Random variables and their distributions

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

So why random variables?

We now have a working understanding of probability.

- This is based on a generic notion of events and (random) outcomes.
- In quantitative approaches we however work with numeric data!

So let's study probabilities of events built around numbers.

(And all the implications...)

Today's outline

Random variables and their distributions

Outline

We want numbers!

In many experiments it is easier to deal with a summary variable than with the original probability structure.

- Say you toss N coins but only care about the total number of heads/tails and not about which coin shows them,
- ... so let us call this number a variable X.
- This simplifies the sample space to the set $\{0, 1, 2, ..., N\}$
- X depends on the outcomes of the experiment
- and is actually a function mapping from the probability space of the experiment to $\{0, 1, 2, ..., N\}$.

Univariate random variable

Definition

Let $\{\mathcal{S}, \mathcal{Y}, \mathrm{P}\}$ be a probability space. If $X : \mathcal{S} \to \mathbb{R}$ (or simply, X) is a real-valued function having as its domain the elements of $\mathcal{S}$, then $X : \mathcal{S} \to \mathbb{R}$ (or X) is called a random variable.

In some experiments random variables are implicitly used:

Experiment	Random variable
Toss two dice	$X =$ sum of the numbers
Toss a coin 50 times	$X =$ number of heads in 50 tosses
Toss a coin 50 times	$X =$ squared number of heads in 50 tosses

Remark on notation

$X(\omega)$	:	denotes the image of $\omega \in \mathcal{S}$ generated by the random variable $X : \mathcal{S} \to \mathbb{R}$.
$x = X(\omega)$	:	(realized) value of the function X

Uppercase letters (X) will be used to denote random variables and corresponding lowercase letters (x) will denote the realized values.

Range of a random variable

Note that by defining a random variable, we have also defined a new sample space, namely, the range of the random variable.

This range, denoted by $\mathrm{R}(X)$, is obtained as the set of all x-values which can be generated on the sample space $\mathcal{S}$ using the function X:

$$\mathrm{R}(X) = \{x : x = X(\omega), \omega \in \mathcal{S}\}.$$

This raises the following important questions:

How can we embed the new sample space $\mathrm{R}(X)$ within a probability space that can be used for assigning probabilities to events in terms of random-variable outcomes?

Hence, what is the probability function on $\mathrm{R}(X)$, say P_X?

Induced probability function

- Suppose we have a discrete sample space

$$\mathcal{S} = \{\omega_1, ..., \omega_n\} \quad \text{with a probability function } \mathrm{P}(\cdot) \; .$$

- Now define a random variable

$$X(\omega) \quad \text{with range } \mathrm{R}(X) = \{x_1, ..., x_m\}.$$

  Assume that we observe $X = x_i$ iff the experiment's outcome is ω_j such that

$$x_i = X(\omega_j).$$

- Since the elementary event $\omega_j \in \mathcal{S}$ is equivalent to the event $x_i \in \mathrm{R}(X)$, both events should have the same probability. Thus

$$\mathrm{P}_X(X = x_i) = \mathrm{P}\left(\{\omega_j : x_i = X(\omega_j), \omega_j \in \mathcal{S}\}\right).$$

Note that the function P_X on the left-hand side is an induced probability set function on $\mathrm{R}(X)$ defined in terms of the original function P.

The eternal coins

Consider the experiment of tossing a fair coin two times.

- Define the random variable X to be the number of heads in the two tosses. Thus

Experiment's outcome $\omega \in \mathcal{S}$	(H,H)	(H,T)	(T,H)	(T,T)
Variable's Realization $x = X(\omega)$	2	1	1	0

- The random variable's range is $\mathrm{R}(X) = \{0, 1, 2\}$
- Since, for example, $\mathrm{P}_X(X = 1) = \mathrm{P}(\{H, T\}) + P(\{T, H\})$, the induced probability function on $\mathrm{R}(X)$ obtains as

x	0	1	2
$\mathrm{P}_X(X = x)$	1/4	1/2	1/4

Outline

Characterizing the probability

Tables are nice, but it is useful to have a representation of the induced probability set function, P_X, in a compact closed-form formula.

- This leads us to the definition of a so-called probability density function.
- They offer a convenient way of conveying the information contained in P_X.

Random variables can be either discrete or continuous. This dichotomy is inherited by the pdfs.

Discrete outcomes

Definition (Discrete random variable)

A random variable X is called discrete iff its range $\mathrm{R}(X)$ is countable.

