Parametric Families

Lecture 2

Please sign in on iClicker

- https://student.iclicker.com/

Outline

1. Review on Properties of Distributions
2. Random Variable Transformations
3. Distribution Families
4. Another Common Discrete Distribution Families

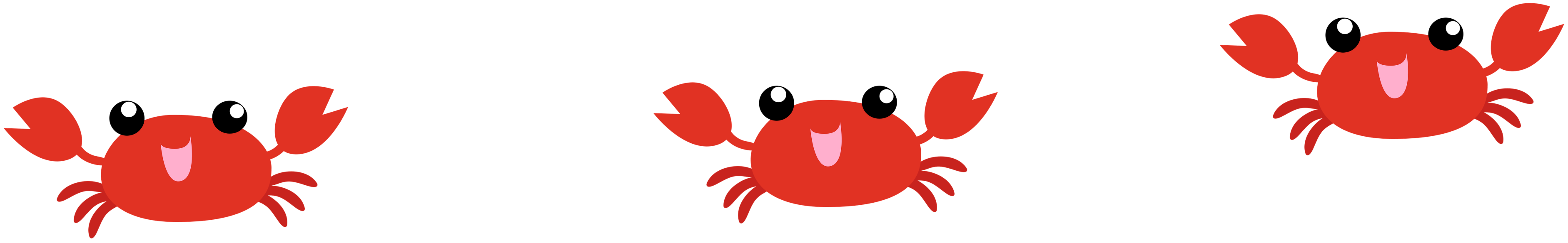
1. Review on Properties of Distributions

- We would start getting familiar with central tendency and uncertainty measures from `lecture1`.
- Let’s practice their computations with some in-class **iClicker**.

1.1. A Single Probability Mass Function

- Suppose X is a discrete random variable denoting the following:

$$X = \text{Number of crabs found at a nest in Spanish Banks.}$$

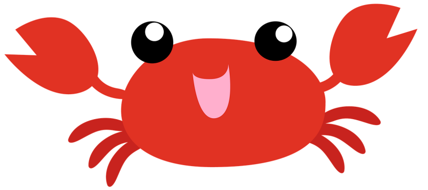


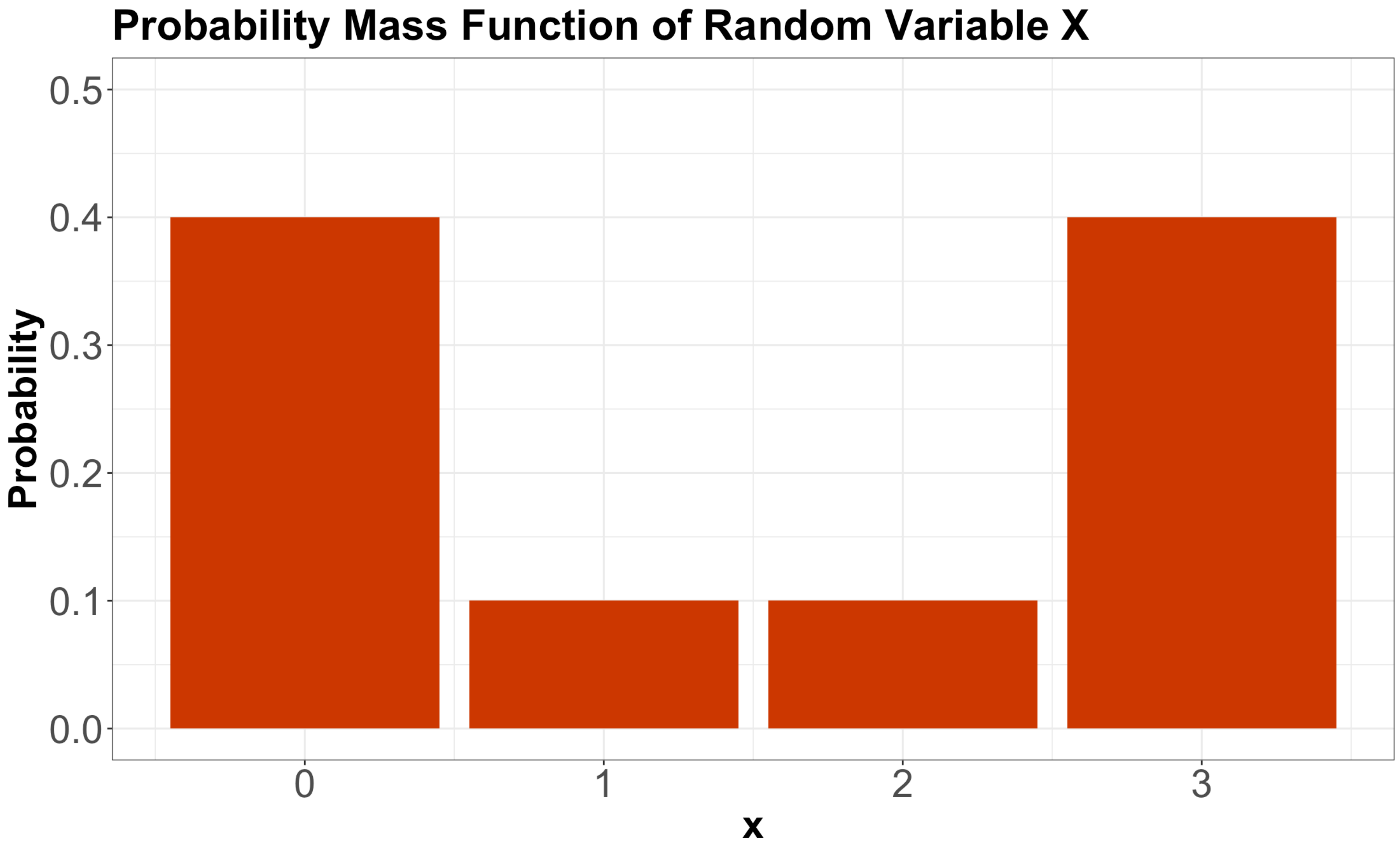
Probability Mass Function (PMF)

x	$P(X = x)$
0	0.4
1	0.1
2	0.1
3	0.4

- Conventionally, upper case letters denote random variables; the lower case letters denote the observed values.

