Statistics: Discrete Random Variables

S. Maset

Dipartimento di Matematica e Geoscienze, Università di Trieste

PEM 2017-2018

Outline

Introduction

- When an experiment is performed, often we are not interested in all the details of the outcome, but rather we are interested only in some numerical quantity determined by the outcome.

  Examples:

  - *In the experiment of the dice, we could be interested only in the sum of the scores and not on the individual scores.*
  - *In the experiment of the 100m run, we could be interested only in the position of the runner at lane* 4*, our favorite runner, and not in the complete arrival order.*
  - *In the experiment of the falling meteor, we could be interested only on the distance of the impact point from our town, and not on the geographical coordinates of the impact point.*

These quantities of interest, that are determined by the outcome of the experiment, are known as random variables.

Definition

Consider an experiment with sample space Ω and a $\sigma-$algebra $\mathcal{F}$ of subset of Ω as the set of the events. A **random variable** for the experiment is a function $X : \Omega \to \mathbb{R}$ such that

(*) for any Borel subset A of $\mathbb{R}$, $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$.

If Ω is discrete, then the condition (*) is trivially satisfied.

If Ω is continuous, then the condition (*) becomes

- for any Borel subset A of $\mathbb{R}$, $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$ is a Borel subset of Ω.

All the functions $X : \Omega \to \mathbb{R}$, Ω continuous, that one encounters in Probability Theory and Statistics satisfy this condition.

By summarizing, we can say that a random variable for an experiment associates a number to each outcome of the experiment.

The random variables considered in the previous example:

- *For the experiment of the dice, the sum of the scores is the random variable*
  $$X : \Omega = \{1, 2, 3, 4, 5, 6\}^2 \to \mathbb{R}$$
  given by
  $$X(\omega) = \omega_1 + \omega_2, \ \omega \in \Omega.$$
- *For the experiment of 100m run, the position of the runner at lane 4 in the arrival order is the random variable*
  $$X : \Omega = \{\omega : \omega \textit{ is a permutation of } 1, 2, 3, 4, 5, 6, 7, 8\} \to \mathbb{R}$$
  given by
  $$X(\omega) = \text{"}\, i \in \{1, 2, 3, 4, 5, 6, 7, 8\} \textit{ such that } \omega_i = 4\text{"}, \ \omega \in \Omega.$$
- *For the experiment of the falling meteor, the distance on the Earth's surface between the impact point and our town, with geographical coordinates* (λ, ϕ)*, is the random variable*
  $$X : \Omega = (-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \to \mathbb{R}$$
  given by (R_E is the Earth's radius)
  $$X(\omega) = R_E \cdot \arccos\left(\sin \omega_2 \sin \phi + \cos \omega_2 \cos \phi \cos |\omega_1 - \lambda|\right), \ \omega \in \Omega.$$

Discrete Random Variables and Probability Mass Functions

Definition

Consider an experiment with sample space Ω. A random variable $X : \Omega \to \mathbb{R}$ for the experiment is called **discrete** if $X(\Omega)$ is a discrete subset of $\mathbb{R}$.
Let X be a discrete random variable for the experiment. The **probability mass function (pmf)** of X is the function

$$f_X : X(\Omega) \to \mathbb{R}$$

given by

$$f_X(x) = \mathbb{P}(X = x), \; x \in X(\Omega).$$

Here, $X = x$ denotes the event "$X(\omega) = x$".

The pmf of X is also called the **distribution** of X.

A random variable for an experiment with a discrete sample space is discrete.

In fact, if

$$\Omega = \{a_i : i \in I\},$$

where $I = \{1, 2, \ldots, n\}$ for some positive integer n or $I = \{1, 2, 3, \ldots\}$, then

$$X(\Omega) = \{X(a_i) : i \in I\}$$

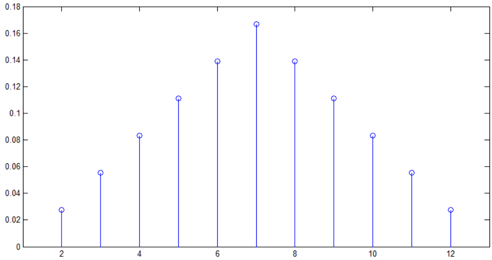
is discrete (we can list the elements of $X(\Omega)$).

For the experiment of the dice, consider the random variable X sum of the scores. We have

$$X(\Omega) = \{2, 3, \ldots, 12\}$$

and the pmf of X is

$$\begin{aligned} f_X(s) &= \mathbb{P}(X = s) = \mathbb{P}(\text{"}X(\omega) = \omega_1 + \omega_2 = s\text{"}) \\ &= \begin{cases} \frac{s-1}{36} \text{ if } s \leq 7 \\ \frac{13-s}{36} \text{ if } s \geq 7 \end{cases}, \ s \in X(\Omega). \end{aligned}$$



For the experiment of the 100m run, consider the random variable X position of the runner at lane 4 in the arrival order. We have

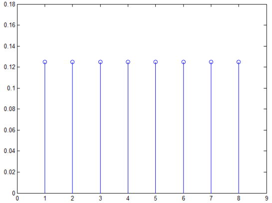
$$X(\Omega) = \{1, 2, 3, 4, 5, 6, 7, 8\}.$$

By assuming that the runners are all of the same strength, and so all the orders of arrival (elementary events) have the same probability, the pmf of X is

$$f_X(k) = \mathbb{P}(X = k) = \mathbb{P}(\text{"}X(\omega) = k\text{"}) = \mathbb{P}(\text{"}\omega_k = 4\text{"})$$

$$= \frac{|\{\text{permutations of } 1,2,3,4,5,6,7,8 \text{ with } 4 \text{ in } k-\text{th position}\}|}{|\Omega|}$$

$$= \frac{|\{\text{permutations of } 1,2,3,5,6,7,8\}|}{|\{\text{permutations of } 1,2,3,4,5,6,7,8\}|} = \frac{7!}{8!} = \frac{1}{8}, \; k \in X(\Omega).$$



For the experiment of the falling meteor, the random variable X distance on the Earth's surface between the impact point and our town is not discrete, since

$$X(\Omega) = [0, \pi R_E],$$

is an interval, where R_E is the Earth's radius.

- Given a discrete random variable X for an experiment of sample space Ω, we have, for the pmf f_X of X,

$$\sum_{x \in X(\Omega)} f_X(x) = 1.$$

In fact,

$$\sum_{x \in X(\Omega)} f_X(x) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{x \in X(\Omega)} X = x\right) = \mathbb{P}(\Omega) = 1.$$

- Here is an interesting interpretation of the pmf of a discrete random variable.

  Consider an experiment with sample space Ω and a discrete random variable $X : \Omega \to \mathbb{R}$.

  Suppose to consider as the new outcome for the experiment $X(\omega)$, rather than ω. So, the new sample space is the discrete set $\Omega^{\text{new}} = X(\Omega)$, rather than Ω.

  The non-negative function $p : \Omega^{\text{new}} \to \mathbb{R}$ giving the probabilities of the elementary events in the new sample space Ω^{new} is the pmf of X:

$$p(x) = \mathbb{P}(X = x) = f_X(x), \; x \in \Omega^{\text{new}} = X(\Omega).$$

- Exercise. In the experiment of flipping a regular coin until Head appears, consider the random variable X giving the remainder of the integer division by 3 of the number of times that T is obtained. Find the pmf f_X of X. Generalize to the integer division by k, where $k \geq 2$ is a positive integer, and check that the sum of the values of the pmf f_X is 1.

Binomial Random Variables

- Consider a Bernoulli process, which is given by n independent trials with the two possible outcomes α and β, with probabilities p and $q := 1 - p$ of obtaining α and β, respectively, at any trial.

  The sample space is
  $$\Omega = \{\alpha, \beta\}^n$$
  and the elementary events have probability
  $$\mathbb{P}(x) = p^k q^{n-k}, \ x \in \Omega,$$
  where k is the number of occurences of α in $x = (x_1, x_2, \ldots, x_n)$.

Now, consider the random variable $X : \Omega \to \mathbb{R}$ given by

$$X(\omega) = \text{"number of occurences of } \alpha \text{ in } \omega = (\omega_1, \omega_2, \ldots, \omega_n)\text{"}$$
$\omega \in \Omega$.

