Probability 1 (MATH 11300) problem class slides

Márton Balázs

School of Mathematics
University of Bristol
Autumn, 2015

December 9, 2015

To know...

- http://www.maths.bris.ac.uk/~mb13434/prob1/
- m.balazs@bristol.ac.uk
- Drop in Sessions: Tuesdays, 13:00–14:00, office 3.7 in the Maths bld.
- Problem classes are *mandatory*, they contain assessed material.
- These slides cover *some* of our material. Most things will be done on blackboard.
-

1. Elementary probability

Sample space
Probability
Equally likely outcomes

Objectives:

- To define events and sample spaces, describe them in simple examples
- To list the axioms of probability, and use them to prove simple results
- To use counting arguments to calculate probabilities when there are equally likely outcomes

Sample space

- We always consider an *experiment*. Ω will denote the set of all possible outcomes of this experiment.

Sample space

- We always consider an *experiment*. Ω will denote the set of all possible outcomes of this experiment.
- An event will be a collection of possible outcomes. Therefore, and event E will be considered a subset of Ω: $E \subseteq \Omega$.

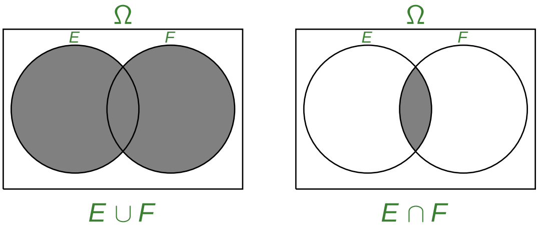
Sample space

- We always consider an *experiment*. Ω will denote the set of all possible outcomes of this experiment.
- An event will be a collection of possible outcomes. Therefore, and event E will be considered a subset of Ω: $E \subseteq \Omega$.
- Sometimes Ω is too large, and not all its subsets can be defined as events. This is where measure theory helps...

Sample space

- We always consider an *experiment*. Ω will denote the set of all possible outcomes of this experiment.
- An event will be a collection of possible outcomes. Therefore, and event E will be considered a subset of Ω: $E \subseteq \Omega$.
- Sometimes Ω is too large, and not all its subsets can be defined as events. This is where measure theory helps...
- It makes perfect sense to define the union $E \cup F$ and the intersection $E \cap F$ of two events, E and F.

Sample space



Notation: sometimes $E \cup F = E + F$, $E \cap F = EF$.

2. The union and the intersection

Inspired by the above:

Remark

The union $E \cup F$ of events E and F always means E OR F.
The intersection $E \cap F$ of events E and F always means E AND F.

Similarly:

Remark

The union $\bigcup_i E_i$ of events E_i always means at least one of the E_i's.
The intersection $\bigcap_i E_i$ of events E_i always means each of the E_i's.

2. The union and the intersection

Definition

If $E \cap F = \emptyset$, then we say that the events E and F are mutually exclusive events.

If the events $E_1, E_2, \ldots$ satisfy $E_i \cap E_j = \emptyset$ whenever $i \neq j$, then we say that the E_i's are mutually exclusive events.

Mutually exclusive events cannot happen at the same time.

3. Inclusion and implication

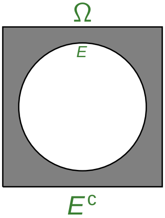
Remark

If the event E is a subset of the event F, $E \subseteq F$, then the occurrence of E implies that of F.

4. Complementary events

Definition

The complement of an event E is $E^c := \Omega - E$. This is the event that E does *not* occur.



Notice: $E \cap E^c = \emptyset$, $\qquad E \cup E^c = \Omega$.

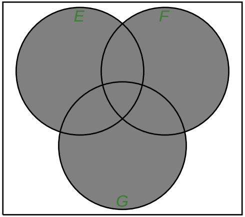
Notation: sometimes $E^c = \bar{E} = E^*$.

5. Simple properties of events

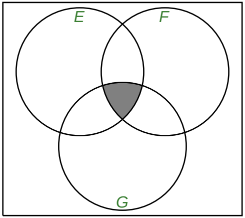
Commutativity: $E \cup F = F \cup E$,
$E \cap F = F \cap E$.

5. Simple properties of events

Associativity: $E \cup (F \cup G) = (E \cup F) \cup G = E \cup F \cup G$,

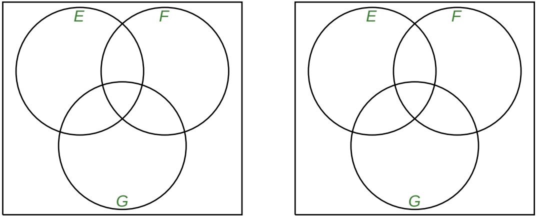


$E \cap (F \cap G) = (E \cap F) \cap G = E \cap F \cap G$.



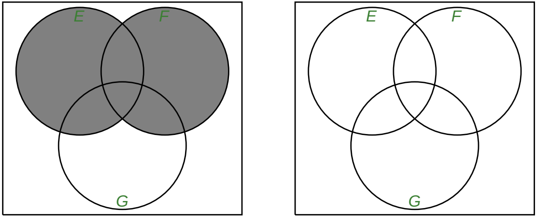
5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,



5. Simple properties of events

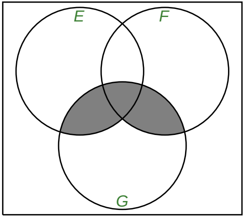
Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,



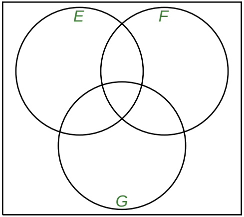
$E \cup F$

5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

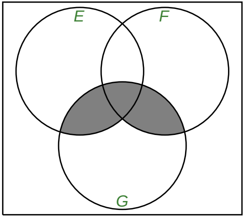


$(E \cup F) \cap G$

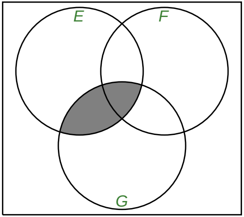


5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,



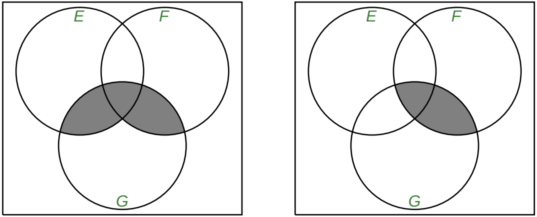
$(E \cup F) \cap G$



$E \cap G$

5. Simple properties of events

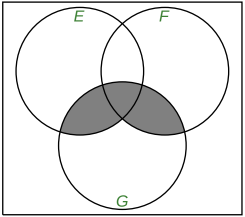
Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,



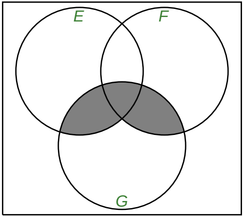
$(E \cup F) \cap G$ $\qquad$ $F \cap G$

5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,



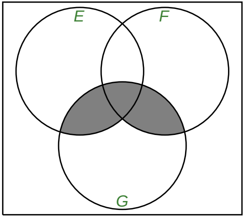
$(E \cup F) \cap G$



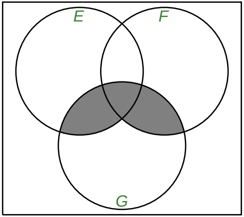
$(E \cap G) \cup (F \cap G)$

5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

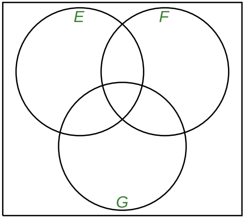


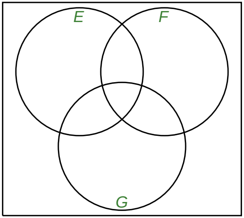
$(E \cup F) \cap G$



$(E \cap G) \cup (F \cap G)$

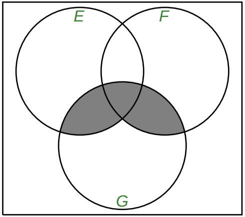
$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.



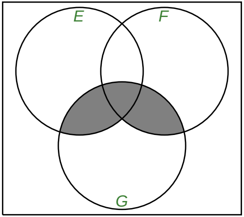


5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

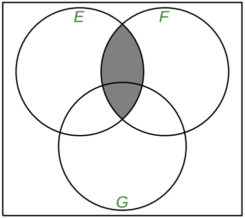


$(E \cup F) \cap G$

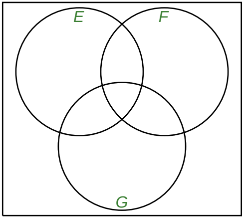


$(E \cap G) \cup (F \cap G)$

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

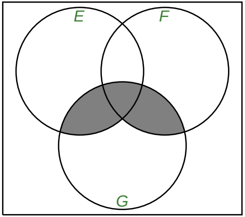


$E \cap F$

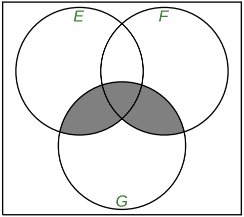


5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

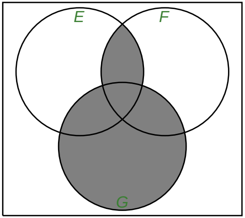


$(E \cup F) \cap G$

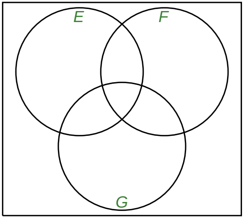


$(E \cap G) \cup (F \cap G)$

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

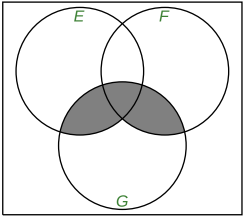


$(E \cap F) \cup G$

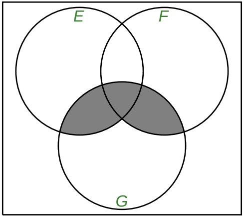


5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

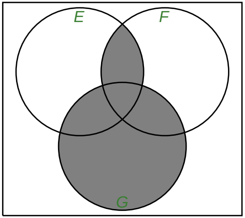


$(E \cup F) \cap G$

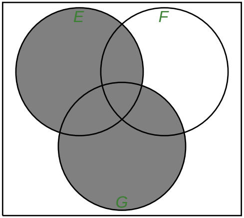


$(E \cap G) \cup (F \cap G)$

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.



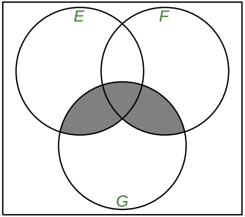
$(E \cap F) \cup G$



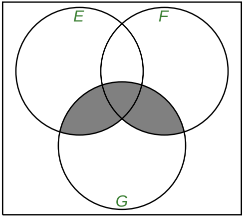
$E \cup G$

5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

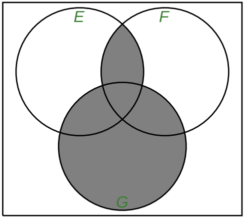


$(E \cup F) \cap G$

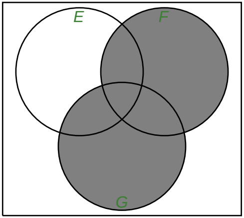


$(E \cap G) \cup (F \cap G)$

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.



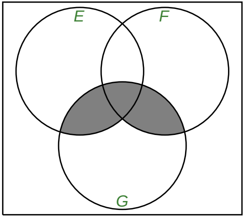
$(E \cap F) \cup G$



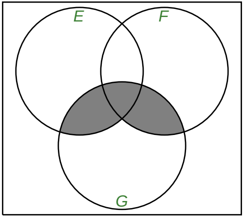
$F \cup G$

5. Simple properties of events

Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

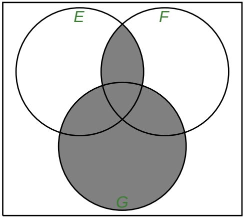


$(E \cup F) \cap G$

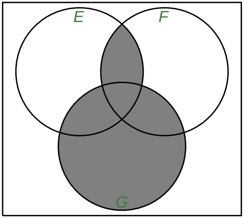


$(E \cap G) \cup (F \cap G)$

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.



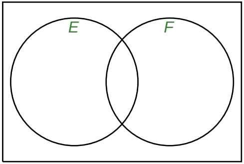
$(E \cap F) \cup G$

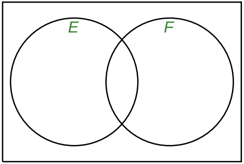


$(E \cup G) \cap (F \cup G)$

5. Simple properties of events

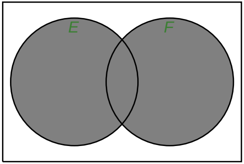
De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.



