Unit 4 Discrete Random Variables

Probability Mass Functions and Expectations

A random variable is loosely speaking, a numerical quantity whose value is determined by the outcome of a probabilistic experiment e.g. the weight of a randomly selected person

Discrete: takes values in finite or countable set
Random variable examples:
Bernoulli
Uniform
Binomial
Geometric

Expected value of a random variable aka Expectation (mean)
weighted average of the values of the random variable
weighted on their probabilities

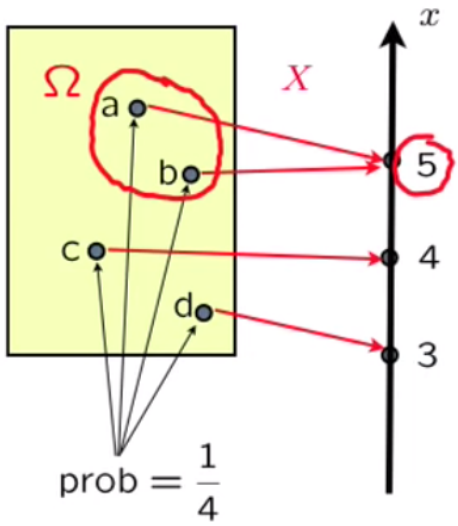
Random variables mathematically: A function from the sample space omega to the real numbers

Notation: random variable X numerical value x

A function of one or more random variables is also a random variable

Probability Mass Function (PMF) aka probability law or probability distribution of X

Probability space of getting different values
P(5) = 1/2 a and b



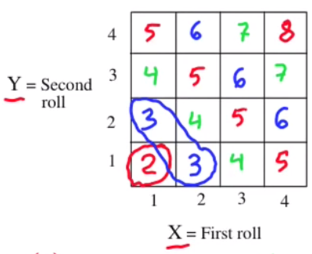
Notation:

$$p_X(x) = \mathbf{P}(X = x) = \mathbf{P}\big(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\}\big)$$

- **Properties:** $p_X(x) \geq 0$ $\qquad \sum_x p_X(x) = 1$

PMF calculation

- Two rolls of a tetrahedral die
- Let every possible outcome have probability 1/16



$Z = X + Y$ $\qquad$ Find $p_Z(z)$ $\quad$ for all z

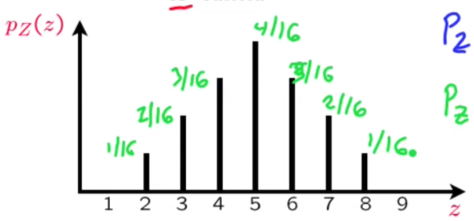
- repeat for all z:
  - collect all possible outcomes for which Z is equal to z
  - add their probabilities

$P_Z(2) = P(Z=2) = 1/16$

$P_Z(3) = P(Z=3) = 2/16$

$P_Z(4) = P(Z=4) = 3/16$

⋮



Exercise: Random variables versus numbers

1/2 points (graded)

Let X be a random variable that takes integer values, with PMF $p_X(x)$. Let Y be another integer-valued random variable and let y be a number.

a) Is $p_X(y)$ a random variable or a number?

Random variable ✖ **Answer:** Number

b) Is $p_X(Y)$ a random variable or a number?

Random variable ✔ **Answer:** Random variable

Solution:

a) Recall that $p_X(\cdot)$ is a function that maps real numbers to real numbers. So, when we give it a numerical argument, y, we obtain a number.

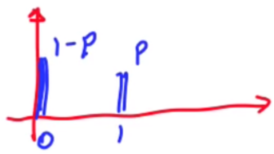
b) In this case, we are dealing with a function, the function being $p_X(\cdot)$, of a random variable Y. And a function of a random variable is a random variable. Intuitively, the "random" value of $p_X(Y)$ is generated as follows: we observe the realized value y of the random variable Y, and then look up the numerical value $p_X(y)$.

The simplest random variable: Bernoulli with parameter $p \in [0, 1]$

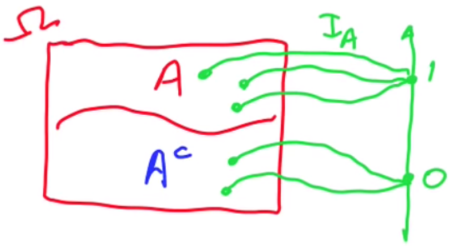
$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases}$$

$p_X(0) = 1-p$

$p_X(1) = p$



- Models a trial that results in success/failure, Heads/Tails, etc.
- Indicator r.v. of an event A: $I_A = 1$ iff A occurs



$$p_{I_A}(1) = \mathbf{P}(I_A = 1) = \underbrace{\mathbf{P}(A)}_{p}$$

Exercise: Indicator variables

2/2 points (graded)

Let A and B be two events (subsets of the same sample space Ω), with nonempty intersection. Let I_A and I_B be the associated indicator random variables.

