MS-A0503 First course in probability and statistics

1B Random variables and distributions

Jukka Kohonen

Deparment of mathematics and systems analysis
Aalto SCI

Academic year 2022–2023
Period III

Contents

Random variable

A random variable (abbrev. rv) is a quantity whose value is determined by the outcome of a random experiment.

- One outcome $s \in S$ *occurs* at random
- Then the outcome s *determines* $X(s)$
- Event $\{X = a\} := \{s \in S : X(s) = a\}$
- The events $\{X = a\}$, for all possible values a, make a *partition* of the sample space. $\longrightarrow$ BLACKBOARD

Example (Two rolls of a die)

Sample space $S = \{(s_1, s_2) : s_1, s_2 = 1, \ldots, 6\}$

- Sum of the two results $N(s) = s_1 + s_2$ is a random variable
- Their maximum $M(s) = \max(s_1, s_2)$ is a random variable
- The first result $X_1(s) = s$ is also a random variable!

Random variables: Theory and practice

Observe the two steps of abstraction:

- The sample space S and its probability function $\mathbb{P}$ describe all of the *randomness* in the situation.
- Once the (random) outcome occurs, it determines all the "random variables" we have defined.

In one random experiment (e.g. "deal one card" or "deal five cards") we may define any number of random variables: whatever functions of the outcome s we are interested in.

Mathematically, a random variable X is a (deterministic) *function* from outcomes s to values $X(s)$.

In practice, we just write X for the (random) value, and e.g. $\{X = 5\}$ or $X = 5$ for the event that it happens to be 5. And $P(X = 5)$ for the probability of this event.

Different kinds of random variables

Typically the values $X(s)$ are real numbers, but they can be something else. Here are some examples.

Name	**Target set**	**Explanation or example**
Random number	$\mathbb{R}$	
Random vector	$\mathbb{R}^n$	$(X_1, X_2, \ldots, X_{10})$ from 10 dice rolled; or (Min,Max) of the dice; now $n = 2$
Random matrix	$\mathbb{R}^{m \times n}$	
Random string	A^n	Random DNA sequence ($A = \{$A,C,T,G$\}$)
Stochastic process	$\mathbb{R}^I$	Real-valued functions on time interval I
Random graph	$\{0,1\}^{V \times V}$	Graphs on vertex set V

On this course we focus in random numbers in $\mathbb{R}$ and random vectors in $\mathbb{R}^2$.

A random variable is discrete if it takes values in a *finite set* like $\{1, 2, 3, 4\}$, or in a *countably infinite* set like "positive integers".

Contents

Distribution

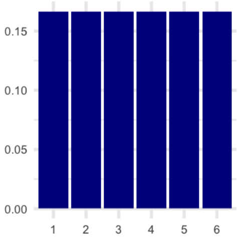
The distribution of a random variable X is a table or a function that determines its possible values and their probabilities.

Example (Two rolls of a die)

The first result X_1 has distribution

k	1	2	3	4	5	6
$\mathbb{P}(X_1 = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

i.e. uniform distribution on $\{1, \ldots, 6\}$.



The second result X_2 has the same *distribution*, **but** it is not the *same* variable! It may well happen, when you roll the dice, that $X_1 \neq X_2$.

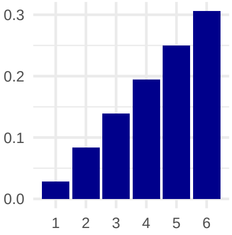
Example. Maximum of two dice

$M = \max(X_1, X_2)$, where X_1 ja X_2 are two rolls of a die.

$$
\begin{aligned}
\mathbb{P}(M = 4) &= \mathbb{P}(M \le 4) - \mathbb{P}(M \le 3) \\
&= \mathbb{P}(X_1 \le 4 \textbf{ and } X_2 \le 4) - \mathbb{P}(X_1 \le 3 \textbf{ and } X_2 \le 3) \\
&= \mathbb{P}(X_1 \le 4) \times \mathbb{P}(X_2 \le 4) - \mathbb{P}(X_1 \le 3) \times \mathbb{P}(X_2 \le 3) \\
&= \left(\frac{4}{6}\right)^2 - \left(\frac{3}{6}\right)^2 \\
&= \frac{16 - 9}{36} = \frac{7}{9}.
\end{aligned}
$$

Other values similarly, so distribution of M:

k	1	2	3	4	5	6
$\mathbb{P}(M = k)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$



New random variables from old

From one or more random variables, you can define a *new* random variable by some rule (function).

