Module 9.1: Random Variables

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Random variables

A **random variable** on (Ω, Pr) is a function $X : \Omega \to \mathbb{R}$.

Denote $\mathsf{Val}(X) = \{x \in \mathbb{R} \mid \exists w \in \Omega \;\; X(w) = x\}$.
(The set of values **taken** by X.)

Like Ω, $\mathsf{Val}(X)$ is also a finite set.

Denote with x the real values that X takes and with $X = x$ the **event** $\{w \in \Omega \mid X(w) = x\}$. Its probability $\mathrm{Pr}[X = x]$ is of particular interest.

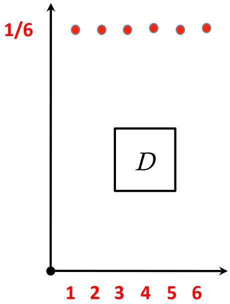
The **distribution** of the random variable X is the function

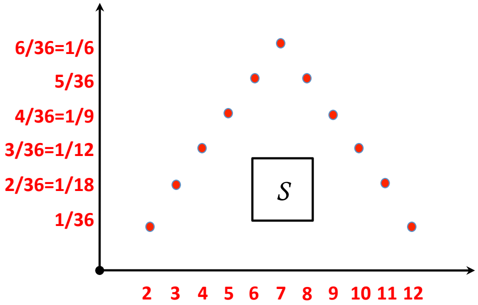
$$f : \mathsf{Val}(X) \to [0, 1] \quad \text{where} \quad f(x) = \mathrm{Pr}[X = x].$$

Examples of random variables I

Problem. We roll a fair die. What is the distribution of the random variable D that returns the number shown by the die? We roll two fair dice. What is the distribution of the random variable S which returns the sum of the numbers the two dice show?

Answer. Note that $\mathrm{Val}(D) = [1..6]$ and $\mathrm{Val}(S) = [2..12]$. Here are the graphs of the distribution functions:

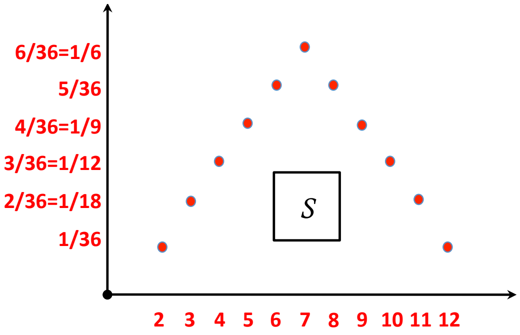




Examples of random variables II

Problem. Let S be the r.v. (short for random variable) which returns the sum of the numbers two fair dice show when rolled. Calculate $\Pr[5 \le S \le 9]$.

Answer. We use the distribution of S:



$.\Pr[5 \le S \le 9] \ =$
$.\Pr[S = 5] + \Pr[S = 6] +$
$.\Pr[S = 7] + \Pr[S = 8] +$
$.\Pr[S = 9] \ =$
$.(1/9) + (5/36) + (1/6) +$
$.(5/36) + (1/9) = 2/3$

Activity : Three-dice slot machine

A casino has a slot machine that shows three fair dice rolled independently. The player wins if the sum of the three values the dice show is in $[5..8]$ or in $[13..17]$. Note that

$$|[5..8]| + |[13..17]| = (8 - 5 + 1) + (17 - 13 + 1) = 4 + 5 = 9$$

out of a total of

$$|[3..18]| = 18 - 3 + 1 = 16$$

thus a player might think this is a good bet since more than half of the possible values the sum takes are winning values. But casinos always win in the long run!

There may be 9 winning values for the sum out of 16 but the values are **not** distributed uniformly! Let's compute the probability that the sum gives a winning value.

ACTIVITY : Three-dice slot machine (continued)

The probability space is uniform with $6 \cdot 6 \cdot 6 = 216$ outcomes.

Therefore, to compute the probabilities of the distribution it suffices to count the number of outcomes corresponding to each value taken by the sum of three dice.

Question. List all the possible ways in which a sum of 5 can be obtained. Count the number of ways. What is the connection with "stars and bars"?

In the video, there is a box here for learners to put in an answer to the question above. As you read these notes, try it yourself using pen and paper!

