

02-680 Module 14

Essentials of Mathematics and Statistics

Aidan Jan

November 13, 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 Ω 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 ω .

Example

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

ω	$p(\omega)$	$C(\omega)$	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

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

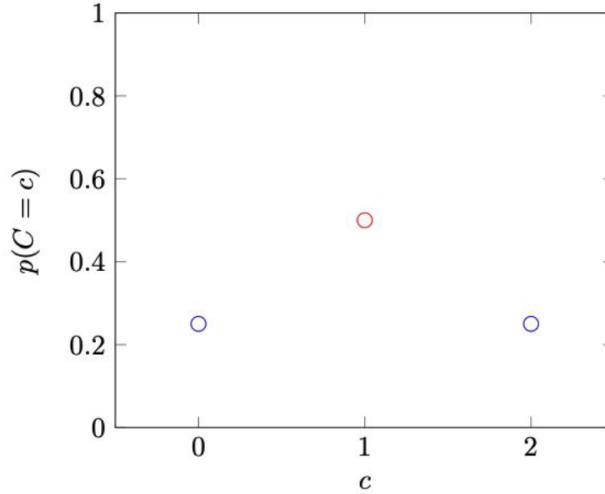
Example

Suppose we toss a coin twice, two coin = $\{HH, HT, TH, TT\}$. Define a random variable $C(\omega)$ as the number of heads in outcome ω . 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

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

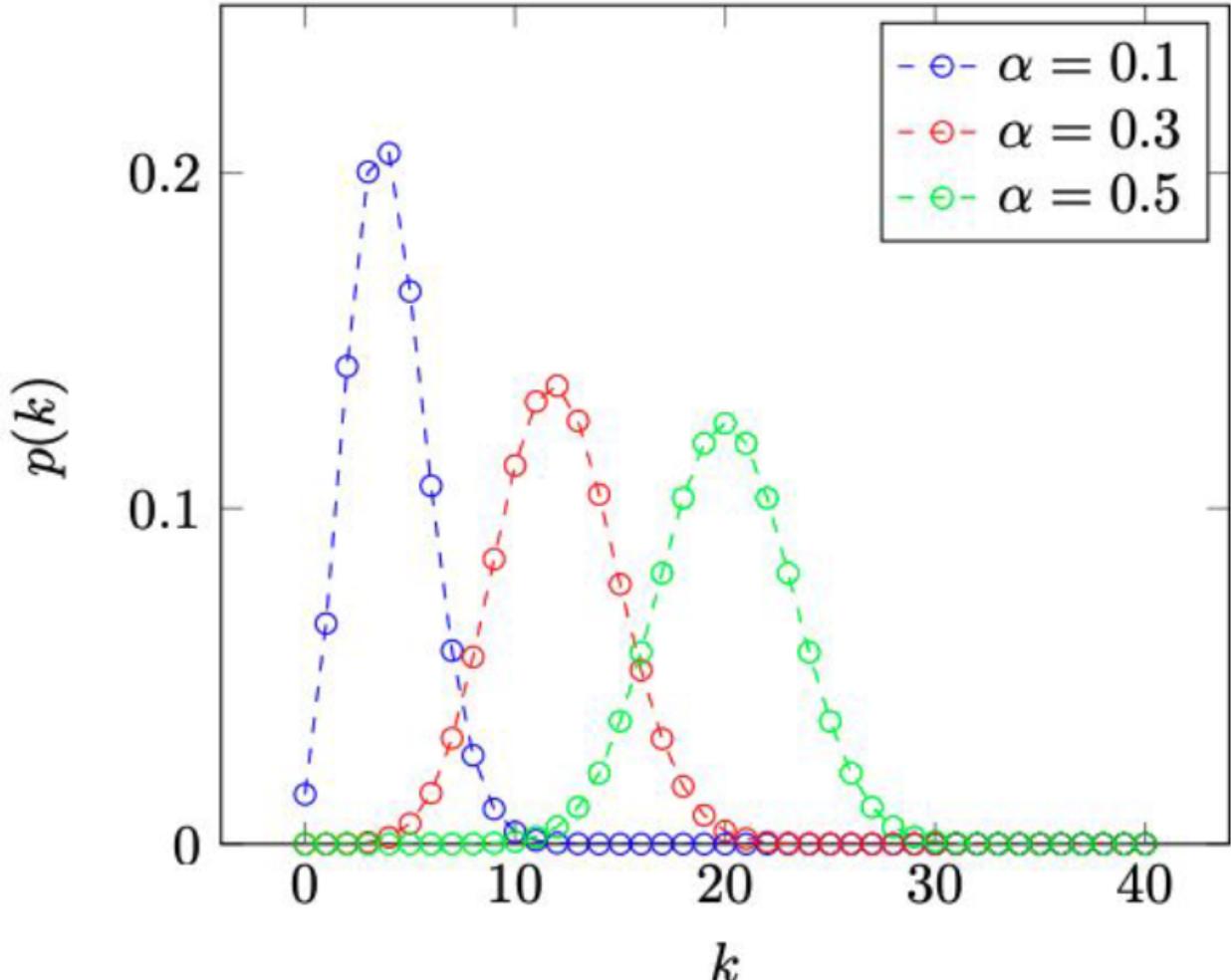
0.1 Well-Known PMFs: Binomial Distribution

