Probability Theory and Statistics
Lecture 4

Bence Csonka

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

September 22

Random Variables

So far: Random experiments $\rightarrow$ outcomes, events, $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

Instead of working with the abstract probability space, it is often simpler to work with real-valued functions.

Definition

A (one-dimensional) random variable is a function $X : \Omega \rightarrow \mathbb{R}$ that assigns to each outcome $\omega \in \Omega$ a real number $X(\omega) \in \mathbb{R}$.

Random Variables

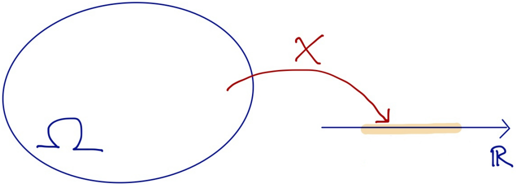
So far: Random experiments $\to$ outcomes, events, $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

Instead of working with the abstract probability space, it is often simpler to work with real-valued functions.

Definition

A (one-dimensional) random variable is a function $X : \Omega \to \mathbb{R}$ that assigns to each outcome $\omega \in \Omega$ a real number $X(\omega) \in \mathbb{R}$.

X has to satisfy certain conditions $\to$ precise definition: coming soon.



Example for random variables

Example: modeling a coin toss, $\Omega = \{H, T\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. $X(H) := 1$, $X(T) := 0$. Then X is a random variable, namely the number of heads obtained in the experiment.

Random Variables

When there are only finitely many possible outcomes, one could often regard any function $X\colon \Omega \to \mathbb{R}$ as a random variable without difficulty. However, this is not what we do, because in more complicated cases of Ω it is important that simple subsets of $\mathbb{R}$ (such as intervals) have *preimages* in Ω that are events (that is, elements of the σ-algebra $\mathcal{F}$).
This motivates the following definitions.

Random Variables

When there are only finitely many possible outcomes, one could often regard any function $X\colon \Omega \to \mathbb{R}$ as a random variable without difficulty. However, this is not what we do, because in more complicated cases of Ω it is important that simple subsets of $\mathbb{R}$ (such as intervals) have *preimages* in Ω that are events (that is, elements of the σ-algebra $\mathcal{F}$).
This motivates the following definitions.

Definition

Let $X\colon \Omega \to \mathbb{R}$ be a function. For $x \in \mathbb{R}$, define

$$\{X < x\} \overset{def.}{=} \{\omega \in \Omega \mid X(\omega) < x\},$$

that is, the set of outcomes $\omega \in \Omega$ for which $X(\omega) < x$. These sets are called the level sets of X.

$$\{X < x\} = X^{-1}((-\infty, x)) \subseteq \Omega.$$

Similarly, we can introduce the following notations (where $x < y$ are real numbers):

$$\begin{aligned}
\{X \leq x\} &= \{\omega \in \Omega \mid X(\omega) \leq x\} = X^{-1}((-\infty, x]),\\
\{X > x\} &= \{\omega \in \Omega \mid X(\omega) > x\} = X^{-1}((x, \infty)),\\
\{X = x\} &= \{\omega \in \Omega \mid X(\omega) = x\} = X^{-1}(\{x\}),\\
\{x < X \leq y\} &= \{\omega \in \Omega \mid x < X(\omega) \leq y\} = X^{-1}((x, y]) \quad \text{etc.}
\end{aligned}$$

In fact, for any set $A \subseteq \mathbb{R}$,

$$\{X \in A\} = \{\omega \in \Omega \mid X(\omega) \in A\} = X^{-1}(A).$$

Using the notation $\{X < x\}$, we can now finally give the definition of a random variable.

Random Variables

Definition

The function $X : \Omega \to \mathbb{R}$ is called a random variable if for every $x \in \mathbb{R}$ we have $\{X < x\} \in \mathcal{F}$.

Random Variables

Definition

The function $X : \Omega \to \mathbb{R}$ is called a random variable if for every $x \in \mathbb{R}$ we have $\{X < x\} \in \mathcal{F}$.

In other words, $X : \Omega \to \mathbb{R}$ is a random variable if each of its level sets is an event.

If X is a random variable, then for any $x < y$, the sets $\{X \leq x\}, \{X \geq x\}, \{X > x\}, \{X = x\}, \{x \leq X \leq y\}, \{x < X < y\}, \{x \leq X < y\}, \{x < X \leq y\}$ are also events.
Countable unions of such sets are also events.

Flip a fair coin three times. Let the sample space Ω be the set of all length-3 heads–tails sequences, and label its elements in the obvious way by the strings $FFF, FIF, \dots$ (with $F =$ heads, $I =$ tails). Define $X : \Omega \to \mathbb{R}$ by $X(FFF) = 0$, and for any other outcome let X be the index of the first occurrence of „tails" (e.g., $X(FIF) = 2$).

