Probability Density Functions (PDF) for a Random Variable

A **probability density function** or **probability distribution function** has two characteristics:

1. Each probability is between zero and one, inclusive.
2. The sum of the probabilities is one.

Combinational Formula

As we have larger numbers of items in the sample space, such as a full deck of 52 cards, the ability to write out the sample space becomes impossible.

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

This is the formula that tells the number of unique unordered subsets of size x that can be created from n unique elements.The formula is read “n combinatorial x”. Sometimes it is read as “n choose x." The exclamation point "!" is called a factorial and tells us to take all the numbers from 1 through the number before the ! and multiply them together thus 4! is 1*2*3*4=24. By definition 0! = 1. The formula is called the Combinatorial Formula. It is also called the Binomial Coefficient, for reasons that will be clear shortly.

Let's find the hard way the total number of combinations of the four aces in a deck of cards if we were going to take them two at a time. The sample space would be:

S={Spade,Heart),(Spade, Diamond),(Spade,Club), (Diamond,Club),(Heart,Diamond),(Heart,Club)}

There are 6 combinations; formally, six unique unordered subsets of size 2 that can be created from 4 unique elements. To use the combinatorial formula we would solve the formula as follows:

$$\binom{4}{2} = \frac{4!}{(4-2)!2!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} = 6$$

If we wanted to know the number of unique 5 card poker hands that could be created from a 52 card deck we simply compute:

$$\binom{52}{5}$$

Hypergeometric Distribution

This is the most basic one because it is created by combining our knowledge of probabilities from Venn diagrams, the addition and multiplication rules, and the combinatorial counting formula.

To find the number of ways to get 2 aces from the four in the deck we computed:

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

And if we did not care what else we had in our hand for the other three cards we would compute:

$$\binom{48}{3} = \frac{48!}{3!45!} = 17{,}296$$

Putting this together, we can compute the probability of getting exactly two aces in a 5 card poker hand as:

$$\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} = .0399$$

This solution is really just the probability distribution known as the Hypergeometric. The generalized formula is:

$$h(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}$$

where x = the number we are interested in coming from the group with A objects.

h(x) is the probability of x successes, in n attempts, when A successes (aces in this case) are in a population that contains N elements.

For the hypergeometric to work,

1. the population must be dividable into two and only two independent subsets (aces and non-aces in our example). The random variable X = the number of items from the group of interest.
2. the experiment must have changing probabilities of success with each experiment (the fact that cards are not replaced after the draw in our example makes this true in this case). Another way to say this is that you sample without replacement and therefore each pick is not independent.
3. the random variable must be discrete, rather than continuous.

Binomial Distribution

Is any case where there are only two possible outcomes in any one trial, called successes and failures. It gets its name from the binary number system where all numbers are reduced to either 1's or 0's, which is the basis for computer technology and CD music recordings.

$$b(x) = \binom{n}{x} p^{x} q^{n - x}$$

where b(x) is the probability of X successes in n trials when the probability of a success in ANY ONE TRIAL is p. And of course q=(1-p) and is the probability of a failure in any one trial.
For the binomial formula to work, the probability of a success in any one trial must be the same from trial to trial, or in other words, the outcomes of each trial must be independent. Flipping a coin is a binomial process because the probability of getting a head in one flip does not depend upon what has happened in PREVIOUS flips.

Mean and Standard Deviation

$$\mu = \mathrm{np}$$

$$\sigma = \sqrt{npq}$$

If a small number is to be drawn from a large population, even if there is no replacement, we can still use the binomial even thought this is not a binomial process. If there is no replacement it violates the independence rule of the binomial. Nevertheless, we can use the binomial to approximate a probability that is really a hypergeometric distribution if we are drawing fewer than 10 percent of the population, i.e. n is less than 10 percent of N in the formula for the hypergeometric function. The rationale for this argument is that when drawing a small percentage of the population we do not alter the probability of a success from draw to draw in any meaningful way. Imagine drawing from not one deck of 52 cards but from 6 decks of cards. The probability of say drawing an ace does not change the conditional probability of what happens on a second draw in the same way it would if there were only 4 aces rather than the 24 aces now to draw from. This ability to use one probability distribution to estimate others will become very valuable to us later.
There are three characteristics of a binomial experiment.
1. There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter n denotes the number of trials.
2. The random variable,x , number of successes, is discrete.
3. There are only two possible outcomes, called "success" and "failure," for each trial. $p + q = 1$.
4. The n trials are independent and are repeated using identical conditions. Think of this as drawing WITH replacement. Another way of saying this is that for each individual trial, the probability, p, of a success and probability, q, of a failure remain the same. For example, randomly guessing at a true-false statistics question has only two outcomes. If a success is guessing correctly, then a failure is guessing incorrectly. Suppose Joe always guesses correctly on any statistics true-false question with a probability $p = 0.6$. Then, $q = 0.4$. This means that for every true-false statistics question Joe answers, his probability of success ($p = 0.6$) and his probability of failure ($q = 0.4$) remain the same.
The mean, μ, and variance, σ^2, for the binomial probability distribution are $\mu = np$ and $\sigma^2 = npq$. The standard deviation, σ, is then $\sigma = \sqrt{npq}$.