Observe that

$$X(\Omega) = \{0, 1, \ldots, n\}.$$

The pmf of X is

$$f_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$$
$$k \in X(\Omega) = \{0, 1, \ldots, n\}.$$

In fact, for $k \in \{0, 1, \ldots, n\}$, we have

$$
\begin{aligned}
&\mathbb{P}(X = k) \\
&= \mathbb{P}(\{\omega \in \Omega : \text{the number of occurences of } \alpha \text{ in } \omega \text{ is } k\}) \\
&= \mathbb{P}\left(\bigcup_{\substack{J \subseteq \{1,\ldots,n\} \\ |J| = k}} \{\omega \in \Omega : \alpha \text{ occurs in } \omega \text{ in the positions of the set } J\} \right) \\
&= \mathbb{P}\left(\bigcup_{\substack{J \subseteq \{1,\ldots,n\} \\ |J| = k}} \{x_J\} \right)
\end{aligned}
$$

$x_J \in \Omega$ has α in the positions of the set J and β in the others

$$
= \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J| = k}} \mathbb{P}(x_J).
$$

Now, for any $J \subseteq \{1, \ldots, n\}$ such that $|J| = k$, we have

$$\mathbb{P}(x_J) = p^k q^{n-k}.$$

Then

$$\begin{aligned}\mathbb{P}(X = k) &= \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=k}} \mathbb{P}(x_J) \\ &= \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=k}} p^k q^{n-k} \\ &= p^k q^{n-k} \cdot \text{"number of subsets of } \{1, \ldots, n\} \text{ with } k \text{ elements"} \\ &= p^k q^{n-k} \cdot \binom{n}{k}.\end{aligned}$$

- This random variable X related to the Bernoulli process is an example of a binomial random variable.

Definition

A random variable $X : \Omega \to \mathbb{R}$ for an experiment with sample space Ω is said to have the **binomial distribution** $\mathrm{Binomial}(n, p)$, where n is a positive integer and $p \in [0, 1]$, if $X(\Omega) = \{0, 1, \ldots, n\}$ and the pmf f_X di X is

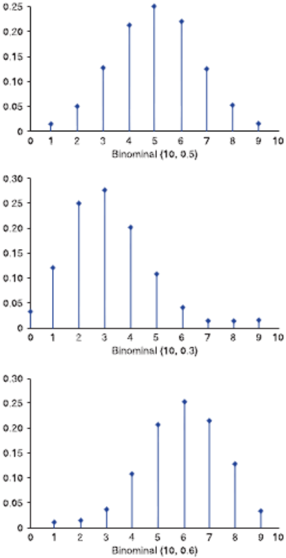
$$f_X(k) = \binom{n}{k} p^k q^{n-k}, \; k \in \{0, 1, \ldots, n\},$$

where $q := 1 - p$.
X is said a **binomial random variable** if it has some binomial distribution $\mathrm{Binomial}(n, p)$.

Note that the distribution $\mathrm{binomial}(n, \frac{1}{2})$, where $p = q = \frac{1}{2}$, is given by:

$$f_X(k) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} 2^{-n}, \; k \in \{0, 1, \ldots, n\}.$$

Examples of binomial distributions.



Exercise. Provide an explanation for the different positions of the peaks in the previous figures of binomial distributions.

In other words: given a random variable X with distribution $\mathrm{binomial}(n, p)$, find an index $k^* \in \{0, 1, \ldots, n\}$ such that $f_X(k)$ increases when k is varying from 0 to k^* and decreases when k is varying from k^* to n.

To this aim, find for which k we have $f_X(k) < f_X(k+1)$ and for which k we have $f_X(k) > f_X(k+1)$.

Exercise. Prove directly by using the expression of the pmf f_X that

$$\sum_{k=0}^{n} f_X(k) = 1$$

for a random variable X with binomial distribution $\mathrm{Binomial}(n, p)$.

Exercise. Let k be a given positive integer and, for any integer $n \geq k$, let X_n be a random variable with binomial distribution $\mathrm{binomial}(n, p)$. Prove that, as a function of n, $f_{X_n}(k)$ increases up to a maximum value and then decreases asymptotically to zero, as $n \to \infty$.
To this aim, find for which n we have $f_{X_n}(k) < f_{X_{n+1}}(k)$ and viceversa. Then, take the limit as $n \to \infty$ of $f_{X_n}(k)$.

Exercise. Consider the Bernoulli process of length n with outcomes α and β at any trial. We have seen that the random variable

$$X(\omega) = \text{"number of occurences of } \alpha \text{ in } \omega\text{"}, \ \omega \in \Omega,$$

has distribution $\mathrm{binomial}(n, p)$, where p is the probability of obtaining α at any trial. What is the distribution of the random variable

$$Y(\omega) = \text{"number of occurences of } \beta \text{ in } \omega\text{"}, \ \omega \in \Omega \ ?$$

- *Example. The Bernoulli process with $p = q = \frac{1}{2}$ includes the experiment of flipping n times a regular coin (or of flipping n regular coins) as well as the situation where a couple of parents decides to have n children.*

  So, the random variable number of Heads in n flips, as well as the random variable number of girls in n children, are binomial random variables X with distribution $\mathrm{binomial}(n, \frac{1}{2})$.

  The pmf of X for $n = 4$ is:

$$\mathbb{P}(X = 0) = \binom{4}{0} \cdot 2^{-4} = \frac{1}{16}, \; \mathbb{P}(X = 1) = \binom{4}{1} \cdot 2^{-4} = \frac{4}{16},$$

$$\mathbb{P}(X = 2) = \binom{4}{2} \cdot 2^{-4} = \frac{6}{16},$$

$$\mathbb{P}(X = 3) = \binom{4}{3} \cdot 2^{-4} = \frac{4}{16}, \; \mathbb{P}(X = 4) = \binom{4}{4} \cdot 2^{-4} = \frac{1}{16}.$$

- *Example from Genetics. Some particular traits (such as eyes color or handedness) of an individual are determined by two genes, a* ***dominant gene*** *d and a* ***recessive gene*** *r.*

  An individual can be:

  - ***pure dominant**, i.e. the individual has two dominant genes d;*
  - ***hybrid**, i.e. the individual has a dominant gene d and a recessive gene r;*
  - ***pure recessive**, i.e. the individual has two recessive genes r.*

  A pure dominant and an hybrid have not the particular trait and are alike in appearance, whereas a pure recessive has the particular trait and she/he is, in some sense, "special".

When two individuals mate, the resulting offspring receives independently and randomly one gene from the mother and one gene from the father.

Is it possible for two non-"special" parents, to have a "special" child? Yes, if both parents are hybrid. In fact, in this case, the child can be pure dominant, hybrid or pure recessive.

Suppose that two hybrid parents have n children. What is the probability that k of these children, $k \in \{0, 1, 2, \ldots, n\}$, are pure recessive?

We can consider the Bernoulli process where each trial is the birth of a child with possible outcomes $\alpha =$"the child is pure recessive" and $\beta =$"the child is not pure recessive".

Since each child receives independently and randomly one gene from the mother and one gene from the father, we have the following table:

Gene from the mother	*Gene from the father*	*Probabilty*
d	d	$\frac{1}{4}$
d	r	$\frac{1}{4}$
r	d	$\frac{1}{4}$
r	r	$\frac{1}{4}$

Thus, we have the probabilities $p = \frac{1}{4}$ *and* $q = \frac{3}{4}$ *in the Bernoulli process.*

The number of pure recessive children is a binomial random variable X *with distribution* $\mathrm{binomial}(n, \frac{1}{4})$*: the pmf is*

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot \left(\frac{1}{4}\right)^k \cdot \left(\frac{3}{4}\right)^{n-k}, k \in \{1, \ldots, n\}.$$

If $n = 4$, the pmf of X is

$$\mathbb{P}(X = 0) = \binom{4}{0} \cdot \left(\frac{3}{4}\right)^4 = 31.64\%$$

$$\mathbb{P}(X = 1) = \binom{4}{1} \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^3 = 42.19\%$$

$$\mathbb{P}(X = 2) = \binom{4}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 21.09\%$$

$$\mathbb{P}(X = 3) = \binom{4}{3} \cdot \left(\frac{1}{4}\right)^3 \cdot \frac{3}{4} = 4.69\%$$

$$\mathbb{P}(X = 4) = \binom{4}{4} \cdot \left(\frac{1}{4}\right)^4 = 0.39\%.$$

Exercise. Suppose that two parents, one hybrid the other pure recessive, have n children. What is the probability that k of these children, $k \in \{0, 1, \ldots, n\}$, are pure recessive?