(a) What is the probability that X is odd?

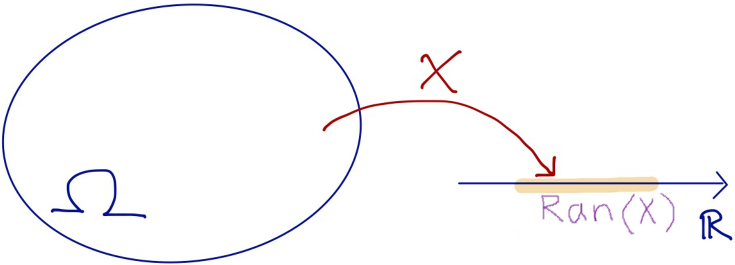
(b) Define Y the same way as X, except that $Y(FFF)$ is randomly either 0 or 1. Is Y a random variable on the sample space Ω?

Random Variables

The range of a random variable X is the set of its possible values:

$$\mathrm{Ran}(X) = \{x \in \mathbb{R} \mid \exists \omega \in \Omega\colon X(\omega) = x\} \subseteq \mathbb{R},$$

just like for any other function.



Discrete Random Variables

Definition

A random variable X is discrete if its range is countable (i.e., finite or countably infinite).

- A non-discrete random variable, for instance, is X „chosen uniformly at random on $[0,1]$" ($\rightarrow$ geometric probability space). We will return to this later.
- The range of X is countably infinite $\Leftrightarrow$ there exists a sequence $(k_1, k_2, \ldots)$ of pairwise distinct real numbers such that
    - $\mathbb{P}(X = k_i) > 0$ for every i, and
    - $\sum_{i=1}^{\infty} \mathbb{P}(X = k_i) = 1$.
- Note: in this case the events $\{X = k_i\}$ form a complete system of events, with the only difference from earlier that there are countably many of them. The Law of Total Probability and Bayes' theorem remain valid for countably many events as well.

Discrete Random Variables: Probability Mass Function

A discrete random variable is most conveniently described by its probability mass function (pmf), which records with what probability the variable takes each possible value.

Definition

Let X be a discrete random variable with range $\mathrm{Ran}(X)$. The probability mass function of X is the map $p_X \colon \mathrm{Ran}(X) \to [0,1]$ defined by

$$x \mapsto p_X(x) \stackrel{def}{=} \mathbb{P}(X = x).$$

Discrete Random Variables: Probability Mass Function

A discrete random variable is most conveniently described by its probability mass function (pmf), which records with what probability the variable takes each possible value.

Definition

Let X be a discrete random variable with range $\mathrm{Ran}(X)$. The probability mass function of X is the map $p_X \colon \mathrm{Ran}(X) \to [0,1]$ defined by

$$x \mapsto p_X(x) \stackrel{def}{=} \mathbb{P}(X = x).$$

- We can extend p_X to a function $\mathbb{R} \to [0,1]$ by setting $p_X(y) = \mathbb{P}(X = y) = 0$ for $y \notin \mathrm{Ran}(X)$.
- Key properties: $p_X(x) > 0$ for all $x \in \mathrm{Ran}(X)$, and $\sum_{x \in \mathrm{Ran}(X)} p_X(x) = \sum_{x \in \mathrm{Ran}(X)} \mathbb{P}(X = x) = 1$.
- The pmf is useful only for discrete random variables! (See later.)

The Simplest Random Variable Whose Value Isn't Random at All

The simplest discrete random variable X is one that is almost surely constant: there exists $x \in \mathbb{R}$ such that

$$\mathbb{P}(X = x) = 1.$$

Thus the value of X is not random: with probability 1 it always equals the single value x.

A rather boring random variable, but it will appear often as a special/degenerate case.

Indicator Variable / Bernoulli Distribution

If $A \in \mathcal{F}$ is an event, then the indicator variable of A is

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Indicator Variable / Bernoulli Distribution

If $A \in \mathcal{F}$ is an event, then the indicator variable of A is

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Since

$$\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\{\omega \in \Omega \mid \mathbb{1}_A(\omega) = 1\}) = \mathbb{P}(\{\omega \mid \omega \in A\}) = \mathbb{P}(A)$$

and similarly

$$\mathbb{P}(\mathbb{1}_A = 0) = \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A),$$

the values of the pmf of $\mathbb{1}_A$ are

$$p_{\mathbb{1}_A}(1) = \mathbb{P}(A), \qquad p_{\mathbb{1}_A}(0) = 1 - \mathbb{P}(A).$$

If $\mathbb{P}(A) = p \in [0, 1]$, then $p_{\mathbb{1}_A}(1) = p$ and $p_{\mathbb{1}_A}(0) = 1 - p$.