Example (A sample statistic)

Three die results X_1, X_2, X_3, their maximum $Y = \max(X_1, X_2, X_3)$, minimum $Z = \min(X_1, X_2, X_3)$. Let us define one more random variable $W = Y - Z$, which tells how *widely* the results were scattered.

If e.g. results were $(4, 2, 5)$, then $Y = 5$, $Z = 2$, and further $W = 5 - 2 = 3$.

- W is now also a *random variable*, with some distribution, i.e. possible values and their probabilities.
- Such a number, computed from data, is called a statistic.
- Such numbers (statistics) are often used to *describe* some properties of the data.
- Another example of a statistic is the *average* of the data.

New from old: Transformation of one random variable

Example (Square of random size)

A machine produces square-shaped tiles, whose side length X is determined (as if) by rolling a die.

i	1	2	3	4	5	6
$\mathbb{P}(X = i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The *area* of a tile $A = X^2$ is also a r.v. What is its distribution? Find out (1) what values X^2 may possibly take and (2) with what probability.

a	?	?	?	?	?	?
$\mathbb{P}(A = a)$	?	?	?	?	?	?

New from old: Transformation of one random variable

Example (Square of random size)

A machine produces square-shaped tiles, whose side length X is determined (as if) by rolling a die.

i	1	2	3	4	5	6
$\mathbb{P}(X = i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The *area* of a tile $A = X^2$ is also a r.v. What is its distribution? Find out (1) what values X^2 may possibly take and (2) with what probability.

a	1	4	9	16	25	36
$\mathbb{P}(A = a)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(More about transformations on Lecture 2A.)

Example. Waiting time for the metro

X = waiting time for the next metro (in minutes, as a real number), where trains arrive at 10-minute intervals. What is the distribution of X?

- $\mathbb{P}(2 \le X \le 3) \;=\; \frac{1}{10} \;=\; 0.1$
- $\mathbb{P}(2.9 \le X \le 3) \;=\; \frac{0.1}{10} \;=\; 0.01$
- $\mathbb{P}(2.999999 \le X \le 3) \;=\; \frac{0.000001}{10} \;=\; 0.0000001$
- $\mathbb{P}(X = 3) \;=\; 0$

Similarly we deduce that $\mathbb{P}(X = t) \;=\; 0$ for all t.

Did calculate something wrong?

No we didn't. Because the X takes values on the *continuous* interval $[0, 10]$, the event $\{X = 3\}$ means that X equals 3 with infinite precision. Surely this is very unlikely (indeed, has probability zero).

The distribution of X must be characterized in some other way.

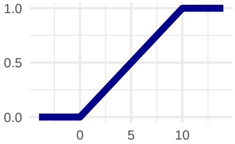
Example. Waiting time for the metro

X = waiting time for the next metro (in minutes, as a real number), where trains arrive at 10-minute intervals. Probabilities of single values are not useful here. Instead, we define probabilities of intervals.

$$\begin{aligned}\mathbb{P}(a \le X \le b) &= \mathbb{P}(a < X \le b) \\ &= \mathbb{P}(X \le b) - \mathbb{P}(X \le a) \\ &= F_X(b) - F_X(a),\end{aligned}$$

where

$$F_X(t) = \begin{cases} 0, & t \le 0, \\ \frac{t}{10}, & 0 < t < 10, \\ 1, & t \ge 10. \end{cases}$$



is the cumulative distribution function for the distribution of X.

Cumulative distribution function (CDF)

The cumulative distribution function (abbrev. CDF) of a random number is $F_X(t) = \mathbb{P}(X \leq t)$.

Fact

The CDF is enough to determine the distribution completely. From it, we can compute the probabilities of all events $\{X \in B\}$.

Example (Metro waiting time)

With what probability is X either in $[1, 2]$ or in $[3, 4]$?