Mean of a Discrete Random Variable

Definition

Let X be a discrete random variable for an experiment with sample space Ω. The **mean** of X is the quantity

$$\mathbb{E}(X) := \sum_{x \in X(\Omega)} x \cdot f_X(x) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x).$$

Other terms for $\mathbb{E}(X)$ are **expected value** and **expectation.**

If a discrete random variable X has **uniform distribution**, i.e.

$$f_X(x) = \mathbb{P}(X = x) = \frac{1}{|X(\Omega)|}, \; x \in X(\Omega),$$

then $\mathbb{E}(X)$ is the average of the values in $X(\Omega)$:

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot f_X(x) = \sum_{x \in X(\Omega)} x \cdot \frac{1}{|X(\Omega)|} = \frac{\sum_{x \in X(\Omega)} x}{|X(\Omega)|}$$

In the general case,

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot f_X(x)$$

is the weighed average of the values $x \in X(\Omega)$ with weights the values $f_X(x)$ of the pmf of X.

Therefore, $\mathbb{E}(X)$ can be thought as the center of mass of a system of particles lying in a straight line at the positions $x \in X(\Omega)$ with masses $f_X(x)$. Exercise. Explain why.

- In the frequentist interpretation of probability, $\mathbb{E}(X)$ has a clear significance.

  Consider to repeat the underlying experiment a very large number n of times and let $\omega_1^{\text{obs}}, \omega_2^{\text{obs}}, \ldots, \omega_n^{\text{obs}}$ be the observed outcomes.

  We have, for any $x \in X(\Omega)$,

$$\begin{aligned}\mathbb{P}(X = x) &\approx \text{Long Time Relative Frequency of } X = x \\ &= \frac{\left|\left\{i \in \{1, 2, \ldots, n\} : X\left(\omega_i^{\text{obs}}\right) = x\right\}\right|}{n}.\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) \\
&\approx \sum_{x \in X(\Omega)} x \cdot \frac{\left|\left\{i \in \{1, 2, \ldots, n\} : X\left(\omega_i^{\mathrm{obs}}\right) = x\right\}\right|}{n} \\
&= \frac{\sum\limits_{x \in X(\Omega)} x \cdot \left|\left\{i \in \{1, 2, \ldots, n\} : X\left(\omega_i^{\mathrm{obs}}\right) = x\right\}\right|}{n} \\
&= \frac{\sum\limits_{i=1}^{n} X\left(\omega_i^{\mathrm{obs}}\right)}{n}
\end{aligned}$$

Therefore, for a very large n, the mean $\mathbb{E}(X)$ is close to the mean of the data $\boldsymbol{x} = \left(X\left(\omega_1^{\mathrm{obs}}\right), X\left(\omega_2^{\mathrm{obs}}\right), \ldots, X\left(\omega_n^{\mathrm{obs}}\right)\right)$.

By considering in the the probabilities as limits as $n \to \infty$, we have:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) \\
&= \sum_{x \in X(\Omega)} x \cdot \lim_{n \to \infty} \frac{\left|\left\{i \in \{1, 2, \ldots, n\} : X\left(\omega_i^{\text{obs}}\right) = x\right\}\right|}{n} \\
&= \lim_{n \to \infty} \frac{\sum\limits_{x \in X(\Omega)} x \cdot \left|\left\{i \in \{1, 2, \ldots, n\} : X\left(\omega_i^{\text{obs}}\right) = x\right\}\right|}{n} \\
&= \lim_{n \to \infty} \frac{\sum\limits_{i=1}^{n} X\left(\omega_i^{\text{obs}}\right)}{n}
\end{aligned}
$$

So, the sum of the observed values of X grows as $\mathbb{E}(X)n$, as $n \to \infty$:

$$
\sum_{i=1}^{n} X\left(\omega_i^{\text{obs}}\right) \sim \mathbb{E}(X)\, n, \ , \ n \to \infty, \ \text{ i.e. } \lim_{n \to \infty} \frac{\sum\limits_{i=1}^{n} X\left(\omega_i^{\text{obs}}\right)}{\mathbb{E}(X)\, n} = 1.
$$

- Observe that if Ω is discrete, then

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega).$$

This is another formula for the mean, where the sum has indices the outcomes ω of the experiment, rather than the values x of the random variable.

In fact

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}\left(\bigcup_{\substack{\omega \in \Omega \\ X(\omega) = x}} \{\omega\} \right) \\
&= \sum_{x \in X(\Omega)} x \cdot \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} \mathbb{P}(\omega) = \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} x \cdot \mathbb{P}(\omega) \\
&= \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} X(\omega) \cdot \mathbb{P}(\omega) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega).
\end{aligned}$$

- Here are two examples of computation of means of discrete random variables.

Example. Consider the experiment where a die is rolled with sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$*. Consider the random variable*

$$X(\omega) = \text{"score of the die"} = \omega, \ \omega \in \Omega.$$

Since X *has uniform distribution,* $\mathbb{E}(X)$ *is the average of the six values in* $X(\Omega)$*:*

$$\mathbb{E}(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

Note that, despite the name "Expected Value", $\mathbb{E}(X)$ *can not be one of the values of* X*.*

In the frequentist interpretation: $\mathbb{E}(X) = 3.5$ *means that after a very large number* n *of rolls of the die, the average of the observed scores is close to* 3.5*, i.e. the sum of the observed scores grows as* $3.5n$*, as* $n \to \infty$*.*

Example. An insurance company sets the annual premium of its life insurance policies in order to have, for the next year, an expected profit for each police of $c\%$ of the amount it would have to pay out for the die of the individual. A reference value can be $c\% = 0.1\%$.

Find the annual premium a for a policy of value v for an individual that will die during the year with probability p.

Here, the experiment is the future next year with two possible outcomes: "the individual will die during the year" and "the individual will not die during the year".

The profit is the random variable:

$$X(\text{"the individual will die during the year"}) = -v + a$$
$$X(\text{"the individual will not die during the year"}) = a.$$

Thus

$$\mathbb{E}\,(X) = X\,(\text{"the individual will die"}) \cdot \mathbb{P}\,(\text{"the individual will die"})$$
$$+X\,(\text{"the individual will not die"}) \cdot \mathbb{P}\,(\text{"the individual will not die"})$$
$$= (-v + a) \cdot p + a \cdot (1 - p) = -v \cdot p + a.$$

Since it is required that

$$c\% \cdot v = \mathbb{E}(X) = -v \cdot p + a,$$

we obtain

$$a = (p + c\%) \cdot v.$$

By the frequentist interpretation of $\mathbb{E}(X)$, we can conclude that the observed sum of the profits of the sold policies asymptotically grows as $n \cdot c\% \cdot v$, where n is the number of sold policies. So, if n is large, then the insurance company has a large profit.

This is the reason for which insurance companies thrive.

Exercise. A bookmaker (Unibet) offers the win of Napoli in this evening football match Napoli-Juventus at 2.04*, i.e. one has to pay* 1 *now for having* 2.04 *in case of a win of Napoli. By assuming that the expected profit of the bookmaker is* $c\% = 1\%$ *of the amount paid out for a win of Napoli, find the probability that Napoli will win the match.*

Properties of the mean

Theorem

*(**Linearity of the mean**). Let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables. We have*

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

and

$$\mathbb{E}(cX) = c\mathbb{E}(X) \text{ for any } c \in \mathbb{R}.$$

Proof.

Assume Ω discrete. We give the proof only for this case. We have

$$\mathbb{E}(X+Y) = \sum_{\omega\in\Omega} (X(\omega) + Y(\omega)) \cdot \mathbb{P}(\omega)$$
$$= \sum_{\omega\in\Omega} X(\omega) \cdot \mathbb{P}(\omega) + \sum_{\omega\in\Omega} Y(\omega) \cdot \mathbb{P}(\omega) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Proof.

and

$$\mathbb{E}(cX) = \sum_{\omega \in \Omega} cX(\omega) \cdot \mathbb{P}(\omega)$$
$$= c \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega) = c\mathbb{E}(X).$$

□

Consequence: for random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n).$$

Exercise. Prove this.

