Recap Continuous Joint Distributions, Law of Total Probability (continuous) Law of Total Expectation

CSE 312 Spring 26

Lecture 22

Agenda

- Recap – joint continuous distns
- Continuous law of total probability
- Law of total expectation

$$f(x)\,dx \approx P(X \in [x, x+dx])$$

$$f_{XY}(x,y)\,dx\,dy \simeq P(X \in [x, x+dx],\ Y \in [y, y+dy])$$

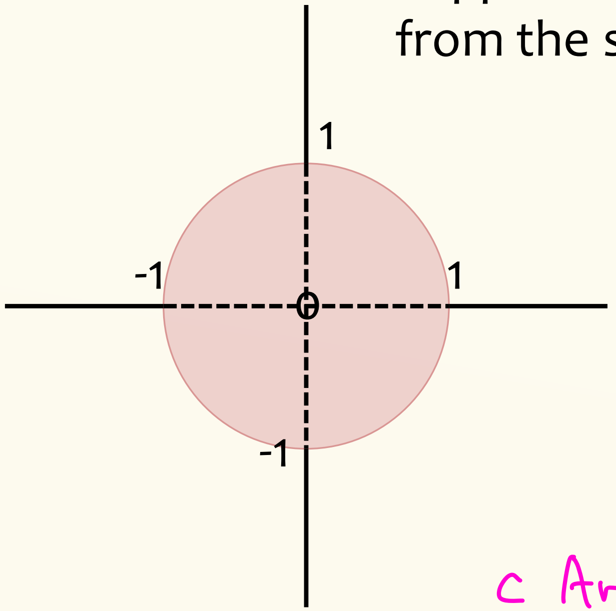
Reference Sheet

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

$\Omega_{X,Y} = \{(x,y) \mid f_{XY}(x,y) > 0\}$

$\Omega_{XY} = \Omega_X \times \Omega_Y$

Example – Uniform distribution on a unit disk



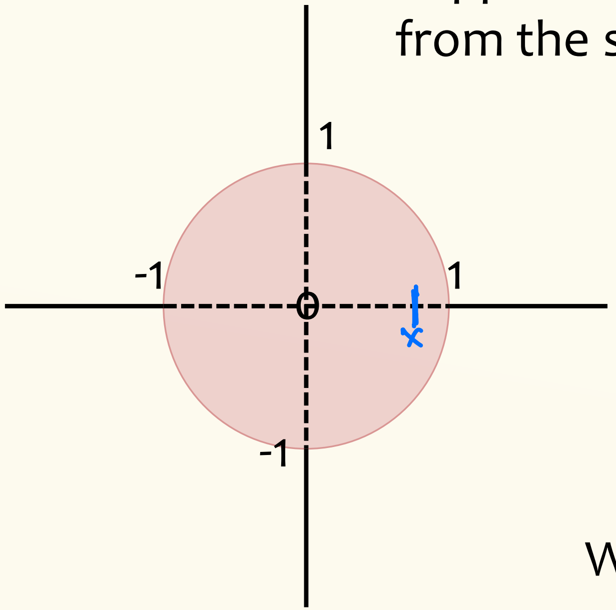
Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

What is the joint density?

$$f_{X,Y}(x, y) = \begin{cases} c & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

c Area of circle $= c\pi = 1$

Example – Uniform distribution on a unit disk



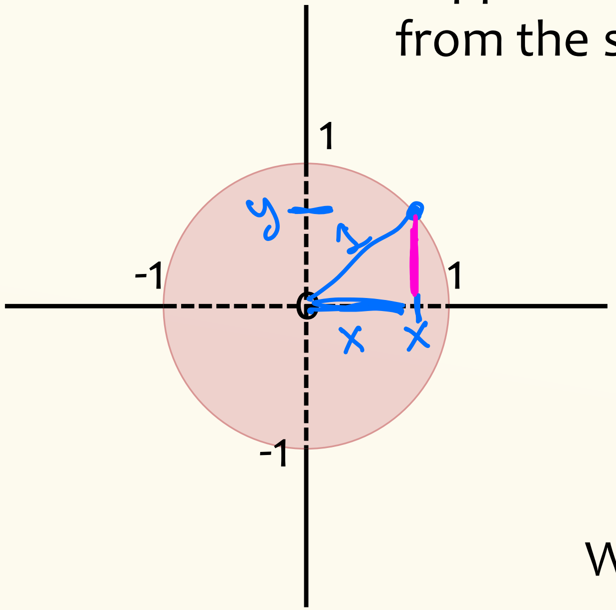
Suppose that a pair of random variabes (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $f_X(x)$?

