Introduction to the Probability.

Chapter 1. Experiments with Random Outcomes.

§ 1.1. Sample Spaces and Probabilities

Definition 1.1. Ingredients of a probability model.

- The sample space Ω is the set of all possible outcomes of the experiment. Elements of Ω are called sample points and usually denoted by ω.
- Subsets of Ω are called events. The collection of events in Ω is denoted by F.
- The probability measure (also called the probability distribution or simply probability) P is a function from F into real numbers. Each event A has a probability $P(A)$, and P satisfies the following axioms (known as Kolmogorov Axioms, 1930s)

(i) $0 \leq P(A) \leq 1, \ \forall A \in F$.

(ii) $P(\Omega) = 1, \ P(\phi) = 0$.

(iii) If $A_1, A_2, A_3, \ldots$ is a sequence of pairwise disjoint events then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The triple (Ω, F, P) is called a probability space.

Fact 1.2. If $A_1, \ldots, A_n$ are pairwise disjoint events then by Axiom (iii) $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$.
$(A_{n+1} = A_{n+2} = \cdots = \phi)$

Example 1.3. Flipping a fair coin. $\Omega = \{H, T\}$, $\mathcal{F} = \{\phi, \{H\}, \{T\}, \{H, T\}\}$. $\mathbb{P}\{H\} = \mathbb{P}\{T\} = 1/2$.

Example 1.4. Rolling a standard six-sided die. $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}\{\omega\} = 1/6$ for all $\omega \in \Omega$. If $A = \{2, 4, 6\}$ then
$\mathbb{P}(A) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = 3/6 = 1/2$.

Example 1.5. Assume that for an unfair coin the heads is three times as likely as tails. So $\mathbb{P}(H) = 3/4$ and $\mathbb{P}(T) = 1/4$.

Similarly, suppose that for a loaded die a six is twice as likely as any other number. In this case the probability measure is $\mathbb{P}\{6\} = 2/7$ and $\mathbb{P}\{1\} = \cdots = \mathbb{P}\{5\} = 1/7$ so that $\mathbb{P}(\Omega) = 1$.

Alternatively, suppose that we scratch away 5 from the original fair die and turn it into a second 2. Then the probability measure is
$\mathbb{P}\{1\} = \mathbb{P}\{3\} = \mathbb{P}\{4\} = \mathbb{P}\{6\} = 1/6$, $\mathbb{P}\{5\} = 0$, $\mathbb{P}\{2\} = \frac{2}{6}$.

Example 1.6. Now assume that the experiment consists of a roll of a pair of dice of different colors, say blue and red. Then $\Omega = \{(i, j) \mid i, j = 1, \ldots, 6\}$, where i denotes the number on the face of the blue die and the j denotes the number on the red die. Assume the dice are fair. Then $\mathbb{P}\{i, j\} = 1/36$ for all $\{i, j\} \in \Omega$.

If $D = \{\text{the sum of the two dice is } 8\} = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$

so that $\mathbb{P}(D) = 5/36$.

__Example 1.7.__ Assume that a fair coined is flipped three times. Let us encode 0 for heads and 1 for tails. Then $\Omega = \{(0,0,0), (1,0,0), \dots, (1,1,1)\}$. Ω has 8 elements. Also $\mathbb{P}\{\omega\} = 1/8$.

More generally, if n experiments are performed with sample spaces $A_1, \dots, A_n$, then the sample space of the "combined" experiment is

$$\begin{aligned}\Omega &= A_1 \times \dots \times A_n. \\ &= \{(x_1, \dots, x_n) \mid x_i \in A_i, \ i = 1, \dots, n\}.\end{aligned}$$

__§1.2. Random Sampling.__

Suppose that Ω, the sample space, is a finite set and each $\omega \in \Omega$ has the same probability. Then $\mathbb{P}(\omega) = 1/\#\Omega$, where $\#A$ denotes the number of elements in the subset $A \subseteq \Omega$. In this case, we say that each outcome $\omega \in \Omega$ is equally likely.

__Fact 1.8.__ In this case, $\mathbb{P}(A) = \sum_{a_i \in A} \mathbb{P}(a_i) = \sum_{a_i \in A} \frac{1}{\#\Omega} = \frac{\#A}{\#\Omega}$.

__Sampling Mechanisms__

- Ordered sample is built by choosing objects one at a time and by keeping track of the order in which these objects were chosen. (e.g. ordered books in a bookshelf.
- Unordered sampling means that we do not care about the order in which these objects were

chosen.

- Sampling with replacement means that once an object is chosen after recording the object is put back.
- Sampling without replacement means that once an object is chosen the object is not put back.

Sampling with replacement, order matters.

Example 1.10. Suppose that an urn contains 5 balls labelled 1, 2, 3, 4, 5. Sample 3 balls with replacement and produce an ordered list of the numbers drawn. The sample space is

$$\Omega = \{1,2,3,4,5\}^3 = \{(s_1, s_2, s_3) \mid s_i \in \{1,2,3,4,5\}\}.$$

So $\#\Omega = 5^3 = 125$. Assume that outcomes are equally likely. Hence, $\mathbb{P}\{\omega\} = 1/125$ for any $\omega \in \Omega$.

Sampling without replacement, order matters

Suppose again an urn contains n balls numbered $1, 2, \dots, n$. The sample space of the experiment of choosing $k \leq n$ balls without replacement, keeping track the order is

$$\Omega = \{(s_1, \dots, s_k) \mid 1 \leq s_i \leq n,\ s_i \neq s_j \text{ if } i \neq j\}.$$

Then clearly, $|\Omega| = n(n-1)\cdots(n-k+1) \doteq (n)_k$

Example 1.11. Consider the urn example with 5 balls. We choose 3 balls without replacement, order matters. Then $|\Omega| = 5\cdot 4\cdot 3 = 60$. So, for example $\mathbb{P}\{2,1,5\} = 1/60$. The outcome $\{2,2,3\}$ is not possible.

Sampling without replacement, order irrelevant.

In this case since the order is irrelevant only the subset of the chosen k balls from the urn with n balls, would matter. So, $\Omega = \{A \subseteq \{1,2,..,n\} \mid \#A = k\}$ and $\#\Omega = \binom{n}{k} = n!/(k!(n-k)!)$.

Hence, for any $\omega \in \Omega$, $\mathbb{P}(\omega) = 1/\binom{n}{k}$.

Further Examples. Clearly, there are several other sampling mechanisms. We'll present 3 examples.

Example 1.13. Suppose we have a class of 24 students. Consider the below three senarios that each involve choosing three children.

(a) Every day a random student is chosen to lead the class to lunch, without regard to previous choices. What is the probability that Cassidy was chose on Monday and Wednesday, and Aaron on Tuesday?

This is a sampling with replacement to produce an ordered sample. Over a period of three days the total number of choices is 24^3. Thus $\mathbb{P}\{(\text{Cassidy}, \text{Aaron}, \text{Cassidy})\} = 1/24^3 = 24^{-3}$.

(b) Three students are chosen randomly to be class president, vice president, and treasurer. No student can hold more than one office. What is the probability that Mary is president, Cory is vice president, and Matt treasurer?

This is a sampling without replacement to produce an ordered sample. Thus

$P\{\text{Mary is president, Cory is vice president, Matt is treasurer}\} = 1/24 \cdot 23 \cdot 22 = 1/12,144.$

Now we are asked instead, for the probability that Ben is either president or vice president.

The number of outcomes in which Ben ends up president is $1 \cdot 23 \cdot 22$ (1 choice for president, 23 choices for vice president and 22 choices for treasurer). Similarly, there are $23 \cdot 1 \cdot 22$ choices which Ben ends up vice president.

So, $P\{\text{Ben is president or vice president}\}$

$$= (1 \cdot 23 \cdot 22 + 23 \cdot 1 \cdot 22)/24 \cdot 23 \cdot 22$$
$$= 1/12.$$

(c) A team of three children is chosen at random. What is the probability that team consists of Shane, Heather and Laura?

This is a sampling without replacement to produce a sample without order.

So, $P = \{\text{the team is } \{\text{Shane, Heather, Laura}\}\} = 1/\binom{24}{3} = 1/2024$

On the other hand, the probability that Mary is on the team is as follows: There are $\binom{23}{2}$ teams that contains Mary. So,

$P\{\text{the team includes Mary}\} = \binom{23}{2}/\binom{24}{3} = \frac{3}{24} = 1/8.$

Example 1.14. Suppose that an urn contains ten marbles labelled 1 to 10 and we sample two marbles without replacement. What is the probability that our sample contains the marble labelled 1? Let A be the event that this happens. So we ask $\mathbb{P}(A)$.

Solution with order. Clearly, $\#\Omega = 10 \cdot 9 = 90$ and $A = \{(1,2), \ldots, (1,10), (2,1), (3,1), \ldots, (10,1)\}$ so that $\#A = 18$. So, $\mathbb{P}(A) = \#A / \#\Omega = 18/90 = 1/5$.

Solution without order. Ω is the set of all two element subsets of $\{1,2,\ldots,10\}$. So $\#\Omega = \binom{10}{2} = 9 \cdot 10/2 = 45$. On the other hand, $A = \{\{1,2\},\{1,3\},\ldots,\{1,10\}\}$ so that $\#A = 9$. Hence, $\mathbb{P}(A) = \#A/\#\Omega = 9/45 = 1/5$.

Example 1.15. Rodney packs 3 shirts for a trip. The closet contains 10 shirts: 5 striped, 3 plaid, and 2 solid colored ones. What is the probability that he chose 2 striped and 1 plaid shirt?

For simplicity let's label the shirts: Striped shirts are labelled 1, 2, 3, 4, 5, the plaid ones 6, 7, 8 and the solid colored ones are 9 and 10.

Solution without considering order:

$\Omega = \{(x_1, x_2, x_3) \mid x_i \in \{1, \ldots, 10\}, x_i \neq x_j\}$. So, $\#\Omega = \binom{10}{3} = 120$. On the other hand, the set of favorable outcomes is $A = \{(x_1, x_2, x_3) \mid x_1, x_2 \in \{1,2,3,4,5\}, x_1 \neq x_2, x_3 \in \{6,7,8\}\}$.

So, $\#A = \binom{5}{2}\cdot\binom{3}{1} = 10\cdot 3 = 30$. Hence, $P(A) = \frac{30}{120} = 1/4$.

Now, solution with ordered samples.

$\tilde{\Omega} = \{(x_1, x_2, x_3) / x_i \in \{1, 2, \ldots, 10\}, x_1, x_2, x_3 \text{ distinct}\}$.
So, $\#\tilde{\Omega} = 10\cdot 9\cdot 8 = 720$.

Now let's count the elements of $\tilde{A}$: (i) Choose the plaid shirt (3 choices), (ii) Choose the position of the plaid shirt in the ordering (3 choices), (iii) Choose the first striped shirt and place it in the first available position (5 choices), (iv) Choose the second striped shirt and place it in the last available position (4 choices).
Hence, $\#\tilde{A} = 3\cdot 3\cdot 5\cdot 4 = 180$ and $P(\tilde{A}) = \frac{\#\tilde{A}}{\#\tilde{\Omega}} = 1/4$.

§1.3. Infinitely Many Outcomes.

Example 1.16. Flip a fair coin until the first tail comes up? Record the number of flips required as the outcome of the experiments. Then $\Omega = \{\infty, 1, 2, \ldots\}$, where ∞ means that tail never comes up? Clearly $P(k) = 2^{-k}$ and thus $1 = P(\Omega) = P\{\infty, 1, 2, \ldots\} = P(\infty) + \sum_{k=1}^{\infty} 2^{-k}$
$\Rightarrow 1 = P(\infty) + 1$ so that $P(\infty) = 0$.

Example 1.17. Pick a real number between 0 and 1. Let X denote the chosen number. Then $\Omega = [0, 1]$. The probability that the chosen number is in a given interval say $[a, b] \subseteq [0, 1]$

is $b-a$, the length of the interval:
$\mathbb{P}\{X$ lies in the interval $[a,b]\} = b-a$.
In particular, $\mathbb{P}\{X = 0.3\} = \mathbb{P}\{X$ lies in $[0.3, 0.3]\}$
$= 0.3 - 0.3 = 0$.

__Example 1.8.__ Consider a circular dart board in the shape of a disk with radius of 9 inches. The bullseye is a disk of diameter 1/4 inches. What is the probability that a dart randomly thrown on the board hits the bullseye?
In this case, $\Omega = \{(x,y)\in\mathbb{R}^2 \mid x^2+y^2 \leq 9^2\}$ and $A = \{(x,y)\in\Omega \mid x^2+y^2 \leq (1/4)^2\}$. So our intuition tells us that $\mathbb{P}(A) = \text{Area}(A)/\text{Area}(\Omega)$
$= 1/4/9^2 = 1/(36)^2 \approx 0.00077$

All these three examples have infinite sample spaces. However, the first one is countable infinite while the last two are uncountable. The computation of the probability for the first example involves still counting while for the others one needs to use the length or the area, which are basically integration that allows us measuring ("counting") uncountable sets.
Finite or countably infinite sample spaces will be both called discrete sample spaces.

§1.4. Consequences of the rules of probability.

Decomposing an event. Suppose that an event $A \subseteq \Omega$ is a disjoint union of events $A_1, A_2, \ldots$. Then $\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$

Example 1.19. An urn contains 30 red, 20 green and 10 yellow balls. Draw two balls without replacement. What is the probability that the sample contains exactly one red or exactly one yellow?

$\mathbb{P}$(exactly one red or exactly one yellow)

$= \mathbb{P}$(red and green) $+ \mathbb{P}$(yellow and green)

$+ \mathbb{P}$(red and yellow).

$$= \frac{30 \cdot 20}{\binom{60}{2}} + \frac{10 \cdot 20}{\binom{60}{2}} + \frac{30 \cdot 10}{\binom{60}{2}} = \frac{110}{177}$$

Example 1.20. Peter and Mary take turns rolling a fair die. If Peter rolls 1 or 2 he wins and the game stops. If Mary rolls 3, 4, 5 or 6, she wins and the game stops. They keep rolling until one of them wins. Suppose Peter rolls first.

(a) What is the possibility that Peter wins and rolls at most 4 times?

Define the following events:

$A = \{$Peter wins and rolls at most 4 times$\}$

$A_k = \{$Peter wins on his k^{th} roll$\}$.

Then $A = A_1 \cup A_2 \cup A_3 \cup A_4$ and the events are mutually disjoint. So, $P(A)=\sum_{k=1}^{4} P(A_k)$.

On the other hand, A_k means that Peter and Mary fail $k-1$ times and Peter wins at his kth trial. So, $P(A_k)=\left(\frac{4}{6}\right)^{k-1}\left(\frac{2}{6}\right)^{k-1}\frac{2}{6}=\frac{1}{3}\left(\frac{2}{9}\right)^{k-1}$

So, $P(A)=\frac{1}{3}\sum_{j=0}^{3}\left(\frac{2}{9}\right)^j=\frac{3}{7}\left(1-\left(\frac{2}{9}\right)^4\right)$

(b) What is the probability that Peter wins?

$$P(\text{Peter wins})=\sum_{k=1}^{\infty}P(A_k)=\lim_{k\to\infty}\frac{3}{7}\left(1-\left(\frac{2}{9}\right)^k\right)=\frac{3}{7}.$$

Similarly one can compute that $P(\text{Mary wins})=\frac{4}{7}$.

<u>Question</u>: What is the average number of rolls for a game to finish?

