Probability Theory and Statistics
Lecture 6

Bence Csonka

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

September 29

Independence of discrete random variables

So far we mainly discussed the distribution of a single (mostly discrete) random variable.

Now: We want to describe how several (typically two) discrete random variables depend on each other, i.e., how they influence each other's behavior, and how this can be captured in a simple, compact form.
In particular, we extend the notion of independence, defined so far only for events, to random variables.

Example: We roll a die twice; let X be the first and Y the second outcome. Knowing X tells us nothing more about Y than without this information, and vice versa. We will see shortly: in this case X and Y are indeed independent random variables.

Discrete joint distribution as a two-dimensional pmf

Definition

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

Independence of random variables

Recall: events A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition

Let $X, Y : \Omega \to \mathbb{R}$ *be random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$. *We say* X *and* Y *are independent if for all* $x, y \in \mathbb{R}$, *the events* $\{X < x\}$ *and* $\{Y < y\}$ *are independent, i.e.,* $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$.

Independence of random variables

Recall: events A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition

Let $X, Y : \Omega \to \mathbb{R}$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X and Y are independent if for all $x, y \in \mathbb{R}$, the events $\{X < x\}$ and $\{Y < y\}$ are independent, i.e., $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$.

This definition also works when the random variables are not necessarily discrete. In the discrete case there is a simpler/more intuitive equivalent definition.

Independence of random variables

Recall: events A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition

Let $X, Y : \Omega \to \mathbb{R}$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X and Y are independent if for all $x, y \in \mathbb{R}$, the events $\{X < x\}$ and $\{Y < y\}$ are independent, i.e., $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$.

This definition also works when the random variables are not necessarily discrete. In the discrete case there is a simpler/more intuitive equivalent definition.

Theorem

Two discrete random variables X and Y are independent if and only if for all $x, y \in \mathbb{R}$ the events $\{X = x\}$ and $\{Y = y\}$ are independent, i.e.,

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

Example: discrete joint distribution and independence

Example: Flip a fair coin twice. Let $X = \mathbb{1}_{\{\text{2nd toss is heads}\}}$ and let Y be the total number of heads.

1. Give the joint distribution of X and Y, i.e., the probabilities $\mathbb{P}(X = k, Y = l)$ for all k, l where this is nonzero.
2. Also give the marginal distributions of X and Y, i.e., the pmfs p_X and p_Y.
3. Are X and Y independent?
4. Compute $\mathbb{E}(XY)$. (Recall: by definition $XY(\omega) = X(\omega)Y(\omega)$, $\omega \in \Omega$.)

(Before computing, think about whether they are independent and what the marginals p_X and p_Y will be.)

Example: discrete joint distribution and independence

$X = \mathbb{1}_{\{\text{2nd toss is heads}\}}$, $Y =$ number of heads.
We give the joint distribution in a table (computation: in the lecture):

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

The probabilities in the table sum to 1.
Where do the marginals appear in the table?

Example: discrete joint distribution and independence

$X = \mathbb{1}_{\{\text{2nd toss is heads}\}}$, $Y =$ number of heads.
We give the joint distribution in a table (computation: in the lecture):

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

The probabilities in the main body of the table sum to 1.
Where do the marginals appear in the table?

- $p_X(k) = \mathbb{P}(X = k) = \sum_{l \in \mathrm{Ran}(Y)} \mathbb{P}(X = k, Y = l)$, so the column sums give the corresponding values of p_X.
- $p_Y(l) = \mathbb{P}(Y = l) = \sum_{k \in \mathrm{Ran}(X)} \mathbb{P}(X = k, Y = l)$, so the row sums give the corresponding values of p_Y.

$\mathbb{E}(XY) = \frac{3}{4}$ (see the lecture). X and Y are not independent since, e.g.,
$\mathbb{P}(X = 0, Y = 2) = 0 \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 2)$.

“Tips and tricks” for joint distributions

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

For any discrete random variables X, Y:

- To show that X and Y are not independent, it suffices to find one pair (k, l) with $\mathbb{P}(X = k, Y = l) \neq \mathbb{P}(X = k)\mathbb{P}(Y = l)$.
  A convenient special case visible in a table: if $0 = \mathbb{P}(X = k, Y = l)$ while $\mathbb{P}(X = k) \neq 0$ and $\mathbb{P}(Y = l) \neq 0$.
- To show that X and Y are independent, you must check every pair (k, l) that $\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k)\mathbb{P}(Y = l)$.
- The order of X and Y matters. If you swap X and Y, then p_X appears in the rows and p_Y in the columns.

