Probability Theory and Statistics
Lecture 7

Bence Csonka

Budapest University of Technology and Economics
csonkab@edu.bme.hu

September 30

Remainder for Devation

Definition

Let X be a random variable with $\mathbb{E}(X^2) < \infty$. Then

$$\mathbb{D}^2(X) := \mathbb{E}[(X - \mathbb{E}(X))^2]$$

is called the variance of X, and $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$ is called the standard deviation of X.

Variances of notable discrete distributions

Theorem

Indicator: Let $X = \mathbb{1}_A$ be the indicator of an event A. Then $\mathbb{D}^2(X) = \mathbb{P}(A)(1 - \mathbb{P}(A))$.

(Proof.)
Equivalently: if X is Bernoulli(p) with $p \in [0, 1]$, then $\mathbb{D}^2(X) = p(1 - p)$.

Variances of notable discrete distributions

Theorem

Indicator: Let $X = \mathbb{1}_A$ be the indicator of an event A. Then $\mathbb{D}^2(X) = \mathbb{P}(A)(1 - \mathbb{P}(A))$.

(Proof.)

Equivalently: if X is Bernoulli(p) with $p \in [0, 1]$, then $\mathbb{D}^2(X) = p(1 - p)$.

Theorem

Binomial: Let $X \sim \mathrm{B}(n; p)$. Then $\mathbb{D}^2(X) = np(1 - p)$.

(Proof.)

Variances of notable discrete distributions

Theorem

Indicator: Let $X = \mathbb{1}_A$ be the indicator of an event A. Then $\mathbb{D}^2(X) = \mathbb{P}(A)(1 - \mathbb{P}(A))$.

(Proof.)

Equivalently: if X is Bernoulli(p) with $p \in [0, 1]$, then $\mathbb{D}^2(X) = p(1 - p)$.

Theorem

Binomial: Let $X \sim \mathrm{B}(n; p)$. Then $\mathbb{D}^2(X) = np(1 - p)$.

(Proof.)

Theorem

Geometric: Let $X \sim \mathrm{Geo}(p)$. Then $\mathbb{D}^2(X) = (1 - p)/p^2$.

(Proof)

Variances of notable discrete distributions

Theorem

Poisson: Let $X \sim \mathrm{Pois}(\lambda)$. Then $\mathbb{D}^2(X) = \lambda$.

(Proof: in practice class!.)
Remark: For the Poisson distribution $\mathbb{D}^2(X) = \mathbb{E}(X)$, which is not typical for other well-known distribution families.

Discrete joint distribution as a two-dimensional pmf

The discrete joint distribution of two random variables X and Y is a *two-dimensional probability mass function*: the function $(x, y) \mapsto p_{(X,Y)}(x, y)$, where by definition

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

This is also called the joint pmf of X and Y.

$Y \backslash X$	0	1
0	$\frac{1}{4}$	0
1	$\frac{1}{4}$	$\frac{1}{4}$
2	0	$\frac{1}{4}$

In the example: $p_{(X,Y)}(0, 1) = \frac{1}{4}$.

Theorem

Two discrete random variables X and Y are independent if and only if their joint pmf factors as the product of the marginals; that is, for all $x, y \in \mathbb{R}$,

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

Convolution in the discrete case

If X and Y are independent random variables, what is the distribution of $X + Y$?

Theorem

Let X and Y be independent, discrete random variables taking values in the nonnegative integers. Then

$$\mathbb{P}(X + Y = k) = \sum_{i=0}^{k} \mathbb{P}(X = i)\mathbb{P}(Y = k - i)$$

for every $k \in \{0, 1, 2, \ldots\}$.

(Proof.)

(Convolution: the distribution of the sum of independent random variables.)
(For general discrete X, Y the formula is analogous.)

Convolution of Poisson

Let X and Y be independent random variables with $X \sim \mathrm{Pois}(\lambda)$ and $Y \sim \mathrm{Pois}(\mu)$. Then $X + Y \sim \mathrm{Pois}(\lambda + \mu)$, i.e., for any $k \in \{0, 1, 2, \ldots\}$,

$$\mathbb{P}(X + Y = k) = \frac{(\lambda + \mu)^k}{k!} \mathrm{e}^{-(\lambda+\mu)}.$$

(Proof: see the lecture.)

Continuous random variables

Reminder: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a function $X\colon \Omega \to \mathbb{R}$ is called a random variable if for every $x \in \mathbb{R}$ the event $\{X < x\} = \{\omega \in \Omega \mid X(\omega) < x\}$ belongs to $\mathcal{F}$.

So far: discrete r.v.'s, where $\mathrm{Ran}(X)$ is countable ($=$ finite or countably infinite).

Examples of continuous random quantities with uncountable range:

- Geometric probability space: "choose a point *at random* on an interval"; probabilities are proportional to "area" (interval length) $\to$ the chosen point is a random variable with a continuous range
- Waiting times (e.g., time until the next bus arrives, or until a light bulb burns out)
- The change in the price of a given stock over a day (this can be negative)
- The height, weight, etc. of a randomly chosen Hungarian citizen

Distribution function

Since $\{X < x\}$ is an event, the probability $\mathbb{P}(\{X < x\})$ is defined; we have already abbreviated this as $\mathbb{P}(X < x)$.
The function $\mathbb{R} \to \mathbb{R}$, $x \mapsto \mathbb{P}(X < x)$ is very useful, especially when $\mathrm{Ran}(X)$ is not finite—let alone when it is uncountable.

