Applied Probability I (STU22004)

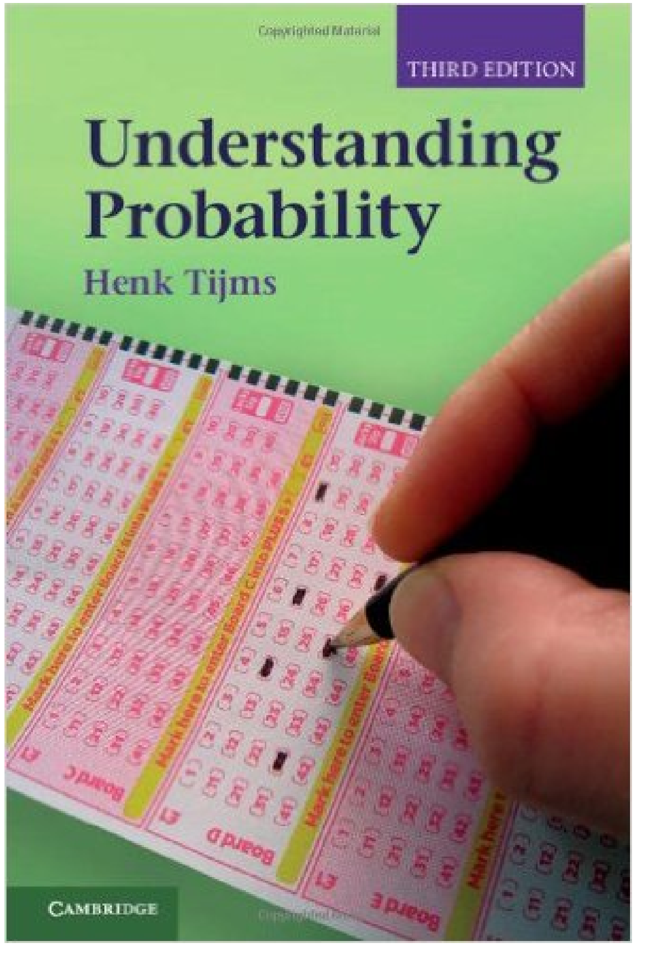
Instructor: James Ng
E-mail: ngja@tcd.ie

Course material

- slides
- lecture recordings
- lab assignments
- tutorial exercises

Textbook

Tijms, “Understanding Probability” (3rd edition, Cambridge 2012).



Assessment

- Exam 2 hours (80%).
- A compulsory group (3-4 people per group) project (20%).
- Reassessment 100% supplemental exam.

What is examinable?

Unless expressly stated otherwise:

- All material presented in class/labs including:
  - material in handouts (e.g. slides)
  - lab assignments
  - tutorial exercises
- Sections of the textbook that will be indicated during the course.

Projects

- Groups of 3 or 4 people
  - Self allocation fine, else random allocation by me.
- Assignment:
  - Investigate some problem by means of random experiments (using computer generated random numbers) and produce a report illustrating the summary results of the experiment.
- More info on assignments and deadlines in the next weeks

Labs

- We will use Excel in labs.
  - Alternative computing platforms would be fine as well, or even better! (for example `R` provides a rich set of statistical tools and is open source).
  - If unfamiliar of syntax/functions templates and help are available.
  - Labs will be useful in view of group project.

Consultation

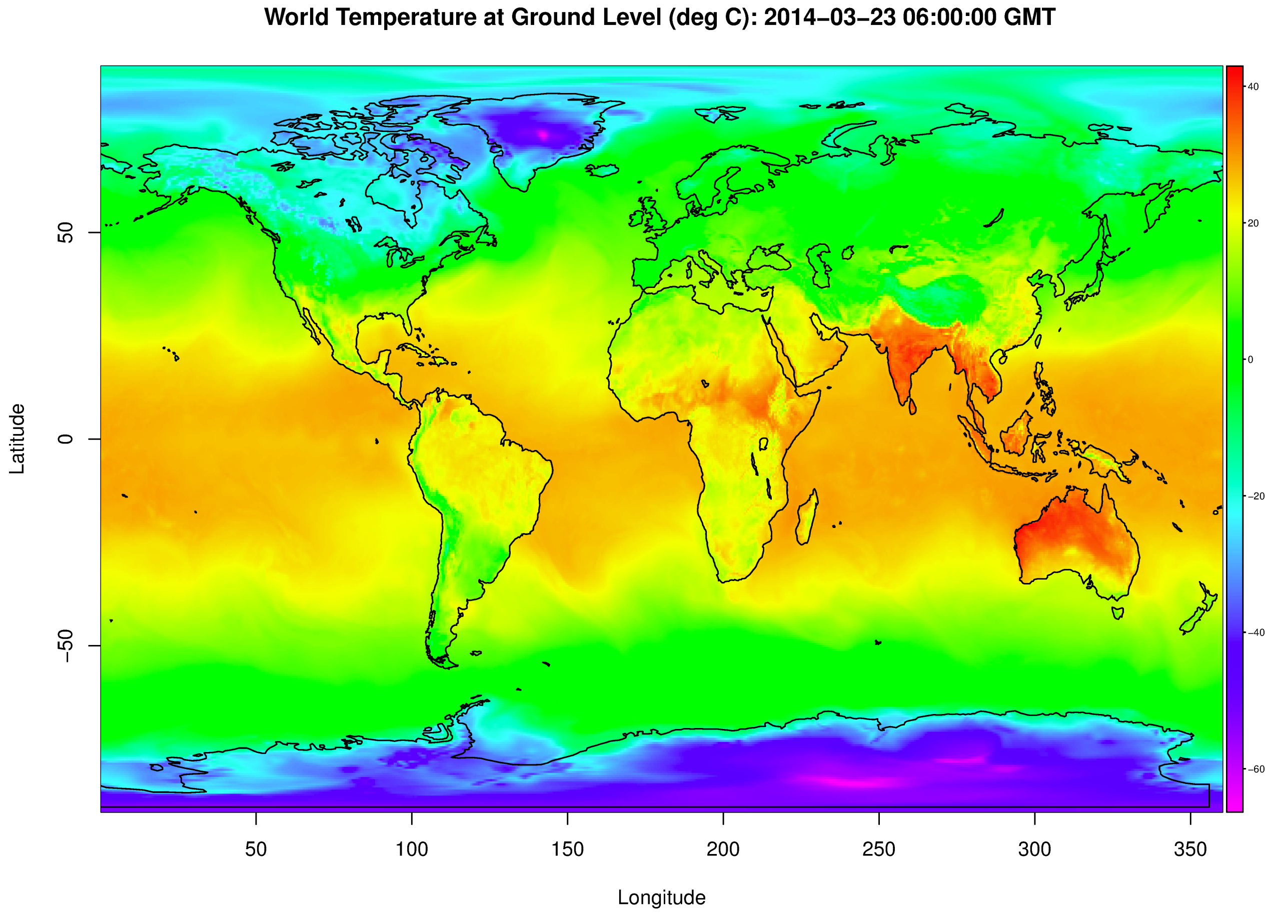
- Thursday 5-7pm every week.
- Send me an email (ngja@tcd.ie) to arrange a consultation.

Spirit of the course

- Basic course in probability with a non-standard structure
- Problems involving uncertain info will be faced in two ways:
  1. By using computers to model uncertainty.
     * Using random numbers to represent uncertain events.
     * Performing many replicates of the same random experiment.
     * Summaries of many replications tell us about how a system with randomness behaves in the long run.
  2. By using probability theory.
     * Precise statements about combining elementary events.
     * Probability distributions for random variables.
     * Summaries of probability distributions tell us about how a system with randomness behaves.

What do we mean by uncertainty?

- Example: weather forecasting. The goal is to predict future weather based on current (and past) conditions.



Example: weather forecasting

- Initial physical conditions are uncertain
  - Lack of information on physical conditions (temperature, position, velocity, etc.) of the system.
- Probabilistic model to deal with uncertainty
  - The evolution of the system is simulated via computer
  - Uncertainty is taken into account by choosing, at random, the initial conditions of the physical system.
- Many replicates of the same random experiment
  - Summary of the outcomes will tell us about the system's evolution.
  - Uncertainty on the outcome of the experiments must be reported.

Sometimes very little uncertainty...

The birthday problem

You go with a friend to a football game. The game involves 22 players of the two teams and one referee. Your friend wagers that, among these 23 people on the field, at least two people will have birthdays on the same day. If this is the case you will have to pay 1 euro to your friend, otherwise he will pay you 1 euro.

Would you accept the bet? Or equivalently who is more likely to win?

- The exact value of the probability of seeing at least two people with the same birthday can be computed exactly by using combinatorics.
- Let's try to compute the same quantity by means of simulation.

A simulation approach

Consider the following random experiment.

- Simplifying assumptions:
  - ignore leap years,
  - assume all days are equally likely (reasonable).

1. Assign at random a birthday to each one of the 23 people on the field. (i.e. assign to each person a number between 1 and 365 at random).
2. Check whether there is at least a repetition among the 23 birthdays.
3. Count this experiment as a *fail* if there is at least one repetition.

- Repeat the experiment several times and compute the relative freq. of fails

$$p = \frac{\#\text{ fails}}{\#\text{ experiments}}$$

Outcome of the experiment

Experiment 1

29, 363, 306, 298, 199, 146, 363, 196, 18, 5, 292, 291, 291, 224, 200, 36, 118, 6, 338, 78, 288, 218, 135

- Two pairs with same birthday, so we count this outcome as a **fail**.

Experiment 2

93, 324, 69, 351, 72, 109, 216, 96, 57, 353, 321, 301, 22, 298, 37, 208, 327, 300, 54, 235, 155, 53, 243

- No coincidences.

Experiment 3

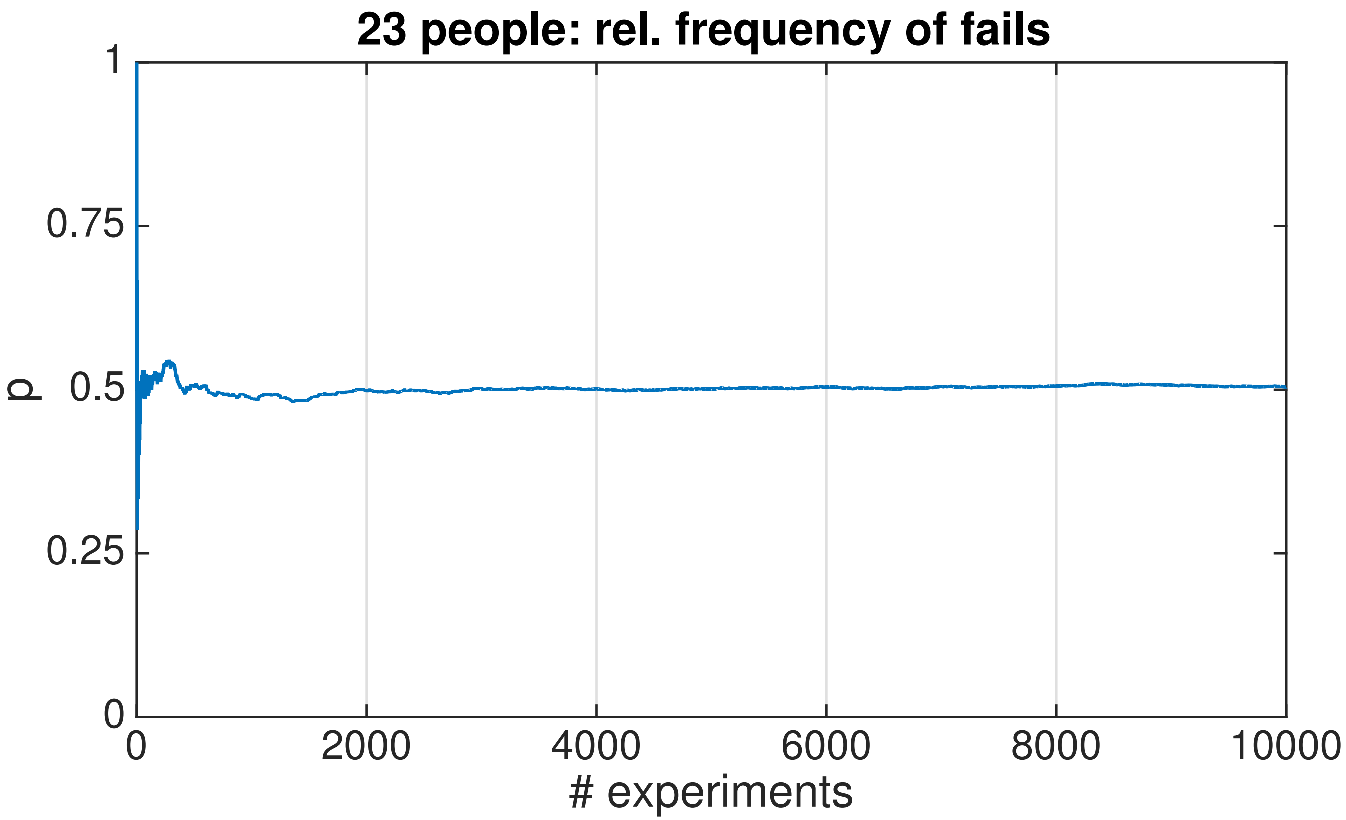
113, 45, 51, 33, 165, 324, 337, 66, 154, 325, 328, 330, 145, 105, 310, 13, 261, 176, 199, 100, 105, 208, 349

- One pair with same birthday, so we count this outcome as a **fail**.

After 3 experiments the relative frequency of fails is $\mathbf{p} = \mathbf{2/3}$.

The long run

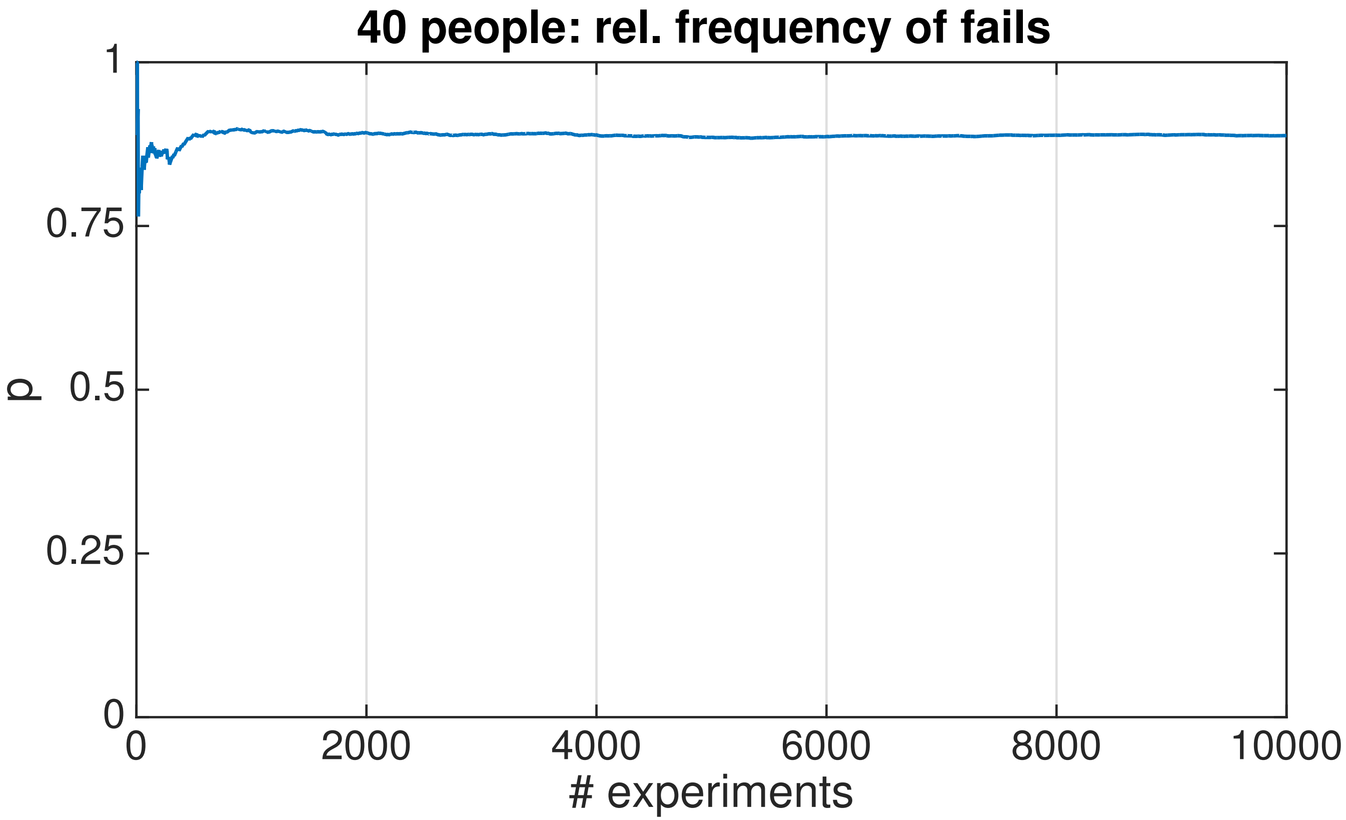
- After running 10000 experiments we get $\mathbf{p} = \mathbf{0.505}$.