For each of the two cases below, select one statement that is true.

a) $I_A + I_B$:

is not the indicator random variable of any event ✔

Answer: is not the indicator random variable of any event

b) $I_A \cdot I_B$:

is the indicator variable of the event $A \cap B$ ✔

Answer: is the indicator variable of the event $A \cap B$

(*Bug warning:* In some browsers, the mathematical content in each choice in the dropdown menu may appear duplicated, e.g. $A \cup B$ may show up twice as $A \cup BA \cup B$.)

Solution:

a) If the outcome of the experiment lies in the intersection of the events A and B, then $I_A + I_B$ takes the value of 2. But indicator random variables can take only the values 0 or 1. Therefore, $I_A + I_B$ is not an indicator random variable.

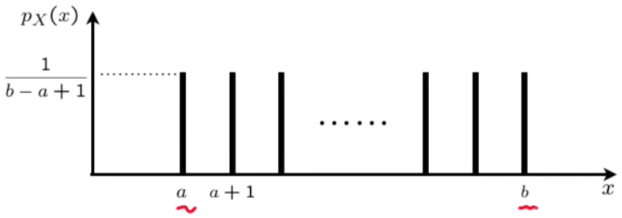
b) Note that $I_A \cdot I_B$ can take only the values 0 or 1. It is equal to 1 if and only if $I_A = 1$ (i.e., event A occurs) and $I_B = 1$ (i.e., event B occurs). Thus, $I_A \cdot I_B$ takes the value of 1 if and only if both A and B occur, and so it is the indicator random variable of the event $A \cap B$.

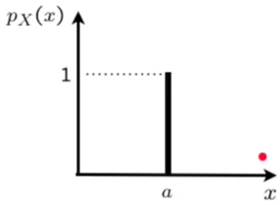
Discrete uniform random variable; parameters $a,\ b$

- **Parameters:** integers $a,\ b$; $\quad a \le b$
- **Experiment:** Pick one of $a, a+1, \ldots, b$ at random; all equally likely
- **Sample space:** $\{a, a+1, \ldots, b\}$ $\quad$ $b - a + 1$ possible values
- **Random variable** X: $X(\omega) = \omega$ $\quad$ 11:52:26 $\{0, 1, \ldots, 59\}$
- **Model of:** complete ignorance

Special case: $a = b$

constant/deterministic r.v.



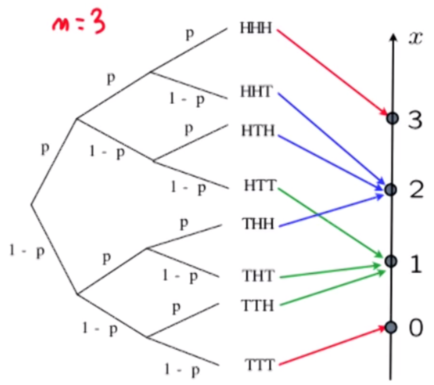


e.g. imagine looking at a clock at time 11:52:26 but only looking at the seconds the probability of getting a number 0 to 59 is equally likely

Special case where a = b is still a random variable in the mathematical sense, it just so happens that it can only be one value in this case

Binomial random variable; **parameters:** positive integer n; $p \in [0, 1]$

- **Experiment:** n independent tosses of a coin with $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of sequences of H and T, of length n
- **Random variable** X**:** number of Heads observed
- **Model of:** number of successes in a given number of independent trials



$$p_X(2) = P(X=2)$$
$$= P(HHT) + P(HTH) + P(THH)$$
$$= 3p^2(1-p) = \binom{3}{2} p^2 (1-p)$$

$$\boxed{p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n}$$

Same formula as earlier but with slightly different notation

Exercise: The binomial PMF

2/2 points (graded)

You roll a fair six-sided die (all 6 of the possible results of a die roll are equally likely) 5 times, independently. Let X be the number of times that the roll results in 2 or 3. Find the numerical values of the following.

a) $p_X(2.5) =$ 0 ✔ **Answer:** 0

b) $p_X(1) =$ 0.3292 ✔ **Answer:** 0.32922

Solution:

a) A value of 2.5 is not possible for X since the number of rolls must be an integer, and therefore $p_X(2.5) = 0$.

b) For each die roll, there is a probability $2/6 = 1/3$ of obtaining a 2 or a 3. Hence, the random variable X is binomial with parameters $n = 5$ and $p = 1/3$, so that $p_X(1) = \binom{5}{1} \cdot (1/3) \cdot (2/3)^4 \approx 0.32922$.

Geometric random variable; parameter p: $0 < p \leq 1$

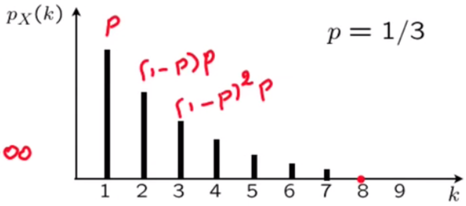
- **Experiment:** infinitely many independent tosses of a coin; $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of infinite sequences of H and T $\quad \underbrace{TTTTH}HT\cdots$
- **Random variable** X**:** number of tosses until the first Heads $\quad X = 5$

- **Model of:** waiting times; number of trials until a success

$$p_X(k) = P(X=k) = P(\underbrace{T\cdots T}_{k-1}H) = (1-p)^{k-1}p \qquad k = 1, 2, 3, \ldots$$

P(no Heads ever) $\leq P(\underbrace{T\cdots T}_{k}) = \underbrace{(1-p)^k}$

$(TTT\cdots)$ "$X = \infty$" $= 0$ $\quad \downarrow k \to \infty \quad 0$



P(Tails forever) becomes 0 as k goes to infinity

Exercise: Geometric random variables

0/2 points (graded)

Let X be a geometric random variable with parameter p. Find the probability that $X \geq 10$. Express your answer in terms of p using standard notation (click on the “STANDARD NOTATION" button below.)

