Math 170E (Introduction to Probability)
University of California, Los Angeles

Aaron Chao

Summer 2020

Course description: Introduction to probability theory with emphasis on topics relevant to applications. Topics include discrete (binomial, Poisson, etc.) and continuous (exponential, gamma, chi-square, normal) distributions, bivariate distributions, distributions of functions of random variables (including moment generating functions and central limit theorem).

These are my lecture notes for Math 170E (Introduction to Probability and Statistics: Part 1 Probability) taught by Enes Ozel. The main textbook for this class is *Probability and Statistical Inference (10th Edition)* by Robert V. Hogg, Elliot Tanis, and Dale Zimmerman.

Contents

1 June 22, 2020

1.1 Properties of Probability

Definition 1.1 (Sample space)
The sample space, denoted by S, is the whole set of possible outcomes.

Definition 1.2 (Event)
Any subset of S is called an event.

Example 1.3
Let A be the event we will get ≥ 1 head. Then

$$A = \{HH, HT, TH\} \subset S$$

Recall 1.4 (Sets) Let A, B be two subsets of S.

- $A \cap B$: intersection $\{x \in S \colon x \in A \wedge x \in B\}$
- $A \cup B$: union $\{x \in S \colon x \in A \vee x \in B\}$
- $A \subset B$: subset $\forall x \in S, x \in A \implies x \in B$
- $\emptyset$: empty set $\forall x \in S, x \notin \emptyset$
- $A' = \overline{A} = A^C$: complement $\{x \in S \colon x \notin A\}$
- $A \setminus B = A \cap B^C = \{x \in S, x \in A \wedge x \notin B\}$
- $A \cap B = \emptyset$ "mutually exclusive"

Example 1.5
Let A be the event we will get ≥ 1 head, B be the event we get both tails. Then

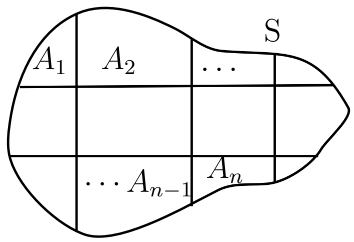
$$A = \{HH, TH, HT\} \quad B = \{TT\}$$

Then

$$A \cup B = S$$

Definition 1.6 (Exhaustive)
$A_1, \ldots, A_n \subseteq S$ are exhaustive if $A_1 \cup A_2 \cup \cdots \cup A_n = S$, i.e. $\bigcup_{i=1}^{n} A_i = S$.

Here, we have mutually exclusive and exhaustive events.



Definition 1.7 (Probability)
Probability is a real-valued set function P that assigns, to each event A in the sample space S, a number $P(A)$, called the probability of the event A, such that the following properties are satisfied:

(a) $P(A) \geq 0$

(b) $P(S) = 1$

(c) For $\{A_i\}_{i=1}^{\infty} \subseteq S$ such that the sets are mutually exclusive,

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

Note, $P(A) \in \mathbb{R}$ and $P(S) \to \mathbb{R}$.

Theorem 1.8
$\forall A \subseteq S, P(A') = 1 - P(A)$.

Proof. Let $A \subseteq S$. Then $S = A \cup A' \implies$ mutually exclusive and exhaustive

$$\begin{aligned} \implies P(A \cup A') &= P(A) + P(A') \\ &= P(S) \\ &= 1 \end{aligned}$$

So

$$P(A) + P(A') = 1 \implies P(A') = 1 - P(A)$$

□

Corollary 1.9
$P(\emptyset) = 0$.

Proof. $\emptyset^C = S$, so

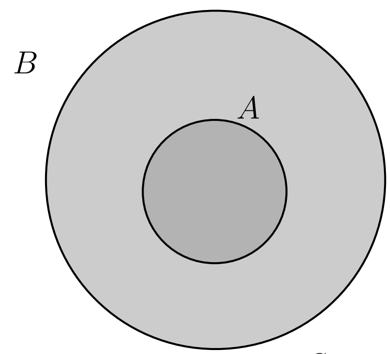
$$\begin{aligned} P(\emptyset) &= 1 - P(S) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

□

Example 1.10
Flip two coins,
$$S = \{HH, HT, TT, TH\}$$
assuming a fair coin toss, all have 1/4 probability.
$A =$ "both heads" so
$$P(A) = \frac{1}{4}$$
$$P(A') = \frac{3}{4}$$
$$A' = \{HT, TT, TH\}$$
where A' is "NOT both heads" = "at least one tail"

Theorem 1.11
Let $A, B \subseteq S$ such that $A \subseteq B$. Then $P(A) < P(B)$.

Proof. $A \subseteq B \implies B = A \cup (B \cap A')$



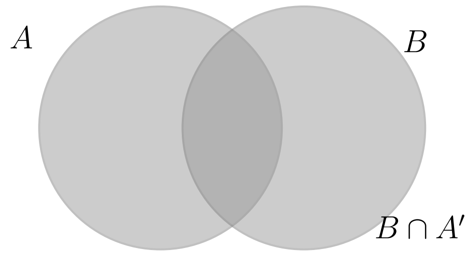
$B \setminus A = B \cap A^C$

$$P(B) = P(A) + \underbrace{P(B \cap A')}_{\geq 0} \geq P(A)$$
$$\implies P(B) \geq P(A)$$
□

Corollary 1.12
$\forall A \subseteq S, 0 \leq P(A) \leq 1$.

Theorem 1.13
If A and B are any two events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof. Decompose the units into two disjoint parts



$$P(A \cup B) = P(A) + P(B \cap A')$$

But

$$P(B) = P(A \cap B) + P(A' \cap B)$$

So

$$P(B \cap A') = P(B) - P(A \cap B)$$

Hence

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \cap A') \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

Example 1.14
We have a fair die, which we roll once. So

$$S = \{1, 2, \ldots, 6\}$$

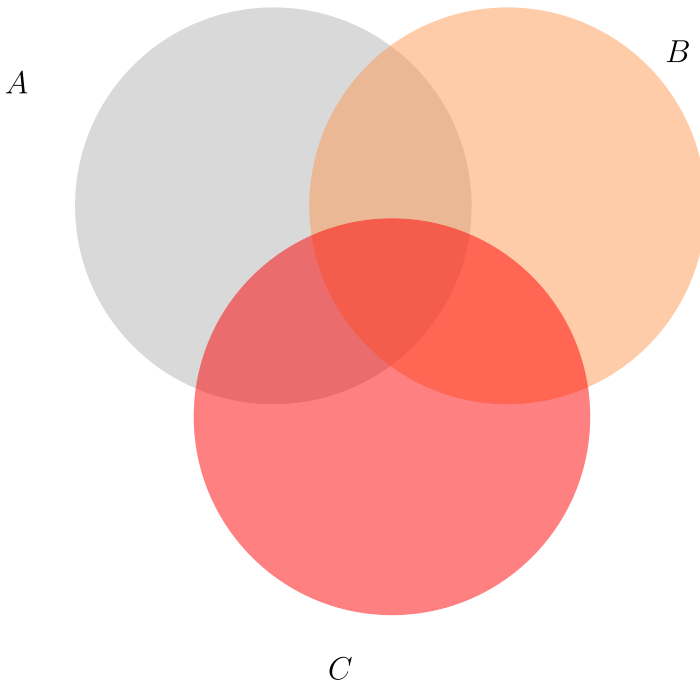
with probability 1/6 each. We call the outcome X. So

$$\begin{aligned} P(2 \mid X \text{ or } 3 \mid X) &= P(2 \mid X) + P(3 \mid X) - P(6 \mid X) \\ &= P(\{2,4,6\}) + P(\{3,6\}) - P(\{6\}) \\ &= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} \\ &= \frac{2}{3} \end{aligned}$$

Theorem 1.15
If A, B, and C are any three events, then

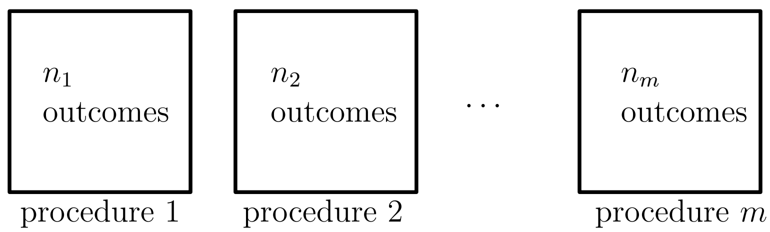
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Proof.



Exercise. □

1.2 Methods of Enumeration



We are interested in the number of ways the overall outcome may be formed, which we calculate using the formula $n_1 \times n_2 \times \cdots \times n_m$.