Example – Uniform distribution on a unit disk



Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

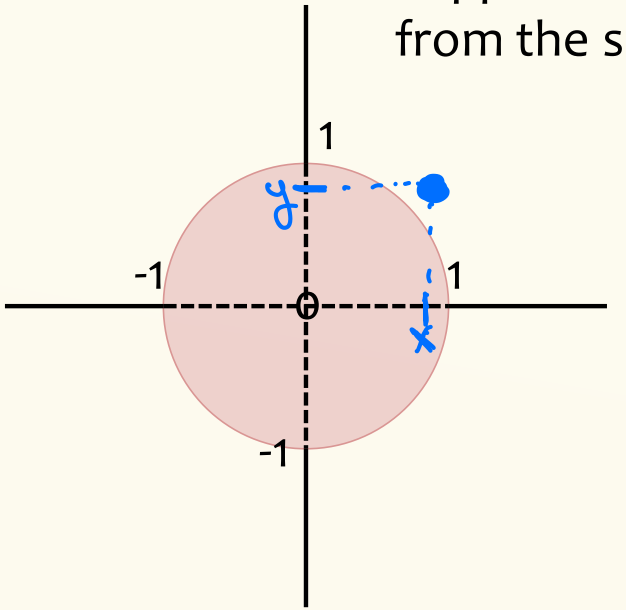
$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $f_X(x)$?

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} \mathrm{d}y$$

$$= 2\sqrt{1-x^2}/\pi$$

Example – Uniform distribution on a unit disk



Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = 2\sqrt{1-x^2}/\pi$$

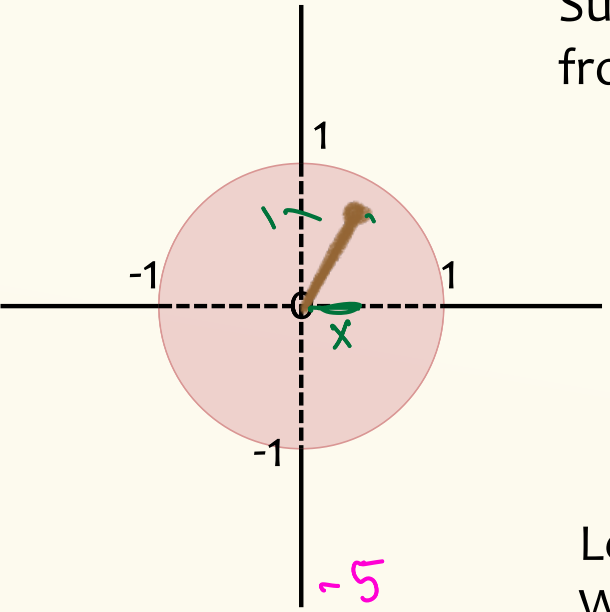
$$f_Y(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi} dx = 2\sqrt{1-y^2}/\pi$$

Are X and Y independent?

Is $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for all $x, y \in \mathbb{R}$?

$x = 1 - \varepsilon$ $y = 1 - \varepsilon$

Example – Uniform distribution on a unit disk



Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let D be the distance of the point (X, Y) from the origin.
What is $\mathrm{P}(\mathrm{D} \leq \mathrm{d})$? What is $f_{\mathrm{D}}(\mathrm{d})$? What is $\mathrm{E}(\mathrm{D})$?

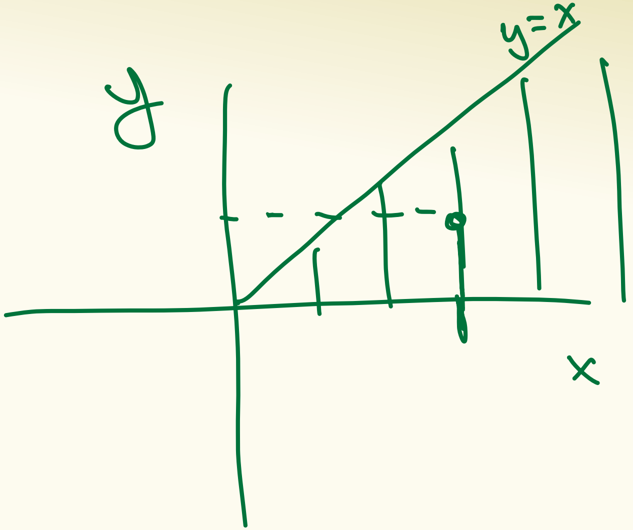
$$P(D \leq d) = P\left(\sqrt{X^2+Y^2} \leq d\right) = P\left(X^2+Y^2 \leq d^2\right) = \iint_{x^2+y^2 \leq d^2} \boxed{f_{XY}(x,y)}\, dx\,dy = \frac{\pi d^2}{\pi} = d^2$$