$\mathbf{P}(X \geq 10) =$ p/(1-((1-p)^9)) ✖ **Answer:** (1-p)^9

$$\frac{p}{1-((1-p)^9)}$$

STANDARD NOTATION

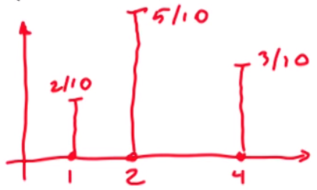
Solution:

We can calculate the desired probability by adding the probabilities of the events $\{X = 10\}$, $\{X = 11\}$, $\{X = 12\}$, etc., and using the formula for the sum of a geometric series. However, we can get the answer in an easier way, using the interpretation of geometric random variables as the number of trials until the first success. The event $\{X \geq 10\}$ is the event that the the first 9 trials resulted in failure, and therefore its probability is $(1-p)^9$.

Expectation/mean of a random variable

- **Motivation:** Play a game 1000 times. Random gain at each play described by:

$$X = \begin{cases} 1, & \text{w.p. } 2/10 \quad \sim 200 \\ 2, & \text{w.p. } 5/10 \quad \sim 500 \\ 4, & \text{w.p. } 3/10 \quad \sim 300 \end{cases}$$



- "Average" gain:

$$\frac{1\cdot 200 + 2\cdot 500 + 4\cdot 300}{1000}$$

$$= 1\cdot\frac{2}{10} + 2\cdot\frac{5}{10} + 4\cdot\frac{3}{10}$$

- **Definition:** $\mathrm{E}[X] = \sum_x x p_X(x)$

- **Interpretation:** Average in large number of independent repetitions of the experiment

- Caution: If we have an infinite sum, it needs to be well-defined. We assume $\sum_x |x|\, p_X(x) < \infty$

Expectation of a Bernoulli r.v.

$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases}$$

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

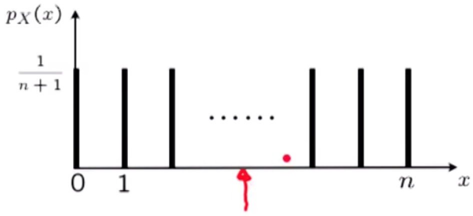
If X is the indicator of an event A, $X = I_A$:

$X = 1$ iff A occurs $\qquad p = P(A)$

$$E[I_A] = P(A)$$

Expectation of a uniform r.v.

- Uniform on $0, 1, \ldots, n$



- **Definition:** $\mathbf{E}[X] = \sum_x x p_X(x)$

$$\mathbf{E}[X] = 0 \cdot \frac{1}{n+1} + 1 \cdot \frac{1}{n+1} + \cdots + n \frac{1}{n+1}$$

$$= \frac{1}{n+1}(0 + 1 + \cdots + n) = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = \frac{n}{2}$$

Elementary properties of expectations

- **Definition:** $\mathbf{E}[X] = \sum_x \underbrace{x}_{\geq 0} \underbrace{p_X(x)}_{\geq 0}$ ≥ 0

- If $X \geq 0$, then $\mathbf{E}[X] \geq 0$

for all ω: $X(\omega) \geq 0$

- If $a \leq X \leq b$, then $a \leq \mathbf{E}[X] \leq b$

for all ω: $a \leq X(\omega) \leq b$

$$E[X] = \sum_x x\, p_X(x) \geq \sum_x a\, p_X(x)$$

$$= a \sum_x p_X(x) = a \cdot 1 = a$$

- If c is a constant, $\mathbf{E}[c] = c$

$$E[c] = c \cdot p(c) = c$$

Exercise: Random variables with bounded range

3/3 points (graded)

Suppose a random variable X can take any value in the interval $[-1, 2]$ and a random variable Y can take any value in the interval $[-2, 3]$.

a) The random variable $X - Y$ can take any value in an interval $[a, b]$. Find the values of a and b:

$a =$ -4 ✔

$b =$ 4 ✔

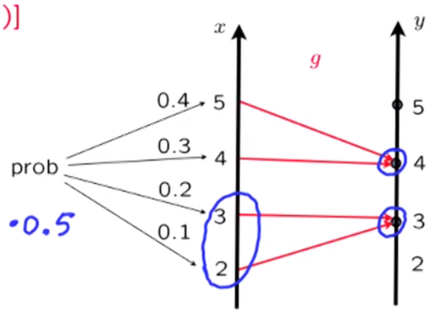
b) Can the expected value of $X + Y$ be equal to 6?