<u>Events and Complements</u>. For any event $A\subseteq\Omega$ $A\cup A^c=\Omega$ so that $1=P(\Omega)=P(A)+P(A^c)$.

Sometime computing $P(A^c)$ might be much easier than $P(A)$, as the example below exhibits.

<u>Example 1.21</u>. Roll a fair dice 4 times. What is the possibility that some number appears more than once?

$A=\{\text{some number apperas more than once}\}$
$A^c=\{\text{all rolls are different}\}$.

$P(A)$ is complicated to compute but $P(A^c)$ is almost trivial: $P(A^c)=6\cdot5\cdot4\cdot3/6^4=5/18$. So $P(A)=13/18$.

This idea may be expanded as follows. For any other event B we may write $B = (B \cap A) \cup (B \cap A^c)$, so that $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$.

Monotonicity of Probability.

If $A \subseteq B$, then write $B = A \cup (B \setminus A)$ so that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.

Example 1.22. Suppose we flip a fair coin. Let's show that the probability that we never see a tail is zero: Let A be the event that we never see a tail and A_k be the event that all of the first k flips are head. Then $A \subseteq A_k$ for all k. So $\mathbb{P}(A) \leq \mathbb{P}(A_k) = 1/2^k, \forall k$. Hence $\mathbb{P}(A) = 0$.

Inclusion-Exclusion Principle.

Fact 1.23. (Inclusion-Exclusion formulas for two or three events)

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

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

General Formula

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n)$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}).$$

Example 1.27. Suppose n people arrive for a show and leave their hats in the cloakroom (vestiyer). Unfortunately, the cloakroom attendant mixes up all the hats completely so that each person leaves with a random hat. Let's assume that all $n!$ assignments of hats are equally likely. What is the probability that no one gets his/her own hat? How does this probability behave as $n \to \infty$.

Define the events: $A = \{$no one gets his/her own hat$\}$ and $A_i = \{$person i gets his/her own hat$\}$.

So, $A = \Omega - A^c = \Omega - \left(\bigcup_{i=1}^{n} A_i\right)$.

$\Rightarrow P(A) = 1 - P\left(\bigcup_{i=1}^{n} A_i\right)$.

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P\{i_1, \dots, i_k \text{ gets their own hats}\}$$
$$= (n-k)!/n!$$

This follows from the following observation: If k persons get their own hats then there are $(n-k)!$ ways to distribute the remaining hats.

On the other hand, there are $\binom{n}{k}$ way of choosing k persons among n people. So the sum

$$\sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!} = 1/k!$$

Hence, $P(A) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!}$

$$\Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + - \cdots + (-1)^n \frac{1}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

So, the limit $\mathbb{P}(A^c)$ as $n \to \infty$ becomes

$\lim_{n \to \infty}$ {no person among n people gets the correct hat}

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} = 1/e.$$

Remark: As $n \to \infty$, the probability that a random permutation of n elements has no fixed point converges to e^{-1}.

§1.5. Random Variables: a first look.

Definition 1.28. Let Ω be a sample space. A random variable is a function from Ω to the real numbers.

Example 1.29. Roll a pair of dice. Let X_1 denote the outcome of the first die, X_2 the second die and S denote the sum of X_1 and X_2.

So, if $(i,j) \in \Omega$ then

$$X_1(i,j) = i, \quad X_2(i,j) = j \text{ and } S(i,j) = i + j.$$

Let's compute $\mathbb{P}(S=8)$ = the probability that the sum is 8.

$$\mathbb{P}(S=8) = \sum_{i+j=8} \mathbb{P}(X_1 = i, X_2 = j) = 5/36. \quad ((2,6),(3,5),(4,4),(5,3),(6,2))$$

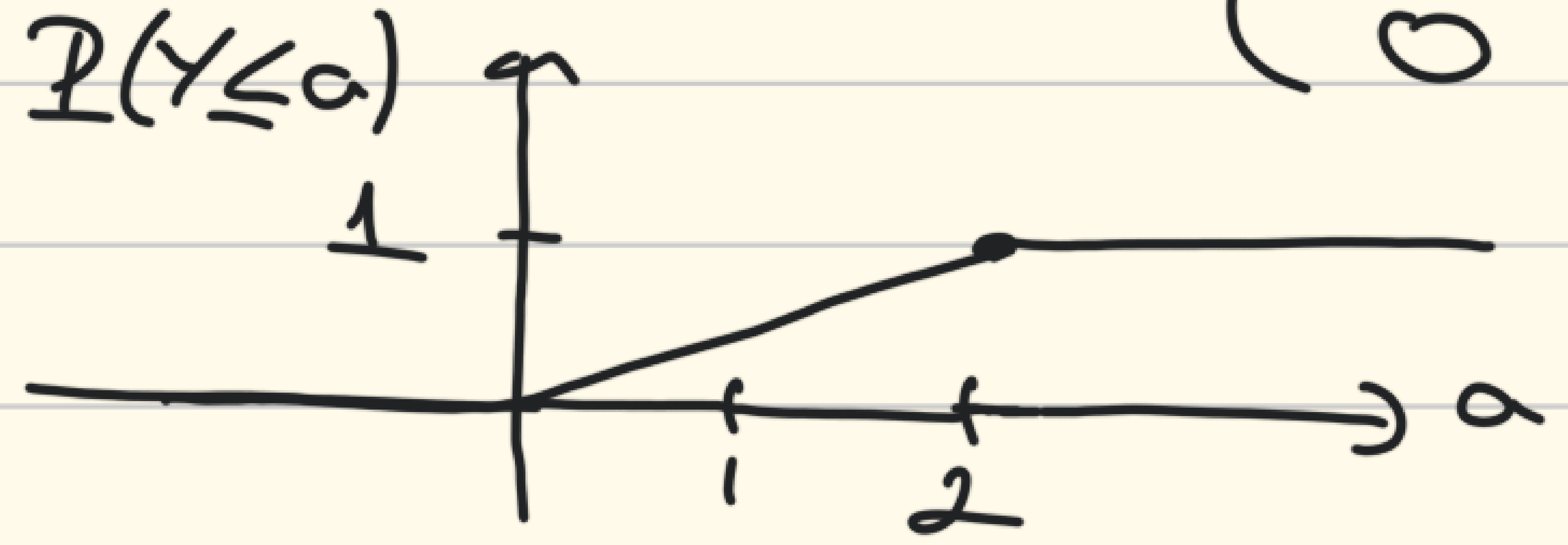
Example 1.31. Select a point uniformly at random

from the interval $[0,1]$. Let Y be equal to twice the chosen point. The sample space is $\Omega=[0,1]$.
$Y:\Omega \to \mathbb{R}$, $Y(\omega)=2\omega$, $\omega\in\Omega$.

For any $a\in\mathbb{R}$ the probability
$$\mathbb{P}(Y\le a)=\{\omega\in[0,1] \mid 2\omega\le a\}$$
$$=\text{length of } [0,1]\cap\left(-\infty,\tfrac{a}{2}\right].$$

So, $\mathbb{P}(Y\le -3)=0$, $\mathbb{P}(Y\le 3)=1$, $P(0.7)=0.35$.

Indeed, $\mathbb{P}(Y\le a)=\begin{cases} 1 & \text{if } a\ge 2 \\ a/2 & \text{if } 0\le a\le 2 \\ 0 & \text{if } a\le 0. \end{cases}$



Example 1.32. A random variable X is degenerate if $\mathbb{P}(X=b)=1$ for some $b\in\mathbb{R}$. Note that any constant random variable is degenerate.

However, not all degenerate random variables are constants.

Definition 1.33. Let X be a random variable. The probability distribution of the random variable X is the collection of probabilities $\mathbb{P}\{X\in B\}$ for sets $B\subseteq\mathbb{R}$.

Definition 1.34. A random variable is a discrete random variable if its image $X(\Omega)$ is either finite or countably infinite.

In particular, if $X(\Omega) = \{k_1, k_2, k_3, \ldots\}$ then

$$\sum_i \mathbb{P}(X=k_i) = \mathbb{P}\Big(\bigcup_i (X=k_i)\Big) = \mathbb{P}(\Omega) = 1.$$

Definition 1.35. The probability mass function (p.m.f) of a discrete random variable X is the function p (or p_X) defined by $p(k) = \mathbb{P}(X=k)$, k any value of X.

Example 1.36. The probability mass functions of X_1 and S of Example 1.29 are as follows:

k	1	2	3	4	5	6
$p_{X_1}(k)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

k	2	3	4	5	6	7	8	9	10	11	12
$p_S(k)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

Example 1.38. Dartboard of radius 9.

$$P(R \leq a) = \frac{\pi a^2}{\pi 9^2} = \frac{a^2}{81}.$$

§1.6. Finer Points: Continuity of the Probability measure.

Fact 1.39. Suppose we have an infinite nested sequence of increasing events $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$. Let $A_\infty = \bigcup_{k=1}^{\infty} A_k$. Then $\mathbb{P}(A_\infty) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Proof: For $n = 2, 3, \ldots$ let $B_n = A_n \setminus A_{n-1}$.
Now we have,

$$A_n = A_{n-1} \cup B_n = A_{n-2} \cup B_{n-1} \cup B_n = \cdots = A_1 \cup B_2 \cup \cdots \cup B_n$$

and hence

$A_\infty = \bigcup_{n=1}^{\infty} A_n = A_1 \cup B_2 \cup B_3 \cup \cdots$, which is a disjoint union of events.

So,

$$\begin{aligned} \mathbb{P}(A_\infty) &= \mathbb{P}(A_1) + \mathbb{P}(B_2) + \mathbb{P}(B_3) + \cdots \\ &= \lim_{n \to \infty} \left(\mathbb{P}(A_1) + \mathbb{P}(B_2) + \cdots + \mathbb{P}(B_n) \right) \\ &= \lim_{n \to \infty} \mathbb{P}(A_n). \end{aligned}$$

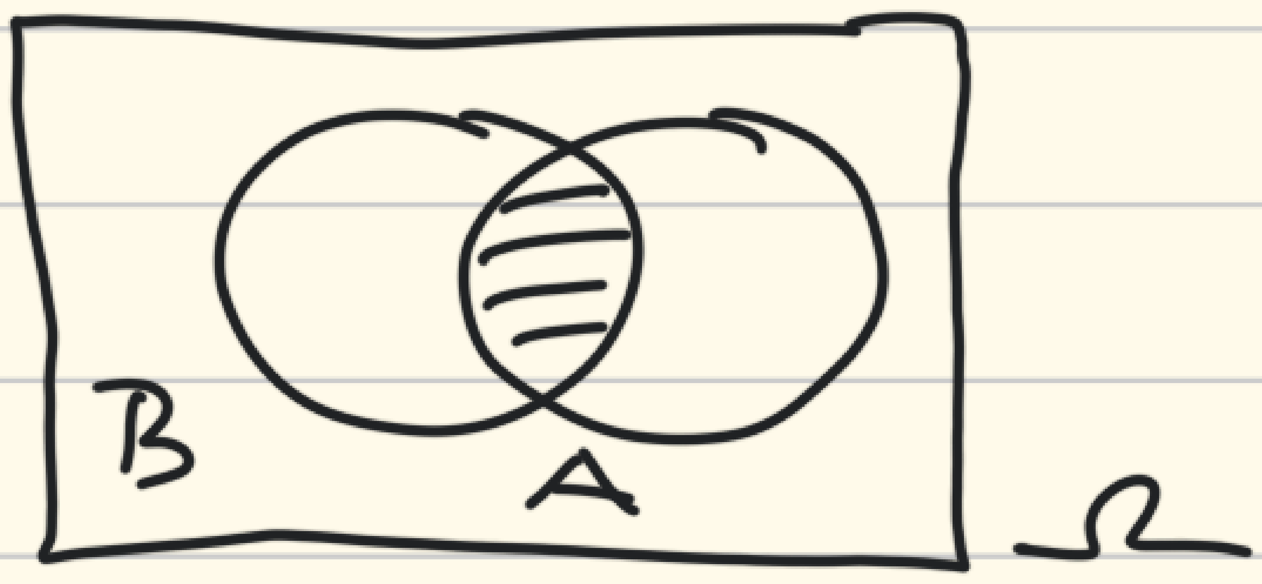
This property of a probability measure is called continuity of the measure.

CHAPTER 2. Conditional Probability and Independence

§ 2.1. Conditional Probability

Definition 2.1. Let B be an event in the sample space Ω such that $\mathbb{P}(B)>0$. Then for any event $A \subseteq \Omega$ the conditional property of A given by is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}, \text{ when } \mathbb{P}(AB) = \mathbb{P}(A\cap B).$$



Fact 2.2. Assume the above setup. Then $\mathbb{P}(\cdot|B)$ satisfies the axioms of Definition 1.1.

Example There are 3 boxes and one of them contains a ball. You are asked to guess the box that contains the ball. Suppose you tell that the 3. box contains the ball. Then the first box is opened and you see that the ball is not in the the first box. Would you change your guess or insist on your first guess that the ball is in the 3. box.

Solution: If you do not change your guess the probability that you will win is 1/2. However, the probability that the ball is in Box 1 or Box 2 was 2/3 and still is. So the probability that Box 2 contains the ball is 2/3 which is larger that 1/2. Hence, it is wise to change your guess to Box 2.

Example 2.3. Recall from Example 1.7 that the probability of getting 2 heads out of three coin flips is 3/8. Suppose that the first coin flip is revealed to be heads. What is the answer of the problem once we have this information?

Solution: $A = \{\text{exactly two heads}\}$
$= \{(0,0,1),(0,1,0),(1,0,0)\}$ so that $P(A) = 3/8$.

$B = \{\text{first flip is heads}\} = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1)\}$

Hence, $P(A|B) = \dfrac{P(AB)}{P(B)} = \dfrac{2}{4} = 1/2$.

Fact 2.4. Suppose that we have an experiment with finitely many equally likely outcomes and B is not the empty set. Then for any set A

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\#AB}{\#B}.$$

Example 2.5. We have an urn with 4 red and 6 green balls. We choose 3 balls without replacements. Find the probability of having exactly 2 red balls in the sample given that at least one red ball in the sample.

Solution: Let A and B denotes the events of having

exactly 2 red balls and at least one red ball, respectively. So, we are asked to compute
$$\mathbb{P}(A|B) = \mathbb{P}(AB)/\mathbb{P}(B).$$

$$\mathbb{P}(B) = 1 - \mathbb{P}(\text{no red balls in the sample})$$
$$= 1 - \binom{6}{3}\Big/\binom{10}{3} = 5/6.$$

$$\mathbb{P}(AB) = \mathbb{P}(A) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{3}{10}. \quad \text{So } \mathbb{P}(A|B) = \frac{3/10}{5/6} = \frac{9}{25}.$$

~o~o~

Note that we have $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$. Iterating this we get
$\mathbb{P}(ABC) = \mathbb{P}(AB)\mathbb{P}(C|AB) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|AB)$, provided that they all make sense.

Fact 2.6. (Multiplication Rule for n events). If $A_1, \ldots, A_n$ are events then
$$\mathbb{P}(A_1A_2\cdots A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1A_2)\cdots\mathbb{P}(A_n|A_1A_2\cdots A_{n-1}),$$
provided that they all make sense.

Example 2.7. Suppose an urn contains 8 red and 4 white balls. Draw two balls without replacement. What is the probability that both are red?