Indicator Variable / Bernoulli Distribution

If $A \in \mathcal{F}$ is an event, then the indicator variable of A is

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Since

$$\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\{\omega \in \Omega \mid \mathbb{1}_A(\omega) = 1\}) = \mathbb{P}(\{\omega \mid \omega \in A\}) = \mathbb{P}(A)$$

and similarly

$$\mathbb{P}(\mathbb{1}_A = 0) = \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A),$$

the values of the pmf of $\mathbb{1}_A$ are

$$p_{\mathbb{1}_A}(1) = \mathbb{P}(A), \qquad p_{\mathbb{1}_A}(0) = 1 - \mathbb{P}(A).$$

If $\mathbb{P}(A) = p \in [0,1]$, then $p_{\mathbb{1}_A}(1) = p$ and $p_{\mathbb{1}_A}(0) = 1 - p$.

Definition

If a random variable X has pmf $p_X(1) = p$ and $p_X(0) = 1 - p$, we say that X has the Bernoulli distribution with parameter p.

Binomial Distribution

A Bernoulli distribution is the distribution of a single trial that succeeds with probability $p \in [0, 1]$ (e.g., for $p = 1/2$ success could be „heads” with a fair coin; for $p = 1/6$ success could be rolling a six with a fair die).

Question: if we perform the same trial n times (independently), with what probabilities do we see $0, 1, \ldots, n$ successes?

Binomial Distribution

A Bernoulli distribution is the distribution of a single trial that succeeds with probability $p \in [0, 1]$ (e.g., for $p = 1/2$ success could be „heads" with a fair coin; for $p = 1/6$ success could be rolling a six with a fair die).

Question: if we perform the same trial n times (independently), with what probabilities do we see $0, 1, \ldots, n$ successes?

Let X denote the number of successes; then X is a random variable with $\mathrm{Ran}(X) = \{0, 1, \ldots, n\}$.
To get exactly $k \in \mathrm{Ran}(X)$ successes, we must choose which k of the n trials succeed $\to \binom{n}{k}$ choices.
The probability that those k particular trials succeed and the remaining $n - k$ fail is $p^k(1-p)^{n-k}$, hence

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k \in \{0, 1, \ldots, n\}. \qquad (\star)$$

Binomial Distribution

Definition

A random variable X has the binomial distribution *with parameters $n \in$ and $p \in [0,1]$ if*

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \forall k \in \{0, 1, \ldots, n\}.$$

Notation: $X \sim \mathrm{B}(n; p)$.

Why is this p_X really a pmf?
Nonnegativity is clear. Moreover,

$$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1,$$

where the penultimate equality is the binomial theorem.

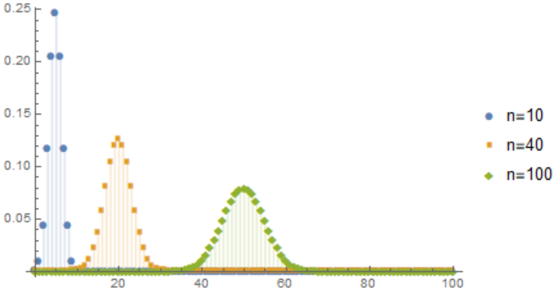
Binomial Distribution

Definition

A random variable X has the binomial distribution with parameters $n \in$ and $p \in [0,1]$ if

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \forall k \in \{0, 1, \ldots, n\}.$$

Notation: $X \sim \mathrm{B}(n; p)$.



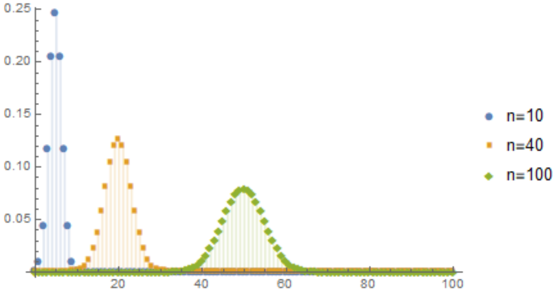
Binomial Distribution

Definition

A random variable X has the binomial distribution *with parameters $n \in$ and $p \in [0, 1]$ if*

$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \forall k \in \{0, 1, \ldots, n\}.$$

Notation: $X \sim \mathrm{B}(n; p)$.



Geometric Distribution

Repeat a trial that succeeds with probability $p \in (0, 1)$ (e.g., toss a biased coin that shows heads with probability p, where „heads” counts as success).

We have seen: after n trials, the number of successes $\sim \mathrm{B}(n; p)$.

Now let X denote the index of the first successful trial (with no cap on the number of repetitions). What is the pmf of X?

Geometric Distribution

Repeat a trial that succeeds with probability $p \in (0,1)$ (e.g., toss a biased coin that shows heads with probability p, where „heads" counts as success).

We have seen: after n trials, the number of successes $\sim \mathrm{B}(n; p)$.