We plot it as a bar chart…





iClicker Question: Mean

Using the PMF for random variable X, compute $\mathbb{E}(X)$.

Select the correct option:

A. 1

B. 1.5

C. 1.9

D. 6

x	$P(X = x)$
0	0.4
1	0.1
2	0.1
3	0.4

Answer

We compute the expected value as follows:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{3} x \cdot P(X = x) \\
&= 0(0.4) + 1(0.1) + 2(0.1) + 3(0.4) \\
&= 1.5.
\end{aligned}
$$

iClicker Question: Variance

Using the PMF for random variable X, compute the variance $\mathrm{Var}(X)$. Select the correct option:

A. 2.6

B. 1.85

C. 4.1

D. -1.85

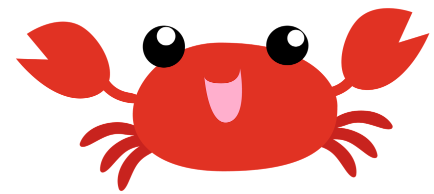
x	$P(X = x)$
0	0.4
1	0.1
2	0.1
3	0.4

Answer

- We can compute the variance of a random variable X in two forms:
  1. $\mathrm{Var}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\}$
  2. $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

And we already know that $\mathbb{E}(X) = 1.5$.

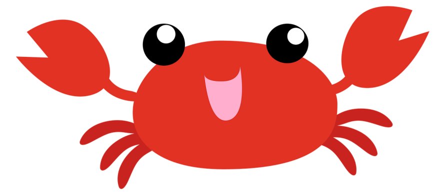
Method 1



Method 1

$$
\begin{aligned}
&\mathrm{Var}(X) \\
&= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} \\
&= \mathbb{E}[(X - 1.5)^2] \qquad\qquad \text{since } \mathbb{E}(X) = 1.5 \\
&= (-1.5)^2(0.4) + (-0.5)^2(0.1) + (0.5)^2(0.1) + (1.5)^2(0.4) \\
&= 1.85.
\end{aligned}
$$

Method 2



Method 2

$$
\begin{aligned}
&\mathrm{Var}(X) \\
&= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \mathbb{E}(X^2) - (1.5)^2 \qquad\qquad \text{since } \mathbb{E}(X) = 1.5 \\
&= (0)^2(0.4) + (1)^2(0.1) + (2)^2(0.1) + (3)^2(0.4) - (1.5)^2 \\
&= 1.85.
\end{aligned}
$$

iClicker Question: Mode

Using the PMF for random variable X, obtain the mode $\text{Mode}(X)$. Select the correct option:

A. 0

B. 3

C. Both 0 and 3

D. Neither

x	$P(X = x)$
0	0.4
1	0.1
2	0.1
3	0.4

Answer

The mode are the outcomes with the largest probabilities in the PMF, i.e.,

$$\text{Mode}(X) = 0 \text{ and } 3.$$

iClicker Question: Entropy

Using the PMF for random variable X, obtain the entropy $H(X)$. Select the correct option:

A. -1.19

B. 0.52

C. -0.52

D. 1.19

x	$P(X = x)$
0	0.4
1	0.1
2	0.1
3	0.4

Answer

We use compute the entropy as follows:

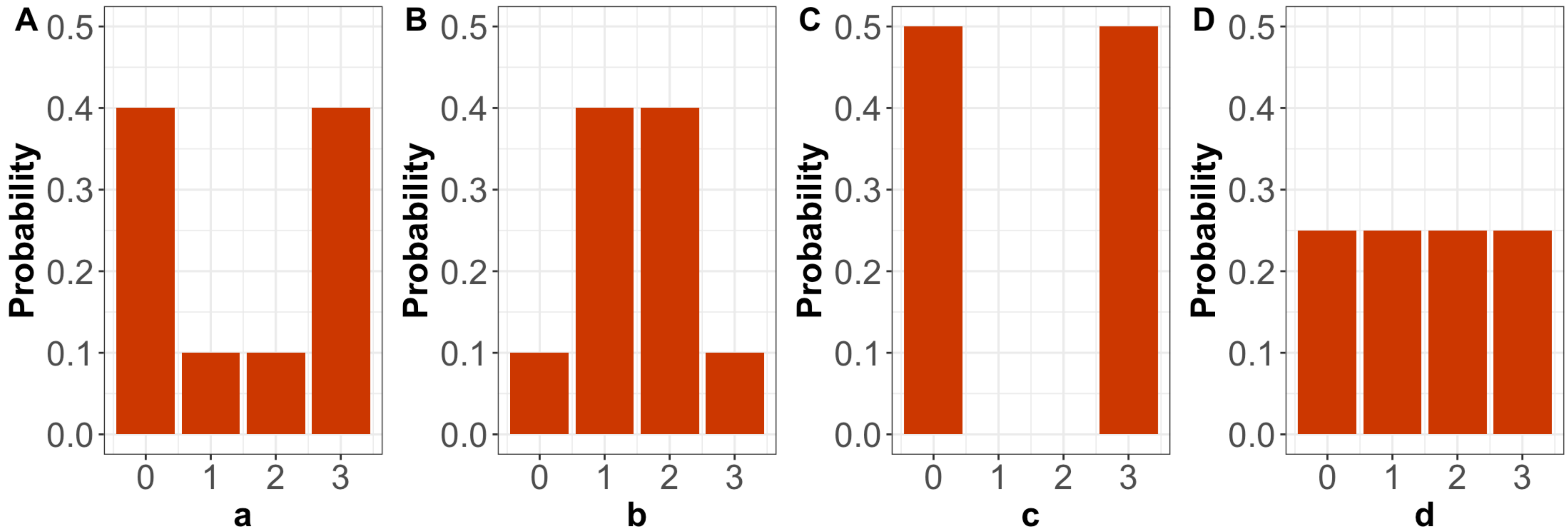
$$
\begin{aligned}
H(X) &= -\sum_{x=0}^{3} P(X = x) \log[P(X = x)] \\
&= -[0.4 \log(0.4) + 0.1 \log(0.1) + \\
&\qquad 0.1 \log(0.1) + 0.4 \log(0.4)] \\
&= 1.19.
\end{aligned}
$$

