Topic 14: Random Variables

02-680: Essentials of Mathematics and Statistics

November 6, 2024

The formal definition of a random variable X, is as a mapping:

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

which assigns a real number $X(\omega)$ to an outcome in our space (note this is not a probability in and of itself).

As an example, let say we toss a coin twice, we know we get $\Omega_{twocoin}$ as defined in previous lectures. We can assign a variable $C(\omega)$ to be the count of the number of heads in each outcome.

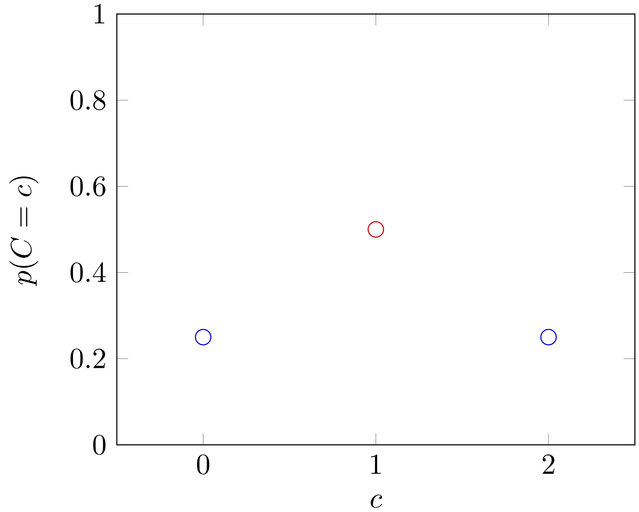
We can then example probabilities of the possible values a random variable X can take on (its range), we usually use a lower case of the came name, x. We can think of this as the sum of the probabilities of all the outcomes that lead to a particular value.

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

c	$p(C = c)$
0	1/4
1	1/2
2	1/4

1 Discrete random variables — Probability Mass Functions

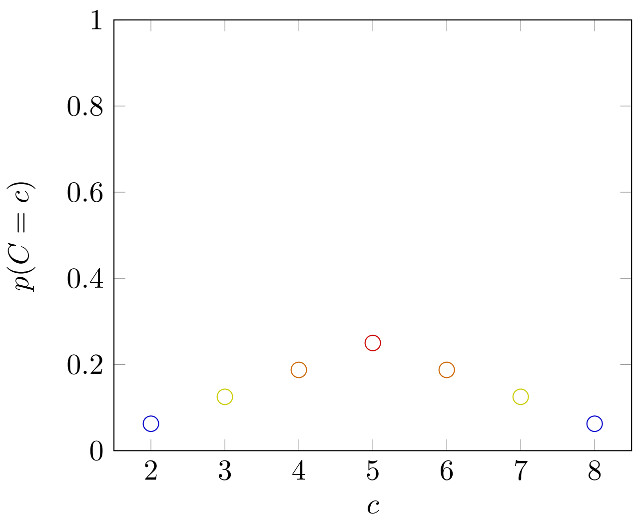
In this case we call the probability distribution a ***probability mass function*** (PMF). Below is a visualization of the PMF for C above.



As another example lets again examine rolling two dice (for simplicity, 4 sided), and let $V(\omega)$ be the value of the sum of the two rolled dice.

ω	$p(\omega)$	$C(\omega)$
11	1/16	2
12	1/16	3
13	1/16	4
14	1/16	5
21	1/16	3
22	1/16	4
23	1/16	5
24	1/16	6
31	1/16	4
32	1/16	5
33	1/16	6
34	1/16	7
41	1/16	5
42	1/16	6
43	1/16	7
44	1/16	8