$$\frac{1}{\pi}$$

$$f_D(d) = \begin{cases} 2d & d \in [0,1] \\ 0 & \text{o.w.} \end{cases}$$

$$E(D) = \int_0^{\infty} x\, f_D(x)\, dx = \frac{2}{3}$$

Example – More joint densities



The joint density of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 \le x, y \le \infty \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X > 1, Y < 1)$? What is $P(X > Y)$? What is $f_X(x)$?

$$= \int_0^1 \int_1^\infty f_{XY}(x,y)\,dx\,dy$$

$$\int_0^\infty \int_0^x f_{XY}(x,y)\,dy\,dx$$

Reference Sheet

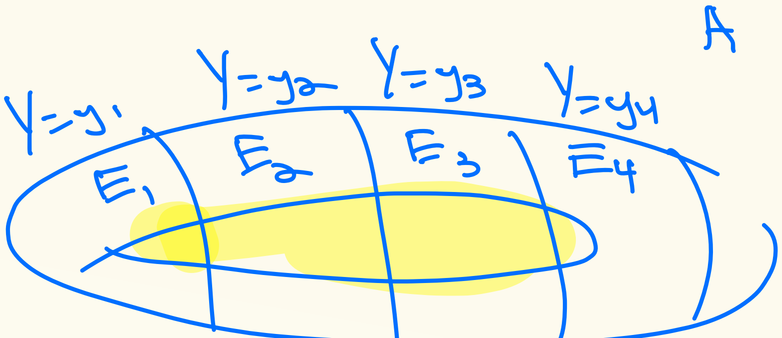
	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

$f_{XYWZ}(x,y,w,z)$

$f_{XY}(\underline{x,y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\underline{XY}\underline{WZ}} \, dw\,dz$

Agenda

- Recap – joint continuous distns
- Continuous law of total probability
- Law of total expectation



Law of total probability

Definition. Let A be an event and Y a discrete random variable. Then

$$P[A] = \sum_{y \in \Omega_Y} P(A|Y = y) p_Y(y)$$

Law of total probability

Definition. Let A be an event and Y a discrete random variable. Then

$$P[A] = \sum_{y \in \Omega_Y} P(A|Y = y) p_Y(y)$$

Definition. Let A be an event and Y a continuous random variable. Then

$$P[A] = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y) \mathrm{d}y$$

Example 1: use of law of total probability

We have a coin with unknown bias.

Specifically, the coin has probability P of heads where $P \sim \mathrm{Unif}(0,1)$.

$$f_P(p) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{o.w.} \end{cases}$$

What is $\boldsymbol{P}$(Next 10 flips are all heads)?

Definition. Let A be an event and Y a continuous random variable. Then

$$P[A] = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y) \mathrm{d}y$$

$$P(\text{next 10 flips are heads} \mid P = p) = p^{10}$$

$$P(\text{next 10 flips heads}) = \int_{-\infty}^{\infty} P(\text{next 10 H} \mid P = p)\, f_P(p)\, dp$$

Example 1: use of law of total probability

We have a coin with unknown bias.

Specifically, the coin has probability P of heads where $P \sim \mathrm{Unif}(0,1)$.

What is $\boldsymbol{P}$(Next 10 flips are all heads)?

$$P[\text{Next 10 flips are all heads}]$$
$$= \int_{-\infty}^{\infty} P(\text{Next 10 flips are all heads} | \mathrm{P} = \mathrm{p}) f_{\mathrm{P}}(\mathrm{p}) \mathrm{dp}$$

$$= \int_0^1 p^{10} \, dp = \frac{1}{11}$$

Example 2: use of law of total probability

A certain worker is able to complete jobs efficiently until they are too tired, which happens after a number of hours T which is exponential, with parameter λ.