Solution: $R_1 = \{\text{first draw is red}\}$
$R_2 = \{\text{second draw is red}\}$.

$$\mathbb{P}(R_1R_2) = \mathbb{P}(R_1)\,\mathbb{P}(R_2|R_1) = \frac{8}{12}\cdot\frac{7}{11} = \frac{14}{33}.$$

Example 2.8. We have two urns. Urn 1 has 2 green and 1 red ball. Urn 2 has 2 red and 3 yellow balls. We perform a two stage experiment. First choose one of the urns with equal probability, then sample one ball uniformly at random from the selected urn. What is the probability that the ball is red?

Solution:
$$\begin{aligned} \mathbb{P}(\text{red}) &= \mathbb{P}(\{\text{red}\} \cap \{\text{urn } 1\}) + \mathbb{P}(\{\text{red}\} \cap \{\text{urn } 2\}) \\ &= \mathbb{P}(\text{red} \mid \text{urn } 1)\, \mathbb{P}(\text{urn } 1) + \mathbb{P}(\text{red} \mid \text{urn } 2)\, \mathbb{P}(\text{urn } 2) \\ &= \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2} = \frac{1}{6} + \frac{1}{5} = \frac{11}{30}. \end{aligned}$$

~o~o~

The general case of the principle used in the above example is
$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(AB) + \mathbb{P}(AB^c) \\ &= \mathbb{P}(A \mid B)\, \mathbb{P}(B) + \mathbb{P}(A \mid B^c)\, \mathbb{P}(B^c) \end{aligned}$$

Definition 2.9. A finite collection of pairwise disjoint events $\{B_1, \dots, B_n\}$ so that $\Omega = B_1 \cup \dots \cup B_n$ is called a partition of Ω.

Fact 2.10. If $B_1, \dots, B_n$ is a partition of Ω, then for any event $A \subseteq \Omega$ we have
$$\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{i=1}^{n} (A \cap B_i)\Big) = \sum_{i=1}^{n} \mathbb{P}(AB_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\, \mathbb{P}(B_i).$$

§2.2. Bayes' Formula

Let's start with an example.

Example 2.12. Consider Example 2.8: We have two urns: urn 1 has 2 green balls and 1 red ball, while urn 2 has 2 red balls and 3 yellow balls. An urn is picked randomly and a ball is drawn from it. Given that the chosen ball is red, what is the probability that the ball came from urn 1?

Solution:
$$\mathbb{P}(\text{ur}1, \text{red}) = \frac{\mathbb{P}(\{\text{red}\}\cap\{\text{urn }1\})}{\mathbb{P}(\text{red})}$$
$$= \frac{\mathbb{P}(\text{red}|\text{urn }1)\,\mathbb{P}(\text{urn }1)}{\mathbb{P}(\text{red}|\text{urn}1)\,\mathbb{P}(\text{urn}1)+\mathbb{P}(\text{red}|\text{urn }2)\,\mathbb{P}(\text{urn}2)}$$
$$= \frac{1/3\cdot 1/2}{1/3\cdot 1/2+2/5\cdot 1/2} = \frac{5}{11}.$$

Fact 2.13. (Bayes' Formula)

If $\mathbb{P}(A), \mathbb{P}(B)$ and $\mathbb{P}(B^c) > 0$, then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\,\mathbb{P}(B)+\mathbb{P}(A|B^c)\,\mathbb{P}(B^c)}, \text{ where}$$

the denominator is obtained as
$$\mathbb{P}(A) = \mathbb{P}(AB) + \mathbb{P}(AB^c)$$
$$= \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c).$$

Example 2.14. Suppose a medical test detects a particular disease 96% of the time, but gives false positive result 2% of the time. Assume that 0.5% of the population carries the disease. If a random person tests positive for the disease, what is the probability that they actually carry the disease.

Solution. Define the events. $D = \{$person has the disease$\}$, $A = \{$person test positive$\}$. Then we are given that

$$\mathbb{P}(A|D) = \frac{96}{100}, \quad \mathbb{P}(A|D^c) = \frac{2}{100} \text{ and } \mathbb{P}(D) = \frac{5}{1000}.$$

So, from the Bayes' Formula,

$$\mathbb{P}(D|A) = \frac{\mathbb{P}(A|D)\,\mathbb{P}(D)}{\mathbb{P}(A|D)\,\mathbb{P}(D) + \mathbb{P}(A|D^c)\,\mathbb{P}(D^c)} = \frac{\frac{96}{100} \times \frac{5}{1000}}{\frac{96}{100} \times \frac{5}{1000} + \frac{2}{100} \times \frac{995}{1000}}$$

$$= \frac{96}{494} \simeq 0.194.$$

Hence only 1 out of 5 positive tests come from the disease.

Now suppose that only the persons believed by the doctors that they have high probability having the disease take the test. For example assume that $\mathbb{P}(D) = 1/2$. Then the same probability becomes

$$\mathbb{P}(D|A) = \frac{\frac{96}{100} \times \frac{1}{2}}{\frac{96}{100} \times \frac{1}{2} + \frac{2}{100} \times \frac{1}{2}} = \frac{96}{98} \simeq 0.980,$$ which

is much bigger than 0.194.

Fact 2.15 (General version of Bayes' Formula)
If $B_1, \ldots, B_n$ is a partition of the sample space Ω such that $\mathbb{P}(B_i) > 0$. Then for any event A with $\mathbb{P}(A) > 0$, and any $k = 1, 2, \ldots, n$,

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(AB_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\,\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}.$$

§2.3. Independence.

An event A is independent of B if $\mathbb{P}(A|B) = \mathbb{P}(A)$. Note that in this case $\mathbb{P}(AB)/\mathbb{P}(B) = \mathbb{P}(A)$ so that $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ and hence, $\mathbb{P}(B|A) = \mathbb{P}(B)$. In other words, if A is independent of B, then B is independent of A.

Definition 2.17. We say that two events A and B are independent if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

Example 2.18. Suppose that we flip a fair coin three times. Let A be the event that we have exactly one tail among the first two coin flips, B the event that we have exactly one tails among the last two coin flips and D the event that we get exactly one tail among the three coin flips.

Show that A and B are independent and the pairs A and D and also B and D aren't independent.

Solution: $A = \{(H,T,H), (H,T,T), (T,H,T), (T,H,H)\}$

$B = \{(H,H,T), (H,T,H), (T,H,T), (T,T,H)\}$

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

$$\mathbb{P}(AB) = 2/8 = \frac{4}{8} \times \frac{4}{8} = \mathbb{P}(A)\,\mathbb{P}(B).$$
$$\mathbb{P}(AD) = \frac{2}{8} \neq \frac{4}{8} \times \frac{3}{8} = \mathbb{P}(A)\,\mathbb{P}(D).$$
$$\mathbb{P}(BD) = 2/8 \neq 4/8 \times 3/8 = \mathbb{P}(B)\,\mathbb{P}(D).$$

__Example 2.19.__ Suppose that we have an urn with 4 red and 7 green balls. We choose two balls with replacement. Let $A=\{\text{first ball is red}\}$ and $B=\{\text{second ball is green}\}$. Is it true that A and B are independent. What if we sample without replacement?

What are your guesses without making any computation?

__Solution for sampling with replacement.__

$\#\Omega = 11^2$, $\#A = 4\times 11$ and $\#B = 7\times 11$. Also $\#AB = 4\times 7$.

So, $P(AB) = \dfrac{4\times 7}{11^2} = \dfrac{4}{11}\times\dfrac{7}{11} = P(A)P(B)$ and hence

A and B are independent.

__Solution for sampling without replacement?__

In this case $\#\Omega = 11\cdot 10$, $\#A = 4\times 10 = 40$,

$\#AB = 4\cdot 7 = 28$ and $\#B = \#(AB) + \#(AB^c) = 4\times 7 + 7\times 6 = 70$

So, $P(AB) = 28/110$ and $P(A)P(B) = \dfrac{40\cdot 70}{110^2} = \dfrac{28}{121}$

so that A and B are __not__ independent.

__Fact 2.20.__ Suppose that A and B are independent. Then the same is true for each of the pairs: A^c and B, A and B^c, and A^c and B^c.

__Proof:__ Let's show only the first one.

$$P(A^cB) = P(B\setminus(A\cap B)) = P(B) - P(A\cap B) = P(B) - P(A)P(B)$$
$$= P(B)(1-P(A)) = P(B)\,P(A^c).$$

__Example 2.2.1.__ Suppose that A and B are independent and $P(A)=1/3$, $P(B)=1/4$. Find the probability that exactly one of the two events is true?

<u>Solution</u>: We are asked to find $P(A^cB \cup AB^c)$. Since the events A^cB and AB^c are disjoint we can write $P(A^cB \cup AB^c) = P(A^cB) + P(AB^c)$. Finally, since the events are independent

$$\begin{aligned} P(A^cB \cup AB^c) &= P(A^cB) + P(AB^c) \\ &= P(A^c)P(B) + P(A)P(B^c) \\ &= \frac{1}{3}\cdot\frac{3}{4} + \frac{2}{3}\frac{1}{4} = \frac{5}{12}. \end{aligned}$$

<u>Definition 2.22</u>. Events $A_1, \dots, A_n$ are independent (or mutually independent) if for every collection $A_{i_1}, \dots, A_{i_k}$ when $2 \le k \le n$, $1 \le i_1 < i_2 < \dots < i_k \le n$,

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

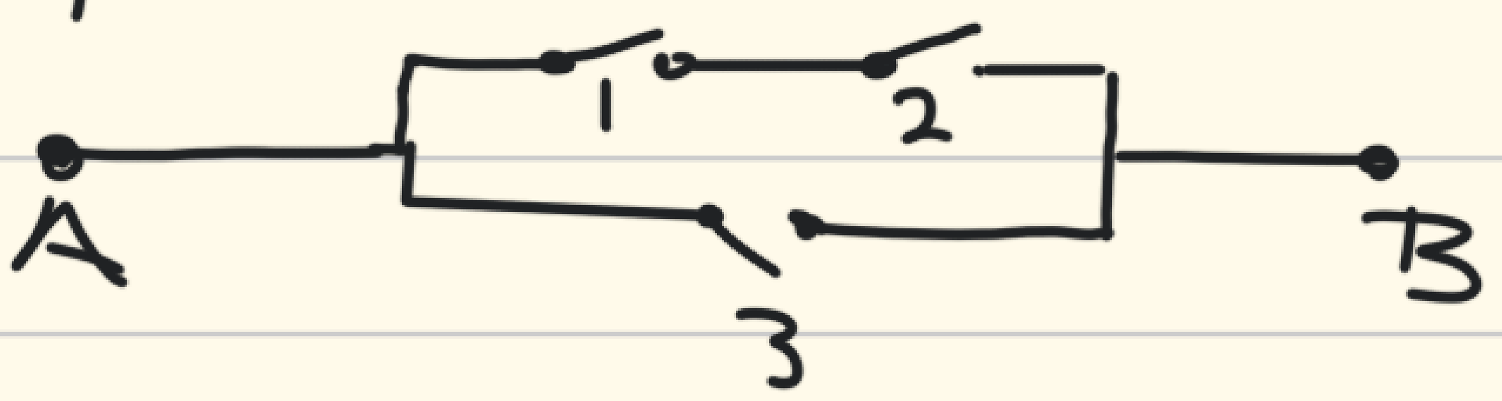
For example for three events A, B, C, they are independent if and only if $P(AB) = P(A)P(B)$, $P(AC) = P(A)P(C)$, $P(BC) = P(B)P(C)$ and $P(ABC) = P(A)P(B)P(C)$.

<u>Fact 2.23</u>. Suppose that $A_1, \dots, A_n$ are mutually independent. Let A^* denotes either A or A^c. Then $A_1^*, \dots, A_n^*$ are mutually independent.

<u>Remark</u>: We say that events $A_1, \dots, A_n$ are pairwise independent if for any $i \neq j$ A_i and A_j are independent.

<u>Example 2.26</u>. Consider the electric network below where the switches are open or

closed independent from each other. Let p_i denote the probability that the i^{th} switch is closed. Find the probability that current can flow from A to B.



<u>Solution:</u> Let C_1 be the event that the current flows from the top branch and C_2 the event that the current flows from the bottom branch. Then $C_1 = S_1 S_2$, $C_2 = S_3$ and we are asked to find $P(D)$ where $D = C_1 \cup C_2$.

$$
\begin{aligned}
\text{So, } P(D) &= P(C_1) + P(C_2) - P(C_1 C_2) \\
&= P(S_1 S_2) + P(S_3) - P(S_1 S_2 S_3) \\
&= P(S_1) P(S_2) + P(S_3) - P(S_1) P(S_2) P(S_3) \\
&= p_1 p_2 + p_3 - p_1 p_2 p_3 .
\end{aligned}
$$

<u>Independence of Random Variables.</u>

<u>Definition 2.27.</u> Let $X_1, \ldots, X_n$ be random variables defined on some probability space Ω. Then $X_1, \ldots, X_n$ are said to independent if

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{k=1}^{n} P(X_k \in B_k), \text{ for all choices } B_1, \ldots, B_k \subseteq \mathbb{R}.$$

In the discrete case this definition reduces to

<u>Fact 2.28.</u> Discrete random variables $X_1, \ldots, X_n$ are independent if and only if

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k), \text{ for all } x_1, \ldots, x_n \in \mathbb{R}.$$

<u>Proof</u>:
$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \sum_{\substack{x_i \in B_i \\ i=1,\ldots,n}} \mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n\}$$
$$= \sum_{\substack{x_i \in B_i \\ i=1,\ldots,n}} \mathbb{P}\{X_1 = x_1\} \cdots \mathbb{P}\{X_n = x_n\}$$
$$= \Big(\sum_{x_1 \in B_1} \mathbb{P}\{X_1 = x_1\}\Big) \cdots \Big(\sum_{x_n \in B_n} \mathbb{P}\{X_n = x_n\}\Big)$$
$$= \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

■

<u>Example 2.29</u>. Assume that we flip three fair coins and let X_i denote the random variable defined by
$$X_i = \begin{cases} 0, & \text{if the } i^{th} \text{ flip is heads} \\ 1, & \text{if the } i^{th} \text{ flip is tails.} \end{cases}$$

Define the events $G_i = \{X_i = 1\}$. Then G_i's are independent:

Clearly, $\mathbb{P}(G_i) = \mathbb{P}(G_i \mid G_j)$ if $i \neq j$.
$$= \mathbb{P}(G_i G_j) / \mathbb{P}(G_j)$$
$$\Rightarrow \mathbb{P}(G_i G_j) = \mathbb{P}(G_i)\mathbb{P}(G_j) \text{ if } i \neq j.$$

Also, $\mathbb{P}(G_1) = \mathbb{P}(G_1 \mid G_2 G_3) = \mathbb{P}(G_1 G_2 G_3) / \mathbb{P}(G_2 G_3)$
$$\Rightarrow \mathbb{P}(G_1 G_2 G_3) = \mathbb{P}(G_1)\,\mathbb{P}(G_2 G_3)$$
$$= \mathbb{P}(G_1)\mathbb{P}(G_2)\mathbb{P}(G_3).$$

Hence, G_1, G_2, G_3 are independent. So, G_1^*, G_2^*, G_3^* are independent, where G_i^* denotes G_i or G_i^c. It follows that the variables X_1, X_2, X_3 are independent since $\{X_i = 1\} = G_i$ and $\{X_i = 0\} = G_i^c$. (This is continuation of Example 2.25, which we skipped!)