Example 1.16 (Cafe, deli sandwich)
Suppose a cafe has

	Bread	Meat	Cheese	Garnishes
	6	4	4	12
# that can be chosen	1	0,1	0,1	$0, 1, 2, \cdots, 12$

So

$$\#\text{ different sandwiches } = 6 \times 5 \times 5 \times 2^{12}$$

Suppose we have n people to be placed. Then the number of ways to arrange them is given by

$$\begin{aligned}\#\text{ ways } &= \underset{1^{st}}{n} \quad \underset{2^{nd}}{n-1} \quad \underset{3^{rd}}{n-2} \quad \cdots \quad \underset{n-1^{st}}{2} \quad \underset{n^{th}}{1}\\ &= n!\end{aligned}$$

Example 1.17
Suppose $S = \{a, b, c, d\}$. Then the number of permutations is $4! = 24$.

$$\left.\begin{matrix} abcd \\ acbd \\ \vdots \end{matrix}\right\} 24$$

If we allow repetitions,

$$4 \times 4 \times 4 \times 4 = 4^4 = 256$$

Definition 1.18 (Permutation)
Suppose we have n objects/people and $r \leq n$ positions, then the number of ways to arrange them is given by

$$\begin{aligned}\#\text{ of ways } &= \underset{1^{st}}{n} \quad \underset{2^{nd}}{n-1} \quad \underset{3^{rd}}{n-2} \quad \cdots \quad \underset{r-1^{st}}{n-r+2} \quad \underset{r^{th}}{n-r+1}\\ &= n \times (n-1) \times (n-2) \times \cdots \times (n-r+2) \times (n-r+1)\\ &= {}_nP_r\end{aligned}$$

which is the number of permutations of n objects taken r at a time.
We define the r^{th} falling factorial of n as

$$(n)_r = n \cdot (n-1) \cdots (n-r+1)$$

and the factorial of n as

$$(n)_n = n \cdot (n-1) \cdots (n-n+1) = n!$$

2 June 24, 2020

2.1 Binomial Coefficients

Definition 2.1 (Binomial coefficient)
We are interested in the number of ways to choose r objects out of n objects. The number is given by

$$\begin{aligned}\# &= \frac{{}_nP_r}{r!} = \frac{(n)_r}{r!} \\ &= \frac{n(n-1)\cdots(n-r+1)}{r!} \\ &= \frac{n(n-1)\cdots(n-r+1)}{r!}\frac{(n-r)!}{(n-r)!} \\ &= \frac{n!}{r!(n-r)!} \\ &= \binom{n}{r}\end{aligned}$$

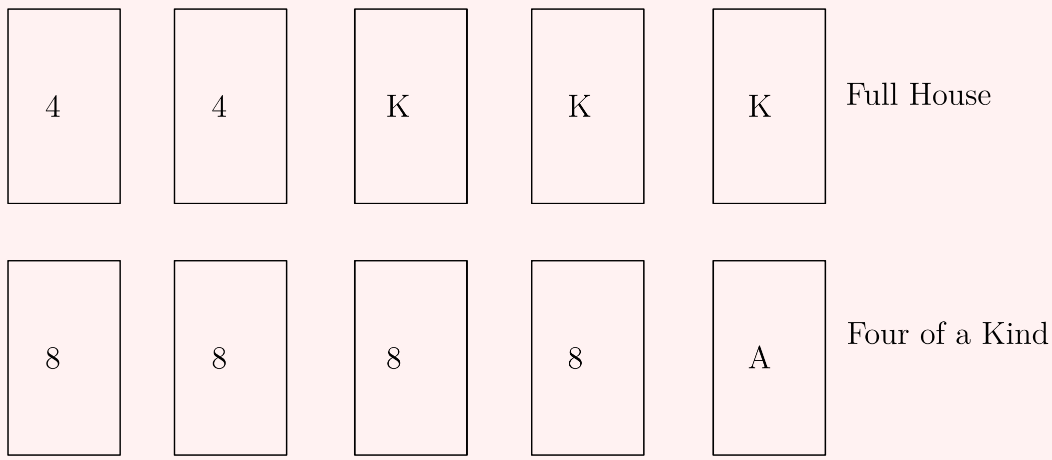
which we call the <u>binomial coefficient</u>.

Example 2.2
We have a group of 5 people. We want to choose 2 to govern, and it does not matter who is president/treasurer.

$$\#\text{ ways } = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

Example 2.3 (Five-card Poker Hands)
We have a deck of 52 cards, where there are 4 suits and 13 denominations. The number of possible five card Poker hands is

$$\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{1 \times 2 \times 3 \times 4 \times 5} = 2598960$$



Which one should beat the other? The one less likely.

$$P(\text{full house}) = \frac{\#\text{ full house hands}}{\binom{52}{5}}$$

$$P(4\text{ of a kind}) = \frac{\#4\text{ kind}}{\binom{52}{5}}$$

So

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}} \approx 0.00144$$

$$P(4\text{ of a kind}) = \frac{\binom{13}{1}\binom{4}{4}\binom{12}{1}\binom{4}{1}}{\binom{52}{5}} \approx 0.00024$$

Example 2.4
Choose from orchestra from 100 students who can play cello, violin, trumpet, clarinet, and ney. We want 10 cellists, 55 violinists, 15 trumpeteers, 18 clarinetists, the rest will be neyzens. How many different orchestras may be formed?

$$\binom{100}{10}\binom{90}{55}\binom{35}{15}\binom{20}{18}\binom{2}{2}$$

The answer may be reformed in terms of a multinomial coefficient:

$$\binom{100}{10, 55, 15, 18, 2} = \frac{100!}{10! \cdot 55! \cdot 15! \cdot 18! \cdot 2!}$$

Definition 2.5 (Multinomial coefficient)
Let $n_1 + \cdots + n_k = n$, then the mutinomial coefficient is

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

Order matters	Repetition allowed	
Yes	Yes	$P(n,r) = n^r$
Yes	No	$P(n,r) = \dfrac{n!}{(n-r)!}$
No	No	$C(n,r) = \dfrac{n!}{r!(n-r)!}$
No	Yes	$C(n+r-1,r) = \dfrac{(n+r-1)!}{r!(n-1)!}$

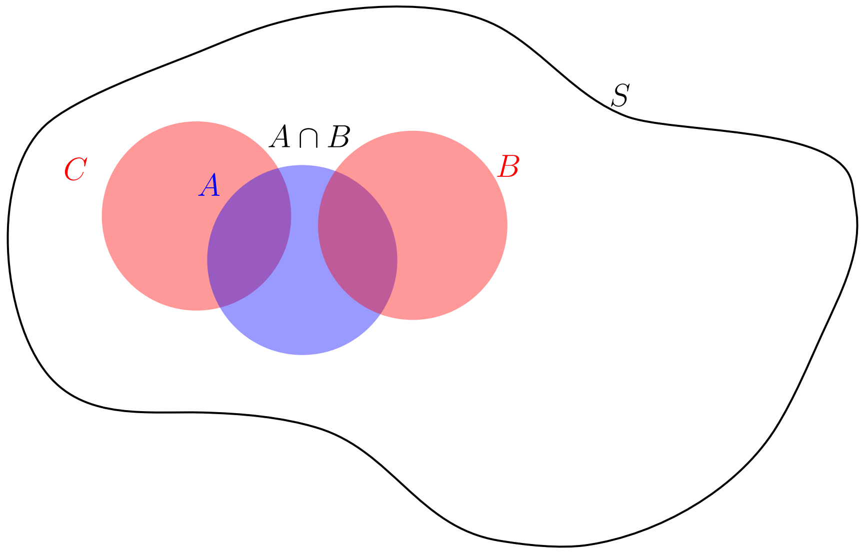
Theorem 2.6 (Binomial and Multinomial Theorem)
If $n \geq 0$ and $s \geq 0$, then

$$(a+b)^n = \sum_{k=0}^{n} a^k b^{n-k} \binom{n}{k}$$

$$(a_1 + a_2 + \cdots + a_s)^n = \sum_{n_1 + n_2 + \cdots + n_s = n} a_1^{n_1} a_2^{n_2} \cdots a_s^{n_s} \cdot \binom{n}{n_1, n_2, \ldots, n_s}$$

2.2 Conditional Probability

Recall the sample space S and let $A \subseteq S$ be any event with probability of occurrence $P(A)$. Assume $B \subseteq S$ is another event and we know B has happened. Should that affect chances of A?



Notice $B \cap C = \varnothing$. So B happens implies C cannot happen.

Definition 2.7 (Conditional probability)
The conditional probability of an event A given that B has occurred is defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

given that $P(B) > 0$.

Note 2.8: If we condition on B, B becomes the new sample space.

Exercise. Prove $P(A \cap B^C \mid B) = 0$.

Example 2.9
Let (d_1, d_2) be two fair dice with 36 outcomes, all with probability $1/36$.
So $P(d_1 = 5) = 1/6$.

$$\begin{aligned} P(d_1 = 5 \mid d_1 + d_2 = 7) &= \frac{P(d_1 = 5 \cap d_1 + d_2 = 7)}{P(d_1 + d_2 = 7)} \\ &= \frac{P(d_1 = 5 \cap d_2 = 2)}{P(d_1 + d_2 = 7)} \\ &= \frac{1/36}{6/36} = 1/6 \end{aligned}$$

The event of $d_1 + d_2 = 7$ did not change the probability of $d_1 = 5$. Now,

$$\begin{aligned} P(d_1 = 5 \mid d_1 + d_2 = 6) &= \frac{P(d_1 = 5 \cap d_1 + d_2 = 6)}{P(d_1 + d_2 = 6)} \\ &= \frac{P(d_1 = 5 \cap d_2 = 1)}{P(d_1 + d_2 = 6)} \\ &= \frac{1/36}{5/36} = 1/5 \end{aligned}$$

The event of $d_1 + d_2 = 6$ changed the probability of $d_1 = 5$.

