Math 105 Chapter 9: Probability

Probability theory is the study of randomness and chance. We want to quantify the likeliness of an event occuring.

eg. What is the chance of it raining today?

What is the chance of me flipping a head in a coin toss?

What is the chance of a man in Vancouver is 6ft tall?

Probability theory aim to answer these types of question and much more.

Note: This is a very deep area of math and we will be just scratching the surface.

Before we begin let's define what an event is.

Definition: An event is a possible outcome occuring. Given an event A, we define the compliment of the event to be the event that A does not occur, it is denoted by A^c.

$$A \quad , \quad A^c$$

A occurs $\qquad$ A does not occur.

eg Let A denote event that a person off the street is at least 6 ft tall

Then A^c is the event that a person off the street is taller than 6ft.

eg. If I toss a coin 2 times, then an example of an event is

$A = 2$ heads come up in a row

$A^c = 2$ heads dont up

$=$ atleast one tail show s up.

Definition: Given an even A, we define the probability of A occuring to be

$$\mathbb{P}(A)$$

And $\mathbb{P}(A)$ satisfies:

① $0 \leq \mathbb{P}(A) \leq 1$

② If A, B are mutually exclusive (i.e can't happen at same time) then

$$\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B)$$

③ If A occurs implies B occurs, then

$$\mathbb{P}(A) \leq \mathbb{P}(B)$$

Remarks:

- A probability is a measure of how likely an event A occuring is.

$\mathbb{P}(A) = 1$ means "A will always occur."

$\mathbb{P}(A) = 0$ means "A will never occur."

- ① says $\mathbb{P}(A)$ is always somewhere in one of the above extremes.
- ② says that if 2 events can't occur at the same time, then the probability of one of them occuring is the sum of the probabilities.

- A special case of ② is $A = A$, $B = A^c$. then both can't occur at the same time, But one of those events always occurs.

Thus $\quad 1 = \mathbb{P}(A \text{ or } A^c)$

$\quad = \mathbb{P}(A) + \mathbb{P}(A^c)$, since A, A^c are mutually exclusive

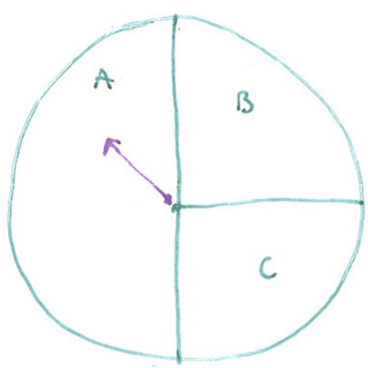
So $\quad \boxed{\mathbb{P}(A^c) = 1 - \mathbb{P}(A)}$ ← important

• ③ says that if A happens implies B happens then B is more likely than A.

eg. A = man is at least 5 ft tall , B = man is at least 6 ft tall.

clearly A implies B so $\mathbb{P}(A) \leq \mathbb{P}(B)$.

eg Suppose I have a spinner below



Assuming the spinner is equally likely to point in each direction, then

$\mathbb{P}(\text{lands in } A) = \frac{1}{2}$

$\mathbb{P}(\text{lands in } B) = \frac{1}{4}$

$\mathbb{P}(\text{lands in } A \text{ or } C) = \mathbb{P}(\text{lands in } A) + \mathbb{P}(\text{lands in } C)$

$= \frac{1}{2} + \frac{1}{4}$

$= \frac{3}{4}$

$\mathbb{P}(\text{lands in } D) = 0$, since it will never land in D because it is not on the spinner.

$$\mathbb{P}(\text{lands in right half of circle})$$

$$= \mathbb{P}(\text{lands in B or C})$$

$$= \mathbb{P}(B) + \mathbb{P}(C)$$

$$= \frac{1}{4} + \frac{1}{4}$$

$$= \frac{1}{2}$$

We now move onto a concept that incredibly important to the study of probability theory.

<u>Definition:</u> A <u>random variable</u> is a "random number" that represents the outcome of some experiment.

eg The height of random person down the street is a random variable.