Activity : Three-dice slot machine (continued)

Answer. This is the same as the number of ways of distributing 5 indistinguishable coins to 3 children in such a way that each child gets at least a coin. We did essentially this in the lecture segment "Stars and bars" except that we distributed 11 coins and here we have 5. The number of ways (here) is:

$$\binom{(5-3)+(3-1)}{3-1} = \binom{4}{2} = 6$$

And here are the six ways of getting a sum of 5:

$$3+1+1 = 1+3+1 = 1+1+3 = 2+2+1 = 2+1+2 = 1+2+2 \; (=5)$$

Activity : Three-dice slot machine (continued)

By thoroughly analyzing sums of three numbers, each from [1..6], we fill in the following table. In the first column is the value of the sum and in the second is the number of outcomes that gives the sum that value (which we can then divide by 216 to obtain the corresponding probability).

sum	# outc.
3	1
4	3
5	6
6	10
7	15
8	21

sum	# outc.
9	25
10	27
11	27
12	25

sum	# outc.
13	21
14	15
15	10
16	6
17	3
18	1

Activity : Three-dice slot machine (continued)

Let U be the random variable that returns the sum of the values shown by three fair dice rolled independently. Using the table we calculate:

$$\Pr[5 \leq U \leq 8] = \frac{6}{216} + \frac{10}{216} + \frac{15}{216} + \frac{21}{216} = \frac{52}{216}$$
$$\Pr[13 \leq U \leq 17] = \frac{21}{216} + \frac{15}{216} + \frac{10}{216} + \frac{6}{216} + \frac{3}{216} = \frac{55}{216}$$

Thus the probability of getting a winning value is:

$$\frac{52}{216} + \frac{55}{216} = \frac{107}{216} < \frac{1}{2}$$

The casino always wins (in the long run).

Probabilities add up to 1

For the distribution of S the sum of **all** the $12 - 2 + 1 = 11$ probabilities is 1. This is true for any random variable X:

Proposition. $\sum_{x \in \mathsf{Val}(X)} \Pr[X = x] \; = \; 1.$

Proof. Recall that $(X = x) \; = \; \{w \in \Omega \mid X(w) = x\}.$

The events $(X = x)$ for $x \in \mathsf{Val}(X)$ are pairwise disjoint.

Also, $\bigcup_{x \in \mathsf{Val}(X)} (X = x) \; = \; \Omega.$

By **P2** (the addition rule) and by **P1**

we have $\sum_{x \in \mathsf{Val}(X)} \Pr[X = x] \; = \; \Pr[\Omega] \; = \; 1.$

Uniform r.v. and distribution

Let $v_1, \ldots, v_n$ be n distinct values in $\mathbb{R}$.

Given $(\Omega, \Pr)$, an r.v. $U : \Omega \to \mathbb{R}$ is **uniform** with these values

when $\text{Val}(U) = \{v_1, \ldots, v_n\}$

and $\Pr[U = v_i] = 1/n$ for $i = 1, \ldots, n$.

The corresponding distribution

. $f : \{v_1, \ldots, v_n\} \to [0, 1] \qquad f(v_i) = 1/n$ for $i = 1, \ldots, n$

is also called **uniform**.

The r.v. D associated with a fair die is uniform

with $n = 6$ and $v_i = i$ for $i = 1, \ldots, 6$.

Bernoulli r.v. and distribution

Recall **Bernoulli trials**. Similarly, we can define:

Given $(\Omega, \Pr)$, an r.v. $X : \Omega \to \mathbb{R}$

with $\mathsf{Val}(X) = \{0, 1\}$

and $\Pr[X = 1] = p$

is called a **Bernoulli random variable** with parameter p.

A Bernoulli r.v. X defines implicitly a Bernoulli trial where "success" is $X = 1$ and "failure" is $X = 0$.

The corresponding distribution

$f : \{0, 1\} \to [0, 1]$ where $f(1) = p$ and $f(0) = 1 - p$

is called a **Bernoulli distribution** with parameter p.

Module 9.2: Expectation

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Expectation (expected value)

Average, or **mean**, value returned by a random variable. Such an average should take into account that some values may "weigh" more than others. The **weights** are given by the probability distribution!