- The bet is pretty fair! Counterintuitive? It is also called Birthday paradox.

The long run

- If we consider 40 people we get $\mathbf{p = 0.890}$.



- If we consider 70 people we get $\mathbf{p = 0.999}$.

Questions

- This type of simulation used to solve problems that contain a random element is called *Monte Carlo simulation* (Fermi in 30's, Ulam & Von Neumann in 40's).

It seems a reasonable way to tackle the birthday problem.

But:

- Can we legitimately use p to measure how likely it is for us to lose the bet?
- How many times we shall repeat the experiment?
- How can we assign numbers *at random*?

Empirical law of large numbers

- A random experiment is repeated several times under identical conditions.
- How likely is for an outcome to feature a certain property A? (we say that A is an *event*).
- The relative frequency of the event A in n repetitions is defined as

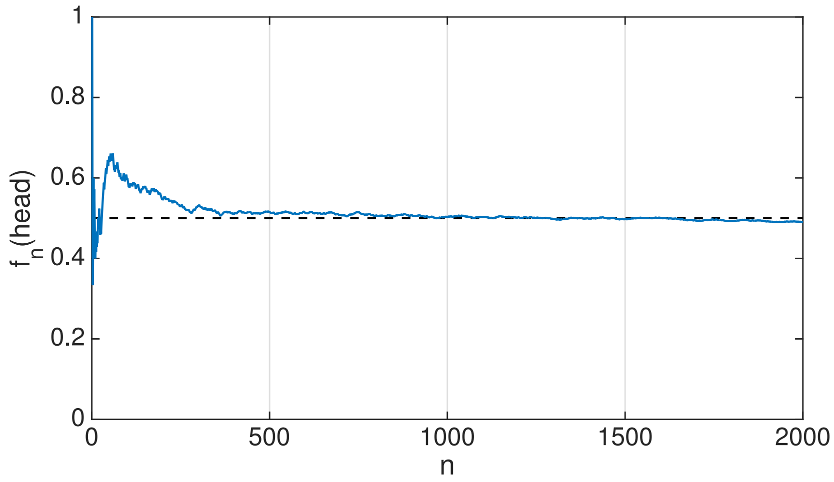
$$f_n(A) = \frac{n(A)}{n}$$

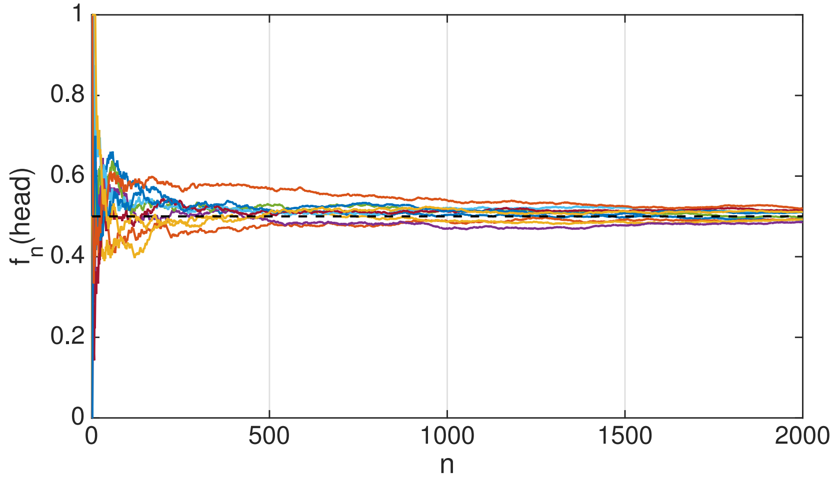
  where $n(A) = \#$ outcomes featuring property A.
- $f_n(A)$ is a number in $[0, 1]$.

The empirical law of large numbers: $f_n(A)$ will fluctuate less and less as the number of repetitions n increases, approaching a limiting value as n goes to infinity.

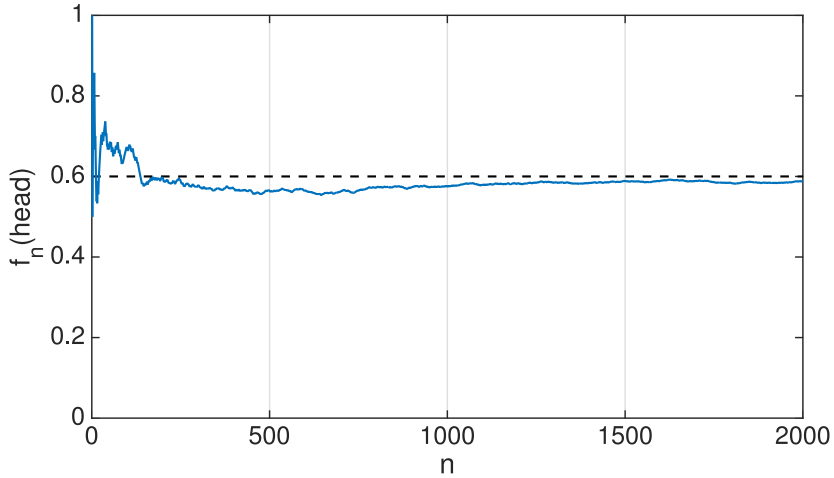
- The *law of large numbers* is the probabilistic formalisation of this same idea.

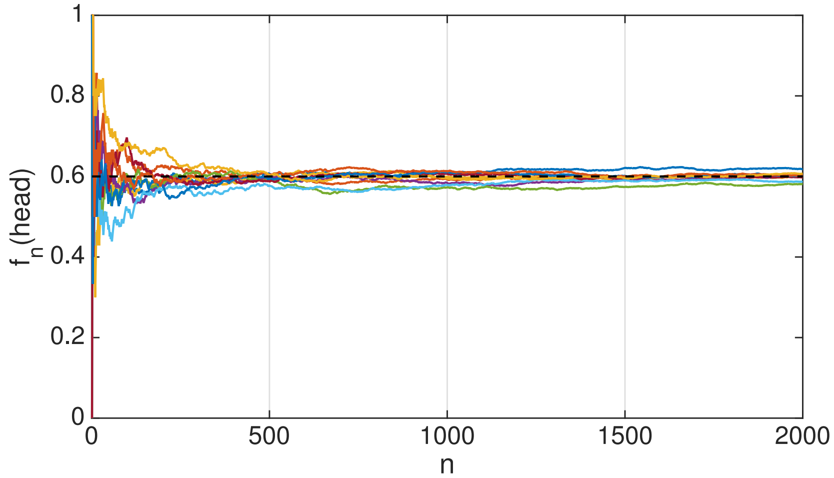
Example: tosses of a coin





- Relative frequency of tossed heads of a fair coin ($\Pr(\text{head}) = \Pr(\text{tail}) = 0.5$)





- Rel. frequency of tossed heads of a non-fair coin ($\Pr(\text{head}) = 1 - \Pr(\text{tail}) = 0.6$)

Random number generation

- What is random?
- Can we reproduce randomness with a computer?
- A very insightful video: *The search for randomness with Persi Diaconis*

  `https://www.youtube.com/watch?v=xit5LDwJVck`

Random number generation

In Monte Carlo simulations random number generation is simply indispensable.

- What is a random number?

One step back:

- Is there true randomness in the world or every event is caused deterministically by prior events?

  (that's a philosophical question –determinism vs indeterminism– we won't investigate)

- We can think of randomness as lack of predictability.

- Thus a random event is an event that cannot be predicted.

True random numbers

- Problem: computers can only implement deterministic procedures that do not involve randomness.
- How can we generate a random number with a computer?
- One way is to exploit unpredictable physical phenomena. Such as:
  - The toss of a coin? Not really... (see Diaconis' video)
  - Physical phenomena that produce "noise" signals, e.g. the atmospheric noise captured by a radio.
- A hardware is used to translate the unpredictable signal into a number.
- Once a sequence of numbers is produced, tests of randomness are performed.
- Check out Random (`https://www.random.org`), a project for true random number generation that started in the 90's at the SCSS in Trinity.

Pseudo random numbers

- Generating *true* random numbers is a safe option (good for example for lotteries).
- Computer simulations typically need more efficient ways to generate random numbers.
- Methods have been developed to generate *deterministically* sequences of numbers that look like randomly generated.
- Statistical tests for randomness are performed to check the goodness of the method.
- Such numbers are called *pseudo random numbers*.
- Today, reliable and efficient methods are available. We will see one of the oldest and simplest methods (*multiplicative congruential generator*).

Random numbers in $(0, 1)$

Basic problem: generating a sequence of random numbers in the interval $(0, 1)$.

- This is a sequence of numbers between 0 and 1 that are picked *at random*.
- By saying *at random*, we mean that any two subintervals of $(0, 1)$ of the same length should have equal probability of containing a generated number.
- For example if we call z the generated number, we want that

$$\Pr(z \in (0, 1/2]) = \Pr(z \in (1/2, 1))$$

or more in general, for any interval $(a, b) \subseteq (0, 1)$,

$$\Pr(z \in (a, b)) = b - a.$$

For example,

$$\Pr(z \in (0.1, 0.3)) = 0.3 - 0.1 = 0.2.$$

We will call such distribution *uniform* on $(0, 1)$.

Pseudo random number generators

- Consider $f : (0,1) \longrightarrow (0,1)$, a **deterministic** procedure used to transform a given number $z_0 \in (0,1)$ in a pseudo-random generated number $z_1 = f(z_0)$.

- The same procedure can be used **iteratively** in order to generate a whole sequence of numbers such that, for any $i \geq 1$

$$z_i = f(z_{i-1}).$$

- Thus,

$$z_1 = f(z_0),$$

$$z_2 = f(z_1),$$

$$\dots$$

$$z_n = f(z_{n-1}).$$

Pseudo random number generators

- If the sequence $\{z_i\}_i$ is statistically undistinguishable from a series of truly random numbers (checked via tests of randomness) then f is a valid pseudo-random number generator.
- If we start our sequence with the same *seed number* z_0 we will obtain the exact same sequence $\{z_i\}_i$.
- This makes a probabilistic simulation reproducible (convenient e.g. when the goal is to compare different experimental designs).

Multiplicative congruential generator (MCG)

- MCG is an instructive example of pseudo-random number generation.
- Consider two positive integers a and m, then we can define the function f as

$$f(z) = az(\text{mod } m).$$

- $f(z)$ coincides with the remainder of az after dividing by m, e.g.

$$17(\text{mod } 5) = 2.$$

Multiplicative congruential generator (MCG)

- We want to generate a sequence of numbers $\{z_1, z_2, \ldots, z_n\} \subset (0, 1)$.
  - Fix a and m
  - Choose an integer seed number x_0 and set

$$x_1 = ax_0 (\text{mod } m)$$

$$x_2 = ax_1 (\text{mod } m)$$

$$\vdots$$

$$x_i = ax_{i-1} (\text{mod } m) \qquad \text{for any } i \geq 1$$

  - By definition the numbers $\{x_1, x_2, \ldots, x_n\}$ are in $\{0, 1, \ldots, m-1\}$
  - If a and m are chosen suitably we can guarantee that $x_i \neq 0$ for every i.
  - Define $z_i = x_i/m$.
  - We obtained $\{z_1, z_2, \ldots, z_n\} \subset (0, 1)$.

MCG: an example

- We set $a = 13$, $m = 31$ and $x_0 = 2$ and get

$$x_1 = x_0 a (\text{mod } m) = 2 * 13 (\text{mod } 31) = 26$$

$$x_2 = x_1 a (\text{mod } m) = 26 * 13 (\text{mod } 31) = 28$$

$$x_3 = \ldots$$

i	seed	**1**	2	3	4	5	6	7	8	9	10
x_i	2	**26**	28	23	20	12	1	13	14	27	10
z_i	0.065	**0.839**	0.903	0.742	0.645	0.387	0.032	0.419	0.452	0.871	0.323

i	11	12	13	14	15	16	17	18	19	20	21
x_i	6	16	22	7	29	5	3	8	11	19	30
z_i	0.194	0.516	0.710	0.226	0.935	0.161	0.097	0.258	0.355	0.613	0.968

i	22	23	24	25	26	27	28	29	30	**31**	32
x_i	18	17	4	21	25	15	9	24	2	**26**	28
z_i	0.581	0.548	0.129	0.677	0.807	0.484	0.290	0.774	0.065	**0.839**	0.903

MCG: an example

- We have the first repetition at the 31st iteration.
- The *period* (number of iterations without repetitions) is equal to $m - 1 = 30$.
- Recall that each number x_i will be in $\{0, 1, \ldots, m - 1\}$ and thus it will take no longer than m steps until some number repeats itself.
- We do not want to have ties among the generated numbers.
- We do not want to generate a 0 otherwise we get stuck in 0. Indeed

$$0 = 0a \text{mod } m.$$

MCG: how to choose a, m?

- Crucial point: how shall integers a and m be chosen?
  1. For any seed in $\{1, 2, \ldots, m-1\}$ the sequence should have the appearance of being generated at random.
  2. For any seed, the period should be large (ideally $m-1$).
  3. The values can be computed efficiently on a computer.
- Number theory provides conditions that guarantee that point 2 is satisfied.
- A combination of values that is known as being convenient is

$$m = 2^{31} - 1 \qquad \text{and} \qquad a = 16807.$$

- First repetition comes after $m-1$ iterations, that is

$$m - 1 = 2147483646 \approx 2.15 * 10^9.$$

MCG: how to choose a, m and x_0?

- The seed x_0 can be fixed (if we want the experiment to be replicable).
- Otherwise the seed is typically chosen based on the computer's clock.
- This was essentially (although not exactly) the random number generator behind the Excel function RAND() up to 2010 release.
- Newer random number generators do not use the MCG scheme and are projected so to have incredibly long periods, to be fast and to have high-quality pseudo-random numbers properties.
- There would be much more to say about pseudo-random number generators. But from now for our simulations we will simply assume that we have a *black box* that gives random numbers from $(0, 1)$ on request.

Pitfalls in random number generation

- In 1999 an online poker room was cracked. The random number generator used to shuffle the deck had some flaws. The team of people that cracked it was able, after a few hands were played, to predict exactly all the cards that would have been dealt to the players at the table.

  `http://pokersafety.blogspot.ie/2005/10/online-poker-room-cracked-not-hoax.html`

- In 1983 Michael Larsen, a contestant on the American television game show Press Your Luck, exploited the poor randomization of the prize wheel used in the game and won 110,237 dollars.

  You can read the story here:

  `http://www.rotten.com/library/conspiracy/press-your-luck`

  And watch the video here:

  `https://www.youtube.com/watch?v=nIZRL4IpB4Y`

Tests for Random Numbers

We want to test whether the generated random numbers are uniformly distributed on $(0, 1)$.

We are going to apply the argument of hypothesis testing:

- we start with the hypothesis that the generated random numbers are uniformly distributed on $(0, 1)$.
- we work under this hypothesis,
- if the observed outcomes are very unlikely, we reject the hypothesis, and conclude that uniformity is not satisfied,
- otherwise, we fail to reject the hypothesis, and conclude that evidence of non-uniformity has not been detected.