Remainder Expected Value

Definition

Let X be a discrete random variable such that

$$\sum_{k \in \mathrm{Ran}(X)} |k|\, p_X(k) < \infty.$$

Then

$$\mathbb{E}(X) \stackrel{\mathrm{def}}{=} \sum_{k \in \mathrm{Ran}(X)} k\mathbb{P}(X = k) = \sum_{k \in \mathrm{Ran}(X)} k\, p_X(k) \in \mathbb{R}.$$

Remainder Expected Value

Definition

Let X be a discrete random variable such that

$$\sum_{k \in \mathrm{Ran}(X)} |k| \, p_X(k) < \infty.$$

Then

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \sum_{k \in \mathrm{Ran}(X)} k \mathbb{P}(X = k) = \sum_{k \in \mathrm{Ran}(X)} k \, p_X(k) \in \mathbb{R}.$$

In most cases $\mathrm{Ran}(X) \subseteq \mathbb{N}$, hence the absolute value does not matter.

Properties of expected value

Theorem

The linearity of expectation, $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, $\mathbb{E}(cX) = c\,\mathbb{E}(X)$, *holds for arbitrary (not necessarily discrete) random variables* X, Y *whose expected values exist.*

(Proof)

Expectation of a product for independent random variables

If X and Y are independent, then the expectation of their product is the product of their expectations:

Theorem

If X and Y are independent random variables and $\mathbb{E}(XY)$, $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{1}$$

(Proof)

Expectation of a product for independent random variables

If X and Y are independent, then the expectation of their product is the product of their expectations:

Theorem

If X and Y are independent random variables and $\mathbb{E}(XY)$, $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{1}$$

(Proof)

The converse is false! It may happen that X and Y are not independent, yet (1) still holds. Example: later.

Independence of several random variables

Recall: independence of $n \geq 2$ events $A_1, \ldots, A_n$ means that for any $2 \leq k \leq n$, and any choice of k events, the probability of their intersection equals the product of their probabilities. Formally: $A_1, \ldots, A_n$ are independent $\Leftrightarrow$ for all $k \in \{2, \ldots, n\}$ and all $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

For $n \geq 3$ it did not suffice, e.g., to require pairwise independence or only the n-fold intersection property.

Independence of several random variables

Recall: independence of $n \geq 2$ events $A_1, \ldots, A_n$ means that for any $2 \leq k \leq n$, and any choice of k events, the probability of their intersection equals the product of their probabilities. Formally: $A_1, \ldots, A_n$ are independent $\Leftrightarrow$ for all $k \in \{2, \ldots, n\}$ and all $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

Similarly, if $X_1, \ldots, X_n$ ($n \geq 2$) are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, their independence means that for any $k \in \{2, \ldots, n\}$, any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, and any $x_{i_1}, \ldots, x_{i_k} \in \mathbb{R}$,

$$\mathbb{P}(X_{i_1} < x_{i_1}, \ldots, X_{i_k} < x_{i_k}) = \mathbb{P}(X_{i_1} < x_{i_1}) \cdot \ldots \cdot \mathbb{P}(X_{i_k} < x_{i_k}).$$

In the discrete case this is equivalent to replacing all the "$<$" by "$=$", that is (with the same conditions)

$$\mathbb{P}(X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k}) = \mathbb{P}(X_{i_1} = x_{i_1}) \cdot \ldots \cdot \mathbb{P}(X_{i_k} = x_{i_k}).$$

Expectation of a transformation in the discrete case

Definition

For a random variable X, a transformation is defined as the function $g(X)\colon \Omega \to \mathbb{R}$ for some $g\colon \mathrm{Ran}(X) \subseteq \mathbb{R} \to \mathbb{R}$.

The transformation is defined „pointwise”:

$$(g(X))(\omega) = g(X(\omega)).$$

For example, if X is the outcome of a die roll and we know that $X = 5$, then $X^2 = 25$.