Notation $\mathsf{E}[X]$. Two candidates, but they give the same answer:

Proposition. For a random variable X defined on $(\Omega, \Pr)$ we have

$$\mathsf{E}[X] = \sum_{x \in \mathsf{Val}(X)} x \cdot \Pr[X = x] = \sum_{w \in \Omega} X(w) \cdot \Pr[w]$$

The first expression corresponds directly to a **weighted average** of the values taken by random variable. Recall that the weights $\Pr[X = x]$ sum up to 1.

The second expression takes the average of values by outcomes they map from. Multiple outcomes may be mapped to the same value taken by the r.v.

Expected number shown by a die

Problem. Compute $\mathrm{E}[D]$ where D is the number shown by a fair die.

Answer.

$$1 \cdot (1/6) + 2 \cdot (1/6) + 3 \cdot (1/6) + 4 \cdot (1/6) + 5 \cdot (1/6) + 6 \cdot (1/6) =$$
$$= (1+2+3+4+5+6)/6 = 3.5$$

Since all weights are the same, this is the usual average. More generally:

Proposition. The expectation of a uniform r.v. that takes values $v_1, \ldots, v_n$ is

$$\frac{v_1 + \cdots + v_n}{n}$$

Expectation of the Bernoulli r.v.

Problem. Compute the expectation of the Bernoulli r.v. X with parameter p.

Answer. Recall that $\mathsf{Val}(X) = \{0, 1\}$

and the distribution is $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$.

Then $\mathsf{E}[X] = 1 \cdot \Pr[X = 1] + 0 \cdot \Pr[X = 0] = 1 \cdot p + 0 \cdot (1 - p) = p$.

Expectation of a constant r.v.

Let $c \in \mathbb{R}$ and let (Ω, Pr) be a probability space.

Consider the r.v. $C : \Omega \to \mathbb{R}$ such that for all outcomes $w \in \Omega$ we have $C(w) = c$.

The distribution is trivial $\mathrm{Pr}[C = c] \;=\; \sum_{w\in\Omega} \mathrm{Pr}[w] \;=\; 1.$

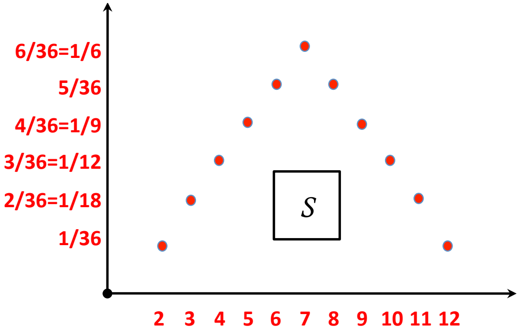
Proposition. $\mathrm{E}[C] \;=\; c$ (Of course!)

Proof.

$$\mathrm{E}[C] \;=\; \sum_{w\in\Omega} C(w)\cdot \mathrm{Pr}[w] = \sum_{w\in\Omega} c\cdot \mathrm{Pr}[w] = c\cdot \sum_{w\in\Omega} \mathrm{Pr}[w] \;=\; c.$$

ACTIVITY : Expected sum of two fair dice

In this activity we will compute the expectation of the random variable S that returns the sum of two fair dice. Recall the distibution of S



Question. List all the outcomes that correspond to $S = 6$. How many are there? What does this imply for $\Pr[S = 6]$?

In the video, there is a box here for learners to put in an answer to the question above. As you read these notes, try it yourself using pen and paper!

Activity : Expected sum of two fair dice (continued)

Answer. The outcomes are $(1,5)$, $(2,4)$, $(3,3)$ $(4,2)$, $(5,1)$. There are 5 of them. Since the probability space is uniform with 36 outcomes it follows that $\Pr[S = 6] = 5/36$, as the figure indicated.

Now, we compute

$$\begin{aligned} E[S] &= 2\cdot(1/36) + 3\cdot(2/36) + 4\cdot(3/36) + 5\cdot(4/36) + 6\cdot(5/36) \\ &+ 7\cdot(6/36) + 8\cdot(5/36) + 9\cdot(4/36) + 10\cdot(3/36) + 11\cdot(2/36) \\ &+ 12\cdot(1/36) = \frac{252}{36} = 7 \end{aligned}$$

Recall that earlier in this segment we computed the expectation of the value a single die shows as 3.5. Now observe that $E[S] = 7 = 3.5 + 3.5$ so it's the sum of the expected values of each die! Is this a coincidence? No, as we will see when we learn about the **linearity of expectation**.