Example 2.30. (Continuation of Example 2.19)

Suppose that we choose one by one a sample of size k from the set $\{1, 2, \dots, n\}$. Denote the outcome of the successive draws by $X_1, X_2, \dots, X_k$. Check if the variables are mutually independent in the following cases:

Sampling with Replacement: $\Omega = \{1, \dots, n\}^k$ and $\# \Omega = n^k$.

$$\mathbb{P}(X_j = x) = \frac{n^{k-1}}{n^k} = \frac{1}{n}, \quad \forall x \in \{1, \dots, n\}.$$

Similarly, $\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \frac{1}{n^k} = \prod_{j=1}^{k} \mathbb{P}(X_j = x_j)$.
Hence, $X_1, \dots, X_k$ are mutually independent.

Sampling without Replacement.

Clearly, $\mathbb{P}(X_1 = 1) = \frac{1}{n}$, $\mathbb{P}(X_2 = 1) = \frac{1}{n} \cdot 0 + \frac{n-1}{n} \cdot \frac{1}{n-1}$
$= 1/n$.

Similarly, $\mathbb{P}(X_j = 1) = 1$, $\forall j = 1, \dots, k$.
However, $\mathbb{P}(X_1 = 1, X_2 = 1) = 0 \neq \frac{1}{n^2} = \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 1)$, so that the variables are not mutually independent.

§2.4. Independent Trials.

Bernoulli Distribution. The Bernoulli random variable records the results of independent trials with two possible outcomes.

Definition 2.31. Let $0 \leq p \leq 1$. A random variable X has the Bernoulli distribution with success probability p if X is $\{0,1\}$-valued and satisfies $\mathbb{P}(X=1)=p$ and $\mathbb{P}(X=0)=1-p$. We abbreviate this by $X \sim Ber(p)$.

For instance, $\mathbb{P}(X_1=0, X_2=1, X_3=X_4=0, X_5=1, X_6=0)$ is $p^2(1-p)^4$ (note that X_i's are all independent.

Binomial Distribution. The binomial distribution arises from counting successes. Let S_n be the number of successes in n independent trials with success probability p. If X_i denotes the outcome of the i^{th} trial then $S_n = X_1 + \cdots + X_n$. Clearly, there are $\binom{n}{k}$ vectors of length n with exactly k ones and $n-k$ zeros. Moreover, each outcome with exactly k ones has probability $p^k(1-p)^{n-k}$. Thus $\mathbb{P}(S_n=k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Definition 2.32. The above random variable $X (= S_n)$ is called the binomial distribution with parameters n and p and it is abbreviated by $X \sim Bin(n,p)$.

Note that $\sum_{k=0}^{n} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n = 1$.

Example 2.33. What is the probability that five rolls of a fair die yield two or three sixes?

Solution: $S_5 \sim Bin(5, 1/6)$.

$$\mathbb{P}(S_5 \in \{2,3\}) = \mathbb{P}(S_5 = 2) + \mathbb{P}(S_5 = 3)$$
$$= \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 + \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0.193.$$

Geometric Distribution.

Now we consider an infinite sequence of independent trials. This is necessary if we are willing to make statements such as "when a fair die is rolled repeatedly the long term frequency of fives is 1/6". So imagine an infinite sequence of independent trials with success probability p. Let X_j denote the outcome of the j^{th} trial, with $X_j = 1$ if trial j is a success and $X_j = 0$ if the trial j is a failure. Again let $S_n = X_1 + \dots + X_n$. Then $\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Now we'll define the geometric distribution N as the number of trials needed to see the first success in a sequence of independent trials with success probability p. Then for any $k \in \mathbb{N}$,
$$P(N = k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1-p)^{k-1} p.$$

Definition 2.34. Let $0 < p \leq 1$. The random variable N above is called the geometric distribution with success parameter p. We abbreviate this by $N \sim Geom(p)$.

Example 2.35. What is the probability that it takes more than seven rolls of a fair die to roll a six?

Solution: The answer is clearly $\left(\frac{5}{6}\right)^7$. Now let's compute the probability in an alternative way:

$$\mathbb{P}(N \geq 7) = \sum_{k=8}^{\infty} \mathbb{P}(N=k) = \sum_{k=8}^{\infty} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \left(\frac{5}{6}\right)^7.$$

Example 2.36. Roll a pair of fair dice until you get either a sum of 5 or a sum of 7. What is the probability that you get 5 first?

Solution. Let A be the event that 5 comes first, and let $A_n = \{$no 5 or 7 in rolls $1, \dots, n-1$, and 5 at rool $n\}$.

$\mathbb{P}$(pair of dice gives 5) $= \frac{4}{36}$ and $\mathbb{P}$(pair of dice gives 7) $= \frac{1}{6}$.

By the independence of trials

$$\mathbb{P}(A_n) = \left(1 - \frac{10}{36}\right)^{n-1} 4/36 = \left(\frac{13}{18}\right)^{n-1} \frac{1}{9}$$ and hence

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} \frac{1}{9} = \frac{1/9}{1-\frac{13}{18}} = \frac{2}{5}.$$

Note that $2/5 = \mathbb{P}$(pair of dice gives 5 | pair of dice gives 5 or 7)

$$= \frac{4/36}{\frac{4+6}{36}} \left(= \frac{4}{10} = \frac{2}{5}\right).$$

§2.5. Further Topics on Sampling and Indepence.

Conditional Independence.

Definition 2.37. Let $A_1, \ldots, A_n$ and B be events and $P(B) > 0$. Then $A_1, \ldots, A_n$ are conditionally independent, given B, if the following condition holds: for any $k \in \{2, \ldots, n\}$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$P(A_{i_1} A_{i_2} \cdots A_{i_k} \mid B) = P(A_{i_1} \mid B) P(A_{i_2} \mid B) \cdots P(A_{i_k} \mid B).$$

Example 2.38. Suppose 90% of coins in circulation are fair and 10% are biased coins that give tails with probability 3/5. I have a random coin and I flip it twice. Denote by A_1 the event that the first flip yields tails and by A_2 the event that the second flip yields tails. Are these events independent?

Solution. Let F denote the event the coin is fair and B that it is biased. We are given that $P(F) = 9/10$ and $P(B) = 1/10$. The conditional probabilities are as follows:

$$P(A_1 \mid F) = P(A_2 \mid F) = 1/2 \text{ and } P(A_1 \mid B) = P(A_2 \mid B) = \frac{3}{5}.$$

Above computation assumes that for a given coin the probability of tails does not change between the first and second flip.

Now, $P(A_i) = P(A_i \mid F) P(F) + P(A_i \mid B) P(B) = 1/2 \times 9/10 + \frac{3}{5} \times \frac{1}{10} = \frac{51}{100}.$

To compute the probability of two tails we need to make an assumption: the successive flips of a

coin are independent. So we have:
$P(A_1A_2|F) = P(A_1|F)P(A_2|F)$ and
$P(A_1A_2|B) = P(A_1|B)\,P(A_2|B)$.

Finally, $P(A_1A_2) = P(A_1A_2|F)\,P(F) + P(A_1A_2|B)\,P(B)$
$= P(A_1|F)\,P(A_2|F)\,P(F) + P(A_1|B)\,P(A_2|B)\,P(B)$
$= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{9}{10} + \frac{3}{5}\cdot\frac{3}{5}\cdot\frac{1}{10} = \frac{261}{1000}$.

Clearly, $\frac{261}{1000} \neq \left(\frac{51}{100}\right)^2$ so that the events A_1 and A_2 are not independent.

Example 2.39. Recall Example 2.14 that a medical test which detects a particular disease 96% of the time, but gives a false positive 2% of the time. It was assumed that 0.5% of the population carry the disease. It was computed that a random person tests positive for the disease, the probability that he actually carries the disease is approximately 0.194.

Suppose that after testing positive, the person is retested for the disease and this test also comes back positive. What is the probability of disease after two positive tests?

Solution. Again let $D = \{$person has the disease$\}$, $A_1 = \{$person tests positive on the first test$\}$ and $A_2 = \{$ " " " " " second " $\}$.

So, we are asked to compute $P(D|A_1A_2)$.
Assuming A_1 and A_2 are conditionally independent, given D,

$$\mathbb{P}(D|A_1A_2) = \frac{\mathbb{P}(DA_1A_2)}{\mathbb{P}(A_1A_2)} = \frac{\mathbb{P}(A_1A_2|D)\mathbb{P}(D)}{\mathbb{P}(A_1A_2|D)\mathbb{P}(D) + \mathbb{P}(A_1A_2|D^c)\mathbb{P}(D^c)}$$

$$= \frac{\mathbb{P}(A_1|D)\mathbb{P}(A_2|D)\mathbb{P}(D)}{\mathbb{P}(A_1|D)\mathbb{P}(A_2|D)\mathbb{P}(D) + \mathbb{P}(A_1|D^c)\mathbb{P}(A_2|D^c)\mathbb{P}(D^c)}$$

$$= \frac{0.96 \times 0.96 \times 0.005}{(0.96 \times 0.96 \times 0.005) + (0.02 \times 0.02 \times 0.995)} = \frac{2304}{2503}$$

$$\approx 0.9205$$

<u>Example 2.40.</u> Roll two fair dice. Define two events $A = \{\text{first die gives 1 or 2}\}$ and $B = \{\text{3 appears at least once}\}$. It can be shown easily that A and B are not independent. On the other hand, if D is the event $D = \{\text{the sum of the dice is 5}\}$, then A and B are condotionally independent given D:

Note that $D = \{(1,4), (2,3), (3,2), (4,1)\}$.

So, $\mathbb{P}(A|D) = \frac{\mathbb{P}(AD)}{\mathbb{P}(D)} = \frac{\mathbb{P}\{(1,4),(2,3)\}}{4/36} = \frac{2/36}{4/36} = 1/2$,

$\mathbb{P}(B|D) = \frac{\mathbb{P}(BD)}{\mathbb{P}(D)} = \frac{\mathbb{P}\{(2,3),(3,2)\}}{4/36} = \frac{2/36}{4/36} = 1/2$, and

$\mathbb{P}(AB|D) = \frac{\mathbb{P}(ABD)}{\mathbb{P}(D)} = \frac{\mathbb{P}\{(2,3)\}}{4/36} = \frac{1/36}{4/36} = 1/4$, so

that $\mathbb{P}(AB|D) = \mathbb{P}(A|D)\,\mathbb{P}(B|D)$.

<u>Independence of Events Contructed from Independent Events.</u>

Let A, B and C be independent events. Then the events A and $B^c \cup C$ are independent:

$$\mathbb{P}(A \cap (B^c \cup C)) = \mathbb{P}((A \cap B^c) \cup (A \cap C))$$
$$= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B^c \cap A \cap C)$$

$$= P(A)P(B^c) + P(A)P(C) - P(A)P(B^c)P(C)$$
$$= P(A)\left(P(B^c) + P(C) - P(B^c)P(C)\right)$$
$$= P(A)P(B^c \cup C).$$

This is true in more generality: Let $A_1, \ldots, A_n$ be independent event and let $B_1, \ldots, B_k$ are obtained from A_i's using set operations so that no A_i is used in more than one B_j. Then $B_1, \ldots, B_k$ are also independent.

A similar statement holds for random variables. For example, if X, Y and Z are independent random variables and $U = g(X,Y)$ for some function g then U and Z are independent.

Hypergeometric Distribution.

We'll revisit sampling without replacement. Suppose there are two types of items, type A and type B. Let N_A and N_B denote the number of items of type A and B, respectively. Also let $N = N_A + N_B$. We sample $n \leq N$ items without replacement. Let X denote the number of type A items in this sampling (without replacement with order not mattering).

Definition 2.42. The random variable X above is called the hypergeometric distributions with parameters (N, N_A, n) if X takes values in the set $\{0, 1, \ldots, n\}$ and has probability mass function

$$P(X = k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n.$$

We abbreviate this by $X \sim \text{Hypergeom}(N, N_A, n)$.

Note that if $k > N_A$ then $\mathbb{P}(X=k) = 0$.

Example 2.43. A basket contains a litter of 6 kittens, 2 males and 4 females. A neighbor comes and picks 3 kittens randomly take home with him. Let X denote the number of male kittens in the group the neighbor chose. Then $X \sim \text{Hypergeom}(6,2,3)$. The probability mass function of X is as follows.

$$\mathbb{P}(X=0) = \frac{\binom{2}{0}\binom{4}{3}}{\binom{6}{3}} = \frac{1}{5}, \quad \mathbb{P}(X=1) = \frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}} = \frac{3}{5}$$

$$\mathbb{P}(X=2) = \frac{\binom{2}{2}\binom{4}{2}}{\binom{6}{3}} = \frac{1}{5}, \quad \mathbb{P}(X=3) = 0.$$

Now consider the following experiment. The neighbor's daughter picks up a random kitten from the litter of six, pets it, and puts it back in the basket. She does this 3 consequtive days. Let Y be the number of times she chose a male kitten. This is sampling with replacement. Note that $Y \sim \text{Bin}(3, 2/6)$ and

$$\mathbb{P}(Y=k) = \binom{3}{k}\left(\frac{2}{6}\right)^k \left(\frac{4}{6}\right)^{3-k} \text{ for } k = 0,1,2,3.$$

The Birthday Problem.

Example 2.44. How large should a randomly selected group of people be to guarantee that with probability at least 1/2 there are two people with the same birthday?

<u>Solution</u>. We can restate the problem as follows: Take a random sample of size k with replacement from the set $\{1,2,\dots,365\}$. Let p_k be the probability that there is repetition in the sample. How large should k be to have $p_k > 1/2$?

Let $A_k = \{$the first k picks are distinct$\}$. So $p_k = \mathbb{P}(A_k^c) = 1 - \mathbb{P}(A_k)$. On the other hand,

$$\mathbb{P}(A_k) = \frac{365 \cdot 364 \cdot \ldots \cdot (365-(k-1))}{365^k} \text{ so that } p_k = \frac{1 - \prod_{i=0}^{k-1}(365-i)}{365^k}.$$

$p_{22} \simeq 0.4757$ and $p_{23} \simeq 0.5073$ so that the answer is 23 (people).

<u>Reasoning with Uncertainty in the real world</u>.

<u>Example 2.45</u>. (The Sally Clark Case)
This is a famous wrongful conviction in England. Sally Clark's two boys died as infants without obvious causes. In 1998 she was charged with murder. At the trial an expert witness made the following calculation. Population statistics indicated that there is about a 1 in 8500 chance of an unexplained infant death in

a family like the Clarks. Hence the chance of two deaths is $(1/8500)^2 \approx 1/72000000$ and this was presented as an indication of how unlikely Clark's innocence was.

Several mistakes were made during the process. Two of those relevant to probability theory are the following: (i) The probabilities of the two deaths were multiplied as if they are independent, and (ii) the resulting number was misinterpreted as the probability of Clark's innocence. Now let's study these points in detail.

(i) Suppose a disease appears in 0.1% of the population. Suppose further that this disease comes from a generic mutation passed from father to son with probability 0.5 and that a carrier of the mutation develops the disease with probability 0.5. If the disease strikes completely at random, the answer is 0.001^2, or 1 in a million. However, the illness of the first son implies that the father carries the mutation. Hence, the conditional probability that the second son also falls ill is $0.5 \times 0.5 = 0.25$. Thus the correct answer is 0.001×0.25, which is 1 in 4000, a much larger probability than 1 in a million.

(ii) The second error is the interpretation of the 1 in 72 million figure. Even it were the correct number, it is not the probability of Sally Clark's innocence. This mistake is known as prosecutor's fallacy. Consider the following example. The probabi-

Loty getting 6 numbers out of 40 exactly right is $\binom{40}{6}^{-1} \approx 1$ in 3.8 million, extremely unlikely. Yet there are plenty of lottery winners, and we do not automatically suspect cheating, just because of the low odds. In a large enough population even a low probability event is likely to happen to somebody.