Observe that if, for some $c \in \mathbb{R}$, we have

$$Y(\omega) = c, \text{ for any } \omega \in \Omega,$$

i.e. Y is a constant random variable, also called a **deterministic random variable**, then

$$\mathbb{E}(Y) = \sum_{x \in Y(\Omega)} x \cdot \mathbb{P}(Y = x) = c \cdot \mathbb{P}(Y = c) = c \cdot 1 = c.$$

Thus, for a random variable X, we have

$$\mathbb{E}(X + c) = \mathbb{E}(X) + c \text{ for any } c \in \mathbb{R}.$$

In fact, let Y be the constant random variable of value c. We have

$$\mathbb{E}(X + c) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mathbb{E}(X) + c.$$

- Here are two examples of use of the formula for the mean of a sum of random variables.

Example. Consider the experiment where n *dice are rolled, whose sample space is* $\Omega = \{1, 2, 3, 4, 5, 6\}^n$, *and consider the random variable*

$$X(\omega) = \text{"sum of the scores"} = \omega_1 + \omega_2 + \cdots + \omega_n, \ \omega \in \Omega.$$

Find the mean of X.

We have

$$X = X_1 + X_2 + \cdots + X_n,$$

where, for $i \in \{1, \ldots, n\}$, X_i *is the random variable*

$$X_i(\omega) = \omega_i = \text{"score of the } i - \text{th die"}, \ \omega \in \Omega.$$

We have

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = 3.5n$$

since X_i, $i \in \{1, 2, \ldots, n\}$, *is the score of a single die and so* $\mathbb{E}(X_i) = 3.5$ *as we have previously seen.*

Exercise. We have just assumed that the pmf of X_i, $i \in \{1, \dots, n\}$, has constant value $\frac{1}{6}$, since X_i is the score of a single die. Prove this by assuming that all the elementary events in $\Omega = \{1, 2, 3, 4, 5, 6\}^n$ have the same probabilities.

Exercise. Suppose that two dice are rolled. What is the mean of the difference of the scores?
Now, suppose the the two dice are rolled n times. What can we say about the sum, over the n rolls, of the differences of the scores when $n \to \infty$.

Exercise. Consider a dice play where three dice are rolled: a red one, a white one and a black one. The score of one roll of the three dice is the sum of the score of the red die, counted two times, the score of the white die, counted one time, and the score of the black die, counted one time but negatively.
Suppose that the three dice are rolled n times. How large is the sum of the scores of the rolls when n is large?

- *Example. Consider a bookmaker that manages n bets on the occurrence of the events $A_1, A_2, \ldots, A_n$ relative to some experiment.*

  Assume that in the $i-$th bet, $i \in \{1, \ldots, n\}$, one pays a_i for having $k_i a_i$, $k_i > 1$, if A_i occurs. The number k_i is called the odd of the i-th bet. Let p_i be the probability of A_i. Find the expected profit of the bookmaker.

  Let X be the random variable profit. We have

$$X = X_1 + X_2 + \cdots + X_n$$

  where X_i, $i \in \{1, 2, \ldots, n\}$, is the profit of the bet on the event A_i.

  Since

$$\mathbb{E}(X_i) = (-k_i a_i + a_i) \cdot p_i + a_i \cdot (1 - p_i) = (1 - k_i p_i) a_i, \ i \in \{1, \ldots, n\},$$

  we have

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} (1 - k_i p_i) a_i.$$

Exercise.

- *By assuming that the bookmaker sets*

$$\frac{1}{k_i} = p_i + c\%, \; i \in \{1, \ldots, n\}, \tag{1}$$

where $c\% > 0$, shows that

$$\mathbb{E}(X) = c\% \cdot \sum_{i=1}^{n} k_i a_i.$$

So the expected profit is $c\%$ of the total paid out by the bookmaker, as in case of the insurance company. Real bookmakers fix the odds by (1).

- *Consider the following k_i, $i \in \{1, 2, 3\}$, for the three events 1X2 relevant to the match Napoli-Juventus:*

$$1 : 2.08, \; X : 3.30, \; 2 : 3.70.$$

Find $c\%$, the probabilities of 1, X and 2 and the expected profit $\mathbb{E}(X)$ of the bookmaker for $a_i = 1$, $i \in \{1, 2, 3\}$.

- We have also the following property of the mean.

Theorem

*(**Monotonicity of the mean**). Let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables. If*

$$X \leq Y, \text{ i.e. } X(\omega) \leq Y(\omega) \text{ for any } \omega \in \Omega,$$

then

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

Proof.

Assume Ω discrete. We give the proof only for this case. We have

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \underbrace{X(\omega)}_{\leq Y(\omega)} \cdot \underbrace{\mathbb{P}(\omega)}_{\geq 0} \leq \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbb{P}(\omega) = \mathbb{E}(Y).$$

□

Indipendence of Discrete Random Variables

- Consider a finite sequence $X_1, X_2, \ldots, X_n$, or an infinite sequence $X_1, X_2, X_3, \ldots$, of discrete random variables for the same experiment with sample space Ω.

  Let I be the set of indices for the sequence: $I = \{1, 2, \ldots, n\}$ if the sequence is finite and $I = \{1, 2, 3, \ldots\}$ if it is infinite.

Definition

The random variables of the sequence X_i, $i \in I$, are called **independent** if for any sequence x_i, $i \in I$, where $x_i \in X_i(\Omega)$ for any $i \in I$, the events of the sequence

$$X_i = x_i, \; i \in I,$$

are independent.

By recalling the definition of independence for the sequence of events $X_i = x_i$, $i \in I$, we can rewrite the definition of independence of the random variables $X_i,\ i \in I$, as follows.

The random variables $X_i,\ i \in I$, are independent if and only if for any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, \ldots, i_{k-1}, i_k \in I$ distinct and for any

$$x_{i_1} \in X_{i_1}(\Omega), \ldots, x_{i_{k-1}} \in X_{i_{k-1}}(\Omega), x_{i_k} \in X_{i_k}(\Omega)$$

such that

$$\mathbb{P}\left(X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_{k-1}} = x_{i_{k-1}}\right) \neq 0,$$

we have

$$\mathbb{P}\left(X_{i_k} = x_{i_k} | X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_{k-1}} = x_{i_{k-1}}\right) = \mathbb{P}\left(X_{i_k} = x_{i_k}\right) = f_{X_{i_k}}(x_{i_k}).$$

In other terms, independence of a sequence of discrete random variables means that knowledge about the values of some of the random variables does not change the pmfs of the others.

Moreover, by recalling the equivalent formulation of the independence of events by probability of an intersection as product of probabilities, we can rewrite the definition of independence of the random variables $X_i,\ i \in I$, as follows.

The random variables $X_i,\ i \in I$, are independent if and only if for any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, \dots, i_k \in I$ distinct and for any

$$x_{i_1} \in X_{i_1}(\Omega), \dots, x_{i_k} \in X_{i_k}(\Omega),$$

we have

$$\begin{aligned} \mathbb{P}\left(X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_1} = x_{i_k}\right) &= \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right) \\ &= f_{X_{i_1}}(x_{i_1}) \cdots f_{X_{i_k}}(x_{i_k}). \end{aligned}$$

- Here are two examples of independence of a sequence of discrete random variables.

Example. Consider a Bernoulli process given by n independent trials with possible outcomes α and β.

Consider, for any $i \in \{1, \dots, n\}$, the random variable

$$X_i(\omega) = \begin{cases} 1 \text{ if } \omega_i = \alpha \\ 0 \text{ if } \omega_i = \beta \end{cases} = \begin{cases} 1 \text{ if the outcome of the } i-\text{th trial is } \alpha \\ 0 \text{ if the outcome of the } i-\text{th trial is } \beta \end{cases}$$

$$\omega \in \Omega = \{\alpha, \beta\}^n.$$

The fact that the trials are independent means that the random variables $X_1, X_2, \dots, X_n$ are independent.

Observe that $X_1 + X_2 + \cdots + X_n$ is the number of outcomes α in the n trials. We known that it has distribution $\mathrm{Binomial}(n, p)$, where p is the probability of α at any trial.