No ✔

The expected value rule, for calculating $\mathbf{E}[g(X)]$

- Let X be a r.v. and let $Y = g(X)$
- Averaging over y: $\mathbf{E}[Y] = \sum_y y p_Y(y)$

  $3\cdot(0.1+0.2)+4\cdot(0.3+0.4)$
- Averaging over x: $3\cdot 0.1+3\cdot 0.2+4\cdot 0.3+4\cdot 0.5$



$$\boxed{\mathbf{E}[Y] = \mathbf{E}[g(X)] = \sum_x g(x) p_X(x)}$$

Proof: $\sum_y \sum_{x:g(x)=y} g(x) p_X(x)$

$= \sum_y \sum_{x:g(x)=y} y\, p_X(x) = \sum_y y \sum_{x:g(x)=y} p_X(x)$

$= \sum_y y\, p_Y(y) = E[Y]$

- $\mathbf{E}[X^2] = \sum_x x^2 p_X(x)$

  $g(x) = x^2$
- Caution: In general, $\mathbf{E}[g(X)] \neq g(\mathbf{E}[X])$

  $E[X^2] \neq (E[X])^2$

Linearity of expectation: $\mathbf{E}[aX+b] = a\mathbf{E}[X]+b$

X = salary $E[X]$ = average salary

Y = new salary $= 2X+100$ $E[Y] = E[2X+100] = 2E[X]+100$

- Intuitive
- **Derivation**, based on the expected value rule: $g(x) = ax+b$, $Y = g(X)$

$$E[Y] = \sum_x g(x)\, p_X(x)$$

$$= \sum_x (ax+b)\, p_X(x) = a\sum_x x\, p_X(x) + b\underbrace{\sum_x p_X(x)}_{1}$$

$$= aE[X] + b$$

$$E[g(X)] = g(E[X])$$

exceptional g

The blue equality is only true for linear functions

Exercise: Linearity of expectations

3/3 points (graded)

The random variable X is known to satisfy $\mathbf{E}[X] = 2$ and $\mathbf{E}[X^2] = 7$. Find the expected value of $8-X$ and of $(X-3)(X+3)$.

a) $\mathbf{E}[8-X] =$ 6

b) $\mathbf{E}[(X-3)(X+3)] =$ -2

Variance; Conditioning on an event; Multiple random variables

Variance - quantifies the spread of a probability mass function (PMF)
We look at the distance from the mean
If we look at the average distance from the mean it is always = 0 so it not informative
However if we look at the average of the distance^2 it is informative

- **Definition of variance:** $\text{var}(X) = \mathbf{E}[(X-\mu)^2]$

Variance is always positive

Standard deviation: $\sigma_X = \sqrt{\text{var}(X)}$

Variance is not in the same units as the original random variable but the square root of it is (standard deviation)

Properties of the variance

- Notation: $\mu = \mathbf{E}[X]$

$$\boxed{\text{var}(aX + b) = a^2 \text{var}(X)}$$

$$\text{var}(3-4x) = (-4)^2 \text{var}(x) = 16 \text{ var}(x)$$

- Let $Y = X + b$ $\quad \nu = E[Y] = \mu + b$

$$\text{var}(Y) = E[(Y-\nu)^2] = E[(X + \cancel{b} - (\mu + \cancel{b}))^2] = E[(X-\mu)^2] = \text{var}(X)$$

- Let $Y = aX$ $\quad \nu = E[Y] = a\mu$

$$\text{var}(Y) = E[(aX - a\mu)^2] = E[a^2(X-\mu)^2] = a^2 E[(X-\mu)^2] = a^2 \text{var}(X)$$

A useful formula: $\text{var}(X) = \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2$

$$\text{var}(X) = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$$

$$= E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - (E[X])^2$$

First calculation in red shows that adding a constant to a random variable does not change its variance (the b’s cancel out)
Second calculation in red shows that multiplying a random variable by a constant would be multiplying the variance by a^2
Useful formula is a quick and easy way to calculate variance

Note: The derivation of the useful alternative formula for the variance, near the end of the video, uses a linearity property for multiple random variables (in this case, the variables X and X^2), which we have not yet discussed.

A derivation that does not rely on this linearity property, goes as follows: This expression is verified as follows:

$$\begin{aligned}
\text{var}(X) &= \sum_x \left(x - \mathbf{E}[X]\right)^2 p_X(x) \\
&= \sum_x \left(x^2 - 2x\mathbf{E}[X] + \left(\mathbf{E}[X]\right)^2\right) p_X(x) \\
&= \sum_x x^2 p_X(x) - 2\mathbf{E}[X] \sum_x x p_X(x) + \left(\mathbf{E}[X]\right)^2 \sum_x p_X(x) \\
&= \mathbf{E}[X^2] - 2\left(\mathbf{E}[X]\right)^2 + \left(\mathbf{E}[X]\right)^2 \\
&= \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2.
\end{aligned}$$

(The sum of all x’s multiplied by their probability is the expected value of x)

Exercise: Variance properties

0/1 point (graded)

Is it always true that $\mathbf{E}[X^2] \geq \left(\mathbf{E}[X]\right)^2$?