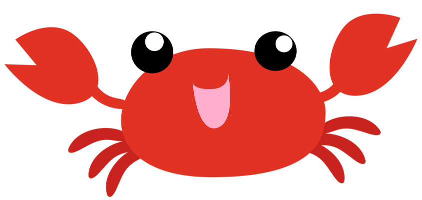
1.2. Comparing Multiple Probability Mass Functions

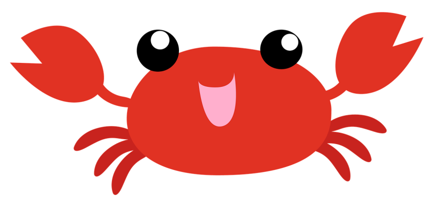
Suppose there are four different random variables related to four locations:

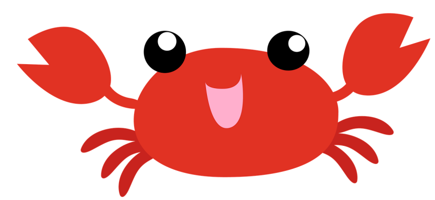
$$
\begin{aligned}
A = &\quad \text{Number of crabs found at a nest at Location A.} \\
B = &\quad \text{Number of crabs found at a nest at Location B.} \\
C = &\quad \text{Number of crabs found at a nest at Location C.} \\
D = &\quad \text{Number of crabs found at a nest at Location D.}
\end{aligned}
$$

Probability Mass Functions (PMFs)









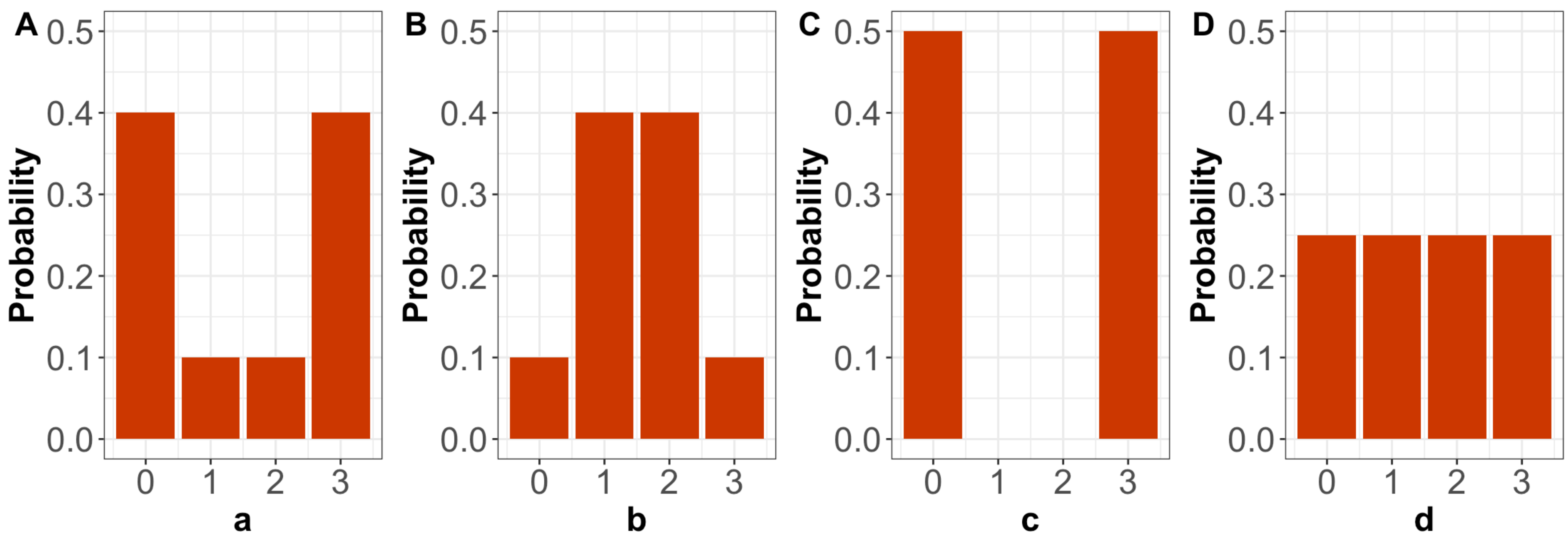
iClicker Question

Answer **TRUE** or **FALSE**:

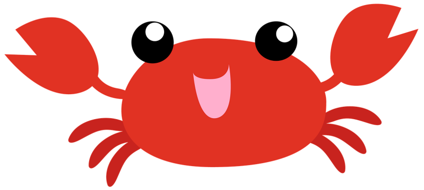
By only looking at the PMFs, A has higher entropy than B.

A. TRUE

B. FALSE



Answer



- It is **FALSE**.
- Both A and B have the same entropy.
- The entropy does not look at the outcome on the horizontal axis.
- It just looks at the probabilities, and both random variables have equivalent sets of probabilities.