Chi-square Test

We apply the **Chi-square test** to test the uniformity of generated random numbers $z_1, \ldots, z_n$:

- divide the unit interval $(0, 1)$ into K equal length subintervals (for concreteness assume $K = 10$),
- the 10 subintervals are $(0, 0.1), (0.1, 0.2), \ldots, (0.9, 1.0)$,
- under the hypothesis of uniformity $\#z_i$ in any subinterval is $E_i = n/10$, $i = 1, \ldots, 10$
- we count the observed $\#z_i$ in each of the subintervals: $O_i, i = 1, \ldots, 10$.

Chi-square Test

We compute the test statistic:

$$\mathcal{X}^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i},$$

which approximately follows a Chi-square distribution with $K-1$ degrees of freedom under the hypothesis of uniformity.

We reject the hypothesis if $\mathcal{X}^2$ is **large**.

Reading

Textbook: Section 2 (pp. 18-25), Section 3 (pp. 75-80, 3.1.4 excluded).

The 1970 draft lottery problem

- During the Vietnam War, the American army used a lottery system based on birth dates to determine who would be called up for service in the army.
- Sampling procedure:
  - Each day of the year (including 29/02) was printed on a slip of paper.
  - Each slip of paper was placed into an individual capsule.
  - Capsules were placed into a large receptacle and mixed.
  - The capsules were drawn one by one out of the receptacle.
  - The first date drawn was assigned a draft number of *one*, the second number *two*, and so on, till all dates had been assigned a number.
- Draftees were called up for service based on the draft number assigned to their dates of birth, those receiving low draft numbers being called up first.
- Doubts about the integrity of the lottery (the way the capsules were mixed) were raised immediately (errors in the randomization procedure).

The 1970 draft lottery problem

- Assume we are unaware of these errors and check the outcome to test the fairness of the procedure.

day	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1	305	086	108	032	330	249	093	111	225	359	019	129
2	159	144	029	271	298	228	350	045	161	125	034	328
3	251	297	267	083	040	301	115	261	049	244	348	157
4	215	210	275	081	276	020	279	145	232	202	266	165
5	101	214	293	269	364	028	188	054	082	024	310	056
6	224	347	139	253	155	110	327	114	006	087	076	010
7	306	091	122	147	035	085	050	168	008	234	051	012
8	199	181	213	312	321	366	013	048	184	283	097	105
9	194	338	317	219	197	335	277	106	263	342	080	043
10	325	216	323	218	065	206	284	021	071	220	282	041
11	329	150	136	014	037	134	248	324	158	237	046	039
12	221	068	300	346	133	272	015	142	242	072	066	314
13	318	152	259	124	295	069	042	307	175	138	126	163
14	238	004	354	231	178	356	331	198	001	294	127	026
15	017	089	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	055	274	120	044	207	254	107	096
17	235	189	033	260	112	073	098	154	255	288	143	304
18	140	292	332	090	278	341	190	141	246	005	146	128
19	058	025	200	336	075	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	063	192	185	135
21	186	363	334	062	250	060	027	291	204	243	156	070
22	337	290	265	316	326	247	153	339	160	117	009	053
23	118	057	256	252	319	109	172	116	119	201	182	162
24	059	236	258	002	031	358	023	036	195	196	230	095
25	052	179	343	351	361	137	067	286	149	176	132	084
26	092	365	170	340	357	022	303	245	018	007	309	173
27	355	205	268	074	296	064	289	352	233	264	047	078
28	077	299	223	262	308	222	088	167	257	094	281	123
29	349	285	362	191	226	353	270	061	151	229	099	016
30	164		217	208	103	209	287	333	315	038	174	003
31	211		030		313		193	011		079		100

Lottery problem: testing the fairness

- The lottery can be thought of as a random permutation of the first 366 integers.
- In the 1970 lottery (see the table) the generated permutation was

$$305, 159, 251, 215, 101, \ldots, 100.$$

- In a fair lottery all possible permutations are equally likely.
- We are going to apply the typical argument of hypothesis testing:
  - we start with the hypothesis that the lottery was fair,
  - we work under this hypothesis,
  - if the outcomes are extremely unlikely, we reject the hypothesis,
  - we conclude that the lottery was most probably unfair.

Lottery problem: testing the fairness

- Let's first focus on the average draft number per month.

January	201.2	July	181.5
February	203.0	August	173.5
March	225.8	September	157.3
April	203.7	October	182.5
May	208.0	November	148.7
June	195.7	December	121.5

- A simple look to the table lets us start doubting about the fairness of the procedure: the last months have a lower average number than the first ones.
- The expected average draft number of a given month is 183.5.
- Why? Intuitively $183.5 = \frac{1+366}{2}$.
- Let's call g_i the average draft number of the ith month (e.g. $g_1 = 201.2$).

- The *sum of the absolute deviations* of the outcomes is given by

$$\sum_{i=1}^{12} | g_i - 183.5 |= 272.4.$$

- Is this number small? Large?
- Under the assumption of a fair lottery (equally likely permutations) we want to compute

$$\Pr\left(\sum_{i=1}^{12} |g_i - 183.5| \geq 272.4\right) \qquad (*)$$

- A small value means that the outcome of 1970 draft lottery was unlikely.
- Computing $(*)$ analytically is extremely complex.
- We can compute $(*)$ via Monte Carlo simulation.

Monte Carlo simulation

- Generate a random permutation of the first 366 integers, say

$$a(1), a(2), \ldots, a(366).$$

- Assign the number $a(i)$ to the ith day of the year, e.g.

$$\begin{aligned} a(1) &\to \text{January, } 1 \\ a(2) &\to \text{January, } 2 \\ &\vdots \\ a(366) &\to \text{December, } 31 \end{aligned}$$

- Compute the monthly averages G_i, for $i = 1, 2, \ldots, 12$, corresponding to the generated permutation.

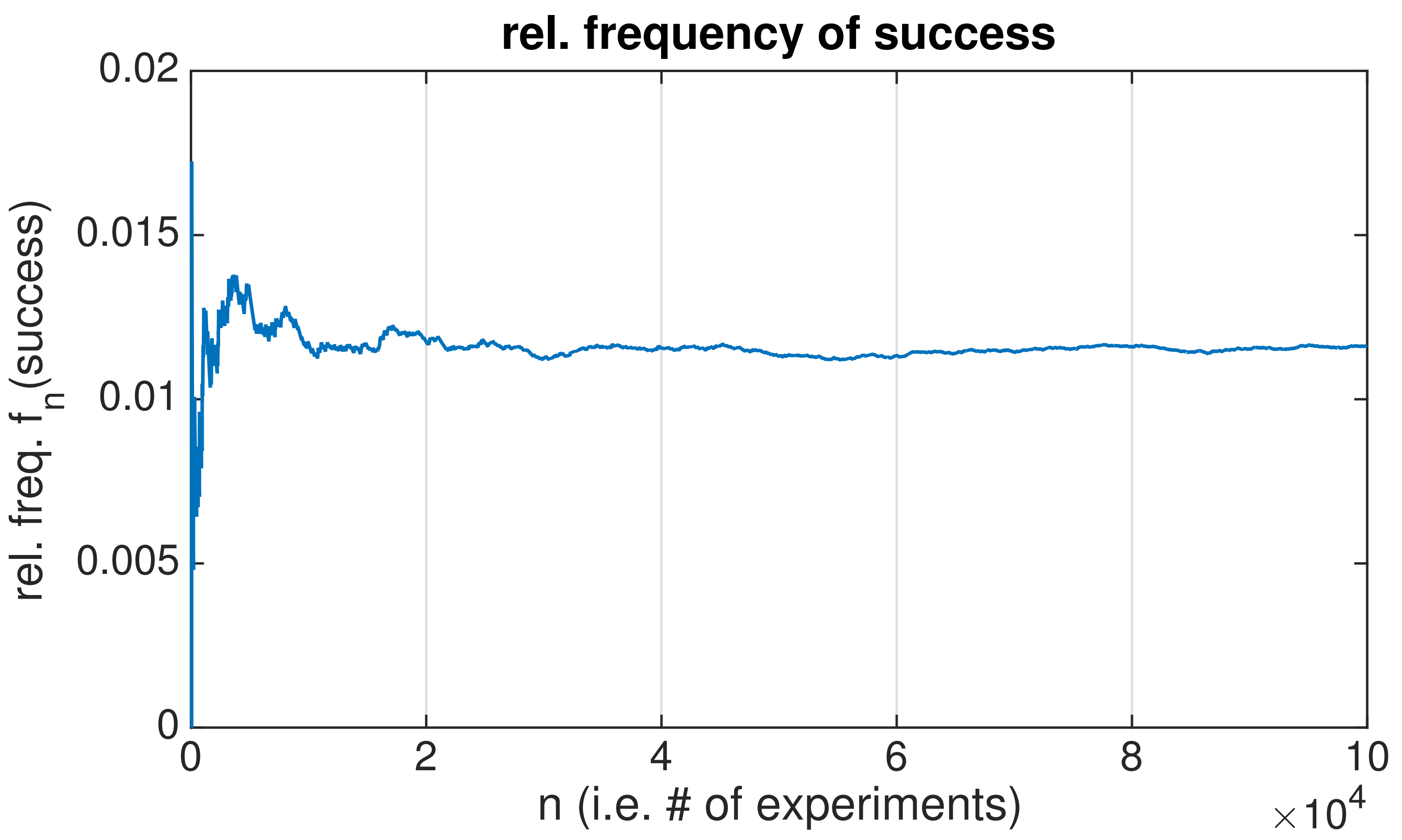
- Count the experiment as a *success* if

$$\sum_{i=1}^{12} |G_i - 183.5| \geq 272.4$$

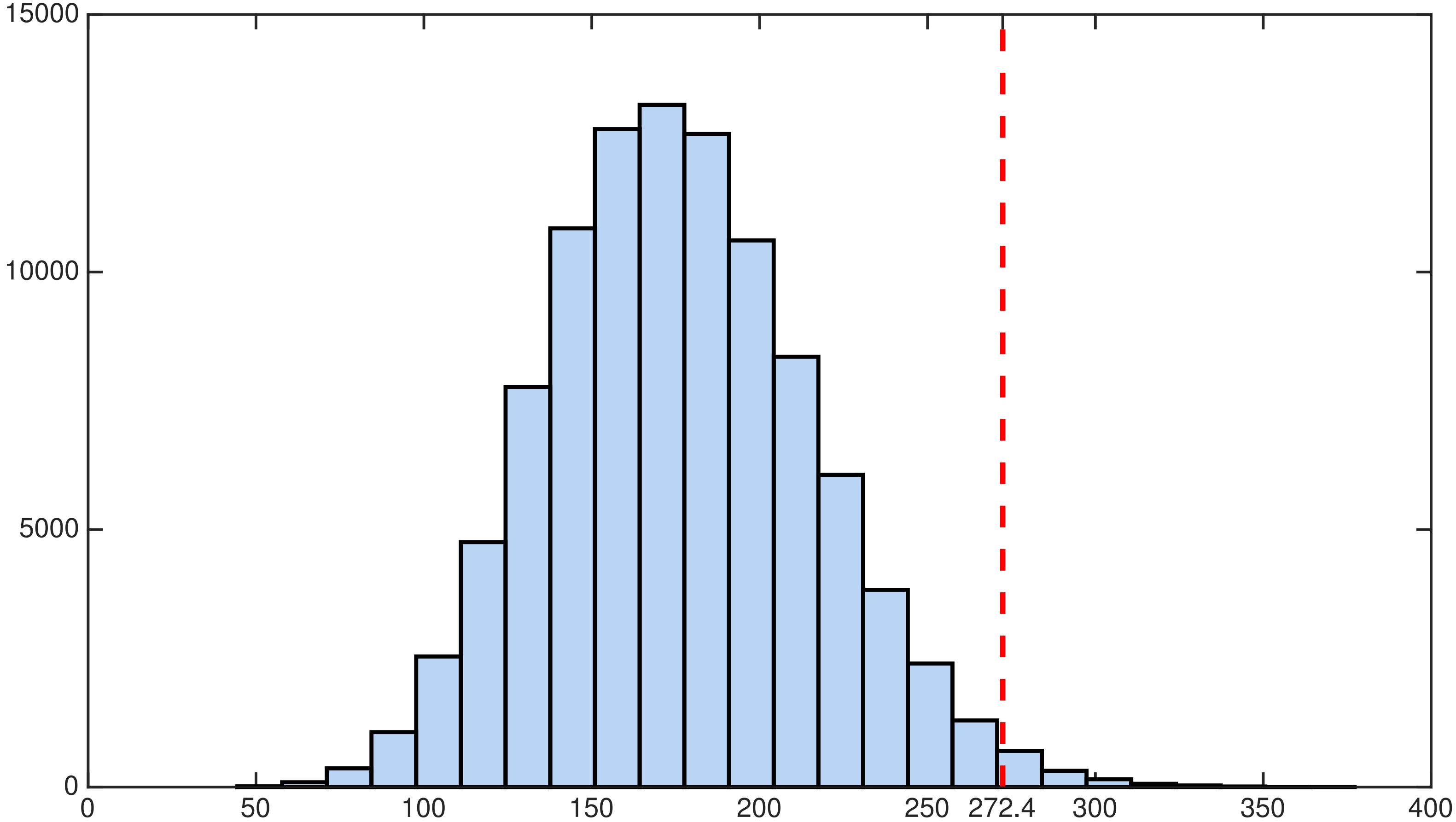
Monte Carlo simulation

- After 100000 replicates we compute the rel. frequency of *success*

$$f_n(\text{success}) = \frac{\#\text{ experiments that lead to a success}}{\#\text{ experiments}} = 0.0116$$



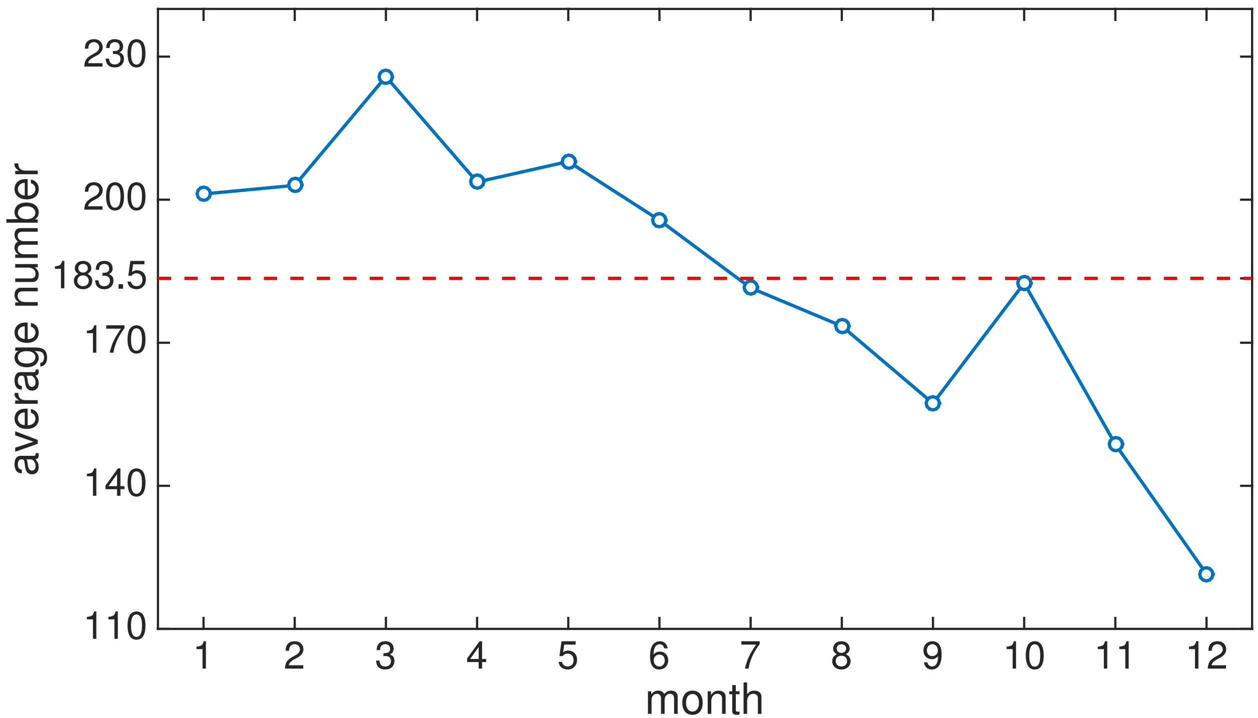
Histogram of $\sum_{i=1}^{12} |G_i - 183.5|$



- We obtain the approximation $\Pr\left(\sum_{i=1}^{12} |g_i - 183.5| \geq 272.4\right) \approx 0.0116$

The lottery problem: another criterion

- We have worked with the sum of the absolute deviations.
- This criterion does not take into account the trend of the monthly average that we have highlighted (last months have a lower monthly average).