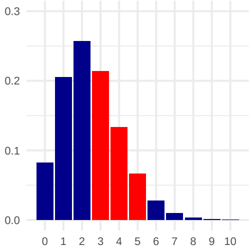
$$\begin{aligned}
\mathbb{P}(X \in [1,2] \textbf{ or } X \in [3,4]) &= \mathbb{P}(X \in [1,2]) + \mathbb{P}(X \in [3,4]) \\
&= (F_X(2) - F_X(1)) + (F_X(4) - F_X(3)) \\
&= \left(\frac{2}{10} - \frac{1}{10}\right) + \left(\frac{4}{10} - \frac{3}{10}\right) \\
&= 0.2.
\end{aligned}$$

Contents

Density function

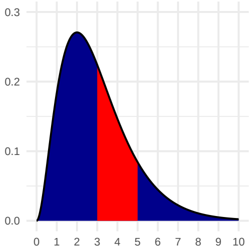
X is discrete, if its distribution can be characterized by a function $f_X(x) \geq 0$ such that

$$\mathbb{P}(X \in A) \;=\; \sum_{x \in A} f_X(x).$$



X is continuous, if its distribution can be characterized by a function $f_X(x) \geq 0$ such that

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



In both cases we can call f_X the (probability) density (function) of X. (Abbreviate PDF.) The subscript X is often dropped.

Density of a discrete distribution

The density function of a discrete distribution is simply

$$f_X(x) \;=\; \mathbb{P}(X = x)$$

and it fulfills conditions

$$f_X(x) \geq 0 \qquad \text{and} \qquad \sum_x f_X(x) \;=\; 1.$$

Also, *any* function that fulfills the above, *is* indeed the density of a discrete distribution.

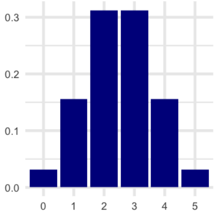
A discrete density function is also called probability mass function (PMF) (= function determining the *mass* of probability at each point).

Density of a discrete distribution

If X takes few different values, its distribution can be presented as a table of the values and their probabilities.

Example (Number of heads from 5 coin tosses)

k	0	1	2	3	4	5
$\mathbb{P}(X = k)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$



For a large target set, a *functional expressions* is more convenient.

Example (Number of heads from $n = 5\ 000\ 000$ coin tosses)

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

This is the so-called binomial distribution with parameters $n = 5\ 000\ 000$ and $p = \frac{1}{2}$.

Density of a continuous distribution

The density of a continuous distribution fulfills

$$f_X(x) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} f_X(x)\, dx \;=\; 1,$$

Also, any function fulfilling those conditions *is* indeed the density of some continuous distribution.

A continuous density at point x does *not* represent the probability of the event $\{X = x\}$. (That probability is zero!).

Instead, if f_X is continuous at x, then $f_X(x)$ approximates the probability of any *small interval* around x, in *proportion* to the interval length. For a small $h > 0$, we have

$$\mathbb{P}(X = x \pm h/2) \;\approx\; f_X(x) \cdot h$$

Note: Density can be arbitrarily large (much bigger than 1), but only over a short interval (why?).

CDF $\leftrightarrow$ PDF

For a continuous distribution,

- CDF is the *integral* of density

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

- density is the *derivative* of CDF

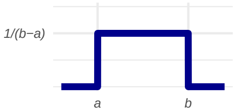
$$f_X(x) = F_X'(x)$$

at the points where the density function is continuous.

Example. Continuous uniform distirbution

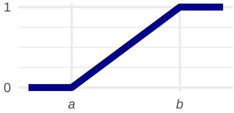
The function

$$f(t) \;=\; \begin{cases} \frac{1}{b-a}, & a < t < b, \\ 0, & \text{otherwise}, \end{cases}$$



is the density of a continuous distribution, namely the (continuous) uniform distribution over the interval $[a, b]$.
The CDF can be calculated as

$$F(t) \;=\; \int_{-\infty}^{t} f(s)\,ds \;=\; \begin{cases} 0, & t < a, \\ \frac{t}{b-a}, & a \le t \le b, \\ 1, & t > b. \end{cases}$$

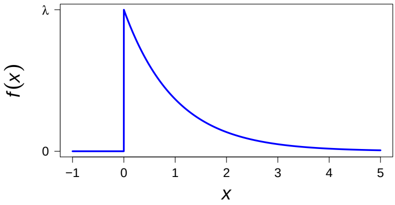


The constants a and b are parameters of the distribution. When you fix their values, you get a particular uniform distribution. (With $a = 0$ and $b = 10$ you get our metro waiting time distribution.)

Another example. Exponential distribution

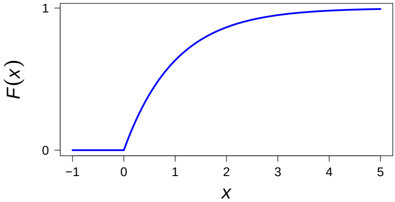
The exponential distribution with parameter $\lambda > 0$ has density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$



By integrating the density, we get the CDF

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$



This is typically used for the waiting time of an event, that has a constant probability λh of happening in any small interval h of time (if it did not happen yet). E.g. insects hitting windshield, or radioactive particles decaying.

