

# 02-680 Module 14

## Essentials of Mathematics and Statistics

Aidan Jan

November 11, 2025

### Random Variables

The formal definition of a random variable  $X$  is a mapping:

- A random variable is a **function** that maps each outcome in a sample space  $\Omega$  to a real number.
- This mapping is not a probability by itself.

$$X : \Omega \rightarrow \mathbb{R}$$

that assigns a real number  $X(\omega)$  to each outcome  $\omega$ .

### Example

Let's say we toss a coin twice, we know we get  $\Omega_{\text{twocoin}}$  as defined previously. We can assign a variable  $C(\omega)$  to be the count of the number of heads in each outcome.

$\omega$	$p(\omega)$	$C(\omega)$	$c$	$p(C = c)$
HH	1/4	2	0	1/4
HT	1/4	1	1	1/2
TH	1/4	1	2	1/4
TT	1/4	0		

We can then sample probabilities of the possible values a random variable  $X$  can take on (its range), we usually use a lower case of the same name,  $x$ .

We can think of this as the sum of the probabilities of all the outcomes that lead to a particular value.

### Random Variable $\neq$ Event

**Common misconception:** A random variable is NOT an event.

Example:

- Toss 2 coins:  $\Omega = \{HH, HT, TH, TT\}$
- Define  $X(\omega) = \text{number of heads}$ 
  - $X(HT) = 1$
- Then the event " $X = 1$ " is the set  $\{HT, TH\}$

## Discrete Random Variables - Probability Mass Functions

$X$  is **discrete** if it takes countably many values  $\{x_1, x_2, \dots\}$ . We define the **probability function** or **probability mass function (PMF)** for  $X$  by  $f_X(x) = P(X = x)$ .

Thus,  $f_X(x) \geq 0$  for all  $x \in \mathbb{R}$ , and  $\sum_i f_X(x_i) = 1$ . Sometimes, we write  $f$  instead of  $f_X$ . The **cumulative distribution function (cdf)** of  $X$  is related to  $f_X$  by

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

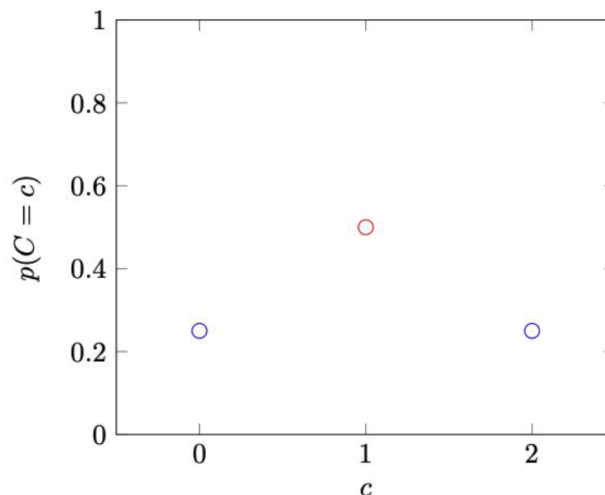
### Example

Suppose we toss a coin twice,  $\text{twocoin} = \{HH, HT, TH, TT\}$ . Define a random variable  $C(\omega)$  as the number of heads in outcome  $\omega$ . Flip a coin twice and let  $X$  be the number of heads.

The probability function is:

$$f_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

We call the probability distribution a **probability mass function (PMF)**. Below is the visualization of the PMF for  $C$  above.



