# Probability Distributions

February 4, 2021

Until now, we have studied only *deterministic* models, in which future states are entirely specified by the current state of the system.

In the real world, however, chance plays a major role in the dynamics of a population. Lightning may strike an individual. A fire may decimate a population. Individuals may fail to reproduce or produce a bonanza crop of offspring. New beneficial mutations may, by happenstance, occur in individuals that leave no children. Drought and famine may occur, or rains and excess.

Probabilistic models include chance events and outcomes and can lead to results that differ from purely deterministic models.

In this section, we will introduce the concept of a random variable and examine its probability distribution, expectation, and variance.

Consider an event or series of events with a variety of possible outcomes. We define a **random variable** that represents a particular outcome of these events.

Random variables are generally denoted by capital letters, e.g., $X$, $Y$, $Z$.

For instance, toss a coin twice and record the number of heads. The number of heads is a random variable and it may take on three values: $X = 0$ (no heads), $X = 1$ (one head), and $X = 2$ (two heads).

First, we will consider random variables, as in this example, which have only a discrete set of possible outcomes (e.g., 0, 1, 2).

In many cases of interest, one can specify the probability distribution that a random variable will follow.

That is, one can give the probability that $X$ takes on a particular value $x$, written as $P(X = x)$, for all possible values of $x$.

E.g., for a fair coin tossed twice, $P(X = 0) = 1/4$, $P(X = 1) = 1/2$, $P(X = 2) = 1/4$.

**Note:** Because $X$ must take on some value, the sum of $P(X = x)$*over* all possible values of $x$ must equal one:

$$\sum_{\text{all } x} P(X = x) = 1$$

Discrete Probability Distributions

The expectation or mean of a random variable $X$, written as $E(X)$ or is equal to the average value of $x$ weighted by the probability that the random variable will equal $x$:

$$E(X) = \sum_{\text{all } x} x \cdot P(X = x)$$

Notice that you can find the expectation of a random variable from its distribution, you don't actually need to perform the experiment.

For example

$$\begin{gathered} E(\text{number of heads in two coin tosses}) \\ = 0 * P(X = 0) + 1 * P(X = 1) + 2 * P(X = 2) \\ = 0 * 1/4 + 1 * 1/2 + 2 * 1/4 = 1 \end{gathered}$$

You expect to see one head on average in two coin tosses.

Some useful facts about expectations:

- $E(cX) = cE(X)$ if $c$ is a constant
- $E(X + Y) = E(X) + E(Y)$ "the expectation of a sum equals the sum of the expectations"
- $E(XY) = E(X)E(Y)$ **only** if $X$ and $Y$ are independent random variables.
- $E[g(X)] = \sum_{\text{all } x} g(x)P(X = x)$

Often, we want to know how dispersed the random variable is around its mean One measure of dispersion is the variance of $X$,

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

Another common measure of dispersion is the standard deviation, $SD(X)$, which equals the square root of $Var(X)$.

But how do we calculate the variance?

Notice that $E[(X - \mu)^2] = E[X^2 - 2X\mu + \mu^2]$. We simplify this using the above rules.

First, because the expectation of a sum equals the sum of expectations: $Var(X) = E[X^2] - E[2X\mu] + E[\mu^2]$.

Then, because constants may be taken out of an expectation:

$$Var(X) = E[X^2] - 2\mu E[X] + \mu^2 E[1] = E[X^2] - 2\mu^2 + \mu^2$$
$$= E[X^2] - \mu^2$$

Finally, notice that $E[X^2]$ can be written as $E[g(X)]$ where $g(X) = X^2$. From the final fact about expectations, we can calculate this:

$$E[X^2] = \sum_{\text{all } x} x^2 P(X = x)$$

For example, for the coin toss,

$$E[X^2] = (0)^2 \cdot P(X = 0) + (1)^2 \cdot P(X = 1) + (2)^2 \cdot P(X = 2)$$

$$= 0 \cdot 1/4 + 1 \cdot 1/2 + 4 \cdot 1/4 = 3/2$$

From this we calculate the variance,

$$Var(x) = E[X^2] - \mu^2 = 3/2 - (1)^2 = 1/2$$

Useful facts about variances:

- $Var(cX) = c^2 Var(X)$ if $c$ is a constant.
- $Var(X + Y) = Var(X) + Var(Y)$ **only** if $X$ and $Y$ are independent.

There are several probability distributions that arise again and again. We'll look at a number of these distributions.