Exponential distribution: Memorylessness

$F(t) = 1 - e^{-\lambda t},\ t \geq 0$

$$
\begin{aligned}
\mathbb{P}(X > s + t \,|\, X > s) &= \frac{\mathbb{P}(X > s + t \quad \textbf{and} \quad X > s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)
\end{aligned}
$$

Thus $\mathbb{P}(X > s + t \,|\, X > s) = \mathbb{P}(X > t)$ for all $s, t \geq 0$.

Interpretation

It does not matter whether we have driven 0 or 5 km after the previous insect hitting the windshield. In both cases we have the same probability of getting another insect in the next e.g. 1 kilometer.

Random numbers — summary

Discrete distribution	**Continuous distribution**
X takes values in a finite set or countably infinite	X takes values continuously in an uncountably infinite set
$\mathbb{P}(X = x) = f_X(x)$	$\mathbb{P}(X = x) = 0$
Density gives probabilities by $$\mathbb{P}(X \in A) \; = \; \sum_{x \in A} f_X(x)$$	Density gives probabilities by $$\mathbb{P}(X \in A) \; = \; \int_A f_X(x)\, dx$$
Density values are probabilities $$f_X(x) \; = \; \mathbb{P}(X = x)$$	Density values are proportional approximate probabilities $$f_X(x) \; \approx \; h^{-1}\mathbb{P}(X = x \pm h/2)$$
E.g. uniform distribution in the set $\{1, \ldots, 6\}$	E.g. uniform distribution over the interval $[0, 10]$

Contents

Joint distribution of two variables 1 (Dice)

The joint distribution of a pair random variables (X, Y), in the same random experiment, is a table or a function that determines the possible values of (X, Y) and their probabilities.

Example (Two dice)

The joint distribution of the two results X_1 and X_2 is

	X_2					
X_1	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

that is, the (discrete) uniform distribution over the product set $\{1, \ldots, 6\} \times \{1, \ldots, 6\}$.

Joint distribution of two variables 2 (Households)

- This table shows the shares of Finnish households, by room count and inhabitant count.
- If one household is picked at random, then its room count R and inhabitant count I are random variables (depending on which household is picked).
- Each cell contains the *probability* of (R, I) getting a particular pair of values. (Source: Tilastokeskus; excluded households where $R > 6$ or $I > 6$)

	I					
R	1	2	3	4	5	6
1	0.1263	0.0129	0.0019	0.0009	0.0003	0.0001
2	0.1961	0.0857	0.0124	0.0048	0.0014	0.0004
3	0.0727	0.0966	0.0336	0.0190	0.0050	0.0013
4	0.0382	0.0793	0.0306	0.0298	0.0100	0.0028
5	0.0154	0.0414	0.0169	0.0206	0.0087	0.0023
6	0.0042	0.0117	0.0055	0.0065	0.0034	0.0010

Discrete joint distribution

For two discrete random numbers X and Y, we can define the (discrete) joint density function

$$\begin{aligned} f_{X,Y}(x,y) &= \mathbb{P}(X = x \textbf{ and } Y = y) \\ &= \mathbb{P}(\{X = x\} \cap \{Y = y\}), \end{aligned}$$

which assigns a probability for each possible pair (x, y).

We can drop the subscripts, and write $f(x, y)$ if it causes no confusion.

Then the probability of *any* event $\{(X, Y) \in A\}$, where A is a collection of pairs, is simply the sum of probablities over A:

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

Continuous joint distribution

A pair of random numbers has a *continuous* joint distribution if the probability of any event is determined by a continuous joint density function

$$\mathbb{P}((X, Y) \in A) = \int_{(x,y) \in A} f_{X,Y}(x, y).$$

Taking an integral over an area is a topic of multivariate calculus. On this course we will not see much of these.

Marginal distributions 1 (Dice)

If we calculate the row sums and column sums from a joint distribution, we get two one-variable distributions, one for each. They are called the marginal distributions.