### Well-Known PMFs: Bernoulli Distribution

A well-known PMF is a Bernoulli Distribution, which is defined as follows with probability of success ( $k = 1$ )  $0 \leq \alpha \leq 1$ , for  $\Omega = \{0, 1\}$ :

$$p(k) = \begin{cases} \alpha & k = 1 \\ 1 - \alpha & k = 0 \end{cases}$$

or put another way:

$$p(k) = \alpha^k (1 - \alpha)^{(1-k)} \quad \forall k \in \Omega$$

## Multinoulli Distribution: Bernoulli

$X$  is a discrete random variable that takes values from  $\{1, \dots, k\}$ . Multinoulli distribution is  $X \sim \text{Multinoulli}(p_1, \dots, p_k)$ .

$$P(X = x) = p_1^{\delta(x,1)} \cdot p_2^{\delta(x,2)} \cdots p_k^{\delta(x,k)}$$

where

$$\delta(x, i) = \begin{cases} 1 & x = i \\ 0 & x \neq i \end{cases}$$

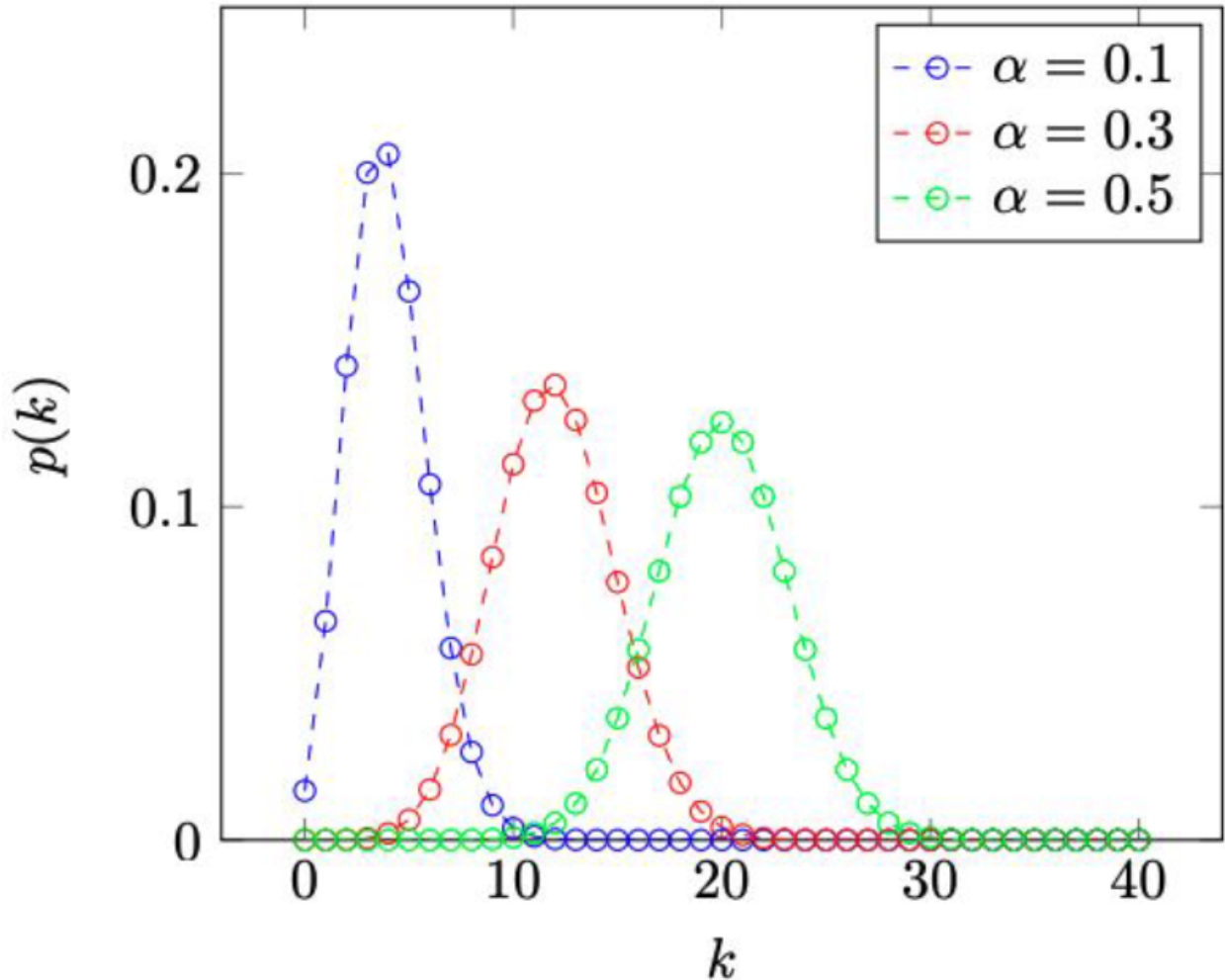
### 0.1 Well-Known PMFs: Binomial Distribution