CHATER 3. Random Variables

§ 3.1. Probability distributions of random variables.

Probability Mass Function.

A random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called discrete if there is a finite or countably infinite set $A = \{k_1, k_2, \ldots\} \subseteq \mathbb{R}$ so that $\mathbb{P}(X \in A) = 1$. We say that k is a possible value of X if $\mathbb{P}(X = k) > 0$.

The probability distribution of a discrete random variable is entirely determined by its probability mass function (p.m.f.) $p(k) = \mathbb{P}(X = k)$. The p.m.f. is a function from the set of possible values of X into $[0,1]$. By assumption

$$\sum_{k \in A} p_X(k) = \sum_{k \in A} \mathbb{P}(X = k) = \mathbb{P}(X \in A) = 1.$$

We may extend the domain of p to real numbers not in A by letting $p(x) = 0$ if $x \notin A$.

Probability Density Function.

Definition 3.1. Let X be a random variable. If a function f satisfies $\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(x)\,dx$ for all values of b, then f is the probability density function (p.d.f) of X.

An important fact is that if f is as above then $\mathbb{P}(X \in B) = \int_B f(x)\,dx$, for any subset $B \subseteq \mathbb{R}$ for which integration makes sense.

The set B can be any interval. For example,

$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$, $\mathbb{P}(X > a) = \int_a^{\infty} f(x)\,dx$.

Fact 3.2. If a random variable X has density function f then point values have probability zero:

$$\mathbb{P}(X=c) = \int_c^c f(x)\,dx = 0, \text{ for any } c \in \mathbb{R}.$$

It follows that the random variable X is not discrete and $\mathbb{P}(a < X \leq b) = P(a < X < b) = \mathbb{P}(a \leq X \leq b)$ $= \int_a^b f(x)\,dx$.

Note that if a function is a p.d.f. for some random variable X, then $f \geq 0$ and $\mathbb{P}(-\infty < X < +\infty) = 1$:

$f(x) \geq 0 \ \forall x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. $(*)$

Any function satisfying $(*)$ is called a probability density function.

Example 3.3. For $a, b, c \in \mathbb{R}$, $a > 0$, consider the functions f_1, f_2 and f_3 below:

$$f_1(x) = \begin{cases} 1/x^2, & x \geq 1 \\ 0, & x < 1 \end{cases}, \quad f_2(x) = \begin{cases} b\sqrt{a^2 - x^2}, & |x| \leq a \\ 0, & |x| > a \end{cases},$$

$$f_3(x) = \begin{cases} c \sin x, & x \in [0, 2\pi] \\ 0, & x \notin [0, 2\pi]. \end{cases}$$

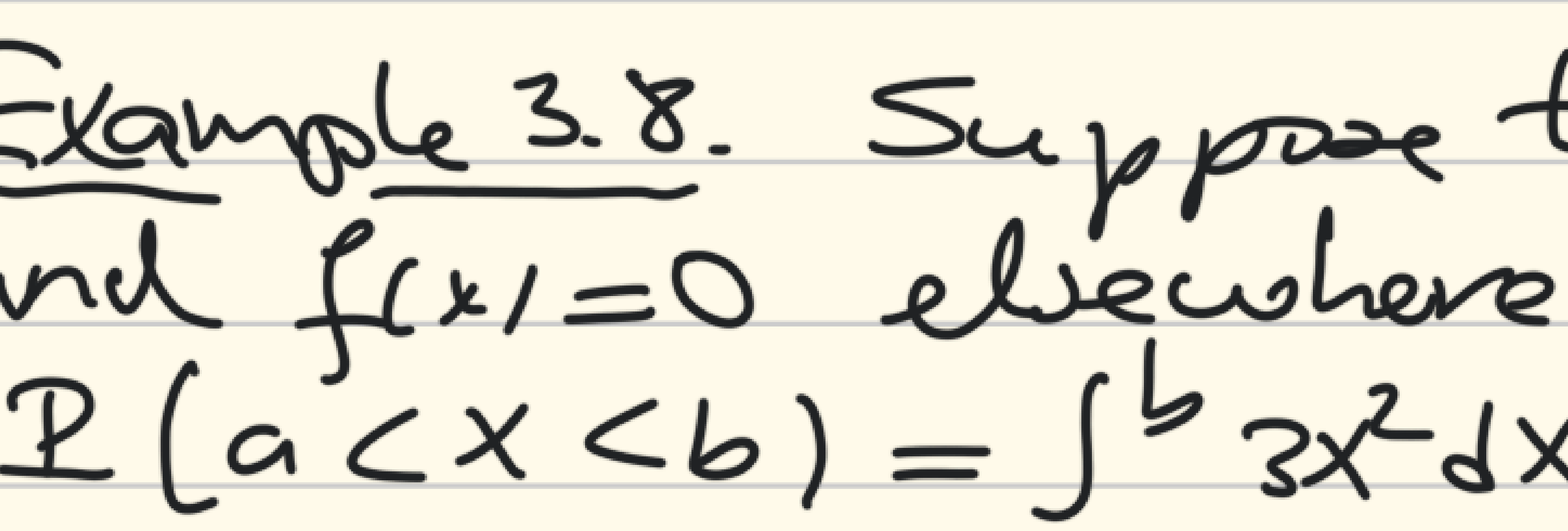
f_1 is a p.d.f, f_2 is a p.d.f. iff $a^2 b \pi = 2$.
f_3 is not a p.d.f. for any $c \in \mathbb{R}$.

Example 3.6. Let Y be a uniform variable on $[-2,5]$. Then the p.d.f. of Y is the constant function $f(x) = 1/7$, $\forall x \in [-2,5]$ and $f(x) = 0$, $\forall x \notin [-2,5]$.

Let's compute
$$\begin{aligned}\mathbb{P}(|Y| \geq 1) &= \mathbb{P}(Y \in [-2,-1] \cup [1,5]) \\ &= \mathbb{P}(Y \in [-2,-1]) + \mathbb{P}(Y \in [1,5]) \\ &= \int_{-2}^{-1} f(x)\,dx + \int_1^5 f(x)\,dx \\ &= \frac{1}{7}(-1-(-2)) + 1/7(5-1) = 5/7.\end{aligned}$$

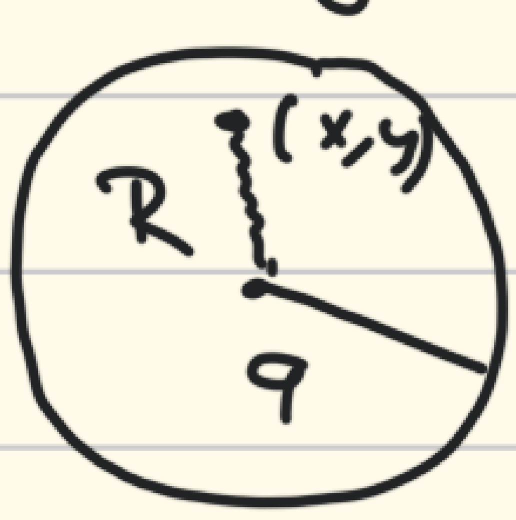
Fact 3.7. Suppose that a random variable X has density function f that is continuous at a point $a \in \mathbb{R}$. Then for small $\epsilon > 0$
$$\mathbb{P}(a < X < a+\epsilon) \simeq f(a)\cdot\epsilon.$$

This follows from $\mathbb{P}(a < X < a+\epsilon) = \int_a^{a+\epsilon} f(x)\,dx \simeq f(a)\cdot\epsilon$, for small $\epsilon > 0$.



Example 3.8. Suppose that $f(x) = 3x^2$, $0 < x < 1$, and $f(x) = 0$ elsewhere. Then
$$\mathbb{P}(a < X < b) = \int_a^b 3x^2\,dx = b^3 - a^3.$$

Example 3.9. Consider the "dart board" example, Example 1.18.



We'll see in the next section that R is a random variable: $R(x,y) = \sqrt{x^2+y^2}$. Consider the p.d.f. of R, f_R.

$$\mathbb{P}(t < R < t+\epsilon) = \frac{(t+\epsilon)^2 - t^2}{9^2} = \frac{2t\epsilon + \epsilon^2}{9^2}.$$

So, $f_R(t) \approx \epsilon^{-1}\,\mathbb{P}(t < R < t+\epsilon) = \dfrac{2t+\epsilon}{9^2}$ and hence

$$f_R(t) = \lim_{\epsilon \to 0} \epsilon^{-1}\,\mathbb{P}(t < R < t+\epsilon) = 2t/9^2, \quad t \in (0,9).$$

§ 3.2. Cumulative Distribution Function.

Definition 3.10. The cumulative distribution function (c.d.f.) of a random variable X is defined by

$$F(s) = \mathbb{P}(X \leq s), \quad \forall s \in \mathbb{R}.$$

Note that the above definition implies that

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a).$$

Indeed these probabilities determine X completely, however proving this fact is beyond the scope of this book.

Cumulative Distribution Function of a Discrete Random Variable.

If X is a discrete random variable then

$$F(s) = \mathbb{P}(X \leq s) = \sum_{k \leq s} \mathbb{P}(X \leq k).$$

Example 3.11. Recall from Definition 2.32 that a random variable $X \sim Bin(2, 2/3)$ is given by

$$X(k) = \binom{2}{k}\left(\frac{2}{3}\right)^k \left(1 - 2/3\right)^{2-k}.$$

So, $P(X=0) = \left(\frac{1}{3}\right)^2$, $P(X=1) = 2 \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{9}$ and $P(X=2) = \frac{4}{9}$.

For example, $F(3/2) = P(X \leq \frac{3}{2}) = P(X=0) + P(X=1) = 5/9$.

Indeed, $F(s) = \begin{cases} 0, & s<0 \\ 1/9 & 0 \leq s < 1 \\ 5/9 & 1 \leq s < 2 \\ 1 & 2 \leq s \end{cases}$.

Cumulative Distribution Function of a Continuous Random Variable.

For a continuous variable random variable X we define $F(s)$ as

$$F(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)\,dx.$$

Fact 3.13. Let the random variable X have c.d.f. F.

(a) Suppose F is piecewise constant. Then X is a discrete random variable. The possible values of X are the locations where F has jumps, and if x is such a point, then $P(X=x)$ equals the magnitude of the jump of F at x.

(b) Suppose F is continuous and the derivative $F'(x)$ exists everywhere on the real line, except possibly at finitely many points. Then X is a continuous random variable and $f(x) = F'(x)$ is the density function of X. If F is not differentiable, then the value $f(x)$ can be set arbitrarily.

(The proof will not be given.)

Example 3.14. Suppose X has the c.d.f. below

$$F(x)=\begin{cases} 0, & x<1 \\ 1/7, & 1\le x<2 \\ 4/7 & 2\le x<5 \\ 5/7 & 5\le x<8 \\ 1 & 1\le x \end{cases}$$

F has jumps at $\{1,2,5,8\}$.

So, $p(1)=\mathbb{P}(X=1)=1/7$, $p(2)=3/7$, $p(5)=1/7$ and $p(8)=2/7$.

Example 3.15. Suppose $F(s)=\begin{cases} 0, & s<1 \\ \frac{s-1}{2}, & 1\le s\le 3 \\ 1, & 3<s. \end{cases}$

Then $F'(s)=\begin{cases} 0, & s<1, s>3 \\ 1/2, & 1<s<3 \end{cases}$.

Hence, the density function $f(s)=\begin{cases} 0, & s\le 1 \text{ or } s\ge 3 \\ 1/2, & 1<s<3. \end{cases}$

Here since F is not differentiable at $s=1$ or $s=3$, we can set the values $f(1)$ and $f(3)$ to be whatever we want. Here we chose zero.

General Properties of Cumulative Distribution Functions.

Fact 3.16. Every cumulative distribution function F has the following properties.

(i) Monotonicity: $s\le t \Rightarrow F(s)\le F(t)$.

(ii) Right Continuity: For any $t\in\mathbb{R}$, $F(t)=\lim_{s\to t^+} F(s)$.

(iii) $\lim_{t\to-\infty} F(t)=0$ and $\lim_{t\to+\infty} F(t)=1$.

By the definition of c.d.f F of a random variable X we have $F(a)=P(X\le a)$.

Writing $(-\infty,a)=\bigcup_{s<a}(-\infty,s]=\bigcup_{n=1}^{\infty}(-\infty,a_n]$, where (a_n) is any sequence with $a_n\uparrow$ and $\lim a_n=a$,

and using the monotonicity of the probability (Fact 1.39) we see that

$$F(-\infty,a) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}(-\infty,a_n]\right) = \lim_{n\to\infty}\mathbb{P}(-\infty,a_n]$$

$$= \lim_{n\to\infty}F(a_n).$$

It follows that $F(-\infty,a) = \lim_{s\to a^-} F(s) = F(a^-)$

and
$$\mathbb{P}(X=a) = \mathbb{P}(X\leq a) - \mathbb{P}(X<a)$$
$$= F(a) - F(a^-)$$
$$= \text{size of the jump in } F \text{ at point } a.$$

Further Examples

Example 3.18. Recall the density function from Example 3.3, with constant $a>0$:

$$f_1(x) = \begin{cases} 1/x^2, & x\geq 1 \\ 0, & x<1 \end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} \frac{2}{\pi a^2}\sqrt{a^2-x^2}, & |x|\leq a \\ 0, & |x|>a. \end{cases}$$

Now, $$F_1(s) = \int_{-\infty}^{s} f_1(x)\,dx = \begin{cases} 0 & \text{if } s<1, \\ \int_1^s \frac{dx}{x^2} = 1-\frac{1}{s}, & s\geq 1. \end{cases}$$

Similarly, for $F_2(s)$, first note that
$$\int\sqrt{1-x^2}\,dx = \tfrac{1}{2}\arcsin x + \frac{x}{2}\sqrt{1-x^2} + C.$$

Now for $-a\leq s\leq a$, $$F_2(s) = \int_{-a}^{s}\frac{2}{\pi a^2}\sqrt{a^2-x^2}\,dx$$

$(x = ay)$ $$= \int_{-1}^{s/a}\frac{2}{\pi}\sqrt{1-y^2}\,dx$$

$$= \frac{1}{\pi}\arcsin\frac{s}{a} + \frac{s}{\pi a}\sqrt{1-\left(\frac{s}{a}\right)^2} + 1/2$$

In particular, $F(-a)=0$ and $F(a)=1$. Hence,

$$F_2(s)=\begin{cases} 0, & s<-a \\ 1/\pi \arcsin \frac{s}{a}+\frac{s}{\pi a}\sqrt{1-(s/a)^2}+1/2, & -a\leq s<a \\ 1, & s\geq a. \end{cases}$$

Example 3.19. Consider the dartboard example (Example 1.18 and 3.9) The probability to hit a point in a circle of radius $t\leq R$, where R is the radius of the board, is

$$F_R(t)=\mathbb{P}(R\leq t)=\mathbb{P}\{\text{the dart landed on the disk of radius } t\}$$
$$=\frac{\pi t^2}{\pi R^2}=t^2/R^2, \quad 0\leq t\leq R.$$

Hence, $f_R(t)=\begin{cases} 2t/R^2, & 0\leq t<R \\ 0, & t<0, t\geq R. \end{cases}$

§3.3. Expectation.