Exercise. $P(d_1 = 5 \mid d_1 + d_2 = 5) = 0$

Definition 2.10 (Multiplication Rule)
The probability that two events, A and B, both occur is given by the Multiplication Rule:

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

assuming $P(A) > 0$ and $P(B) > 0$.

Example 2.11 (Card example)
Assuming a deck of 52 cards:

$$P(\text{1st card is a queen}) = P(\text{1st} = \text{Q}) = \frac{4}{52}$$

$$P(\text{2nd} = \text{queen} \mid \text{1st} = \text{queen}) = \frac{3}{51}$$

$$P(\text{2nd} = \text{queen} \mid \text{1st} \neq \text{queen}) = \frac{4}{51}$$

$$P(\text{2nd} = \text{queen}) = \frac{4}{52}$$

$$P(\text{26th} = \text{H} \mid \text{1st} = \text{Queen of Hearts, 5th} = \text{Ace of spades}) = \frac{12}{50}$$

$$P(\text{1st} = \text{A} \cap \text{50th} = \text{Queen of H}) = P(\text{50th} = \text{Queen of H} \mid \text{1st} = \text{A}) \cdot P(\text{1st} = \text{A}) = \frac{1}{51} \cdot \frac{4}{52}$$

Theorem 2.12

$$\begin{aligned} P(A \cap B \cap C) &= P(C \mid A \cap B) \cdot P(A \cap B) \\ &= P(C \mid A \cap B) \cdot P(B \mid A) \cdot P(A) \\ &= \underbrace{P(A) \cdot P(B \mid A)}_{P(A \cap B)} \cdot P(C \mid A \cap B) \end{aligned}$$

given $P(A), P(A \cap B) > 0$.

Example 2.13 (Card example (Cont'd))

$$\begin{aligned} &P(\text{1st} = \text{J} \cap \text{2nd} = \text{J} \cap \text{4th} = \text{J of Spades}) = P(\text{4th} = \text{J of Spades}) \\ &\quad \cdot P(\text{2nd} = \text{J} \mid \text{4th} = \text{J of Spades}) \cdot P(\text{1st} = \text{J} \mid \text{2nd} = \text{J} \cap \text{4th} = \text{J of Spades}) \\ &\qquad = \frac{1}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \end{aligned}$$

2.3 Independent Events

Suppose

$$P(d_1 = 5) = 1/6$$

$$P(d_1 = 5 \mid d_1 + d_2 = 7) = 1/6$$

but

$$P(d_1 = 5 \mid d_1 + d_2 = 6) = 1/5$$

Then,

- $\{d_1 = 5\}$ and $\{d_1 + d_2 = 7\}$ are independent

- $\{d_1 = 5\}$ and $\{d_1 + d_2 = 6\}$ are dependent

Definition 2.14 (Independent vs. dependent events)
Events A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$, otherwise A and B are called dependent events.

Theorem 2.15
Assume A and B are independent and both have nonzero probabilities. Then,

$$\begin{aligned} P(A \cap B) &= P(A \mid B)P(B) = P(B \mid A)P(A) \\ &= P(A)P(B) \end{aligned}$$

As a result, we get

$$P(A \mid B) = P(A)$$

and

$$P(B \mid A) = P(B)$$

In the case $P(A) = 0$ (or $P(B) = 0$) as $A \cap B \subseteq A$,

$$P(A \cap B) \leq P(A) = 0$$

$$P(A \cap B) = 0 = P(A)P(B)$$

As a result,

Theorem 2.16
A trivial event (meaning probability 0) is always independent from any other event.

$$0 < P(A) < 1 \implies 0 < P(A') < 1$$

Theorem 2.17
If A and B are independent events, then the following events are also independent:

a) A and B'

b) A' and B

c) A' and B'

Definition 2.18 (Mutually independent)
Events A, B, and C are mutually independent if and only if the following two conditions hold:

i. Pairwise independent

$$P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C), P(B \cap C) = P(B)P(C)$$

ii. Triple wise independent

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

2.4 Bayes' Theorem

Example 2.19
Consider a very rare disease and a diagnosis test proposed by a very famous pharmaceutical company.

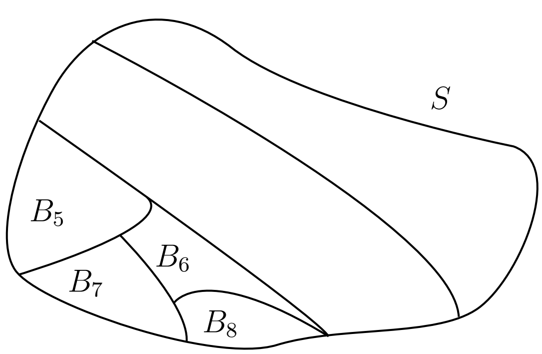
$$\begin{gathered} D\text{: have the disease} \\ D^C\text{: not have the disease} \\ T\text{: test positive} \\ T^C\text{: test negative} \end{gathered}$$

Suppose $P(D) = \dfrac{1}{1000000}$. The company assures us the test is accurate.

$$P(T \mid D) = 0.999$$

$$P(T^C \mid D^C) = 0.9999$$

Let A be any event. Let $\{B_i\}_{i=1}^n$ be mutually exclusive and exhaustive events.



Theorem 2.20 (Bayes' Theorem)
Let A and $\{B_i\}_{i=1}^n$ be defined as above. Then for each $k \in \{1, 2, \ldots, n\}$:

$$P(B_k \mid A) = \frac{P(A \mid B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A \mid B_i) \cdot P(B_i)}$$

Proof.

$$P(B_k \mid A) = \frac{P(A \cap B_k)}{P(A)}$$

$$A = \underbrace{(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)}_{\text{mutually exclusive}}$$

Substitute in the denominator,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

So,

$$P(B_k \mid A) = \frac{P(A \cap B_k)}{\sum_{i=1}^{n} P(A \cap B_i)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$

□

Example 2.21 (Company testing (Cont'd))

The company's test:

$$P(T \mid D) = 0.999$$

$$P(T^C \mid D^C) = 0.9999$$

We conduct the test and it says positive. What is the probability that they are really sick?

$$\begin{aligned} P(D \mid T) &= \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid D^C)P(D^C)} \\ &= \frac{0.999(0.000001)}{0.999(0.000001) + 0.0001(0.999999)} \\ &\approx 0.00989 < 1\% \end{aligned}$$

Example 2.22
A life insurance company has the following policies: standard, preferred, ultrapreferred. We have a fixed age x: Policyholders who are of x years old:

$$P(S) = 0.6$$

$$P(P) = 0.3$$

$$P(U) = 0.1$$

1-year mortality probabilities for x year old policyholders:

S	0.01
P	0.008
U	0.007

$D = \{x \text{ dies within a year}\}$.

$$P(D \mid S) = 0.01$$

$$P(D \mid P) = 0.008$$

$$P(D \mid U) = 0.007$$

Given that a policyholder dies within a year, what are the probabilities that their policy was of each type?

$$P(S \mid D) = ? \quad P(P \mid D) = ? \quad P(U \mid D) = ?$$

$$\begin{aligned} P(S \mid D) &= \frac{P(D \mid S)P(S)}{P(D \mid S)P(S) + P(D \mid P)P(P) + P(D \mid U)P(U)} \\ &= \frac{0.01(0.6)}{0.01(0.6) + 0.008(0.3) + 0.007(0.1)} \\ &= \frac{0.006}{0.0091} \end{aligned}$$

so $P(D) = 0.0091$

$$\approx 0.65934$$

$$P(P \mid D) = \frac{P(D \mid P)P(P)}{0.0091} = \frac{0.008(0.3)}{0.0091} \approx 0.26374$$

$$P(U \mid D) = \frac{P(D \mid U)P(U)}{0.0091} = \frac{0.007(0.1)}{0.0091} = 0.07692$$

Example 2.23 (Birthday problem)
We have 80 students in lecture. What is the probabilities that $\geq$ two students have the same birthday?
Assume 365 days in a year, for each student, each day is equally probable, and birthdays are independent. Then

$$\begin{aligned} P(\geq 1 \text{ coincide}) &= 1 - P(\text{all birthdays differ}) \\ &= 1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{286}{365} \\ &\approx 0.99991433 \end{aligned}$$

If there are 23 students,

$$P(\geq 1 \text{ coincide}) > 0.5$$

$$P_{30} > 0.7$$

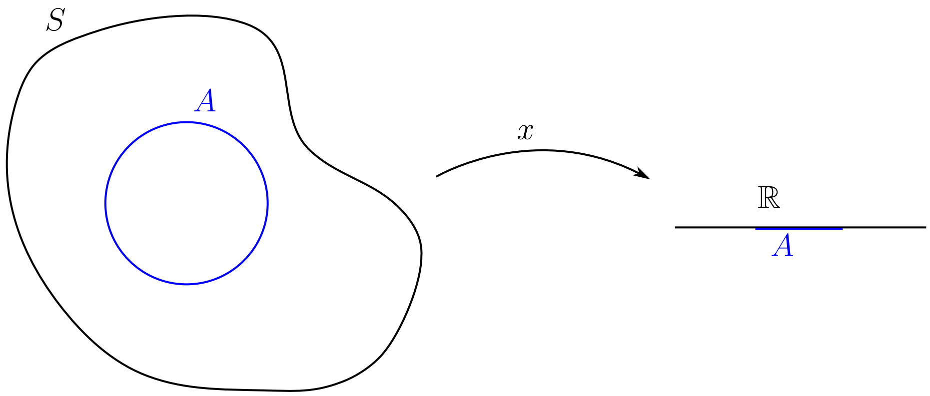
$$P_{40} > 0.89$$

$$P_{50} > 0.97$$

3 June 26, 2020

3.1 Discrete Distributions

The sample space S may consist of many strange outcomes wherever numbers are not numerical, analysis becomes difficult. Random variables are functions that carry the events $A \subseteq S$ into subsets of $\mathbb{R}$.