Example. Consider the experiment where n dice are rolled. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}^n$*.*

Consider, for any $i \in \{1, \dots, n\}$*, the random variable*

$$X_i(\omega) = \omega_i = \text{"score of the } i - \text{th die"}, \; \omega \in \Omega.$$

The single rolls are independent and this fact means that the random variables $X_1, X_2, \dots, X_n$ *are independent.*

Clearly $X_1 + X_2 + \cdots + X_n$ *is the total score.*

- As a consequence of the definition of independence we have the following property.

Theorem

Let X_i, $i \in I$, be a sequence of independent discrete random variables. For any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, i_2, \ldots, i_k \in I$ distinct and for any

$$U \subseteq X_{i_1}(\Omega) \times \cdots \times X_{i_k}(\Omega),$$

we have

$$\mathbb{P}\left((X_{i_1}, \ldots, X_{i_k}) \in U\right) = \sum_{(x_{i_1}, \ldots, x_{i_k}) \in U} \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right). \tag{2}$$

Proof.

We have

$$
\begin{aligned}
&\mathbb{P}\left((X_{i_1}, \ldots, X_{i_k}) \in U\right) \\
&= \mathbb{P}\left(\bigcup_{(x_{i_1}, \ldots, x_{i_k}) \in U} X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_k} = x_{i_k} \right) \\
&= \sum_{(x_{i_1}, \ldots, x_{i_k}) \in U} \mathbb{P}\left(X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_k} = x_{i_k}\right) \\
&= \sum_{(x_{i_1}, \ldots, x_{i_k}) \in U} \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right).
\end{aligned}
$$

□

Exercise. Consider the situation where the independent random variables X_i, $i \in I$, have the same distribution, i.e. they have the same range Z, $Z = X_i(\Omega)$ for any $i \in I$, and the same pmf f, $f = f_{X_i}$ for any $i \in I$. Compute $\mathbb{P}\left(\left(X_{i_1}, \ldots, X_{i_k}\right) \in U\right)$ for

$$U = \{(z, \ldots, z) : z \in Z\}.$$

Exercise. Show that in the previous two examples of the Bernoulli process and the roll of the dice the situation of the previous exercise holds. What is the event $\left(X_{i_1}, \ldots, X_{i_k}\right) \in U$ and what is its probability?

Exercise. Consider a Bernoulli process of length $n = 3k$, where k is even. Divide the Bernoulli process in three consecutive parts of length k. Compute the probability that in each part there is an equal number of outcomes α and β. To this aim, define

$$U = \{x \in \{0,1\}^n : \text{ there is an equal number of components 0 and 1 in } (x_1, \ldots, x_k), (x_{k+1}, \ldots, x_{2k}) \text{ and } (x_{2k+1}, \ldots, x_{3k})\}$$

$$V = \{x \in \{0,1\}^k : \text{ there is an equal number of components 0 and 1 in } x\}$$

and show, by using (2), that

$$\mathbb{P}((X_1, \ldots, X_n) \in U) = \mathbb{P}((X_1, \ldots, X_k) \in V) \cdot \mathbb{P}((X_{k+1}, \ldots, X_{2k}) \in V) \cdot \mathbb{P}((X_{2k+1}, \ldots, X_{3k}) \in V).$$

where $X_1, \ldots, X_n$ are the random variables previously defined for the Bernoulli process.

We have also the following other property.

Theorem

Let X_i, $i \in I$, be a sequence of independent discrete random variables. For any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, i_2, \ldots, i_k \in I$ distinct and for any

$$U_{i_1} \subseteq X_{i_1}(\Omega), \ldots, U_{i_k} \subseteq X_{i_k}(\Omega),$$

we have

$$\mathbb{P}(X_{i_1} \in U_{i_1} \cap \cdots \cap X_{i_k} \in U_{i_k}) = \mathbb{P}(X_{i_1} \in U_{i_1}) \cdots \mathbb{P}(X_{i_k} \in U_{i_k}).$$

Proof.

We use $U = U_{i_1} \times \cdots \times U_{i_k}$ in the previous formula (2) and obtain

$$\begin{aligned}
&\mathbb{P}(X_{i_1} \in U_{i_1} \cap \cdots \cap X_{i_k} \in U_{i_k}) \\
&= \mathbb{P}((X_{i_1}, \ldots, X_{i_k}) \in U_{i_1} \times \cdots \times U_{i_k}) \\
&= \sum_{(x_{i_1}, \ldots, x_{i_k}) \in U_{i_1} \times \cdots \times U_{i_k}} \mathbb{P}(X_{i_1} = x_{i_1}) \cdots \mathbb{P}(X_{i_k} = x_{i_k}) \\
&= \sum_{x_{i_1} \in U_{i_1}} \cdots \sum_{x_{i_k} \in U_{i_k}} \mathbb{P}(X_{i_1} = x_{i_1}) \cdots \mathbb{P}(X_{i_k} = x_{i_k}) \\
&= \Big(\sum_{x_{i_1} \in U_{i_1}} \mathbb{P}(X_{i_1} = x_{i_1}) \Big) \cdots \Big(\sum_{x_{i_k} \in U_{i_k}} \mathbb{P}(X_{i_k} = x_{i_k}) \Big) \\
&= \mathbb{P}(X_{i_1} \in U_{i_1}) \cdots \mathbb{P}(X_{i_k} \in U_{i_k}) .
\end{aligned}$$

□

By recalling the formulation of the independence of events by probability of an intersection as product of the probabilities, this last property can be restated as follows.

Theorem

Let X_i, $i \in I$, be a sequence of independent discrete random variables. For any sequence U_i, $i \in I$, where $U_i \subseteq X_i(\Omega)$ for any $i \in I$, the events of the sequence

$$X_i \in U_i, \ i \in I,$$

are independent.

Exercise. Explain why the property in the definition of independent discrete random variables is a particular case of this property.

Exercise. Consider the experiment where we roll n dice. What is the probability that the dice of odd index has an even score and the dice of even index has an odd score?

Operations preserving independence

- We present four operations on a sequence $X_i,\ i \in I$, of independent discrete random variables preserving the independence relationship.

  First operation. Let X_i, $i \in I$, be a sequence of independent discrete random variables and let

$$Y_i = f_i(X_i),\ i \in I,$$

  where $f_i : X_i(\Omega) \to \mathbb{R}$. The discrete random variables Y_i, $i \in I$, are independent.

  In fact, consider a sequence $y_i \in Y_i(\Omega)$, $i \in I$. The events $Y_i = y_i$, $i \in I$, are independent since, for $i \in I$,

$$Y_i = y_i \ \Leftrightarrow\ X_i \in U_i = f_i^{-1}(\{y_i\}) = \{x \in X_i(\Omega) : f_i(x) = y_i\}$$

  and the events $X_i \in U_i$, $i \in I$, are independent.

  So, as an example, if X_i, $i \in I$, are independent, also X_i^2, $i \in I$, are independent.

- Second operation. Let $X_1, \dots, X_n, X_{n+1}, X_{n+2}, \dots$ be a finite or infinite sequence of discrete random variables and let

$$Y = f(X_1, \dots, X_n),$$

where $f : X_1(\Omega) \times \cdots \times X_n(\Omega) \to \mathbb{R}$. The discrete random variables $Y, X_{n+1}, X_{n+2}, \dots$ are independent.