Definition (Discrete probability density function)

The discrete probability density function (pdf) of a discrete random variable X, denoted by f, is defined by

$$f : \mathbb{R} \to [0,1] \quad \text{such that} \quad f(x) = \begin{cases} \mathrm{P}_X(X = x) & \text{if} \quad x \in \mathrm{R}(X) \\ 0 & \text{else.} \end{cases}$$

- The discrete pdf is also called probability mass function (pmf).
- $\mathrm{R}(X)$ may be countable, but the domain of the pmf is $\mathbb{R}$.
- This convention works for continuous rvs as well, but the discrete pdf is zero “almost everywhere”.

Working with the discrete pdf

The pdf allows us to obtain the probability for an event in $\mathrm{R}(X)$.

- Consider the event $A \subset \mathrm{R}(X)$, written as a union of elementary events $A = \cup_{x \in A}\{x\}$.
- Since elementary events are disjoint, we know from Axiom 1.3 that

$$\mathrm{P}_X(A) = \mathrm{P}_X(\cup_{x \in A}\{X = x\}) \overset{(Ax.3)}{=} \sum_{x \in A} \mathrm{P}_X(x) = \sum_{x \in A} f(x).$$

- Thus, we can use the pdf to calculate probabilities for events on $\mathrm{R}(X)$ by summing the probabilities of the elementary events given by the pdf.

Example: Counting the dots I

Consider the experiment of tossing two fair dice and observing the number of dots facing up.

- The sample space is $\mathcal{S} = \{(i, j) : i = 1, \ldots, 6; j = 1, \ldots, 6\}$, where i, j are the number of dots. $\mathcal{S}$ consists of 36 elementary events.
- Define a random variable X to be the sum of the dots, such that $x = X((i, j)) = i + j$.
- We can derive the pdf of X using elementary arguments.

Example: Counting the dots II

We obtain the following correspondence between outcomes of X and events in $\mathcal{S}$:

$x = X((i,j))$	$B_x = \{(i,j) : x = i + j, (i,j) \in \mathcal{S}\}$	$\mathrm{P}_X(x) = f(x) = \mathrm{P}(B_x)$
2	$\{(1,1)\}$	1/36
3	$\{(1,2),(2,1)\}$	2/36
4	$\{(1,3),(2,2),(3,1)\}$	3/36
	$\vdots$	
12	$\{(6,6)\}$	1/36

- Consider the event $X \in \{3,4\}$. The probability is given as $\mathrm{P}_X(A) = \sum_{x \in A} f(x) = f(3) + f(4) = 5/36$.
- A compact algebraic form for the pdf f is $f(x) = \frac{6-|x-7|}{36} \mathbb{I}_{\{2,3,\ldots,12\}}(x)$.

Continuous distributions

Definition (Continuous random variable)

A random variable X is called continuous iff its range $\mathrm{R}(X)$ is not countable.

Problem:

- The range $\mathrm{R}(X)$ is continuous with events A defined as intervals in $\mathrm{R}(X) \subset \mathbb{R}$
- But can't use summation to add uncountably many probabilities!

(Heuristic) Solution: Substitute the summation operation $\sum_{x \in A}$ by integration $\int_{x \in A}$.

The "genuine" probability density function

Definition (Continuous probability density function)

A random variable X is called continuous iff

- its range $\mathrm{R}(X)$ is uncountably infinite and
- there exists a function

$$f : \mathbb{R} \to [0, \infty) \quad \text{such that for any event } A, \quad \mathrm{P}_X(A) = \int_{x \in A} f(x)\mathrm{d}x$$

and

$$f(x) = 0 \;\forall\; x \notin \mathrm{R}(X).$$

The function f is called a continuous probability density function.

Cars (& Laplace)

Consider a Formula 1 circuit of 10 km. Suppose that accidents are equally likely to occur at each point of the circuit.

So define the continuous random variable X to be
the point of a potential accident with range $\mathrm{R}(X) = [0, 10]$.

In order to obtain the pdf for X,

- consider the event A of an accident between two points a and b, such that $A = [a, b]$.
- Since all points are equally likely, $\mathrm{P}_X(A) = \frac{\text{length of } A}{\text{length of } \mathrm{R}(X)} = \frac{b-a}{10}$.

... and their accidents

According to the definition, the pdf f for X has to satisfy

$$\int_{x \in A} f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x \overset{!}{=} \mathrm{P}_X(A) = \frac{b-a}{10}, \quad \forall \quad 0 \leq a \leq b \leq 10,$$

with

$$\frac{\partial[\int_a^b f(x)\mathrm{d}x]}{\partial b} = f(b) \overset{!}{=} \frac{\partial[\frac{b-a}{10}]}{\partial b} = \frac{1}{10}, \quad \forall \quad b \in [0, 10].$$

Hence,

- the function $f(x) = \frac{1}{10}\mathbb{I}_{[0,10]}(x)4$ can be used as a pdf for X,
- and for any event A on $\mathrm{R}(X)$ we obtain $\mathrm{P}_X(A) = \int_{x \in A} \frac{1}{10}\mathrm{d}x$.
- E.g., the probability for $X \in A = [0, 5]$ is $\mathrm{P}_X(A) = \int_0^5 \frac{1}{10}\mathrm{d}x = 1/2$.