Another (maybe the most) well-known PMF is the **Binomial Distribution**, again we say the probability of a (single) trial's success is  $0 \leq \alpha \leq 1$ , but now we have  $n$  trials and we want to know the probability of getting  $k \leq n$  successes (this makes  $\Omega = \{0, 1, 2, \dots, n\}$ ).

In this case, the **probability mass function** is:

$$p(k) = \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \quad \forall k \in \Omega$$

For example, consider  $n$  independent coin flips, each with a probability  $\alpha$  of landing heads. The total number of heads  $k$  follows a Binomial distribution.



The plot shows how the distribution shifts as  $\alpha$  increases: with  $\alpha = 0.1$ , most outcomes are near 0. With  $\alpha = 0.3$ , the peak moves right, and with  $\alpha = 0.5$ , the distribution is symmetric around  $n/2$ .

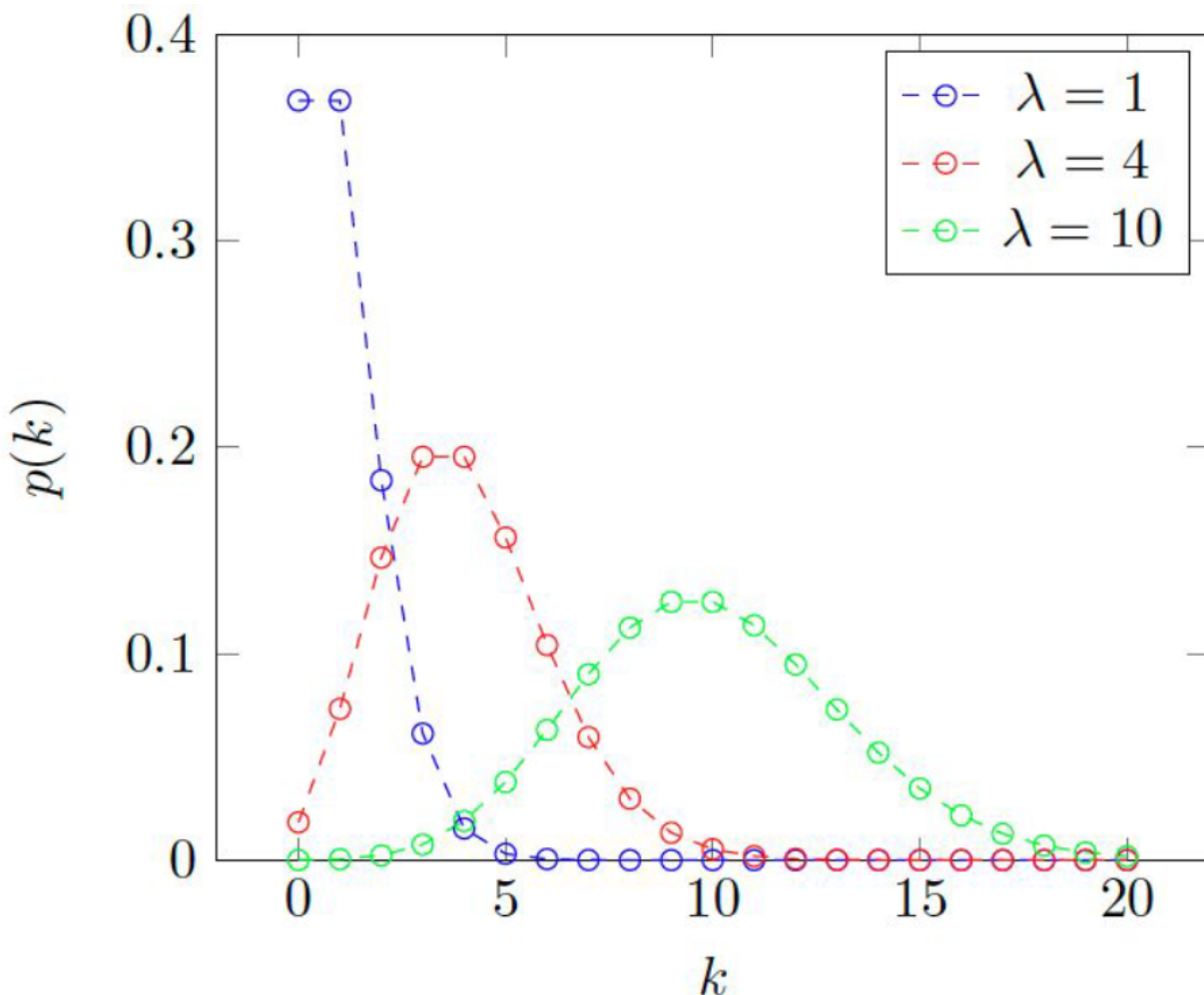
## Well-Known PMFs: Poisson Distribution