Expectation of a transformation in the discrete case

Definition

For a random variable X, a transformation is defined as the function $g(X)\colon \Omega \to \mathbb{R}$ for some $g\colon \mathrm{Ran}(X) \subseteq \mathbb{R} \to \mathbb{R}$.

The transformation is defined „pointwise”:

$$(g(X))(\omega) = g(X(\omega)).$$

For example, if X is the outcome of a die roll and we know that $X = 5$, then $X^2 = 25$.

If X is a discrete random variable, then any of its transformations is also a random variable (i.e. $\{g(X) < x\} \in \mathcal{F}$ holds for all $x \in \mathbb{R}$) and is also discrete. For general random variables, the situation is more complicated.

Question: how do we compute the expectation of the transformation of a discrete random variable (if it exists)?

Expectation of a transformation in the discrete case

Example with die roll $\rightarrow$ we computed the probability mass function of X^2 and from this $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \sum_{k\in\{1,4,9,16,25,36\}=\mathrm{Ran}(X^2)} k\mathbb{P}(X^2 = k).$$

Question: do we really need to compute the distribution of X^2 for this?
Alternative: take the probability mass function of X and weight with k^2 instead of k:

$$\mathbb{E}(X^2) = \sum_{k\in\{1,2,\dots,6\}=\mathrm{Ran}(X)} k^2\mathbb{P}(X = k).$$

In general:
If X is a discrete random variable, then for any transformation $g(X)$ for which the expectation exists:

$$\mathbb{E}(g(X)) = \sum_{k\in\mathrm{Ran}(X)} g(k)\mathbb{P}(X = k).$$

Moments of random variables

Theorem

If X is a discrete random variable, then for any transformation $g(X)$ for which the expectation exists:

$$\mathbb{E}(g(X)) = \sum_{k \in \mathrm{Ran}(X)} g(k)\mathbb{P}(X = k).$$

According to the earlier definition, $\mathbb{E}(g(X))$ exists if

$$\infty > \sum_{k \in \mathrm{Ran}(g(X))} |k|\mathbb{P}(g(X) = k) = \sum_{l \in \mathrm{Ran}(X)} |g(l)|\mathbb{P}(X = l),$$

and the right-hand side is precisely $\mathbb{E}(|g(X)|)$, the expectation of the absolute value of the transformation.

Moments of random variables

Theorem

If X is a discrete random variable, then for any transformation $g(X)$ for which the expectation exists:

$$\mathbb{E}(g(X)) = \sum_{k \in \mathrm{Ran}(X)} g(k)\mathbb{P}(X = k).$$

According to the earlier definition, $\mathbb{E}(g(X))$ exists if

$$\infty > \sum_{k \in \mathrm{Ran}(g(X))} |k|\mathbb{P}(g(X) = k) = \sum_{l \in \mathrm{Ran}(X)} |g(l)|\mathbb{P}(X = l),$$

and the right-hand side is precisely $\mathbb{E}(|g(X)|)$, the expectation of the absolute value of the transformation.

Special case: $g(x) = x^n$, $n = 1, 2, \ldots$.

Moments of random variables

Special case: $g(x) = x^n$, $n = 1, 2, \ldots$.

Definition

$\mathbb{E}[|X|^n]$ *is called the nth* *absolute moment* *of* X.
If X *has a finite nth absolute moment, then* $\mathbb{E}(X^n)$ *is called the nth* *moment* *of* X.

Thus, the first moment is the expectation (if it exists).
(Example: moments of an indicator random variable)

Variance and standard deviation

It is often said that the expectation is "the most information one can express about a distribution/random variable with a single number." This information is obviously not complete. (E.g., "a fair die shows 3.5 on average" $\rightarrow$ this is true, but it tells us nothing about the support of the die outcomes.)

Variance and standard deviation

It is often said that the expectation is "the most information one can express about a distribution/random variable with a single number." This information is obviously not complete. (E.g., "a fair die shows 3.5 on average" $\rightarrow$ this is true, but it tells us nothing about the support of the die outcomes.)

Example: To participate in a lottery we must pay 1 euro. The referee flips a coin; if the result is heads, we receive 2 euros, otherwise we lose our 1 euro as well.
Then our expected gain is 0 euros.