	X_2						
X_1	1	2	3	4	5	6	sum
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
sum	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

Row sums are the distribution of X_1
Column sums are the distribution of X_2

Marginal distributions 2 (Households)

	I						
R	1	2	3	4	5	6	Sum
1	0.1263	0.0129	0.0019	0.0009	0.0003	0.0001	0.1424
2	0.1961	0.0857	0.0124	0.0048	0.0014	0.0004	0.3010
3	0.0727	0.0966	0.0336	0.0190	0.0050	0.0013	0.2282
4	0.0382	0.0793	0.0306	0.0298	0.0100	0.0028	0.1908
5	0.0154	0.0414	0.0169	0.0206	0.0087	0.0023	0.1053
6	0.0042	0.0117	0.0055	0.0065	0.0034	0.0010	0.0324
Sum	0.4530	0.3277	0.1010	0.0816	0.0289	0.0078	

Row sums are the distribution of R
Column sums are the distribution of I

Formulas for the marginal densities

Marginals from a discrete joint distribution (i.e. taking the row or column sum):

$$
\begin{aligned}
f_X(x) &= \sum_{y \in S_Y} f_{X,Y}(x,y) \\
f_Y(y) &= \sum_{x \in S_X} f_{X,Y}(x,y),
\end{aligned}
$$

where S_X and S_Y are the sets of possible values. Marginals from a continuous joint distribution:

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.
\end{aligned}
$$

Contents

Conditional distribution (Households)

Let's look at the first row of the joint distribution: the case $R = 1$ (one-room households, six different inhabitant counts).

	I						
R	1	2	3	4	5	6	Sum
1	0.1263	0.0129	0.0019	0.0009	0.0003	0.0001	0.1424

The row sum is $0.1424 \neq 1$, so this row cannot be a probability distribution. In fact, the cells contain the shares of different one-room households *out of all households*.

Now divide by row sum $\longrightarrow$ shares *out of the one-room households*.

	I						
R	1	2	3	4	5	6	Sum
1	0.8870	0.0909	0.0131	0.0062	0.0021	0.0006	1.0000

This is a valid distribution: the conditional distribution of I, when $R = 1$.

Conditional distribution (Dice)

From rolling two *fair* dice, the conditional distributions would be very simple (uniform).

More interesting with unfair dice. Consider two dice (X_1, X_2) with:

- $P(X_1 = 2) = 0.5$, and $P(X_1 = i) = 0.1$ for $i \neq 2$
- $P(X_2 = 6) = 0.9$, and $P(X_1 = i) = 0.02$ for $i \neq 6$
- The two dice are independent from each other, i.e. what happens with one die, does not change the conditional distribution of the other; on every row $X_1 = i$ we have the same conditional distribution for X_2.

Compare now the *conditional distributions* and the *joint distribution* (BLACKBOARD).

Conditional distribution (discrete)

The conditional density function of Y, given the value of X, is defined as

$$f_{Y|X}(y|x) \;=\; \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

In the discrete case, it simply gives the *conditional probabilities*

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y \text{ and } X = x)}{X = x}.$$

If $f_X(x) > 0$, we observe that $f_{Y|X}$ is a density function.
It defines the conditional distribution of Y *in the case that* $X = x$.

For continuous variables we define the conditional density with the same formula, but it has a slightly different interpretation.

Stochastic dependence / independence

Random variables X and Y are (stochastically) independent, if for all sets A, B it is true that

$$\mathbb{P}(X \in A \,\textbf{and}\, Y \in B) \;=\; \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

Or (equivalently) if either of the following are true for all A, B:

$$\mathbb{P}(Y \in B \,|\, X \in A) \;=\; \mathbb{P}(Y \in B)$$

or

$$\mathbb{P}(X \in A \,|\, Y \in B) \;=\; \mathbb{P}(X \in A).$$

Then the event $X \in A$ (whether it is true or not) does not affect the distribution of Y; knowing X does not help to predict Y.

If (for any A, B) these equations do not hold, then A and B are dependent.

Stochastic independence / dependence

Fact

Random variables X, Y (whether discrete or continuous) are independent if *their joint density function can be expressed as*

$$f_{X,Y}(x, y) \ = \ f_X(x) f_Y(y)$$

(for all values of x, y).

An equivalent condition is

$$f_{Y|X}(y|x) \ = \ f_Y(y),$$

i.e. the conditional distribution of Y given X is *equal* to the "unconditional" distribution of Y as such.

Example. Random sampling

How many of the students in the room have been to Argentina?

- $S =$ "All students, $\#S = 80$
- $A =$ "Those who have been to Argentina, $\#A = 3$.