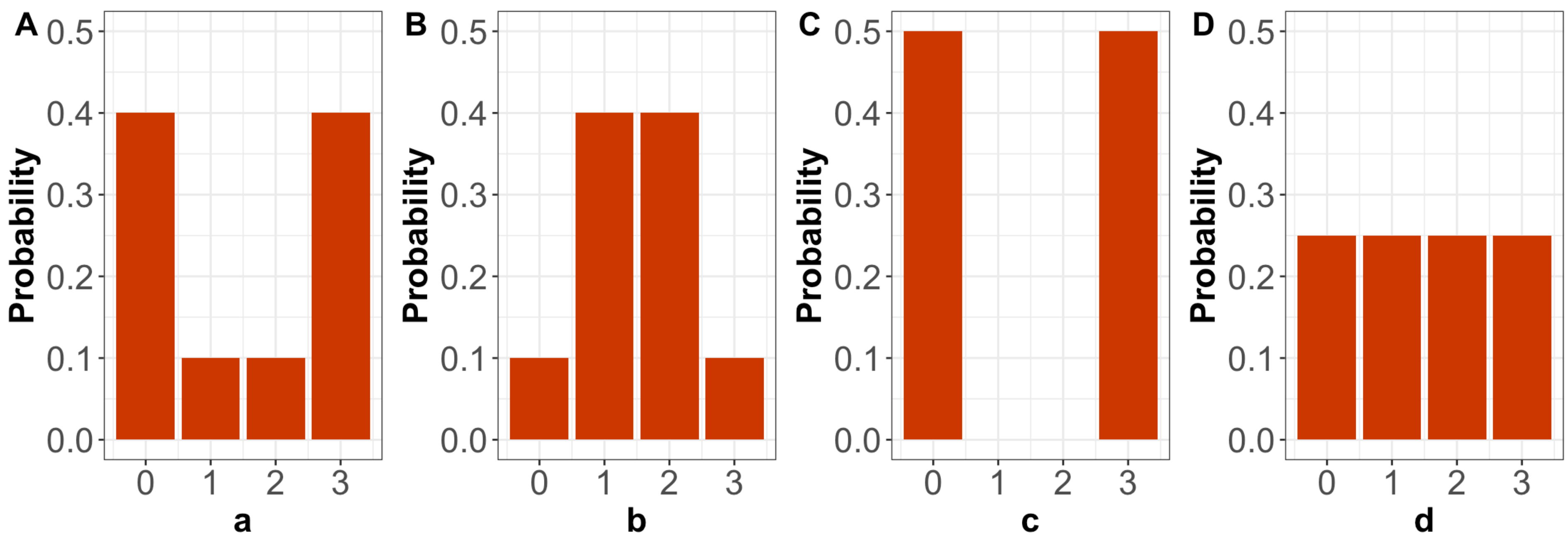
iClicker Question

Answer **TRUE** or **FALSE**:

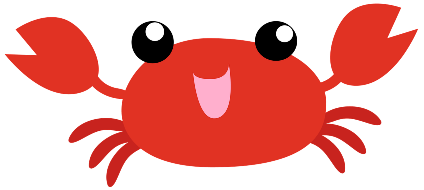
By only looking at the PMFs, A has higher variance than B.

A. TRUE

B. FALSE



Answer

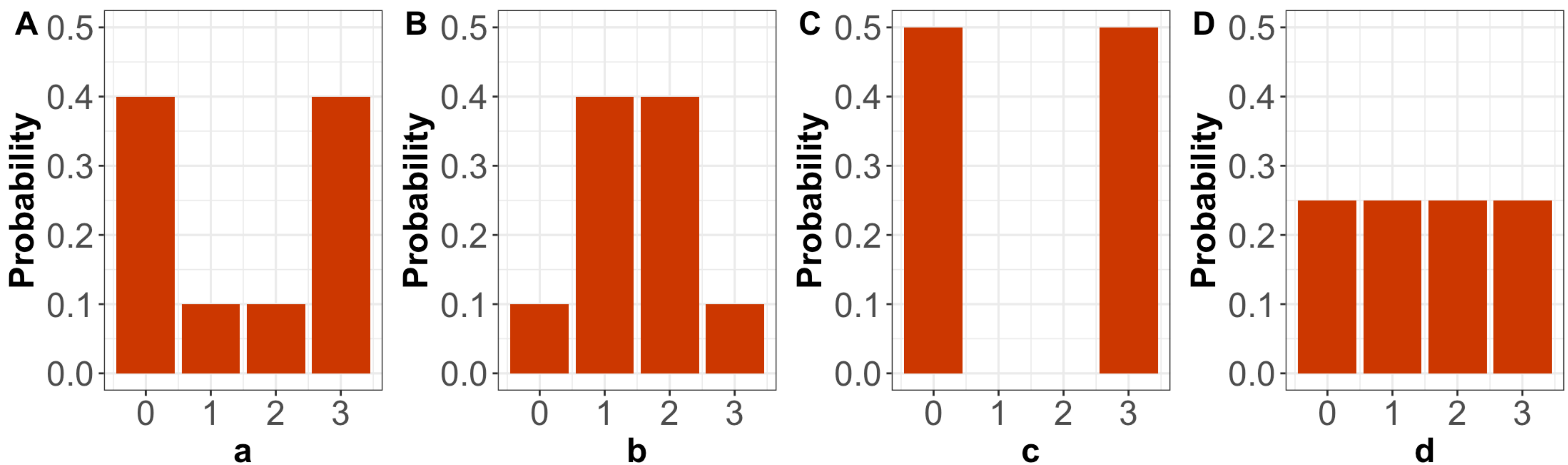


- It is **TRUE**.
- Variance measures how much a random variable deviates from its mean.
- In both A and B, the mean is 1.5.
- However, in A, higher probabilities are associated with values further from the mean than in B.
- Therefore, we have a larger variance in A.

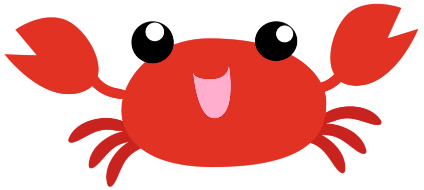
iClicker Question

By only looking at the PMFs, which RV has the highest variance?