For a random variable X we'll define three associated concepts, its expectation (also called mean), variance and moments.

Expectation of a discrete Random Variable

Definition 3.21. The expectation or mean of a discrete random variable X is defined by $\mathbb{E}[X]=\sum_k k\,\mathbb{P}(X=k)$, where the sum over all the possible values of X.

Notation. The expectation of X, also called the first moment, is denoted by $\mu=\mathbb{E}[X]$.

Example 3.22. Let Z denote the number from the roll of a fair die. Then $E[Z] = 1\cdot\frac{1}{6} + \cdots + 6\cdot\frac{1}{6} = 7/2$.

<u>Example 3.23.</u> Let $X \sim Bin(n,p)$. Recall from Definition 2.32 that
$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k=0,1,\dots,n.$$

Now, let's determine $E[X]$.
$$E[X] = \sum_{k=0}^{n} k\,\mathbb{P}(X=k) = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$
$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$
$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$
$$= np\,(p + (1-p))^{n-1} = np.$$

So, we've computed the mean of the Binomial random variable $X \sim Bin(n,p)$ as np.

<u>Example 3.24.</u> (Mean of the Bernoulli random variable) For $0 \le p \le 1$, let $X \sim Ber(p)$ (Defn. 2.31, $\mathbb{P}(X=1)=p$, $\mathbb{P}(X=0) = 1-p$). The expectation is then
$$E[X] = 0\cdot\mathbb{P}(X=0) + 1\cdot\mathbb{P}(X=1) = p.$$

~o~o~

Let A be an event on a sample space Ω. Its indicator random variable is defined as
$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}.$$
Then $E[\mathbb{1}_A] = 1\cdot\mathbb{P}(A) + 0\cdot\mathbb{P}(\Omega\setminus A) = \mathbb{P}(A)$.

<u>Example 3.25.</u> (Mean of a geometric random variable) Let $0 < p < 1$ and $X \sim Geom(p)$. Recall from

Defn. 2.34 that p.m.f. of X is $P(X=k)=pq^{k-1}$, $k=1,2,3,\dots$, where $q=1-p$. Then

$$E[X]=\sum_{k=1}^{\infty} kpq^{k-1}=\frac{1}{p}.$$

Expectation of a continuous random variable.

Definition 3.26. The expectation or mean of a continuous random variable X with density function f is

$$E[X]=\int_{-\infty}^{+\infty} x f(x)\,dx.$$

Example 3.27. (Mean of a Uniform Random Variable)
Let $X\sim \mathrm{Unif}[a,b]$. Then

$$E[X]=\int_{-\infty}^{\infty} x f(x)\,dx=\int_a^b x\frac{1}{b-a}\,dx=\frac{a+b}{2}.$$

Example 3.28. (Continuation of Example 3.19)
Recall that $f_R(t)=\begin{cases}2t/R^2 & 0\le t\le R\\ 0, & t<0 \text{ or } t>R.\end{cases}$

Then $E[f_R]=\int_{-\infty}^{+\infty} t f_R(t)\,dt=\int_0^R t\frac{2t}{R^2}\,dt=\frac{2R}{3}.$

Infinite and Nonexistent Expectation

One has to be careful about expectations of random variables which are not finite and discrete, since they may not be even defined.

Example 3.29. (Discrete Example of an infinite expectation).

Consider the following gamble. You flip a fair coin. If it comes up heads, you win 2 \$ and the game is over. Otherwise, the prize is doubled and you flip again. Let Y be the amount the prize you win. Find $E[Y]$.

Note that $P(Y=2^n) = 2^{-n}$ for positive integer n. Then $E[Y] = \sum_{n=1}^{\infty} 2^n 2^{-n} = \sum_{n=1}^{\infty} 1 = +\infty$.

This is known as the St. Petersburg paradox.

<u>Example 3.30.</u> Let X have the density function

$$f(x) = \begin{cases} x^{-2}, & x \geq 1 \\ 0, & x < 1 \end{cases}.$$

Then $E[X] = \int_{-\infty}^{+\infty} x \, P(X=x)\, dx = \int_{-\infty}^{+\infty} x \, f(x)\, dx$

$$= \int_1^{\infty} \frac{1}{x}\, dx = \infty.$$

<u>Expectation of a function of a Random Variable.</u>

If X is a random variable on a sample space Ω and g is a real valued function defined on the range $X(\Omega)$ then $g(X)$ is also a random variable on Ω. So, we may talk about $E[g(X)]$.

<u>Example 3.32.</u> Consider a fair die and W is the random variable on Ω given by

$$W = \begin{cases} -1, & \text{roll is } 1, 2 \text{ or } 3 \\ 1, & \text{roll is } 4 \\ 3, & \text{roll is } 5, 6. \end{cases}$$

Let X denote the outcome of the die roll. Then $W = g(X)$ where $g(1) = g(2) = g(3) = -1$,

$g(4)=1$ and $g(5)=g(6)=3$.

In particular, $E[g(X)]=E[W]=(-1)P(W=-1)$
$+1\cdot P(W=1)+3P(W=3)$

$$\Rightarrow E[g(X)]=(-1)\cdot\frac{1}{2}+1\cdot\frac{1}{6}+3\cdot\frac{1}{3}=\frac{2}{3}.$$

Alternatively, this could be computed using the following fact.

Fact 3.33. Let X and g be as above. If X is a discrete random variable then
$$E[g(X)]=\sum_k g(k)P(X=k),$$
while if X is continuous with density function f then
$$E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)\,dx.$$

Proof: $$E[g(X)]=\sum_y yP(g(X)=y)=\sum_y y\sum_{k:g(k)=y}P(X=k)$$
$$=\sum_y\sum_{k:g(k)=y} yP(X=k)$$
$$=\sum_y\sum_{k:g(k)=y} g(k)P(X=k)$$
$$=\sum_k g(k)P(X=k).$$

Example 3.34. A stick of length l is broken at a uniformly chosen random location. What is the expected length of the longer piece?

Let X denote the position where the stick is broken. Then $X \sim \text{Unif}[0, l]$. X has the density function $f(x) = 1/l$. Let $g(X)$ be the length of the longer piece. Then

$$g(X) = \begin{cases} l - x & \text{if } 0 \leq x \leq l/2 \\ x & \text{if } l/2 < x \leq l. \end{cases}$$

Then $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x)\, dx = \int_0^{l/2} \frac{l-x}{l}\, dx + \int_{l/2}^{l} \frac{x}{l}\, dx$

$= \frac{3l}{4}.$

Fact 3.35. The n^{th} moment of the random variable X is the expectation $E[X^n]$. In the discrete and continuous case it is given by

$$E[X^n] = \sum_k k^n P(X=k) \quad \text{and}$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx, \text{ respectively, where}$$

f is the density function of X.

The second moment, $E[X^2]$, is also called the mean square.

Example 3.37. Let Y be a uniformly chosen random integer from $\{1, 2, \ldots, m\}$. Find the first and second moments of Y.

$$E[Y] = \sum_{k=1}^{m} k P(Y=k) = \sum_{k=1}^{m} k/m+1 = \frac{m(m+1)}{2(m+1)} = \frac{m}{2}.$$

$$E[Y^2] = \sum_{k=1}^{\infty} k^2 P(Y=k) = \sum_{k=1}^{m} k^2/m+1 = \frac{m(m+1)(2m+1)}{6(m+1)} = \frac{m(2m+1)}{6}.$$

Median and Quantiles.

Definition 3.39. The median of a random variable X is any real value m that satisfies
$$P(X \geq m) \geq 1/2 \text{ and } P(X \leq m) \geq 1/2.$$

Example 3.40. Let X be uniformly distributed on the $\Omega = \{-100, 1, 2, 3, \ldots, 9\}$, so X has p.m.f. $P_X(\omega) = 1/10$, for $\omega \in \Omega$. $E[X] = -5.5$, while the median is any number between 4 and 5.

Example 3.41. This example show that when the c.d.f. is continuous and strictly increasing on the range of the random variable, the median is unique.

Consider the dartboard example again. We know that the cumulative distribution function of R (R being the distance between the dart and the center of the board) is
$$F_R(t) = \begin{cases} t^2/9^2 & 0 \leq t \leq 9 \\ 0 & t < 0 \text{ or } t > 9. \end{cases}$$

Then the equation $F_R(m) = 1/2$ has a unique solution: $m^2/9^2 = 1/2 \Rightarrow m = 9/\sqrt{2} \approx 6.36$, the median.

Note also that $E[R] = \frac{2}{3} \cdot 9 = 6$.

The following generalizes the median.

Definition 3.43. For $0<p<1$, the pth quantile of a random variable X is any real number x satisfying
$$P(X\geq x)\geq 1-p \quad \text{and} \quad P(X\leq x)\geq p.$$

§3.4. Variance

Definition 3.44. Let X be a random variable with mean μ. The variance of X is defined by
$$\mathrm{Var}[X]=E[(X-\mu)^2].$$
An alternative symbol is $\sigma^2=\mathrm{Var}[X]$.

$\sigma=\sqrt{\mathrm{Var}[X]}$ is called the standard deviation.

Fact 3.45. Let X be the random variable with mean μ. Then $\mathrm{Var}[X]=\sum_k (k-\mu)^2 P(X=k)$ if X is discrete and
$$\mathrm{Var}[X]=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx,$$ if X has the density function f.

Example 3.47. (Variance of a Bernoulli and indicator random variable)
Let $0\leq p\leq 1$. Recall that for $X\sim \mathrm{Ber}(p)$ p.d.f. $P(X=1)=p$ and $P(X=0)=1-p$ and expectation $E[X]=p$. Hence, its variance is
$$\mathrm{Var}[X]=E[(X-p)^2]=(1-p)^2P(X=1)+(0-p)^2P(X=0)$$
$$=(1-p)^2p+p^2(1-p)=p(1-p).$$

For the indication function $I_A(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \omega\notin A\end{cases}$, we have $E[I_A] = 1 \cdot P(A) = P(A)$ and

$$Var[I_A] = E[(I_A - \mu)^2] = E[(I_A - P(A))^2]$$
$$= (1 - P(A))^2 P(A) + (0 - P(A))^2 (1 - P(A)).$$
$$= P(A)(1 - P(A))$$
$$= P(A) P(A^c).$$

<u>Fact 3.48.</u> (Alternative Formula for $Var[X]$.

Note that we have

$$Var[X] = \sum_k (k^2 - 2\mu k + \mu^2) P(X=k)$$
$$= \sum_k k^2 P(X=k) - 2\mu \sum_k k P(X=k) + \mu^2 \sum_k P(X=k)$$
$$= E[X^2] - 2E[X]^2 + E[X]^2 \cdot 1$$
$$= E[X^2] - E[X]^2.$$

<u>Example 3.49.</u> (Varience of a binomial random variable)

Let $X \sim Bin(n, p)$. We have computed that $E[X] = np$. Similarly, we may compute $E[X^2]$:

$$E[X^2] = \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$
$$= \sum_{k=1}^{\infty} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$
$$= \sum_{k=2}^{n} (k-1) \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$
$$= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$
$$= n(n-1)p^2 \cdot 1 + np.$$

So, $Var[X] = E[X^2] - E[X]^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$

Fact 3.52. Let X be a random variable and $a, b \in \mathbb{R}$. Then $E[aX+b] = aE[X]+b$, and
$$Var[aX+b] = a^2 Var[X].$$

Proof: Let f be the p.d.f. of X. Then,
$$E[aX+b] = \int_{-\infty}^{+\infty} (ax+b) f(x)\,dx = a\int_{-\infty}^{+\infty} x f(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx$$
$$= a\,E[X] + b.$$

Also,
$$\begin{aligned}
Var[aX+b] &= E\left[(aX+b - E[aX+b])^2\right] \\
&= E\left[(aX - aE[X])^2\right] \\
&= E\left[a^2X^2 + a^2E[X]^2 - 2a^2E[X]X\right] \\
&= a^2E[X^2] + a^2E[X]^2 - 2a^2E[X]^2 \\
&= a^2\left(E[X^2] - E[X]^2\right) \\
&= a^2\,Var[X]. \quad \blacksquare
\end{aligned}$$

Example 3.53. Let $Z \sim Bin(10, 1/5)$. Find $E[3Z+2]$ and $Var[3Z+2]$.

From Examples 3.23 and 3.49 we have
$E[Z] = 10 \cdot \frac{1}{5} = 2$ and $Var[Z] = 10 \cdot \frac{1}{5} = \frac{8}{5}$.

So, $E[3Z+2] = 3 \cdot 2 + 2 = 8$ and

$$Var[3Z+2] = 9\,Var[Z] = \frac{72}{5}.$$

Fact 3.54. For a random variable X, $Var[X] = 0$ if and only if $P(X=a) = 1$ for some real number $a \in \mathbb{R}$.

Proof: Suppose first that $\mathbb{P}(X=a)=1$, then X is discrete and $E[X]=a$ and
$Var[X] = E[(X-a)^2] = (a-a)^2\, \mathbb{P}(X=a) = 0$.

For the other direction, we'll assume that X is discrete. Suppose $\mu = E[X]$. Then
$0 = Var[X] = \sum_k (k-\mu)^2\, \mathbb{P}(X=k)$, which

implies that $(k-\mu)^2\, \mathbb{P}(X=k) = 0$, $\forall k$. So if $k \neq \mu$ then $\mathbb{P}(X=k)=0$. So, we see that $\mathbb{P}(X=\mu) = 1$. This finishes the proof. ∎

§ 3.5. Gaussian Distribution.

Definition 3.55. A random variable Z has the standard normal distribution (also called standard Gaussian distribution) if Z has the density function
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$ on the real line.

We abbreviate this by $Z \sim \mathcal{N}(0,1)$.

Its cumulative distribution function has also a special notation
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2}\, ds, \quad x \in \mathbb{R}.$$

Fact 3.56. $\int_{-\infty}^{+\infty} e^{-s^2/2}\, ds = \sqrt{2\pi}$.

Example 3.57. Let $Z \sim N(0,1)$. Find $\mathbb{P}(-1 \leq Z \leq 1.5)$ approximately.