Expected sum of three dice

Problem. Compute $\mathsf{E}[T]$ where T is the sum of the numbers shown by **three** fair dice rolled together.

Answer. Clearly $\mathsf{Val}(T) = [3..18]$ so T takes $18 - 3 + 1 = 16$ values. For values $T = 3, 4, 5, 6, 7, 8$ we can use stars and bars. Let's find $\Pr[T = 9]$.

$9 = 1 + 2 + 6 = 1 + 3 + 5 = 1 + 4 + 4 = 2 + 2 + 5 = 2 + 3 + 4 = 3 + 3 + 3$

$3 + 3 + 3$ is one outcome, probability $1/216$. $1 + 2 + 6$, $1 + 3 + 5$, and $2 + 3 + 4$ each correspond to $3!$=6 outcomes so each has probability $6/216$. $1 + 4 + 4$ and $2 + 2 + 5$ each correspond to $3!/2!1!$=3 outcomes so each has probability $3/216$.

$\Pr[T = 9]$ is therefore $(1/216)$+3$(6/216)$+2$(3/216)$=$25/216$.

This is too hard! I give up! Is there an easier way?

Module 9.3: Linearity of Expectation

MCIT Online - CIT592 - Professor Val Tannen

LECTURE NOTES

Sum of two r.v.'s

Let $X, Y : \Omega \to \mathbb{R}$ be two random variables defined on the same probability space $(\Omega, \Pr)$. Their **sum**, notation $X + Y$, is the random variable on the same space defined by $(X + Y)(w) = X(w) + Y(w)$ for all $w \in \Omega$.

In general there is no simple relation between the distribution of $X + Y$ and those of X and Y (see examples).

Example. In the green-purple dice space consider the r.v. G that takes as values the numbers shown by the green die.

G has the same distribution as D (see earlier). Same with P (purple die).

The sum $G + P$ is the same as the r.v. S (see earlier).

However, S is **not** $D + D$ because S and D are defined on **different** probability spaces.

Scalar multiplication of an r.v.

Let $c \in \mathbb{R}$ and $X : \Omega \to \mathbb{R}$ be an r.v. The **multiplication by** c of X, notation $c\,X$ is the random variable on the same space defined by $(c\,X)(w) = c \cdot X(w)$ for all $w \in \Omega$. Its distribution is closely related to that of X (unless $c = 0$).

Example. Let D be the single die r.v. we saw earlier. $2D$ takes as values $2, 4, 6, 8, 10, 12$ with the uniform distribution.

Example. We flip a fair coin n times. Let H, respectively T, be the r.v. that returns the number of heads, respectively tails, observed.

Consider H, T and $(-1)T$.

$H + T$ is quite trivial: takes one value, n, with probability 1!

But $H - T = H + (-1)T$ is very interesting: intuitively $\mathrm{E}[H - T] = 0$.

Linearity of expectation

Proposition. Let $X_1, \ldots, X_n$ be random variables on the same probability space (Ω, Pr) and let $c_1, \ldots, c_n \in \mathbb{R}$. Then:

$$\mathsf{E}[c_1 X_1 + \cdots + c_n X_n] = c_1 \mathsf{E}[X_1] + \cdots + c_n \mathsf{E}[X_n]$$

Proof.

$$\begin{aligned}
\mathsf{E}\left[\sum_{i=1}^{n} c_i X_i\right] &= \sum_{w \in \Omega} \left(\sum_{i=1}^{n} c_i X_i\right)(w) \cdot \mathrm{Pr}[w] \\
&= \sum_{w \in \Omega} \left(\sum_{i=1}^{n} c_i X_i(w)\right) \cdot \mathrm{Pr}[w] \\
&= \sum_{i=1}^{n} c_i \left(\sum_{w \in \Omega} X_i(w) \cdot \mathrm{Pr}[w]\right) = \sum_{i=1}^{n} c_i \mathsf{E}[X_i]
\end{aligned}$$

Applications of linearity of expectation

Problem. Compute $\mathrm{E}[S]$ where S is the r.v. that returns the sum of numbers shown by two fair dice rolled together.

Answer. Assume green-purple dice. Since G and P have the same values and distribution as D: $\mathrm{E}[G] = \mathrm{E}[P] = \mathrm{E}[D] = 3.5$. Now, using linearity of expectation, $\mathrm{E}[S] = \mathrm{E}[G + P] = \mathrm{E}[G] + \mathrm{E}[P] = 3.5 + 3.5 = 7$.