Is any $x\colon S \to \mathbb{R}$ a random variable? No. It must be "measurable".

Definition 3.1 (Random variable and space)
Given a random experiment with a sample space S, a function x that assigns a real number to each element $s \in S$, $X(s) = x \in \mathbb{R}$ is called a random variable. The space of X is the set of real numbers

$$\{x\colon X(s) = x, \text{ for some } s \in S\}$$

Definition 3.2 (Discrete random variable)
If $\exists n \in \mathbb{N}$ such that $|S| = \aleph$, or S can be put into a bijection with $\mathbb{N}$, also called countably infinite, then x is called a discrete random variable.

Example 3.3
Let S be the outcomes of 5 coin tosses. So $P(H) = p \in (0,1)$. Then

$$S = \{HHHHH, HTTTH, TTHTH, \dots\}$$

$$|S| = 2^5 = 32$$

$$X: S \to \mathbb{R}$$

$$X(s) = \#H's$$

So

$$X(HHHHH) = 5$$

$$X(THTHT) = 2$$

So $X \sim \text{Binomial}(n, P)$. The support/space is given by $S(x) = \{0, 1, \dots, 5\}$.
Now let

$$Y: S \to \mathbb{R}$$

$$\forall s \in S, Y(S) = \#H - \#T$$

So

$$Y(HHHHH) = 5$$

$$Y(THTHT) = -1$$

Hence,

$$S(Y) \in \{-5, \dots, 5\}$$

Wherever we consider a discrete random variable, the values it may take, its discrete space will be associated with masses. We call $P(X = x)$ the probability mass of X at x.

$$p_x(x) = f(x) = P(X = x)$$

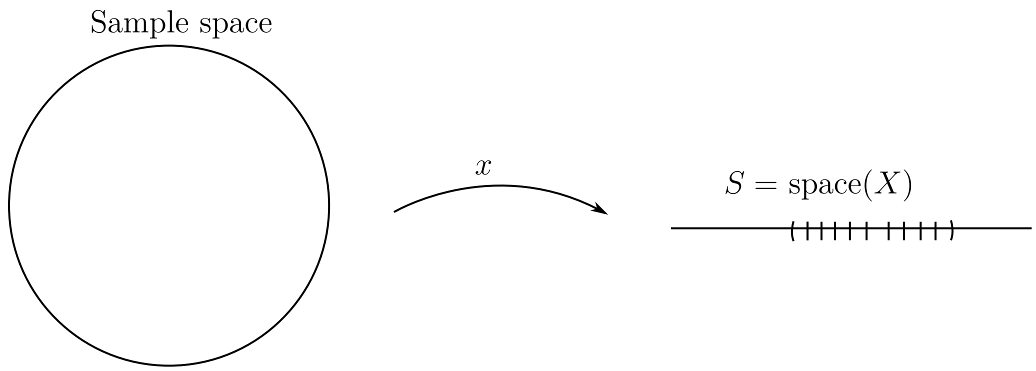
Definition 3.4 (Probability mass function (PMF))
Let X be a given discrete random variable, S be its space, i.e. $x \in S$. Then any function $f: S \to [0,1]$ satisfying the following properties is a well-defined PMF:

a) $f(x) > 0, \forall x \in S$

b) $\sum_{x \in S} f(x) = 1$

c) $P(X \in A) = \sum_{x \in A} f(x)$, for any $A \subseteq S$

Suppose we have our sample space S from Example 3.3, then:



$$S(x) = \{0, 1, \ldots, 5\}$$

Definition 3.5 (Cumulative distribution function (CDF))
The cumulative distribution function, which is the distribution of a random variable, is given by

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

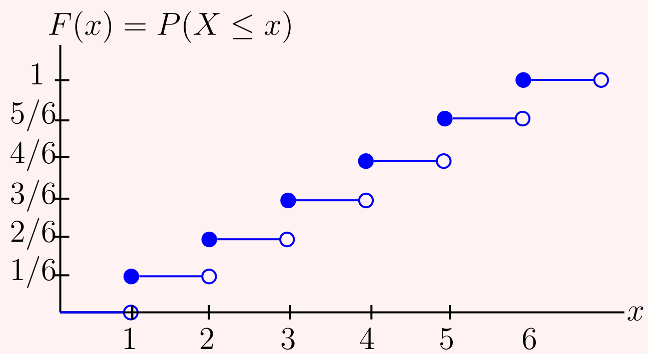
Example 3.6
Recall $\{1, 2, 3, 4, 5, 6\}$, a fair die roll. Then the PMF is

$$f(x) = P(X = x) = 1/6 \text{ for } x \in \{1, 2, 3, \ldots, 6\}$$

$$f(x) = 0 \text{ for all other } x$$

The CDF is

$$F_x(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/6 & \text{if } 1 \leq x < 2 \\ 2/6 & \text{if } 2 \leq x < 3 \\ 3/6 & \text{if } 3 \leq x < 4 \\ 4/6 & \text{if } 4 \leq x < 5 \\ 5/6 & \text{if } 5 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$



$F(6.5) = P(X \leq 6.5) = P(X = 1, X = 2, \ldots, X = 6)$.

4 June 29, 2022

4.1 Discrete Distributions (Cont'd)

Example 4.1

Roll a fair 4-sided die twice: d_1, d_2. Let $x = \max(d_1, d_2)$. So,

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ (d_1, d_2) & (1,1) & (1,2) & (1,3) & (1,4) \\ & (2,1) & (2,2) & (2,3) & (2,4) \\ & (3,1) & (3,2) & (3,3) & (3,4) \\ & (4,1) & (4,2) & (4,3) & (4,4) \end{array}$$

$$\text{Support}(X) = \{1, 2, 3, 4\}$$

$$f(1) = 1/16$$

$$f(2) = 3/16$$

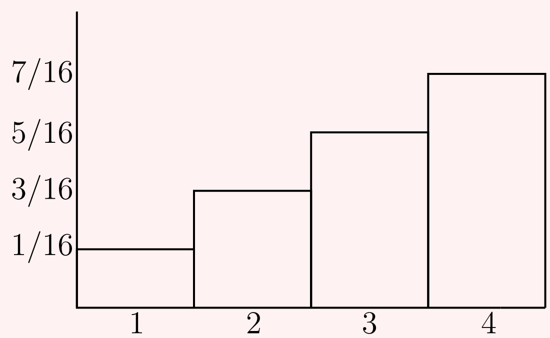
$$f(3) = 5/16$$

$$f(4) = 7/16$$

$$f(x) = 0 \text{ for any other } x \in \mathbb{R}$$

The PMF is

$$f(x) = \frac{2x-1}{16}, \quad x = 1, 2, 3, 4$$



Example 4.2
Let $x = \#$ accidents per week in a factory with PMF

$$f(x) = \frac{1}{(x+1)(x+2)} = \frac{1}{x-1} - \frac{1}{x+2} \quad x = 0, 1, 2, \ldots$$

Find $P(X \geq 4 \mid x \geq 1)$.

$$\begin{aligned} P(X \geq 4 \mid x \geq 1) &= \frac{P(X \geq 4 \cap X \geq 1)}{P(X \geq 1)} \\ &= \frac{P(X \geq 4)}{P(X \geq 1)} \\ &= \frac{1 - f(0) - f(1) - f(2) - f(3)}{1 - f(0)} \end{aligned}$$

Example 4.3
Consider two nonstandard but fair dice:

$$d_1 \in \{1, 2, 2, 3, 3, 4\}$$

$$d_2 \in \{1, 3, 4, 5, 6, 8\}$$

What's the PMF of $x = d_1 + d_2$?