A. A **B.** B **C.** C **D.** D



Answer

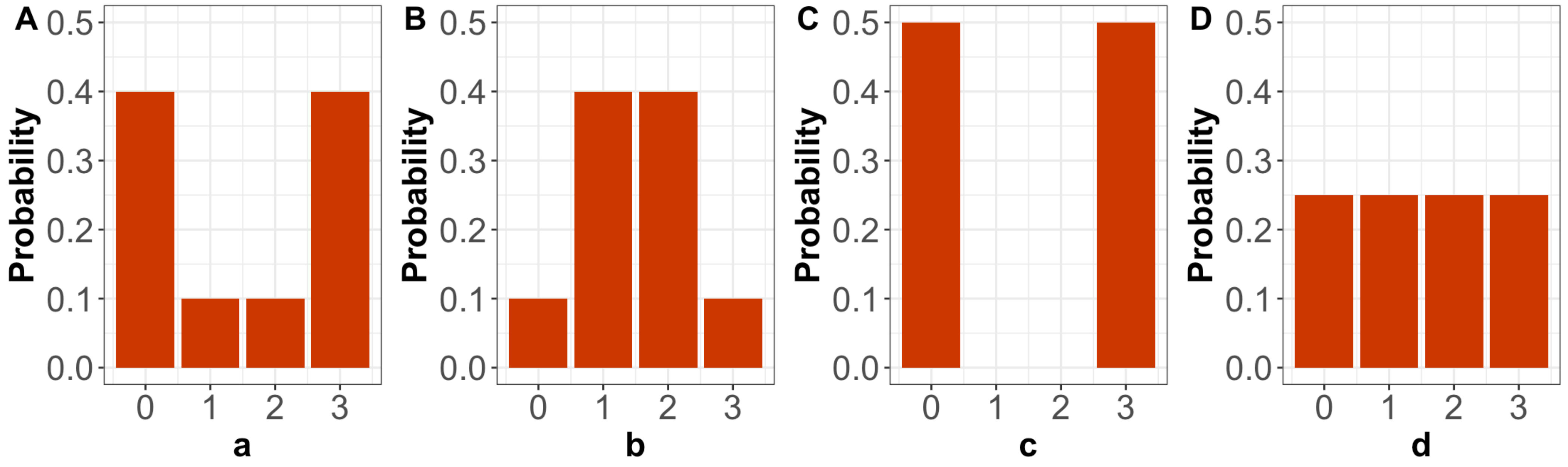


- It is **C**.
- Again, the four distributions have a mean of 1.5.
- C only has two extreme possible outcomes: 0 and 3.

iClicker Question

By only looking at the PMFs, which RV has the highest entropy?

A. A **B.** B **C.** C **D.** D



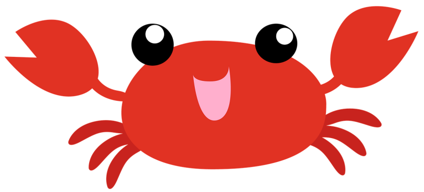
Answer

- It is **D**.
- To maximize entropy, you need equal probabilities for all the outcomes, which is one quarter in the case of D.
- This indicates we have a uniform uncertainty over the whole range of possible outcomes.

2. Random Variable Transformations

- A random variable can be transformed into other random variables via mathematical operations.
- This feature is crucial in data modelling!