Singletons

The definition of the continuous pdf implies that the probability for an elementary event $A = \{a\}$ is zero, since

$$\mathrm{P}_X(A) = \int_a^a f(x)\mathrm{d}x = 0.$$

Still, some outcome *will* occur!

We may interpret this to mean that A is 'relatively impossible', relative to all other outcomes that can occur in $\mathrm{R}(X) \setminus A$.

E.g., since $\{a\}, \{b\}$ and (a, b) are disjoint and $\mathrm{P}_X(\{a\}) = \mathrm{P}_X(\{b\}) = 0$,

$$\mathrm{P}_X([a, b]) = \mathrm{P}_X((a, b]) = \mathrm{P}_X([a, b)) = \mathrm{P}_X((a, b)) = \int_a^b f(x)\mathrm{d}x.$$

Some comparisons

The interpretation of the function value of a continuous pdf $f(x)$ is fundamentally different from that of a discrete pdf:

- If f is discrete, $f(x) = \mathrm{P}_X(x) =$ probability of the outcome x.
- If f is continuous, $f(x)$ is not the probability of outcome x, which is $\mathrm{P}_X(x) = 0$. (If $f(x)$ was a probability, we would have $f(x) = 0 \ \forall x$.)
- For a unified interpretation, imagine the discrete pdf as having point probability mass.[1]

[1] We'll discuss this later in the course but we need some additional motivation, so please be patient for now.

Requirements for pdfs

Pdfs should be such that the probabilities obtained from f adhere to the probability axioms.

Definition (Class of discrete pdfs)

The function $f : \mathbb{R} \to \mathbb{R}$ is a member of the class of discrete pdfs iff

(i_a) the set $C = \{x : f(x) > 0, x \in \mathbb{R}\}$ is countable;

(ii_a) $f(x) = 0 \ \forall \ x \in \bar{C}$;

(iii_a) $\sum_{x \in C} f(x) = 1$.

Definition (Class of continuous pdfs)

The function $f : \mathbb{R} \to \mathbb{R}$ is a member of the class of continuous pdfs iff

(i_b) $f(x) \geq 0 \ \forall \ x \in \mathbb{R}$;

(ii_b) $\int_{x \in \mathbb{R}} f(x) \mathrm{d}x = 1$.

Some checks

1) Consider the function $f(x) = (0.3)^x\,(0.7)^{1-x}\mathbb{I}_{\{0,1\}}(x)$. Can this f serve as pdf?

   Since (i) $f(x) > 0$ on the countable set $\{0,1\}$, and (ii) $\sum_{x=0}^{1} f(x) = 1$, and (iii) $f(x) = 0\ \forall\ x \notin \{0,1\}$, the function f can serve as a pdf.

2) Consider the function $f(x) = (x^2+1)\mathbb{I}_{[-1,1]}(x)$. Can this f serve as pdf?

   While $f(x) \geq 0\ \forall\ x \in \mathbb{R}$, f does not integrate to 1:

   $$\int_{\mathbb{R}} f(x)\mathrm{d}x = \int_{-1}^{1}(x^2+1)\mathrm{d}x = \frac{8}{3} \neq 1.$$

   Thus, f can not serve as a pdf. (Normalization gets us from f to a function which can serve as a pdf)

Outline

Another description of probability

Definition (Cumulative distribution function)

The cumulative distribution function (cdf) of a random variable X, denoted by F, is defined by

$$F : \mathbb{R} \to [0, 1] \quad \text{such that} \quad F(b) = \mathrm{P}_X(X \leq b), \quad \forall b \in \mathbb{R}.$$

For a discrete random variable the cdf is obtained as

$$F(b) = \sum_{x \leq b} f(x), \quad \forall b \in \mathbb{R},$$

and for a continuous random variable as

$$F(b) = \int_{-\infty}^{b} f(x) \mathrm{d}x, \quad \forall b \in \mathbb{R}.$$

A continuous example

Let the random variable X be the duration of a telephone call (in min), with range $\mathrm{R}(X) = \{x : x > 0\}$.