(d_1, d_2)						
	$(1,4)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$	$(1,8)$
	$(2,1)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$	$(2,8)$
	$(2,1)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$	$(2,8)$
	$(3,1)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$	$(3,8)$
	$(3,1)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$	$(3,8)$
	$(4,1)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$	$(4,8)$

4.2 Expected Value/Expectation

Let X be any random variable. We need to identify two parts:

- What values can X take?
- What are the corresponding probabilities?

Random variables have other properties/invariants we may know that help identify or use them.

Definition 4.4 (Mathematical expectation/expected value)
Let X be a discrete random variable with PMF $f_X(x)$. Then we define mathematical expectation of X or the expected value of X to be

$$\mathbb{E}[X] = \sum_{x \in S} x f(x)$$

Example 4.5
Let $Y = u(x)$ be another random variable. What's the expected value of Y?

$$\mathbb{E}[Y] = \sum_{y} y f_Y(y)$$

$$\mathbb{E}[Y] = \mathbb{E}[u(x)] = \sum_{x \in S} u(x) f_X(x)$$

Example 4.6
Let X be a random variable with support $S_X = \{-1, 0, 1\}$, and $f_X(x) = 1/3$ for $x \in S_X$ and zero otherwise. Consider $u(X) = X^2$. Calculate $\mathbb{E}[u(X)] = \mathbb{E}[X^2]$.
Consider X^2 as a separate random variable.

$$X \in \{-1, 0, 1\} \implies X^2 \in \{0, 1\}, S_{X^2} = \{0, 1\}$$

$f_{X^2}(x) = P(X^2 = X)$ is to be determined for $X = 0, X = 1$.

$$f_{X^2}(0) = P(X^2 = 0) = P(X = 0) = 1/3$$

$$f_{X^2}(1) = P(X^2 = 1) = P(X = -1 \text{ or } X = 1) = 2/3$$

So,

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{y \in S_{X^2}} y \cdot f_{X^2}(y) \\ &= 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right) \\ &= \frac{2}{3}\end{aligned}$$

Alternatively,

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{x \in S_X} x^2 \cdot f_X(x) \\ &= (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \left(\frac{1}{3}\right) + (1)^2 \cdot \frac{1}{3} \\ &= \frac{2}{3}\end{aligned}$$

Theorem 4.7
When it exists, the mathematical expectation $\mathbb{E}$, satisfies

a) If c is constant, then
$$\mathbb{E}[c] = c$$

b) If c is a constant, and u is a function, then
$$\mathbb{E}[c \cdot u(X)] = c\mathbb{E}[u(X)]$$

c) If c_1 and c_2 are constants and u_1 and u_2 are functions, then
$$\mathbb{E}[c_1u_1(X) + c_2u_2(X)] = c_1\mathbb{E}[u_1(X)] + c_2\mathbb{E}[u_2(X)]$$

Proof. Textbook. □

Example 4.8
Consider $u(x) = (x-b)^2$, where $b \in R$. Among those values for which $\mathbb{E}[u(X)]$ exists (if any), which value of b could minimize it?
Let $g(b) \coloneqq \mathbb{E}[(x-b)^2]$, for which b is $g(b)$ is minimized.

$$g(b) = \mathbb{E}[X^2 - 2bX + b^2] = \mathbb{E}[X^2] - 2b\mathbb{E}[X] + b^2$$

$$\frac{d}{db}(g(b)) = 0 \implies \text{critical point}$$

$$\frac{d^2}{db^2}(g(b)) > 0 \implies \text{concave up}$$

$$g'(b) = -2\mathbb{E}[X] + 2b = 0 \implies b = \mathbb{E}[X] = \mu$$

$$g''(b) = 2 > 0 \implies b \text{ is a minimizer}$$

$$\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[(X-b)^2] \quad \forall b \in \mathbb{R}$$

where $\mathbb{E}[(X - \mathbb{E}[X])^2]$ measures how much X deviates from its expected value.

Example 4.9

Consider a coin toss with $P(H) = p \in (0,1)$. Tosses are independent and continue until we get an H. Let X be the number of trials at the same time we stop. The space of X is $S_X = \{1, 2, 3, 4, \dots\}$. Observe,

$$P(T) = 1 - p = q$$

The PMF is

$$f(x) = P(X = x) = q^{x-1}p \quad x \in S_X$$

This is a PMF because,

$$\begin{aligned}
\sum_{x \in S_X} f(x) = f(1) + f(2) + f(3) + \cdots &= \sum_{x=1}^{\infty} q^{x-1}p \\
&= p\sum_{x=1}^{\infty} q^{x-1} \\
&= \underbrace{p\sum_{x=0}^{\infty} q^{x}}_{p\in(0,1) \implies q\in(0,1)} \\
&= p\underbrace{\frac{1}{1-q}}_{p} = 1
\end{aligned}$$

The mean of this geometric distribution is given by

$$\begin{aligned}
q\mu &= \sum_{x=1}^{\infty} xq^x p \\
\mu - q\mu &= \sum_{x=1}^{\infty} xq^{x-1}p - \sum_{x=1}^{\infty} xq^x p \\
\mu(1-q) &= p\sum_{x=1}^{\infty} x(q^{x-1} - q^x) \\
&= (1-q) + 2(q - q^2) + 3(q^2 - q^3) + \dots \\
&= 1 + q + q^2 + q^3 + \cdots \\
&= \frac{1}{1-q} = \frac{1}{p}
\end{aligned}$$

So

$$\mu(1-q) = p \cdot \frac{1}{p} = 1 \implies \mu = \frac{1}{1-q} = \frac{1}{p}$$

Definition 4.10 (Variance, standard deviation (of a random variable))
The variance of a random variable X, denoted $\text{Var}(X)$ is given by

$$\text{Var}(X) = \mathbb{E}[(x-\mu)^2] = \mathbb{E}\underbrace{[x-\mathbb{E}[x])^2]}_{\text{avg squared distance}}$$

The standard deviation of X is

$$SD(X) = \sqrt{\text{Var}(X)}$$

$$\sigma^2 = \text{Var}(X)$$

We can show,

$$\begin{aligned}\sigma^2 &= \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2 \\ &= \mathbb{E}[X^2] - [\mathbb{E}(X)]^2\end{aligned}$$

Example 4.11
Let X be the random variable with $S_X = \{1, 2, \ldots, m\}$ and

$$f(x) = \frac{1}{m} \quad \forall x \in S_X$$

and $X \sim \text{Uniform}(\{1, 2, \ldots, m\})$.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=1}^{m} x \cdot f(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m}(1 + 2 + \cdots + n) \\ &= \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2} \end{aligned}$$

Now compute $\mathbb{E}[X^2]$,

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=1}^{m} x^2 \cdot f(x) = \sum_{x=1}^{m} x^2 \cdot \frac{1}{m} \\ &= \frac{1}{m} \sum_{x=1}^{m} x^2 = \frac{1}{m} \left(\frac{m \cdot (m+1) \cdot (2m+1)}{6} \right) \\ &= \frac{(m+1)(2m+1)}{6} \end{aligned}$$

Therefore,

$$\begin{aligned} \sigma^2 = \text{Var}(X) = \mathbb{E}[X^2] - \mu^2 &= \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2} \right)^2 \\ &= (m+1) \cdot \left(\frac{2m+1}{6} - \frac{m+1}{4} \right) \\ &= (m+1) \frac{(4m+2) - (3m+3)}{12} \\ &= \frac{(m+1)(m-1)}{12} \end{aligned}$$

Therefore, $X \sim \text{Uniform}(\{1, 2, \ldots, m\})$ implies

$$\mu = \frac{m+1}{2}$$

$$\sigma^2 = \frac{m^2 - 1}{12}$$

$$X \underset{\text{tranform}}{\rightarrow} h(X)$$

Imagine

$$Y = aX + b \quad a, b \in \mathbb{R}$$

So,

$$\mu_Y = \mathbb{E}[Y] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = a\mu_X + b$$

And

$$\begin{aligned}\sigma_{Y^2} = \mathrm{Var}(Y) &= \mathbb{E}[(Y-\mu_Y)^2] \\ &= \mathbb{E}[(aX+b-(a\mu_X+b))^2] \\ &= \mathbb{E}[(aX-a\mu_X)^2] \\ &= a^2\mathbb{E}[(X-\mu_X)^2] \\ &= a^2\cdot\mathrm{Var}(X) \\ &= a^2\sigma_X^2\end{aligned}$$

So

$$\begin{aligned}\mu_Y &= a\mu_X + b \\ \sigma_Y^2 &= a^2\cdot\sigma_X^2\end{aligned}$$

Definition 4.12 (Moment)
$\mathbb{E}[X^k]$ is known as the <u>kth raw moment</u> and $\mathbb{E}[(X-\mu_X)^k]$ is known as the <u>kth central moment</u>.

5 Jul 1, 2022

5.1 Moment Generating Functions

Definition 5.1 (Moment generating functions (MGF))
Let X be a random variable of the discrete type with pmf $f(x)$ and space S. If there is a positive number h such that

$$\mathbb{E}[e^{tx}] = \sum_{x \in S} e^{tx} f(x)$$

exists and is finite for $-h < t < h$, then the function defined by

$$M(t) = \mathbb{E}(e^{tX})$$

is called the moment-generating function of X.