Problem. We flip a fair coin n times. Compute $\mathrm{E}[H - T]$, the difference between the number of heads and of tails observed.

Answer. H and T have the same distribution, therefore $\mathrm{E}[H] = \mathrm{E}[T]$. Then, by linearity of expectation,
$\mathrm{E}[H - T] = \mathrm{E}[H + (-1)T] = \mathrm{E}[H] + (-1)\mathrm{E}[T] = \mathrm{E}[H] - \mathrm{E}[T] = 0$.

Quiz

Recall that we tried to compute the expectation of the sum of three fair dice and gave up because the calculations were too onerous? Now, armed with **linearity of expectation**, answer the following.

We roll a fair die r times and add up the numbers shown. What is the expected value of this sum?

(A) 10.5

(B) $3.5/r$

(C) $7r/2$

Answer

We roll a fair die r times and add up the numbers shown. What is the expected value of this sum?

(A) 10.5

Incorrect. This is the expectation of the sum of **three** dice, not r.

(B) $3.5/r$

Incorrect. No reason to divide by r.

(C) $7r/2$

Correct. We add up r times the expectation of the value of a single die: $(r)(3.5) = (r)(7/2) = 7r/2$.

More information

We roll a fair die r times independently. Let D_i be the rv that returns the value shown by the i'th roll and let $W = D_1 + \ldots + D_r$ be the sum of the r values shown.

By linearity of expectation $\mathrm{E}[W] = \mathrm{E}[D_1] + \cdots + \mathrm{E}[D_r]$.

So far we did not use the fact that the rolls are independent (linearity of expectation does not require any such assumption).

But now, because the rolls are independent we can assert that the probability distribution of each D_i is the same as the probability distribution of a single die roll, D. And we have calculated earlier that $\mathrm{E}[D] = 3.5$.

Therefore $\mathrm{E}[D_i] = 3.5$ for $i = 1, \ldots, r$ thus $\mathrm{E}[W] = (r)(3.5)$.

Module 9.4: Indicators

MCIT Online - CIT592 - Professor Val Tannen

LECTURE NOTES

Indicator random variables

Let A be an event in a probability space $(\Omega, \Pr)$. The **indicator** random variable of the event A, notation I_A, is defined by

$$I_A(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases}$$

Note that I_A is a Bernoulli random variable with success probability $\Pr[I_A = 1] = \Pr[A]$.

As we have shown before, its expectation is $E[I_A] = \Pr[I_A = 1] = \Pr[A]$.

Number of heads in n coin flips I

Problem. We flip a biased coin n times with heads probability p. Let H be the r.v. that returns the number of heads observed. Compute $\mathrm{E}[H]$.

Answer (first attempt). We will try to use the formula for expectation.

The outcomes are 2^n sequences of length n of `H`'s and `T`'s. An outcome with k `H`'s has probability $p^k q^{n-k}$ where $q = 1 - p$.

There are $\binom{n}{k}$ outcomes with k `H`'s. Therefore the probability of the event "k heads observed" is $\binom{n}{k} p^k q^{n-k}$.

Using the formula for expectation:

$$\mathrm{E}[H] = \sum_{k=0}^{n} k \cdot \Pr[H = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

Now what?

Number of heads in n coin flips II

Answer (second attempt). We are going to use **linearity of expectation**.

In the probability space of the n flips that we just saw, let H_k be the event "the k'th flip is $\mathtt{H}$" for $k = 1, \ldots, n$ and let I_k be the **indicator** random variable of the event H_k.

Clearly, $H = I_1 + \cdots + I_n$.

By linearity of expectation $\mathrm{E}[H] = \mathrm{E}[I_1] + \cdots + \mathrm{E}[I_n]$.

We established earlier that $\mathrm{E}[I_k] = \Pr[H_k]$.

Since the flips are independent $\Pr[H_k] = p$.

In conclusion $\mathrm{E}[H] = p + \cdots + p = n \cdot p$.

Interestingly,

$$\sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = np$$

Balls to bins "on average"

Problem. We throw k balls into n bins. What is the number of balls that end up in Bin 1, "on average"?