c	$p(C=c)$
2	1/16
3	1/8
4	3/16
5	1/4
6	3/16
7	1/8
8	1/16



1.1 Bernoulli

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 anther way

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

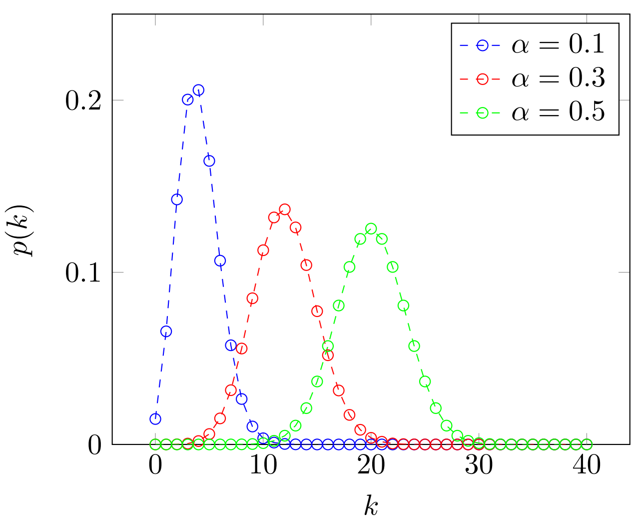
Example Scenario. A single (possibly biased) coin flip with probability of heads α.

1.2 Binomial.

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 $\Omega = \{0, 1, 2, ...n\}$). In this case

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

Example Scenario. n repeated coin flips, each one bring independent and having probability of heads α.



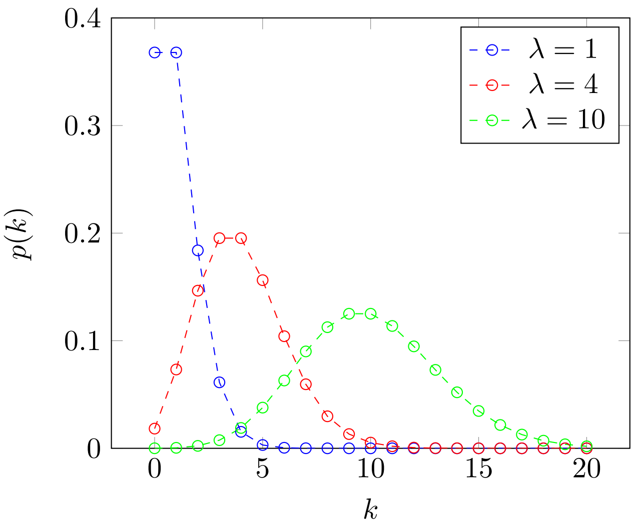
1.3 Poisson.

Poisson distributions are used to model rare events over time.

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

Example 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.*

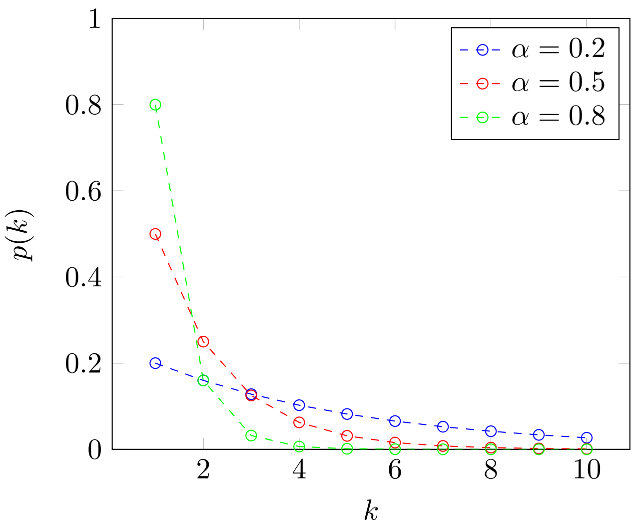


1.4 Geometric.

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

Example Scenario. Number of coin flips (probability of heads α) needed to get k heads.