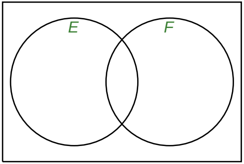


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.

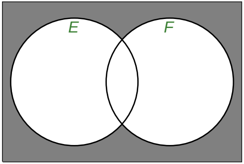


$E \cup F$

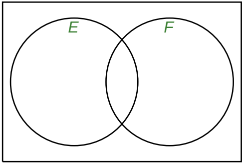


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.

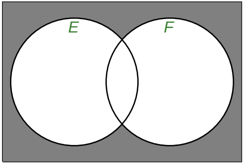


$(E \cup F)^c$

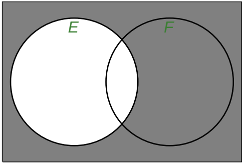


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.



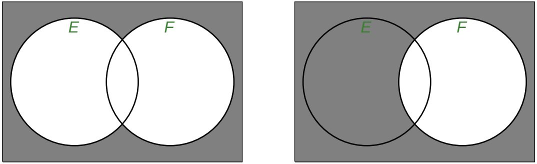
$(E \cup F)^c$



E^c

5. Simple properties of events

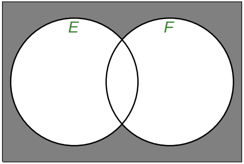
De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.



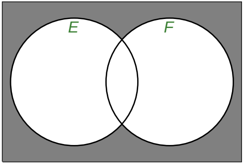
$(E \cup F)^c$ F^c

5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c.$



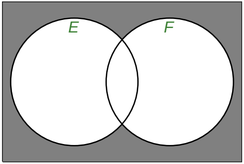
$(E \cup F)^c$



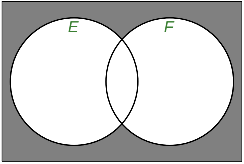
$E^c \cap F^c$

5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c.$

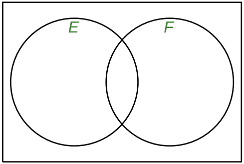


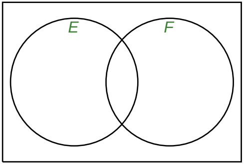
$(E \cup F)^c$



$E^c \cap F^c$

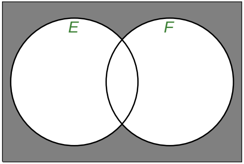
$$(E \cap F)^c = E^c \cup F^c.$$



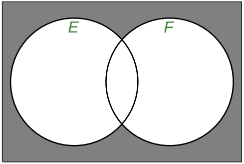


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.

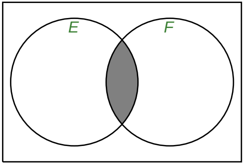


$(E \cup F)^c$

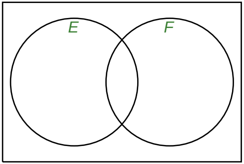


$E^c \cap F^c$

$(E \cap F)^c = E^c \cup F^c$.

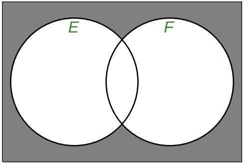


$E \cap F$

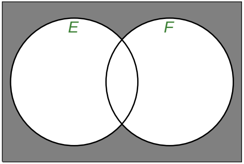


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c.$

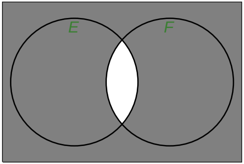


$(E \cup F)^c$

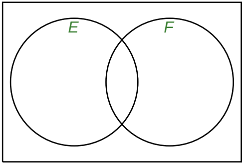


$E^c \cap F^c$

$$(E \cap F)^c = E^c \cup F^c.$$

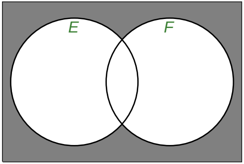


$(E \cap F)^c$

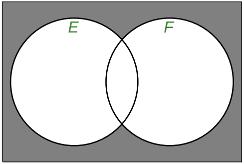


5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c.$

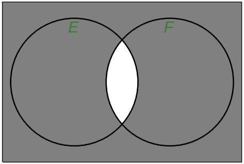


$(E \cup F)^c$

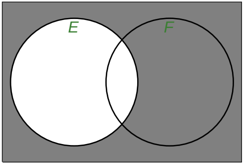


$E^c \cap F^c$

$(E \cap F)^c = E^c \cup F^c.$



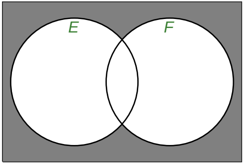
$(E \cap F)^c$



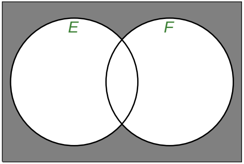
E^c

5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c.$

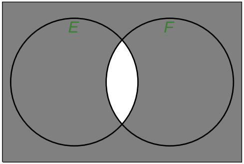


$(E \cup F)^c$

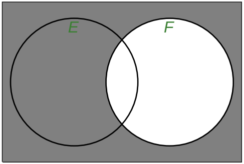


$E^c \cap F^c$

$(E \cap F)^c = E^c \cup F^c.$



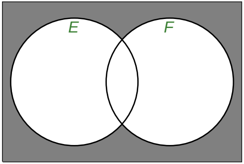
$(E \cap F)^c$



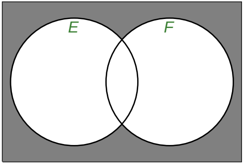
F^c

5. Simple properties of events

De Morgan's Law: $(E \cup F)^c = E^c \cap F^c$.

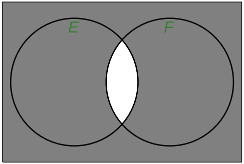


$(E \cup F)^c$

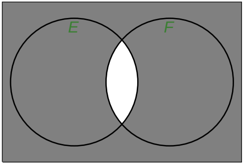


$E^c \cap F^c$

$(E \cap F)^c = E^c \cup F^c$.



$(E \cap F)^c$



$E^c \cup F^c$

Probability

Finally, we can now define what probability is.

Definition (axioms of probability)

The probability $\mathbf{P}$ on a sample space Ω assigns *numbers* to *events* of Ω in such a way, that:

1. the probability of any event is non-negative: $\mathbf{P}\{E\} \geq 0$;
2. the probability of the sample space is one: $\mathbf{P}\{\Omega\} = 1$;
3. for any finitely or countably infinitely many *mutually exclusive* events $E_1, E_2, \ldots$,

$$\mathbf{P}\Big\{\bigcup_i E_i\Big\} = \sum_i \mathbf{P}\{E_i\}.$$

Probability

Notation:

$$\bigcup_{i=1}^{n} E_i = E_1 \cup E_2 \cup \cdots \cup E_n \quad \text{, or}$$

$$\bigcup_{i=1}^{\infty} E_i = E_1 \cup E_2 \cup \dots \quad ,$$

$$\sum_{i=1}^{n} \mathbf{P}\{E_i\} = \mathbf{P}\{E_1\} + \mathbf{P}\{E_2\} + \cdots + \mathbf{P}\{E_n\} \quad \text{, or}$$

$$\sum_{i=1}^{\infty} \mathbf{P}\{E_i\} = \mathbf{P}\{E_1\} + \mathbf{P}\{E_2\} + \dots \quad .$$

A few simple facts

Proposition

For any event, $\mathbf{P}\{E^c\} = 1 - \mathbf{P}\{E\}$.

Corollary

We have $\mathbf{P}\{\emptyset\} = \mathbf{P}\{\Omega^c\} = 1 - \mathbf{P}\{\Omega\} = 1 - 1 = 0$.

For any event E, $\mathbf{P}\{E\} = 1 - \mathbf{P}\{E^c\} \le 1$.

A few simple facts

Proposition (Inclusion-exclusion principle)

For any events E *and* F, $\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F\} - \mathbf{P}\{E \cap F\}$.

Proposition (Boole's inequality)

For any events $E_1, E_2, \ldots, E_n$,

$$\mathbf{P}\Big\{\bigcup_{i=1}^{n} E_i\Big\} \le \sum_{i=1}^{n} \mathbf{P}\{E_i\}.$$

A few simple facts

Proposition (Inclusion-exclusion principle)

For any events E, F, G,

$$\begin{aligned}\mathbf{P}\{E\cup F\cup G\} = \mathbf{P}\{E\} + \mathbf{P}\{F\} + \mathbf{P}\{G\} \\ -\mathbf{P}\{E\cap F\} - \mathbf{P}\{E\cap G\} - \mathbf{P}\{F\cap G\} \\ +\mathbf{P}\{E\cap F\cap G\}.\end{aligned}$$

A few simple facts

Proposition

If $E \subseteq F$*, then* $\mathbf{P}\{F - E\} = \mathbf{P}\{F\} - \mathbf{P}\{E\}$.

Corollary

If $E \subseteq F$*, then* $\mathbf{P}\{E\} \leq \mathbf{P}\{F\}$.

Equally likely outcomes

A very special but important case is when the sample space is finite: $|\Omega| = N < \infty$, and each outcome of our experiment has equal probability. Then necessarily this probability equals $\frac{1}{N}$:

$$\mathbf{P}\{\omega\} = \frac{1}{N} \qquad \forall \omega \in \Omega.$$

Definition

These outcomes $\omega \in \Omega$ are also called <u>elementary events</u>.

Let $E \subseteq \Omega$ be an event that consists of k elementary events: $|E| = k$. Then

$$\mathbf{P}\{E\} = \frac{|E|}{|\Omega|} = \frac{k}{N}.$$

We thus see why *counting* will be important.

Equally likely outcomes

Example

An urn contains n balls, one of which is red, all others are black. We draw k balls at random (without replacement). What is the chance that the red ball will be drawn?

Solution (without order)

As before, Ω will be k-combinations of the n balls, and our event E is picking the red ball plus $k-1$ other balls. Thus,

$$\mathbf{P}\{E\} = \frac{|E|}{|\Omega|} = \frac{\binom{1}{1}\cdot\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{(n-1)!\cdot k!\cdot(n-k)!}{(k-1)!\cdot(n-1-k+1)!\cdot n!} = \frac{k}{n}.$$

The answer is so simple, there must be something behind this. . .

Equally likely outcomes

Example

An urn contains n balls, one of which is red, all others are black. We draw k balls at random (without replacement). What is the chance that the red ball will be drawn?

Solution (without order)

As before, Ω will be k-combinations of the n balls, and our event E is picking the red ball plus $k-1$ other balls. Thus,

$$\mathbf{P}\{E\} = \frac{|E|}{|\Omega|} = \frac{\binom{1}{1}\cdot\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{(n-1)!\cdot k!\cdot(n-k)!}{(k-1)!\cdot(n-1-k+1)!\cdot n!} = \frac{k}{n}.$$

The answer is so simple, there must be something behind this. . .

Equally likely outcomes

Example

An urn contains n balls, one of which is red, all others are black. We draw k balls at random (without replacement). What is the chance that the red ball will be drawn?

Solution (without order)

As before, Ω will be k-combinations of the n balls, and our event E is picking the red ball plus $k-1$ other balls. Thus,

$$\mathbf{P}\{E\} = \frac{|E|}{|\Omega|} = \frac{\binom{1}{1}\cdot\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{(n-1)!\cdot k!\cdot(n-k)!}{(k-1)!\cdot(n-1-k+1)!\cdot n!} = \frac{k}{n}.$$

The answer is so simple, there must be something behind this. . .

Equally likely outcomes

Solution (less combinatorics, more probability)

Imagine the draws in order, and let E_i be the event that our i^{th} draw is the red ball, $i = 1, \dots, k$. Clearly, each ball has equal chance of being drawn at a given time, thus $\mathbf{P}\{E_i\} = \frac{1}{n}$ for each i. Also, these events are mutually exclusive as there is only one red ball in the urn. Finally, we are looking for the union of these events, being that the red ball is drawn at one of our trials. With all this, the answer is

$$\mathbf{P}\Big\{\bigcup_{i=1}^{k} E_i\Big\} = \sum_{i=1}^{k} \mathbf{P}\{E_i\} = \sum_{i=1}^{k} \frac{1}{n} = \frac{k}{n}.$$

Equally likely outcomes

Solution (the elegant one)

We slightly extend our probability space, and we consider random permutations of all of the n balls. We imagine that the balls at the first k positions will be drawn. Clearly, this is a uniform random choice of k balls, realising our original problem. Now, the red ball will be drawn if and only if it's within the first k slots. As it has equal chance of being anywhere, the answer is $\frac{k}{n}$.

A second (third) look at the problem sometimes saves a lot of work.

Equally likely outcomes

Example

Three married couples are seated randomly around a round table. What is the chance that no husband sits next to his wife?