2.2. Distribution Mapping

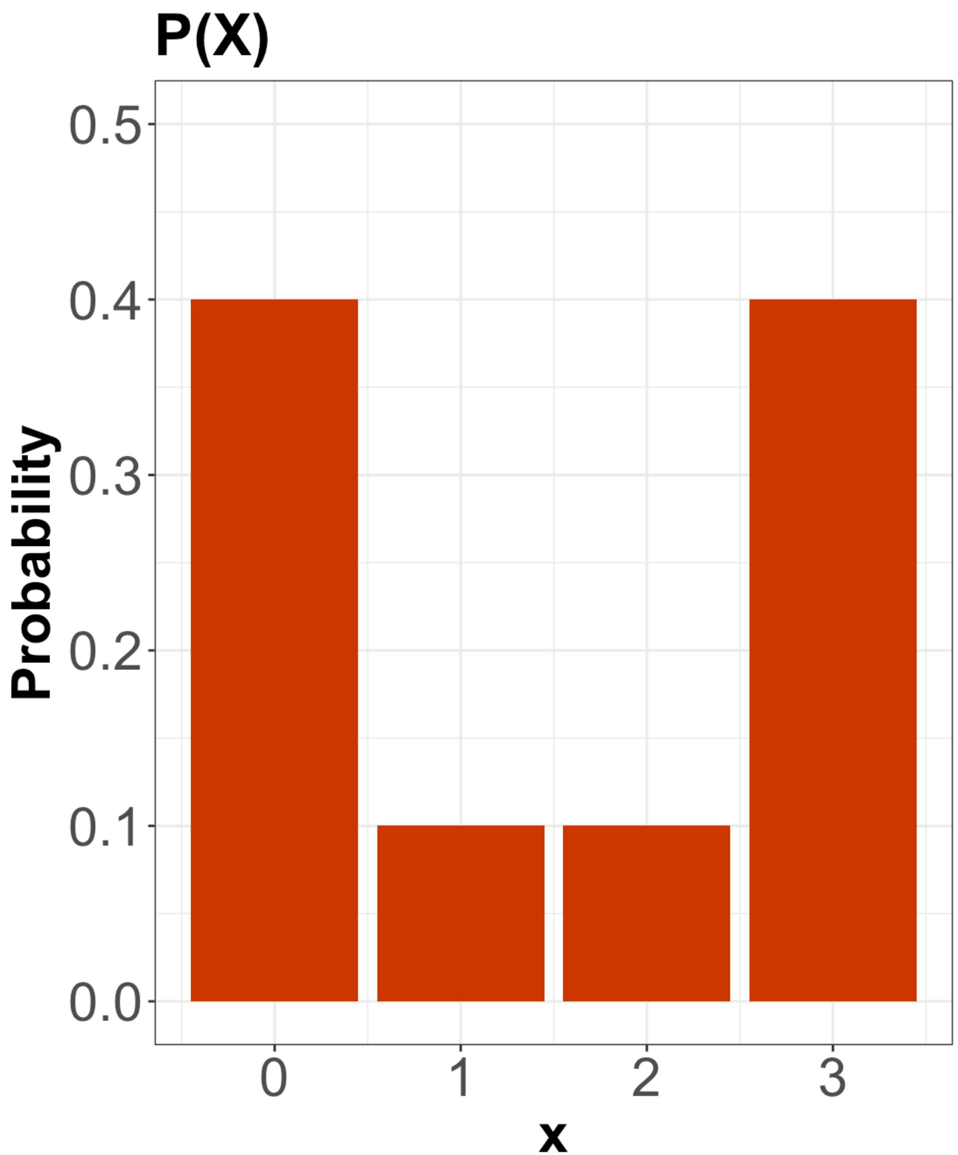


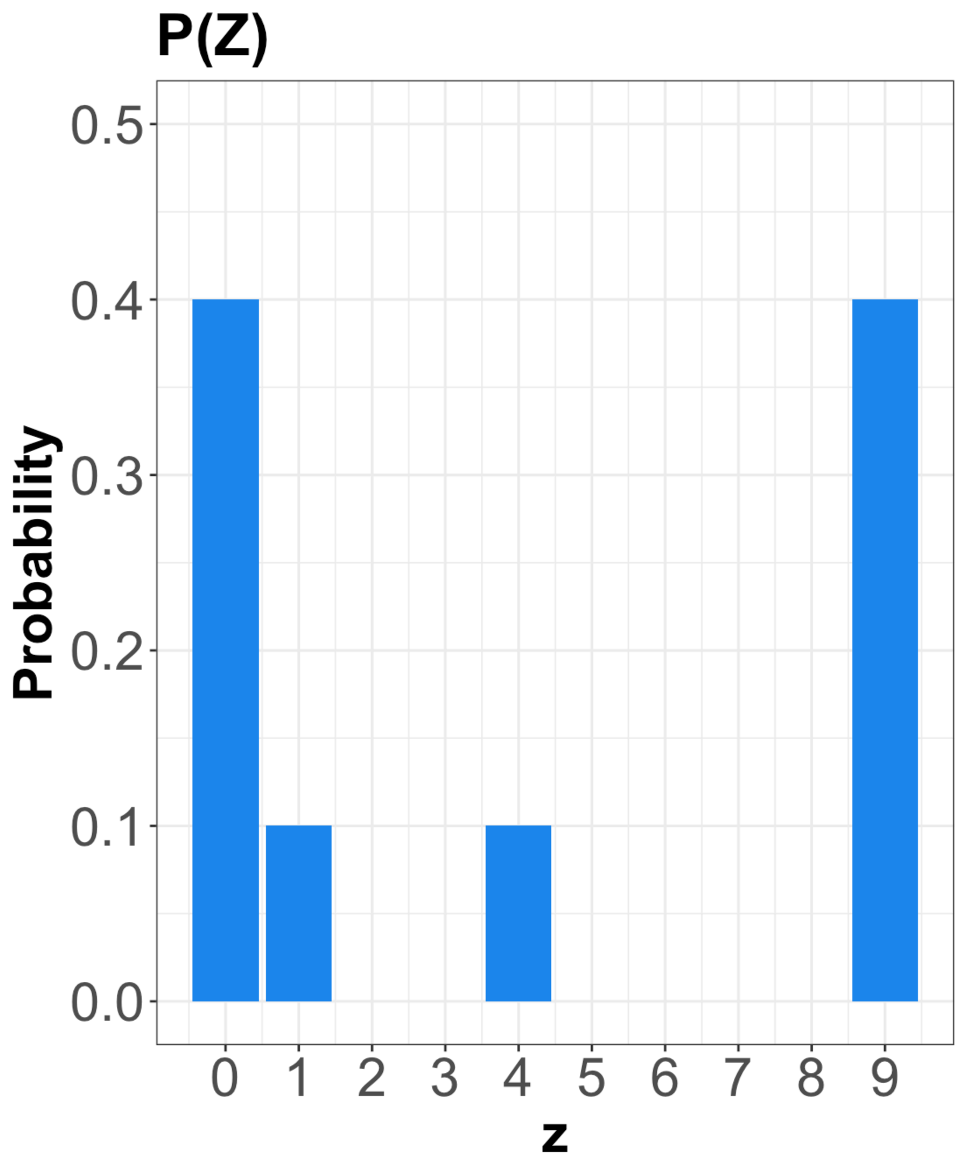
- Following up with the crabs PMF, let us focus on $\mathbb{E}(X^2)$.
- More specifically, what does X^2 mean?
- We can define a new RV as

$$Z = X^2.$$

- This is a random variable transformation.

Comparing PMFs





The Expected Value

- We can compute the expected value of Z by its defintion

$$\mathbb{E}(Z) = \sum_z zP(Z = z)$$

- Or we can use the PMF of X to do so:

$$\mathbb{E}(Z) = \mathbb{E}(X^2) = \sum_x x^2 P(X = x)$$

- The second approach is more straightforward sicne we don't need to calculate the PMF of Z

UBC Master of Data Science

Law of the Unconscious Statistician (LOTUS)

- More generally, for any function g

$$\mathbb{E}(g(X)) = \sum_{x} g(x)P(X = x)$$

- $Z = X^2$ is a special case for $g(x) = x^2$.
- It is so intuitive that many statisticians were using this method without fully realizing it!

2.3. Expected Value Properties

- Expected values have certain useful properties under linear transformations.
- If a and b are constants, with X and Y as random variables, then we can obtain the expected value of the following expressions as:

$$\mathbb{E}(aX) = a\mathbb{E}(X)$$
$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Caution

- The operator $\mathbb{E}(\cdot)$ does not follow the usual algebraic rules.
- For instance, if no further assumptions are made for random variables X and Y, then

$$\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y).$$

- And

$$\mathbb{E}(X^2) \neq [\mathbb{E}(X)]^2.$$

2.3. Variance Properties

- If a and b are constants, with X and Y as independent random variables, then we can obtain the variance of the following expressions as:

$$
\begin{gathered}
\operatorname{Var}(aX) = a^2 \operatorname{Var}(X) \\
\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) \\
\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y).
\end{gathered}
$$

3. Distribution of Families

- A significant part of Data Science is to model data as random variables.
  - Example: The number of tickets sold for a Vancouver Canucks game (i.e., a discrete count random variable).
  - A Poisson distribution can be used to model the count data.
- Many probability distributions that are important in theory or applications have been given specific names.

3.1. Bernoulli

- Consider an experiment where the outcome is a “success” with probability p
- Let X be a binary random variable as follows:

$$X = \begin{cases} 1 & \text{if success,} \\ 0 & \text{otherwise.} \end{cases}$$

PMF

- This is called a Bernoulli distribution:

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

- Its PMF is

$$P(X = x) = p^x (1 - p)^{1-x} \quad \text{for} \quad x = 0, 1.$$

Mean

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{1} x \cdot P(X = x) \\
&= 0 \cdot p^0 (1 - p)^{1-0} + 1 \cdot p^1 (1 - p)^{1-1} \\
&= p
\end{aligned}
$$

Variance

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \mathbb{E}(X^2) - p^2 \qquad\qquad \text{since } \mathbb{E}(X) = p \\
&= \sum_{x=0}^{1} x^2 \cdot P(X = x) - p^2 \\
&= p(1-p).
\end{aligned}$$

3.2. Binomial

- Consider an experiment where each trial is a "success" with probability p
- Let X be the number of successes in n independent trials.
- X is said to follow a Binomial distribution, written as

$$X \sim \text{Binomial}(n, p).$$

PMF

- A Binomial distribution is characterized by the PMF

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for} \quad x = 0, 1, \ldots, n.$$