It is worth knowing what type of distribution you expect under different circumstances, because if you expect a particular distribution you can determine the mean and variance that you expect to observe.

If a single event or trial has two possible outcomes (say $X_i$ can be 0 or 1 with $P(X_i = 1) = p$), the probability of getting $k$ "ones" in $n$ independent trials is given by the binomial distribution.

If $n = 1$, the probability of getting a zero in the trial is $P(X = 0) = 1 - p$. The probability of getting a one in the trial is $P(X = 1) = p$.

These are the only possibilities $[P(X = 0) + P(X = 1) = 1]$.

If $n = 2$, the probability of getting two zeros is $P(X = 0) = (1 - p)^2$ (getting a zero on the first trial and then independently getting a zero on the second trial), the probability of getting a single one is $P(X = 1) = p(1 - p) + (1 - p)p = 2p(1 - p)$, and the probability of getting two ones is $P(X = 2) = p^2$.

These are the only possibilities $[P(X = 0) + P(X = 1) + P(X = 2) = 1]$.

For general $n$, we use the binomial distribution to determine the probability of $k$ "ones" in $n$ trials:

$$P(X = k) = E[\sum_{i=1}^{n} X_i] = \binom{n}{k} p^k (1-p)^{n-k}$$

where

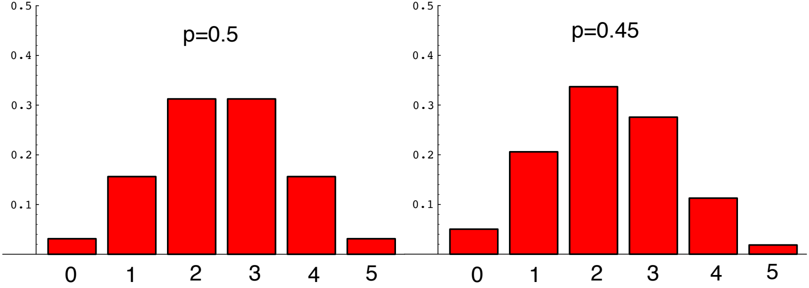
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is called "n choose k".

"n choose k" is the number of ways that you can arrange $k$ ones and $n-k$ zeros in a row. For instance, if you wrote down, in a row, the results of $n$ coin tosses, the number of different ways that you could write down all the outcomes and have exactly $k$ heads is "n choose k".

Example: If the probability of having a daughter is $p$, what is the probability of having three daughters among five children?

We can draw the probability distribution given specific values of $p$:



The **mean** of the binomial distribution equals $E(X) = np$.

Proof: The expectation of the sum of $X_i$ equals the sum of the expected values of $X_i$, but $E[X_i]$ equals $p$ so the sum over $n$ such trials equals $np$.

The variance of the binomial distribution is equal to $Var(X) = np(1-p)$.

Proof: Because each trial is independent, the variance of the sum of $X_i$ equals the sum of the variances of $X_i$. Because $Var(X_i)$ equals

$$Var(X_i) = E[X_i^2] - p^2$$

$$= (0)^2 \cdot P(X = 0) + (1)^2 \cdot P(X = 1) - p^2$$

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

and thus the sum of the variances will equal $np(1-p)$.

If a single event or trial has two possible outcomes (say $X_i$ can be 0 or 1 with $P(X_i = 1) = p$), the probability of having to observe $k$ trials before the first "one" appears is given by the geometic distribution.

The probability that the first "one" would appear on the first trial is $p$.

The probability that the first "one" appears on the second trial is $(1 - p) * p$, because the first trial had to have been a zero followed by a one.

By generalizing this procedure, the probability that there will be $k - 1$ failures before the first success is:

$$P(X = k) = (1 - p)^{k-1} p$$

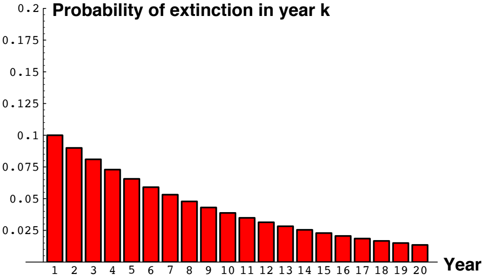
This is the geometric distribution.

A geometric distribution has mean = $1/p$ and variance = $(1-p)/p^2$.

Example: If the probability of extinction of a population is estimated to be 0.1 every year, what is the expected time until extinction?

Notice that the variance in this case is nearly 100! This means that the actual year in which the population will go extinct is very hard to predict accurately.