Geometric Distribution

The experiment continues until either a success or a failure occurs rather than for a set number of trials. There are three main characteristics of a geometric experiment.
1.There are one or more Bernoulli trials with all failures except the last one, which is a success. In other words, you keep repeating what you are doing until the first success. Then you stop. For example, you thro

w a dart at a bullseye until you hit the bullseye. The first time you hit the bullseye is a "success" so you stop throwing the dart. It might take six tries until you hit the bullseye. You can think of the trials as failure, failure, failure, failure, failure, success, STOP.
2. In theory, the number of trials could go on forever.
3. The probability, p, of a success and the probability, q, of a failure is the same for each trial. $p + q = 1$ and $q = 1 - p$.
4. X = the number of independent trials until the first success.

Read this as "X is a random variable with a **geometric distribution**." The parameter is p; p = the probability of a success for each trial.

$$\mathrm{P}(X = x) = (1 - p)^{x - 1}p$$

for $x = 0, 1, 2, 3, ...$
In this case the trial that is a success is not counted as a trial in the formula: x = number of failures. The expected value,mean

$$\mu = \frac{(1 - p)}{p}.$$

The formula for the variance is $\sigma^2 = \left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)$

The standard deviation is $\sigma = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)}$

Poisson Distribution

Another useful probability distribution is the Poisson distribution, or waiting time distribution.
There are two main characteristics of a Poisson experiment.
1. The **Poisson probability distribution** gives the probability of a number of events occurring in a **fixed interval** of time or space if these events happen with a known average rate.
2. The events are independently of the time since the last event. For example, a book editor might be interested in the number of words spelled incorrectly in a particular book. It might be that, on the average, there are five words spelled incorrectly in 100 pages. The interval is the 100 pages and it is assumed that there is no relationship between when misspellings occur.
3. The random variable X = the number of occurrences in the interval of interest.

Notation for the Poisson: P = Poisson Probability Distribution Function

$X \sim P(\mu)$
Read this as "X is a random variable with a Poisson distribution." The parameter is μ (or); μ (or) = the mean for the interval of interest. The mean is the number of occurrences that occur on average during the interval period.
The formula for computing probabilities that are from a Poisson process is:

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

where $P(X)$ is the probability of X successes, μ is the expected number of successes based upon historical data, e is the natural logarithm approximately equal to 2.718, and X is the number of successes per unit, usually per unit of time.
In order to use the **Poisson distribution, certain assumptions must hold**. These are: the probability of a success, μ, is unchanged within the interval, there cannot be simultaneous successes within the interval, and finally, that the probability of a success among intervals is independent, the same assumption of the binomial distribution.
The Poisson is asking for the probability of a number of successes during a period of time while the binomial is asking for the probability of a certain number of successes for a given number of trials.

Standard Deviation = $\sigma = \sqrt{\mu}$

Estimating the Binomial Distribution with the Poisson Distribution

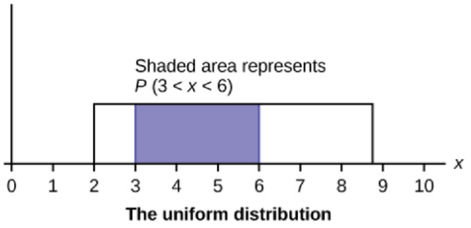
We found before that the binomial distribution provided an approximation for the hypergeometric distribution. Now we find that the Poisson distribution can provide an approximation for the binomial. We say that the binomial distribution approaches the Poisson. The binomial distribution approaches the Poisson distribution is as n gets larger and p is small such that np becomes a constant value. There are several rules of thumb for when one can say they will use a Poisson to estimate a binomial. One suggests that np, the mean of the binomial, should be less than 25. Another author suggests that it should be less than 7. And another, noting that the mean and variance of the Poisson are both the same, suggests that np and npq, the mean and variance of the binomial, should be greater than 5. There is no one broadly accepted rule of thumb for when one can use the Poisson to estimate the binomial.