Definition

Let X be a random variable. Then $F_X : \mathbb{R} \to \mathbb{R}$,

$$F_X(x) \stackrel{\text{def}}{=} \mathbb{P}(X < x) \in [0, 1]$$

is called the distribution function (cdf) of X.

Simple example: toss a fair coin, $\Omega = \{T, H\}$, $X(T) = 1$, $X(H) = 0$.
Plot the distribution function of X ($X(T) = 1$, $X(H) = 0$)! (See lecture.)
Note that the cdf is not continuous—and this will always be the case when $\mathrm{Ran}(X)$ is countable. However, it is left-continuous.

Properties of the distribution function

For any real numbers $a < b$,

$$F_X(b) - F_X(a) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = \mathbb{P}(a \le X < b),$$

by additivity of probability.

CDFs can also be characterized:

Theorem

A function $F : \mathbb{R} \to \mathbb{R}$ is the distribution function of some random variable $\Leftrightarrow$

(1) *F is (not necessarily strictly) nondecreasing,*

(2) *F is left-continuous, i.e., for every $x \in \mathbb{R}$, $\lim_{y \uparrow x} F(y) = F(x)$,*

(3) *and $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.*

- (Proof: the $\Rightarrow$ direction)
- (Remark: convention regarding the definition of the cdf differs across countries)

Distribution function

Definition

Let X be a random variable. Then $F_X : \mathbb{R} \to \mathbb{R}$,

$$F_X(x) \stackrel{\text{def}}{=} \mathbb{P}(X < x) \in [0, 1]$$

is called the distribution function of X.

This is defined for every random variable.

In the discrete case it was more convenient to use the probability mass function $p_X : k \mapsto p_X(k) = \mathbb{P}(X = k)$ than the cdf.

Indeed, if $\mathrm{Ran}(X)$ is finite, then the cdf F_X has upward jumps at the k's in the range of X (jump size: $\mathbb{P}(X = k)$), and is otherwise constant. (See the coin example.)

Distribution function

Definition

Let X be a random variable. Then $F_X : \mathbb{R} \to \mathbb{R}$,

$$F_X(x) \stackrel{\text{def}}{=} \mathbb{P}(X < x) \in [0, 1]$$

is called the distribution function of X.

This is defined for every random variable.

In the discrete case it was more convenient to use the probability mass function $p_X : k \mapsto p_X(k) = \mathbb{P}(X = k)$ than the cdf.

Problem: if the cdf is *continuous* (as for a point chosen “at random” on $[0, 1]$), then for every fixed $x \in \mathbb{R}$ we have $\mathbb{P}(X = x) = 0$. Thus the pmf is trivial—we need a different quantity $\to$ the density function.

Density function

Thus, for a “continuous” random variable, the derivative of the cdf (or a function whose antiderivative is the cdf) gives an approximation to the probability of the variable falling in a small neighborhood of a point. With this motivation we introduce the following notions.

Density function

Thus, for a “continuous” random variable, the derivative of the cdf (or a function whose antiderivative is the cdf) gives an approximation to the probability of the variable falling in a small neighborhood of a point. With this motivation we introduce the following notions.

Definition

A random variable X is called continuous if there exists a nonnegative real function $f_X \colon \mathbb{R} \to \mathbb{R}$ such that the improper Riemann integral $\int_{-\infty}^{\infty} f_X(z)\mathrm{d}z$ is finite, and for every $x \in \mathbb{R}$,

$$F_X(x) = \int_{-\infty}^{x} f_X(z)\mathrm{d}z,$$

where F_X is the cdf of X and the integral is an improper Riemann integral. The function f_X is called the density function of X.

Density function

By the Newton–Leibniz theorem, since the integral of the density is the cdf, the derivative of the cdf is the density.
It is not a problem if the derivative of F_X fails to exist at finitely many points—this will in fact be useful in practice:

Theorem

If F_X is continuous and differentiable everywhere except at finitely many points, then X is a continuous random variable, and

$$f(x) = \begin{cases} F_X'(x), & \text{if } F_X \text{ is differentiable at } x, \\ 0, & \text{otherwise,} \end{cases}$$

is a density function of X.

Instead of 0 one could write, say, 42 here; for integration it makes no difference if we change the density at finitely many points (as long as it remains nonnegative).

Properties of the density

Theorem

Let X be a continuous random variable. Then for all real numbers $a < b$,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\mathrm{d}x.$$

(Proof)

In the proof we saw: if X is continuous, then for every $a \in \mathbb{R}$,

$$\mathbb{P}(X = a) = \int_a^a f_X(x)\mathrm{d}x = 0.$$

Hence for all a, b,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b).$$

This is *not* the case in the discrete setting!!!