We can see this from the distribution:



This is an extension of the geometric distribution, describing the waiting time until $r$ "ones" have appeared. The probability of the $r^{th}$ "one" appearing on the $k^{th}$ trial is given by the negative binomial distribution:

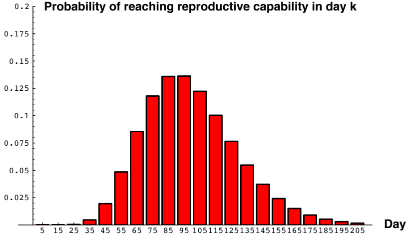
$$P(X = k) = \underbrace{\binom{k-1}{r-1} p^{r-1} (1-p)^{k-r}}_{\substack{\text{probability of } r-1 \text{ "ones" in} \\ \text{the previous } k-1 \text{ trials}}} \cdot p$$

$$= \binom{k-1}{r-1} p^{r} (1-p)^{k-r}$$

A negative binomial distribution has mean = $r/p$ and variance = $r(1-p)/p^2$.

Example: If a predator must capture 10 prey before growing enough to reproduce, what would the mean age of onset of reproduction be if the probability of capturing a prey on any given day is 0.1?

Notice that again the variance in this case is quite high ($\sim 1000$) and that the distribution looks quite skewed (=not symmetric). Some predators will reach reproductive age much sooner and some much later than the average:



The Poisson distribution arises in two important instances.

First, it is an approximation to the binomial distribution when $n$ is large and $p$ is small.

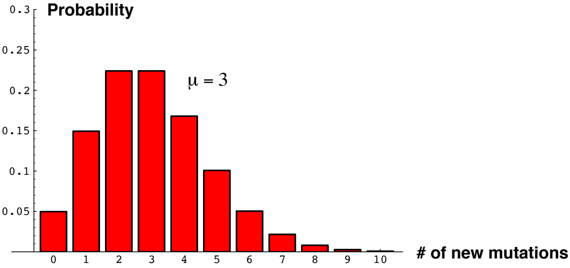
Secondly, the Poisson describes the number of events that will occur in a given time period when the events occur randomly and are independent of one another. Similarly, the Poisson distribution describes the number of events in a given area when the presence or absence of a point is independent of occurrences at other points.

The Poisson distribution looks like:

$$P(X = k) = \frac{e^{-\mu}\mu^k}{k!}$$

A Poisson distribution has the unique property that its variance equals its mean, $\mu$. When the Poisson is used as an approximation to the binomial, the mean and the variance both equal $np$.

Example: If there are $3 \times 10^9$ basepairs in the human genome and the mutation rate per generation per basepair is $10^{-9}$, what is the mean number of new mutations that a child will have? What is the variance in this number? What will the distribution look like?



Example: If hummingbirds arrive at a flower at a rate $\lambda$ per hour, how many visits are expected in $x$ hours of observation and what is the variance in this expectation? If significantly more variance is observed than expected, what might this tell you about hummingbird visits?

Example: If bacteria are spread across a plate at an average density of 5000 per square inch, what is the chance of seeing no bacteria in the viewing field of a microscope if this viewing field is $10^{-4}$ square inches? What, therefore, is the probability of seeing at least one cell?

# Continuous Probability Distributions

The distributions discussed so far have only a discrete set of possible outcomes (e.g., $0, 1, 2, \ldots$). Next, we'll discuss several common continuous distributions, whose outcomes lie along the real line.

One interesting point about continuous probability distributions is that, because an infinite number of points lie on the real line, the probability of observing any particular point is effectively zero.

Continuous distributions are described by *probability density functions*, $f(x)$, which give the probability that an observation falls near a point $x$:

$$P(\text{observation lies within } dx \text{ of } x) = f(x)dx$$

One can, therefore, find the probability that a random variable $X$ will fall between two values by integrating $f(x)$ over the interval:

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

The total integral over the real line must equal one:

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

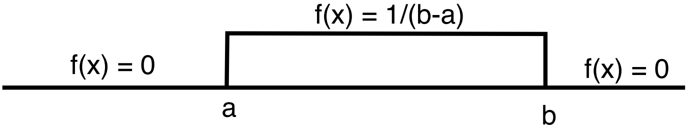
As in the discrete case, $E[X]$ may be found by integrating the product of $x$ and the probability density function over all possible values of $x$:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

The $Var(X)$ equals $E[X^2] - (E[X])^2$, where the expectation of $X^2$ is given by:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

This is the simplest of continuous distributions. The probability density function is $f(x) = 1/(b-a)$ if $x$ lies between $a$ and $b$ and zero otherwise:




The expected value of the uniform distribution equals:

$$E[X] = \int_a^b x \cdot \frac{1}{(b-a)} dx \quad = \quad \frac{1}{(b-a)} \int_a^b x \cdot dx \quad = \quad \frac{a+b}{2}$$

which is the midpoint between $a$ and $b$.