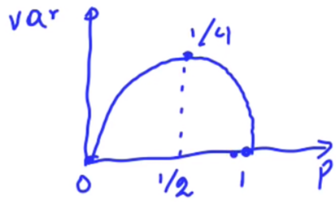
No ✖ **Answer:** Yes

Solution:

We know that variances are always nonnegative and that $\text{Var}(X) = \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2$. Therefore, $0 \leq \text{Var}(X) = \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2$, or, equivalently, $\mathbf{E}[X^2] \geq \left(\mathbf{E}[X]\right)^2$.

Variance of the Bernoulli

$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases} \qquad E[x] = p$$



$$\text{var}(X) = \sum_x (x - \mathbf{E}[X])^2 p_X(x) = (1-p)^2 p + (0-p)^2 \cdot (1-p)$$
$$= p - 2p^2 + \cancel{p^3} + p^2 - \cancel{p^3} = p - p^2 = p(1-p)$$

$$\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = E[x] - (E[x])^2 = p - p^2 = \boxed{p(1-p)}$$

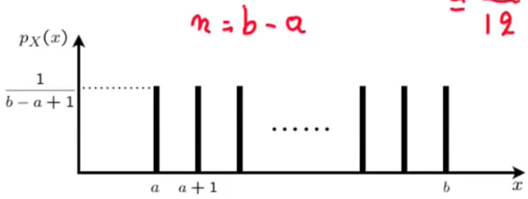
$$X^2 = X$$

X^2 = X here only because it is a Bernoulli random variable i.e. is a 0 or 1
Both methods show what the variance is
Plotting variance as a function of p we see that variance peaks when p is 0.5 i.e. when a coin toss is fair it is the most random

Variance of the uniform

$$\text{var}(x) = E[x^2] - (E[x])^2 = \frac{1}{n+1} \overbrace{(0^2 + 1^2 + 2^2 + \dots + n^2)}^{\frac{1}{6} n(n+1)(2n+1)} - \left(\frac{n}{2}\right)^2$$
$$= \frac{1}{12} n(n+2)$$

$$n = b - a$$



$$\text{Var}(x) = \frac{1}{12}(b-a)(b-a+2)$$

second is a more general form where it doesn't start from 0 necessarily
This shift is equivalent to adding a constant to the r.v. therefore the variance does not change
We just have to substitute b-a in for n

Exercise: Variance of the uniform

2/2 points (graded)

Suppose that the random variable X takes values in the set $\{0, 2, 4, 6, \ldots, 2n\}$ (the even integers between 0 and $2n$, inclusive), with each value having the same probability. What is the variance of X? *Hint:* Consider the random variable $Y = X/2$ and recall that the variance of a uniform random variable on the set $\{0, 1, \ldots, n\}$ is equal to $n(n+2)/12$.

Express your answer in terms of n using standard notation. Remember to write '*' for all multiplications and to include parentheses where necessary.

$\text{Var}(X) =$ (n*(n+2))/3 ✔

$$\frac{n \cdot (n+2)}{3}$$

Conditioning

Creating a new PMF based on information that an event has happened such as A
This is just another PMF and follows the same rules as ordinary PMFs except probabilities are replaced by conditional probabilities

Conditional PMF and expectation, given an event

- Condition on an event $A \Rightarrow$ use conditional probabilities

assume $P(A) > 0$

$$p_X(x) = \mathbf{P}(X = x) \qquad p_{X|A}(x) = \mathbf{P}(X = x \mid A)$$

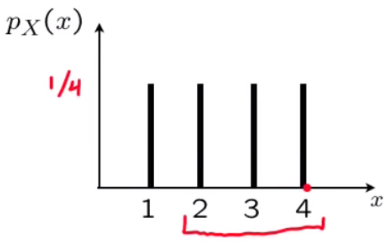
$$\sum_x p_X(x) = 1 \qquad \sum_x p_{X|A}(x) = 1$$

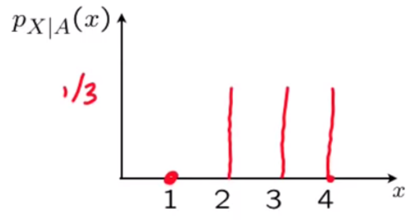
$$\mathbf{E}[X] = \sum_x x p_X(x) \qquad \mathbf{E}[X \mid A] = \sum_x x p_{X|A}(x)$$

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x) \qquad \mathbf{E}[g(X) \mid A] = \sum_x g(x) p_{X|A}(x)$$

Example of conditioning

- Let $A = \{X \geq 2\}$





$\mathbf{E}[X] = 2.5$

$\mathbf{E}[X \mid A] = 3$

$$\text{var}(X) = \frac{1}{12}(b-a)(b-a+2) = \frac{1}{12} 3 \cdot 5 = \frac{5}{4}$$

$$\text{var}(X \mid A) = \frac{1}{3}(4-3)^2 + \frac{1}{3}(3-3)^2 + \frac{1}{3}(2-3)^2 = \frac{2}{3}$$

In the last example, we saw that the conditional distribution of X, which was a uniform over a smaller range (and in some sense, less uncertain), had a smaller variance, i.e., Var $(X \mid A) \leq$ Var (X). Here is an example where this is not true. Let Y be uniform on $\{0, 1, 2\}$ and let B be the event that Y belongs to $\{0, 2\}$.

a) What is the variance of Y?