Let H denote the hieght of a person on the street.

$$\mathbb{P}(H \leq 6\text{ ft}) = \mathbb{P}(\text{height of person is atleast 6 ft})$$

eg Suppose I toss a fair coin 2 times.

Let X denote the number of heads.

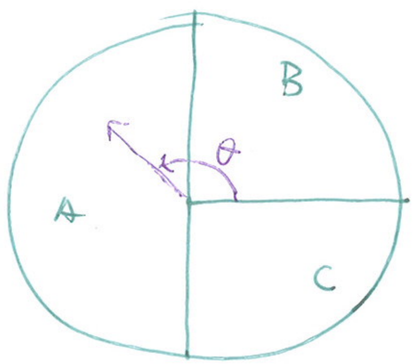
$$\mathbb{P}(X=2) = \mathbb{P}(\text{HH occurs})$$
$$= \frac{1}{4}$$

$$P(X = 1 = \mathbb{P}(\text{HT or TH occurs})$$
$$= \mathbb{P}(\text{HT occurs}) + \mathbb{P}(\text{TH occurs})$$
$$= \frac{1}{4} + \frac{1}{4}$$
$$= \frac{1}{2}$$

$$\mathbb{P}(X=0) = 1 - \mathbb{P}(X \geq 1)$$
$$= 1 - \left[\mathbb{P}(X=1) + \mathbb{P}(X=2)\right]$$
$$= 1 - \left[\frac{1}{2} + \frac{1}{4}\right]$$
$$= \frac{1}{4}$$

eg. Let us go back to the spinner.

Let θ denote the angle made from the positive x-axis, $\theta \in [0, 2\pi)$.



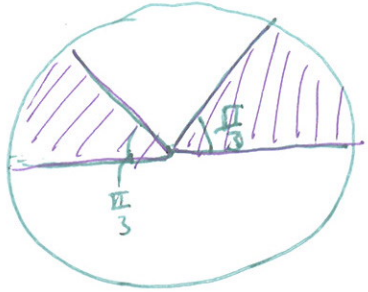
So θ is a random variable between 0 and 2π.

$$\mathbb{P}\left(\frac{\pi}{2} < \theta < \frac{3\pi}{2}\right) = \mathbb{P}(A \text{ occurs})$$
$$= \frac{1}{2}$$

$$\mathbb{P}(\theta > 10\pi) = 0 \quad , \text{ since } 0 \leq \theta < 2\pi$$

$$\mathbb{P}\left(\frac{\pi}{4} < \theta < \frac{3\pi}{4}\right) = \frac{\frac{3\pi}{4} - \frac{\pi}{4}}{2\pi}$$
$$= \frac{1}{4}$$

$$\mathbb{P}\left(0 \leq \sin\theta \leq \frac{\sqrt{3}}{2}\right) = \mathbb{P}\left(0 \leq \theta \leq \frac{\pi}{3} \text{ or } \frac{2\pi}{3} \leq \theta < \pi\right)$$



So $\mathbb{P}\left(0 \leq \sin\theta \leq \frac{\sqrt{3}}{2}\right) = \%$ of circle shaded

$$= \frac{\pi/3 + \pi/3}{2\pi}$$

$$= \frac{1}{3}$$

We will now define a function that is very useful to study random variables.

Definition! Given a random variable X we define the cumulative distribution function (CDF) of X by

$$F_X(t) = \mathbb{P}(X \leq t)$$

So lets compute some CDF's for 2 of the examples above.

eg If I flip 2 fair coins and X is the number of heads then lets find F_X. Now there are a few cases.

- If $t < 0$ then

$$F_X(t) = \mathbb{P}(X \leq t)$$
$$= 0$$

since you cant have a negative number of heads.