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 α of landing heads. The total number of heads k follows a Binomial distribution.



The plot shows how the distribution shifts as α 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 $n/2$.

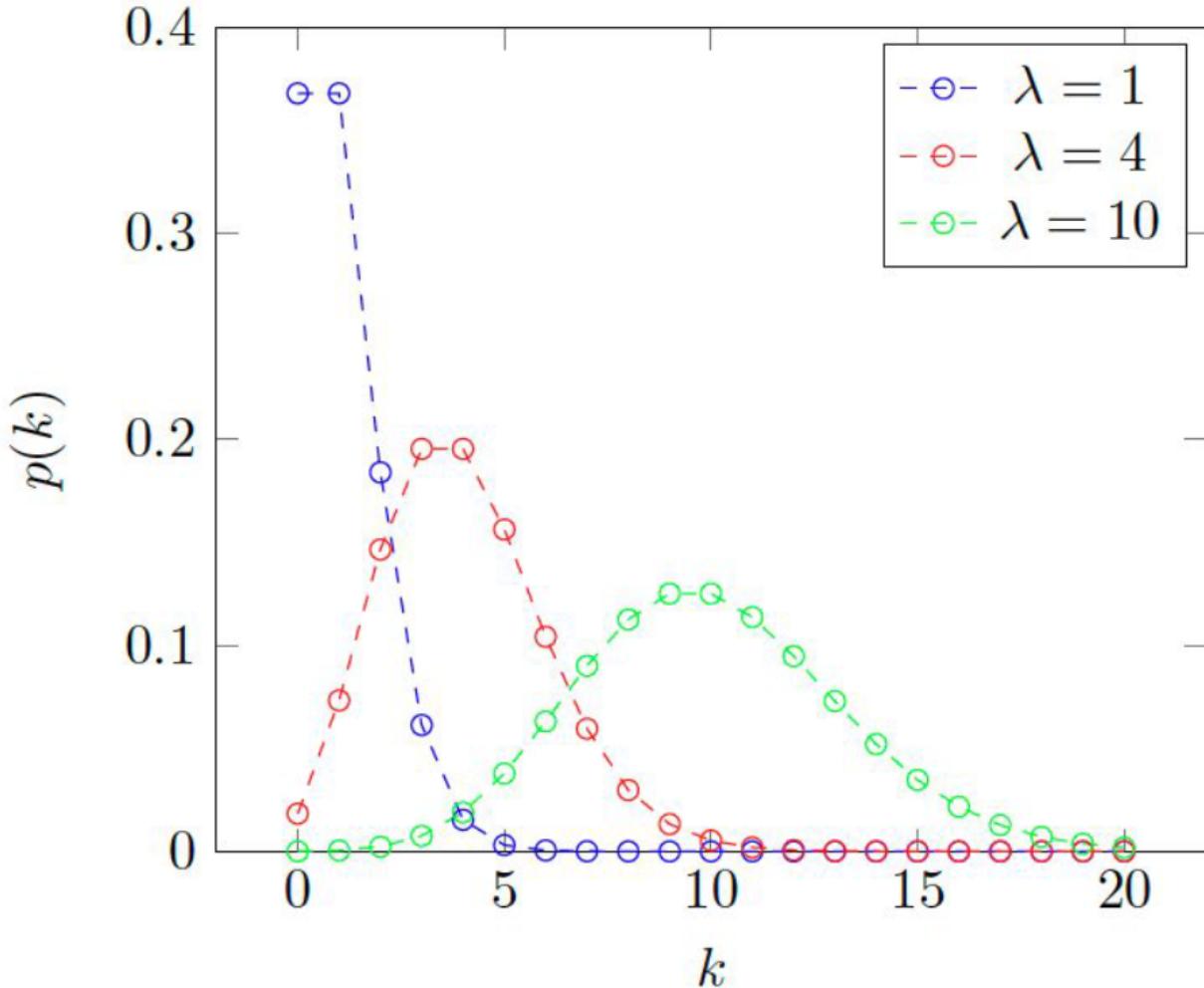
Well-Known PMFs: Poisson Distribution

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 λ . 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
- λ is the expected number of events in the interval