Var $(Y) =$ 8/12 ✔

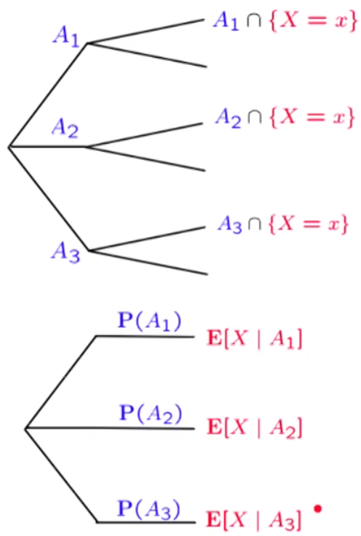
b) What is the conditional variance Var $(Y \mid B)$?

Var $(Y \mid B) =$ 1 ✔

Conditional probabilities allow us to divide and conquer

Reminder of Total Probability Theorem
translated to total expectation theorem

Total expectation theorem



$$\mathbf{P}(B) = \mathbf{P}(A_1)\,\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\,\mathbf{P}(B \mid A_n)$$

$B = \{X = x\}$

$$p_X(x) = \mathbf{P}(A_1)\,p_{X|A_1}(x) + \cdots + \mathbf{P}(A_n)\,p_{X|A_n}(x)$$

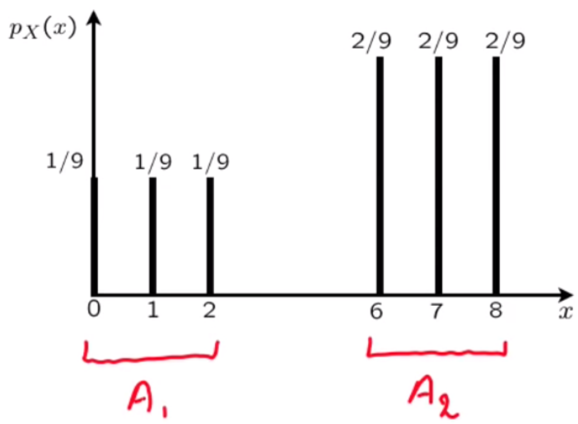
for all x

$$\underbrace{\sum_x x\, p_X(x)}_{E[X]} = P(A_1) \underbrace{\sum_x x\, p_{X|A_1}(x)}_{E[X|A_1]} + \cdots$$

$$\boxed{\mathbf{E}[X] = \mathbf{P}(A_1)\,\mathbf{E}[X \mid A_1] + \cdots + \mathbf{P}(A_n)\,\mathbf{E}[X \mid A_n]}$$

Each event is weighted by its probability

Total expectation example



$P(A_1) = \frac{1}{3}$

$P(A_2) = \frac{2}{3}$

$E[X|A_1] = 1$

$E[X|A_2] = 7$

$E[X] = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 7$

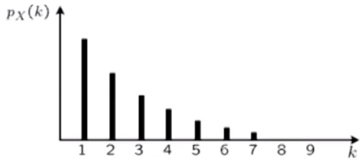
Geometric PMG, memorylessness and expectation

memorylessness: coin tosses have no 'memory', one coin toss isn't affected by the last

Conditioning a geometric random variable

- X: number of independent coin tosses until first head; $\mathbf{P}(H) = p$

$$p_X(k) = (1-p)^{k-1}p, \qquad k = 1, 2, \ldots$$



Memorylessness:
Number of **remaining** coin tosses, conditioned on Tails in the first toss, is **Geometric**, with parameter p

Conditioned on $X > \underset{=}{n}$, $X - \underset{=}{n}$ is geometric with parameter p

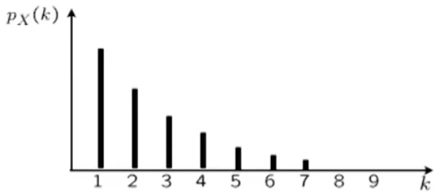
$$p_{X-1|X>1}(3) = \mathbf{P}(X-1=3 \mid X>1) = \mathbf{P}(T_2 T_3 H_4 \mid T_1) = \mathbf{P}(T_2 T_3 H_4)$$

$$= (1-p)^2 p = p_X(3)$$

$$p_{X-1|X>1}(k) = p_X(k) = p_{X-n|X>n}(k).$$

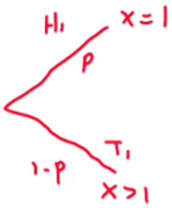
X>n is saying that the first n tosses were tails

The mean of the geometric



$$\mathbf{E}[X] = \sum_{k=1}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$\boxed{\mathbf{E}[X] = \frac{1}{p}.}$$



$$E[X] = 1 + E[X-1]$$
$$= 1 + p \cdot E[X-1 \mid X=1] + (1-p) E[X-1 \mid X>1]$$
$$= 1 + 0 + (1-p) E[X]$$

Exercise: Total expectation calculation

2/2 points (graded)

We have two coins, A and B. For each toss of coin A, we obtain Heads with probability $1/2$; for each toss of coin B, we obtain Heads with probability $1/3$. All tosses of the same coin are independent. We select a coin at random, where the probabilty of selecting coin A is $1/4$, and then toss it until Heads is obtained for the first time.