• If $0 \le t < 1$ then

$$F_X(t) = \mathbb{P}(X \le t) = \mathbb{P}(X \le 0) = \frac{1}{4}$$

• If $1 \le t < 2$ then

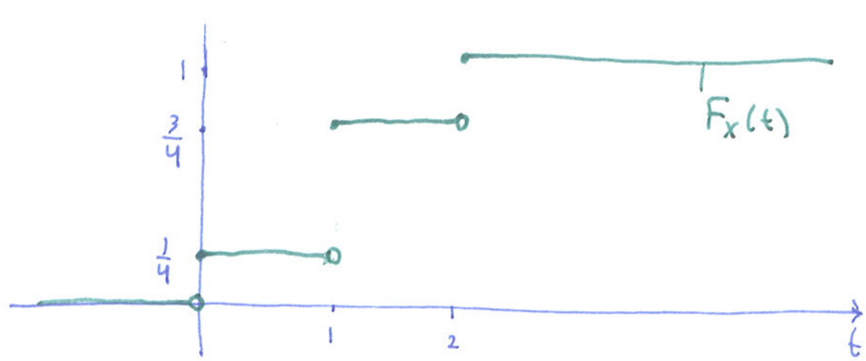
$$F_X(t) = \mathbb{P}(X \le t) = \mathbb{P}(X \le 1) = \mathbb{P}(X=0) + \mathbb{P}(X=1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

• If $2 \le t$ then

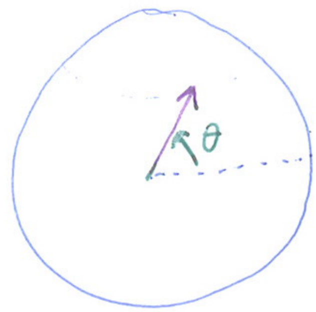
$$F_X(t) = \mathbb{P}(X \le t) = \mathbb{P}(X \le 2) = 1$$

So

$$F_X(t) = \begin{cases} 0 & , t<0 \\ \frac{1}{4} & , 0 \le t < 1 \\ \frac{3}{4} & , 1 \le t < 2 \\ 1 & , 2 \le t \end{cases}$$



eg Back to our spinner example we have θ defined as the angle between $0 \leq \theta < 2\pi$ of the spinner from the positive real axis.



Again we have a few cases.

- If $t < 0$ then

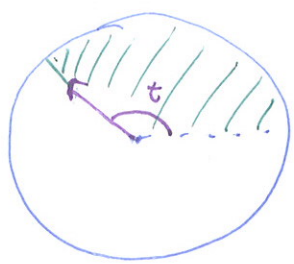
$$F_\theta(t) = \mathbb{P}(\theta \leq t) \\ = 0$$

- If $t \geq 2\pi$ then

$$F_\theta(t) = \mathbb{P}(\theta \leq t) \\ = \mathbb{P}(\theta \leq 2\pi) \\ = 1$$

So the interesting case is when $0 \leq t < 2\pi$

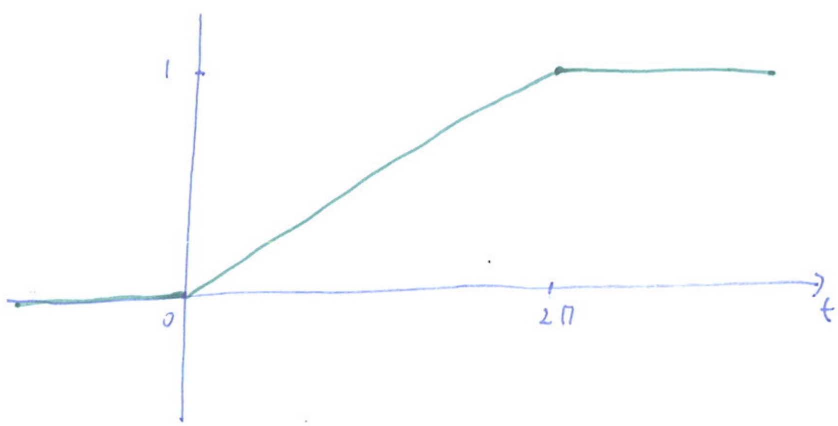
- If $0 \leq t < 2\pi$ then $\theta \leq t$ implies the spinner landed in the shaded section.