The lottery problem: another criterion

- Let's consider another test that takes into account this monthly trend.
- Consider how the months are ranked in terms of average draft number:

month	1	2	3	4	5	6	7	8	9	10	11	12
average number	201.2	203.0	225.8	203.7	208.0	195.7	181.5	173.5	157.3	182.5	148.7	121.5
rank	5	4	1	3	2	6	8	9	10	7	11	12

- The lottery can be thought of as a random permutation of the first 12 integers (the 12 months).
- In the 1970 lottery (see the table) the generated permutation a^* was

$$5, 4, 1, 3, 2, 6, 8, 9, 10, 7, 11, 12$$

- In a fair lottery all possible permutations are equally likely.
- Let's test how likely it is to get the permutation a^* assuming a fair lottery.

The lottery problem: another criterion

- Consider a random perm. $a = (a(1), a(2), \ldots, a(12))$ of the first 12 integers.
- Define the distance
$$d(a) = \sum_{1}^{12} |a(i) - i|.$$
- It is possible to verify that $0 \leq d(a) \leq 72$.
- For the permutation a^* we have $d(a^*) = 18$.
- Is this number small? Large?
- Under the assumption of a fair lottery (equally likely permutations) we want to compute
$$\Pr\left(\sum_{i=1}^{12} |a^*(i) - i| \leq 18\right) \qquad (***)$$
- A small value means that the outcome of 1970 draft lottery was unlikely.
- We can compute $(***)$ via Monte Carlo simulation.

Monte Carlo simulation

- Generate a random permutation of the first 12 integers, say

$$a(1), a(2), \ldots, a(12).$$

- Assign the number $a(i)$ to the ith month of the year, e.g.

$$\begin{aligned} &a(1) \to \text{January} \\ &a(2) \to \text{February} \\ &\qquad \vdots \\ &a(12) \to \text{December} \end{aligned}$$

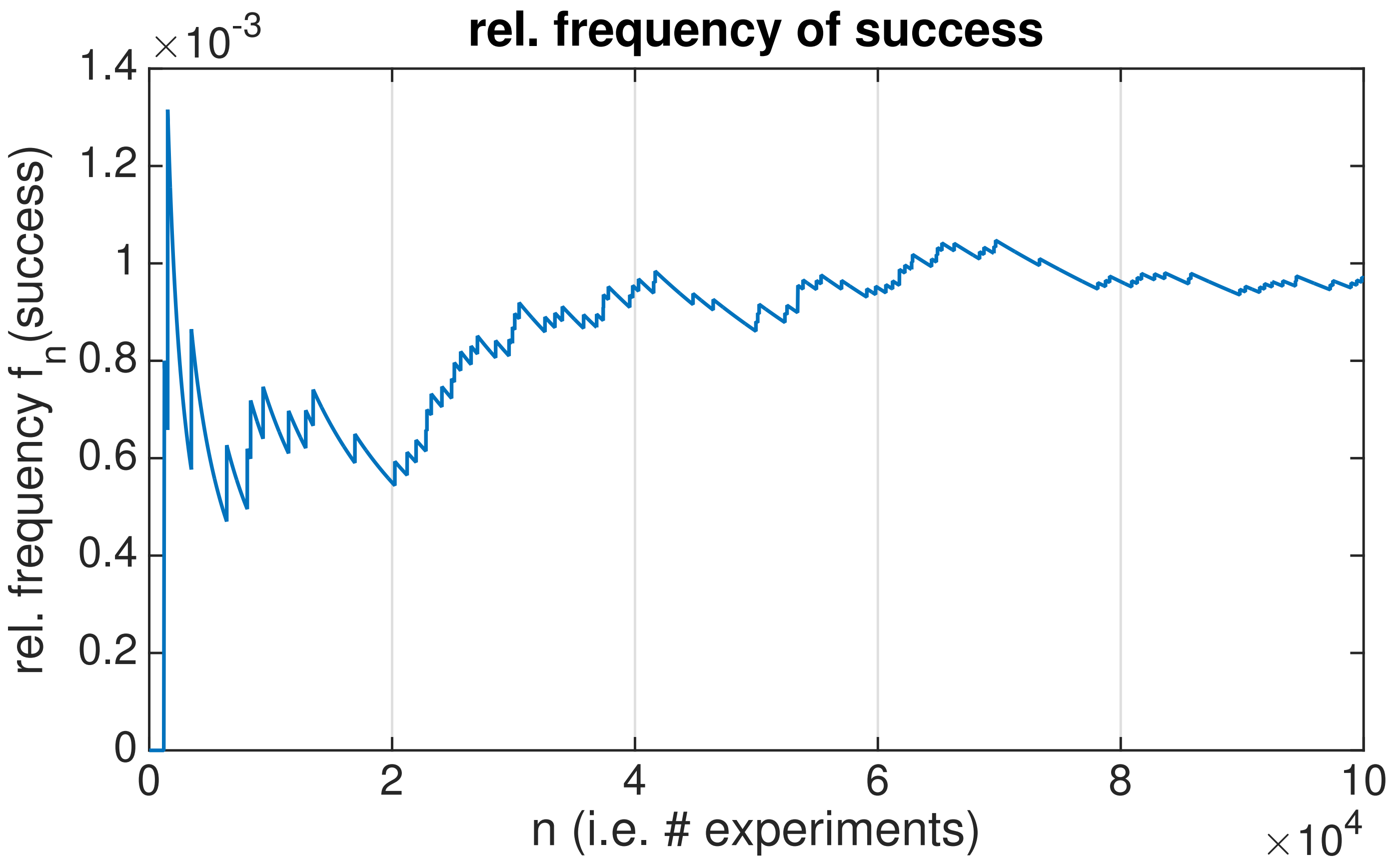
- Compute the distance $d(a) = \sum_1^{12} |a(i) - i|$.
- Count the experiment as a *success* if

$$d(a) \leq 18$$

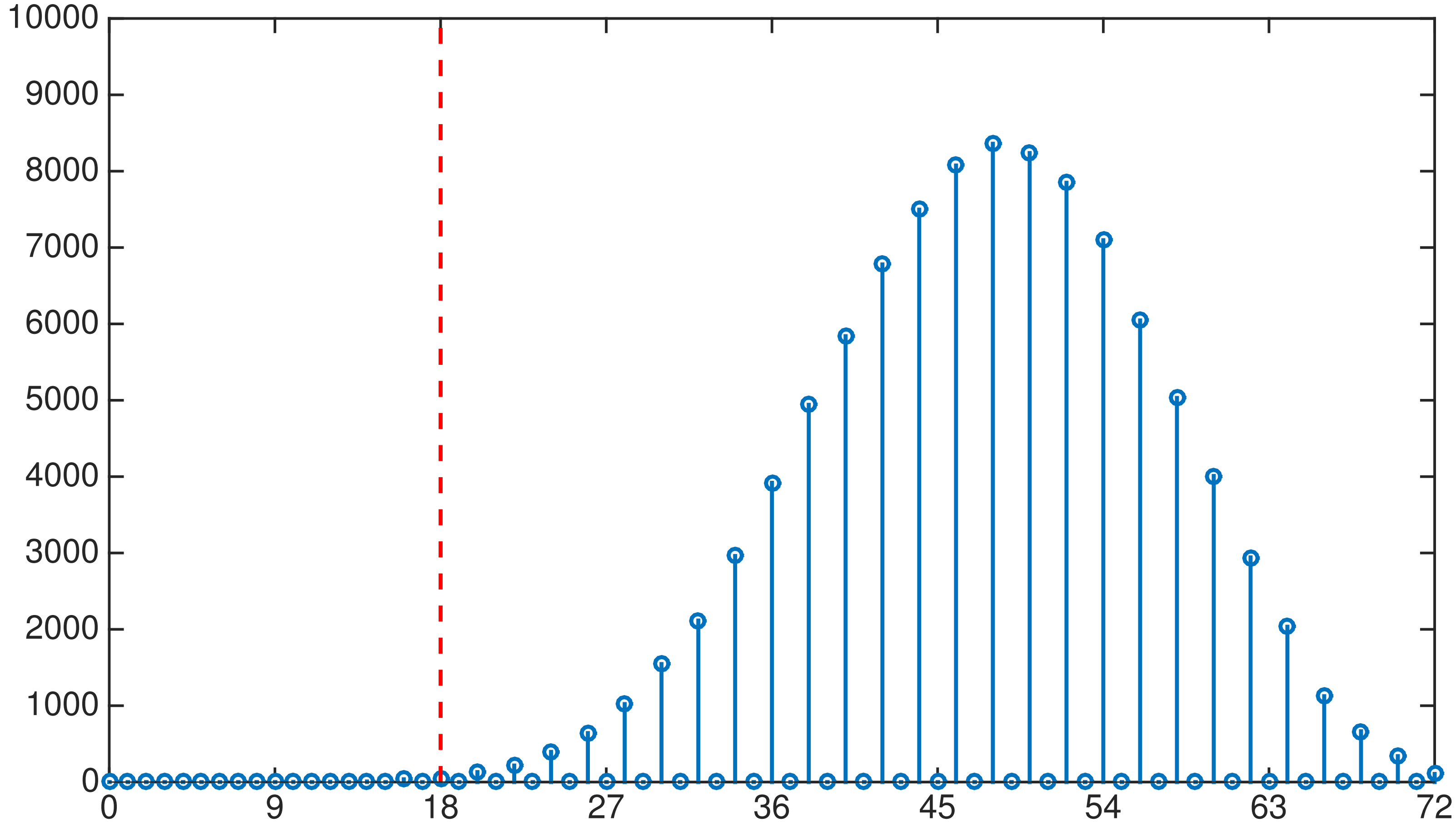
Monte Carlo simulation

- After 100000 replicates we compute the rel. frequency of *success*

$$f_n(\text{success}) = \frac{\#\text{ experiments that lead to a success}}{\#\text{ experiments}} = 0.0009$$



Frequencies of $\sum_{i=1}^{12} |a(i) - i|$



- We obtain the approximation $\Pr\left(\sum_{i=1}^{12} |a(i) - i| \le 18\right) \approx 0.0009$

Some comments

- Marginal comment: $\Pr(d(a) \text{ is an odd number}) = 0$
- This second test provides stronger evidence that the 1970 draft lottery was not fair.
- We have seen two tests based on random permutation sampling (that is sampling without replacement). These methods go under the name of *bootstrap*.
- When performing a statistical test, it is important to report the criterion that was used and, if the hypothesis is rejected, the degree of confidence.

Reading

Textbook: Section 2.8.1 (pp. 55-56), Section 3.7 (pp. 98-101).

Probability

A running example:

A fair die is rolled once. The possible outcomes are:

$$\{1, 2, 3, 4, 5, 6\}$$

Preliminary definitions

An experiment is an act or process of observation that leads to a single outcome that cannot be predicted with certainty.

- Example: The experiment is rolling the die.

A sample point is an outcome of an experiment

- Example: There are 6 sample points in this experiment:

$$1, 2, 3, 4, 5, 6.$$

The sample space of an experiment is the set of ALL its sample points.

- Example: The sample space of this experiment is:

$$\{1, 2, 3, 4, 5, 6\}.$$

Probability axioms

Let S be the sample space, and let $\mathbb{P}$ be a probability on the sample space. For each sample point s_i in S, we let p_i denote the probability of sample point i, i.e. $p_i := \mathbb{P}(\{s_i\})$.

$\mathbb{P}$ must satisfy the following properties:

- $0 \leq p_i \leq 1$.
- $\mathbb{P}(S) = 1$.
- The probabilities of **all** sample points within a sample space must sum to 1.

$$\sum_i p_i = 1$$

Events

An event is a collection of sample points. It is a subset of the sample space S.

- An event $\mathcal{A}$ occurs if any one of the sample points in $\mathcal{A}$ occur.
- Let $\mathcal{A}$ be the event that an even number comes up when rolling the die.
- $\mathcal{A}$ occurs if 2, 4, or 6 comes up.

The probability of an event $\mathcal{A}$ is calculated by summing the probabilities of the sample points in $\mathcal{A}$.

- Let $\mathcal{A}$ be the event that 2 or 3 comes up when rolling the die.
- Suppose the die is fair, what is the probability of $\mathcal{A}$?

Loaded die

- A die has been *loaded* so that the probability of side i coming up is proportional to i.
- If $\mathcal{A}$ is the event that either a 2 or a 3 comes up.
- What is the probability of event $\mathcal{A}$?

Equally likely events

If the sample space S has a finite number of sample points, and all are equally likely, then the probability of each event $\mathcal{A}$ is given by:

$$\mathbb{P}(\mathcal{A}) = \frac{|\mathcal{A}|}{|S|}$$

where $|\cdot|$ denotes the cardinality (i.e. number of elements) of a set.

Two dice are rolled independently. The sample space is:

$$\{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}.$$

What is the probability that the sum of two dice is equal to 7?

Reading

Textbook: Section 7.1 (pp. 229-232).

How many events?

- Consider a sample space $S = \{s_1, \ldots, s_N\}$.
- Recall that an event is a collection of sample points.
- In other words, an event is any subset of S.
- How many events exist?

Example: toss of two coins

The sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Define $\mathcal{A} :=$ Tossing of at least one H.

- Is $\mathcal{A}$ an event?
- Yes, $\mathcal{A} = \{(H, H), (H, T), (T, H)\} \subset S$.

Define $\mathcal{B} :=$ Tossing of both H.

- Is $\mathcal{B}$ an event?
- Yes, $\mathcal{B} = \{(H, H)\} \subset S$.

Example: toss of two coins

We can list all events:

$$\begin{aligned}
\{\emptyset, \{(H,H)\}, \{(H,T)\}, \{(T,H)\}, \{(T,T)\}, \\
\{(H,H),(H,T)\}, \{(H,H),(T,H)\}, \{(H,H),(T,T)\}, \\
\{(H,T),(T,H)\}, \{(H,H),(T,T)\}, \{(T,H),(T,T)\}, \\
\{(H,H),(H,T),(T,H)\}, \{(H,H),(H,T),(T,T)\}, \\
\{(H,H),(T,H),(T,T)\}, \{(H,T),(T,H),(T,T)\}, \\
\{(H,H),(H,T),(T,H),(T,T)\}\}
\end{aligned}$$

- This is called the power set of S, denoted $\mathcal{P}(S)$.
- $|\mathcal{P}(S)| = 16$.
- In general, $|S| = m$ implies $|\mathcal{P}(S)| = 2^m$.

Union and intersection

- The union of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs if either $\mathcal{A}$ or $\mathcal{B}$ (or both) occur. Denoted $\mathcal{A} \cup \mathcal{B}$.
- The intersection of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs if both $\mathcal{A}$ and $\mathcal{B}$ occur. Denoted $\mathcal{A} \cap \mathcal{B}$.

Dice example Consider an experiment of rolling two dice. Let $\mathcal{A}$ be the event that the sum of two dice is greater than 10 and $\mathcal{B}$ be the event that the sum of two dice is even.

- What is $\mathcal{A} \cup \mathcal{B}$?
- What is $\mathcal{A} \cap \mathcal{B}$?

Dice example

S:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Additive law

- The probability of the union of two events $\mathcal{A}$ and $\mathcal{B}$ is:

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

- Two events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if they cannot occur at the same time.
- We can write $\mathcal{A} \cap \mathcal{B} = \emptyset$ when the events are mutually exclusive.
- If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events, $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = 0$ and so

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B})$$

Complementary events

The **complement** of an event $\mathcal{A}$ is the event that $\mathcal{A}$ does not occur. Denoted $\mathcal{A}^c$.