Solution

Firstly, define the sample space as the set of circular permutations of the 6 people. This is the same as ordinary permutations, except that configurations that can be rotated into each other are not distinguished. There are $|\Omega| = 5!$ circular permutations of the 6 people.

$\rightsquigarrow$

Equally likely outcomes

Solution (...cont'd)

Next, we define E_i as the event that Couple i sit next to each other, $i = 1, 2, 3$. We use De Morgan's Law and inclusion-exclusion to find the probability that no husband sits next to his wife:

$$\begin{aligned}\mathbf{P}\Big\{\bigcap_{i=1}^{3} E_i^c\Big\} &= \mathbf{P}\Big\{\Big(\bigcup_{i=1}^{3} E_i\Big)^c\Big\} = 1 - \mathbf{P}\Big\{\bigcup_{i=1}^{3} E_i\Big\} \\ &= 1 - \mathbf{P}\{E_1\} - \mathbf{P}\{E_2\} - \mathbf{P}\{E_3\} \\ &\quad + \mathbf{P}\{E_1 \cap E_2\} + \mathbf{P}\{E_1 \cap E_3\} + \mathbf{P}\{E_2 \cap E_3\} \\ &\quad - \mathbf{P}\{E_1 \cap E_2 \cap E_3\}.\end{aligned}$$

Equally likely outcomes

Solution (. . . cont'd)

Start with $E_1 \cap E_2 \cap E_3$: each couple is together. We have $2!$ relative orders of the 3 couples, then 2^3 for husband-wife or wife-husband within each couple. Therefore
$\mathbf{P}\{E_1 \cap E_2 \cap E_3\} = \frac{2! \cdot 2^3}{5!} = \frac{2}{15}$.
Then for $E_1 \cap E_2$ we say that there are 4 "objects": the two couples and two remaining people. These can be arranged in $3!$ different ways, then 2^2 for the orders within the couples:
$\mathbf{P}\{E_1 \cap E_2\} = \frac{3! \cdot 2^2}{5!} = \frac{1}{5}$, and of course the other two-intersections are similar.
Finally, for E_1 we have one couple and four people, a total of 5 objects, $\mathbf{P}\{E_1\} = \frac{4! \cdot 2}{5!} = \frac{2}{5}$, and the same for the other singletons.
Combining, the answer is $1 - \frac{2}{5} - \frac{2}{5} - \frac{2}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} - \frac{2}{15} = \frac{4}{15}$.

$\rightsquigarrow$

2. Conditional probability

Conditional probability

Objectives:

- To understand what conditioning means, reduce the sample space
- To use the conditional probability, Law of Total Probability and Bayes' Theorem
- To understand and use independence and conditional independence

2. The formal definition

Definition

Let F be an event with $\mathbf{P}\{F\} > 0$ (we'll assume this from now on). Then the conditional probability of E, given F is defined as

$$\mathbf{P}\{E \mid F\} := \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}}.$$

Definition

The event F that is given to us is also called a reduced sample space. We can simply work in this set to figure out the conditional probabilities given this event.

3. It's well-behaved

Proposition

The conditional probability $\mathbf{P}\{\cdot\,|F\}$ is a proper probability (it satisfies the axioms):

1. *the conditional probability of any event is non-negative:* $\mathbf{P}\{E\,|\,F\} \ge 0$;
2. *the conditional probability of the sample space is one:* $\mathbf{P}\{\Omega\,|\,F\} = 1$;
3. *for any finitely or countably infinitely many mutually exclusive events* $E_1, E_2, \ldots$,

$$\mathbf{P}\Big\{\bigcup_i E_i\,|\,F\Big\} = \sum_i \mathbf{P}\{E_i\,|\,F\}.$$

3. It's well-behaved

Corollary

All statements remain valid for $\mathbf{P}\{\cdot \,|\, F\}$*. E.g.*

- $\mathbf{P}\{E^c \,|\, F\} = 1 - \mathbf{P}\{E \,|\, F\}$.
- $\mathbf{P}\{\emptyset \,|\, F\} = 0$.
- $\mathbf{P}\{E \,|\, F\} = 1 - \mathbf{P}\{E^c \,|\, F\} \le 1$.
- $\mathbf{P}\{E \cup G \,|\, F\} = \mathbf{P}\{E \,|\, F\} + \mathbf{P}\{G \,|\, F\} - \mathbf{P}\{E \cap G \,|\, F\}$.
- *If* $E \subseteq G$*, then* $\mathbf{P}\{G - E \,|\, F\} = \mathbf{P}\{G \,|\, F\} - \mathbf{P}\{E \,|\, F\}$.
- *If* $E \subseteq G$*, then* $\mathbf{P}\{E \,|\, F\} \le \mathbf{P}\{G \,|\, F\}$.

Remark

BUT: Don't change the condition! E.g., $\mathbf{P}\{E \,|\, F\}$ and $\mathbf{P}\{E \,|\, F^c\}$ have nothing to do with each other.

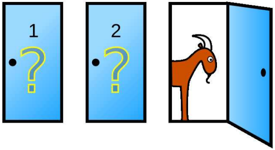
3. It's well-behaved

Example (Monty Hall problem)

We have three doors. Behind one of them is a car, behind the others, goats.

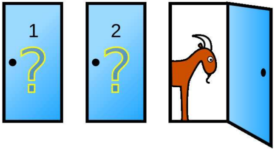
1. You pick a door (assume it's 1 WLOG).
2. Monty opens *another door with a goat behind it* (e.g., 3).
3. Now you pick one of the two closed doors (repeat your choice, or switch to the other one).
4. Whatever is behind this door is yours.

Would you repeat your choice or switch?



(http://en.wikipedia.org/wiki/File:Monty_open_door.svg)

3. It's well-behaved



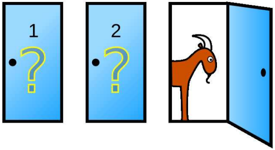
Wrong answer: The car is behind either door 1 or 2 now, so the chance is $\frac{1}{2}$ – $\frac{1}{2}$.

Define $E = \{\text{car is behind door 1}\}$, $F = \{\text{door 3 has a goat}\}$. Then

$$\mathbf{P}\{E \mid F\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}} = \frac{1/3}{2/3} = \frac{1}{2}. \qquad \leftarrow \text{WRONG answer.}$$

What's the problem?

3. It's well-behaved



Correct answer: define $G = \{\text{Monty has opened door 3}\}$. This is exactly what's given! Then

$$\mathbf{P}\{E \mid G\} = \frac{\mathbf{P}\{E \cap G\}}{\mathbf{P}\{G\}} = \frac{1/3 \cdot 1/2}{1/2} = \frac{1}{3}$$

(symmetry).

What is going on here?

3. It's well-behaved

$F = \{\text{door 3 has a goat}\}$, $G = \{\text{Monty has opened door 3}\}$.

$$G \subseteq F, \qquad \text{but} \qquad G \neq F!$$

$$\mathbf{P}\{E \,|\, F\} = \frac{1}{2},$$
$$\mathbf{P}\{E \,|\, G\} = \frac{1}{3}.$$

Remember: Don't change the condition!

Simple alternate argument for $\frac{1}{3}$: $\rightsquigarrow$

3. Discrete random variables

Mass function
Expectation, variance
Bernoulli, Binomial
Poisson
Geometric

Objectives:

- To build a mathematical model for discrete random variables
- To define and get familiar with the probability mass function, expectation and variance of such variables
- To get experience in working with some of the basic distributions (Bernoulli, Binomial, Poisson, Geometric)

Random variables

The best way of thinking about random variables is just to consider them as random numbers.

But *random* means that there must be some kind of experiment behind these numbers. They actually fit well in our framework:

Definition

A random variable is a function from the sample space Ω to the real numbers $\mathbb{R}$.

The usual notation for random variables is X, Y, Z, etc., we often don't mark them as functions: $X(\omega), Y(\omega), Z(\omega)$, etc.

Discrete random variables

Definition

A random variable X that can take on finitely or countably infinitely many possible values is called discrete.

Mass function

The *distribution* of a random variable will be the object of central importance to us.

Definition

Let X be a discrete random variable with possible values $x_1, x_2, \ldots$. The probability mass function (pmf), or distribution of a random variable tells us the probabilities of these possible values:

$$\mathfrak{p}_X(x_i) = \mathbf{P}\{X = x_i\},$$

for all possible x_i's.

Often the possible values are just integers, $x_i = i$, and we can just write $\mathfrak{p}_X(i)$ for the mass function.
We also omit the subscript $_X$ if it's clear which random variable we are considering and simply put $\mathfrak{p}(i)$.

Mass function

Proposition

For any discrete random variable X,

$$\mathfrak{p}(x_i) \geq 0, \qquad \textit{and} \qquad \sum_i \mathfrak{p}(x_i) = 1.$$

Remark

Vice versa: any function $\mathfrak{p}$ which is only non-zero in countably many x_i values, and which has the above properties, is a probability mass function. There is a sample space and a random variable that realises this mass function.

Expectation, variance

Once we have a random variable, we would like to quantify its *typical* behaviour in some sense. Two of the most often used quantities for this are the *expectation* and the *variance*.

1. Expectation

Definition

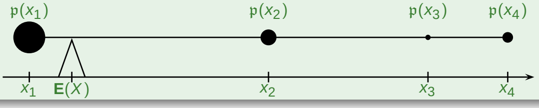
The expectation, or mean, or expected value of a discrete random variable X is defined as

$$\mathbf{E}X := \sum_i x_i \cdot \mathfrak{p}(x_i),$$

provided that this sum exists.

Remark

The expectation is nothing else than a weighted average of the possible values x_i with weights $\mathfrak{p}(x_i)$. A center of mass, in other words.



1. Expectation

Example

Let X be a positive integer random variable with mass function

$$\mathfrak{p}(i) = \frac{6}{\pi^2} \cdot \frac{1}{i^2} \qquad i = 1,\, 2,\, \dots .$$

Its expectation is

$$\mathbf{E}X = \sum_{i=1}^{\infty} i \cdot \mathfrak{p}(i) = \sum_{i=1}^{\infty} \frac{6}{\pi^2} \cdot \frac{1}{i} = \infty.$$

That's fine, don't worry.

1. Expectation

Example

Let X be an integer valued random variable with mass function

$$\mathfrak{p}(i) = \frac{3}{\pi^2} \cdot \frac{1}{i^2} \qquad \text{whenever } i \neq 0.$$

Its expectation is

$$\mathbf{E}X = \sum_{\substack{i=-\infty \\ i\neq 0}}^{\infty} i \cdot \mathfrak{p}(i) = \sum_{\substack{i=-\infty \\ i\neq 0}}^{\infty} \frac{3}{\pi^2} \cdot \frac{1}{i} = \textit{ooops!}$$

This expectation does not exist. $\rightsquigarrow$

For most cases we'll have nice finite expectations. In fact, we'll assume that from now on.

2. A few properties of expectation

Proposition (expectation of a function of a r.v.)

Let X be a discrete random variable, and $g : \mathbb{R} \to \mathbb{R}$ function. Then

$$\mathbf{E}g(X) = \sum_i g(x_i) \cdot \mathfrak{p}(x_i),$$

if exists...

2. A few properties of expectation

Corollary (linearity of expectations, first version)

Let X be a discrete random variable, a and b fixed real numbers. Then

$$\mathbf{E}(aX + b) = a \cdot \mathbf{E}X + b.$$

2. A few properties of expectation

Definition (moments)

Let n be a positive integer. The n^{th} moment of a random variable X is defined as

$$\mathbf{E}X^n.$$

The n^{th} absolute moment of X is

$$\mathbf{E}|X|^n.$$

Remark

Our notation in this definition and in the future will be

$$\mathbf{E}X^n := \mathbf{E}(X^n) \neq (\mathbf{E}X)^n \,!!$$

3. Variance

Definition (variance, standard deviation)

The variance and the standard deviation of a random variable are defined as $\mathbf{Var}X := \mathbf{E}(X - \mathbf{E}X)^2$ and $\mathrm{SD}\,X := \sqrt{\mathbf{Var}X}$.

4. A few properties of the variance

Proposition (an equivalent form of the variance)

For any X, $\mathbf{Var}X = \mathbf{E}X^2 - (\mathbf{E}X)^2$.

Corollary

For any X, $\mathbf{E}X^2 \geq (\mathbf{E}X)^2$, *with equality only if* $X = const.$ a.s.

New notation a.s. (almost surely) means *with probability one*.

4. A few properties of the variance

Example (an important one)

The variance of the indicator variable X of the event E is

$$\mathbf{Var}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 1^2 \cdot \mathbf{P}\{E\} - (\mathbf{P}\{E\})^2 = \mathbf{P}\{E\} \cdot (1 - \mathbf{P}\{E\})$$

and the standard deviation is $\mathrm{SD}\,X = \sqrt{\mathbf{P}\{E\} \cdot (1 - \mathbf{P}\{E\})}$.