(In reality $\#A$ would be unknown, we would try to estimate it)

Take a random sample of $n = 2$ students, ask them, and let

$$X_1 = \begin{cases} 1, & \text{if 1st student} \in A \\ 0, & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if 2nd student} \in A \\ 0, & \text{otherwise} \end{cases}$$

What is the joint distribution of X_1, X_2? For example,
$\mathbb{P}(X_1 = 1, X_2 = 1) \ = ?$

Sampling with and without replacement

- With replacement = Second student chosen *again from the same population* "replace" = "put back"
- Without replacement = Second student chosen *from the remaining population*

With replacement

	X_2		
X_1	0	1	Sum
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

Without replacement

	X_2		
X_1	0	1	Sum
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

Surprise: Both cases have same *marginals*.

However the *joint* distributions are different.

Sampling with and without replacement

With

X_1	$X_2 = 0$	$X_2 = 1$	Sum
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_1,X_2}(i,j) = f_{X_1}(i) f_{X_2}(j)$$

Without

X_1	$X_2 = 0$	$X_2 = 1$	Sum
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_1,X_2}(i,j) \neq f_{X_1}(i) f_{X_2}(j)$$

Marginal distributions are same in both cases.

With replacement, X_1 and X_2 are *independent*.
Without replacement, X_1 are X_2 are *dependent*.

Conditional distribution (with replacement)

What is the *conditional* distribution of X_2 if $\{X_1 = 0\}$ occurs?

	X_2		
X_1	0	1	Sum
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_2|X_1}(0|0) = \frac{\frac{77}{80} \times \frac{77}{80}}{\frac{77}{80}} = \frac{77}{80}.$$

$$f_{X_2|X_1}(1|0) = \frac{\frac{77}{80} \times \frac{3}{80}}{\frac{77}{80}} = \frac{3}{80}.$$

Now conditional and unconditional distributions of X_2 are the same.

Conditional distribution (without replacement)

What is the *conditional* distribution of X_2 if $\{X_1 = 0\}$ occurs?

	X_2		
X_1	0	1	Sum
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Sum	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_2|X_1}(0|0) = \frac{\frac{77}{80} \times \frac{76}{79}}{\frac{77}{80}} = \frac{76}{79}.$$

$$f_{X_2|X_1}(1|0) = \frac{\frac{77}{80} \times \frac{3}{79}}{\frac{77}{80}} = \frac{3}{79}.$$

Now the conditional distribution of X_2 is different from the unconditional distribution.

Contents

Discrete distribution on an infinite set

A discrete random variable can have infinitely many possible values.

Example (Rolling dice until six)

Roll a die repeatedly *until* you get a six. Let N be the number of rolls done.

$$\begin{aligned}\mathbb{P}(N=k) &= \mathbb{P}(X_1 \neq 6, \ldots, X_{k-1} \neq 6, X_k = 6) \\ &= \mathbb{P}(X_1 \neq 6) \cdots \mathbb{P}(X_{k-1} \neq 6)\mathbb{P}(X_k = 6) \\ &= \left(1 - \frac{1}{6}\right)^{k-1} \left(\frac{1}{6}\right)\end{aligned}$$

The random variable N has a geometric distribution (with parameter "success probability" $p = 1/6$) over all positive integers $\{1, 2, \ldots\}$. It is a discrete distribution with density

$$f_N(k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots$$

Mixed discrete-continuous distribution

Y = waiting time (minutes) when trains arrive each 10 minutes, and stay 1 minute. If you arrive during that minute, no waiting.

X = time after *previous* train arrived is uniform over $[0, 10]$.
For $t \in [0, 9]$,

$$\begin{aligned} \mathbb{P}(Y \leq t) &= \mathbb{P}(Y = 0) + \mathbb{P}(0 < Y \leq t) \\ &= \mathbb{P}(X \leq 1) + \mathbb{P}(0 < 10 - X < t). \end{aligned}$$

$$\Longrightarrow \quad F_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{10} + \frac{t}{10}, & 0 \leq t \leq 9, \\ 1, & t > 9. \end{cases}$$

Is Y discrete or uniform? Neither! It is a *mixture* of a discrete distribution and a continuous one.

- with probability 0.1, we have $Y = 0$ exactly
- with probability 0.9, Y is uniformly distributed over $[0, 9]$.

(Further details omitted for now.)

Next lecture concerns the expected value of a random variable...