Characterization of densities

Theorem

A nonnegative function $f : \mathbb{R} \to \mathbb{R}$ is a density of some continuous random variable X if and only if f is Riemann-integrable and

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$$

Proof:

- Easy direction: if X is continuous, then f_X satisfies the equation; see lecture.

Characterization of densities

Theorem

A nonnegative function $f : \mathbb{R} \to \mathbb{R}$ is a density of some continuous random variable X if and only if f is Riemann-integrable and

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$$

Proof:

- For the other direction (not in the notes): if $f \geq 0$ is Riemann-integrable and $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$, then for $x \in \mathbb{R}$ define

$$F(x) := \int_{-\infty}^{x} f(z)\mathrm{d}z.$$

It is easy to check that F satisfies the cdf properties (correct limits at $\pm\infty$, nondecreasing, left-continuous).

First Example: the Uniform Distribution

Definition

The density function of a random variable X uniformly distributed on the interval (a, b) is

$$f_X(x) = \frac{1}{b-a},$$

if $a < x < b$, and $f_X(x) = 0$ otherwise. By integration, we obtain

$$F_X(x) = \begin{cases} 0, & \text{if } x \le a, \\ \frac{x-a}{b-a}, & \text{if } a < x < b, \\ 1, & \text{if } x \ge b. \end{cases}$$

Notation: $X \sim \mathrm{U}(a;b)$. For the $\mathrm{U}(a;b)$ distribution the density function is constant (on (a, b), and zero outside), while the distribution function is linear (on (a, b), and constant outside).

Example: Exponential Distribution

Definition

A random variable Z has an exponential distribution with parameter $\lambda > 0$ if

$$f_Z(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Equivalently,

$$F_Z(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1 - \mathrm{e}^{-\lambda x}, & \text{if } x > 0. \end{cases}$$

Notation: $X \sim \mathrm{Exp}(\lambda)$.
Occurrence: waiting times ($Z \geq 0$ with probability 1), e.g. the remaining lifetime of a light bulb until it burns out.

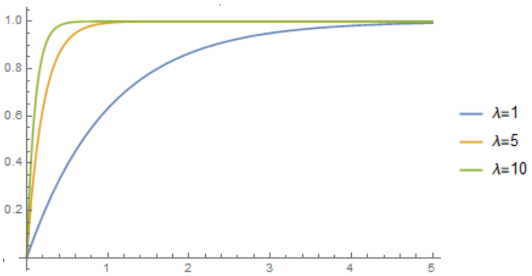
Example for exponential distribution

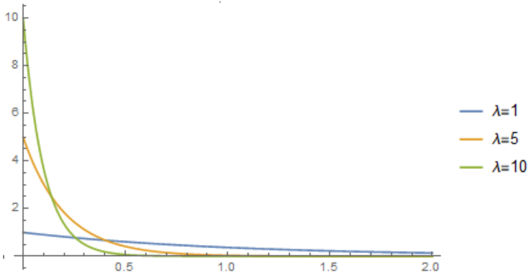
Let X be an exponentially distributed random variable such that $\mathbb{P}(X > 3) = e^{-6}$.

Example for exponential distribution

Let X be an exponentially distributed random variable such that $\mathbb{P}(X > 3) = e^{-6}$.
a) What is the parameter λ of the distribution of X?
b) Compute $\mathbb{P}(X < 2)$.

Exponential Distribution: CDF and PDF





Exponential Distribution: Memorylessness

Reminder: the geometric distribution is memoryless: if $X \sim \mathrm{Geo}(p)$, then

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \qquad \forall s, t \in \{1, 2, \dots\}. \tag{1}$$

Note: if at least one of s, t is not an integer, then (1) need not hold. For instance, let $t = 3/4$, $s = 1/2$, then

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > 5/4 \mid X > 3/4) = \mathbb{P}(X \geq 2) = 1 - p,$$

but

$$\mathbb{P}(X > s) = \mathbb{P}(X > 1/2) = 1.$$

However, if $X \sim \mathrm{Exp}(\lambda)$ with $\lambda > 0$, then (1) holds for all $s, t \in [0, \infty)$. (Proof: see lecture.)

Exponential Distribution: Memorylessness

Reminder: the geometric distribution is memoryless: if $X \sim \mathrm{Geo}(p)$, then

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \qquad \forall s, t \in \{1, 2, \dots\}. \tag{1}$$

Note: if at least one of s, t is not an integer, then (1) need not hold. For instance, let $t = 3/4$, $s = 1/2$, then

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > 5/4 \mid X > 3/4) = \mathbb{P}(X \geq 2) = 1 - p,$$

but

$$\mathbb{P}(X > s) = \mathbb{P}(X > 1/2) = 1.$$

However, if $X \sim \mathrm{Exp}(\lambda)$ with $\lambda > 0$, then (1) holds for all $s, t \in [0, \infty)$. (Proof: see lecture.)

Modeling question: for which types of waiting times does (1) hold? E.g. for light bulbs, measurements suggest it is *approximately* true. For human lifetimes, it is not true!
Light bulbs hardly age; humans do.