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

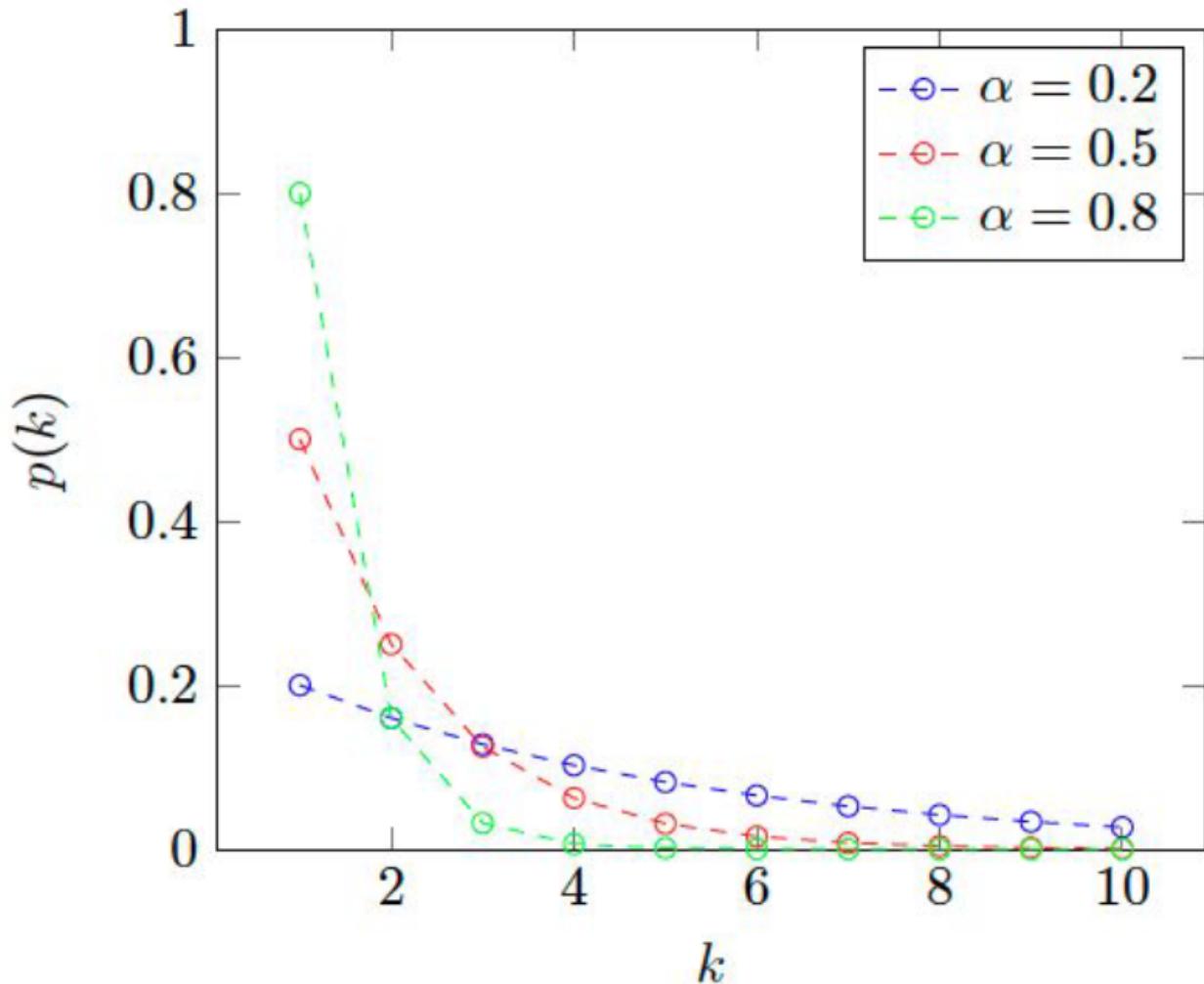
- Parameter: $\alpha \in (0, 1)$, the probability of success in each trial.
- Support: $k = 1, 2, 3, \dots$ - the trial on which the first success occurs.
- Probability mass function (PMF)

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

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 α) 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 α :

- Larger α (e.g., 0.8): success is more likely early →sharp drop.
- Smaller α (e.g., 0.2): more trials are expected →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 λ_i is the expected expression level.