Poisson distributions are used to model rare events over time. The Poisson distribution models the probability of a given number of events ( $k$ ) occurring in a fixed interval of time or space, when these events happen independently and with a constant average rate  $\lambda$ . The **probability mass function** is:

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \forall k \in \Omega$$

- $k$  is the number of events observed
- $\lambda$  is the expected number of events in the interval

Poisson distributions for different  $\lambda$  values are shown below.  $\lambda$  represents the average number of events. As  $\lambda$  increases, the distribution shifts right and spreads out. The  $x$ -axis shows event counts  $k$ , and the  $y$ -axis shows their probabilities. Smaller  $\lambda$  means most events are rare, while larger  $\lambda$  centers around higher  $k$ .



## Poisson Distribution: Real World Scenario

One of the first applications of the Poisson distribution was by statistician Ladislaus Bortkiewicz. In the late 1800s, he investigated accidental deaths by horse kick of soldiers in the Prussian army. He analyzed 20

years of data for 10 army corps, equivalent to 200 years of observations of one corps.

He found that a mean of 0.61 soldiers per corps died from horse kicks each year. However, most years, no soldiers died from horse kicks. On the other end of the spectrum, one tragic year there were four soldiers in the same corps who died from horse kicks.

This scenario can be modeled using the Poisson distribution, where the number of deaths per corps per year is a discrete random variable.

Let  $X \sim \text{Poisson}(\lambda = 0.61)$  denote the number of deaths per year for a single corps.

## Well-Known PMFs: Geometric Distribution

The geometric distribution models the number of **Bernoulli Trials** needed to get the **first success**.