Theorem 5.2
If the moment-generating function exists, then

$$M'(0) = \mathbb{E}[X] = \mu$$

$$M''(0) - [M'(0)]^2 = \mathbb{E}[X^2] - [E[X]]^2 = \sigma^2$$

Example 5.3
Let x equal the number of flips of a fair coin that is required to observe the same face on consecutive flips.

1. Find the PMF of X.

$$S_X = \{2, 3, 4, 5, \dots\}$$

   $k \in S_X$. So

$$f(k) = \frac{1}{2^k} + \frac{1}{2^k} = \frac{1}{2^{k-1}}$$

2. Find $M(t)$.

$$\mathbb{E}[e^{tx}] = \sum_{x=2}^{\infty} e^{tx} \cdot \frac{1}{2^{x-1}} = 2 \sum_{x=2}^{\infty} \left(\frac{e^t}{2}\right)^x$$
$$= 2\left(\frac{1}{1 - \frac{e^t}{2}} - 1 - \frac{e^t}{2}\right)$$

So the moment generating function is

$$\begin{aligned}
M(t) &= \frac{2}{1-\frac{e^t}{2}} - 2 - e^t \\
&= 2\left(1-\frac{e^t}{2}\right)^{-2} \cdot \left(\frac{-e^t}{2}\right) - e^t \\
&= \frac{e^t}{\left(1-\frac{e^t}{2}\right)^2} - e^t \\
&= e^t\left(\frac{1}{\left(1-\frac{e^t}{2}\right)^2} - 1\right)
\end{aligned}$$

3. Find $\mathbb{E}[X]$ and $\operatorname{Var}(X)$.

$$\mathbb{E}[X] = M'(0) = 1\left(\frac{1}{\left(1-\frac{1}{2}\right)^2} - 1\right) = 4 - 1 = 3$$

And

$$\begin{aligned}
M''(t) &= e^t\left(\frac{1}{\left(1-\frac{e^t}{2}\right)^2} - 1\right) + e^t \cdot (-2) \cdot \frac{1}{\left(1-\frac{e^t}{2}\right)^3} \cdot \frac{-e^t}{2} \\
&= e^t\left(\frac{1}{\left(1-\frac{e^t}{2}\right)^2} - 1\right) + \frac{e^{2t}}{\left(1-\frac{e^t}{2}\right)^3}
\end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E}[X^2] = M''(0) &= \left(\frac{1}{\left(\frac{1}{2}\right)^2} - 1\right) + \frac{1}{\left(1-\frac{1}{2}\right)^3} \\
&= (4-1) + 8 \\
&= 11
\end{aligned}$$

Therefore,

$$\operatorname{Var}(X) = 11 - 3^2 = 2$$

Hence,

$$\mathbb{E}[X] = 3 \quad \operatorname{Var}(X) = 2$$

4. Find $P(X \le 3), P(X \ge 5)$, and $P(X = 3)$.

$$P(X \le 3) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

$$P(X \ge 5) = 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right) = \frac{1}{8}$$

$$P(X = 3) = \frac{1}{4}$$

5.2 The Binomial Distribution

Definition 5.4 (Bernoulli Distribution)
A random trial where $X(\text{Success}) = 1$ and $X(\text{failure}) = 0$. The space of X is $\text{Space}(X) = \{0, 1\}$. The pmf is

$$f(x) = p^x(1-p)^{1-x} \quad x \in \{0, 1\}$$

The notation is $X \sim \text{Bernoulli}(p)$.

Note 5.5:

$$\mathbb{E}[X] = 0(1-p) + 1 \cdot p = p$$

$$\mathbb{E}[X^2] = 0^2(1-p) + 1^2 \cdot p = p$$

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p) = pq$$

Example 5.6
Suppose we have instant lottery tickets and with each ticket, we either win or lose.

$$P(\text{Win}) = 0.2$$

$$P(\text{Loss}) = 0.8$$

If 5 tickets are purchased,

$$P(0, 0, 0, 1, 0) = (0.8)(0.8)(0.8)(0.2)(0.8) = (0.8)^4(0.2)$$

Definition 5.7 (Binomial distribution)
$X \sim \text{Binomial}(n, p)$, where we have n trials and the probability of success, p, is independent, the space of X is $\text{Space}(X) = \{0, 1, 2, \dots, n\}$, and the pmf is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \dots, n\}$$

Example 5.8
We have 5 lottery tickets, each independent with probability of success 0.2. Then

$$P(0, 0, 0, 1, 0) = (0.8)^4(0.2)$$

$$P(1 \text{ Winning}) = \binom{5}{1}(0.8)^4(0.2)$$

$$P(2\text{Winning}) = \binom{5}{2}(0.8)^3(0.2)^2$$

Example 5.9
Mithrandir is practicing a difficult spell with probability of success 0.1. He exercises for 100 times and each casting attempt is independent.

$$\begin{aligned} P(X \leq 98) &= P(X=0) + P(X=1) + \cdots = P(X=98) \\ &= 1 - P(X > 98) \\ &= 1 - P(X=99) - P(X=100) \\ &= 1 - \binom{100}{99}(0.1)^{99}(0.9)^1 - \binom{100}{100}(0.1)^{100}(0.9)^0 \end{aligned}$$

Note 5.10: Let $x \in \mathbb{R}^+$, then the CDF of a binomial distribution is

$$F(x) = P(X \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}.$$

For $x \in \mathbb{R}^-$,

$$F(x) = 0$$

for $x \in (n, \infty)$,

$$F(x) = 1$$

Theorem 5.11
If $X \sim \text{Binomial}(n, p)$, then

$$\mathbb{E}[X] = np$$

$$\text{Var}(X) = np(1-p).$$

And its MGF is

$$M(t) = (1 - p + pe^t)^n.$$

Proof. Find $\mathbb{E}[X]$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} p^{k+1} (1-p)^{n-k-1} \\
&= \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} p^{k+1} (1-p)^{n-k-1} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^{k} (1-p)^{n-k-1} \\
&= np \cdot (p + (1-p))^{n-1} = np \cdot 1^{n-1} = np
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n-2} \frac{n!}{k!(n-k-2)!} p^{k+2} (1-p)^{n-k-2} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-k-2)!} p^{k} (1-p)^{n-k-2} \\
&= n(n-1)p^2 \cdot (p + (1-p))^{n-2} \\
&= n(n-1)p^2
\end{aligned}$$

Next,

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

Hence,

$$\begin{aligned}\operatorname{Var}(X) &= n(n-1)p^2 + np - (np)^2 \\ &= n^2p^2 - np^2 + np - (np)^2 \\ &= np - np^2 \\ &= np(1-p) \\ &= npq\end{aligned}$$

Finally, the MGF is

$$\begin{aligned}M_X(t) &= \mathbb{E}[e^{tx}] \\ &= \sum_{k=0}^{n} e^{kt}\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \\ &= (1-p+pe^t)^n\end{aligned}$$

□

5.3 The Negative Binomial Distribution

Recall 5.12 In a binomial distribution, we fix the number of trials and count the number of successes.

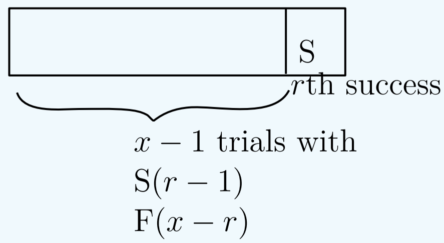
What if we fix the number of successes and count the number of trials to achieve?

Definition 5.13 (Negative binomial distribution)
If $X \sim NB(r, p)$, where r is the number of successes that we fix, and p is the probability of a success, and is independent, then the values X takes is

$$\text{Space}(X) = \{r, r+1, r+2, \dots\}.$$

If we have r successes and x trials, then the values x can take is $x = r, r+1, r+2, \dots$.

S r
F $x - r$



So,

$$\# \text{ arrangements } = \binom{x-1}{r-1}$$

Hence, the pmf is

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, r+2, \dots$$

Definition 5.14 (Geometric distribution)
The special case $r = 1$ is called a geometric random variable. $X \sim \text{Geometric}(p)$, where the space is $\text{Space}(X) = \{1, 2, 3, \dots\}$, and the pmf is

$$f(x) = (1-p)^{x-1} p \quad x = 1, 2, 3, \dots$$

Note 5.15:
$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} p = p \sum_{x=1}^{\infty} (1-p)^{x-1} = p \cdot \frac{1}{1-(1-p)} = 1$$

And,

$$\begin{aligned} P(X > k) &= P(\text{1st success happens after } k \text{ trials}) \\ &= P(\text{first } k \text{ trials are failures}) \\ &= (1-p)^k \end{aligned}$$

6 Jul 6, 2022

6.1 The Poisson Distribution

Definition 6.1 (Poisson process)
Let number of occurrences of some event in a given continuous interval be counted. Then we have an approximate Poisson process with parameter $\lambda > 0$ if the following conditions are satisfied:

a) Number of occurrences in non-overlapping subintervals are independent.

b) Probability of exactly one occurrence in a short subinterval of length h is approximately λh.

c) Probability of two or more occurrences in a short subinterval is essentially zero.