*see **Ladislaus von Bortkiewicz** "The Law of Small Numbers", 1898. pp. 23-25



2 Continuous Random Variables – Probability Density Functions

In the case of continuous random variable (think grades, weight, height, rainfall [in.], 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:

$$\int_{-\infty}^{\infty} p(X = x)dx = 1$$

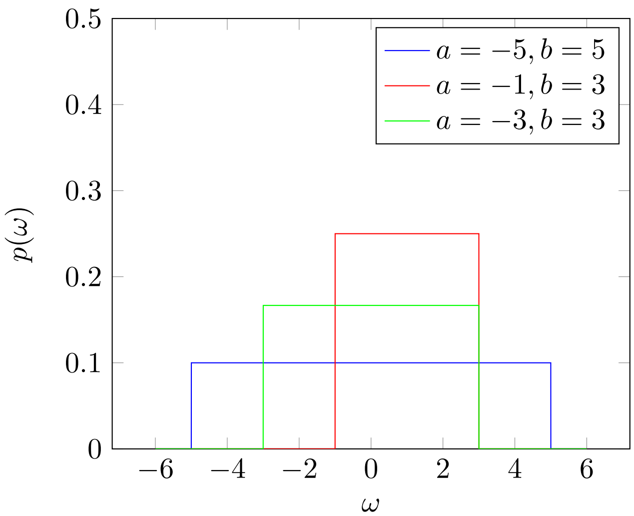
therefore to test something like $p(a \leq X \leq b)$ we need to find

$$p(a \leq X \leq b) = \int_{a}^{b} p(X = x)dx$$

2.1 Uniform.

All outcomes with value between a and $b > a$ are equally probable:

$$p(k) = \begin{cases} \frac{1}{b-a} & a \leq k \leq b \\ 0 & \text{othwerwise} \end{cases}$$

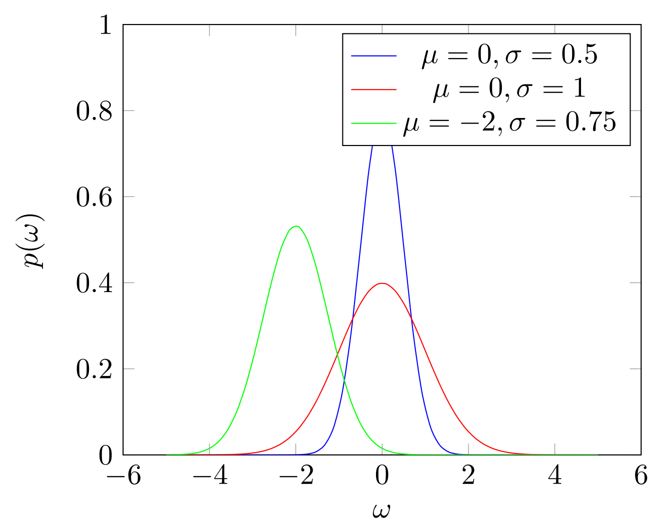


2.2 Gaussian

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

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \;\; e^{\frac{(x-\mu)^2}{-2\sigma^2}}$$

Example Scenario. Most natural phenomena (average daily temperature, height of adults, etc.).

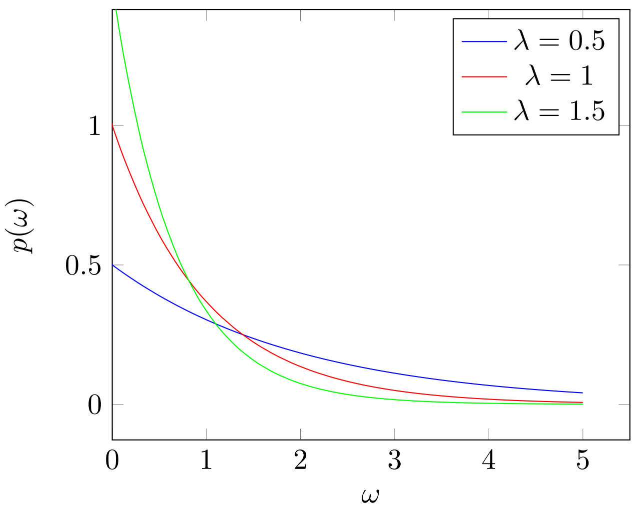


2.3 Exponential.

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

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

Example Scenario. The ***waiting time*** between rare events (notice this can also thought of as the time between Poisson events, Sec. 1.3).