Now raise the entry fee to 1000 euros and the possible prize to 2000. The expectation stays the same, but we feel the difference.

General question: How can we measure the distance between a random variable and its expectation?

Variance and standard deviation

The (absolute) deviation of a random variable X from its expectation is $|X - \mathbb{E}(X)|$.

We could study its expectation, but the absolute value is often inconvenient to handle.

Instead we study the square of this distance, namely

$$\mathbb{E}[(X - \mathbb{E}(X))^2].$$

This quantity has many interesting and useful properties—besides being easier to compute.

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.

Variance and standard deviation

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

Remarks:

- The variance (as the expectation of a nonnegative quantity) is always nonnegative, so taking the square root makes sense.
- If $\mathbb{D}^2(X) = 0$, then $(X - \mathbb{E}(X))^2$ is a nonnegative r.v. with expectation 0, hence $(X - \mathbb{E}(X))^2 = 0$ with probability 1, i.e., $\mathbb{P}(X = \mathbb{E}(X)) = 1$.
  The converse is also true, therefore:
  X has variance/standard deviation 0 if and only if X is almost surely constant.

Variance and standard deviation

$\mathbb{D}^2(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$, $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$.
In practice, variance is computed via the following formula (though sometimes the definition is simpler to use):

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

(Proof:

$$\begin{aligned}\mathbb{D}^2(X) &= \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2,\end{aligned}$$

by linearity of expectation.)

A consequence of the above properties:
if $\mathbb{E}(X^2) < \infty$, then $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$, with equality if and only if X is almost surely constant.

Elementary properties of (standard deviation and) variance

Example for the standard deviation: fair die

Elementary properties of (standard deviation and) variance

Example for the standard deviation: fair die

Theorem

If $\mathbb{E}(X^2) < \infty$ *and* $a, b \in \mathbb{R}$, *then*

$$\mathbb{D}^2(aX + b) = a^2\mathbb{D}^2(X).$$

(Proof.)

Elementary properties of (standard deviation and) variance

Example for the standard deviation: fair die

Theorem

If $\mathbb{E}(X^2) < \infty$ *and* $a, b \in \mathbb{R}$, *then*

$$\mathbb{D}^2(aX + b) = a^2\mathbb{D}^2(X).$$

(Proof.)

Corollary

If $\mathbb{E}(X^2) < \infty$ *and* $c \in \mathbb{R}$, *then*

- $\mathbb{D}(cX) = |c|\mathbb{D}(X)$ *(standard deviation is absolutely homogeneous)*
- $\mathbb{D}(X + c) = \mathbb{D}(X)$ *(standard deviation is translation invariant).*

Elementary properties of (standard deviation and) variance

$$\mathbb{D}^2(aX + b) = a^2\mathbb{D}^2(X).$$

Caution: unlike expectation, variance is generally not additive!
E.g., let X be a nonconstant r.v. with existing standard deviation, and let $Y = -X$. Then

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2(X - X) = 0,$$

yet

$$\mathbb{D}^2(X) + \mathbb{D}^2(Y) = 2\mathbb{D}^2(X) \neq \mathbb{D}^2(X + Y).$$

Elementary properties of (standard deviation and) variance

$$\mathbb{D}^2(aX+b) = a^2\mathbb{D}^2(X).$$

Caution: unlike expectation, variance is generally not additive!
E.g., let X be a nonconstant r.v. with existing standard deviation, and let $Y=-X$. Then

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X-X) = 0,$$

yet

$$\mathbb{D}^2(X) + \mathbb{D}^2(Y) = 2\mathbb{D}^2(X) \neq \mathbb{D}^2(X+Y).$$

To come: If X and Y are independent random variables, then $\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
We will first introduce independence for random variables—coming soon.

Theorem

If X and Y are independent random variables with $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, then

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y).$$

- (Proof)
- (Reminder: for non-independent X, Y this is generally false!)

Theorem

If X and Y are independent random variables with $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, then

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y).$$

- (Proof)
- (Reminder: for non-independent X, Y this is generally false!)

(Example: If X, Y are independent, $\mathbb{D}^2(X) = 4$, $\mathbb{D}^2(Y) = 9$, what is $\mathbb{D}(2X + Y + 9)$?)