In fact, consider a sequence

$$y \in Y(\Omega), x_{i_1} \in X_{i_1}(\Omega), \dots, x_{i_k} \in X_{i_k}(\Omega),$$

where $i_1, \dots, i_k \in \{n+1, n+2, \dots\}$ are distinct indices. We show that

$$\mathbb{P}(Y = y \cap X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_k} = x_{i_k}) = \mathbb{P}(Y = y)\mathbb{P}(X_{i_1} = x_{i_1}) \cdots \mathbb{P}(X_{i_k} = x_{i_k})$$

To this aim, observe that

$$\mathbb{P}(Y = y \cap X_{i_1} = x_{i_1} \cap \cdots \cap X_{i_k} = x_{i_k}) = \mathbb{P}((X_1, \dots, X_n, X_{i_1}, \dots, X_{i_k}) \in U)$$

where U is the subset of $X_1(\Omega) \times \cdots \times X_n(\Omega) \times X_{i_1}(\Omega) \times \cdots \times X_{i_k}(\Omega)$ given by

$$U = \{(x_1, \dots, x_n, x_{i_1}, \dots, x_{i_k}) : (x_1, \dots, x_n) \in V\}$$
$$V = f^{-1}(\{y\}) = \{(x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega) : f(x_1, \dots, x_n) = y\}.$$

Now

$$\mathbb{P}\left((X_1, \ldots, X_n, X_{i_1}, \ldots, X_{i_k}) \in U\right)$$
$$= \sum_{(x_1,\ldots,x_n)\in V} \mathbb{P}\left(X_1 = x_1\right) \cdots \mathbb{P}\left(X_n = x_n\right) \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right)$$

by the previous formula (2)

$$= \left(\sum_{(x_1,\ldots,x_n)\in V} \mathbb{P}\left(X_1 = x_1\right) \cdots \mathbb{P}\left(X_n = x_n\right) \right) \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right)$$
$$= \mathbb{P}\left((X_1, \ldots, X_n) \in V\right) \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right)$$

by the previous formula (2)

$$= \mathbb{P}\left(Y = y\right) \mathbb{P}\left(X_{i_1} = x_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} = x_{i_k}\right)$$

since $(X_1, \ldots, X_n) \in V \Leftrightarrow Y = y$.

So, as an example, if $X_1, X_2, X_3, X_4, X_5, \ldots$ are independent, also $X_1 + X_2 + X_3, X_4, X_5, \ldots$ and $X_1 \cdot X_2 \cdot X_3, X_4, X_5, \ldots$ are independent.

- Third operation. By the definition of independence for events, independent events

$$A_1, A_2, A_3, \ldots$$

presented here in the order: first A_1, then A_2 and so on, remain independent if they are presented in any other order.

In fact, the notion of independence is related to all the finite sequences

$$A_{i_1}, \ldots, A_{i_{k-1}}, A_{i_k}$$

of events taken in $A_1, A_2, A_3, \ldots$, with not necessarily increasing distinct indices $i_1, \ldots, i_{k-1}, i_k$ arbitrarily chosen in $\{1, 2, 3, \ldots\}$. So, for any order with which the events $A_1, A_2, A_3, \ldots$ are presented, we always consider the same finite sequences.

Therefore, independent discrete random variables $X_1, X_2, X_3, \ldots$ remain independent if they are presented in any other order.

Exercise. Let $X_1, X_2, X_3, X_4, X_5, X_6$ be independent discrete random variables. Show that $(X_1 + X_2)^2, X_3^2 \cdot X_4^2, X_5^2 - X_6^2$ are independent.

- Fourth operation. By the definition of independence for events, a subsequence

$$A_i,\ i \in J \subseteq I,$$

of a sequence $A_i,\ i \in I$, of independent events is a sequence of independent events.

In fact, any finite sequence

$$A_{i_1}, \ldots, A_{i_{k-1}}, A_{i_k}$$

of events taken in $A_i,\ i \in J$, with distinct indices $i_1, \ldots, i_{k-1}, i_k \in J$ is also a finite sequence of events taken in $A_i,\ i \in I$, since $i_1, \ldots, i_{k-1}, i_k \in I$.

Therefore, a subsequence $X_i,\ i \in J \subseteq I$, of a sequence $X_i,\ i \in I$, of independent discrete random variables is a sequence of independent discrete random variables.

So, as an example, if $X_1, X_2, X_3, X_4, X_5, X_6, \ldots$ are independent, also $X_2, X_4, X_6, \ldots$ are independent.

Mean of the product

- Observe that the property

$$\mathbb{E}\,(X \cdot Y) = \mathbb{E}\,(X) \cdot \mathbb{E}((Y)\,,$$

where $X, Y : \Omega \to \mathbb{R}$ are discrete random variables, does not hold in general.

In fact, consider a discrete random variable X with

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Since X has mean 0 and X^2 is a constant random variable of value 1, we have

$$\mathbb{E}\,(X \cdot X) = \mathbb{E}\left(X^2\right) = 1 \neq 0 = \mathbb{E}\,(X)^2 = \mathbb{E}\,(X) \cdot \mathbb{E}\,(X)\,.$$

However, the property holds if X and Y are independent.

Theorem

Let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables. If X and Y are independent, then

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

Proof.

We have

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{z \in (XY)(\Omega)} z \cdot \mathbb{P}(XY = z) \\
&= \sum_{z \in (XY)(\Omega)} z \cdot \mathbb{P}\Big(\bigcup_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ xy = z}} X = x \cap Y = y \quad \Big) \\
&= \sum_{z \in (XY)(\Omega)} z \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ xy = z}} \mathbb{P}(X = x \cap Y = y) \\
&= \sum_{z \in (XY)(\Omega)} \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ xy = z}} \underbrace{z}_{= xy} \mathbb{P}(X = x \cap Y = y) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} xy \cdot \mathbb{P}(X = x \cap Y = y)
\end{aligned}$$

Proof.

Now, since X and Y are independent we have

$$
\begin{aligned}
\mathbb{E}(XY) &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbb{P}(X = x \cap Y = y) \\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbb{P}(X = x)\,\mathbb{P}(Y = y) \\
&= \Big(\sum_{x\in X(\Omega)} x\mathbb{P}(X = x)\Big)\Big(\sum_{y\in Y(\Omega)} y\mathbb{P}(Y = y)\Big) \\
&= \mathbb{E}(X)\cdot\mathbb{E}(Y).
\end{aligned}
$$

□

Consequence: for independent random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}(X_1 X_2 \cdots X_n) = \mathbb{E}(X_1)\,\mathbb{E}(X_2)\cdots\mathbb{E}(X_n).$$

Exercise. Prove this.

- Exercise. Consider a Bernoulli process of lenght n. For $i \in \{1, \ldots, n\}$, let X_i be the random variable with value 1 if the outcome of the $i-$th trial is α and 0 otherwise. Without using the independence of $X_1, X_2, \ldots, X_n$, verify that

$$\mathbb{E}\,(X_1 X_2 \cdots X_n) = \mathbb{E}\,(X_1)\,\mathbb{E}\,(X_2) \cdots \mathbb{E}(X_n).$$

Exercise. Consider the experiment where n dice are rolled. Compute the mean of the random variable product of the scores. Compute the mean of the product between the sum of the scores of the dice of odd index and the sum of the scores of the dice of even index.

Variance

- The mean is a number that we associate to each discrete random variable, whose meaning is clear in the frequentist interpretation.

  Apart this interpretation, we can say that the mean summarizes in a single number (as center of mass) all the possible values that can be assumed by the random variable. But the crucial question is: how are close the values of the random variable to the mean?

  For example, consider these three discrete random variables U, V and W with pmfs

$$\mathbb{P}(U=0)=1$$
$$\mathbb{P}\left(V=-1\right)=\mathbb{P}\left(V=1\right)=\frac{1}{2}$$
$$\mathbb{P}\left(W=-10\right)=\mathbb{P}\left(W=10\right)=\frac{1}{2}.$$

  All three random variables have mean 0, but U surely assumes the value 0 and the two values of V are closer to 0 than the two values of W.

In order to measure how close are the values of a discrete random variable X to the mean $\mu := \mathbb{E}(X)$, we consider the random variable $|X - \mu|$ distance between X and μ.

Then, we could consider, as a measure of closeness of the values of X to μ, the mean $\mathbb{E}(|X - \mu|)$. But, it is matematically more convenient to use the mean of $|X - \mu|^2 = (X - \mu)^2$.

Definition

Let X be a discrete random variable. The **variance** of X is the quantity

$$\mathrm{Var}(X) := \mathbb{E}\left((X - \mu)^2\right).$$

The variance is a measure of how close are the random values of X to μ.

Exercise. By using the monotonicity property of the mean, prove that $\mathrm{Var}(X) \geq 0$ for a discrete random variable X.

For the three previous random variables U, V and W, where $\mu = 0$, the squares U^2, V^2 and W^2 are deterministic random variables of value 0, 1 and 100, respectively, and so

$$\begin{aligned} \mathrm{Var}(U) &= \mathbb{E}\left(U^2\right) = 0, \\ \mathrm{Var}(V) &= \mathbb{E}\left(V^2\right) = 1 \\ \mathrm{Var}(W) &= \mathbb{E}\left(W^2\right) = 100. \end{aligned}$$

- Observe that

$$\mathrm{Var}(X) = \mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}\left(X^2\right) - \mu^2.$$

In fact,

$$\begin{aligned}
\mathbb{E}\left((X-\mu)^2\right) &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\
&= \mathbb{E}(X^2 - 2\mu X) + \mu^2 \\
&= \mathbb{E}(X^2) + \mathbb{E}(-2\mu X) + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}\left(X\right) + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - \mu^2.
\end{aligned}$$

- We have defined the variance as a measure of the closeness of the random values of X to μ, but a better measure is given by the standard deviation.