So we have

$$F_\theta(t) = \mathbb{P}(\theta \le t)$$
$$= \text{\% of region shaded}$$
$$= \frac{t}{2\pi}$$

Thus
$$F_\theta(t) = \begin{cases} 0 & , \ t < 0 \\ \frac{t}{2\pi} & , \ 0 \le t < 2\pi \\ 1 & , \ 2\pi < t \end{cases}$$



Definition: We say a random variable X is

- Continous if $F_X(t)$ is continuous.
- Discrete if $F_X(t)$ has a jump discontinuity.

So X was discrete, and θ was continous. From here on out we will only be working with continous random variables.

Now let's talk about some properties of CDF's.

Properties of CDF's:

Let X be a random variable with CDF $F_X(t)$.

① For all $t \in \mathbb{R}$,

$$0 \leq F(t) \leq 1$$

② • $\lim_{t \to \infty} F_X(t) = 1$

• $\lim_{t \to -\infty} F_X(t) = 0$

③ $F_X(t)$ is non-decreasing

theorem: Given a $F(t)$ that satisfies the above 3 properties, there is a random variable X such that $F(t)$ is the CDF of X.

That is a deep result that tells us that if we study the CDF's we can study random variables.

④ $P(a < X \leq b) = F(b) - F(a)$, $a < b$.

⑤ If X is cts then

$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= \mathbb{P}(a \leq X < b) \\ &= \mathbb{P}(a < X \leq b) \\ &= \mathbb{P}(a < X < b) \\ &= F(b) - F(a) \end{aligned}$$

Proof: ① $F_X(t) = \mathbb{P}(X \le t) \in [0,1]$

② • $\lim_{t\to\infty} F_X(t) = \lim_{t\to\infty} \mathbb{P}(X \le t)$

$= \mathbb{P}(X < \infty)$

$= 1$

• $\lim_{t\to-\infty} F_X(t) = \lim_{t\to-\infty} \mathbb{P}(X \le t)$

$= \mathbb{P}(X \le -\infty)$

$= 0$

③ If $s < t$ then $X \le s$ implies $X \le t$, so

$F_X(s) = \mathbb{P}(X \le s)$

$\le \mathbb{P}(X \le t)$ since $X \le s \Rightarrow X \le t$

$= F_X(t)$

④ We have if $a < b$

$F_X(b) = \mathbb{P}(X \le b)$

$= \mathbb{P}(X \le a \text{ or } a < X \le b)$

$= \mathbb{P}(X \le a) + \mathbb{P}(a < X \le b)$

$= F_X(a) + \mathbb{P}(a < X \le b)$

⑤ is a deep fact and is thus excluded.

eg. Let $F(x) = \frac{1}{1+e^{-kx}}$ for $k>0$.

To check that F is a CDF

① Note $1 + e^{-kx} > 1$, for all x so

$$F(x) = \frac{1}{1+e^{-kx}} < 1$$

so $F(x) \in [0,1]$.

② • $\lim_{x\to\infty} F(x) = \frac{1}{1+\lim_{x\to\infty} e^{-kx}}$

$$= \frac{1}{1+0}$$

$$= 1$$

• $\lim_{x\to-\infty} F(x) = \frac{1}{1+\lim_{x\to-\infty} e^{-k\infty}}$

$$= \frac{1}{1+\infty}$$

$$= 0$$

③ $\frac{dF}{dx} = \frac{ke^{-kx}}{(1+e^{-kx})^2} > 0$

so F is increasing.