- All sample points in S are either in $\mathcal{A}$ or $\mathcal{A}^c$, no sample point can be in both.
- Thus,

$$\begin{aligned} & \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{A}^c) = 1 \\ \Rightarrow \quad & \mathbb{P}(\mathcal{A}) = 1 - \mathbb{P}(\mathcal{A}^c) \end{aligned}$$

- This is a useful formula for computation.

Example: Dice

Two fair dice are rolled. Event $\mathcal{A}$ is that we observe a 5. Event $\mathcal{B}$ is that the dice sum to 7. Calculate:

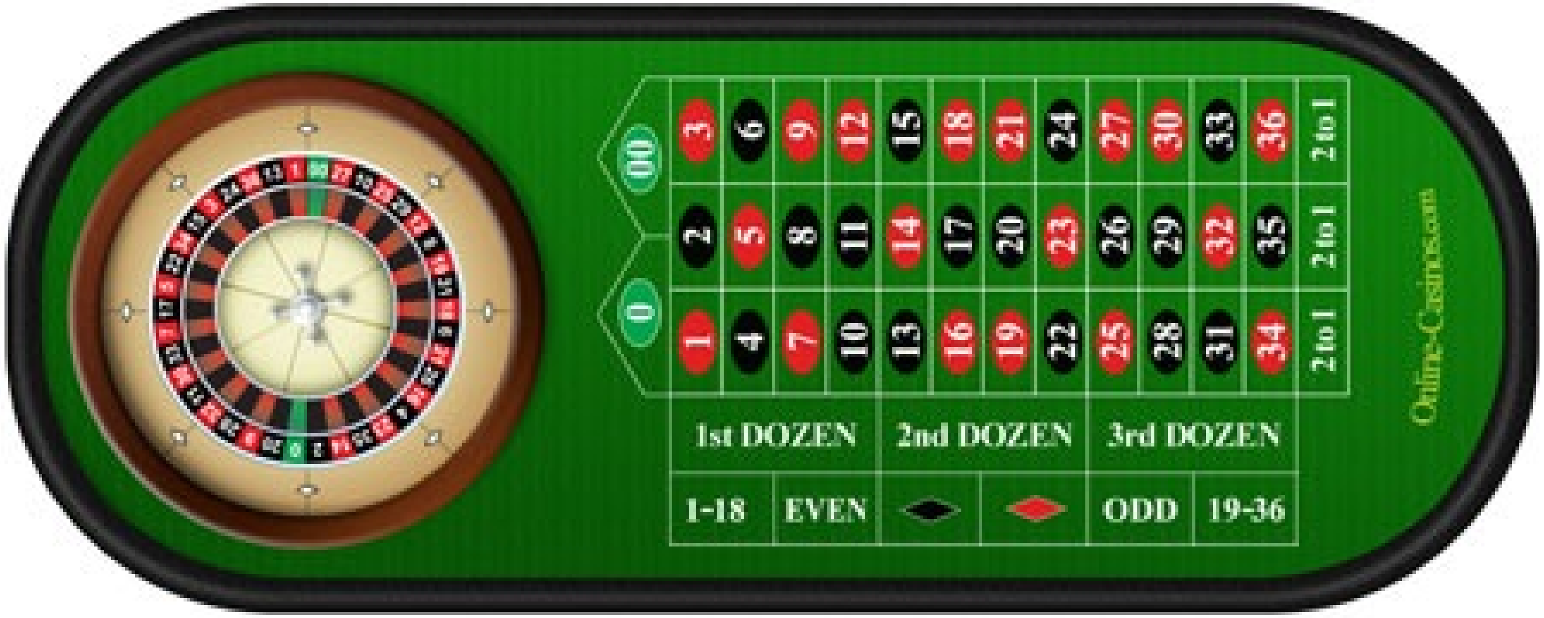
- $\mathbb{P}(\mathcal{A} \cap \mathcal{B})$ and $\mathbb{P}(\mathcal{A} \cup \mathcal{B})$.
- $\mathbb{P}(\mathcal{A}^c)$.

S:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Example: Dice

Exercise: Roulette



- 38 slots, 18 red, 18 black, 2 green, 18 even, 18 odd.

Exercise: Roulette

- A – outcome is an odd number (0 and 00 are neither odd nor even).
- B – outcome is a red number.
- C – outcome is in the first dozen (1-12).

1. Define the events $A \cap B$ and $A \cup B$ as a specific sets of sample points.
2. Find $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$, $\mathbb{P}(A \cup B)$ and $\mathbb{P}(C)$ by summing the probabilities of the appropriate sample points.
3. Find $\mathbb{P}(A \cup B)$ using the additive rule. Are events A and B mutually exclusive?
4. Find $\mathbb{P}(A \cap B \cap C)$.

Sample space

- The sample space S can be **finite**, **countably infinite**, or **uncountably infinite**.
- Finite: $S = \{s_1, \ldots, s_N\}$.
- Countably infinite: $S = \{s_1, s_2, \ldots\}$.

An example of countably infinite sample space

- Consider an experiment of repeatedly tossing a coin until the first head shows up.
- The sample space is $S = \{(H), (T, H), (T, T, H), \ldots\}$.
- If the coin is far, $\mathbb{P}(\{(H)\}) = \frac{1}{2}$, $\mathbb{P}(\{(T, H)\}) = \frac{1}{4}, \mathbb{P}(\{(T, T, H)\}) = \frac{1}{8}, \cdots$.

Example of countably infinite sample space

For each event $\mathcal{A} \subset S$,

$$\mathbb{P}(\mathcal{A}) = \sum_{i:s_i \in \mathcal{A}} \mathbb{P}(\{s_i\}).$$

Let $\mathcal{A}$ be the event that it takes at most three tosses for the first head to show up. Calculate $\mathbb{P}(\mathcal{A})$.

- $\mathcal{A} = \{(H), (T, H), (T, T, H)\}$.
- $\mathbb{P}(\mathcal{A}) = \mathbb{P}(\{(H)\}) + \mathbb{P}(\{(T, H)\}) + \mathbb{P}(\{(T, T, H)\}) = 1/2 + 1/4 + 1/8 = 7/8$.

Let $\mathcal{B}$ be the event that it takes at least four tosses for the first head to show up. Calculate $\mathbb{P}(\mathcal{B})$.

Uncountably infinite sample space

Examples of uncountably infinite sample space:

- The waiting time until the arrival of the next bus. $S = (0, \infty)$.
- Picking a number randomly between 0 and 1. $S = (0, 1)$.

For uncountably infinite S,

- it is not possible to define $\mathbb{P}$ on the power set of S: $\mathcal{P}(S)$.
- but it is not a big deal!
- Informally, an event is a subset of S, to which probabilities will be assigned.

Example: dartboard

- You randomly throw a dart at a circular dartboard with radius R.
- Assume that the dart lands on a completely random point on the dartboard.
- What is the probability of the dart hitting the bull's-eye having radius b?

Solution:

- The sample space is $S = \{(x, y) : x^2 + y^2 \leq R^2\}$ where (x, y) denotes the point at which the dart hits the dartboard.
- Uncountably infinite sample space.
- Observation: the probability that the dart lands exactly on a prespecified point is zero. Why?

Example: dartboard

Solution (cont):

- The area of S is πR^2.
- Let $\mathcal{A}$ be any region in S
- By assumption of complete randomness, the probability of the dart hitting inside the region $\mathcal{A}$ is

$$\mathbb{P}(\mathcal{A}) = \frac{\text{area of the region}\mathcal{A}}{\pi R^2}.$$

- Consider $\mathcal{A} = \{(x, y) : x^2 + y^2 \leq b^2\}$.

$$\mathbb{P}(\mathcal{A}) = \frac{\pi b^2}{\pi R^2} = b^2/R^2.$$

Reading

Textbook: Section 7.1.1.

Slide 4
Conditional probability

- Sometimes we are aware of extra information which might affect the outcome of an experiment. This extra information may then alter the probability of a particular event of interest.
- Suppose we are interested in evaluating the probability that event $\mathcal{B}$ happens given that we know that event $\mathcal{A}$ has happened.
- We write $\mathbb{P}(\mathcal{B}|\mathcal{A})$ for this.
- It is called the *conditional probability of* $\mathcal{B}$ *given* $\mathcal{A}$.

Conditional probability and Bayes theorem

- The conditional probability is defined as:

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})},$$

  if $\mathbb{P}(\mathcal{A}) \neq 0$.

- The following multiplication rules can be derived:

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{B}|\mathcal{A})\mathbb{P}(\mathcal{A}).$$

- This leads to the Bayes theorem:

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})}{\mathbb{P}(\mathcal{A})},$$

  provided $\mathbb{P}(\mathcal{A}) \neq 0$.

Very useful identity

- Let's consider the probability of event $\mathcal{A}$.

$$\begin{aligned}\mathbb{P}(\mathcal{A}) &= \mathbb{P}\{(\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}^c)\} \\ &= \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}^c) \\ &= \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{A}|\mathcal{B}^c)\mathbb{P}(\mathcal{B}^c)\end{aligned}$$

- Hence, a very useful form of Bayes Theorem can be written as

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})}{\mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{A}|\mathcal{B}^c)\mathbb{P}(\mathcal{B}^c)}.$$

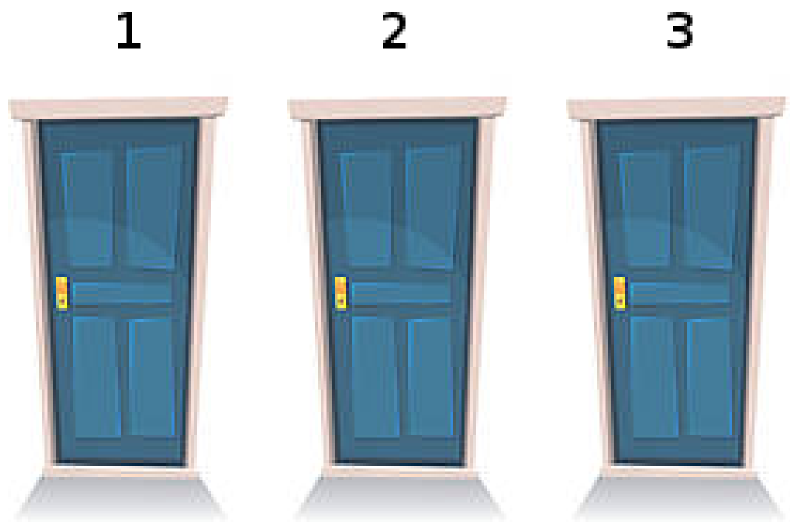
Example: Diagnostic Test

Suppose there is a rare disease which affects 1 person in every 1000 of the population. Fortunately a diagnostic medical test exists for the disease. It is a good test in that, if you have the disease, the test will be positive 95% of the time and if you do not have the disease it will be negative 99% of the time. If a patient tests positive for the disease, what is the probability that they actually have the disease?

Example: Children

Suppose a family has two children and suppose one of the children is a boy. What is the probability that both children are boys?

Monty Hall game



In a tv show you're given the choice of three doors: behind one door is a car; behind the others, goats. Let's say you pick door No. 1. The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you: "Do you want to pick door No. 2?"

Is it to your advantage to switch your choice?

Monty Hall game

- Based on the American television game show *Let's Make a Deal*
- One car behind one of three doors, the other two have a goat behind them.
- Player selects one, say Door No. 1.
- Before opening this door, the host (who knows what is behind each door), opens one of the other two doors, say door No. 3, and shows a goat.
- Host now offers to change selection.
- Issue: Is there any point in changing?
- Vote:

  A. the car is equally likely to be behind door No. 1 and door No. 2

  B. the car is more likely to be behind door No. 2

Very popular problem

- It got popular as a question from a reader's letter in a magazine in 1990 (see pp. 213-215 in the textbook)
- Explained by Kevin Spacey (actually one of his students) in the movie 21 `http://www.youtube.com/watch?v=Zr_xWfThjJ0`
- A (pretty boring!) game online `http://math.ucsd.edu/~crypto/Monty/monty.html`
- Known also under other variants: e.g. the problem of the three prisoners (see p. 216 in the textbook)
- Paul Erdos, one of the most famous and active mathematicians in history, got convinced about the solution only after seeing a Monte Carlo simulation.

Monty Hall game

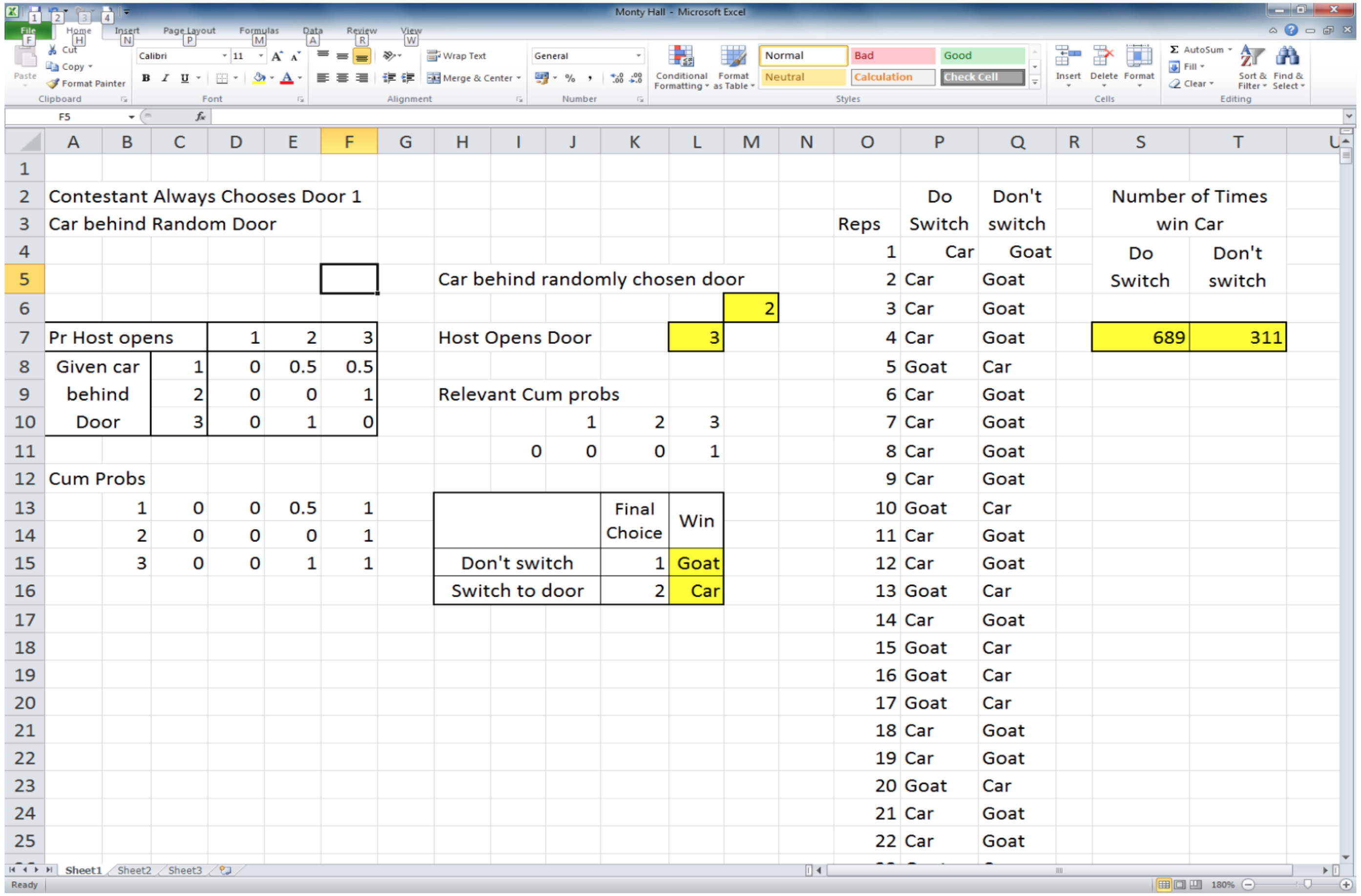
- Let's say that you choose door 1
- What is random?
  - which door has the car
  - which door the host opens

		Pr Host opens door		
		1	**2**	**3**
	1	0	0.5	0.5
given car behind door	**2**	0	0	1
	3	0	1	0

where we implicitly introduced the notion of conditional probability

- Which door to choose is a policy, so not random (switch *vs* don't switch)
- We first run a *simulation*, and then use *probability arguments*

Monty Hall game



	A	B	C	D	E	F	H	I	J	K	L	M	O	P	Q	S	T
1																	
2	Contestant Always Chooses Door 1													Do	Don't	Number of Times	
3	Car behind Random Door												Reps	Switch	switch	win Car	
4													1	Car	Goat	Do	Don't
5							Car behind randomly chosen door						2	Car	Goat	Switch	switch
6												2	3	Car	Goat		
7	Pr Host opens			1	2	3	Host Opens Door				3		4	Car	Goat	689	311
8	Given car		1	0	0.5	0.5							5	Goat	Car		
9	behind		2	0	0	1	Relevant Cum probs						6	Car	Goat		
10	Door		3	0	1	0			1	2	3		7	Car	Goat		
11								0	0	0	1		8	Car	Goat		
12	Cum Probs												9	Car	Goat		
13		1	0	0	0.5	1				Final Choice	Win		10	Goat	Car		
14		2	0	0	0	1							11	Car	Goat		
15		3	0	0	1	1	Don't switch			1	Goat		12	Car	Goat		
16							Switch to door			2	Car		13	Goat	Car		
17													14	Car	Goat		
18													15	Goat	Car		
19													16	Goat	Car		
20													17	Goat	Car		
21													18	Car	Goat		
22													19	Car	Goat		
23													20	Goat	Car		
24													21	Car	Goat		
25													22	Car	Goat		

Check the syntax by downloading *Monty_Hall.xlsx* from the course webpage.