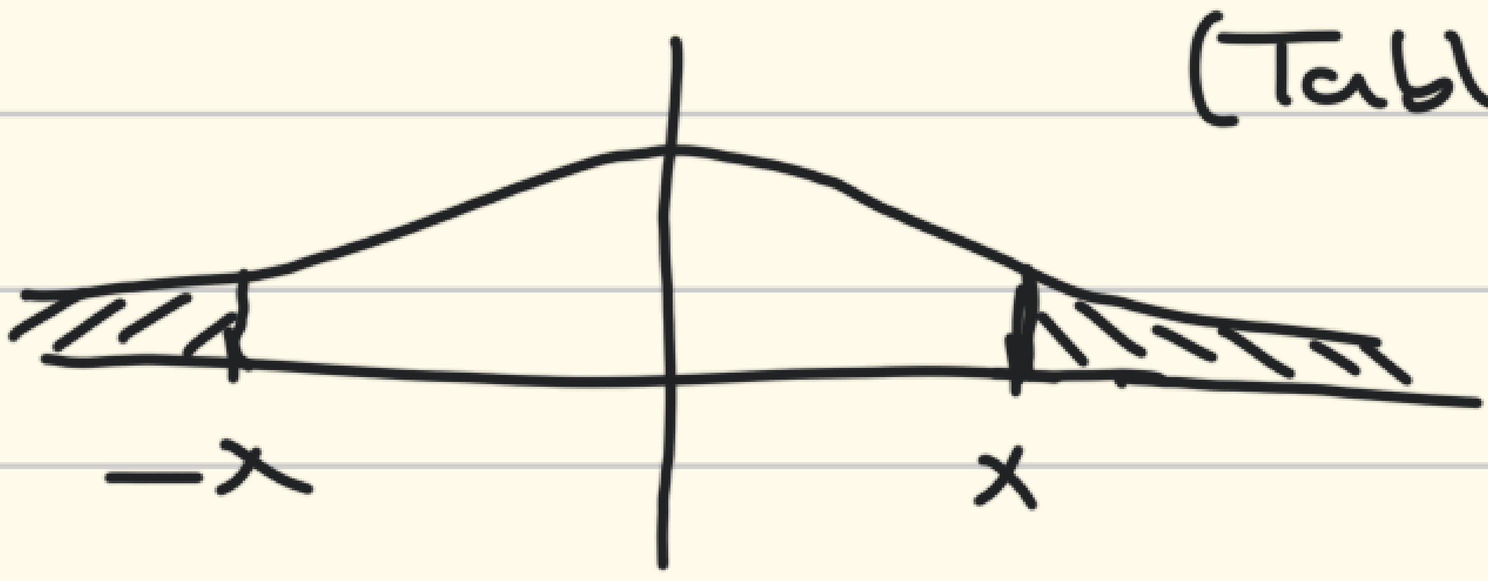
$$\mathbb{P}(-1 \leq Z \leq 1.5) = \int_{-1}^{1.5} \varphi(s)\,ds = \int_{-\infty}^{1.5} \varphi(s)\,ds - \int_{-\infty}^{-1} \varphi(s)\,ds$$

$$= \Phi(1.5) - \Phi(-1)$$

$$= \Phi(1.5) - (1 - \Phi(1))$$

$$\approx 0.9332 - (1 - 0.8413) = 0.7745$$

(Table in Appendix E.)



$\Phi(-x) = 1 - \Phi(x)$, for all $x \in \mathbb{R}$.

<u>Fact 3.59.</u> Let $Z \sim \mathcal{N}(0,1)$. Then $E(Z) = 0$ and
$$Var(Z) = E(Z^2) = 1.$$

<u>Proof</u> $E(Z) = \int_{-\infty}^{+\infty} x\, e^{-x^2/2}\,dx = 0$ since the integrant is odd.

For $Var(Z) = E(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx$ (Use integration by parts)

$$= \frac{-1}{\sqrt{2\pi}} \left\{ x e^{-x^2/2} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-x^2/2}\,dx \right\}$$

$$= \frac{-1}{\sqrt{2\pi}} \left\{ 0 - \sqrt{2\pi} \right\} = 1.$$

$E(Z) = 0$ and $Var(Z) = 1$ explains the notation $Z \sim Nor(0,1)$. Similarly, for any $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$ the variable $X = \sigma Z + \mu$ satisfies $E(X) = \mu$ and $Var(Z) = \sigma^2$ (by Fact 3.52).

Note that the c.d.f and p.d.f. of X are given by $F(x) = \mathbb{P}(X \leq x) = P(\sigma Z + \mu) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

and $f(x) = F'(x) = \frac{d}{dx} \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) \Rightarrow$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

We'll write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Fact 3.61. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$. Then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. In particular, $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable.

Example 3.63. From the table from Appendix E,

$\mathbb{P}(|X-\mu| > 2\sigma) \approx 0.0456$ so that we may say that "A normal random variable is within two standard deviations of its mean with probability over 95%.

CHAPTER 4. Approximations of Binomial Distribution.

Recall that a binomial distribution $X \sim \mathrm{Bin}(n,p)$ has mean np and varience $np(1-p)$. Let S_n denote its probability mass function. Then $(X-np)/\sqrt{np(1-p)}$ has mean zero and varience 1.

It turns out that this variable converges to the normal distribution of type $\mathcal{N}(0,1)$ as n gets large. The below fact describes this in terms of probability mass functions / probability density functions, whose proof will be given later.

Theorem 4.1. (Central Limit Theorem for Binomial Random Variables)

Let $0<p<1$ be fixed and suppose that $S_n \sim \mathrm{Bin}(n,p)$. Then for any fixed $-\infty \le a \le b \le \infty$ we have the limit

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

Taking $a=-\infty$, $\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n-np}{\sqrt{np(1-p)}} \le b\right) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \Phi(b)$.

In particular, $\mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) \approx \Phi(b) - \Phi(a)$, as $n \to \infty$.

It turns out that approximation is "good" if $np(1-p) > 10$.

Example 4.2. A fair coin is flipped 10,000 times. Estimate the probability that the number of heads is between 4850 and 5100.

Solution: Let S be the probability mass function, $S \sim Bin(10{,}000, \frac{1}{2})$, with $E(S) = 10{,}000 \times \frac{1}{2} = 5{,}000$ and $Var(S) = 10{,}000 \times \frac{1}{2^2} = 2{,}500$.

Now, $$P(4850 \leq S \leq 5100) = P\left(\frac{4850-5000}{\sqrt{2500}} \leq \frac{S-5000}{\sqrt{2500}} \leq \frac{5100-5000}{\sqrt{2500}}\right)$$
$$= P\left(-3 \leq \frac{S-5000}{50} \leq 2\right).$$

So by the Central Limit Theorem (Theorem 4.1) this is $\approx \Phi(2) - \Phi(-3) \approx 0.9759$.

On the other hand, the actual probability is $P(4850 \leq S \leq 5100) = \sum_{k=4850}^{5100} \binom{10{,}000}{k} 2^{-k} \approx 0.9765$ which is quite close to the approximate value.

Example 4.3. Suppose a game piece moves along a board according to the following rule: If a roll of a die 1 or 2 then take two steps ahead, otherwise take 3 steps. Approximate the probability that after 120 rolls the piece has moved 315 steps.

Solution. Let X_n denote the number of steps taken up to and including the n^{th} die roll. If S_n is the total number of ones and twos among the first n roll, then $X_n = 2S_n + 3(n - S_n) = 3n - S_n$. Using this we can write the probability we need to estimate
$$P(X_{120} > 315) = P(3 \times 120 - S_{120} > 315) = P(S_{120} < 45).$$

We know that $S_{120} \sim Bin(120, 1/3)$ (note that $1/3$ is the probability p that the roll comes 1 or 2).

Also note that $E(S_{120}) = np = 120 \cdot \frac{1}{3} = 40$ and $Var(S_{120}) = np(1-p) = 120 \times 1/3 \times 2/3 = \frac{80}{3}$.

So, $\mathbb{P}(S_{120} < 45) = \mathbb{P}\left(\frac{S_{120}-40}{\sqrt{80/3}} < \frac{5}{\sqrt{80/3}}\right) \approx \Phi\left(\frac{5}{\sqrt{80/3}}\right)$

$\approx 0.834.$

<u>Example 4.4.</u> (The three sigma rule)

The probability that a standard normal random variable lies in $[-3,3]$ is $\Phi(3) - \Phi(-3) \approx 0.9974$. So by the normal approximation of a binomial, this means that with probability 0.997 a $Bin(n,p)$ random variable is within $3\sqrt{np(1-p)}$ of its mean np, provided n is large enough and p is not too close to 0 or 1. $3\sqrt{np(1-p)} = 3\sigma$ explain why this is called the three sigma rule.

<u>Continuity Correction.</u>

Since $S_n \sim Bin(n,p)$ can take only integer values, if k_1 and k_2 are integers then one has $\mathbb{P}(k_1 \leq S_n \leq k_n) = \mathbb{P}(k_1 - 1/2 \leq S_n \leq k_2 + 1/2)$. It turns out this gives slightly a better approximation of the exact binomial probability. In other words,

$$\Phi\left(\frac{k_2 + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 1/2 - np}{\sqrt{np(1-p)}}\right) \text{ gives a better}$$

estimate for $\mathbb{P}(k_1 \leq S_n \leq k_2)$ than the formula without the $1/2$ terms. Switching from $[k_1, k_2]$

to $[k_1 - 1/2, k_2 + 1/2]$ is called the continuity correction.

<u>Example 4.5.</u> Roll a fair die 720 times. Estimate the probability that we have exactly 113 sixes ($\mu = 120$, $\sigma = 10$)

<u>Solution:</u> To estimate $P(S=113)$ we write it as

$$P(S=113) = P(112.5 \leq S \leq 113.5)$$

$$\approx \Phi\left(\frac{113.5-120}{10}\right) - \Phi\left(\frac{112.5-120}{10}\right)$$

$$= \Phi(-0.65) - \Phi(-0.75) \approx 0.0312.$$

On the other hand, the exact probability is

$P(S=113) \approx \binom{720}{113} \frac{5^{607}}{6^{720}} \approx 0.0318$, which is 0.2% close to the estimation.

On the other hand, a mindless computation would give $P(S=113) = P\left(\frac{113-120}{10} \leq \frac{S-120}{10} \leq \frac{113-120}{10}\right)$

$$= P\left(-0.7 \leq \frac{S-120}{10} \leq -0.7\right)$$

$$\approx \Phi(-0.7) - \Phi(-0.7) = 0.$$

Example 4.6. shows that when the number of trials is large, the continuity correction does not make much of a difference.

<u>A partial Proof of the CLT for the Binomial.</u>

Let q denote $1-p$. Then the L.H.S. of the equation in the statement of the theorem becomes

$$\mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \mathbb{P}(np + a\sqrt{npq} \le S_n \le np + b\sqrt{npq})$$

$$= \sum_{np + a\sqrt{npq} \le k \le np + b\sqrt{npq}} n!/(n-k)!\,k! \; p^k q^{n-k}, \text{ where } k \in \mathbb{Z}.$$

In what follows $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

Next we use so called Stirling's Formula to get an estimation for $n!$.

Fact 4.7. (Stirling's Formula)
As $n \to \infty$, $n! \sim n^n e^{-n} \sqrt{2\pi n}$.

(For a proof see Exercise D.13 and 4.57.)

Now, $n!/(n-k)!\,k! \; p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$.

This implies, $\mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) \approx \sum_{np + a\sqrt{npq} \le k \le np + b\sqrt{npq}} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$ (*)

The sum above can be viewed as a Riemann sum for an integral over $[a,b]$ by defining partition points $x_k = (k-np)/\sqrt{npq}$. The limits of the summation specify exactly $a \le x_k \le b$ and the length of the subintervals is $\Delta x = 1/\sqrt{npq}$. The approximation above becomes

$$\mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) \approx \sum_{k:\, a \le x_k \le b} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} \Delta x \xrightarrow[n\to\infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

One has to study the error in (*) to make this proof rigorous.

§4.2. Law of Large Numbers.

For $0<p<1$ and let $X_1, X_2, \dots, X_n, \dots$ be the outcomes of independent repetitions of a trial with success probability p. So, $P(X_i=1)=p$ and $P(X_i=0)=1-p$. Let $S_n = X_1+X_2+\dots+X_n$, the number of successes in the first n trials.

Theorem 4.8. (Law of Large Numbers for Binomial random Variables)
Assume the above setup. For any $\epsilon>0$ we have $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-p\right|<\epsilon\right)=1$.

Before we give a proof of the theorem let's consider two examples

Example 4.9. Let S_n denote the number of sixes in n rolls of a fair die. Then S_n/n is the observed frequence of sixes in n rolls. Let $\epsilon = 0.0001$. Then by the above theorem
$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-1/6\right|<0.0001\right)=1.$$

Example 4.10. Show that the probability that fair coin flips yield 51% or more tails converge to zero as the number of flips tends to infinity.

$P(\text{at least } 51\% \text{ tails in } n \text{ flips}) = P\left(\frac{S_n}{n}\geq 0.51\right)$
$= P(S_n/n - 0.5 \geq 0.01) \leq P\left(\left|\frac{S_n}{n}-0.5\right|\geq 0.01\right) \to 0$ as $n\to\infty$

<u>Proof of Teorem 4.8.</u> Fix any $\epsilon > 0$ and some big constant $C > 0$. Consider n large enough so that $\epsilon\sqrt{n}/\sigma > C$. Then

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) = \mathbb{P}(-n\epsilon < S_n - np < n\epsilon)$$
$$= \mathbb{P}\left(-\frac{\epsilon\sqrt{n}}{\sigma} < \frac{S_n - np}{\sigma\sqrt{n}} < \frac{\epsilon\sqrt{n}}{\sigma}\right)$$
$$\geq \mathbb{P}\left(-C < \frac{S_n - np}{\sigma\sqrt{n}} < C\right).$$

Now by the CLT this last probability converges:
$$\mathbb{P}\left(-C < \frac{S_n - np}{\sigma\sqrt{n}} < C\right) \xrightarrow[n\to\infty]{} \int_{-C}^{C} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

As taking $C \to +\infty$ we see that the integral on RHS goes to 1 and this finishes the proof. ■

§4.3. Applications of the Normal Approximation.

<u>Confidence Interval.</u> Suppose that we have a biased coin and we don't know the true probability p that it lands on heads. How can we estimate p?
Let $\hat{p} = \frac{S_n}{n}$ the observed probability for some possibly large n. So it is important to know how close $\hat{p}$ and p are. In other words, we are interested in $\mathbb{P}(|\hat{p} - p| < \epsilon) = \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right)$

$$= \mathbb{P}(-n\epsilon < S_n - np < n\epsilon)$$
$$= \mathbb{P}\left(\frac{-\epsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n - np}{\sqrt{np(1-p)}} < \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right).$$

$$\Rightarrow \mathbb{P}(|\hat{p}-p|<\epsilon) \approx \Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\epsilon\sqrt{n}}{p(1-p)}\right)$$
$$= 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1.$$

At this point since we do not know p we cannot proceed and hence instead, we try to get a lower bound for it that works for all p. Note that the maximum of $p(1-p)$ is $1/4$. Hence, $\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq 2\epsilon\sqrt{n}$ so that, using the fact that Φ is increasing,
$$2\Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 2\Phi(2\epsilon\sqrt{n}) - 1.$$

So, $\mathbb{P}(|\hat{p}-p|<\epsilon) \geq 2\Phi(2\epsilon\sqrt{n}) - 1$. The examples below shows how can we use this.

<u>Example 4.11.</u> How many times should we flip a coin with unknown success probability p so that the estimate $\hat{p} = S_n/n$ is within 0.05 of the true p, with probability at least 0.99?

<u>Solution:</u> By above work, we need to find n so that $\mathbb{P}(|\hat{p}-p|<\epsilon) \geq 2\Phi(2\epsilon\sqrt{n}) - 1 \geq 0.99$.
$\Rightarrow \Phi(2\epsilon\sqrt{n}) \geq 0.995$ and from the table we see that $2\epsilon\sqrt{n} \geq 2.58 \Rightarrow n \geq \frac{2.58}{4\epsilon^2}$
$\Rightarrow n \geq \frac{2.58^2}{4\cdot(0.05)^2} \approx 665.64$. So, we may take $n = 666$.

<u>Example 4.12.</u> We repeat a trial 1000 times and observe 450 successes. Find the 95% confidence interval for the true success probability p.