Now suppose in general F is differentiable and

$$F'(t) = p(t)$$

Then lets find $\int_{-\infty}^{t} p(s)ds$.

$$\int_{-\infty}^{t} p(s)\,ds = \lim_{a \to -\infty} \int_{a}^{t} p(s)\,ds$$

$$= \lim_{a \to -\infty} F(t) - F(a) \quad \text{, FTC II}$$

$$= F(t) \quad \text{, since } \lim_{a \to -\infty} F(x) = 0$$

Thus $$F(t) = \int_{-\infty}^{t} p(s)\,ds$$

<u>Definition</u> Given a CDF $F_X(t)$ for a random variable continuous X, we say that $p_X(t)$ is a probability density function (PDF) of X if

$$F_X(t) = \int_{-\infty}^{t} p(s)\,ds$$

Properties of PDF's:

Let X be a continuous random variable:

1. $p_X(t) \geq 0$
2. $\int_{-\infty}^{\infty} p_X(t)\,dt = 1$
3. $\int_{a}^{b} p_X(t)\,dt = F_X(b) - F_X(a)$

<u>proof:</u> 1. F_X is increasing so $p_X(t) = F_X'(t) \geq 0$

2. $\int_{-\infty}^{\infty} p_X(t)\,dt = \lim_{b \to \infty} \int_{-\infty}^{b} p_X(t)\,dt = \lim_{b \to \infty} F(b) = 1$

3. $F_X'(t) = p_X(t)$ so F_X is an antiderivative of p_X and thus 3 is just FTC II.

Intuition: So unlike discrete random variables, when X is a continous RV,

$$\mathbb{P}(X=a)=0, \text{ for all } a\in\mathbb{R}.$$

So finding a probability at a particular point is meaningless. What we are interested in is finding the area in a small region.

Suppose I want to find

$$\mathbb{P}(t \le X \le t+\Delta t)$$

Where Δt is small. Well...

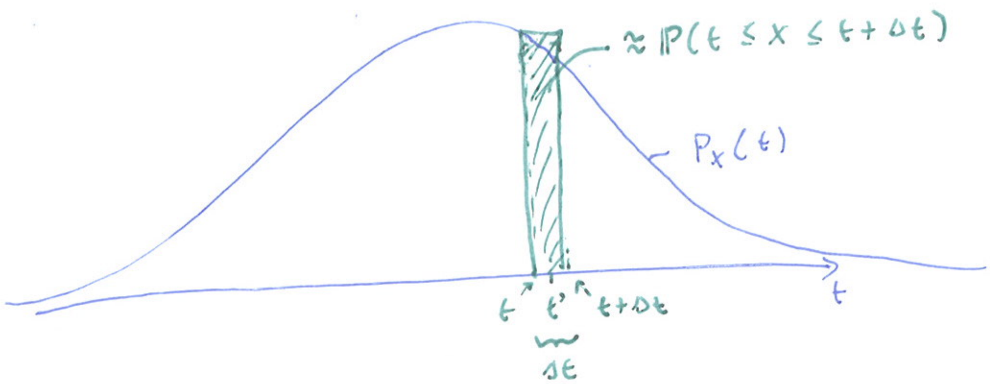
$$\mathbb{P}(t \le X \le t+\Delta t)$$

$= F_X(t+\Delta t) - F_X(t)$, since X is conti[n]ous

$= F_X'(t^*)\Delta t$, for some $t \in [t, t+\Delta t]$, by mean value theorem

$= p_X(t^*)\Delta t$

$\approx p_X(t)\Delta t$, when Δt is small.



So $\quad p_X(t) \approx \dfrac{\mathbb{P}(t \le X \le t+\Delta t)}{\Delta t}$

$\approx$ "average" probability that $X \in [t, t+\Delta t]$.

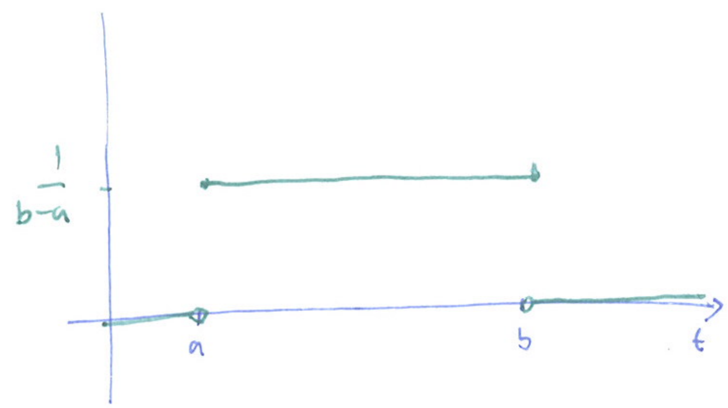
So we can think of $p_X(t)$ as the averge probabity of finding X in an infinitesimal neighbourhood around t.

