. \end{cases}$$

Hence $c := \frac{1}{b-a}$ and distribution of X is called uniform on $[a,b]$.
We write $X \sim U[a,b]$.

Exponential distribution

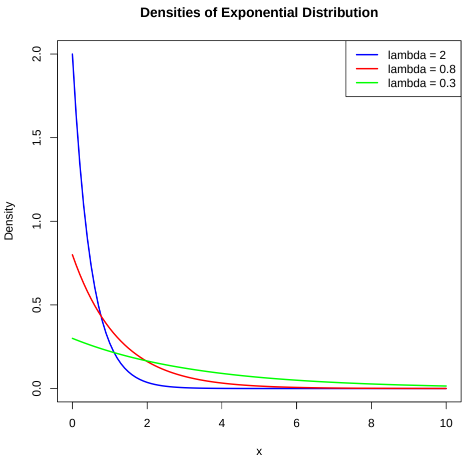
Example 9: The distribution with the density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0, \end{cases}$$

is called exponential with parameter $\lambda > 0$.
We write $X \sim EXP(\lambda)$.

Exponential distribution: Example

Example 10: Consider $X \sim EXP(\lambda)$ for values $\lambda = 0.3, 0.8, 2$.
The graph of the density functions

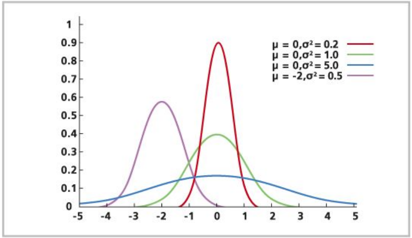


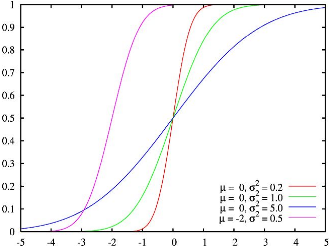
Normal distribution

Example 11: The distribution with the density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \ x \in \mathbb{R},$$

is called normal with parameters μ and σ. We write $X \sim N(\mu, \sigma^2)$.





Standard normal distribution

For any $a < b$

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} \mathrm{d}z = \Phi(z_2) - \Phi(z_1),$$

where $z_1 = \frac{a-\mu}{\sigma}, z_2 = \frac{b-\mu}{\sigma}$ and Φ is the cumulative distribution function of the standard normal distribution denoted as $N(0, 1)$.

Moreover,

$$\Phi(-z) = 1 - \Phi(z).$$

Quantiles

Definition 8: Let X be a random variable with cdf F_X, and let $0 < p < 1$. The pth (lower) quantile ξ_p of X is defined as a minimum of solutions of inequalities

$$P(X \leq x) \geq p, \;\; P(X \geq x) \geq 1 - p.$$

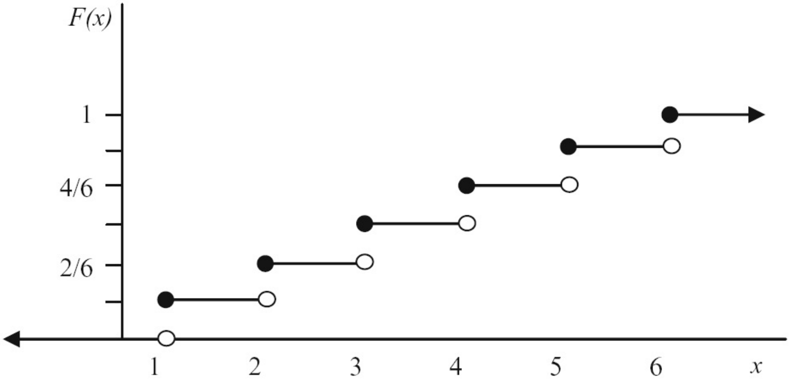
- Alternatively, the pth quantile ξ_p has to satisfy

$$F_X(\xi_p) \geq p \text{ and } F_X(\xi_p)_- \leq p.$$

- for $p = 0.25$, ξ_p is called the lower quartile
- for $p = 0.5$, ξ_p is called the median
- for $p = 0.75$, ξ_p is called the upper quartile
- alternatively, pth upper quantile is the $(1 - p)$th lower quantile

Quantiles: Example

Example 12: Consider rolling of a die.
The graph of the cdf:



Find median and upper and lower quartiles.