3 Cumulative Distribution Functions

Sometimes we need to know the probability over a range, where the lower bound is $-\infty$ (or upper/∞). Something like:

$$p(X \leq x).$$

3.1 Discrete

For discrete this is somewhat intuitive:

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

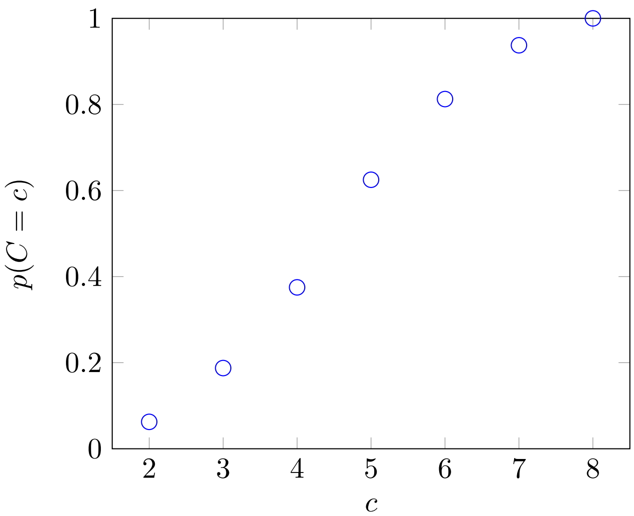
In the example above, the probability

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

so if we want to know

$$p(C \leq 4) = \sum_{d \in \{2,3,4\}} p(C = d) = p(C = 2) + p(C = 3) + p(C = 4) = \frac{1}{16} + \frac{1}{8} + \frac{3}{16} = \frac{3}{8}$$

This in and of itself can also be visualized:



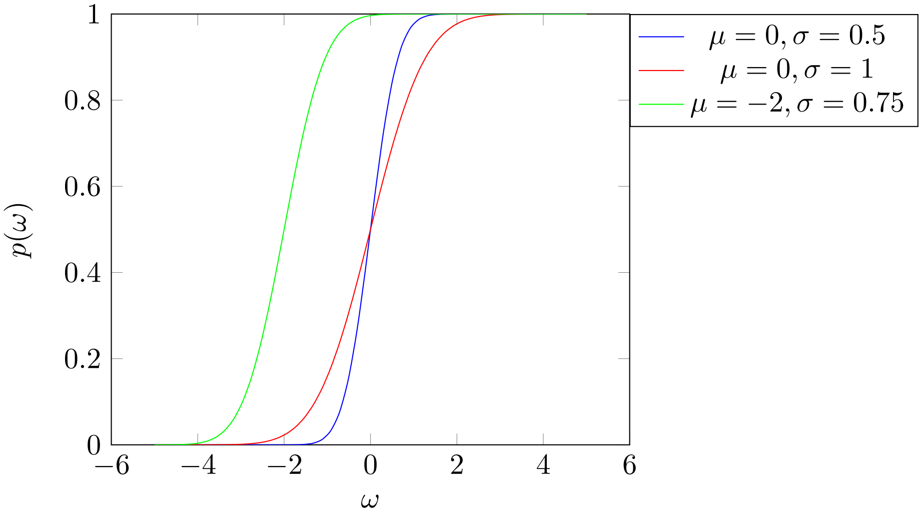
3.2 Continuous

For continuous random variables, it looks similar to the property we had earlier:

$$p(X \leq x) = \int_{-\infty}^{x} p(X = x) dx$$

So for instance, if we look at the exponential distribution:

$$(pX \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \; e^{\frac{(x-\mu)^2}{-2\sigma^2}} dx$$



Useful References

Wasserman. "All of Statistics: A Concise Course in Statistical Inference" §§2.1-2.4