Let $X \sim B(n, \lambda/n)$ so

$$P(X = x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

What happens as $n \to \infty$?

$$\begin{aligned}\lim_{n\to\infty} P(X = x) &= \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\frac{\lambda^x}{n^x}\left(1 - \frac{\lambda}{n}\right)^n\left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \lim_{n\to\infty} \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \cdot \frac{\lambda^x}{x!}\underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\to 1} \\ &= \frac{e^{-\lambda}\lambda^x}{x!}\end{aligned}$$

Observe,

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda \cdot \lambda^x}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$

The MGF is

$$\begin{aligned}M(t) &= \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda}e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

Hence,

$$\text{Var}(X) = \lambda$$

$$\mathbb{E}[X] = \lambda$$

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

Example 6.2
Let $X \sim \text{Poisson}(5)$. Calculate $F(6)$.

$$F(6) = P(X \leq 6) = \sum_{x=0}^{6} P(X = x) = e^{-5} \sum_{x=0}^{6} \frac{5^x}{x!} = 0.762$$

6.2 Continuous Distributions

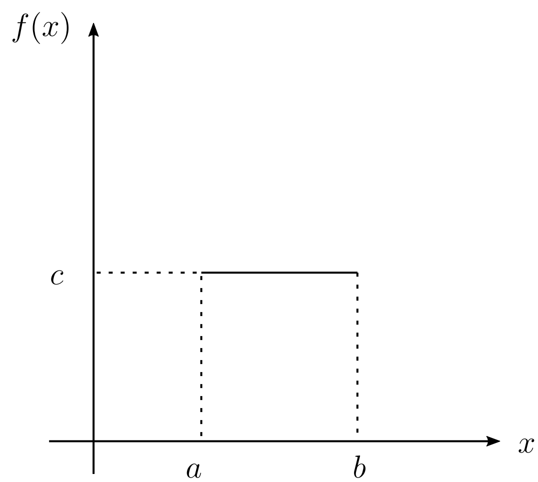
Let $X \colon S \to \mathbb{R}$ be a random variable, where $\text{Range}(X) = S_X$ is the support of X. The discrete random variable S_X is finite and countable infinite, so

$$\sum_{x \in S_X} f(x) = \sup_{\substack{N \subseteq S_X \\ \text{countable}}} \sum_{x \in N} f(x) > M, \quad \forall M \in R$$

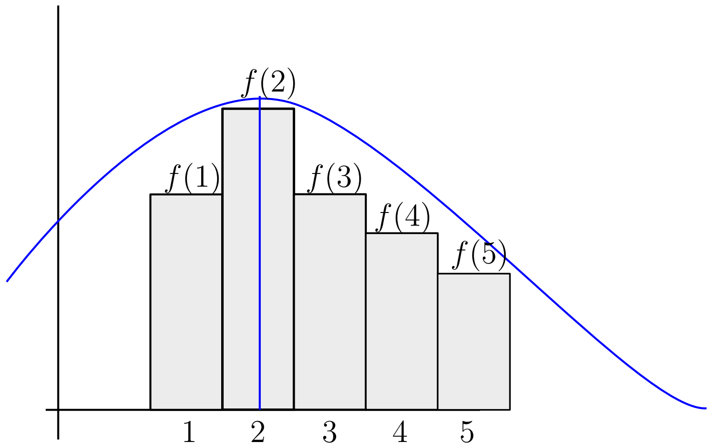
Example 6.3
Consider $S_X = [a, b]$ and x is a randomly selected real number from S_X.

$$P(x \in [c, d]) = \frac{d - c}{b - a}$$

How dense a random variable at any place over its support is expressed by its probability density function (pdf).



For discrete pmf, recall the histogram:



The area under the pdf of a continuous random variable must equal to 1.

$$c(b-a) = 1$$

$$c = \frac{1}{b-a}$$

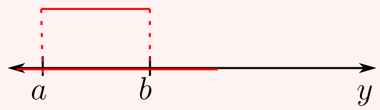
Definition 6.4 (Uniform distribution)
If $X \sim \text{Uniform}[a, b]$, then

i. $S_X = [a, b]$

ii. $f(x) = \dfrac{1}{b-a}$ for $x \in [a, b]$, and 0 otherwise

Question 6.5: What about CDF?

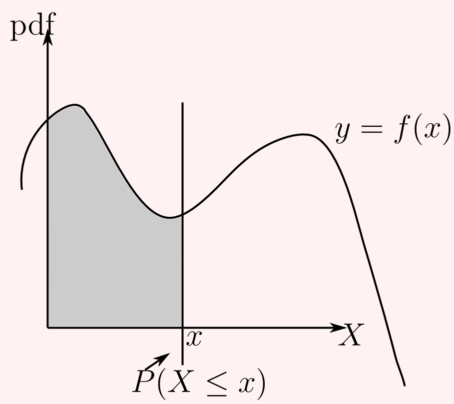
Example 6.6 (Uniform distribution)
Suppose $F(x) = P(X \le x)$, where $x \in \mathbb{R}$.



$$S_x = [a, b] \text{ so } \begin{cases} F(x) = 0 \text{ for } x < a \\ F(x) = 1 \text{ for } x \ge b \end{cases}$$

If $x \in [a, b]$, then $F(x) = \dfrac{x-a}{b-a}$. So,

CDF	PDF
$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$	$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$



So,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Definition 6.7 (Continuous random variable)
Any function $f\colon S_X \to \mathbb{R}$ will be a pdf of a random variable X if

i. $f(x) \geq 0 \quad \forall x \in S_X$

ii. $\int_{S_X} f(x)dx = 1$

iii. For any interval $(a, b) \subseteq S$,

$$P(a < X < b) = \int_a^b f(x)dx$$

X is called a <u>continuous random variable</u>.

Example 6.8

Let Y be a continuous random variable with pdf $g(y) = 2y$, $0 < y < 1$. So $S_Y = (0, 1)$. The cdf is defined by:

$$\begin{cases} \text{if } y < 0, G(y) = 0 \\ \text{if } y \geq 1, G(y) = 1 \end{cases}$$

If $y \in [0, 1)$, then

$$\begin{aligned} G(y) = \int_{-\infty}^{y} g(t)dt &= \int_{0}^{y} 2tdt \\ &= t^2 \Big|_0^y \\ &= y^2 - 0^2 = y^2 \end{aligned}$$

So,

$$G(y) = \begin{cases} 0 & \text{if } y < 0 \\ y^2 & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

So,

$$\begin{aligned} P\left(\frac{1}{2} < Y \leq \frac{3}{4}\right) &= P\left(Y \leq \frac{3}{4}\right) - P\left(Y \leq \frac{1}{2}\right) \\ &= G\left(\frac{3}{4}\right) - G\left(\frac{1}{2}\right) \\ &= \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 \\ &= \frac{5}{16} \end{aligned}$$

Now find $P\left(Y = \frac{3}{4}\right)$.

$$P\left(Y = \frac{3}{4}\right) = \int_{\frac{3}{4}}^{\frac{3}{4}} g(t)dt = 0$$

for any continuous random variable, $P(Y = c) = 0$ for all $c \in \mathbb{R}$. Finally,

$$\begin{aligned} P\left(\frac{1}{4} \leq Y < 2\right) = P\left(\frac{1}{4} < Y < 2\right) &= P\left(\frac{1}{4} \leq Y \leq 2\right) \\ &= G(2) - G\left(\frac{1}{4}\right) \\ &= 1^2 - \left(\frac{1}{4}\right)^2 = \frac{15}{16} \end{aligned}$$

7 Jul 8, 2022

7.1 Continuous Distribution

Recall 7.1 The expected value is given by:

$$\mu = E[X]$$

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

while the variance is given by:

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

So,

$$E[\mu(X)] = \int_{-\infty}^{\infty} \mu(X) f(x) dx = \int_{S_X} \mu(X) \cdot f(x) dx$$

The MGF is given by

$$M_X(t) = E[e^{tx}] = \int_{S_x} e^{tx} f(x) dx$$

Example 7.2
Suppose that $S_Y = (0,1)$ and

$$g(y) = 2y, \quad 0 < y < 1$$

then,

$$\begin{aligned}\mu = E[Y] &= \int_{-\infty}^{\infty} y g(y) dy \\ &= \int_0^1 y 2y \cdot dy = \int_0^1 2y^2 dy \\ &= \left[\frac{2y^3}{3}\right]_0^1 = \frac{2}{3} \\ E[Y^2] &= \int_0^1 y^2 \cdot 2y dy = \int_0^1 2y^3 dy = \left[\frac{y^4}{2}\right]_0^1 \\ &= \frac{1}{2} \\ \sigma^2 = \text{Var}(Y) &= \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}\end{aligned}$$

Example 7.3
Let

$$f(x) = |x|, \quad \text{for } -1 < x < 1$$

then,

$$\begin{aligned}
P\left(-\frac{1}{2} < X < \frac{3}{4}\right) &= \int_{-1/2}^{3} |x| dx = \int_{-1/2}^{0} (-x) dx + \int_{0}^{3/4} x dx \\
&= \left[-\frac{x^2}{2}\right]_{-1/2}^{0} + \left[\frac{x^2}{2}\right]_{0}^{3/4} \\
&= \frac{13}{32} \\
E[X] &= \int_{-1}^{1} x \cdot |x| dx = \int_{-1}^{0} x(-x) dx + \int_{0}^{1} x \cdot x dx \\
&= \left(-\frac{x^3}{3}\right)_{-1}^{0} + \left(\frac{x^3}{3}\right)_{x=0}^{x=1} \\
&= 0 \\
\operatorname{Var}(X) &= E[(X-\mu)^2] = E[X^2] \\
&= \int_{-1}^{0} -x^3 dx + \int_{0}^{1} x^3 dx \\
&= \frac{1}{2}
\end{aligned}$$

Definition 7.4 (Percentile)
The $\underline{100p^{th} \text{ percentile}}$ is a number π_p such that the area under $f(x)$ to the left of π_p is p:

$$p = \int_{-\infty}^{\pi_p} f(x) dx = F(\pi_p)$$

The 50^{th} percentile is called the median while the 25^{th} and 75^{th} percentiles are called the first and third quartiles, respectively.