The expected number of tosses until the first Heads is: 2.75 ✔

For A E[h] = 1 / (1/2) x P(selecting A) = 1/4 = 0.5
For B E[h] = 1 / (1/3) x P(selecting B) = 3/4 = 2.25
0.5 + 2.25 = 2.75

Exercise: Memorylessness of the geometric

2/2 points (graded)

Let X be a geometric random variable, and assume that Var $(X) = 5$.

a) What is the conditional variance Var $(X - 4 \mid X > 4)$?

Var $(X - 4 \mid X > 4) =$ 5 ✔ **Answer:** 5

b) What is the conditional variance Var $(X - 8 \mid X > 4)$?

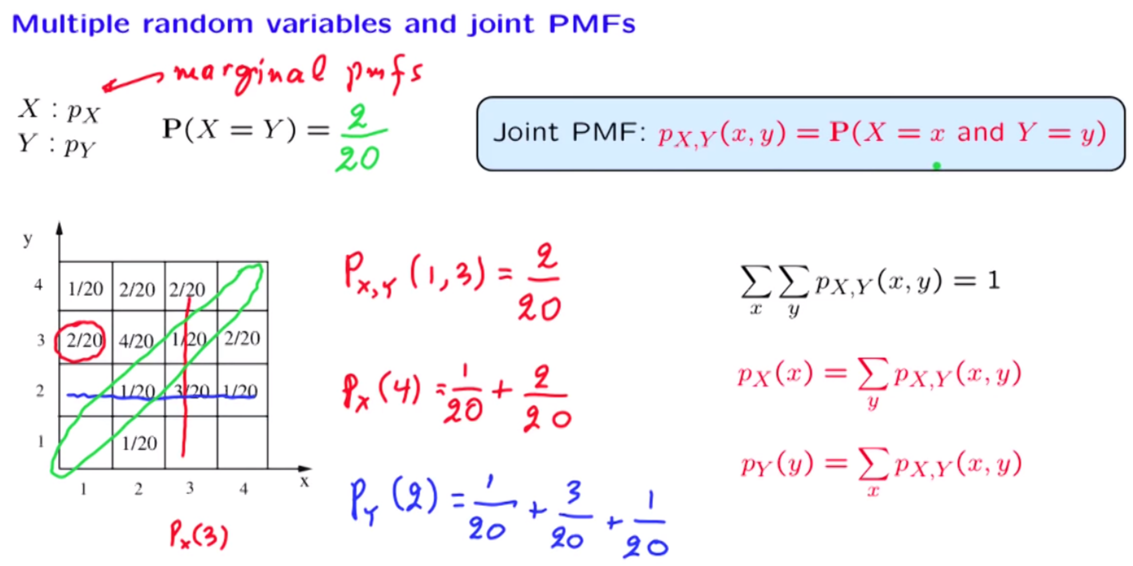
Var $(X - 8 \mid X > 4) =$ 5 ✔ **Answer:** 5

Solution:

a) The conditional distribution of $X - 4$ given $X > 4$ is the same geometric PMF that describes the distribution of X. Hence Var $(X - 4 \mid X > 4) =$ Var $(X) = 5$.

b) In the conditional model (i.e., given that $X > 4$), the random variables $X - 4$ and $X - 8$ differ by a constant. Hence they have the same variance and the answer is again 5.

Join PMFs and the expected value rule



Once we have a join PMF we can answer questions about the marginal PMFs (single)

More than two random variables

$$\boxed{p_{X,Y,Z}(x,y,z) = \mathbf{P}(X = x \text{ and } Y = y \text{ and } Z = z)}$$

$$\sum_x \sum_y \sum_z p_{X,Y,Z}(x,y,z) = 1$$

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x,y,z)$$

$$p_{X,Y}(x,y) = \sum_z p_{X,Y,Z}(x,y,z)$$

Functions of multiple random variables

$Z = g(X,Y)$

PMF: $p_Z(z) = \mathbf{P}(Z = z) = \mathbf{P}\big(g(X,Y) = z\big) = \sum_{(x,y):\, g(x,y)=z} p_{X,Y}(x,y)$

$$\boxed{\textbf{Expected value rule: } \mathbf{E}[g(X,Y)] = \sum_x \sum_y \underline{g(x,y)}\, \underline{p_{X,Y}(x,y)}}$$

$E[g(x)]$

Summing over x,y pairs to give an expected value of a random variable given by the function g(x,y)

Exercise: Joint PMF calculation

2/2 points (graded)

The random variable V takes values in the set $\{0, 1\}$ and the random variable W takes values in the set $\{0, 1, 2\}$. Their joint PMF is of the form

$$p_{V,W}(v, w) = c \cdot (v + w),$$

where c is some constant, for v and w in their respective ranges, and is zero everywhere else.

a) Find the value of c.