Definition

Let X be a discrete random variable. The **standard deviation** of X is the quantity

$$\mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)}.$$

Unlike the variance, the standard deviation has the same dimensions of the values of X.

- In order to compute variances and standard deviations, the following formula is useful.

  Let $X : \Omega \to \mathbb{R}$ be a discrete random variable and let $h : \mathbb{R} \to \mathbb{R}$. The random variable $h(X)$ is discrete. Exercise. Why? We have

$$\mathbb{E}\left(h(X)\right) = \sum_{x \in X(\Omega)} h(x) \cdot \mathbb{P}\left(X = x\right).$$

  In fact

$$\begin{aligned}
\mathbb{E}\left(h(X)\right) &= \sum_{y \in h(X)(\Omega)} y \cdot \mathbb{P}\left(h(X) = y\right) \\
&= \sum_{y \in h(X)(\Omega)} y \cdot \mathbb{P}\Big(\bigcup_{\substack{x \in X(\Omega) \\ h(x) = y}} X = x \Big) \\
&= \sum_{y \in h(X)(\Omega)} y \cdot \sum_{\substack{x \in X(\Omega) \\ h(x) = y}} \mathbb{P}\left(X = x\right) = \sum_{y \in h(X)(\Omega)} \sum_{\substack{x \in X(\Omega) \\ h(x) = y}} \underbrace{y}_{= h(x)} \cdot \mathbb{P}\left(X = x\right) \\
&= \sum_{x \in X(\Omega)} h(x) \cdot \mathbb{P}\left(X = x\right).
\end{aligned}$$

Observe that, if Ω is discrete, then

$$\mathbb{E}(h(X)) = \sum_{\omega \in \Omega} h\left(X(\omega)\right) \mathbb{P}\left(\omega\right)$$

by using the formula for the mean of $h(X)$ with the outcomes as indices of the sum.

In particular, for $h(X) = X^2$, we have

$$\mathbb{E}\left(X^2\right) = \sum_{x \in X(\Omega)} x^2 \cdot \mathbb{P}\left(X = x\right).$$

and, if Ω is discrete,

$$\mathbb{E}\left(X^2\right) = \sum_{\omega \in \Omega} X\left(\omega\right)^2 \mathbb{P}(\omega).$$

Moreover, for $h(X) = (X - \mu)^2$, we have

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mu)^2\right) = \sum_{x \in X(\Omega)} (x - \mu)^2 \cdot \mathbb{P}\left(X = x\right). \qquad (3)$$

and, if Ω is discrete,

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mu)^2\right) = \sum_{\omega \in \Omega} \left(X\left(\omega\right) - \mu\right)^2 \mathbb{P}(\omega).$$

Exercise. By using (3), show that the variance of X can be interpreted as a momentum of inertia. In this context, explain why the relation

$$\mathrm{Var}\left(X\right) = \mathbb{E}\left(X^2\right) - \mu^2$$

is the Parallel Axis Theorem.

Exercise. By using (3), show that $\mathrm{Var}(X) = 0$ if and only if $\mathbb{P}\left(X \neq \mu\right) = 0$.

- *As an example of computation of variance and standard deviation, consider the experiment roll of a single die and the random variable X score. We have*

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=1}^{6} k^2 \cdot \mathbb{P}(X = k) \\ &= \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6} = 15.1667\end{aligned}$$

and so

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mu^2 = 15.1667 - 3.5^2 = 2.9167 \\ \mathrm{SD}(X) &= \sqrt{\mathrm{Var}(X)} = \sqrt{2.1967} = 1.7078.\end{aligned}$$

- Exercise. Compute the standard deviation for the random variable profit of the insurance company on its life insurance policy.

The Chebyshev's inequality

- We have seen that the variance of a discrete random variable X is a measure of how close are the random values of X to the mean of X: we can say qualitatively that the smaller the variance, the closer to the mean are the values.

  The **Chebyshev's inequality** specifies quantitatively the closeness of the values of X to the mean in terms of the variance.

Theorem

*(**Chebyshev's inequality**) Let $X : \Omega \to \mathbb{R}$ be a discrete random variable with mean μ. For any $c > 0$, we have*

$$\mathbb{P}\left(|X - \mu| \geq c\right) \leq \frac{\mathrm{Var}\left(X\right)}{c^2}.$$

Fixed $c > 0$, the smaller the variance, the smaller the probability that the value of the random variable is far at least c from the mean

Proof.

Let $c > 0$. Consider the event $|X - \mu| \geq c$, i.e. the event

$$F = \{\omega \in \Omega : |X(\omega) - \mu| \geq c\}$$

and the random variable (**indicator function** of the event F)

$$I_F = \begin{cases} 1 \text{ if } F \text{ occurs, i.e. } \omega \in F \\ 0 \text{ if } F \text{ does not occur, i.e. } \omega \notin F. \end{cases}$$

Note that

$$\mathbb{E}(I_F) = 1 \cdot \mathbb{P}(F) + 0 \cdot \mathbb{P}(F^c) = \mathbb{P}(F).$$

Moreover we have

$$(X - \mu)^2 \geq c^2 I_F.$$

In fact: for $\omega \in F$, we have

$$(X(\omega) - \mu)^2 \geq c^2 = c^2 I_F(\omega)$$

and, for $\omega \notin F$, we have

$$(X(\omega) - \mu)^2 \geq 0 = c^2 I_F(\omega).$$

Proof.

Thus, by the monotonicity property of the mean,

$$\mathrm{Var}(X) = \mathbb{E}\left((X-\mu)^2\right) \geq \mathbb{E}\left(c^2 I_F\right) = c^2 \mathbb{E}(I_F) = c^2 \mathbb{P}(F).$$

We conclude that

$$\mathbb{P}(|X-\mu| \geq c) = \mathbb{P}(F) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

□

- The Chebyshev's inequality can be restated as follows: for any $k > 0$, we have

$$\mathbb{P}\left(|X-\mu| \geq k\mathrm{SD}\left(X\right)\right) \leq \frac{1}{k^2}.$$

In fact

$$\mathbb{P}\left(|X-\mu| \geq k\mathrm{SD}\left(X\right)\right) \leq \frac{\mathrm{Var}\left(X\right)}{\left(k\mathrm{SD}\left(X\right)\right)^2} = \frac{1}{k^2}.$$

So, for any $k > 0$,

$$\mathbb{P}\left(|X-\mu| < k\mathrm{SD}\left(X\right)\right) \geq 1-\frac{1}{k^2}.$$

In fact

$$\mathbb{P}\left(|X-\mu| < k\mathrm{SD}\left(X\right)\right) = 1-\mathbb{P}\left(|X-\mu| \geq k\mathrm{SD}\left(X\right)\right) \geq 1-\frac{1}{k^2}.$$

In particular

$$\begin{aligned}\mathbb{P}\left(|X-\mu| < 2\mathrm{SD}\left(X\right)\right) &\geq \frac{3}{4}\\ \mathbb{P}\left(|X-\mu| < 3\mathrm{SD}\left(X\right)\right) &\geq \frac{8}{9}.\end{aligned}$$

Exercise. For the random variable X score when a single die is rolled, verify that

$$\mathbb{P}\left(|X-\mu| < 2\mathrm{SD}\left(X\right)\right) \geq \frac{3}{4}.$$

Properties of the variance

Theorem

Let $X : \Omega \to \mathbb{R}$ be a discrete random variable. We have

$$\mathrm{Var}\,(cX) = c^2 \mathrm{Var}\,(X) \quad \textit{for any } c \in \mathbb{R}$$

and

$$\mathrm{Var}\,(X + c) = \mathrm{Var}\,(X) \quad \textit{for any } c \in \mathbb{R}.$$

Proof.