7.2 Exponential Distributions

Definition 7.5 (Exponential distribution)
Let Let W be a random variable such that $S_w = (0, \infty)$ and

$$F_W(w) = \begin{cases} 0 & \text{if } w < 0 \\ 1 - e^{-\lambda w} & \text{if } w \geq 0 \end{cases}$$

Then $W \sim \exp(\lambda)$ is called an exponential random variable.

λ is called a parameter of w, so the reparametrization is $\theta = \frac{1}{\lambda}$.

Definition 7.6 (Exponential random variable)
Let X be a random variable such that $S_X = (0, \infty)$, and

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x/\theta} & \text{if } x \geq 0 \end{cases}$$

then $X \sim \exp(\theta)$ is called an exponential random variable.

So why use $\theta = \frac{1}{\lambda}$ instead of λ? Poisson(λ) is the rate of occurrence, so as λ increases, the number of occurrences increases while the time of the first occurrence decreases.

- The number of occurrences and time to the first occurrence are inversely related.
- For exponential random variables, it makes sense to use the inverse parameter $\theta = 1/\lambda$.

So if $X \sim \exp(\theta)$, then

$$F_X(x) = 1 - e^{-x/\theta} \quad x \geq 0$$

therefore,

$$f_X(x) = F'_X(x) = -e^{-x/\theta} \cdot \left(-\frac{1}{\theta}\right)$$

so,

$$f_X(x) = \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}}, \quad x \geq 0$$

The MGF is

$$\begin{aligned} M_X(t) &= E[e^{tx}] = \int_0^\infty e^{tx} \cdot \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}} dx \\ M_X(t) &= \frac{1}{\theta} \cdot \int_0^\infty e^{x \cdot \left(t - \frac{1}{\theta}\right)} dx \\ &= \frac{1}{\theta \cdot \left(t - \frac{1}{\theta}\right)} \lim_{b \to \infty} \left[e^{x\left(t - \frac{1}{\theta}\right)} \right]_{x=0}^{x=b} \\ &= \frac{1}{t\theta - 1} \lim_{b \to \infty} \left(\underbrace{e^{b(t - 1/\theta)}}_{\text{only exists if } t - 1/\theta < 0} - 1 \right) \\ &= \frac{1}{t\theta - 1}(0 - 1) \\ &= \frac{1}{1 - t\theta} \end{aligned}$$

So,

$$\begin{aligned} M'_X(t) = \frac{d}{dt}(1 - t\theta)^{-1} &= (-1) \cdot (1 - t\theta)^{-2} \cdot (-\theta) \\ &= \frac{\theta}{(1 - t\theta)^2} \end{aligned}$$

So,

$$E[X] = \mu = M_X'(0) = \frac{\theta}{(1-0)^2} = \theta$$

And,

$$\begin{aligned} M_X''(t) = \frac{d}{dt}\Big(\theta \cdot (1-t\theta)^{-2}\Big) &= (-2)\theta \cdot (1-t\theta)^{-3} \cdot (-\theta) \\ &= \frac{2\theta^2}{(1-t\theta)^3} \end{aligned}$$

Hence

$$E[X^2] = M_X''(0) = 2\theta^2 \implies \mathrm{Var}(X) = 2\theta^2 - \theta^2 = \theta^2$$

Theorem 7.7
Let $X \sim \exp(\theta)$, then

$$E[X] = \theta$$
$$\mathrm{Var}(X) = \theta^2$$

Note: With the original parametrization, if $Y \sim \exp(\lambda)$, then

$$E[Y] = \frac{1}{\lambda}$$
$$\mathrm{Var}(Y) = \frac{1}{\lambda^2}$$

Note 7.8: If $\lambda = 7$ is the average number of occurrences within the unit interval, then $\theta = 1/7$ is the average time to the first occurrence.

Example 7.9

Suppose $X \sim \exp(20)$, then

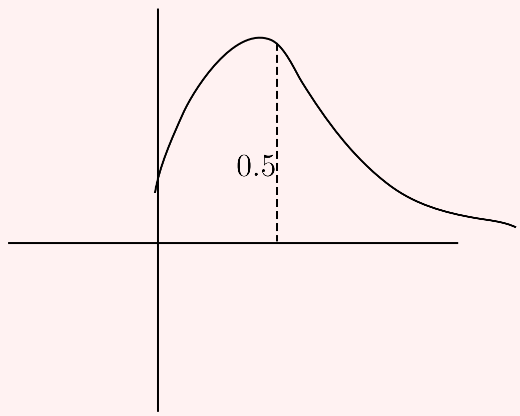
$$f_X(x) = \frac{1}{20}e^{-x/20}, \quad \text{for } x \geq 0$$

$$F_X(x) = 1 - e^{-x/20}$$

Then,

$$\begin{aligned} P(X \geq 15) &= 1 - P(X < 15) = 1 - (P(X \leq 15) - \underbrace{P(X = 15)}_{0}) \\ &= 1 - F_X(15) = e^{-15/20} \approx 0.5276 \end{aligned}$$

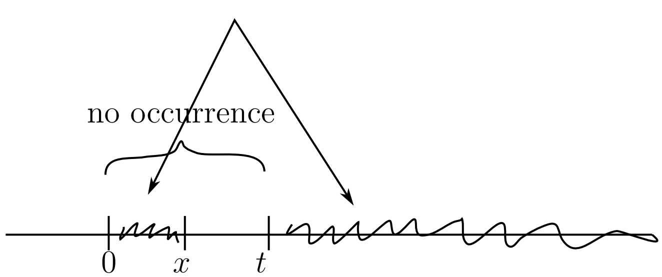
What's the median m?



We have

$$\begin{aligned} F_X(m) &= 0.5 \\ 1 - e^{-m/20} &= 0.5 \\ 0.5 &= e^{-m/20} \\ -\ln 2 &= -m/20 \\ m &= 20\ln(2) \\ m &\approx 13.8629 \end{aligned}$$

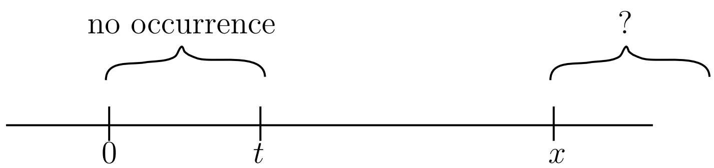
Say we know for some $t > 0$, no occurrence has happened yet.



Then,

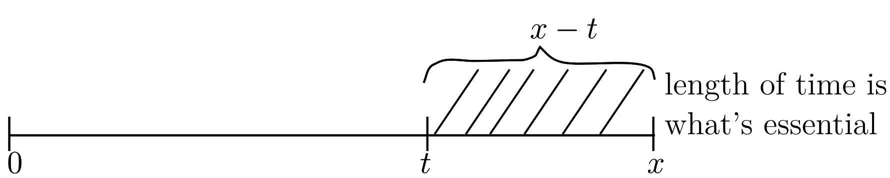
$$P(X < x | X > t) = 0 \quad \text{for } x \leq t$$

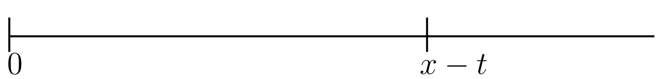
What about $x > t$?



Hence,

$$\begin{aligned} P(X > x | X > t) &= \frac{P(X > x \cap X > t)}{P(X > t)} \\ &= \frac{P(X > x)}{P(X > t} \\ &= \frac{1 - F_X(x)}{1 - F_X(t)} \\ &= \frac{1 - (1 - e^{-x/\theta})}{1 - (1 - e^{-t/\theta})} \\ &= \frac{e^{-x/\theta}}{e^{-t/\theta}} \\ &= e^{-(x-t)/\theta} \\ &= P(X > x - t) \end{aligned}$$





which is called the memorylessness property.

The exponential random variable models time to first failure, independent from how much time has elapsed.