Monty Hall game (Bayes)

Independence

Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if the occurence of $\mathcal{B}$ does not alter the probability that $\mathcal{A}$ has occurred. i.e. $\mathcal{A}$ and $\mathcal{B}$ are independent if:

$$\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$$

Events which are not independent are said to be dependent.

- Combining the definition above with the multiplicative rule, it can be seen that if $\mathcal{A}$ and $\mathcal{B}$ are independent then:

  $$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$$

  The converse is also true, i.e. if $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$ then the events $\mathcal{A}$ and $\mathcal{B}$ are independent.

- Suppose $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events. If $\mathcal{B}$ occurs then $\mathcal{A}$ cannot occur simultaneously so $\mathbb{P}(\mathcal{A}|\mathcal{B}) = 0$.

$\Rightarrow$ Mutually exclusive events are dependent events.

Example: Corrosion

The independence of corrosion and the functional status of a machine component are to be investigated. Are they independent?

	Functioning	Malfunctioning
Corroded	0.2	0.4
Not corroded	0.3	0.1

Reading

Textbook: Section 6.1, Section 8.1 (page 256-258), Section 8.1.2, Section 8.2 (page 266-267).

Slide 5
Law of total probability

Suppose the sample size $\mathcal{S}$ can be divided into *mutually exclusive and exhaustive* events $\mathcal{A}_1, \ldots, \mathcal{A}_n$ - that is, every sample point falls into exactly one of those events.

Then for any event $\mathcal{B}$,

$$\mathbb{P}(\mathcal{B}) = \sum_{i=1}^{n} \mathbb{P}(\mathcal{B}|\mathcal{A}_i)\mathbb{P}(\mathcal{A}_i).$$

Example: Urn

An urn contains 5 red balls and 2 green balls. Two balls are drawn one after the other. What is the probability that the second ball is red? Sample space is $\mathcal{S} = \{rr, rg, gr, gg\}$.

Let R_1 be the event that the first ball is red, G_1 first ball is green, R_2 second ball is red, G_2 second ball is green. We want $\mathbb{P}(R_2)$. We have $\mathbb{P}(R_2|R_1) = 4/6$, $\mathbb{P}(R_2|G_1) = 5/6$. Therefore,

$$\mathbb{P}(R_2) = \mathbb{P}(R_2|R_1)\mathbb{P}(R_1) + \mathbb{P}(R_2|G_1)\mathbb{P}(G_1) = 4/6 \cdot 5/7 + 5/6 \cdot 2/7 = 5/7.$$

Random variables

Examples of random variables

- A business manager is interested in whether sales will reach \$5 million next year.
- An economist may be interested in if GDP growth rate will reach 6% next year.
- In a much simpler case, one may be interested in whether the sum of two dice rolls will exceed 10.

Random variables - dice

We will motivate the definition of random variables using a simpler example.

- Suppose we roll two dice independently, and let Y denote the sum of two dice.
- Because the value of Y will vary depending on the outcome of the experiment, it is called a random variable.
- In particular, for each outcome (sample point) in the sample space $\mathcal{S} = \{(1,1), \ldots, (1,6), \ldots, (6,6)\}$, a real number is assigned.
- For example, if the outcome $(2,3)$ is observed, then $Y = 5$. We write $Y((2,3)) = 5$.
- If the outcome $(3,5)$ is observed, then $Y((3,5)) = 8$.

Random variables - dice

- The value assigned to Y will vary from one outcome to another, but some outcomes may be assigned the same value.

- For example,

$$Y((2,3)) = Y((3,2)) = Y((1,4)) = Y((4,1)) = 5$$

$$Y((5,6)) = Y((6,5)) = 11,$$

$$Y((1,2)) = Y((2,1)) = 3.$$

- Y is an example of random variable.

Random variables - dice

The sample space $\mathcal{S}$ can be partitioned into subsets so that outcomes in the same subset are all assigned the same value of Y.

- We let $\{Y = y\}$ be the subset of sample points in $\mathcal{S}$ assigned the value y by the random variable Y.

- For example,

$$\{Y = 2\} = \{(1, 1)\},$$

$$\{Y = 3\} = \{(1, 2), (2, 1)\},$$

$$\{Y = 5\} = \{(1, 4), (2, 3), (3, 2), (4, 1)\},$$

$$\{Y = 1\} = \emptyset.$$

Random variables - dice

In general, let A be a subset of $\mathbb{R}$, we write

$$\{Y \in A\} = \{s \in \mathcal{S} : Y(s) \in A\}.$$

$A_1 = \{11, 12\}$, $A_2 = (10, +\infty)$, $A_3 = (0, 4)$. We have

$$\{Y \in A_1\} = \{(5,6), (6,5), (6,6)\} \ ,$$

$$\{Y \in A_2\} = \{(5,6), (6,5), (6,6)\} \ ,$$

$$\{Y \in A_3\} = \{(1,1), (1,2), (2,1)\} \ .$$

Random variables - dice

We can compute the probability of events of the form $\{Y \in A\}$:

$$\mathbb{P}(\{Y \in A\}) = \mathbb{P}(\{s \in \mathcal{S} : Y(s) \in A\}).$$

- $$\mathbb{P}(\{Y \in A_1\}) = \mathbb{P}(\{(5,6),(6,5),(6,6)\}) = 3/36,$$

- $$\mathbb{P}(\{Y \in A_2\}) = \mathbb{P}(\{(5,6),(6,5),(6,6)\}) = 3/36,$$

- $$\mathbb{P}(\{Y \in A_3\}) = \mathbb{P}(\{(1,1),(1,2),(2,1)\}) = 3/36.$$

Random variables - definition

We now provide the formal definition of random variable.

- A random variable is a function from the sample space $\mathcal{S}$ to $\mathbb{R}$.
- In other words, a random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (and only one) numerical value is assigned to each sample point.

Random variables

It may not be convenient or necessary to specify a model for a random experiment via a complete description of $\mathcal{S}$ and $\mathbb{P}$. Consider the following examples:

1. Number of defective items in a batch.
2. Number of cars crossing a bridge in a day.
3. Highest daily temperature in Dublin.
4. Trading price of a gold bullion each day.

Types of random variables

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

Discrete probability distribution

- We sometimes denote $\mathbb{P}(Y = y)$ by $p(y)$.
- Because $p(y)$ is a function which assigns probabilities to each value y of the random variable Y, $p(y)$ is called the **probability function** (or probability mass function) for Y.

For any discrete probability distribution the following must be true:

1. $0 \leq p(y) \leq 1$ for all y.
2. $\sum_y p(y) = 1$ where the summation is over all possible values of y.

Example of discrete probability distribution

- X – Number observed after rolling a fair die:

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Y – Number of heads observed in two coin tosses:

	$Y = 0$	$Y = 1$	$Y = 2$
$p(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Exercise

Suppose we roll two dice.

- X = max of two outcomes.
- Y = min of two outcomes.

Find $\mathbb{P}(X = x)$ and $\mathbb{P}(Y = y)$.

Expected value

While the probability mass functions of X and Y fully characterize their distributions, it may be convenient to have more compact summaries of X and Y.

For example, what value for X and Y should we expect before tossing the dice?

The *expected value* for a discrete random variable X is defined as the *weighted average of all the values that X can take.*

Expected value

The mean or expected value of a discrete random variable is:

$$\mu = \mathbb{E}(X) = \sum_x x p(x)$$

For example, consider rolling a die again and let X be the number observed:

$$\mathbb{E}(X) = \sum_x x p(x) = 1.\left(\frac{1}{6}\right) + 2.\left(\frac{1}{6}\right) + \ldots + 6.\left(\frac{1}{6}\right) = 3.5$$

Expected value

The following (linearity) property for expected value holds (not only for discrete random variables):

- Let X and Y be two random variables, then $Z = X + Y$ is also a random variable and
$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y).$$
- More generally, for random variable $X_1, \ldots, X_n$, $Z = \sum_{i=1}^{n} X_i$ is a random variable and
$$\mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(X_i).$$

Example: 52 cards deck

Suppose we randomly shuffle a 52 cards deck. What is the expected number of times two adjacent cards of the same suit?

Reading

Textbook: Section 9.1, (page 283-286), Section 9.2 (page 287 - 288), Section 9.3 (page 290-292)

Slide 6
Expected Value (Example)

- Suppose we toss a die.
- Let X be the random varaible deonting the outcome of the toss.
- Suppose we get a prize equal to X^2 EURO.
- What is our expected win in this game?

Solution

- Let X be the outcome of the toss of a die.
- Let $Y = g(X) = X^2$ be the amount we win. Y is a random variable. Why?
- The new random variable Y has range $\mathcal{R}_y := \{1, 4, 9, 16, 25, 36\}$.
- We have $p(y) = 1/6$ for all $y \in \mathcal{R}$. Why?

Therefore,

$$\mathbb{E}(Y) = \sum_{y \in \mathcal{R}_y} y p(y) = \sum_{y \in \mathcal{R}_y} y \frac{1}{6} = \frac{91}{6}.$$

Does $\mathbb{E}(g(X)) = g(\mathbb{E}(X))$?

Expected value of $g(X)$

Result:

Let X be a discrete random variable with probability function $p(x)$ and $g(X)$ be a real-valued funciton of X. Then the expected value of $g(X)$ is given by

$$E[g(X)] = \sum_x g(x)p(x).$$

Expected value of $g(X)$

Result:

If $g(x)$ is a linear function of x, that is, $g(x) = ax + b$ for some constant a, b. Then

$$\mathbb{E}(g(X)) = g(\mathbb{E}(X)).$$

That is,

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

Variance

- Recall that the expected value of a random variable measures its centrality.
- However, it does not tell us the variability of a random variable.
- To capture the variablity of a random variable, we need a *measure of dispersion*.

Variance (example)

Let X be a discrete random variable with probability mass function

$$\mathbb{P}(X = 4) = \mathbb{P}(X = 6) = \frac{1}{4}, \quad \mathbb{P}(X = 5) = \frac{1}{2}.$$

Let Y be another discrete random variable with probability mass function

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 9) = 0.025, \quad \mathbb{P}(Y = 2) = \mathbb{P}(Y = 8) = 0.05,$$

$$\mathbb{P}(Y = 3) = \mathbb{P}(Y = 7) = 0.1, \quad \mathbb{P}(Y = 4) = \mathbb{P}(Y = 6) = 0.2,$$

$$\mathbb{P}(Y = 5) = 0.25.$$

Variance (example)

Both X and Y have the same expected value

$$\mathbb{E}(X) = \mathbb{E}(Y) = 5.$$

But Y has more *variability* than X, i.e. Y is less *concentrated* around its expected value than X.

Variance and standard deviation (definition)

If X is a random variable with mean $\mathbb{E}(X) = \mu$, the variance of X is defined to be the expected value of $(X - \mu)^2$:

$$Var(X) = \mathbb{E}[(X - \mu)^2].$$

The standard deviation of X is defined as the square root of $Var(X)$.

Useful rules for variance:

- $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
- For any constant a, b,

$$Var(aX + b) = a^2 Var(X).$$

Independence of random variables

Two random variables X and Y are said to be independent if

$$\mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \mathbb{P}(\{X \leq x\})\mathbb{P}(\{Y \leq y\}),$$

for any real numbers x, y.

An equivalent definition is the following

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

for any sets $A, B \subset \mathbb{R}$.

Independence of random variables: discrete case

Two discrete random variables X and Y are independent if and only if

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}),$$

for all real numbers x, y.

Useful rules for independence of random variables

If X and Y are independent random variables, then the random variables $f(X)$ and $g(Y)$ are independent for any two functions f and g.

Useful rules for independent random variables

If X and Y are independent random variables, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

The converse of the result is not true!

Useful rules for independent random variables

If X and Y are independent random variables, then

$$Var(X + Y) = Var(X) + Var(Y).$$

Useful rules for independent random variables

A simple corollary:

If $X_1, \ldots, X_n$ are n independent random variables, then

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n).$$

If $X_1, \ldots, X_n$ are independent with the same variance σ^2, then

$$Var(X_1 + \cdots + X_n) = n\sigma^2.$$

Important discrete random variables - Bernoulli

A random variable X is said to have a Bernoulli distribution with parameter p with $0 < p < 1$ if

$$\mathbb{P}(X = 1) = p,$$

$$\mathbb{P}(X = 0) = 1 - p.$$

We write $X \sim \text{Ber}(p)$.
The expected value and variance of X are given by

$$\mathbb{E}(X) = p, \quad Var(X) = p(1 - p).$$

Important discrete random variables - Binomial

A binomial random variable X can be obtained from the sum of n *independent and identically distributed* Bernoulli random variables $X_1, \dots, X_n$.
If X_i has Bernoulli distribution with parameter p, for $i = 1, \dots, n$, then $X = \sum_{i=1}^{n} X_i$ is a binomial random variable with

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for $k = 0, 1, \dots, n$. We write $X \sim Bin(n, p)$.

Binomial distribution

A binomial random variable X has expected value and variance

$$\mathbb{E}(X) = np,$$

$$Var(X) = np(1-p).$$

Reading

Textbook: Section 9.4 (page 293 - 295), Section 9.5 (page 299 - 301), Section 9.6.1

Slide 7
Discrete uniform random variable

A random variable X is said to have a discrete uniform distribution on integers $a, a+1, \ldots, b$ if

$$\mathbb{P}(X = k) = \frac{1}{b - a + 1},$$

for $k = a, a+1, \ldots, b$.

Discrete uniform random variable

A random variable X with a discrete uniform distribution on integers $a, a+1, \ldots, b$ has expected value and variance

$$\mathbb{E}(X) = \frac{a+b}{2},$$

$$Var(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Geometric random variable

A random variable X is said to have a geometric distribution with parameter p if

$$\mathbb{P}(X = k) = p(1 - p)^{k-1},$$

for $k = 1, 2, \ldots$.

- The random variable X can be interpreted as the number of independent Bernoulli trials with success probability p required to obtain the first success.
- We write $X \sim \text{Geo}(p)$.

Geometric random variable

A random variable X with a geometric distribution with parameter p has expected value and variance

$$\mathbb{E}(X) = \frac{1}{p},$$

$$Var(X) = \frac{1-p}{p^2}.$$

Example

Suppose products are produced by a machine with a 3% defective rate.

- What is the probability that the first defective occurs in the fifth item inspected?
- What is the probability that the first defective occurs in the first five inspections?