For $c \in \mathbb{R}$, we have

$$\begin{aligned}
\mathrm{Var}\,(cX) &= \mathbb{E}\left((cX - \mathbb{E}\,(cX))^2\right) = \mathbb{E}\left((cX - c\mathbb{E}\,(X))^2\right) \\
&= \mathbb{E}\left((c\,(X - \mathbb{E}\,(X)))^2\right) = \mathbb{E}\left(c^2\,(X - \mathbb{E}\,(X))^2\right) \\
&= c^2\mathbb{E}\left((X - \mathbb{E}\,(X))^2\right) = c^2\mathrm{Var}\,(X)
\end{aligned}$$

Proof.

and

$$\begin{aligned}\mathrm{Var}\,(X+c) &= \mathbb{E}\left((X+c-\mathbb{E}\,(X+c))^2\right)\\ &= \mathbb{E}\left((X+c-(\mathbb{E}\,(X)+c))^2\right)\\ &= \mathbb{E}\left((X-\mathbb{E}\,(X))^2\right) = \mathrm{Var}\,(X)\,.\end{aligned}$$

□

Consequence:

$$\mathrm{SD}\,(cX) = |c|\cdot\mathrm{SD}\,(X) \quad \text{for any } c\in\mathbb{R}$$

and

$$\mathrm{SD}\,(X+c) = \mathrm{SD}\,(X) \quad \text{for any } c\in\mathbb{R}.$$

Exercise. Prove this.

- Observe that the property

$$\mathrm{Var}\,(X+Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)\,,$$

where $X, Y : \Omega \to \mathbb{R}$ are discrete random variables, does not hold in general.

In fact, if $\mathrm{Var}\,(X) \neq 0$, then

$$\mathrm{Var}\,(X+X) = \mathrm{Var}\,(2X) = 4\mathrm{Var}\,(X) \neq 2\mathrm{Var}\,(X) = \mathrm{Var}\,(X) + \mathrm{Var}\,(X)\,.$$

However, the property holds if X and Y are independent.

Theorem

Let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables. If X and Y are independent, then

$$\mathrm{Var}\,(X+Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)\,.$$

Proof.

$$
\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}\left((X+Y)^2\right) - \mathbb{E}(X+Y)^2 \\
&= \mathbb{E}\left(X^2 + 2XY + Y^2\right) - \mathbb{E}(X+Y)^2 \\
&= \mathbb{E}\left(X^2\right) + 2\mathbb{E}(XY) + \mathbb{E}\left(Y^2\right) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}\left(X^2\right) + 2\mathbb{E}(XY) + \mathbb{E}\left(Y^2\right) - \left(\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2\right) \\
&= \underbrace{\mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2}_{=\operatorname{Var}(X)} + \underbrace{\mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2}_{=\operatorname{Var}(Y)} + 2\underbrace{(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))}_{=0 \text{ since } X \text{ e } Y \text{ are independent}} \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y).
\end{aligned}
$$

□

Consequence: for independent random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$, we have

$$\operatorname{Var}(X_1 + X_2 + \cdots + X_n) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n).$$

Exercise. Prove this.

- *As an example of use of the formula for the variance of the sum of independent random variables, consider the experiment where n dice are rolled and the random variable X sum of the scores.*

  We have
  $$X = X_1 + X_2 + \cdots + X_n,$$
  where X_i, $i \in \{1, 2, \ldots, n\}$, is the score of the $i-$th die.

  By the independence of $X_1, X_2, \ldots, X_n$, we obtain
  $$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = 2.9167n$$
  $$\mathrm{SD}(X) = 1.7078\sqrt{n},$$
  since $\mathrm{Var}(X_i) = 2.9167$, $i \in \{1, 2, \ldots, n\}$, as we have previously seen.

  By recalling that $\mathbb{E}(X) = 3.5n$, we obtain
  $$\mathbb{P}\left(|X - 3.5n| < 1.7078k\sqrt{n}\right) \geq 1 - \frac{1}{k^2},\ k > 0.$$

  Exercise. In case of $n = 100$, find a lower bound for
  $$\mathbb{P}(300 < X < 400) = \mathbb{P}\left(|X - 350| < 50\right).$$

- Exercise. Derive a formula for the standard deviation of the bookmaker's profit in case of independent events $A_1, A_2, \ldots, A_n$. Can you apply this formula to the events $1X2$ of a football match?

Mean and Variance of a Binomial Random Variable

- Consider a Bernoulli process given by n independent trials, where each trial has the possible outcomes α with probability p and β with probability q.

  We have seen that

$$\begin{aligned} X &= \text{"number of outcomes } \alpha \text{ in the trials"} \\ &= X_1 + X_2 + \cdots + X_n, \end{aligned}$$

  where, for $i \in \{1, \ldots, n\}$,

$$X_i = \begin{cases} 1 \text{ if the outcome of the } i-\text{th trial is } \alpha \\ 0 \text{ if the outcome of the } i-\text{th trial is } \beta. \end{cases}$$

  and that the random variables $X_1, X_2, \ldots, X_n$ are independent.

Note that, for $i \in \{1, \ldots, n\}$,

$$\begin{aligned} \mathbb{E}(X_i) &= 1 \cdot p + 0 \cdot q = p \\ \mathbb{E}\left(X_i^2\right) &= 1^2 \cdot p + 0^2 \cdot q = p \\ \mathrm{Var}(X_i) &= \mathbb{E}\left(X_i^2\right) - \mathbb{E}(X_i)^2 = p - p^2 = pq. \end{aligned}$$

Thus

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = np$$

and, since $X_1, X_2, \ldots, X_n$ are independent,

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = npq.$$

- We know that X has distribution $\mathrm{Binomial}(n, p)$. What about mean and variance of a general discrete random variable with distribution $\mathrm{Binomial}(n, p)$?

  Note that mean and variance of a random variable Y depend only on the pmf f_Y of Y: in fact

$$\mathbb{E}(Y) = \sum_{y \in Y(\Omega)} y \cdot f_Y(y)$$

$$\mathrm{Var}(Y) = \mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2$$

$$= \sum_{y \in Y(\Omega)} y^2 \cdot f_Y(y) - \left(\sum_{y \in Y(\Omega)} y \cdot f_Y(y)\right)^2$$

  Thus, two random variables with the same distribution have the same mean and the same variance.

  We conclude that a random variable Y of distribution $\mathrm{Binomial}(n, p)$ has the same mean and the same variance of the random variable X related to the above Bernoulli process:

$$\mathbb{E}(Y) = np \quad \text{and} \quad \mathrm{Var}(Y) = npq.$$

- *Example. Suppose that a machine produces defective pieces with probability $p = 1\%$. Find the mean and the standard deviation of the number of defective pieces in a shipment of size $n = 1000$.*

  The random variable X number of defective pieces in the shipment has distribution $Binomial(n, p)$ $=Binomial(1000, 1\%)$.

  Thus
  $$\mathbb{E}\,(X) = np = 1000 \cdot \frac{1}{100} = 10$$
  and
  $$\mathrm{SD}\,(X) = \sqrt{npq} = \sqrt{1000 \cdot \frac{1}{100} \cdot \frac{99}{100}} = \sqrt{9.9} = 3.1464.$$
  So
  $$\mathbb{P}\,(|X - 10| < 3.1464k) \geq 1 - \frac{1}{k^2},\ k > 0.$$

Exercise. Given $C\% \in (0,1)$*, find a number* M*, as a function of* $C\%$*, such that*

$$\mathbb{P}\left(|X-10| < M\right) \geq C\%.$$

Compute M *for* $C\% = 75\%$ *and* $C\% = 90\%$*.*

Exercise. By using the fact that X *has distribution* $\mathit{Binomial}\,(1000, 1\%)$ *compute the exact probability*

$$\mathbb{P}(|X-10| < M)$$

for the values of M *computed in the previous exercise for* $C\% = 75\%$ *and* $C\% = 90\%$*.*

- Exercise.
  - Find the mean and the standard deviation of the number X of Heads when a regular coin is flipped n times.
  - Given $C\% \in (0, 1)$, find a number n of flips, as a function of $C\%$, such that

$$\mathbb{P}\left(\left| \frac{\frac{X}{n}}{\frac{1}{2}} - 1 \right| < 0.1 \right) \geq C\%.$$

    Observe that $\frac{X}{n}$ is the percentage of Heads in the n flips of the coin. Compute n for $C\% = 75\%$ and $C\% = 95\%$.