- Let the pdf be: $f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \cdot \mathbb{I}_{(0,\infty)}(x)$, with $\lambda > 0$.
- The cdf is then $F(b) = \int_0^b \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \mathrm{d}x = (1 - e^{-\frac{b}{\lambda}}) \cdot \mathbb{I}_{(0,\infty)}(b)$
- Assume that $\lambda = 100$ (average duration). Then the probability that the duration is less than 50 min is: $F(50) = 1 - e^{-\frac{50}{100}} = 0.39$.

A discrete example

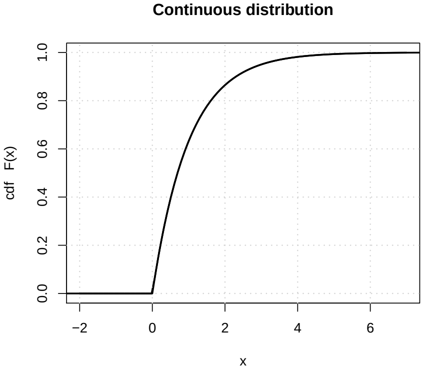
Let the random variable X be the number of dots observed rolling a die, with range $\mathrm{R}(X) = \{1, 2, \ldots, 6\}$.

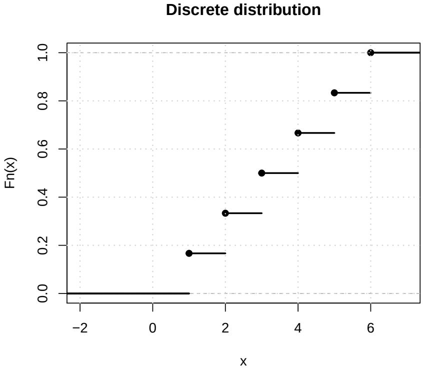
- The pdf is: $f(x) = \frac{1}{6} \cdot \mathbb{I}_{\{1,\ldots,6\}}(x)$.
- The cdf is obtained as:

$$F(b) = \sum_{x \leq b} \frac{1}{6} \cdot \mathbb{I}_{\{1,\ldots,6\}}(x) = \frac{1}{6} \lfloor b \rfloor \cdot \mathbb{I}_{[1,\ldots,6]}(b) + \mathbb{I}_{(6,\infty)}(b)$$

($\lfloor b \rfloor$ denotes the integer part of the number b) – see following figure.

Cdfs of continuous vs. discrete RVs





(And we may have mixtures of the two – nothing to be scared of.)

Properties

Theorem (2.1)

For any cdf F, we have that

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

(ii) $F(x)$ *is a non decreasing function on* x*; that is,* $F(a) \leq F(b)$ *for* $a < b$;

(iii) $F(x)$ *is right-continuous; that is,* $\lim_{h\downarrow 0} F(x+h) = F(x)$.

Relation to pdfs

Theorem (2.2)

Let $x_1 < x_2 < x_3 < \cdots$ be the countable set of outcomes in the range of the discrete random variable X. Then the pdf for X obtains as

$$f(x_i) = \begin{cases} F(x_i), & i = 1 \\ F(x_i) - F(x_{i-1}), & i = 2, 3, \ldots \\ 0, & x \notin \mathrm{R}(X). \end{cases}$$

Theorem (2.3)

Let $f(x)$ and $F(x)$ denote the pdf and cdf of a continuous random variable X. Then the pdf for X obtains as

$$f(x) = \begin{cases} \frac{\mathrm{d}F(x)}{\mathrm{d}x}, & \textit{wherever } f(x) \textit{ is continuous} \\ 0, & \textit{elsewhere.} \end{cases}$$

Jumps & co.

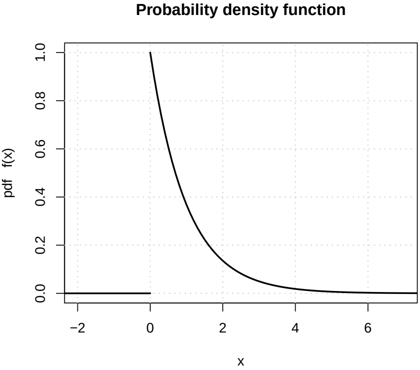
Recall X the duration of a telephone call, with cdf

$$F(x) = (1 - e^{-\frac{x}{\lambda}}) \cdot \mathbb{I}_{(0,\infty)}(x).$$

A pdf for X is given by

$$f(x) = \begin{cases} F'(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & x > 0 \\ 0 \quad \text{(say)} & x = 0 \\ 0 & x < 0 \end{cases}$$

Note the (rather arbitrary) choice at $x = 0$.



Outline

Coming up

Multivariate random variables