- where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Example

- Let us derive the probability of winning exactly two games out of five, with winning probability $p = 0.25$
- That is, we want to know $P(X = 2)$ for $X \sim \text{Binomial}(5, 0.25)$:

$$
\begin{aligned}
P(X = 2) &= \binom{5}{2}(0.25)^2(1 - 0.25)^{5-2} \\
&= \frac{5!}{2!(5-2)!}(0.25)^2(1 - 0.25)^{5-2} \\
&= 0.26.
\end{aligned}
$$

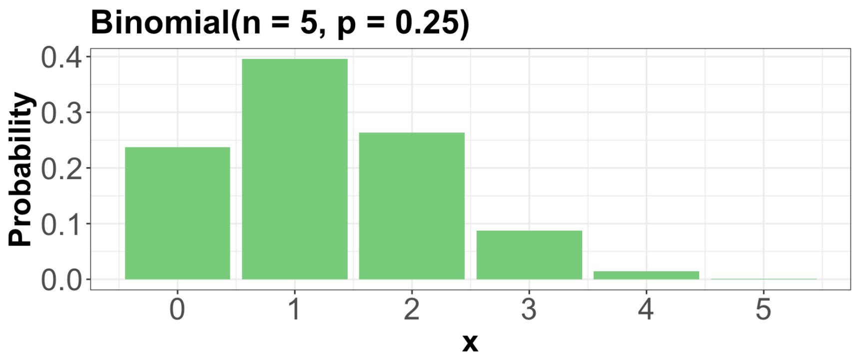
Mean and Variance

$$\mathbb{E}(X) = np$$

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

3.3. Families Versus Distributions

- Specifying a value for both p and n results in a unique Binomial distribution.

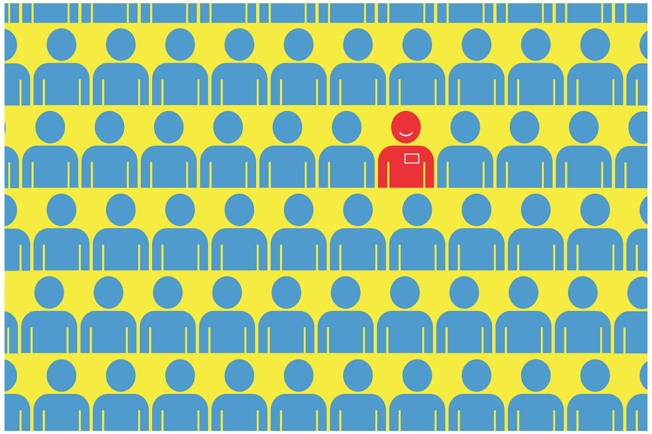


- There are, in fact, infinite Binomial distributions.

- We refer to the entire set of probability distributions as the Binomial family of distributions.

3.4. Parameters

- A parameter is a specific variable that determines the characteristics of a distribution within a distribution family.
- Parameters narrow down the set of possible distributions to a unique one within the family.
  - p and n fully specify a Binomial distribution, we call them parameters of the Binomial family.



3.5. Parameterization

- Parameterization refers to the set of parameters used to identify a distribution within a family.
- The Binomial distribution is usually parameterized according to n and p.

Parameterization for Bernoulli Distribution

- How many parameters do we need to fully specify a Bernoulli Distribution?
- There are many ways in which a distribution family can be parameterized. We can use p or $q = 1 - p$ as the parameter.
- There is often a “usual” parameterization, which we call the canonical (natural) parameterization.

4. Another Common Discrete Distribution Families

- Aside from the Binomial family of distributions, many other families come up in data modelling.
- In practice, it is rare to encounter situations that a distribution family exactly describes, but distribution families still act as useful approximations.

4.1. Geometric

- Consider an experiment where each trial is a "success" with probability p
- Let X be the number of failures before the first success in a sequence of independent trials.
- X is said to have a Geometric distribution, written as

$$X \sim \text{Geometric}(p).$$

PMF

- A Geometric distribution is characterized by the PMF

$$P(X = x) = p(1 - p)^x \quad \text{for} \quad x = 0, 1, \ldots$$

- Since there is only one parameter, this means that if you know the mean, you also know the variance!
- It has an infinite support.

Mean and Variance

$$\mathbb{E}(X) = \frac{1-p}{p}$$

$$\text{Var}(X) = \frac{1-p}{p^2}.$$

4.2. Negative Binomial (a.k.a. Pascal)

- Consider an experiment where each trial is a "success" with probability p
- Let X be the number of failures before experiencing k success in a sequence of independent trials.
- X is said to have a Negative Binomial distribution, written as

$$X \sim \text{Negative Binomial}(k, p).$$

PMF

- A Negative Binomial distribution is characterized by the PMF

$$P(X = x) = \binom{k - 1 + x}{x} p^k (1 - p)^x \quad \text{for} \quad x = 0, 1, \ldots$$

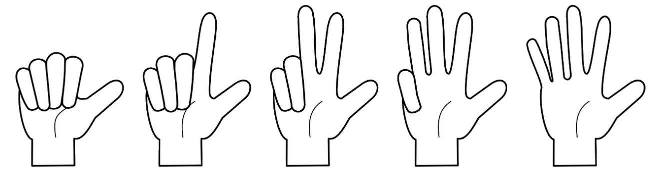
- It has two parameters: k and p.
- The Geometric family is a special case with $k = 1$.

Mean and Variance

$$\mathbb{E}(X) = \frac{k(1-p)}{p}$$

$$\mathrm{Var}(X) = \frac{k(1-p)}{p^2}.$$

4.3. Poisson



- A Poisson RV gives the probability of a given number of events in a fixed interval of time (or space).
- Suppose customers independently arrive at a store at some average rate λ.
- Let X denotes the total number of customers arriving after a pre-specified length of time. Then X follows a Poisson distribution:

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

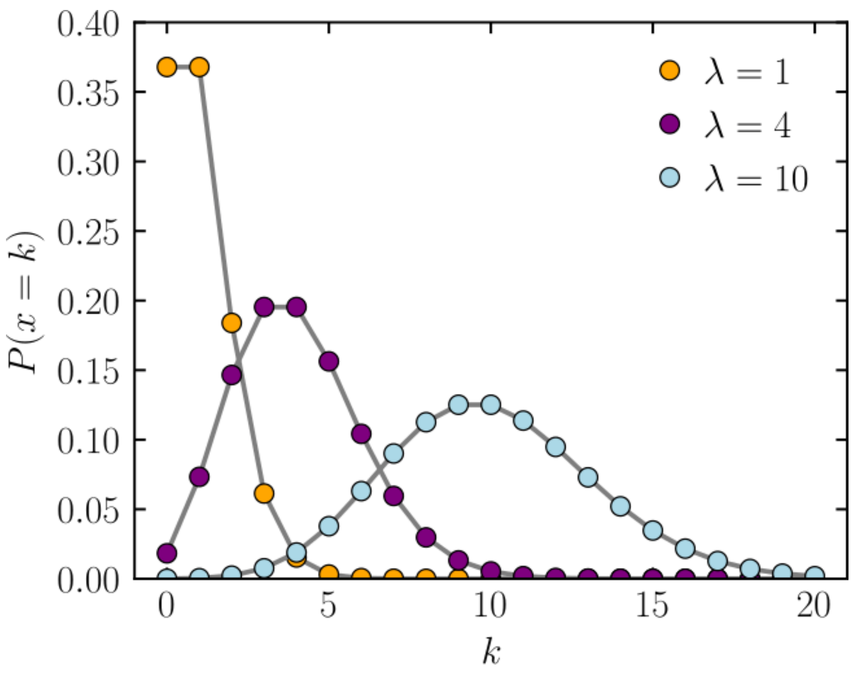
More Examples

- We can find other examples where the Poisson distribution serves as a good approximation:
  - The number of ships arriving at Vancouver port on a given day.
  - The number of emails you receive on a given day.

PMF

- A Poisson distribution is characterized by the PMF

$$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!} \quad \text{for} \quad x = 0, 1, \ldots$$



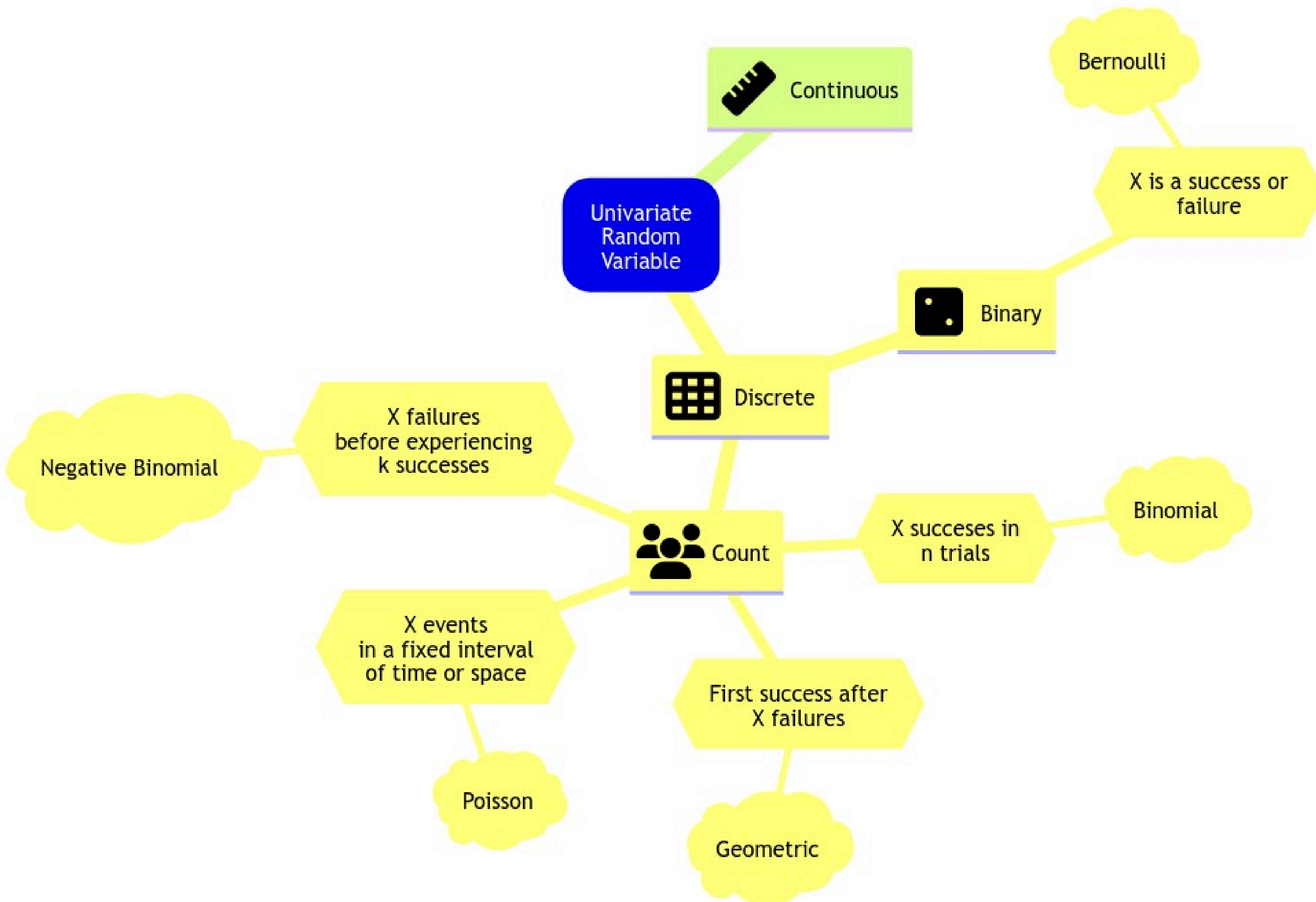
Mean and Variance

$$\mathbb{E}(X) = \lambda$$

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

- A notable property of this family is that the mean is equal to the variance!

4.5. Finally, let us check this mindmap…



Today’s Learning Objectives

- Understand random variable transformations, e.g., X^2.
- Calculate expectations and variances of random variable transformations.
- Distinguish between a family of distributions and a distribution.
- Calculate probabilities, mean, and variance of a distribution belonging to a distribution family.
- Identify whether a set of parameters is enough to specify a distribution from a family of distributions.
- Match a physical process to a distribution family (Binomial, Geometric, Negative Binomial, Poisson, and Bernoulli).