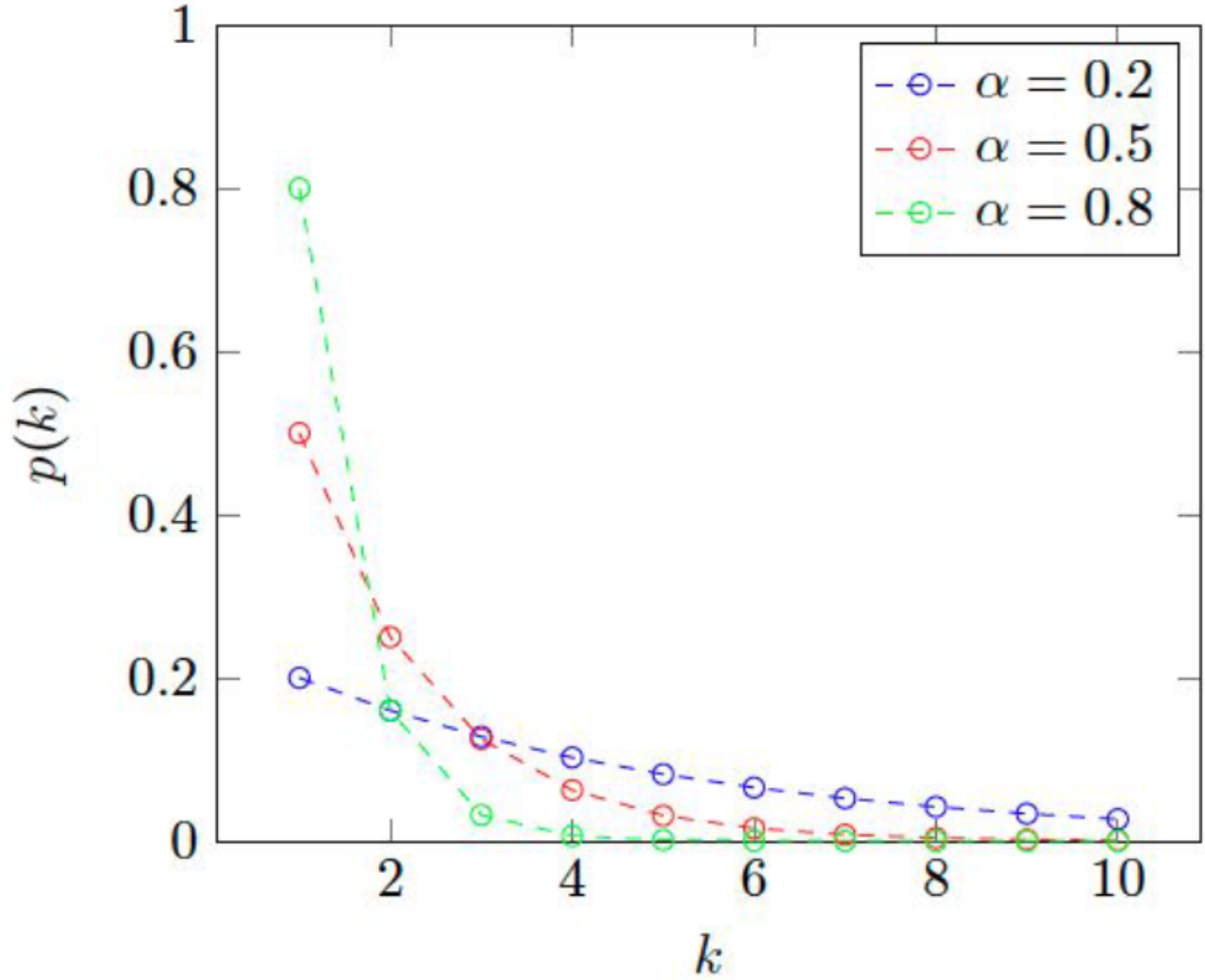
- Parameter:  $\alpha \in (0, 1)$ , the probability of success in each trial.
- Support:  $k = 1, 2, 3, \dots$  - the trial on which the first success occurs.
- Probability mass function (PMF)

$$p(k) = (1 - \alpha)^{k-1} \alpha \quad \forall k \in \Omega$$

The probability that the first success happens on the  $k$ -th trial. For example, consider the probability that the  $k$ -th coin-flip is the first head (probability of heads  $\alpha$ ) with the previous throws being tails.

The plot shows how the probability  $p(k)$  of getting the first head on the  $k$ -th coin flip changes with different values of success probability  $\alpha$ :

- Larger  $\alpha$  (e.g., 0.8): success is more likely early  $\rightarrow$  sharp drop.
- Smaller  $\alpha$  (e.g., 0.2): more trials are expected  $\rightarrow$  slower decay.



This reflects how the geometric distribution models the waiting time until the first success.

## Real-world Application

### Modeling Gene Expression as Random Variables

In RNA-seq, the number of reads mapping to each gene is modeled as a discrete random variable.

For gene  $G_i$ , the read count  $X_i \sim \text{Poisson}(\lambda_i)$ , where  $\lambda_i$  is the expected expression level.

For example:

$X_i$  = observed count of gene  $G_i$

$$P(X_i = k) = \frac{e^{-\lambda_i} \lambda_i^k}{k!}$$