4. A few properties of the variance

Proposition (*non*linearity of the variance)

Let X be a random variable, a and b fixed real numbers. Then

$$\mathbf{Var}(aX + b) = a^2 \cdot \mathbf{Var}X\,.$$

Notice the square on a^2 and also that, in particular, $\mathbf{Var}(X + b) = \mathbf{Var}X = \mathbf{Var}(-X)$: the variance is invariant to shifting the random variable by a constant b or to reflecting it.

Bernoulli, Binomial

In this part we'll get to know the Bernoulli and the Binomial distributions.

The setting will be that a fixed number of *independent* trials will be made, each succeeding with probability p. We will be counting the number of successes.

1. Definition

Definition

Suppose that n independent trials are performed, each succeeding with probability p. Let X count the number of successes within the n trials. Then X has the Binomial distribution with parameters n and p or, in short, $X \sim \text{Binom}(n, p)$.

The special case of $n = 1$ is called the Bernoulli distribution with parameter p.

Notice that the Bernoulli distribution is just another name for the *indicator variable* from before.

2. Mass function

Proposition

Let $X \sim \mathit{Binom}(n, p)$. Then $X = 0, 1, \ldots, n$, and its mass function is

$$\mathfrak{p}(i) = \mathbf{P}\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \qquad i = 0, 1, \ldots, n.$$

In particular, the $\mathit{Bernoulli}(p)$ variable can take on values 0 or 1, with respective probabilities

$$\mathfrak{p}(0) = 1 - p, \qquad \mathfrak{p}(1) = p.$$

2. Mass function

Remark

That the above is indeed a mass function we verify via the Binomial Theorem ($\mathfrak{p}(i) \geq 0$ is clear):

$$\sum_{i=0}^{n} \mathfrak{p}(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1.$$

3. Expectation, variance

Proposition

Let $X \sim Binom(n, p)$. *Then*

$$\mathbf{E}X = np, \qquad \textit{and} \qquad \mathbf{Var}X = np(1-p).$$

Proof.

We first need to calculate

$$\mathbf{E}X = \sum_i i \cdot \mathfrak{p}(i) = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-i}.$$

To handle this, here is a cute trick: $i = \frac{\mathrm{d}}{\mathrm{d}t} t^i \big|_{t=1}$. $\rightsquigarrow$

3. Expectation, variance

Proof.

$$\begin{aligned}
\mathbf{E}X &= \sum_{i=0}^{n} \binom{n}{i} i \cdot p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{\mathrm{d}}{\mathrm{d}t} t^i \big|_{t=1} \cdot p^i (1-p)^{n-i} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \Big(\sum_{i=0}^{n} \binom{n}{i} (tp)^i (1-p)^{n-i} \Big) \Big|_{t=1} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} (tp + 1 - p)^n \big|_{t=1} = n(tp + 1 - p)^{n-1} \cdot p \big|_{t=1} = np.
\end{aligned}$$

3. Expectation, variance

Proof.

$$\begin{aligned}
\mathbf{E}X &= \sum_{i=0}^{n} \binom{n}{i} i \cdot p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{\mathrm{d}}{\mathrm{d}t} t^i |_{t=1} \cdot p^i (1-p)^{n-i} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \Big(\sum_{i=0}^{n} \binom{n}{i} (tp)^i (1-p)^{n-i} \Big) \Big|_{t=1} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} (tp + 1 - p)^n |_{t=1} = n(tp + 1 - p)^{n-1} \cdot p|_{t=1} = np.
\end{aligned}$$

3. Expectation, variance

Proof.

For the variance we'll need the second moment too. Observe first

$$i(i-1) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} t^i |_{t=1}.$$

$\rightsquigarrow$

This enables us to compute the second factorial moment

3. Expectation, variance

Proof.

$$\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{i=0}^{n} \binom{n}{i} i(i-1)\cdot p^i(1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{\mathrm{d}^2}{\mathrm{d}t^2} t^i|_{t=1}\cdot p^i(1-p)^{n-i} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big(\sum_{i=0}^{n} \binom{n}{i} t^i\cdot p^i(1-p)^{n-i}\Big)\Big|_{t=1} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2}(tp+1-p)^n|_{t=1} \\
&= n(n-1)(tp+1-p)^{n-2}\cdot p^2|_{t=1} = n(n-1)p^2.
\end{aligned}$$

3. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{i=0}^{n} \binom{n}{i} i(i-1) \cdot p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{\mathrm{d}^2}{\mathrm{d}t^2} t^i \big|_{t=1} \cdot p^i (1-p)^{n-i} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2} \Big(\sum_{i=0}^{n} \binom{n}{i} t^i \cdot p^i (1-p)^{n-i} \Big) \Big|_{t=1} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2} (tp+1-p)^n \big|_{t=1} \\
&= n(n-1)(tp+1-p)^{n-2} \cdot p^2 \big|_{t=1} = n(n-1)p^2.
\end{aligned}
$$

3. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{i=0}^{n} \binom{n}{i} i(i-1) \cdot p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n}{i} \frac{\mathrm{d}^2}{\mathrm{d}t^2} t^i \big|_{t=1} \cdot p^i (1-p)^{n-i} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2} \Big(\sum_{i=0}^{n} \binom{n}{i} t^i \cdot p^i (1-p)^{n-i} \Big) \Big|_{t=1} \\
&= \frac{\mathrm{d}^2}{\mathrm{d}t^2} (tp + 1 - p)^n \big|_{t=1} \\
&= n(n-1)(tp+1-p)^{n-2} \cdot p^2 \big|_{t=1} = n(n-1)p^2.
\end{aligned}
$$

3. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\
&= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}
$$

□

3. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\
&= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}
$$

$\square$

3. Expectation, variance

Proof.

$$\begin{aligned}\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2 \\ &= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2 \\ &= n(n-1)p^2 + np - (np)^2 = np(1-p).\end{aligned}$$

$\square$

3. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\
&= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2 \\
&= n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}
$$

□

Poisson

The Poisson distribution is of central importance in Probability. We won't see immediately why, we'll just start with defining its distribution. Later we'll see how it comes from the Binomial.

1. Mass function

Definition

Fix a positive real number λ. The random variable X is Poisson distributed with parameter λ, in short $X \sim \text{Poi}(\lambda)$, if it is non-negative integer valued, and its mass function is

$$\mathfrak{p}(i) = \mathbf{P}\{X = i\} = e^{-\lambda} \cdot \frac{\lambda^i}{i!}, \qquad i = 0, 1, 2, \ldots$$

We have already seen in an example that this is indeed a mass function.

Ok, nice, but why this distribution?

2. Poisson approximation of Binomial

Proposition

Fix $\lambda > 0$, and suppose that $Y_n \sim Binom(n, p)$ with $p = p(n)$ in such a way that $n \cdot p \to \lambda$. Then the distribution of Y_n converges to $Poisson(\lambda)$:

$$\forall i \geq 0 \qquad \mathbf{P}\{Y_n = i\} \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}.$$

That is, take $Y \sim$ Binom(n, p) with large n, small p, such that $np \simeq \lambda$. Then Y is approximately Poisson(λ) distributed.

3. Expectation, variance

Proposition

For $X \sim Poi(\lambda)$, $\mathbf{E}X = \mathbf{Var}X = \lambda$.

Recall np and $np(1-p)$ for the Binomial...

Proof.

$$\mathbf{E}X = \sum_{i=0}^{\infty} i\mathfrak{p}(i) = \sum_{i=1}^{\infty} i \cdot \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}$$

$$= \lambda \sum_{i=1}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} = \lambda \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^j}{j!} = \lambda.$$

3. Expectation, variance

Proof.

For the variance, start with the *second factorial moment again*:

$$\begin{aligned}\mathbf{E}[X(X-1)] &= \sum_{i=0}^{\infty} i(i-1)\mathfrak{p}(i) = \sum_{i=2}^{\infty} i(i-1)\cdot \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}\\ &= \lambda^2 \sum_{i=2}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^j}{j!} = \lambda^2.\end{aligned}$$

$$\begin{aligned}\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2\\ &= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

$\square$

3. Expectation, variance

Proof.

For the variance, start with the *second factorial moment again*:

$$\mathbf{E}[X(X-1)] = \sum_{i=0}^{\infty} i(i-1)\mathfrak{p}(i) = \sum_{i=2}^{\infty} i(i-1)\cdot \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}$$

$$= \lambda^2 \sum_{i=2}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^j}{j!} = \lambda^2.$$

$$\begin{aligned}\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2\\ &= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

$\square$

3. Expectation, variance

Proof.

For the variance, start with the *second factorial moment again*:

$$\mathbf{E}[X(X-1)] = \sum_{i=0}^{\infty} i(i-1)\mathfrak{p}(i) = \sum_{i=2}^{\infty} i(i-1)\cdot \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}$$

$$= \lambda^2 \sum_{i=2}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^j}{j!} = \lambda^2.$$

$$\begin{aligned} \mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2 \\ &= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

$\square$

3. Expectation, variance

Proof.

For the variance, start with the *second factorial moment again*:

$$\begin{aligned}\mathbf{E}[X(X-1)] &= \sum_{i=0}^{\infty} i(i-1)\mathfrak{p}(i) = \sum_{i=2}^{\infty} i(i-1)\cdot \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}\\ &= \lambda^2\sum_{i=2}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^{i-2}}{(i-2)!} = \lambda^2\sum_{j=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^j}{j!} = \lambda^2.\end{aligned}$$

$$\begin{aligned}\mathbf{Var}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2\\ &= \mathbf{E}(X^2 - X) + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2\\ &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

$\square$

Geometric

In this setting we again perform independent trials. However, the question we ask is now different: we'll be waiting for the first success.

1. Mass function

Definition

Suppose that independent trials, each succeeding with probability p, are repeated until the first success. The total number X of trials made has the Geometric(p) distribution (in short, $X \sim \text{Geom}(p)$).

Proposition

X can take on positive integers, with probabilities $\mathfrak{p}(i) = (1-p)^{i-1} \cdot p$, $i = 1, 2, \ldots$.

That this is a mass function, we verify by $\mathfrak{p}(i) \geq 0$ and

$$\sum_{i=1}^{\infty} \mathfrak{p}(i) = \sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p = \frac{p}{1-(1-p)} = 1.$$

1. Mass function

Remark

For a Geometric(p) random variable and any $k \geq 1$ we have $\mathbf{P}\{X \geq k\} = (1-p)^{k-1}$ (we have at least $k-1$ failures).

Corollary

The Geometric random variable is (discrete) memoryless: for every $k \geq 1$, $n \geq 0$

$$\mathbf{P}\{X \geq n+k \mid X > n\} = \mathbf{P}\{X \geq k\}.$$

2. Expectation, variance

Proposition

For a Geometric(p) random variable X,

$$\mathbf{E}X = \frac{1}{p}, \qquad \mathbf{Var}X = \frac{1-p}{p^2}.$$

2. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{E}X &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p = \sum_{i=0}^{\infty} i \cdot (1-p)^{i-1} p \\
&= \sum_{i=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} t^i \big|_{t=1} \cdot (1-p)^{i-1} p = \frac{\mathrm{d}}{\mathrm{d}t} \Big(\sum_{i=0}^{\infty} t^i \cdot (1-p)^{i-1} p \Big) \Big|_{t=1} \\
&= \frac{p}{1-p} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{1-(1-p)t} \Big|_{t=1} \\
&= \frac{p}{1-p} \cdot \frac{1-p}{(1-(1-p))^2} = \frac{1}{p}.
\end{aligned}
$$

Variance: $\rightsquigarrow$ $\square$

2. Expectation, variance

Proof.

$$
\begin{aligned}
\mathbf{E}X &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p = \sum_{i=0}^{\infty} {\color{red} i} \cdot (1-p)^{i-1} p \\
&= \sum_{i=0}^{\infty} {\color{red} \frac{\mathrm{d}}{\mathrm{d}t} t^i \big|_{t=1}} \cdot (1-p)^{i-1} p = \frac{\mathrm{d}}{\mathrm{d}t} \Big(\sum_{i=0}^{\infty} t^i \cdot (1-p)^{i-1} p \Big) \Big|_{t=1} \\
&= \frac{p}{1-p} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{1-(1-p)t} \Big|_{t=1} \\
&= \frac{p}{1-p} \cdot \frac{1-p}{(1-(1-p))^2} = \frac{1}{p}.
\end{aligned}
$$

Variance: $\rightsquigarrow$ $\square$

4. Continuous random variables

Distribution, density
Uniform
Exponential
Normal
Transformations

Objectives:

- To build a mathematical model of continuous random variables
- To define and get familiar with the cumulative distribution function, probability density function, expectation and variance of such variables
- To get experience in working with some of the basic distributions (Uniform, Exponential, Normal)
- To find the distribution of a function of a random variable

Distribution function

Definition

The cumulative distribution function (cdf) of a random variable X is given by

$$F : \mathbb{R} \to [0, 1], \qquad x \mapsto F(x) = \mathbf{P}\{X \le x\}.$$

Notice that this function is well defined for *any* random variable.