$T \sim exp(\lambda)$

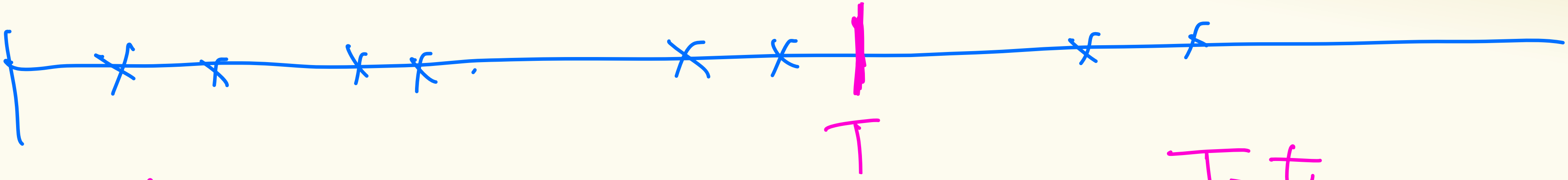
Jobs to be performed arrive according to a Poisson distribution with a rate of μ jobs per hour. Any job that arrives before time T (i.e., before the worker is tired) is successfully completed.

What is the probability that the worker completes k jobs before they are too tired?

jobs in an hour $\sim$ Poisson (μ)

jobs in 2 hours $\sim$ Poisson (2μ)

jobs in t hours $\sim$ Poisson(μt)



$T = t$

Poisson(μt)

P(worker completes k jobs)

$$P(\text{complete } k \text{ jobs} \mid T = t) = e^{-\mu t}\frac{(\mu t)^k}{k!}$$

$$P(\text{complete } k \text{ jobs}) = \int_0^\infty P(\text{complete } k \text{ jobs} \mid T = t)\, \underline{f_T(t)}\, dt$$

$$= \int_0^\infty e^{-\mu t}\frac{(\mu t)^k}{k!}\, \lambda e^{-\lambda t}\, dt = \left(\frac{\lambda}{\lambda+\mu}\right)\left(\frac{\mu}{\lambda+\mu}\right)^k \approx \text{Geom}\left(\frac{\lambda}{\lambda+\mu}\right)$$

Definition. Let A be an event and Y a continuous random variable. Then

$$P[A] = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y) \mathrm{d}y$$

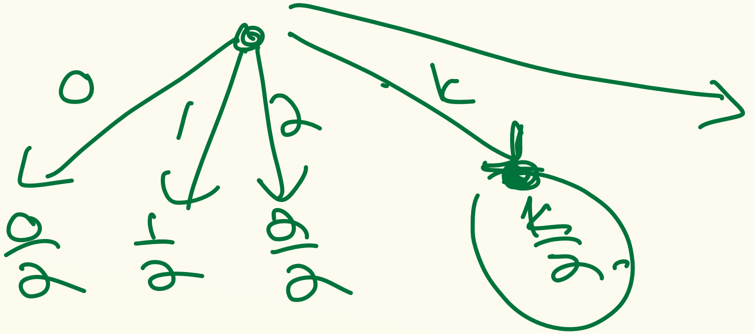
Agenda

- Recap – joint continuous distns
- Continuous law of total probability
- Conditional expectation and Law of Total Expectation

Conditional Expectation and Law of Total Expectation

Suppose someone gave us $Y \sim \text{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

$E(X) = ?$



$E(X \mid Y = k)$

$$E(X) = \sum_{k=0}^{\infty} E(X \mid Y=k)\, P(Y=k)$$

Conditional Expectation

Definition. Let X be a discrete random variable then the **conditional expectation** of X given event A is

$$\mathbb{E}[X \mid A] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid A)$$

Notes:

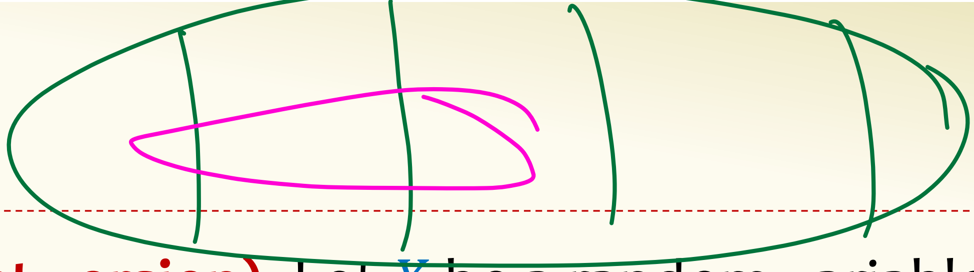
- Can be phrased as a "random variable version"

$$\mathbb{E}[X \mid Y = y]$$

- Linearity of expectation still applies here

$$\mathbb{E}[aX + bY + c \mid A] = a\,\mathbb{E}[X \mid A] + b\,\mathbb{E}[Y \mid A] + c$$