More properties of continuous random variables

- Let X be a continuous random variable. Then the pth quantile ξ_p of X is defined as a minimum value of $x \in \mathbb{R}$ satisfying
$$P(X \leq x) = p, 0 < p < 1.$$

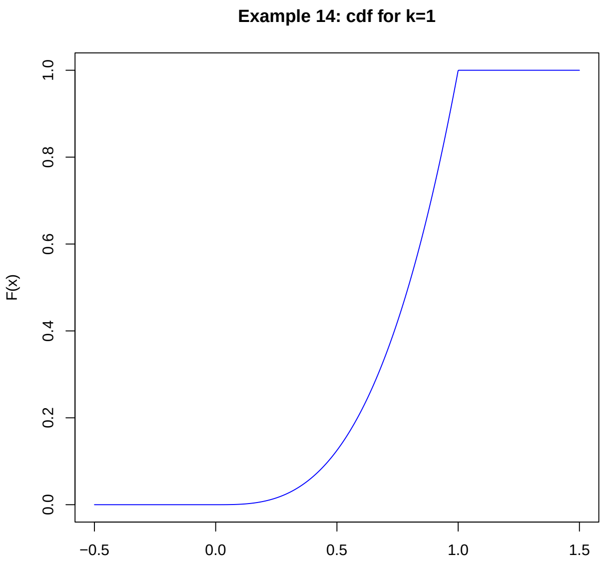
Theorem 4: Let X be a random variable with continuous cdf F_X. Then
$$P(\xi_a \leq X \leq \xi_b) = b - a \quad \text{for } 0 < a < b < 1.$$

Example 13: If F is continuous, there is always 50% probability that a random variable with cdf F will assume a value between its upper and lower quartile.

Example 14: Determine the median, lower and upper quartiles for random variable X with the following cdf, $k > 0$:

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0; \\ kx^3, & \text{for } 0 \leq x \leq \frac{1}{\sqrt[3]{k}}; \\ 1, & \text{for } x > \frac{1}{\sqrt[3]{k}}. \end{cases}$$

Example 14



More properties of cdfs

Theorem 5: (without proof) For any function F satisfying a) - c) (Theorem 2) there is a probability space $(S, \mathcal{A}, P)$ and a random variable X such that $F_X = F$.

- Note that different random variables may have the same cdf.

Example 15: Let us consider the experiment consisting of three tosses of a coin, and two random variables: X the total number of heads and Y the total number of tails.

A simple count shows that $P(X = k)$ and $P(Y = k)$ are the same for all k and hence the distribution functions F_X and F_Y are identical.

Transformation of random variable

- in practical situations we often deal with a function (transformation) of a random variable

Example 16: A computer can generate a random number from distribution uniform on $[0, 1]$. Then if the desired random number is from a different range $[a, b], a < b$, one just need to apply a linear transformation

$$\varphi(x) = (b - a)x + a.$$

- functions of random variables are very important in statistics

Transformation of discrete random variables

- Let X assume values in the set $U = \{x_1, x_2, \dots\}$, with corresponding probabilities $p_i = P(X = x_i)$, such that $\sum p_i = 1$. Then $Y = \varphi(X)$, where φ is a real-valued function, also has a discrete distribution.
- function φ may not be one-to-one, thus

$$P(Y = y) = P(\varphi(X) = y) = \sum_{x:\varphi(x)=y} P(X = x).$$

Example 17: Suppose that X has a distribution

values	-2	-1	0	1	2	3	4
probabilities	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

If $\varphi(x) = x^2$ then $Y = X^2$ with the distribution

values	0	1	4	9	16
probabilities	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

Transformation of continuous random variables

- Let F and f denote the cdf and density of continuous random variable X and let $Y = \varphi(X)$, where φ is assumed to be at least piecewise differentiable.
- Consider first the case of φ strictly monotone
- If φ is strictly increasing then

$$F_Y(y) = P(\varphi(X) \leq y) = P(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y)).$$

- If φ is strictly decreasing then

$$F_Y(y) = P(\varphi(X) \leq y) = P(X \geq \varphi^{-1}(y)) = 1 - F_X(\varphi^{-1}(y)).$$

Theorem 6: If φ is a continuous differentiable function with inverse ψ and X is a continuous random variable with density f_X, then the density of $Y = \varphi(X)$ is

$$f_Y(y) = f_X(\psi(y))|\psi'(y)|, y \in \mathbb{R}.$$

Probability integral transformation

Example 18: Let $Y = aX + b$. If $a > 0$ then for $y \in \mathbb{R}$,

$$F_Y(y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Theorem 7: (Probability integral transformation)
Let X be a continuous random variable with strictly increasing cdf F. Then the distribution of $Y = F(X)$ is uniform on $[0, 1]$.

Transformation of a continuous random variable

- when φ is not monotone we still have

$$F_Y(y) = P(\varphi(X) \le y),$$

but since φ has no inverse the inequality $\varphi(X) \le y$ is usually not equivalent to a single inequality for X

Example 19: Let X be a continuous random variable and $\varphi(x) = x^2$. Then $F_Y(y) = 0$ for $y \le 0$ and for $y > 0$

$$F_Y(y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Also, for $y > 0$

$$f_Y(y) = F_Y'(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$