Remark

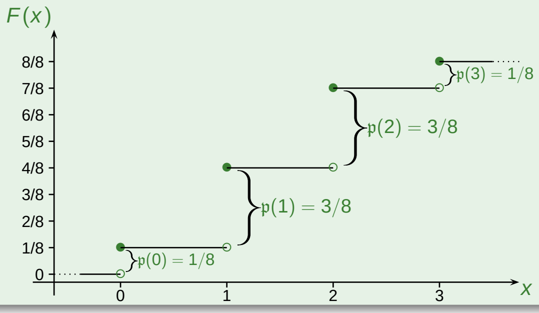
The distribution function contains all relevant information about the distribution of our random variable. E.g., for any fixed $a < b$,

$$\mathbf{P}\{a < X \le b\} = \mathbf{P}\{X \le b\} - \mathbf{P}\{X \le a\} = F(b) - F(a).$$

Distribution function

Example

Flip a coin three times, and let X be the number of Heads obtained. Its distribution function is given by



Distribution function

Definition

A random variable with piecewise constant distribution function is called *discrete*. Its mass function values equal to the jump sizes in the distribution function.

And this is equivalent to our earlier definition (taking on countably many possible values).

Distribution function

Proposition

A cumulative distribution function F

- *is non-decreasing;*
- *has limit* $\lim_{x\to-\infty} F(x) = 0$ *on the left;*
- *has limit* $\lim_{x\to\infty} F(x) = 1$ *on the right;*
- *is continuous from the right.*

$\rightsquigarrow$

Vice versa: any function F with the above properties is a cumulative distribution function. There is a sample space and a random variable on it that realises this distribution function.

1. Density function

Definition

Suppose that a random variable has its distribution function in the form of

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

with a function $f \geq 0$. Then the distribution is called (absolutely) continuous, and f is the probability density function (pdf).

We'll assume that X is continuous for the rest of this chapter.

1. Density function

Proposition

A probability density function f

- *is non-negative;*
- *has total integral* $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

$\rightsquigarrow$

Vice versa: any function f with the above properties is a probability density function. There is a sample space and a continuous random variable on it that realises this density.

2. Properties of the density function

Proposition

For any subset* $B \subseteq \mathbb{R}$,*

$$\mathbf{P}\{X \in B\} = \int_B f(x)\,\mathrm{d}x.$$

$\rightsquigarrow$

Corollary

Indeed, for a continuous random variable X,

$$\mathbf{P}\{X = a\} = \int_{\{a\}} f(x)\,\mathrm{d}x = 0 \qquad (\forall a \in \mathbb{R}).$$

2. Properties of the density function

Corollary

For a small ε,

$$\mathbf{P}\{X \in (a, a+\varepsilon]\} = \int_a^{a+\varepsilon} f(x)\,\mathrm{d}x \simeq f(a) \cdot \varepsilon.$$

There is no particular value that X can take on with positive chance. We can only talk about intervals, and the density tells us the likelihood that X is *around a point* a.

2. Properties of the density function

Corollary

To get to the density from a(n absolutely continuous!) distribution function,

$$f(a) = \frac{\mathrm{d}F(a)}{\mathrm{d}a} \qquad (a.e.\ a \in \mathbb{R}).$$

New notation a.e. (almost every): for all but a zero-measure set of numbers, so it's no problem for any integrals.

3. Expectation, variance

The way of defining the expectation will be no surprise for anyone (c.f. the discrete case):

Definition

The expected value of a continuous random variable X is defined by

$$\mathbf{E}X = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x,$$

if the integral exists.

3. Expectation, variance

In a way similar to the discrete case,

Proposition

Let X be a continuous random variable, and g an $\mathbb{R} \to \mathbb{R}$ function. Then

$$\mathbf{E}g(X) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,\mathrm{d}x$$

if exists.

From here we can define moments, absolute moments

$$\mathbf{E}X^n = \int_{-\infty}^{\infty} x^n \cdot f(x)\,\mathrm{d}x, \qquad \mathbf{E}|X|^n = \int_{-\infty}^{\infty} |x|^n \cdot f(x)\,\mathrm{d}x,$$

variance $\mathbf{Var}X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2$ and standard deviation $\mathrm{SD}\,X = \sqrt{\mathbf{Var}X}$ as in the discrete case. These enjoy the same properties as before.

Uniform

We are given real numbers $\alpha < \beta$, and wish to define a random variable X that's *equally likely to fall anywhere* in this interval. Thinking about the definitions, we can do that by assuming a constant density on this interval.

1. Density, distribution function

Definition

Fix $\alpha < \beta$ reals. We say that X has the uniform distribution over the interval (α, β), in short, $X \sim \mathrm{U}(\alpha, \beta)$, if its density is given by

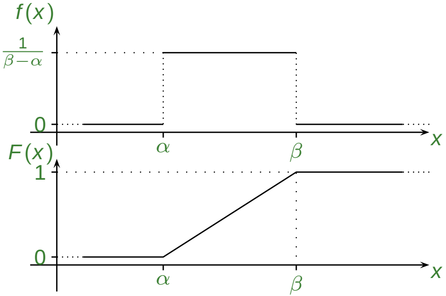
$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{if } x \in (\alpha, \beta), \\ 0, & \text{otherwise.} \end{cases}$$

Notice that this is exactly the value of the constant that makes this a density.

1. Density, distribution function

Integrating this density,

$$F(x) = \begin{cases} 0, & \text{if } x \le \alpha, \\ \dfrac{x-\alpha}{\beta-\alpha}, & \text{if } \alpha \le x \le \beta, \\ 1, & \text{if } \beta \le x. \end{cases}$$



1. Density, distribution function

Remark

If $X \sim \mathrm{U}(\alpha, \beta)$, and $\alpha < a < b < \beta$, then

$$\mathbf{P}\{a < X \le b\} = \int_a^b f(x)\,\mathrm{d}x = \frac{b-a}{\beta-\alpha}.$$

Probabilities are computed by proportions of lengths.

2. Expectation, variance

Proposition

For $X \sim U(\alpha,\,\beta)$,

$$\mathbf{E}X = \frac{\alpha+\beta}{2}, \qquad \mathbf{Var}X = \frac{(\beta-\alpha)^2}{12}.$$

“12” is the only non-trivial part of this formula. $\rightsquigarrow$

Proof.

$$\mathbf{E}X = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha}\,\mathrm{d}x = \frac{\frac{\beta^2}{2}-\frac{\alpha^2}{2}}{\beta-\alpha} = \frac{\alpha+\beta}{2}.$$

2. Expectation, variance

Proof.

$$\mathbf{E}X^2 = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_{\alpha}^{\beta} \frac{x^2}{\beta-\alpha}\,\mathrm{d}x$$

$$= \frac{1}{3}\cdot\frac{\beta^3-\alpha^3}{\beta-\alpha} = \frac{\beta^2+\alpha\beta+\alpha^2}{3},$$

hence

$$\mathbf{Var}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4}$$

$$= \frac{\beta^2-2\alpha\beta+\alpha^2}{12} = \frac{(\beta-\alpha)^2}{12}.$$

□

Exponential

The Exponential is a very special distribution because of its *memoryless* property. It is often considered as a waiting time, and is widely used in the theory of stochastic processes.

1. Density, distribution function

Definition

Fix a positive parameter λ. X is said to have the Exponential distribution with parameter λ or, in short, $X \sim \mathrm{Exp}(\lambda)$, if its density is given by

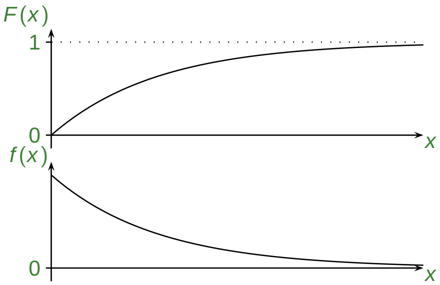
$$f(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \lambda \mathrm{e}^{-\lambda x}, & \text{if } x \ge 0. \end{cases}$$

Remark

Its distribution function can easily be integrated from the density:

$$F(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 1 - \mathrm{e}^{-\lambda x}, & \text{if } x \ge 0. \end{cases}$$

1. Density, distribution function



$$F(x) = 1 - \mathrm{e}^{-\lambda x}; \qquad f(x) = \lambda \mathrm{e}^{-\lambda x}.$$

2. Expectation, variance

Proposition

For $X \sim Exp(\lambda)$,

$$\mathbf{E}X = \frac{1}{\lambda}; \qquad \mathbf{Var}X = \frac{1}{\lambda^2}.$$

Thinking about X as a waiting time, we now see that λ describes how fast the event we wait for happens. Therefore λ is also called the rate of the exponential waiting time.

2. Expectation, variance

Proof.

We need to compute

$$\mathbf{E}X = \int_0^\infty x\lambda e^{-\lambda x}\,dx \qquad \text{and} \qquad \mathbf{E}X^2 = \int_0^\infty x^2\lambda e^{-\lambda x}\,dx$$

using integration by parts. $\rightsquigarrow$ □

Normal

The Normal, or Gaussian, is a very nice distribution on its own, but we won't see why it is useful until a bit later.

1. Density, distribution function

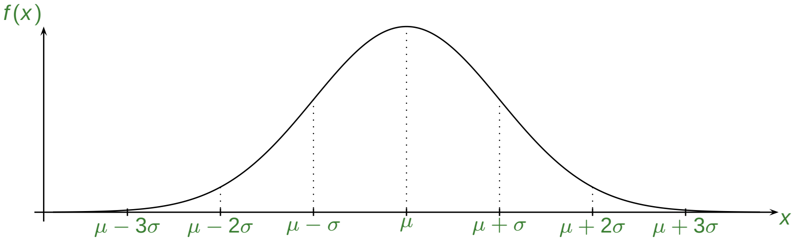
Definition

Let $\mu \in \mathbb{R}$, $\sigma > 0$ be real parameters. X has the Normal distribution with parameters μ and σ^2 or, in short $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density is given by

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (x \in \mathbb{R}).$$

To prove that this is a density, 2-dim. polar coordinates are needed, anyone interested come and see me after class.

1. Density, distribution function



Definition

The case $\mu = 0$, $\sigma^2 = 1$ is called standard normal distribution ($\mathcal{N}(0, 1)$). Its density is denoted by φ, and its distribution function by Φ:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-y^2/2} \,\mathrm{d}y \quad (x \in \mathbb{R}).$$

1. Density, distribution function

Remark

The standard normal distribution function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-y^2/2}\,\mathrm{d}y$$

has *no closed form*, its values will be looked up in tables.

Next we'll establish some tools that will enable us to use such tables to find probabilities of normal random variables.

2. Symmetry

Proposition

For any $z \in \mathbb{R}$,

$$\Phi(-z) = 1 - \Phi(z).$$

That's why most tables only have entries for positive values of z.

3. Linear transformations

Proposition

Let $X \sim \mathcal{N}(\mu, \sigma^2)$*, and* $\alpha, \beta \in \mathbb{R}$ *fixed numbers. Then* $\alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$.

3. Linear transformations

Corollary

If $X \sim \mathcal{N}(\mu, \sigma^2)$*, then its* *standardised version* $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Just use $\alpha = \frac{1}{\sigma}$ *and* $\beta = -\frac{\mu}{\sigma}$.

4. Expectation and variance

Proposition

If $X \sim \mathcal{N}(0,\,1)$ is standard normal, then its mean is 0 and its variance is 1.

Proof.

That the mean is zero follows from symmetry. For the variance we need to calculate

$$\mathbf{E}X^2 = \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \cdot \mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

using integration by parts. $\rightsquigarrow$ $\square$

4. Expectation and variance

Corollary

If $X \sim \mathcal{N}(\mu,\ \sigma^2)$, then its mean is μ and its variance is σ^2.

Proof.

$$\mathbf{E}X = \sigma \cdot \mathbf{E}\Big(\frac{X-\mu}{\sigma}\Big) + \mu = 0 + \mu = \mu,$$

$$\mathbf{Var}X = \sigma^2 \cdot \mathbf{Var}\Big(\frac{X-\mu}{\sigma}\Big) = \sigma^2 \cdot 1 = \sigma^2$$

as $\frac{X-\mu}{\sigma}$ is standard. □

$\mathcal{N}(\mu,\ \sigma^2)$ is also said to be the normal distribution with mean μ, variance σ^2.