Often in practice, one is given the PDF instead of the CDF.

eg If $a < b$. We say that X is uniformly distributed if,

$$p_X(t) = \begin{cases} \frac{1}{b-a} & , t \in [a,b] \\ 0 & , \text{otherwise} \end{cases}$$

So



X is uniform is the average probability of finding X in an infinitesimal neighbourhood around $t \in [a,b]$ is the same for all t.

It is clear that: ① $p_X \geq 0$

② $\int_{-\infty}^{\infty} p_X(t)\,dt = 1$

So p_X is a pdf.

In our spinner example θ is normally distributed with $a = 0$, $b = 2\pi$

Also $F_X(t) = \int_{-\infty}^{t} p(s)\,ds = \begin{cases} 0 & \text{if } t < a \\ \frac{t-a}{b-a} & \text{if } a \leq t \leq b \\ 1 & \text{if } t > b \end{cases}$

eg Given $\lambda > 0$, we say that X is exponentially distributed with rate λ if

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

Let us check that p_X is a pdf:

① $p_X \geq 0$ is clear.

② $$\int_{-\infty}^{\infty} p_X(t)\,dt = \int_0^{\infty} \lambda e^{-\lambda t}\,dt \quad , \text{ since } p_X(t) = 0,\ t < 0$$

$$= \lim_{b \to \infty} \int_0^b \lambda e^{-\lambda t}\,dt$$

$$= \lim_{b \to \infty} \left[e^{-\lambda t} \right]_0^b$$

$$= \lim_{b \to \infty} 1 - e^{-\lambda b}$$

$$= 1$$

Fun fact: Suppose X is the life span of device (or more morbidly a person). with an average life span m. Then a good model is that X is exponentially distributed with rate $\frac{1}{m}$.

$$F_X(t) = \int_{-\infty}^{t} p(s)\,ds$$

$$= \begin{cases} 0 & , t < 0 \\ \int_0^t \lambda e^{-\lambda s}\,ds & , t \geq 0 \end{cases}$$

$$= \begin{cases} 0 & , t < 0 \\ 1 - e^{-\lambda t} & , t \geq 0 \end{cases}$$

Not all pdf's need to have a physical meaning.

eg Find a constant k that makes $f(x)$ into a pdf.

$$f(x) = \begin{cases} kx^2, & 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Well ① $f(x) \geq 0$ for $k \geq 0$

② $1 = \int_{-\infty}^{\infty} f(x)\,dx$

$$= \int_1^2 kx^2\,dx$$

$$= \frac{kx^3}{3}\Big|_1^2$$

$$= k\left[\frac{8}{3} - \frac{1}{3}\right]$$

$\Rightarrow k = \frac{3}{7}$

So $$f(x) = \begin{cases} \frac{3}{7}x^2, & 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

The constant is called a normalization constant eg. Let X be a RV taking values in $[1,\infty)$. With PDF

$$p_X(t) = \begin{cases} A t^{-5} & , t \geq 1 \\ 0 & , t < 0 \end{cases}$$

a) Find A

b) Find F_X, the CDF

c) Find $\mathbb{P}(1 \leq X \leq 2)$, $\mathbb{P}(X \geq 4)$

Sol^n a) $$1 = \int_{-\infty}^{\infty} p_X(t)\,dt$$

$$= \int_1^{\infty} A t^{-5}\,dt$$

$$= A \lim_{b\to\infty} \left[\frac{t^{-4}}{-4} \right]_1^b$$

$$= A \lim_{b\to\infty} \left[\frac{1}{4} - \frac{b^{-4}}{4} \right]$$

$$= \frac{A}{4}$$

So $A = 4$, $p_X(t) = \begin{cases} 4t^{-5} & , t \geq 1 \\ 0 & , t < 0 \end{cases}$