Answer. We want $\mathrm{E}[X]$ where X is the random variable that returns the number of balls that end up in Bin 1.

We can express X as $X = I_1 + \cdots + I_k$.

where I_i is the **indicator** r.v. of the event $L_i =$ "ball i ends up in Bin 1".

Recall from the discussion that we introduced the model that $\Pr[L_i] = 1/n$.

Therefore $\mathrm{E}[I_i] = \Pr[L_i] = 1/n$.

Using **linearity of expectation** we obtain $\mathrm{E}[X] = (1/n) + \ldots + (1/n) = k/n$.

ACTIVITY : Balls in bins and biased coins

The balls into bins problem we just saw can be seen as a particular case of the biased coin one that precedes it.

Indeed for each ball throw consider only two outcomes: (1) falls in Bin 1, and (2) does not fall in Bin 1. Since each of the n bins is equally likely to receive the ball, outcome (1) has probability $1/n$ (while outcome (2) has probability $(n-1)/n$).

Therefore each ball throw is like flipping biased coin with $1/n$ probability of showing heads.

We can then apply the calculation of the expected number of heads in multiple flips of biased coin. Here we have k throws so the expected number of balls in Bin 1 will be $(k)(1/n) = k/n$, the same answer.

Module 9.5: More Expectations

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Expectation: success followed by failure I

Problem. Consider again n IID Bernoulli trials, each with probability of success p. Let C be the random variable that returns the number of successes which are followed immediately (in the next trial) by failure. Compute the expectation of C.

Answer. Using **indicator** r.v.'s and **linearity of expectation**.

Notice that $C = I_1 + \cdots + I_{n-1}$

where, for $k = 1, \ldots, n-1$, I_k is the indicator r.v. of the event A_k = "success in trial k and failure in trial $k+1$"

Recall that $\mathrm{E}[I_k] = \Pr[I_k = 1] = \Pr[A_k]$.

Expectation: success followed by failure II

Answer (continued).

Since trial k is independent of trial $k+1$, $\Pr[A_k] = p(1-p)$.

Hence, $\mathrm{E}[C] = p(1-p) + \cdots + p(1-p) = (n-1)p(1-p)$.

Notice that for $p \neq 0$ or 1 the events $A_1, \ldots, A_{n-1}$ are not mutually independent. They are not even pairwise independent, because $A_k \not\perp A_{k+1}$!

Indeed, $\Pr[A_k] \cdot \Pr[A_{k+1}] = p^2(1-p)^2$.

However, $\Pr[A_k \cap A_{k+1}] = \Pr[\emptyset] = 0$.

The hats and gangsters problem I

Problem. n hat-wearing gangsters leave their distinguishable hats with a restaurant cloakroom attendant. After the meal, the attendant gives them back their hats **randomly**. How many of the gangsters get back their own hat, "on average"?

Answer. We assume that the returned hats form a **random permutation** of n elements. We studied them in a previous segment.

By definition, these form a uniform probability space: each outcome has probability $1/n!$.

We also saw that the probability that a given element occurs in a given position of a random permutation is $1/n$.

The hats and gangsters problem II

Answer (continued). Let X be the random variable that returns the number of gangsters that get back their own hat. We wish to compute $\mathrm{E}[X]$.

Figuring out the distribution X is complicated. Therefore, computing $\mathrm{E}[X]$ directly from the definition of expectation is also complicated. **Linearity of expectation** gives us, again, an easy solution.

Clearly $X = X_1 + \cdots + X_n$

where X_k is the indicator r.v. of the event E_k = "Gangster k gets back his own hat".

Recall again that $\mathrm{E}[X_k] = \Pr[X_k = 1] = \Pr[E_k]$

As I reminded you, we earlier calculated $\Pr[E_k] = 1/n$.

Therefore, by linearity of expectation $\mathrm{E}[X] = n \cdot (1/n) = 1$

On average, only **one** gangster gets back his own hat!

“Sorting” by random permutation

The hats and gangsters problem has a computer science equivalent!

Suppose I want to sort a set of n distinct real numbers.

I try to sort them by producing a **random permutation** of the elements of the set.

My brilliant intuition tells me that this cannot be **too** bad, can it? I should get, on average, maybe half of the numbers in their place? Am I right?

I am wrong, as we saw. On average, only **one element** ends up in the right place. Did you guess that it would be that bad?