Transformations

Let X be a random variable, and $g(X)$ a function of it. If X is discrete then the distribution of $g(X)$ is rather straightforward. In the continuous case the question is more interesting.

We have in fact seen an example before: an *affine transformation* $g(x) = ax + b$ of Normal keeps it Normal.

Transformations

Example

A torch, pointing in a random direction downwards, is 1 yard high above the origin of an infinite table. What is the distribution of the position of the lightened point on the table?

$\rightsquigarrow$

Transformations

Solution

Let $\Theta\sim\mathrm{U}(-\pi/2,\,\pi/2)$ be the angle between the torch and vertical. The lightbeam touches the table at $X=\tan\Theta$. Its distribution function and density are

$$F_X(x)=\mathbf{P}\{X<x\}=\mathbf{P}\{\tan\Theta<x\}=\mathbf{P}\{\Theta<\arctan x\}$$
$$=F_\Theta(\arctan x)=\frac{\arctan x-(-\pi/2)}{\pi/2-(-\pi/2)}=\frac{1}{\pi}\arctan x+\frac{1}{2},$$
$$f_X(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

This distribution is called *standard Cauchy*.

What is its expectation. . . ?

Transformations

Solution

Let $\Theta \sim \mathrm{U}(-\pi/2, \pi/2)$ be the angle between the torch and vertical. The lightbeam touches the table at $X = \tan\Theta$. Its distribution function and density are

$$F_X(x) = \mathbf{P}\{X < x\} = \mathbf{P}\{\tan\Theta < x\} = \mathbf{P}\{\Theta < \arctan x\}$$
$$= F_\Theta(\arctan x) = \frac{\arctan x - (-\pi/2)}{\pi/2 - (-\pi/2)} = \frac{1}{\pi}\arctan x + \frac{1}{2},$$
$$f_X(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

This distribution is called *standard Cauchy*.

What is its expectation. . . ?

Transformations

Solution

Let $\Theta \sim \mathrm{U}(-\pi/2, \pi/2)$ be the angle between the torch and vertical. The lightbeam touches the table at $X = \tan\Theta$. Its distribution function and density are

$$F_X(x) = \mathbf{P}\{X < x\} = \mathbf{P}\{\tan\Theta < x\} = \mathbf{P}\{\Theta < \arctan x\}$$

$$= F_\Theta(\arctan x) = \frac{\arctan x - (-\pi/2)}{\pi/2 - (-\pi/2)} = \frac{1}{\pi}\arctan x + \frac{1}{2},$$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

This distribution is called *standard Cauchy*.

What is its expectation. . . ?

Transformations

Solution

Let $\Theta \sim \mathrm{U}(-\pi/2, \pi/2)$ be the angle between the torch and vertical. The lightbeam touches the table at $X = \tan\Theta$. Its distribution function and density are

$$\begin{aligned} F_X(x) &= \mathbf{P}\{X < x\} = \mathbf{P}\{\tan\Theta < x\} = \mathbf{P}\{\Theta < \arctan x\} \\ &= F_\Theta(\arctan x) = \frac{\arctan x - (-\pi/2)}{\pi/2 - (-\pi/2)} = \frac{1}{\pi}\arctan x + \frac{1}{2}, \end{aligned}$$

$$f_X(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

This distribution is called *standard Cauchy*.

What is its expectation. . . ?

Transformations

There is a general formula along the same lines:

Proposition

Let X be a continuous random variable with density f_X, and g a continuously differentiable function with nonzero derivative. Then the density of $Y = g(X)$ is given by

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|}.$$

Proving this can be done in a way very similar to the scheme above.

The *scheme* is important.

5. Joint distributions

Joint distributions

Independence, convolutions

Objectives:

- To build a mathematical model for several random variables on a common probability space
- To get familiar with joint, marginal and conditional discrete distributions
- To understand discrete convolutions
- To get familiar with the Gamma distribution (via a continuous convolution)

Joint distributions

Often an experiment can result in several random quantities at the same time. In this case we have several random variables defined on a common probability space. Their relations can be far from trivial, and are described by *joint distributions*. Here we'll familiarise ourselves with the basics of joint distributions.

For most part we restrict our attention to the discrete case, as the jointly continuous case would require multivariable calculus and more time.

At the end of this chapter we introduce the Gamma distribution, purely motivated by joint distributions.

1. Joint mass function

Most examples will involve two random variables, but everything can be generalised for more of them.

Definition

Suppose two discrete random variables X and Y are defined on a common probability space, and can take on values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, respectively. The joint probability mass function of them is defined as

$$\mathfrak{p}(x_i, y_j) = \mathbf{P}\{X = x_i, Y = y_j\}, \qquad i = 1, 2, \ldots, j = 1, 2, \ldots.$$

This function contains all information about the joint distribution of X and Y.

1. Joint mass function

Definition

The marginal mass functions are

$$\mathfrak{p}_X(x_i) := \mathbf{P}\{X = x_i\}, \qquad \text{and} \qquad \mathfrak{p}_Y(y_j) := \mathbf{P}\{Y = y_j\}.$$

It is clear from the Law of Total Probability that

Proposition

$$\mathfrak{p}_X(x_i) = \sum_j \mathfrak{p}(x_i, y_j), \qquad \textit{and} \qquad \mathfrak{p}_Y(y_j) = \sum_i \mathfrak{p}(x_i, y_j).$$

1. Joint mass function

Proposition

Any joint mass function satisfies

- $\mathfrak{p}(x, y) \geq 0$, $\forall x, y \in \mathbb{R}$;
- $\sum_{i,j} \mathfrak{p}(x_i, y_j) = 1$.

Vice versa: any function $\mathfrak{p}$ which is only non-zero in countably many (x_i, y_j) values, and which has the above properties, is a joint probability mass function. There is a sample space and random variables that realise this joint mass function.

2. Conditional mass function

Definition

Suppose $p_Y(y_j) > 0$. The
conditional mass function of X, given $Y = y_j$ is defined by

$$\mathfrak{p}_{X\,|\,Y}(x\,|\,y_j) := \mathbf{P}\{X = x\,|\,Y = y_j\} = \frac{\mathfrak{p}(x,\,y_j)}{\mathfrak{p}_Y(y_j)}.$$

As the conditional probability was a proper probability, this is a proper mass function: $\forall x,\,y_j$,

$$\mathfrak{p}_{X\,|\,Y}(x\,|\,y_j) \ge 0, \qquad \sum_i \mathfrak{p}_{X\,|\,Y}(x_i\,|\,y_j) = 1.$$

Independence, convolutions

An important special case of joint distributions is the one of *independent* variables: whatever the value of some of them is, it does not influence the distribution of the others. We'll make this precise in this part, and then use it to determine the distribution of the sum of independent variables.

As a slight generalisation and application, we'll also introduce the *Gamma* distribution.

1. Independent r.v.'s

Definition

Random variables X and Y are independent, if events formulated with them are so. That is, if for every $A, B \subseteq \mathbb{R}$

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

Similarly, random variables $X_1, X_2, \ldots$ are independent, if events formulated with them are so. That is, if for every $A_{i_1}, A_{i_2}, \ldots, A_{i_n} \subseteq \mathbb{R}$

$$\mathbf{P}\{X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \ldots X_{i_n} \in A_{i_n}\}$$
$$= \mathbf{P}\{X_{i_1} \in A_{i_1}\} \cdot \mathbf{P}\{X_{i_2} \in A_{i_2}\} \cdots \mathbf{P}\{X_{i_n} \in A_{i_n}\}.$$

Recall *mutual independence* for events...

1. Independent r.v.'s

Remark

People use the abbreviation i.i.d. for independent and identically distributed random variables.

Proposition

Two random variables X and Y are independent if and only if their joint mass function factorises into the product of the marginals:

$$\mathfrak{p}(x_i, y_j) = \mathfrak{p}_X(x_i) \cdot \mathfrak{p}_Y(y_j), \qquad (\forall x_i, y_j).$$

1. Independent r.v.'s

Example (marking of the Poisson random variable)

Suppose that the number of customers entering the post office is of $\text{Poi}(\lambda)$ distribution. Assume also that each person, independently of each other, is female with probability p and male with probability $1-p$. Let us find the joint distribution of the number X of females and Y of males.

Solution

Everything we can ask for is described by the joint mass function. To compute it, we are going to make a strange conditioning:

$$\begin{aligned}\mathfrak{p}(i,j) &= \mathbf{P}\{X=i,\ Y=j\}\\ &= \mathbf{P}\{X=i,\ Y=j\,|\,X+Y=i+j\}\cdot\mathbf{P}\{X+Y=i+j\}.\end{aligned}$$

1. Independent r.v.'s

Solution (...cont'd)

The reason for this step is that we can now make use of the information given. First, the total number of people, $X + Y \sim \text{Poi}(\lambda)$, hence

$$\mathbf{P}\{X + Y = i + j\} = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^{i+j}}{(i+j)!}.$$

1. Independent r.v.'s

Solution (…cont'd)

Second, given the total number $X+Y$ of people, each of them is independently female with probability p, or male with probability $1-p$. Thus, given $X+Y=i+j$, $(X \mid X+Y=i+j) \sim \text{Binom}(i+j,\, p)$. Therefore,

$$\begin{aligned}\mathbf{P}\{X=i,\, Y=j \mid X+Y=i+j\} &= \mathbf{P}\{X=i \mid X+Y=i+j\} \\ &= \binom{i+j}{i} p^i (1-p)^j.\end{aligned}$$

As a preparation for what's coming later, we can summarise the above as $(X \mid X+Y) \sim \text{Binom}(X+Y,\, p)$.

1. Independent r.v.'s

Solution (...cont'd)

Second, given the total number $X + Y$ of people, each of them is independently female with probability p, or male with probability $1 - p$. Thus, given $X + Y = i + j$, $(X \mid X + Y = i + j) \sim \text{Binom}(i + j, p)$. Therefore,

$$\begin{aligned}\mathbf{P}\{X = i, Y = j \mid X + Y = i + j\} &= \mathbf{P}\{X = i \mid X + Y = i + j\} \\ &= \binom{i+j}{i} p^i (1 - p)^j.\end{aligned}$$

As a preparation for what's coming later, we can summarise the above as $(X \mid X + Y) \sim \text{Binom}(X + Y, p)$.

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i,j) = \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}$$
$$= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i,j) = \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}$$
$$= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i, j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i, j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i, j) = \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}$$
$$= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned}\mathfrak{p}(i,j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).\end{aligned}$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = e^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i,j) = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}$$
$$= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned}\mathfrak{p}(i,j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).\end{aligned}$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i, j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i, j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned}\mathfrak{p}(i, j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).\end{aligned}$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned}\mathfrak{p}(i,j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).\end{aligned}$$

1. Independent r.v.'s

Solution (...cont'd)

Combining,

$$\mathfrak{p}(i,j)=\binom{i+j}{i}p^i(1-p)^j\cdot \mathrm{e}^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}=\mathrm{e}^{-\lambda}p^i(1-p)^j\cdot\frac{\lambda^i\cdot\lambda^j}{i!\cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j)=f(i)\cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i,j)=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^i}{i!}\cdot\mathrm{e}^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}$$
$$=\mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i)\cdot\mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i, j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i, j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned} \mathfrak{p}(i, j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j). \end{aligned}$$

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\begin{aligned}\mathfrak{p}(i,j) &= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).\end{aligned}$$

1. Independent r.v.'s

Solution (…cont'd)

Combining,

$$\mathfrak{p}(i,j) = \binom{i+j}{i} p^i (1-p)^j \cdot \mathrm{e}^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = \mathrm{e}^{-\lambda} p^i (1-p)^j \cdot \frac{\lambda^i \cdot \lambda^j}{i! \cdot j!}.$$

What does this tell us? It's incriminating that the right hand-side is of *product form*: $\mathfrak{p}(i,j) = f(i) \cdot g(j)$. *We need to group constants in a proper way to see product of marginal mass functions* from here.

$$\mathfrak{p}(i,j) = \mathrm{e}^{-\lambda p} \frac{(\lambda p)^i}{i!} \cdot \mathrm{e}^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}$$
$$= \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

1. Independent r.v.'s

Solution (...cont'd)

$$\mathfrak{p}(i,j) = \mathfrak{p}_{\mathrm{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\mathrm{Poi}(\lambda(1-p))}(j).$$

It follows that

- $X \sim \mathrm{Poi}(\lambda p)$;
- $Y \sim \mathrm{Poi}(\lambda(1-p))$;
- X and Y are independent.

