Probability and Statistics (EPRST)

Lecture 8

A revision - variance

Recall: the variance of X is

$$\operatorname{Var} X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Properties of the variance:

- $\operatorname{Var} X \geq 0$,
- $\operatorname{Var} X = 0$ iff X has a one-point distribution,
- $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$ for all real numbers a, b.

Variances of the important named distributions

- if $X \sim \mathrm{bin}(n, p)$, then $\operatorname{Var} X = np(1-p)$
- if $X \sim \mathrm{geom}(p)$, then $\operatorname{Var} X = (1-p)/p^2$
- if $X \sim \mathrm{Poiss}(\lambda)$, then $\operatorname{Var} X = \lambda$
- if $X \sim \mathrm{U}(a, b)$, then $\operatorname{Var} X = (b-a)^2/12$
- if $X \sim \mathrm{Exp}(\lambda)$, then $\operatorname{Var} X = 1/\lambda^2$
- if $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then $\operatorname{Var} X = \sigma^2$

Sample variance and sample standard deviation

If $x_1, \ldots, x_n$ are some (random) values of a random variable X (independently generated) then a natural estimate of $\mathbb{E}g(X)$ (g is any function) is the arithmetic mean of the values $g(x_1), \ldots, g(x_n)$:

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i).$$

However, the **sample variance** is defined as

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}_n\right)^2$$

The **sample standard deviation** is the square root of the sample variance.

Sample variance - cont'd

The idea of the definition is to mimic the formula

$$\operatorname{Var} X = \mathbb{E}\,(X - \mathbb{E}X)^2$$

by averaging the squared distances of the x_i from the sample mean, except with $n-1$ rather than n in the denominator.
The motivation for the $n-1$ is that this makes S_n^2 **unbiased** for estimating $\operatorname{Var} X$, that is it is correct on average.

Quantiles

Definition

*Let $q \in (0,1)$ and X - a random variable. Number a_q is q-**quantile** of the distribution of X, if*

$$\mathbb{P}(X \leq a_q) \geq q \quad and \quad \mathbb{P}(X \geq a_q) \geq 1-q,$$

or, equivalently,

$$\mathbb{P}(X < a_q) \leq q \leq \mathbb{P}(X \leq a_q).$$

- *For $q = 1/2$, q-quantile is called **median** (denoted: $\operatorname{med} X$).*
- ***Quartiles** q-quantiles with $q = 1/4, 1/2, 3/4$.*
- *Also **deciles** and **percentiles** are frequently considered.*

Quantiles - some examples

Example

- *if* $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/4$, $\mathbb{P}(X = 0) = 1/2$, *then* $\operatorname{med} X = ?$
- *if* X *has a discrete uniform distribution on the set* $\{1, 2, 3, 4\}$, *then* $\operatorname{med} X = ?$

Quantiles - cont'd

If the cumulative distribution function of the distribution of a random variable X is a function which is continuous and strictly increasing on an interval (a, b), with $-\infty \leq a < b \leq \infty$, then the definition of quantile becomes simpler - a number c is a q-quantile, if

$$\mathbb{P}(X \leq c) = F_X(c) = q,$$

so

$$c = F_X^{-1}(q).$$

Example

- *If* $X \sim \mathcal{N}(0, 1)$, *then* $\operatorname{med} X = ?$
- *If* X *has a Cauchy distribution, then* $\operatorname{med} X = ?$

Median vs expectation

The following assertions hold:

- if X is a random variable such that $\mathbb{E}|X| < \infty$, then function

  $$f_1(a) = \mathbb{E}|X - a|$$

  attains its minimal value at

  $$a = \operatorname{med} X.$$

- if X is a random variable such that $\mathbb{E}X^2 < \infty$, then function

  $$f_2(a) := \mathbb{E}(X - a)^2$$

  attains its minimal value at

  $$a = \mathbb{E}X.$$

Random vectors

Random vectors

Definition

Consider two random variables $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$, defined on the same sample space.

- *Function $\boldsymbol{X} : \Omega \to \mathbb{R}^2$ given by*

$$\boldsymbol{X}(\omega) = \begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix}$$

  *is called a **random vector** or a **two-dimensional random variable**.*

- ***Joint distribution of the random vector $\boldsymbol{X}$** is a function which assigns a number*

$$\begin{aligned} \mathbb{P}(\boldsymbol{X} \in B) &= \mathbb{P}((X, Y) \in B) \\ &= \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}) \end{aligned}$$

  to any set $B \subset \mathbb{R}^2$.

Random vectors - some examples

Example

We flip two symmetric coins (heads=0, tails=1). Let X denote the sum of the ourcomes, and Y - the absolute value of their difference. Find the joint distribution of $\boldsymbol{X} = (X, Y)$.

Example

We randomly pick a point from the triangle

$$D = \left\{(x, y) \in \mathbb{R}^2 : x, y \geq 0,\ x + y \leq 1\right\}.$$

Let X and Y denote (random) coordinates of the picked point, and $\boldsymbol{X} = (X, Y)$. What is the joint distribution of $\boldsymbol{X}$?