Solution. This time we are given n and we look for ϵ such that $\mathbb{P}(|\hat{p}-p|<\epsilon) \geq 0.95$

Since $\mathbb{P}(|\hat{p}-p|<\epsilon) \geq 2\Phi(2\epsilon\sqrt{n})-1$ the inequality we need to ~~solve~~ reduces to $2\Phi(2\epsilon\sqrt{n})-1 \geq 0.95$

$\Rightarrow \Phi(2\epsilon\sqrt{n}) \geq 0.975 \Leftrightarrow 2\epsilon\sqrt{n} \geq 1.96 \Leftrightarrow \epsilon \geq 0.031.$

Thus if $n=1000$, then with probability at least 0.95 the random quantity $\hat{p}$ satisfies $|\hat{p}-p|<0.031$. So, since $\hat{p}=450/1000=0.45$, we say that 95% confidence interval for the true success probability p is $(0.45-0.031, 0.45+0.031)$.

Polling means estimating public opinion by interviewing a sample of people, which naturally involves confidence interval.

Example 4.14. Suppose that a fraction of a population who like broccoli is p. We wish to estimate p. Instead of asking every individual we take a random sample, say n individual and record their answer. Say $\hat{p}$ is the ratio of people who say yes. What is the accuracy of this estimate?

Taking a poll is a sampling without replacement. Hence, the number of people who said yes to broccoli is not exactly $\mathrm{Bin}(n,p)$. So, the method used above to estimate p for independent trials is not valid now.

However, if the sample size n is small compared to the size of the population even if we sampled with replacement, the changes of asking the same person twice would be small. As a result, sampling with or without replacement are very close to each other. With this justification we can pretend that the sample of the poll was taken with replacement and thus results in a $Bin(n,p)$ random variable. So, we may use the above estimation for $|\hat{p}-p|$. (See Remark 4.16 for justification.)

Suppose we interviewed 100 people and 20 of them liked broccoli. So $\hat{p} = 20/100 = 0.2$. Let's find the 90% confidence interval for the true p.

$$
\begin{aligned}
\mathbb{P}(|\hat{p}-p|<\epsilon) \geq 0.9 &\iff 2\Phi(2\epsilon\sqrt{n})-1 \geq 0.90 \\
&\iff \Phi(2\epsilon\sqrt{n}) \geq 0.95 \\
\text{(Table)} \quad &\iff 2\epsilon\sqrt{n} \geq 1.645 \\
(n=100) \quad &\iff 2\epsilon\sqrt{100} \geq 1.645 \\
&\iff \epsilon \geq 0.082.
\end{aligned}
$$

Thus the 90% confidence interval for p is $(0.2-0.082, 0.2+0.082) = (0.118, 0.282)$.

Suppose now, we want $\epsilon = 0.05$, so what should be n, the number of people to be interviewed?

$$
\begin{aligned}
2\Phi(2\epsilon\sqrt{n})-1 \geq 0.9 &\iff \Phi(2\epsilon\sqrt{n}) \geq 0.95 \\
&\iff 2\epsilon\sqrt{n} \geq 1.645 \\
&\iff 2\times 0.05\sqrt{n} \geq 1.645 \\
&\iff n \geq 16.45^2 \simeq 270.6.
\end{aligned}
$$

So, an additional 171 should be interviewed.

Random Variable. Suppose we flip a coin with probability of giving heads $1 > p > 0$ repeatedly. If heads come she take one step to the right otherwise to the left.

A mathematical model for this is as follows: Let $X_1, X_2, \dots, X_n \dots$ be independent random variables with $P(X_j = 1) = p$ and $P(X_j = -1) = 1-p$ and $S_n = X_1 + \dots + X_n$.

If $p = 1/2$ it is called a symmetric random walk and $p \neq 1/2$ it is an asymmetric random walk.

Let T_n count the number of times the coin comes up heads in the first n flops. Then

$S_n = T_n - (n - T_n) = 2T_n - n$. Since $T_n \sim Bin(n,p)$ we obtain

$$E[S_n] = 2E[T_n] - n = 2pn - n = n(2p-1) \text{ and}$$

$$Var[S_n] = 4 Var[T_n] = 4pn(1-p).$$

So the CLT gives

$$P\left(\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} \le b\right) = P\left(\frac{(2T_n - n) - (2pn - n)}{\sqrt{4pn(1-p)}} \le b\right)$$

$$= P\left(\frac{T_n - pn}{\sqrt{np(1-p)}} \le b\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-x^2/2} dx.$$

§4.4. Poisson Approximation

Definition 4.17. Let $\lambda > 0$. A random X has the Poisson distribution with parameter λ if X is nonnegative integer valued and has the probability mass function

$$P(X=k) = e^{-\lambda} \lambda^k / k!, \quad k = 0, 1, 2, \ldots$$

We abbreviate this $X \sim \text{Poisson}(\lambda)$.

The fact that $\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} e^{-\lambda} \lambda^k/k! = e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k/k! = 1$

show that this is indeed a probability mass function.

Fact 4.18. If $X \sim \text{Poisson}(\lambda)$ then $E[X] = Var[X] = \lambda$.

This fact follows from easily by direct computations.

Theorem 4.19. Let $\lambda > 0$ and consider integers n with $\lambda/n < 1$. Let $S_n \sim Bin(n, \lambda/n)$. Then

$$\lim_{n \to \infty} P(S_n = k) = e^{-\lambda} \lambda^k / k!, \quad k = 0, 1, 2, \ldots$$

Remark: The above theorem tells us that if S_n counts the number of successes in n independent trials and the mean $E[S_n] = \lambda$ does not change with n, then as $n \to \infty$ the distribution of S_n approaches the Poisson(λ) distribution. This theorem, after CLT, is the second limit theorem in distributions.

Proof: $\mathbb{P}(S_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$

$$\Rightarrow \mathbb{P}(S_n = k) = \frac{n(n-1)\dots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left[1-\frac{\lambda}{n}\right]^n \frac{1}{\left(1-\frac{\lambda}{n}\right)^k}$$

$$= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left[1\cdot\left(1-\tfrac{1}{n}\right)\left(1-\tfrac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]\frac{1}{\left(1-\lambda/n\right)^k}$$

$$\xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}\cdot 1\cdot 1 = \frac{\lambda^k e^{-\lambda}}{k!}$$

(Note that k is kept fixed as $n\to\infty$.)

Poisson Approximation of counts of rare events

The above theorem tells us that the distribution $B(n, \lambda/n)$ gets closer and closer to Poisson(λ) as $n \to \infty$. The theorem below gives a statement for fixed n, whose proof will be skipped.

Theorem 4.20. Let $X \sim Bin(n,p)$ and $Y \sim$ Poisson(np). Then for any subset $A \subseteq \{0,1,2,\dots\}$, we have

$$|\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)| \leq np^2.$$

(Note that $E[X] = np = \lambda$ for Y.)

Note that this result has a meaning if np^2 is small. Since n is fixed this is possible only if p is very small. In other words, if the occurence is rare!

Poisson approximation for counting rare events.

Assume that the random variable X counts the occurences of "rare events" that are "not strongly dependent" on each other. Then the distribution of X can be approximated with a Poisson(λ) distribution for $\lambda = \mathbb{E}[X]$. That is

$\mathbb{P}(X=k)$ is close to $e^{-\lambda}\lambda^k/k!$, for $k=0,1,2,\ldots$

Remark: Note that the terms "rare events" and "strong dependence" are not mathematically defined!

Example 4.21. Suppose a factory experiences on average 3 accidents each month. What is the probability that during a particular month there are exactly 2 accidents?

If X is the number of accidents during a month then we may assume $X \sim$ Poisson(3). Hence, $\mathbb{P}(X=2) = e^{-3}3^2/2! \approx 0.224$.

Example 4.22. The proofreader of an undergraduate probability textbook noticed that a randomly chosen page of the book has no typos with probability 0.9. Estimate the probability that a randomly chosen page contains exactly two typos.

Let X denote the number of typos on a randomly chosen page. Suppose X is close to Poisson(λ) where $\lambda = E[X]$. However, $E[X]$ is not known. On the other hand, it is given that $\mathbb{P}(X=0) \approx 0.9 = e^{-\lambda}\lambda^k/k!$, where $k=0$. So $e^{-\lambda} = 0.9 \implies \lambda = -\ln 0.9 \approx 0.105$. Hence

$P(X=2) \approx e^{-\lambda} \lambda^2/2! \approx 0.9 \times \frac{(0.105)^2}{2!} \approx 0.005.$

Remark: Example 4.24 and 4.25 give a comparison of the normal and Poisson approximation of the binomial.

§4.5. Exponential Distribution.

The geometric distribution is a discrete probability distribution that models waiting times, such as, the first time in a sequence of coin flips that tails appears. Below we develops a continuous counterpart, for modelling waiting times such as the first arrival of a customer at a post office.

Definition 4.26. Let $0<\lambda<\infty$. A random variable X has the exponential distribution with parameter λ if X has the density function

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

on the real line.

This will be abbreviated by $X \sim Exp(\lambda)$, which is also called as the exponential distribution with rate λ.

The cumulative density function of $Exp(\lambda)$ is

$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \int_0^t \lambda e^{-\lambda x}\,dx = 1-e^{-\lambda t}, \text{ for } t \geq 0$$

and $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

Also note that, for $t>0$, $P(X>t) = 1-P(X\leq t) = e^{-\lambda t}.$

On the other hand, $E[X] = \int_0^\infty x\lambda e^{-\lambda x}\,dx = 1/\lambda$ and $E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} = 2/\lambda^2$ so that $Var[X] = 1/\lambda^2$.

<u>Example 4.27.</u> Let $X \sim Exp(1/2)$. Find $P(X > 7/2)$. Also find the median of X, that is, the value m for which $P(X \leq m) = 1/2$.

$P(X > 7/2) = \int_{7/2}^\infty e^{-x/2}\,dx = e^{-7/4}$. Also $P(X \leq m) = 1 - e^{-m/2}$.

So, for the median $1 - e^{-m/2} = 1/2 \Rightarrow m = 2\ln 2$.

<u>Example 4.29.</u> Suppose that the length of a phone call, in minutes, is well modelled by an exponential random variable with mean 10 minutes. So if X is the length of a phone call then $X \sim Exp(1/10)$. Compute $P(X > 8)$ and $P(8 < X < 22)$.

Since $P(X > t) = e^{-\lambda t}$, $P(X > 8) = e^{-8/10} \approx 0.4493$ and $P(8 < X < 22) = e^{-8/10} - e^{-22/10} \approx 0.3385$.

<u>Example 4.30.</u> Consider a particular protein found in human cells. Let T denote the time, in seconds, required until one of these proteins is degraded (destroyed). A common modelling assumption in biomolecular chemistry is that $T \sim Exp(\lambda n)$ for some $\lambda > 0$, where n is the number of copies of the protein in the cell. Suppose that for a particular protein $\lambda = 2$. How many proteins have to be present so that $P(T > 10^{-2}) < 1/10$?

$$1/10 > \mathbb{P}(T > 10^{-2}) = \int_{10^{-2}}^{\infty} 2n e^{-2nx}\,dx = e^{-2n/100}.$$

$$\Rightarrow n > \frac{100}{2}\ln(10) \simeq 115.1 \Rightarrow n = 116 \text{ has to be present.}$$

<u>Fact 4.31.</u> (Memoryless property of the exponential distribution) Suppose that $X \sim \mathrm{Exp}(\lambda)$. Then, for any $s,t > 0$ we have $\mathbb{P}(X > s+t \mid X > t) = \mathbb{P}(X > s)$.

<u>Proof:</u>
$$\mathbb{P}(X > s+t \mid X > t) = \frac{\mathbb{P}(X > t+s,\ X > t)}{\mathbb{P}(X > t)}$$
$$= \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda s}$$
$$= \mathbb{P}(X > s).$$

<u>Example 4.32.</u> Animals of a forest need to cross a remote highway. From experience, it is known that, from the moment someone arrives at the roadside, the time till the next car is an exponential random variable with expected value 30 minutes. The turtle needs 10 minutes to cross the road. (a) What is the probability that the turtle can cross the road safely? (b) Now suppose that when the turtle arrives at the roadside, the fox tells her that he has been there for 5 minutes without seeing a car go by. What is the probability now that the turtle can cross safely?

<u>Solution:</u> Let X denote the time in minutes the arrival of the next car, measured from the

moment when the turtle arrived on the side of the road. Then, $X \sim Exp(\lambda)$, where $1/\lambda = E[X] = 30 \Rightarrow \lambda = 1/30$. So for (a)

$\mathbb{P}(X > 10) = e^{-10\lambda} = e^{-1/3} \approx 0.7165$. (b) Since X is memoryless, $\mathbb{P}(X > 5+10 \mid X > 5) = \mathbb{P}(X > 10) \approx 0.7165$ again.

<u>Derivation of the Exponential Distribution.</u>

We'll use the discrete time scale model for the time when the first customer arrives at a post office. Suppose that there exists a constant $\lambda > 0$ such that the probability at least one customer arrives during a time interval of length $1/n$ minutes is approximately λ/n, for very large n.

For each large n we record a discretized random arrival time T_n as follows: For $k = 1, 2, 3, \ldots$, if the first customer arrives during time interval $[(k-1)/n, k/n)$, we set $T_n = k/n$. Assuming the arrivals in disjoint time intervals are independent, the probability mass function is then $\mathbb{P}(T_n = k/n) = (1 - \lambda/n)^{k-1} \frac{\lambda}{n}$, for $k \geq 1$.

Note that $nT_n \sim Geom(\lambda/n)$.

<u>Theorem 4.33.</u> Fix $\lambda > 0$ and consider n with $\lambda/n < 1$. Suppose that for each large enough n, the random variable T_n satisfies $nT_n \sim Geom(\lambda/n)$.

Then $\lim_{n\to\infty} \mathbb{P}(T_n > t) = e^{-\lambda t}$, for all non-negative real t.

Proof: For $t>0$ let's compute $\mathbb{P}(T_n > t)$.

$$\mathbb{P}(T_n > t) = \mathbb{P}(nT_n > nt) = \sum_{k:\, k \geq nt} \mathbb{P}(nT_n = k)$$

$$= \sum_{k:\, k\geq nt} \mathbb{P}(T_n = k/n)$$

$$= \sum_{k:\, k \geq n} \left(1-\frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n}$$

$$= \sum_{k = \lfloor nt \rfloor + 1} \left(1-\frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n}$$

$$= \left(1-\frac{\lambda}{n}\right)^{\lfloor nt \rfloor} = \left(1-\frac{\lambda}{n}\right)^{nt} \left(1-\frac{\lambda}{n}\right)^{\lfloor nt \rfloor - nt}$$

$$= \left(1-\frac{\lambda t}{nt}\right)^{nt} \frac{1}{\left(1-\frac{\lambda}{n}\right)^{nt-\lfloor nt \rfloor}}.$$

Since $0 \leq nt - \lfloor nt \rfloor \leq 1$ for all n and t we see that $n \to \infty$, $\left(1-\frac{\lambda}{n}\right)^{nt-\lfloor nt \rfloor} \to 1$.

So, as $n \to \infty$, $\mathbb{P}(T_n > t) \to \lim_{n\to\infty} \left(1 - \frac{\lambda t}{nt}\right)^{nt} = e^{-\lambda t}$. ■

As a consequence, the c.d.f. of T_n, F_{T_n} satisfies
$$\lim_{n\to\infty} F_{T_n}(t) = \lim_{n\to\infty} (1-\mathbb{P}(T_n > t)) = 1 - e^{-\lambda t} = F(t),$$
where $F(t)$ is the c.d.f. of the distribution $\mathrm{Exp}(\lambda)$.

The precise statement is that the random variable T_n converges in distribution to an $\mathrm{Exp}(\lambda)$ random variable. So this is another example of limit in distribution.

Chapter 5. Transforms and Transformations.

§5.1. Moment Generating Function.