Poisson random variable

Poisson distribution is a limiting case of the binomial distribution.

- Consider a binomial distribution with parameters n, p.
- Suppose n is large, p is small, and $np = \lambda$ (λ a constant).

Then

$$e^{-\lambda}\frac{\lambda^k}{k!}$$

is a good approximation to

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

More precisely,

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Poisson random variable

A random variable X is said to have a Poisson distribution with parameter $\lambda > 0$ if

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

for $k = 0, 1, 2, \ldots$.

- e is a mathematical constant, $e \approx 2.718282$.
- We write $X \sim \text{Poi}(\lambda)$.

Example

Suppose a machine is known to produce 1% defective components is used for a production run of 40 components. Calculate the probability that two defective items are produced using binomial distribution and its Poisson approximation.

Examples of Poisson distribution

The Poisson distribution is suitable for modelling event counts.

Possible applications of the Poisson distribution

- Number of deaths from heart attack in a particular day.
- Number of suicides reported in a particular city in a year.
- Number of traffic accidents in a month.

Poisson random variable

A random variable X with a Poisson distribution with rate parameter $\lambda > 0$ has expected value and variance

$$\mathbb{E}(X) = \lambda,$$

$$Var(X) = \lambda.$$

Example

Births in a hospital occur randomly according to a Poisson distribution with an average rate of 1.8 births per hour.

- What is the probability of observing 4 births in a given hour at the hospital?
- What is the probability of observing at least 2 births in a given hour at the hospital?

Sum of Poisson random variables

Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2, respectively. Then $Z = X + Y$ is a Poisson random variable with parameter $\lambda = \lambda_1 + \lambda_2$.

Sum of Poisson random variables

If $X_1, X_2, \ldots, X_n$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $X = X_1 + X_2 + \cdots + X_n$ is a Poisson random variable with parameter $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Example

Births in a hospital occur randomly according to a Poisson distribution with an average rate of 1.8 births per hour.

What is the probability that we observe 5 births in a given 2-hour interval?

Poisson random variables

Let X be a Poisson random variable with rate parameter λ. Let $a > 0$ be a constant. Is aX a Poisson random variable?

Reading

Textbook: Section 9.6.2, 9.6.4, 9.6.5

Slide 8
Cumulative Distribution Function

Let X be a random variable (discrete or continuous), the *cumulative distribution function* (CDF) of X is the function $F : \mathbb{R} \to [0, 1]$ defined as

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

F is monotonically increasing with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Example

Let X be a discrete random variable with pmf

$$\mathbb{P}(X = -1) = \frac{1}{4}$$

$$\mathbb{P}(X = 0) = \frac{1}{2}$$

$$\mathbb{P}(X = 1) = \frac{1}{4}.$$

Compute the CDF of X.

CDF of a geometric random variable

Let X be a geometric random variable with parameter p, show that the CDF of X is given by

$$F(k) = 1 - (1-p)^k, \quad k = 1, 2, \ldots.$$

Memoryless property of geometric distribution

For all non-negative integers s and t,

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s)$$

or, equivalently

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t).$$

Memoryless property of geometric distribution

If X is a discrete random variable taking values $\{1, 2, 3, \ldots\}$ with probabilities $\{p_1, p_2, p_3, \ldots\}$ and satisfies the memoryless property, then X must follow a geometric distribution.

Continuous random variables

Recall that a continuous random variable is a random variable that can take uncountably infinitely many values.

We have encountered continuous random variables in this module.

- We have considered generating random numbers in $(0, 1)$.
- The generated random numbers are realizations of a random variable with a (continuous) uniform distribution on $(0, 1)$.

Continuous uniform distribution

For the uniform distribution on $(0,1)$, any two subintervals of $(0,1)$ of the same size should have the same probability. This implies that for $0 < a < b < 1$,

$$\mathbb{P}(X \in (a,b)) = b - a.$$

For example,

$$\mathbb{P}\Big(X \in \Big(0, \frac{1}{2}\Big)\Big) = \mathbb{P}\Big(X \in \Big(\frac{1}{2}, 1\Big)\Big) = \frac{1}{2}$$

$$\mathbb{P}\Big(X \in \Big(0, \frac{1}{4}\Big)\Big) = \mathbb{P}\Big(X \in \Big(\frac{1}{2}, \frac{3}{4}\Big)\Big) = \frac{1}{4}$$

CDF of continuous uniform distribution

Let $X \sim \text{Unif}(0, 1)$, what is the CDF of X?

- For $x \leq 0$, $F(x) = \mathbb{P}(X \leq x) = 0$.
- For $x \in (0, 1)$, $F(x) = x$.
- For $x \geq 1$, $F(x) = 1$.

PDF of continuous random variables

Let $X : S \to \mathbb{R}$ be a continuous random variable with CDF $F(x)$. Then there exists a function f such that

- $f(x) \geq 0, \quad \forall x$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $F(x) = \int_{-\infty}^{x} f(s)ds$

The function f is called the probability density function (PDF) of X (or simply the density of X).

PMF vs PDF

- Unlike the probability mass function (PMF) $p(x)$ for a discrete random variable, the probability density function (PDF) $f(x)$ is not a probability.
- Since $f(x)$ is not a probability, there is no restriction that $f(x)$ be less than or equal to 1.

Probability calculation for continuous random variable

Let X be a continuous random variable with CDF $F(x)$ and PDF $f(x)$.
How do we compute probability of the form $\mathbb{P}(c \leq X \leq d)$?

$$\begin{aligned}
\mathbb{P}(c \leq X \leq d) &= \mathbb{P}(X \leq d) - \mathbb{P}(X \leq c) \\
&= F(d) - F(c) \\
&= \int_{-\infty}^{d} f(s)ds - \int_{-\infty}^{c} f(s)ds \\
&= \int_{c}^{d} f(s)ds
\end{aligned}$$

$\mathbb{P}(c \leq X \leq d)$ is the area under the graph $f(x)$ between c and d.

Questions

Let X be any continuous random variable:

- What is $\mathbb{P}(X = x)$ for any x?
- Does $\mathbb{P}(X = x) = 0$ implies that X can never equal to x?

PDF of uniform random variables

Let $X \sim \text{Unif}(0, 1)$, what is the density of X?

We show that

$$f(x) = 1_{[0,1]}(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x \in (0, 1) \\ 0, & x \geq 1 \end{cases}$$

CDF to PDF

How can we find the density of a random variable X if we know its CDF?

$$F(x) = \int_{-\infty}^{x} f(s)ds \implies F'(x) = f(x)$$

by fundamental theorem of calculus.

The density is the derivative of the CDF.

Density of uniform random variables

Let $X \sim \text{Unif}(0,1)$. Recall the CDF of X is

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & x \in (0,1) \\ 1, & x \geq 1 \end{cases}$$

Differentiating, we have

$$f(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x \in (0,1) \\ 0, & x \geq 1 \end{cases}$$

Exercise

Let X be a uniform random variable on $(0, 2)$.

- Find the CDF of X.
- Find the PDF of X.
- Compute the probabilities $\mathbb{P}(X < 1.5), \mathbb{P}(X \geq 1), \mathbb{P}(0.25 < X < 1.35)$.

Example

The lifetime of a battery is described by a random variable X with CDF

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{4}x^2, & x \in (0,2) \\ 1, & x > 2. \end{cases}$$

Find the density of X.

Reading

Textbook: Section 10.1

Slide 9
Cumulative Distribution Function (Revision)

Let X be a random variable (discrete or continuous), the *cumulative distribution function* (CDF) of X is the function $F : \mathbb{R} \to [0, 1]$ defined as

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

F is monotonically increasing with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

PDF of continuous random variables

Let $X : S \to \mathbb{R}$ be a continuous random variable with CDF $F(x)$. Then there exists a function f such that

- $f(x) \geq 0, \quad \forall x$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $F(x) = \int_{-\infty}^{x} f(s)ds$

The function f is called the probability density function (PDF) of X (or simply the density of X).

Expectation of a continuous random variable

Recall that for a discrete random variable Y,

$$\mathbb{E}(Y) = \sum_y y\mathbb{P}(Y = y)$$

$$\mathbb{E}(g(Y)) = \sum_y g(y)\mathbb{P}(Y = y)$$

For a continuous random variable X with density function f,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Variance of a continuous random variable

The variance of a continuous random variable X with density f is defined as

$$\begin{aligned} Var(X) &= \mathbb{E}((X-\mu)^2) \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \end{aligned}$$

where $\mu = \mathbb{E}(X)$.

Exercise

Show that for a continuous random variable X,

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Expectation and Variance of uniform random variables

Let X be a (continuous) uniform distribution on (a, b),

$$\mathbb{E}(X) = \frac{1}{2}(a + b)$$

$$Var(X) = \frac{1}{12}(b - a)^2$$

Exponential distribution

A random variable X is said to have an **exponential distribution** with rate parameter $\lambda > 0$ if the density function of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x < \infty \\ 0\ , & \text{elsewhere.} \end{cases}$$

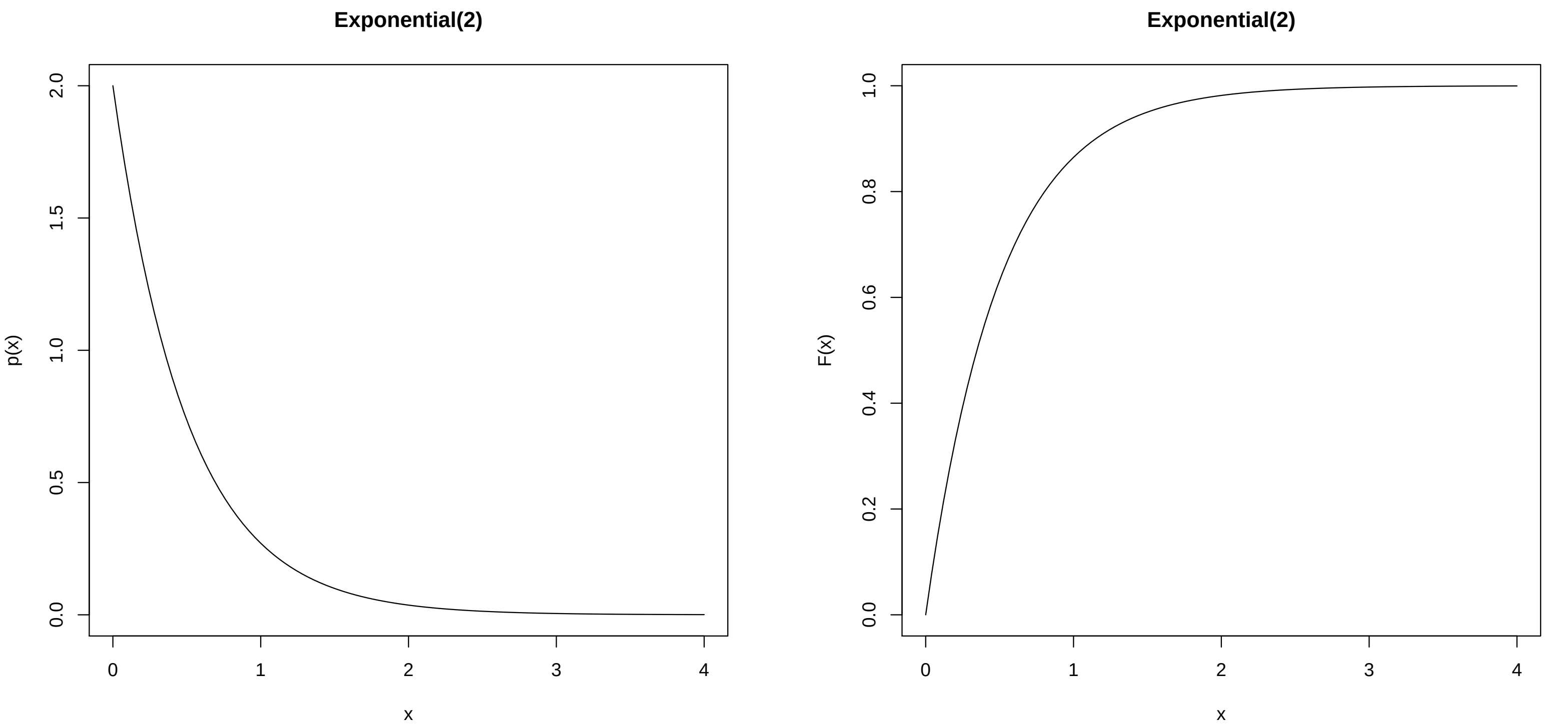
We write $X \sim \mathrm{Exp}(\lambda)$.
The exponential distribution is often used as a model for times, and is a good first model to consider.

Exponential distribution

The cumulative distribution function (cdf) of an exponential distribution is of the form

$$F(x) = 1 - \exp(-\lambda x), \text{ for } x > 0.$$

Thus, the pdf and cdf look like this:



Expected value and variance of exponential distribution

The expected value and variance of an exponential random variable X have the form

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = \frac{1}{\lambda^2}.$$

Example: Country hospital

Let X be the interval between births at a country hospital, for which the average time between births is seven days. We assume that the distribution of time between births follows an exponential distribution.

- Write down the density function for X.
- Compute the expected value and variance of X.
- What is the chance that there is a birth in the next 10 days? 10 hours? 10 minutes?

Memoryless property of exponential distribution

Let X be an exponential random variable.

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s), \quad t, s > 0$$

Memoryless property

- The memoryless property is a little perplexing.
- Let's consider it in the case of the exponential distribution as a model for human life expectancy.
- At birth an Irish male has life expectancy of 78.4 years (female=82.8).
- If life times follow an exponential distribution, we would find:

Age	Male Remaining	
	Exponential	Actual
0	78.4	78.4
10	78.4	68.8
20	78.4	58.9
30	78.4	49.4
40	78.4	39.9
50	78.4	30.6
60	78.4	21.8
70	78.4	13.9
80	78.4	7.8

Age	Female Remaining	
	Exponential	Actual
0	82.8	82.8
10	82.8	73.1
20	82.8	63.2
30	82.8	53.4
40	82.8	43.6
50	82.8	34.0
60	82.8	25.0
70	82.8	16.5
80	82.8	9.3

Waiting time paradox and exponential distribution

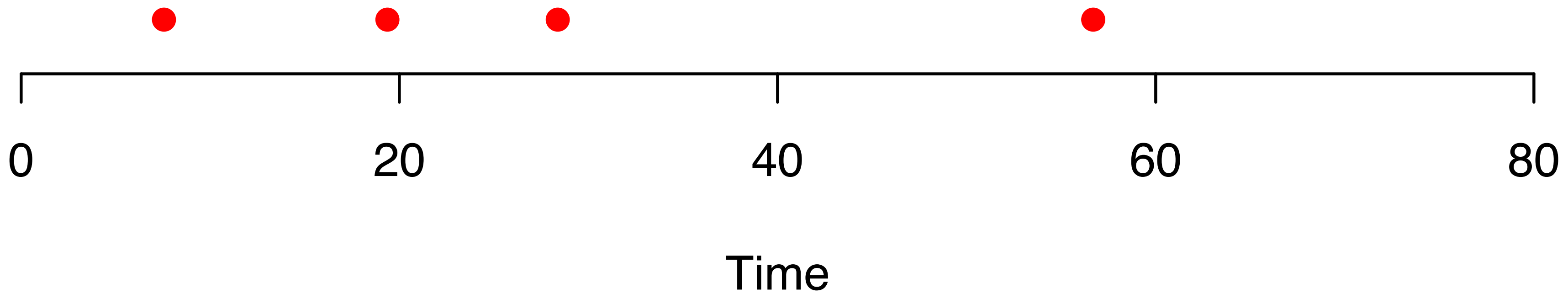
You walk to a bus stop, wait for the bus. Somebody told you that the buses, on average, arrive every 10 minutes.
Now you have been waiting for 10 minutes.

- Shouldn't the bus have arrived by now?
- How much longer are you going to wait for?

The answer is **10 minutes** if you assume the time between the arrival of any two buses to be exponentially distributed.

Poisson Process

- Suppose we observe events over a fixed time interval $(0, T]$ where $T > 0$.