b) $$F_X(t) = \int_{-\infty}^{t} p_X(s)\,ds$$

$$= \begin{cases} 0 & , t < 1 \\ \int_1^t 4 s^{-5}\,ds & , t \geq 1 \end{cases}$$

$$= \begin{cases} 0 & , t < 1 \\ 1 - \frac{1}{t^4} & , t \geq 1 \end{cases}$$

c) $\mathbb{P}(1 \leq X \leq 2) = F_X(2) - F_X(1)$

$= \left(1 - \frac{1}{2^4}\right) - 0$

$= \frac{15}{16}$

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

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

$= 1 - F_X(4)$

$= 1 - \left(1 - \frac{1}{4^4}\right)$

$= \frac{1}{4^4}$

$= \frac{1}{256}$

Expectation

Let's play a game. Suppose I flip a coin. If it is heads you given me $1, if it is a tails I give you $2. Suppose a head appears with probability $\frac{3}{4}$ and tails with probability with $\frac{1}{4}$.

Should you play this game? Well let's see how much I expect to win per game. Your expected gains are:

$$(*) \qquad (\$-1)\frac{3}{4} + (\$2)\frac{1}{4} = \$-\frac{1}{4} = -25¢.$$

So you expect to lose 25 cents per game and you should not play.

Let us see what we did. The above computation is intuitive, but how do we bring it into the framework of probability.

Let X denote the out come of the toss. Let Y denote the gains. So

$$Y(X) := \begin{cases} -1, & X = \text{heads} \\ 2, & X = \text{tails} \end{cases}$$

So the expected gains are the possibible outcomes of X weighted by the probability of them occuring. So

$$(*) = Y(H)\,\mathbb{P}(X=H) + Y(T)\,\mathbb{P}(X=T).$$

Definition: Given a discrete random variable X, we define the **expectation** of X to be the average of the outcomes of X weighted by the probability of the outcome occurring. Mathematically

$$\mathbb{E}(X) = \sum_{\substack{\text{outcomes} \\ t}} t \, \mathbb{P}(X = t)$$

We can analogously define the expectation for a continuous random variable X by replacing the sum with an integral and probability with the PDF.

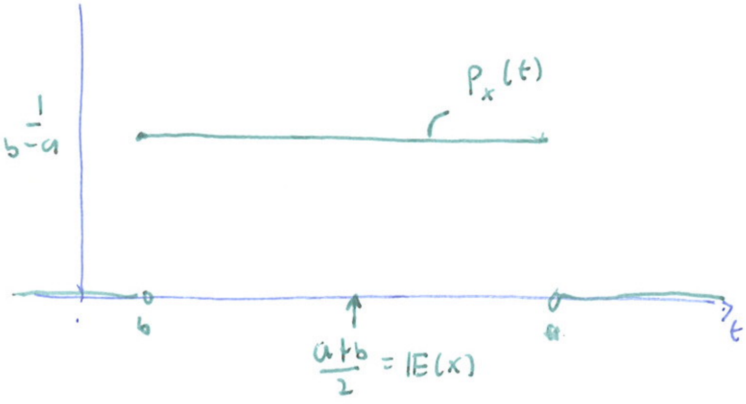
$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t \, P_X(t) \, dt$$

If $f(X)$ is a function of X then the expectation of f is.

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(t) \, P_X(t) \, dt$$

Eg. Let X be uniformly distributed, on $[a, b]$, find $\mathbb{E}(X)$.