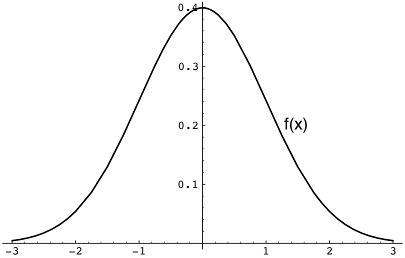
The variance of the uniform distribution equals:

$$Var[X] = \int_a^b x^2 \cdot \frac{1}{(b-a)} dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

This is the most familiar of continuous distributions. The probability density function of the normal distribution is given by:

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

where $\mu$ is the mean and $\sigma^2$ is the variance of the distribution. The mean, for example, may be found by integrating $x \cdot f(x)$ over all values of $x$.




$\mu=0$

$\sigma^2 = 1$

The normal distribution arises repeatedly in biology.

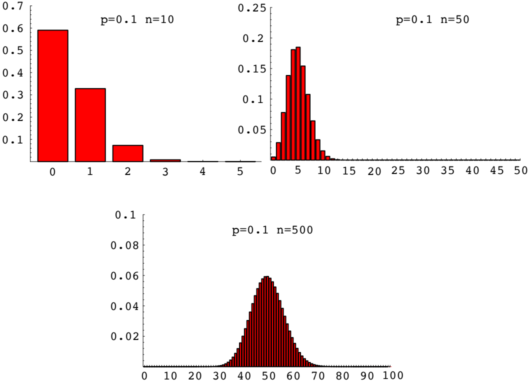
Gauss and Laplace noticed that measurement errors tend to follow a normal distribution.

Quetelet and Galton observed that the normal distribution fits data on the heights and weights of human and animal populations. This holds true for many other characters as well.

Why does the normal distribution play such a ubiquitous role?

**The Central Limit Theorem:** For independent and identically distributed random variables, their sum (or their average) tends towards a normal distribution as the number of events summed (or averaged) goes to infinity.

For example, as $n$ increases in the binomial distribution, the sum of outcomes approaches a normal distribution:




In particular, we expect that if several genes contribute to a trait, trait values should be normally distributed. The random variable being summed or averaged is the contribution of each gene to the trait.

If events occur randomly over time at a rate $\lambda$, then the time until the first event occurs has an exponential distribution:

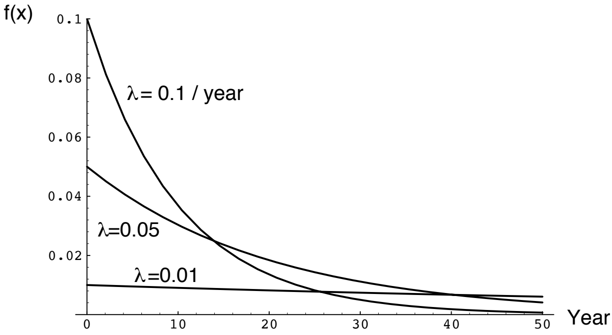
$$f(x) = \lambda e^{-\lambda x}$$

This is the equivalent of the geometric distribution for events that occur continuously over time.

$E[X]$ can be found be integrating $x \cdot f(x)$ from 0 to $\infty$, leading to the result that $E[X] = 1/\lambda$.

The variance of the exponential distribution is $1/\lambda^2$.

For example, let $\lambda$ equal the instantaneous *death rate* of an individual. The lifespan of the individual would be described by an exponential distribution (assuming that $\lambda$ does not change over time).



As the negative binomial generalizes the geometric distribution, the gamma distribution generalizes the exponential distribution. It describes the waiting time until the $r^{th}$ event for a process that occurs randomly over time at a rate $\lambda$:

$$f(x) = \underbrace{\frac{e^{-\lambda x}(\lambda x)^{r-1}}{(r-1)!}}_{\substack{\text{probability of } r-1 \text{ "ones"} \\ \text{in } x \text{ time } (=\text{Poisson})}} \cdot \underbrace{\lambda}_{+1 \text{ more "one"}}$$

The mean of the gamma distribution is $r/\lambda$ and the variance is $r/\lambda^2$.