Law of Total Expectation



Law of Total Expectation (event version). Let X be a random variable and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot P(A_i)$$

Law of Total Expectation (random variable version). Let X be a random variable and Y be a discrete random variable. Then,

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y)$$

Proof of Law of Total Expectation

Follows from Law of Total Probability and manipulating sums

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$= \sum_{x \in \Omega_X} x \cdot \sum_{i=1}^{n} P(X = x \mid A_i) \cdot P(A_i)$$ (by LTP)

$$= \sum_{i=1}^{n} P(A_i) \sum_{x \in \Omega_X} x \cdot P(X = x \mid A_i)$$ (change order of sums)

$$= \sum_{i=1}^{n} P(A_i) \cdot \mathbb{E}[X \mid A_i]$$ (def of cond. expect.)

Example – Flipping a Random Number of Coins

Suppose someone gave us $Y \sim \mathrm{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

By the Law of Total Expectation

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{E}[X \mid Y = i] \cdot P(Y = i) = \sum_{i=0}^{\infty} \frac{i}{2} \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \sum_{i=0}^{\infty} i \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \mathbb{E}[Y] = \frac{1}{2} \cdot 5 = 2.5$$

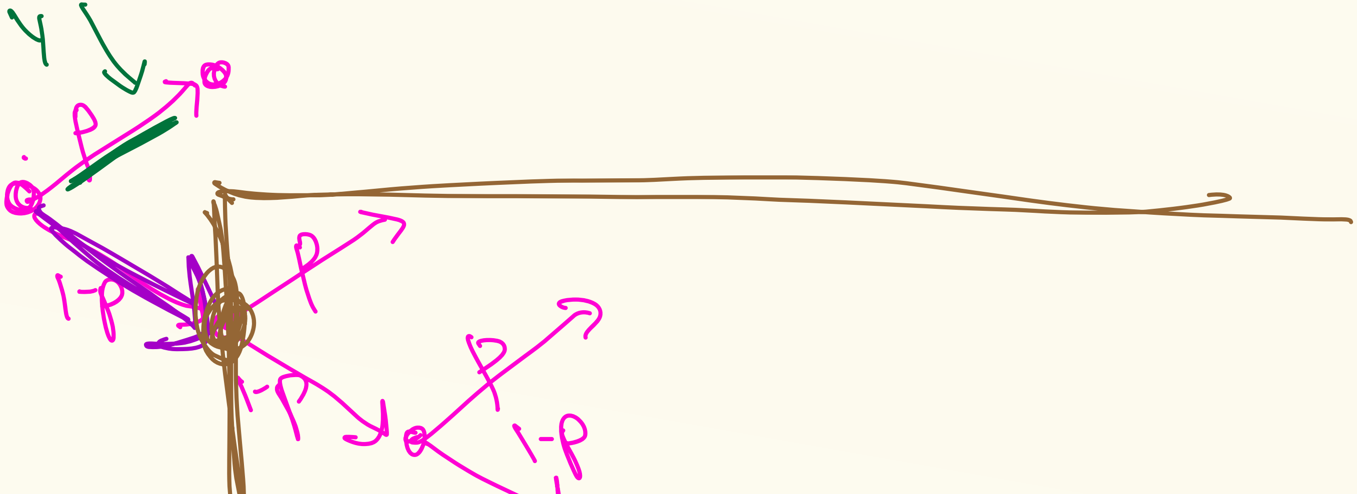
Example – Computer Failures (a familiar example)

Suppose your computer operates in a sequence of steps, and that at each step i your computer will fail with probability p (independently of other steps).

Let X be the number of steps it takes your computer to fail. $X \sim \mathrm{Geo}(p).$

What is $\mathbb{E}[X]$?

$$E(X) = \sum_{k=1}^{\infty} k\, p\, (1-p)^{k-1} = \frac{1}{p}$$



Let Y be the indicator random variable for the event of failure in step 1

Then by LTE, $\mathbb{E}[X] = \mathbb{E}[X \mid Y = 1] \cdot P(Y = 1) + \mathbb{E}[X \mid Y = 0] \cdot P(Y = 0)$

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

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

since if $Y = 0$ experiment starting at step 2 looks like original experiment

Solving we get $\mathbb{E}[X] = 1/p$