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} t \, P_X(t) \, dt \\
&= \int_a^b \frac{t}{b-a} \, dt \\
&= \left[\frac{t^2}{2(b-a)} \right]_a^b \\
&= \frac{b^2 - a^2}{2(b-a)} \\
&= \frac{a+b}{2}
\end{aligned}$$



$\mathbb{E}(X)$ is a measure of the average value that X takes on.

eg Let X be a random variable with pdf

$$P_X(t) = \begin{cases} 2t & , \ t \in [0,1] \\ 0 & , \ \text{otherwise} \end{cases}$$

Find a) $\mathbb{E}(X)$

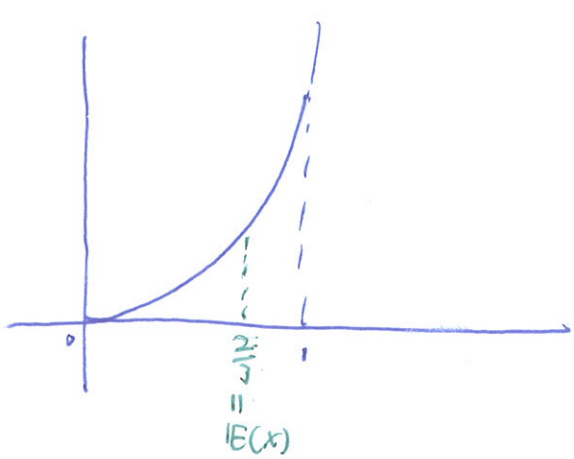
b) $\mathbb{E}(e^X)$

a) $\mathbb{E}(X) = \int_{-\infty}^{\infty} t \, P_X(t) \, dt$

$= \int_0^1 t \, 2t \, dt$

$= 2 \int_0^1 t^2 \, dt$

$= \frac{2}{3}$



So on average X is $\frac{2}{3}$.

b) $\mathbb{E}(e^X) = \int_{-\infty}^{\infty} e^t p_X(t) dt$

$= \int_0^1 2 e^t t \, dt$, $u = 2t$ $dv = e^t dt$
$du = 2dt$ $v = e^t$

$= 2te^t \Big|_0^1 - \int_0^1 2e^t dt$

$= 2e - [2e^t]_0^1$

$= e - [2e - 2]$

$= 2$

So on averge e^X is 2. Note!

$$\mathbb{E}(e^X) \neq e^{\mathbb{E}(X)}$$

Variance

Note if X is uniformly distributed on $[-1, 1]$
Y " " " " $[-2, 2]$

Both $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, but X is going to be closer to 0 on average than Y. So a question is how to measure how far away a random variable is from its mean.

We want to know how far away X is from $\mathbb{E}(X)$ on average: ie

$$\underbrace{\mathbb{E}}_{\text{on average.}} \big(\underbrace{| X - \mathbb{E}(X) |}_{\text{distance from } X \text{ to } \mathbb{E}(X)} \big)$$

It turns out to be easier to ~~find~~ measure the average square distance between X and $\mathbb{E}(X)$, i.e.

$$\mathbb{E}\left[(X - \mathbb{E}(X))^2\right]$$

Definition: Given a random variable X the variance of X is

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

The standard deviation is

$$\sigma(X) = \sqrt{Var(X)}$$

It can be shown (you will in your assignment) that

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

eg Let X be random variable with pdf:

$$p_X(t) = \begin{cases} 2x & , \quad x \in [0,1] \\ 0 & , \quad \text{otherwise.} \end{cases}$$

We had that $\mathbb{E}(X) = \frac{2}{3}$, so now lets find $Var(X)$.

$$\begin{aligned} Var(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \mathbb{E}\left((X - \tfrac{2}{3})^2\right) \\ &= \int_{-\infty}^{\infty} (t - \tfrac{2}{3})^2 p_X(t)\, dt \\ &= \int_0^1 2(t - \tfrac{2}{3})^2 t \, dt \end{aligned}$$

$$\mathrm{Var}(X) = \int_0^1 2\left[t^3 - \frac{4}{3}t^2 + \frac{4}{9}t\right]dt$$

$$= 2\left[\frac{1}{4} - \frac{4}{9} + \frac{2}{9}\right]$$

$$= \frac{1}{18}$$

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

$$= \frac{1}{3\sqrt{2}}$$

So on average X is a distance $\frac{1}{3\sqrt{2}}$ a from $\frac{2}{3}$.

Excercise: • If X is uniformly distributed on $[a, b]$, show

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

• If X is exponentially distributed with rate λ,

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$