CONTINUOUS RANDOM VARIABLES

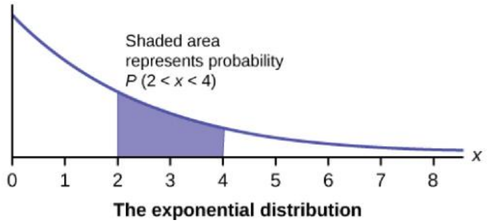
Continuous random variables have many applications. Baseball batting averages, IQ scores, the length of time a long distance telephone call lasts, the amount of money a person carries, the length of time a computer chip lasts, rates of return from an investment, and SAT scores are just a few. The field of reliability depends on a variety of continuous random variables, as do all areas of risk analysis.

- The outcomes are measured, not counted.
- The entire area under the curve and above the x-axis is equal to one.
- Probability is found for intervals of x values rather than for individual x values.
- $P(c < x < d)$ is the probability that the random variable X is in the interval between the values c and d. $P(c < x < d)$ is the area under the curve, above the x-axis, to the right of c and the left of d.
- $P(x = c) = 0$ The probability that x takes on any single individual value is zero. The area below the curve, above the x-axis, and between $x = c$ and $x = c$ has no width, and therefore no area (area = 0). Since the probability is equal to the area, the probability is also zero.
- $P(c < x < d)$ is the same as $P(c = x = d)$ because probability is equal to area.

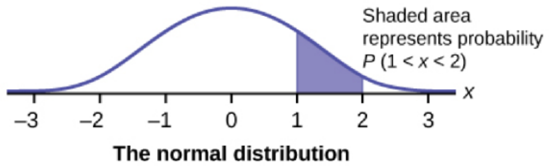
Types of Continuous Probability Distributions



The graph shows a Uniform Distribution with the area between x = 3 and x = 6 shaded to represent the probability that the value of the random variable X is in the interval between three and six.



The graph shows an Exponential Distribution with the area between x = 2 and x = 4 shaded to represent the probability that the value of the random variable X is in the interval between two and four.



The graph shows the Standard Normal Distribution with the area between x = 1 and x = 2 shaded to represent the probability that the value of the random variable X is in the interval between one and two.

For continuous probability distributions, PROBABILITY = AREA.

The Uniform Distribution

The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur. When working out problems that have a uniform distribution, be careful to note if the data is inclusive or exclusive of endpoints.

$$f(x) = \frac{1}{b-a} \text{ for } a \le x \le b$$

where a = the lowest value of x and b = the highest value of x.

$$\mu = \frac{a+b}{2} \text{ and } \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

The Exponential Distribution

The **exponential distribution** is often concerned with the amount of time until some specific event occurs. For example, the amount of time (beginning now) until an earthquake occurs has an exponential distribution. Other examples include the length of time, in minutes, of long distance business telephone calls, and the amount of time, in months, a car battery lasts.
Exponential distributions are commonly used in calculations of product reliability, or the length of time a product lasts.

The random variable for the exponential distribution is continuous and often measures a passage of time, although it canbe used in other applications. Typical questions may be, "what is the probability that some event will occur within the next x hours or days, or what is the probability that some event will occur between $x1$ hours and $x2$ hours, or what is the probability that the event will take more than $x1$ hours to perform?" In short, the random variable X equals (a) the time between events or (b) the passage of time to complete an action, e.g. wait on a customer. The probability density function is given by:

$$f(x) = \frac{1}{\mu} e^{-\frac{1}{\mu}x}$$

where μ is the historical average waiting time. and has a mean and standard deviation of 1/μ.
An alternative form of the exponential distribution formula recognizes what is often called the decay factor. The decay factor simply measures how rapidly the probability of an event declines as the random variable X increases. When the notation using the decay parameter m is used, the probability density function is presented as:

$$f(x) \quad = \quad me^{-mx} \qquad \text{where } m = \frac{1}{\mu}$$

In order to calculate probabilities for specific probability density functions, the cumulative density function is used. The cumulative density function (cdf) is simply the integral of the pdf and is:

$$F\left(x\right) = \int_0^{\infty} \left[\frac{1}{\mu} e^{-\frac{x}{\mu}}\right] = 1 - e^{-\frac{x}{\mu}}$$

Relationship between the Poisson and the Exponential Distribution

There is an interesting relationship between the exponential distribution and the Poisson distribution. Suppose that the time that elapses between two successive events follows the exponential distribution with a mean of μ units of time. Also assume that these times are independent, meaning that the time between events is not affected by the times between previous events.
If these assumptions hold, then the number of events per unit time follows a Poisson distribution with mean μ. Recall that if X has the Poisson distribution with mean μ, then

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

The formula for the exponential distribution: $P(X = x) = me^{-mx} = \frac{1}{\mu}e^{-\frac{1}{\mu}x}$ Where m = the rate parameter, or μ = average time between occurrences.