1. Independent r.v.'s

Solution (...cont'd)

$$\mathfrak{p}(i,j) = \mathfrak{p}_{\text{Poi}(\lambda p)}(i) \cdot \mathfrak{p}_{\text{Poi}(\lambda(1-p))}(j).$$

It follows that

- $X \sim \text{Poi}(\lambda p)$;
- $Y \sim \text{Poi}(\lambda(1-p))$;
- X and Y are independent. Surprise!

This is a particular feature of the Poisson distribution (and, in your further studies, the *Poisson process*).

2. Discrete convolution

We restrict ourselves now to integer valued random variables. Let X and Y be such, and also independent. What is the distribution of their sum?

Proposition

Let X and Y be independent, integer valued random variables with respective mass functions $\mathfrak{p}_X$ and $\mathfrak{p}_Y$. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i), \qquad (\forall k \in \mathbb{Z}).$$

This formula is called the discrete convolution of the mass functions $\mathfrak{p}_X$ and $\mathfrak{p}_Y$.

2. Discrete convolution

Example

Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot \mathrm{e}^{-\mu} \frac{\mu^i}{i!}$$

$$= \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \text{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot e^{-\mu} \frac{\mu^i}{i!}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \text{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot \mathrm{e}^{-\mu} \frac{\mu^i}{i!}$$

$$= \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \text{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot \mathrm{e}^{-\mu} \frac{\mu^i}{i!}$$

$$= \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \text{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot \mathrm{e}^{-\mu} \frac{\mu^i}{i!}$$

$$= \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \mathrm{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ be independent. Then

$$\mathfrak{p}_{X+Y}(k) = \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i) \cdot \mathfrak{p}_Y(i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot \mathrm{e}^{-\mu} \frac{\mu^i}{i!}$$

$$= \mathrm{e}^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

from where we conclude $X + Y \sim \text{Poi}(\lambda + \mu)$.

2. Discrete convolution

Example

Let X, Y be i.i.d. Geom(p) variables. Then

$$\begin{aligned}
\mathfrak{p}_{X+Y}(k) &= \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i)\cdot\mathfrak{p}_Y(i) \\
&= \sum_{i=1}^{k-1} (1-p)^{k-i-1}p\cdot(1-p)^{i-1}p \\
&= (k-1)\cdot(1-p)^{k-2}p^2,
\end{aligned}$$

$X+Y$ is *not* Geometric.

(It's actually called *Negative Binomial*...)

2. Discrete convolution

Example

Let X, Y be i.i.d. Geom(p) variables. Then

$$\begin{aligned}\mathfrak{p}_{X+Y}(k) &= \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i)\cdot \mathfrak{p}_Y(i)\\ &= \sum_{i=1}^{k-1} (1-p)^{k-i-1}p\cdot(1-p)^{i-1}p\\ &= (k-1)\cdot(1-p)^{k-2}p^2,\end{aligned}$$

$X+Y$ is *not* Geometric.

(It's actually called *Negative Binomial*. . .)

2. Discrete convolution

Example

Let X, Y be i.i.d. Geom(p) variables. Then

$$\begin{aligned}
\mathfrak{p}_{X+Y}(k) &= \sum_{i=-\infty}^{\infty} \mathfrak{p}_X(k-i)\cdot\mathfrak{p}_Y(i) \\
&= \sum_{i=1}^{k-1} (1-p)^{k-i-1}p\cdot(1-p)^{i-1}p \\
&= (k-1)\cdot(1-p)^{k-2}p^2,
\end{aligned}$$

$X+Y$ is *not* Geometric.

(It's actually called *Negative Binomial*. . .)

2. Discrete convolution

Example (of course. . .)

Let $X \sim \text{Binom}(n, p)$ and $Y \sim \text{Binom}(m, p)$ be independent (notice the same p!). Then $X + Y \sim \text{Binom}(n + m, p)$:

$$p_{X+Y}(k) = \sum_{i=0}^{k} \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} \cdot \binom{m}{i} p^{i}(1-p)^{m-i}$$

$$= p^k \cdot (1-p)^{m+n-k} \sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i}.$$

Now,

$$\sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i} = \binom{n+m}{k} \qquad \rightsquigarrow \quad ,$$

thus $p_{X+Y}(k) = \binom{n+m}{k} p^k \cdot (1-p)^{m+n-k}$, and we are done.

2. Discrete convolution

Example (of course. . .)

Let $X \sim \text{Binom}(n, p)$ and $Y \sim \text{Binom}(m, p)$ be independent (notice the same p!). Then $X + Y \sim \text{Binom}(n + m, p)$:

$$p_{X+Y}(k) = \sum_{i=0}^{k} \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} \cdot \binom{m}{i} p^i (1-p)^{m-i}$$

$$= p^k \cdot (1-p)^{m+n-k} \sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i}.$$

Now,

$$\sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i} = \binom{n+m}{k} \qquad \rightsquigarrow \quad ,$$

thus $p_{X+Y}(k) = \binom{n+m}{k} p^k \cdot (1-p)^{m+n-k}$, and we are done.

2. Discrete convolution

Example (of course...)

Let $X \sim \text{Binom}(n, p)$ and $Y \sim \text{Binom}(m, p)$ be independent (notice the same p!). Then $X + Y \sim \text{Binom}(n + m, p)$:

$$p_{X+Y}(k) = \sum_{i=0}^{k} \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} \cdot \binom{m}{i} p^i (1-p)^{m-i}$$

$$= p^k \cdot (1-p)^{m+n-k} \sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i}.$$

Now,

$$\sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i} = \binom{n+m}{k} \qquad \rightsquigarrow \quad ,$$

thus $p_{X+Y}(k) = \binom{n+m}{k} p^k \cdot (1-p)^{m+n-k}$, and we are done.

2. Discrete convolution

Example (of course. . .)

Let $X \sim \mathrm{Binom}(n,\, p)$ and $Y \sim \mathrm{Binom}(m,\, p)$ be independent (notice the same p!). Then $X + Y \sim \mathrm{Binom}(n + m,\, p)$:

$$p_{X+Y}(k) = \sum_{i=0}^{k} \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} \cdot \binom{m}{i} p^{i}(1-p)^{m-i}$$

$$= p^{k} \cdot (1-p)^{m+n-k} \sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i}.$$

Now,

$$\sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i} = \binom{n+m}{k} \qquad \rightsquigarrow \quad ,$$

thus $p_{X+Y}(k) = \binom{n+m}{k} p^{k} \cdot (1-p)^{m+n-k}$, and we are done.

2. Discrete convolution

Example (of course. . .)

Let $X \sim \text{Binom}(n, p)$ and $Y \sim \text{Binom}(m, p)$ be independent (notice the same p!). Then $X + Y \sim \text{Binom}(n + m, p)$:

$$
\begin{aligned}
p_{X+Y}(k) &= \sum_{i=0}^{k} \binom{n}{k-i} p^{k-i} (1-p)^{n-k+i} \cdot \binom{m}{i} p^{i} (1-p)^{m-i} \\
&= p^{k} \cdot (1-p)^{m+n-k} \sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i}.
\end{aligned}
$$

Now,

$$
\sum_{i=0}^{k} \binom{n}{k-i} \cdot \binom{m}{i} = \binom{n+m}{k} \qquad \rightsquigarrow ,
$$

thus $p_{X+Y}(k) = \binom{n+m}{k} p^{k} \cdot (1-p)^{m+n-k}$, and we are done.

6. Expectation, covariance

Properties of expectations

Covariance

Objectives:

- To explore further properties of expectations of a single and multiple variables
- To define covariance, and use it for computing variances of sums
- To explore and use conditional expectations
- To define and use moment generating functions

Properties of expectations

Recall the respective definitions

$$\mathbf{E}X = \sum_i x_i \mathfrak{p}(x_i) \qquad \text{or} \qquad \mathbf{E}X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

for the discrete and continuous cases. In this chapter we'll explore properties of expected values. We'll always assume that the expectations we talk about exist. Most proofs will be done for the discrete case, but everything in this chapter is very general even beyond discrete and continuous. . .

1. A simple monotonicity property

Proposition

Suppose that $a \le X \le b$ a.s. *Then* $a \le \mathbf{E}X \le b$.

Recall: a.s. means *with probability one*.

2. Expectation of functions of variables

Proposition

Suppose that X and Y are discrete random variables, and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ function. Then

$$\mathbf{E}g(X, Y) = \sum_{i,j} g(x_i, y_j) \cdot \mathfrak{p}(x_i, y_j).$$

There is a very analogous formula for continuous random variables, using *joint densities*, beyond the scope of this unit.

A similar formula holds for functions of 3, 4, etc. random variables.

3. Expectation of sums and differences

Corollary (a very important one)

Let X and Y be any random variables. Then

$$\mathbf{E}(X+Y)=\mathbf{E}X+\mathbf{E}Y \qquad \textit{and} \qquad \mathbf{E}(X-Y)=\mathbf{E}X-\mathbf{E}Y.$$

3. Expectation of sums and differences

Corollary

Let X and Y be such that $X \le Y$ a.s. Then $\mathbf{E}X \le \mathbf{E}Y$.

3. Expectation of sums and differences

Example (sample mean)

Let $X_1, X_2, \dots, X_n$ be identically distributed random variables with mean μ. Their sample mean is

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Its expectation is

$$\mathbf{E}\bar{X} = \mathbf{E}\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\mathbf{E}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}X_i = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

3. Expectation of sums and differences

Example

Let N be a non-negative integer random variable, and

$$X_i := \begin{cases} 1, & \text{if } N \ge i, \\ 0, & \text{if } N < i. \end{cases}$$

Then

$$\sum_{i=1}^{\infty} X_i = \sum_{i=1}^{N} X_i + \sum_{i=N+1}^{\infty} X_i = \sum_{i=1}^{N} 1 + \sum_{i=N+1}^{\infty} 0 = N,$$

$$\mathbf{E} \sum_{i=1}^{\infty} X_i = \sum_{i=1}^{\infty} \mathbf{E} X_i = \sum_{i=1}^{\infty} \mathbf{P}\{N \ge i\} = \mathbf{E} N.$$

3. Expectation of sums and differences

Example

Let N be a non-negative integer random variable, and

$$X_i := \begin{cases} 1, & \text{if } N \geq i, \\ 0, & \text{if } N < i. \end{cases}$$

Then

$$\sum_{i=1}^{\infty} X_i = \sum_{i=1}^{N} X_i + \sum_{i=N+1}^{\infty} X_i = \sum_{i=1}^{N} 1 + \sum_{i=N+1}^{\infty} 0 = N,$$

$$\mathbf{E}\sum_{i=1}^{\infty} X_i = \sum_{i=1}^{\infty} \mathbf{E}X_i = \sum_{i=1}^{\infty} \mathbf{P}\{N \geq i\} = \mathbf{E}N.$$

Covariance

In this part we investigate the relation of independence to expected values. It will give us some (not perfect) way of measuring independence.

Again, we assume that all the expectations we talk about exist.

1. Independence

We start with a simple observation:

Proposition

Let X and Y be independent random variables, and g, h functions. Then

$$\mathbf{E}\big(g(X)\cdot h(Y)\big) = \mathbf{E}g(X)\cdot\mathbf{E}h(Y).$$

2. Covariance

Then, the following is a natural object to measure independence:

Definition

The covariance of the random variables X and Y is

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}X) \cdot (Y - \mathbf{E}Y)].$$

Before exploring its properties, notice

$$\mathbf{Cov}(X, Y) = \mathbf{E}XY - \mathbf{E}X \cdot \mathbf{E}Y.$$

2. Covariance

Remark

From either forms

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}X) \cdot (Y - \mathbf{E}Y)] = \mathbf{E}XY - \mathbf{E}X \cdot \mathbf{E}Y$$

it is clear that for independent random variables,

$$\mathbf{Cov}(X, Y) = 0.$$

2. Covariance

Example

This is not true the other way around: let

$$X := \begin{cases} -1, \text{ with prob. } \frac{1}{3}, \\ 0, \text{ with prob. } \frac{1}{3}, \\ 1, \text{ with prob. } \frac{1}{3}, \end{cases} \qquad Y := \begin{cases} 0, \text{ if } X \neq 0, \\ 1, \text{ if } X = 0. \end{cases}$$

Then $X \cdot Y = 0$ and $\mathbf{E}X = 0$, thus $\mathbf{Cov}(X, Y) = 0$, but these variables are clearly not independent.

2. Covariance

Proposition (properties of covariance)

Fix a_i, b, c_j, d real numbers. Covariance is

- *positive semidefinite:* $\mathbf{Cov}(X, X) = \mathbf{Var}X \geq 0$,
- *symmetric:* $\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$,
- *almost bilinear:*

$$\mathbf{Cov}\Big(\sum_i a_i X_i + b, \sum_j c_j Y_j + d\Big) = \sum_{i,j} a_i c_j \mathbf{Cov}(X_i, Y_j).$$

3. Variance

Now we can answer a long overdue question: what happens to the variance of sums or random variables?