$c =$ 0.1111 ✔ **Answer:** 0.11111

b) Find $p_V(1)$.

$p_V(1) =$ 2/3 ✔ **Answer:** 0.66667

Solution:

a) The sum of the entries of the PMF is $c \cdot (0+0) + c \cdot (0+1) + c \cdot (0+2) + c \cdot (1+0) + \ldots = 9c$. Since this sum must be equal to 1, we have $c = 1/9$.

b)

$$p_V(1) = \sum_{w=0}^{2} p_{V,W}(1, w) = p_{V,W}(1, 0) + p_{V,W}(1, 1) + p_{V,W}(1, 2) = \frac{1}{9}(1 + 2 + 3) = \frac{6}{9}.$$

Let X and Y be discrete random variables. For each one of the formulas below, state whether it is true or false.

a) $\mathbf{E}[X^2] = \sum_x x p_X(x^2)$

False ✔ **Answer:** False

b) $\mathbf{E}[X^2] = \sum_x x^2 p_X(x)$

True ✔ **Answer:** True

c) $\mathbf{E}[X^2] = \sum_x x^2 p_{X,Y}(x)$

True ✖ **Answer:** False

d) $\mathbf{E}[X^2] = \sum_x x^2 p_{X,Y}(x, y)$

False ✔ **Answer:** False

e) $\mathbf{E}[X^2] = \sum_x \sum_y x^2 p_{X,Y}(x, y)$

False ✖ **Answer:** True

f) $\mathbf{E}[X^2] = \sum_z z p_{X^2}(z)$

True ✔ **Answer:** True

a) False. This does not follow from any of our formulas.

b) True. This is the expected value rule for a function of a single random variable.

c) False. This is syntactically wrong since the function $p_{X,Y}$ needs two arguments.

d) False. The left-hand side is a number whereas the right-hand side is actually a function of y.

e) True. This is the expected value rule

$$\mathbf{E}[g(X, Y)] = \sum_x \sum_y g(x, y)\, p_{X,Y}(x, y),$$

for the function $g(x, y) = x^2$.

f) True. This is just the definition of the expectation $\mathbf{E}[Z] = \sum_z z p_Z(z)$, where Z is the random variable X^2.

Linearity of expectations and the mean of the binomial

Derivation of linearity of expectations

Linearity of expectations

$$\mathbf{E}[aX+b] = a\mathbf{E}[X]+b$$

$$\mathbf{E}[X+Y] = \mathbf{E}[X]+\mathbf{E}[Y]$$

$$E[X+Y] = E[g(X,Y)]$$

$$(g(x,y) = x+y)$$

$$= \sum_x \sum_y (x+y)\, p_{X,Y}(x,y)$$

$$= \underbrace{\sum_x \sum_y x\, p_{X,Y}(x,y)} + \sum_x \sum_y y\, p_{X,Y}(x,y)$$

$$= \sum_x x \underbrace{\sum_y p_{X,Y}(x,y)} + \sim\sim$$

$$= \sum_x x\, p_X(x) + \sum_y y\, p_Y(y) = E[X] + E[Y]$$

And for multiple variables

$$\boxed{\mathbf{E}[X_1+\cdots+X_n] = \mathbf{E}[X_1]+\cdots+\mathbf{E}[X_n]}$$

The mean of the binomial

- X: binomial with parameters n, p
  - number of successes in n independent trials

$$\mathbf{E}[X] = \sum_{k=0}^{n} k \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{p_X(k)}$$

$$\boxed{\mathbf{E}[X] = np}$$

$X_i = 1$ if ith trial is a success; — p

$X_i = 0$ otherwise — $1-p$

(indicator variable)

$$X = X_1 + \cdots + X_n$$

$$E[X] = \underbrace{E[X_1]}_{p} + \cdots + \underbrace{E[X_n]}_{p} = np$$

Exercise: Linearity of expectations drill

1/1 point (graded)

Suppose that $\mathbf{E}[X_i] = i$ for every i. Then,

$\mathbf{E}[X_1 + 2X_2 - 3X_3] =$ -4

Exercise: Using linearity of expectations

2/2 points (graded)

We have two coins, A and B. For each toss of coin A, we obtain Heads with probability $1/2$; for each toss of coin B, we obtain Heads with probability $1/3$. All tosses of the same coin are independent.

We toss coin A until Heads is obtained for the first time. We then toss coin B until Heads is obtained for the first time with coin B.

The expected value of the total number of tosses is: 5 **Answer:** 5

Solution:

Let T_A and T_B be the number of tosses of coins A and B, respectively. We know that T_A is geometric with parameter $p = 1/2$, so that $\mathbf{E}[T_A] = 1/p = 1/(1/2) = 2$. Similarly, $\mathbf{E}[T_B] = 3$. The total number of coin tosses is $T_A + T_B$. Using linearity,

$$\mathbf{E}[T_A + T_B] = \mathbf{E}[T_A] + \mathbf{E}[T_B] = 2 + 3 = 5.$$