For example:

$$X_i = \text{observed count of gene } G_i$$

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

Continuous Random Variables - Probability Density Functions

In the case of **continuous random variables** (e.g., height, weight, rainfall, etc.) the probability distribution is referred to as a Probability Density Function (PDF). We still have the same properties as before, everything sums to 1, all the probabilities are non-negative, etc. but we typically consider a range rather than a single value to test.

In the case of PDFs, the probability is the area under the curve along a range (the total area is going to end up being one), but when we're talking about area under the curve, we're really considering the integral:

$$p(a \leq x \leq b) = \int_a^b f(x)dx$$

Notice that $f(x)$ is not a probability in and of itself, but the probability comes from the integral. That's why the probability of a single value ($p(X = x)$) is (usually) zero.

Well-Known PDFs: Uniform Distribution

X has a $\text{Uniform}(a, b)$ distribution, written $X \sim \text{Uniform}(a, b)$, if

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

where $a < b$.

The cumulative distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

This is basically a rectangle on a graph; all outcomes with a value between a and $b > a$ are equally probable.

Well-Known PDFs: Gaussian Distribution

X has a Normal (or Gaussian) distribution with parameters μ and σ , denoted by $X \sim N(\mu, \sigma^2)$ if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x \in \mathbb{R}$$

Is parameterized by the mean (center) value of the samples $\mu \in \mathbb{R}$ and the standard deviation (spread) of the samples $\sigma > 0$.

The Normal plays an important role in probability and statistics. Many phenomena in nature have approximately Normal distributions. Later, we shall study the Central Limit Theorem, which says that the distribution of a sum of random variables can be approximated by a Normal distribution.

We say that X has a standard Normal distribution if $\mu = 0, \sigma = 1$.

Well-Known PDFs: Exponential Distribution

Parameterized by a value of $\lambda > 0$, which is the average rate of arrivals over time.

$$f(k) = \lambda e^{-k\lambda}$$

The exponential distribution is used to model the lifetimes of electronic components and the waiting times between rare events.

Cumulative Distribution Functions

PMF vs. PDF

For review, if Ω is a discrete sample space, consider a singleton event $\omega \in A$, where $\omega \in \Omega$.

$$P(\{\omega\}) = p(\omega)$$

If Ω is a continuous sample space, consider an interval event $A = [x, x + \Delta x]$, where Δx is small.

$$P(A) = \int_x^{x+\Delta x} p(\omega) d\omega \approx p(x)\Delta x$$

Discrete CDF

A cumulative distribution function (CDF) is the probability that the random variable is lower than a given threshold. For discrete, this is somewhat intuitive:

$$p(X \leq x) = \sum_{y \in \{\omega \in \Omega | \omega \leq x\}} p(X = y)$$

In the example above, the probability

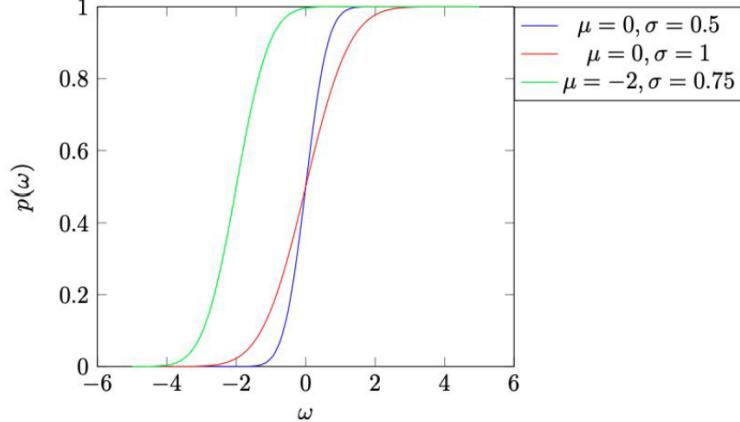
$$p(C \leq c) = \sum_{d \in \{x \in \mathbb{Z} | x \geq 2 \wedge x \leq c\}} p(C = d)$$

So, for example, if we want to know $p(C \leq 4)$, then that can be written as:

$$p(C \leq 4) = \sum_{d \in \{2, 3, 4\}} p(C = d) = p(C = 2) + p(C = 3) + p(C = 4)$$

Continuous CDF

For continuous random variables, it looks similar to the property we had earlier. So for instance, if we look at the gaussian, we get



Summary

PMF	PDF	CDF
$p(x) = P(X = x)$	$P(a \leq X \leq b) = \int_a^b f(x) dx$	$F(x) = P(X \leq x)$
Direct probability	Density (area under curve gives probability)	Accumulated probability up to x
$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x) dx = 1$	$F(x) \in [0, 1]$
Discrete variables	Continuous variables	Both