Proposition (variance of sums)

Let $X_1, X_2, \ldots, X_n$ be random variables. Then

$$\mathbf{Var}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathbf{Var} X_i + 2 \sum_{1\le i<j\le n} \mathbf{Cov}(X_i, X_j).$$

In particular, variances of independent random variables are additive.

No additivity, however, of variances in general.

3. Variance

Remark

Notice that for independent variables,

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

3. Variance

Example (variance of the sample mean)

Suppose that X_i's are i.i.d., each of variance σ^2. Recall the definition

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

of the sample mean. Its variance is

$$\begin{aligned}\mathbf{Var}\bar{X} &= \mathbf{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\mathbf{Var}\Big(\sum_{i=1}^{n} X_i\Big) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{Var}X_i = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.\end{aligned}$$

Decreases with n, that's why we like sample averages.

3. Variance

Example (unbiased sample variance)

Suppose we are given the values of $X_1, X_2, \ldots X_n$ of an i.i.d. sequence of random variables with mean μ and variance σ^2. We know that the *sample mean* $\bar{X}$

- has mean μ, and
- small variance ($\frac{\sigma^2}{n}$).

Therefore it serves as a good *estimator* for the value of μ. But what should we use to estimate the variance σ^2? This quantity is the *unbiased sample variance*:

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

3. Variance

Example (unbiased sample variance)

We'll compute its expected value (and take it for granted that it doesn't fluctuate much):

$$\begin{aligned}\mathbf{E}S^2 = \mathbf{E}\Big(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\Big) &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbf{E}(X_i-\bar{X})^2\\ &= \frac{n}{n-1}\mathbf{E}(X_1-\bar{X})^2\end{aligned}$$

by symmetry. Next notice that $\mathbf{E}(X_1-\bar{X}) = \mathbf{E}X_1 - \mathbf{E}\bar{X} = \mu - \mu = 0$, therefore

$$\mathbf{E}(X_1-\bar{X})^2 = \mathbf{Var}(X_1-\bar{X}) = \mathbf{Var}X_1 + \mathbf{Var}\bar{X} - 2\mathbf{Cov}(X_1,\bar{X}).$$

3. Variance

Example (unbiased sample variance)

From here, $\mathbf{Var}X_1 = \sigma^2$, $\mathbf{Var}\bar{X} = \frac{\sigma^2}{n}$, only need to calculate

$$\mathbf{Cov}(X_1, \bar{X}) = \mathbf{Cov}\Big(X_1, \frac{1}{n}\sum_j X_j\Big) = \frac{1}{n}\mathbf{Cov}\Big(X_1, \sum_j X_j\Big)$$
$$= \frac{1}{n}\sum_j \mathbf{Cov}\Big(X_1, X_j\Big) = \frac{1}{n}\mathbf{Cov}(X_1, X_1) = \frac{\sigma^2}{n}.$$

Putting everything together,

$$\mathbf{E}S^2 = \frac{n}{n-1}\Big(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\Big) = \frac{n}{n-1}\frac{(n-1)\sigma^2}{n} = \sigma^2.$$

3. Variance

Example (unbiased sample variance)

From here, $\mathbf{Var}X_1 = \sigma^2$, $\mathbf{Var}\bar{X} = \frac{\sigma^2}{n}$, only need to calculate

$$\mathbf{Cov}(X_1, \bar{X}) = \mathbf{Cov}\Big(X_1, \frac{1}{n}\sum_j X_j\Big) = \frac{1}{n}\mathbf{Cov}\Big(X_1, \sum_j X_j\Big)$$
$$= \frac{1}{n}\sum_j \mathbf{Cov}\big(X_1, X_j\big) = \frac{1}{n}\mathbf{Cov}(X_1, X_1) = \frac{\sigma^2}{n}.$$

Putting everything together,

$$\mathbf{E}S^2 = \frac{n}{n-1}\Big(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\Big) = \frac{n}{n-1}\frac{(n-1)\sigma^2}{n} = \sigma^2.$$

3. Variance

Example (unbiased sample variance)

From here, $\mathbf{Var}X_1 = \sigma^2$, $\mathbf{Var}\bar{X} = \frac{\sigma^2}{n}$, only need to calculate

$$\mathbf{Cov}(X_1, \bar{X}) = \mathbf{Cov}\Big(X_1, \frac{1}{n}\sum_j X_j\Big) = \frac{1}{n}\mathbf{Cov}\Big(X_1, \sum_j X_j\Big)$$

$$= \frac{1}{n}\sum_j \mathbf{Cov}\big(X_1, X_j\big) = \frac{1}{n}\mathbf{Cov}(X_1, X_1) = \frac{\sigma^2}{n}.$$

Putting everything together,

$$\mathbf{E}S^2 = \frac{n}{n-1}\Big(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\Big) = \frac{n}{n-1}\,\frac{(n-1)\sigma^2}{n} = \sigma^2.$$

3. Variance

Example (unbiased sample variance)

From here, $\mathbf{Var}X_1 = \sigma^2$, $\mathbf{Var}\bar{X} = \frac{\sigma^2}{n}$, only need to calculate

$$\mathbf{Cov}(X_1, \bar{X}) = \mathbf{Cov}\Big(X_1, \frac{1}{n}\sum_j X_j\Big) = \frac{1}{n}\mathbf{Cov}\Big(X_1, \sum_j X_j\Big)$$

$$= \frac{1}{n}\sum_j \mathbf{Cov}\big(X_1, X_j\big) = \frac{1}{n}\mathbf{Cov}(X_1, X_1) = \frac{\sigma^2}{n}.$$

Putting everything together,

$$\mathbf{E}S^2 = \frac{n}{n-1}\Big(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\Big) = \frac{n}{n-1}\,\frac{(n-1)\sigma^2}{n} = \sigma^2.$$

7. Law of Large Numbers, Central Limit Theorem

Markov's, Chebyshev's inequality
Central Limit Theorem

Objectives:

- To get familiar with general inequalities like Markov's and Chebyshev's
- To (almost) prove and use the Weak Law of Large Numbers
- To (almost) prove and use the Central Limit Theorem

Markov's, Chebyshev's inequality

Knowing a distribution of a random variable makes it possible to compute its moments. Vice-versa, knowing a few moments gives some bounds on certain probabilities. We'll explore such bounds in this part.

Our bounds here will be very general, and that makes them very useful in theoretical considerations. The price to pay is that they are often not sharp enough for practical applications.

1. Markov's inequality

Theorem (Markov's inequality)

Let X be a non-negative random variable. Then for all $a > 0$ reals,

$$\mathbf{P}\{X \geq a\} \leq \frac{\mathbf{E}X}{a}.$$

Of course this inequality is useless for $a \leq \mathbf{E}X$.

2. Chebyshev's inequality

Theorem (Chebyshev's inequality)

Let X be a random variable with mean μ and variance σ^2 both finite. Then for all $b > 0$ reals,

$$\mathbf{P}\{|X - \mu| \geq b\} \leq \frac{\mathbf{Var}X}{b^2}.$$

Of course this inequality is useless for $b \leq \mathrm{SD}\,X$.

3. Examples

Example (it's not sharp)

Let $X \sim \mathrm{U}(0, 10)$. Then $\mu = 5$ and $\mathbf{Var}X = \frac{10^2}{12}$, and Chebyshev's inequality tells us

$$\mathbf{P}\{|X - 5| \geq 4\} \leq \frac{\mathbf{Var}X}{4^2} = \frac{10^2}{12 \cdot 4^2} = \frac{25}{48} \simeq 0.52.$$

This is certainly valid, but the truth is

$$\mathbf{P}\{|X - 5| \geq 4\} = \mathbf{P}\{X \leq 1\} + \mathbf{P}\{X \geq 9\} = \frac{1}{10} + \frac{1}{10} = 0.2.$$

3. Examples

Example (it's not sharp)

Let $X \sim \mathrm{U}(0,\, 10)$. Then $\mu = 5$ and $\mathbf{Var}X = \frac{10^2}{12}$, and Chebyshev's inequality tells us

$$\mathbf{P}\{|X - 5| \geq 4\} \leq \frac{\mathbf{Var}X}{4^2} = \frac{10^2}{12 \cdot 4^2} = \frac{25}{48} \simeq 0.52.$$

This is certainly valid, but the truth is

$$\mathbf{P}\{|X - 5| \geq 4\} = \mathbf{P}\{X \leq 1\} + \mathbf{P}\{X \geq 9\} = \frac{1}{10} + \frac{1}{10} = 0.2.$$

3. Examples

Example (it's not sharp)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. According to Chebyshev's inequality

$$\mathbf{P}\{|X - \mu| \geq 2\sigma\} \leq \frac{\mathbf{Var}X}{(2\sigma)^2} = \frac{\sigma^2}{4\sigma^2} = 0.25,$$

while in fact

$$\begin{aligned}\mathbf{P}\{|X - \mu| \geq 2\sigma\} &= \mathbf{P}\Big\{\frac{X - \mu}{\sigma} \leq -2\Big\} + \mathbf{P}\Big\{\frac{X - \mu}{\sigma} \geq 2\Big\}\\ &= \Phi(-2) + 1 - \Phi(2) = 2 - 2\Phi(2) \simeq 0.0456.\end{aligned}$$

4. A turbo version

While Markov's inequality is not sharp in many cases, there are many ways of strengthening it. Here is a commonly used argument called *Chernoff bound*. Let X be any random variable with finite moment generating function M. Then for any $c \in \mathbb{R}$ and $\lambda > 0$,

$$\mathbf{P}\{X \geq c\} = \mathbf{P}\{e^{\lambda X} \geq e^{\lambda c}\} \leq \frac{\mathbf{E}e^{\lambda X}}{e^{\lambda c}} = e^{-\lambda c}M(\lambda),$$

where Markov's inequality was applied on $e^{\lambda X} \geq 0$. Minimising the right hand-side in λ can give rather sharp estimates in many cases.

Central Limit Theorem

The WLLN tells us that the sample mean of an i.i.d. sequence is close to the expectation of the variables. A second, finer approach will be the Central Limit Theorem. It will tell us the order of magnitude of the distance between the sample mean and the true mean of our random variables.

Central Limit Theorem

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_2, \dots$ be i.i.d. random variables with both their mean μ and variance σ^2 finite. Then for every real $a < b$,

$$\mathbf{P}\left\{a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} \le b\right\} \underset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a).$$

Remark

Notice the mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$ of the sum $X_1 + X_2 + \cdots + X_n$.

Central Limit Theorem

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_2, \ldots$ be i.i.d. random variables with both their mean μ and variance σ^2 finite. Then for every real $a < b$,

$$\mathbf{P}\left\{a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} \le b\right\} \underset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a).$$

Remark

Notice the mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$ of the sum $X_1 + X_2 + \cdots + X_n$.

Central Limit Theorem

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_2, \ldots$ be i.i.d. random variables with both their mean μ and variance σ^2 finite. Then for every real $a < b$,

$$\mathbf{P}\left\{a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} \le b\right\} \underset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a).$$

Remark

Notice the mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$ of the sum $X_1 + X_2 + \cdots + X_n$.

Central Limit Theorem

Example

The number of students who apply to a course is $X \sim \mathrm{Poi}(100)$. Estimate the probability that more than 120 apply.

Solution

Working with sums of 121 terms with huge factorials is not very good for one's health. Try CLT instead. But: where are the i.i.d. random variables?

Central Limit Theorem

Example

The number of students who apply to a course is $X \sim \text{Poi}(100)$. Estimate the probability that more than 120 apply.

Solution

Working with sums of 121 terms with huge factorials is not very good for one's health. Try CLT instead. But: where are the i.i.d. random variables?

Recall that the Poisson has the nice convolution property. Thus,

$$X \stackrel{\text{d}}{=} \sum_{i=1}^{100} X_i,$$

where X_i's are i.i.d. $\text{Poi}(1)$. These are completely hypothetical variables, they are nowhere present in the problem.

Central Limit Theorem

Solution (…cont'd)

Then ($\mathbf{E}X_i = 1$, $\mathbf{Var}X_i = 1$):

$$\mathbf{P}\{X > 120\} = \mathbf{P}\Big\{\sum_{i=1}^{100} X_i > 120\Big\} = \mathbf{P}\Big\{\sum_{i=1}^{100} X_i > 120.5\Big\}$$

$$= \mathbf{P}\left\{\frac{\sum_{i=1}^{100} X_i - 100 \cdot 1}{\sqrt{100 \cdot 1}} > \frac{120.5 - 100 \cdot 1}{\sqrt{100 \cdot 1}}\right\}$$

$$\simeq 1 - \Phi(2.05) \simeq 0.0202.$$