General random vectors

For simplicity, we are going to discuss (mainly) two-dimensional random vectors. But the concept of a random vector can be applied to any dimension: if $X_1, \ldots, X_n$ are some random variables (defined on the common sample space), then

$$\boldsymbol{X} = (X_1, \ldots, X_n)$$

is an n-dimensional random vector. Its joint distribution is a probability distribution on $\mathbb{R}^n$.

Types of distributions of random vectors

As in the case of ordinary one-dimensional probability distributions, we will consider two types of two-dimensional distributions:

- discrete and
- continuous distributions.

As before, discrete distributions will have finite or countable supports, and continuous distributions will have uncountable supports.

Discrete random vectors

If X and Y are random variables (on a common sample space Ω) and

- X takes values from a finite or countably infinite set $\mathcal{X} := \{x_1, x_2, \ldots\} \subset \mathbb{R}$,
- Y takes values from a finite or countably infinite set $\mathcal{Y} := \{y_1, y_2, \ldots\} \subset \mathbb{R}$,

then we define the joint distribution of the random vector $\boldsymbol{X} = (X, Y)$ giving the probability

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}).$$

for every pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Discrete random vectors - cont'd

If sets X and Y take only finite numbers of values, it is often convenient to present their probabilities in a table.

Example

We flip a symmetric coin twice. Let X_i be the outcome of the i-th toss (we identify $\mathrm{H} \to 1$, $\mathrm{T} \to -1$*). Present the (joint) distribution of the random vector $\boldsymbol{X} = (X_1, X_2)$ in a table.*

Continuous random vectors

Definition

A random vector $\boldsymbol{X} = (X, Y)$ has a ***continuous joint distribution****, if there exists a function $f : \mathbb{R}^2 \to [0, \infty)$ such that*

$$\mathbb{P}(\boldsymbol{X} \in B) = \mathbb{P}((X, Y) \in B) = \iint_B f(x, y) dx dy$$

for $B \subset \mathbb{R}^2$. Function f is called ***joint density****.*

As in dimension one, function $f : \mathbb{R}^2 \to \mathbb{R}$ is joint density of a certain random vector, if f takes non-negative values and

$$\iint_{\mathbb{R}^2} f(x, y) \mathrm{d}x \mathrm{d}y = 1.$$

Joint density - an example

Example

Set $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = \begin{cases} cxy, & (x,y) \in [0,1]^2, \\ 0, & (x,y) \notin [0,1]^2. \end{cases}$$

- *For what values of* c *is this function a two-dimensional density?*
- *If* f *is a joint density of* $\boldsymbol{X} = (X, Y)$, *what is* $\mathbb{P}(\boldsymbol{X} \in [0, 1/2]^2)$*?*

Marginal distributions

Regardless of whether a random vector $\boldsymbol{X} = (X, Y)$ is discrete or continuous, distributions of random variables X and Y can be derived from its joint distribution, because for $A, B \subset \mathbb{R}$

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(\boldsymbol{X} = (X, Y) \in A \times \mathbb{R}), \\ \mathbb{P}(Y \in B) &= \mathbb{P}(\boldsymbol{X} = (X, Y) \in \mathbb{R} \times B). \end{aligned}$$

The distributions of (one-dimensional) random variables X and Y are called **marginal distributions of random vector $\boldsymbol{X}$**.

How to determine marginal distributions of discrete vectors?

Suppose X, Y are (one-dimensional) discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Given the joint distribution of the discrete random vector $\boldsymbol{X} = (X, Y)$, that is, probabilities

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

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, one determines the marginal distributions from

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) \ \forall x$$

and

$$\mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y) \ \forall y.$$

If $\boldsymbol{X}$ has a discrete joint distribution, then its marginal distributions are discrete as well.

Marginal distributions of a discrete random vector - an example

Example

Find the marginal distributions for random vector $\boldsymbol{X} = (X, Y)$ *with the joint distribution given by the table:*

-

$X \backslash Y$	0	1
0	1/4	1/4
1	1/4	1/4

-

$X \backslash Y$	0	1
0	1/8	3/8
1	3/8	1/8

Joint vs marginal distributions

The above example illustrates an important fact (which is true regardless of distribution types):

- knowledge of the joint distribution allows to determine the marginal distributions,
- in general (without some additional assumptions), knowledge of the marginal distributions does not allow to determine the joint distribution.

The joint distribution contains more information than both marginal distributions together.

How to determine marginal distributions of continuous vectors?

If function f (defined on $\mathbb{R}^2$) is the joint density of random vector $\boldsymbol{X} = (X, Y)$, then the marginal distributions of $\boldsymbol{X}$ are continuous, and the marginal densities can be computed from the following formulas

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y, \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x. \end{aligned}$$

Example

Compute the marginal densities of random vector $\boldsymbol{X} = (X, Y)$, uniformly distributed on a set D, $\boldsymbol{X} \sim \mathcal{U}(D)$, where

- $D = [0,1]^2$,
- $D = \{(x,y) \in \mathbb{R}^2 : x, y \geq 0, x + y \leq 1\}$.