Now let X denote the index of the first successful trial (with no cap on the number of repetitions). What is the pmf of X?

First: $\mathrm{Ran}(X) = \{1, 2, \dots\}$, the set of positive integers. So X is a discrete random variable with a countably infinite range.

For the first success to occur on trial $k \geq 1$, the first $k-1$ trials must fail and the k-th must succeed. By independence,

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1} p, \qquad \forall k \in \{1, 2, \dots\}. \tag{1}$$

Definition

A random variable X has the geometric distribution with parameter $p \in (0,1)$ if (1) holds. Notation: $X \sim \mathrm{Geo}(p)$.

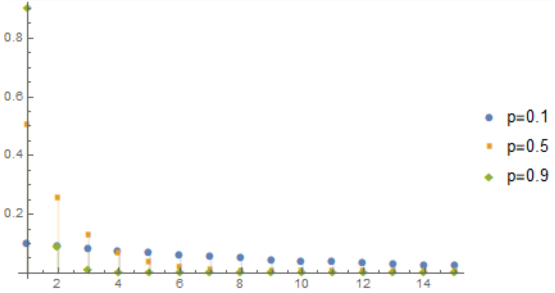
Geometric Distribution

Definition

A random variable X has the geometric distribution *with parameter $p \in (0,1)$ if*

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1}p, \qquad \forall k \in \{1, 2, \ldots\}. \tag{2}$$

Notation: $X \sim \mathrm{Geo}(p)$.



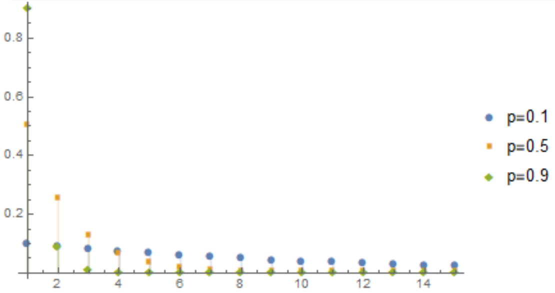
Geometric Distribution

Definition

A random variable X has the geometric distribution *with parameter $p \in (0,1)$ if*

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1}p, \qquad \forall k \in \{1, 2, \ldots\}. \tag{2}$$

Notation: $X \sim \mathrm{Geo}(p)$.



Remarks:

Geometric Distribution: Memorylessness

In a roll-and-move board game we have been waiting for a six for a very long time. We might hope that our next roll is now surely a six. Is this hope justified? What is your experience?

Geometric Distribution: Memorylessness

In a roll-and-move board game we have been waiting for a six for a very long time. We might hope that our next roll is now surely a six. Is this hope justified? What is your experience?

Let X be the index of the first roll of six; then $X \sim \mathrm{Geo}(p)$ with $p = 1/6$. Let $t, s \in \{1, 2, \ldots\}$. The probability that no six appears within the first t trials is

$$\mathbb{P}(X > t) = \mathbb{P}(\text{first } t \text{ trials all fail}) = (1-p)^t.$$

For $s \geq 1$, the probability that no six appears by trial $s+t$, given that none appeared by trial s, is

$$\mathbb{P}(X > s+t \mid X > s) = \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)} = \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t.$$

Thus the two probabilities are equal:

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

In words: if no six has appeared in the first s trials, it is as if the sequence of rolls were just starting now.

Geometric Distribution: Memorylessness

The geometric distribution satisfies the

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \qquad \forall s, t \in \{1, 2, \ldots\} \tag{3}$$

memoryless property.

We have just proved that the geometric distribution is memoryless. A kind of converse also holds:

Theorem

If a non-constant random variable X satisfies memoryless and $\mathrm{Ran}(X) = \{1, 2, \ldots\}$, then $X \sim \mathrm{Geo}(p)$ where $p = \mathbb{P}(X = 1)$.

(Proof)

Among discrete distributions, the geometric is the only memoryless family. In the continuous world we will encounter the exponential distribution, which is also memoryless and closely related to the geometric distribution.

Discrete Uniform Distribution

Definition

A random variable X is (discrete) uniformly distributed on an n-element set $S \subset \mathbb{R}$ if

$$p_X(k) = \mathbb{P}(X = k) = \frac{1}{n}$$

for every $k \in S$.

Examples:

- $n = 6$, $S = \{1, 2, \ldots, 6\}$: X is the outcome of a fair die roll.
- $n = 2$, $S = \{0, 1\}$: X is the outcome of a fair coin toss with the encoding heads$= 1$, tails$= 0$, i.e., $X = \mathbb{1}_{\{\text{heads}\}}$.
  (In general, a Bernoulli distribution is not discrete uniform—only when $p = 1/2$.)
- $n = 1$, $S = \{x\}$: X is the a.s. constant random variable with value x.