- We make the following assumptions about the event process:
  - For a short time period h, at most one event can occur.
  - The probability of exactly one event occurring in the time period of length h is λh.
  - The number of events that occur in non-overlapping time periods are independent.
- If these assumptions hold, then the events follow a *homogeneous* Poisson process.

Poisson Distribution and Poisson Process

- Suppose we have a monitor a Poisson process for a unit time interval.
- Suppose we let X equal the number of events observed.
- The resulting random variable has a Poisson distribution with rate parameter λ.

  We write this as $X \sim \text{Poisson}(\lambda)$.
- The probability mass function is

$$\mathbb{P}\{X = k\} = \frac{\lambda^k \exp(-\lambda)}{k!}.$$

Poisson Distribution and Poisson Process

- Suppose we have a monitor a Poisson process for a time interval of length T.
- Suppose we let X equal the number of events observed.
- The resulting random variable has a Poisson distribution with rate parameter λT.

  We write this as $X \sim \text{Poisson}(\lambda T)$.
- The probability mass function is

$$\mathbb{P}\{X = k\} = \frac{(\lambda T)^k \exp(-\lambda T)}{k!}.$$

Exponential Distribution

- If we consider the time between any two consecutive events in a Poisson process (with rate λ), this is called the *inter event time*.
- It turns out that the inter event time X follows an exponential distribution with rate λ.

  That is, $X \sim \text{Exp}(\lambda)$.
- Therefore, the expected inter-event time is $1/\lambda$.

Poisson process example

In a certain manufacturing plant accidents have occurred in the past according to a Poisson process at the rate of 1 every 3 months.

1. What is the expected number of accidents per year? What is the standard deviation of the number of accidents per year?
2. What is the probability that there are no accidents in a given month?

Implications of Poisson Process

- Consider the arrivals of customers at a metro station. Clearly there will be many more arrivals between 5 P.M. and 6 P.M. than between 3 A.M. and 4 A.M. The stationary increments assumption of the Poisson process seems too restrictive.
- Independent increments may also be violated (e.g. occurrence of events may trigger additional events), but it will be a reasonably accurate representation of reality in many cases.

Reading

Textbook: Section 10.2, 10.3, 10.4.1, 10.4.3

Probability inequalities

- We have seen a number of probability distributions (continuous and discrete).
- In particular, we have seen the pdf, CDF, $\mathbb{E}(X)$ and $Var(X)$ for many families.
- We will now look at a number of inequalities that help us understand the connection between some of these quantities.

Markov Inequality

- Suppose that X is a positive random variable.
- Suppose we are interested in the probability that the random variable exceeds some quantity.
- The Markov inequality tells us that for $a > 0$

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

- Thus, if we know the expected value of the random variable, then we can upper bound the probability.

Markov inequality (Application)

- Suppose a financial investor knows that the expected time between stock market crashes is 12 months.
- The investor wants to know how likely it is that the stock market will go five years (60 months) without a crash.
- Markov's inequality tells us that,

$$\mathbb{P}(X > 60) \le \frac{12}{60} = \frac{1}{5}.$$

- Further, it tells us that

$$\mathbb{P}(X \le 60) \ge 1 - \frac{1}{5} = \frac{4}{5}.$$

Markov inequality (Application)

- Suppose an engineer wants to construct a barrier that will protect a city from floods.
- The engineer knows that the expected height of the greatest flood in each year is 3 metres.
- He wants to construct the wall so that it is exceeded with probability 1/10 in any one year.
- Markov's inequality tells us that,

$$\mathbb{P}(X > a) \leq \frac{3}{a}.$$

- If the wall is built to be 30 metres high, then the engineer knows that

$$\mathbb{P}(X > 30) \leq \frac{3}{30}.$$

- The Markov inequality holds for any distribution, hence the need for a large value of a.

Markov Inequality (Application)

- The value derived from the Markov inequality is very conservative, because the bound is general.
- Suppose the engineer assumed the greatest floods are exponentially distributed, what value of a gives

$$\mathbb{P}(X > a) = \frac{1}{10}?$$

  Answer: 6.9m

- This is much lower than the value required for the general bound.

Chebyshev's Inequality

- The Chebyshev inequality is an extension of the Markov inequality that applies to any random variables.
- The inequality looks at the deviations between a random variable and its expected value.
- It states that
$$\mathbb{P}(|X - \mathbb{E}(X)| > c) \leq \frac{Var(X)}{c^2}.$$

Chebyshev's Inequality (Application)

- A quality assurance engineer is monitoring a process.

  The temperature of the process is recorded every hour and has expected value μ and variance σ^2.

- What is the probability that the temperature deviates from the expected value by more than three standard deviations?

- Chebyshev's inequality tells us that

$$\mathbb{P}(|X - \mu| > 3\sigma) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9}.$$

- What is the probability that the temperature deviates from the expected value by more than six standard deviations?

- Chebyshev's inequality tells us that

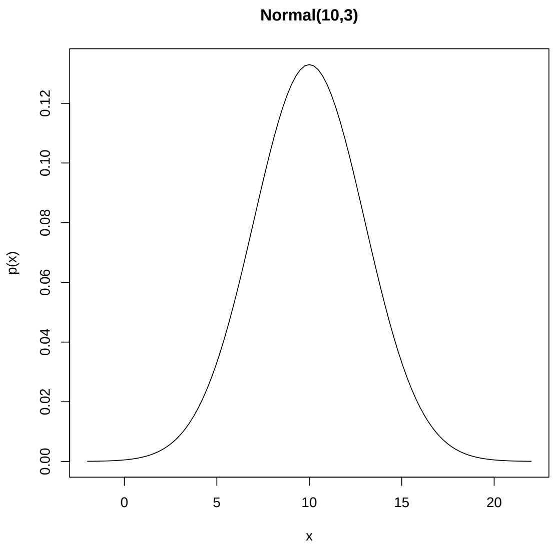
$$\mathbb{P}(|X - \mu| > 6\sigma) \leq \frac{\sigma^2}{(6\sigma)^2} = \frac{1}{36}.$$

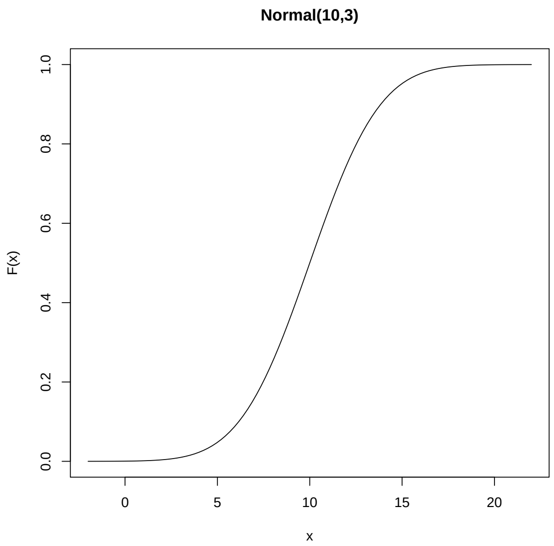
Normal distribution

- The normal distribution is the most important probability distribution.
- It is a continuous distribution and has two parameters μ and σ and the pdf of the normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

The CDF does not have a closed form expression.





Normal Distribution

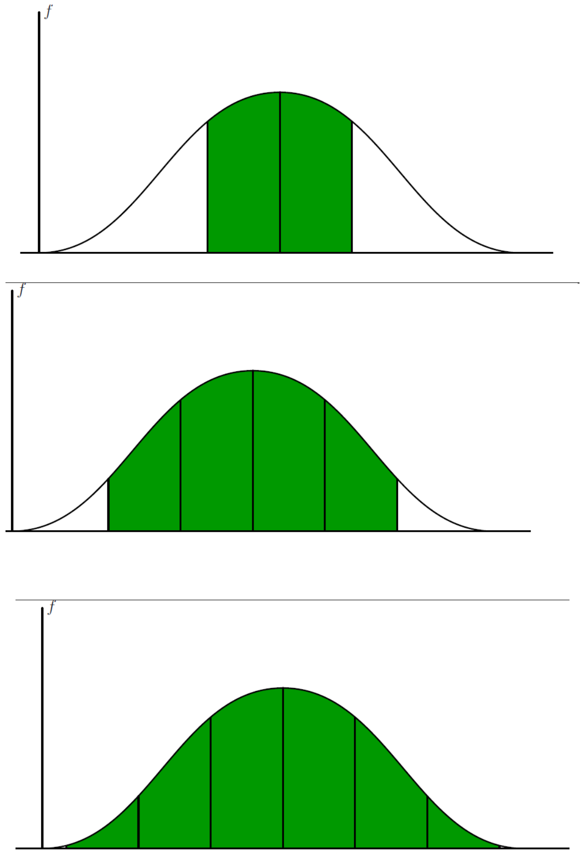
- The normal distribution has expected value and variance equal to

$$\mathbb{E}(X) = \mu \text{ and } Var(X) = \sigma^2.$$

- The distribution has a peak (mode) at μ.
- The median is also μ, that is,

$$\mathbb{P}\{X \leq \mu\} = \frac{1}{2} \text{ and } \mathbb{P}\{X \geq \mu\} = \frac{1}{2}.$$

Normal distribution



- **68%** of values lie within 1 sd from mean
- **95%** of values lie within 2 sd from mean
- **99.7%** of values lie within 3 sd from mean

Normal Distribution

- The normal distribution has been successfully applied to model measurement data.
- It has been surprisingly good at modeling:
  - weights
  - heights
  - lengths
  - temperatures
  - fluctuations in measurements

Example: Paper Friction

In a study to investigate the paper feeding process in a photocopier the coefficient of friction is a proportion which measures the degree of friction between adjacent sheets of paper in the paper stack. This coefficient is assumed to be normally distributed with mean 0.55 and standard deviation 0.013. During system operation the friction coefficient is measured at randomly selected times.

1. Find the probability that the friction coefficient falls between 0.53 and 0.56.

2. Is it likely to observe a friction coefficient below 0.52? Explain.

Example: Skyscraper

The lifetime of steel joists in a skyscraper are independent and normally distributed with a mean of 50 years and a standard deviation of 6 years.

1. What is the probability that a steel joist lasts more than 55 years?
2. What proportion of the steel joists would you expect to last between 42 and 58 years?
3. In a random sample of 4 joists what is the probability that 2 last less than 55 years?

Standard Normal

Let X be a normal random variable with mean μ and variance σ^2, i.e. $X \sim N(\mu, \sigma^2)$, what is the distribution of $Z = \frac{X-\mu}{\sigma}$?

It turns out that Z is a normal random variable with mean 0 and variance 1, i.e. $Z \sim N(0, 1)$.

- The operation $\frac{X-\mu}{\sigma}$ is called standardization.
- The random variable $Z \sim N(0, 1)$ is called the standard normal random variable.

Standard Normal

- Let $Z \sim N(0, 1)$ be a standard normal random variable.
- Define $X = Z\sigma + \mu$.
- Then $X \sim N(\mu, \sigma^2)$.

Tail Probabilities

- The normal distribution has relatively light tail probability,

$$\mathbb{P}\{|X - \mu| > k\sigma\}$$

  is relatively small.

- We know by the Chebyshev Inequality that

$$\mathbb{P}\{|X - \mu| > k\sigma\} \leq \frac{1}{k^2}.$$

- If $X \sim N(\mu, \sigma^2)$,

$$\mathbb{P}\{|X - \mu| > k\sigma\} \leq 2\frac{\exp(-k^2/2)}{k\sqrt{2\pi}}.$$

Reading

Textbook: Section 10.4.7

Joint distribution (discrete case)

Let X and Y be two discrete random variables. The joint probability mass function of X and Y is denoted by

$$p(x, y) := \mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

Let $p_X(x)$ and $p_Y(y)$ be the marginal probability mass functions:

$$p_X(x) := \mathbb{P}(X = x), \quad p_Y(y) := \mathbb{P}(Y = y).$$

The following property holds:

$$p_X(x) = \sum_y p(x, y).$$

Similary,

$$p_Y(y) = \sum_x p(x, y).$$

Example: two dice

- Let D_1, D_2 be the outcome of tossing two dice independently.
- Define $X = \max(D_1, D_2)$, $Y = |D_1 - D_2|$.
- Determine the joint distribution of X and Y.
- Determine the marginal distributions of X and Y.

Independence

Recall that two discrete random variables X and Y are independent if

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \forall (x, y).$$

That is, $p(x, y) = p_X(x)p_Y(y)$ for all (x, y).

To prove X and Y are independent, we should check the above equality holds for all combinations (x, y).
If we want to prove X and Y are not independent, it suffices to find a counterexample, i.e. a pair (x, y) for which

$$p(x, y) \neq p_X(x)p_Y(y).$$

Joint distribution (continuous case)

Two continuous random variables X and Y have joint density function $f(x, y)$ if the joint CDF admits this representation

$$F(x, y) := \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dt ds,$$

where

- $f(s, t) \geq 0, \forall s, t.$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s, t) dt ds = 1.$

Joint probability and marginal densities

$$
\begin{aligned}
\mathbb{P}((X,Y) \in [a,b] \times [c,d]) &= \mathbb{P}(X \in [a,b], Y \in [c,d]) \\
&= \int_a^b \int_c^d f(x,y)dydx
\end{aligned}
$$

The marginal density of X can be obtained as

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy.$$

Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

Example

Let X, Y be two continuous random variables with joint density

$$f(x,y) = \frac{1}{12}(1 + x + y)$$

for $0 < x < 2, 0 < y < 2$, and

$$f(x,y) = 0$$

otherwise.

Compute the probability $\mathbb{P}(\min(X,Y) \leq 1)$.

Independence and expectation

If X, Y are continuous, X and Y are independent if and only if

$$f(x,y) = f_X(x) f_Y(y).$$

Recall for the univariate case, $X \sim f$,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Analogously, if $(X,Y) \sim f(x,y)$,

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dy dx.$$

Exercise

A random variable X is strictly larger than -100. Suppose $\mathbb{E}(X) = -60$. Use Markov's inequality to find an upper bound on $\mathbb{P}(X > -20)$.

Exercise

A biased coin has 20% probability of landing on heads. Suppose the coin is flipped 20 times and each flip is independent.

1. Use Markov's inequality to bound the probability it lands on heads at least 16 times.
2. Use Chebyshev's inequality to bound the probability it lands on heads at least 16 times.

Exercise

The height of a randomly selected basketball player is assumed to be normally distributed with mean 200 cm and variance 64 cm. What is the probability that a randomly selected basketball player is over 210 cm?

Exercise

Suppose X is a normal random variable with mean 6. If $\mathbb{P}(X > 16) = 0.0228$, find the standard deviation of X.

Exercise

Suppose X and Y are continuous random variables with joint pdf

$$f(x, y) = \frac{3}{2}(x^2 + y^2), \quad x \in [0, 1], y \in [0, 1].$$

1. Show that $f(x, y)$ is a valid pdf.
2. Let A be the event that $X > 0.3$ and B be the event that $Y > 0.5$. Find the probability of $A \cap B$.
3. Find the cdf $F(x, y)$.
4. Find the marginal pdf $f_X(x)$ of X.
5. Find the probability $X < 0.5$.

Sample Exam Question 1

1. A fair die is cast until a 1 appears. Determine
   (a) the probability that it must be cast exactly 2 times,
   (b) the probability that it must be cast less than 2 times,
   (c) the probability that it must be cast at least 2 times,
   (d) the expected number of times the die has to be cast.
2. Assume that X and Y have Bernoulli distribution with parameters p and $1-p$, respectively, and compute $\mathrm{Var}(Z)$, where $Z = (X - Y)/2$.

Sample Exam Question 2

Let (X, Y) have the distribution defined by the following table of probabilities:

$X \backslash Y$	1	2	3
1	$1/6$	$1/12$	0
2	0	0	$1/6$
3	$1/6$	$1/12$	$1/12$
4	$1/12$	0	$1/6$

Determine the following:

1. the marginal probability distributions of X and Y,
2. $\mathbb{P}(X = Y)$,
3. $\mathbb{P}(X + Y \leq 4)$,
4. $\mathbb{P}(X = 3 \,|\, Y = 1)$,
5. $\mathbb{E}(XY)$.

Reading

Textbook: Section 11.1-11.3