

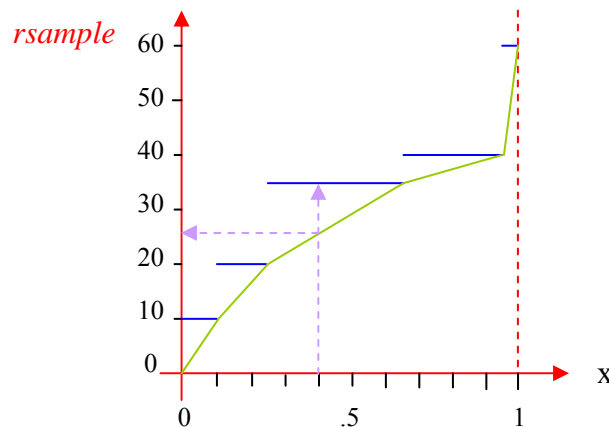
## Empirical Distributions

An **empirical distribution** is one for which each possible event is assigned a probability derived from experimental observation. It is assumed that the events are independent and the sum of the probabilities is 1.

An empirical distribution may represent either a continuous or a discrete distribution. If it represents a discrete distribution, then sampling is done “**on step**”. If it represents a continuous distribution, then sampling is done via “**interpolation**”. The way the table is described usually determines if an empirical distribution is to be handled discretely or continuously; e.g.,

<i>discrete description</i>		<i>continuous description</i>	
value	probability	value	probability
10	.1	$0 - 10^-$	.1
20	.15	$10 - 20^-$	.15
35	.4	$20 - 35^-$	.4
40	.3	$35 - 40^-$	.3
60	.05	$40 - 60^-$	.05

To use **linear interpolation** for continuous sampling, the discrete points on the end of each step need to be connected by line segments. This is represented in the graph below by the green line segments. The steps are represented in blue:



In the discrete case, sampling on step is accomplished by accumulating probabilities from the original table; e.g., for  $x = 0.4$ , accumulate probabilities until the cumulative probability exceeds 0.4; **rsample** is the event value at the point this happens (i.e., the cumulative probability  $0.1+0.15+0.4$  is the first to exceed 0.4, so the **rsample** value is 35).

In the continuous case, the end points of the probability accumulation are needed, in this case  $x=0.25$  and  $x=0.65$  which represent the points  $(.25,20)$  and  $(.65,35)$  on the graph. From basic college algebra, the **slope** of the line segment is  $(35-20)/(.65-.25) = 15/.4 = 37.5$ . Then **slope**  $= 37.5 = (35-\text{rsample})/(.65-.4)$  so **rsample**  $= 35 - (37.5 \times .25) = 35 - 9.375 = 25.625$ .

## Discrete Distributions

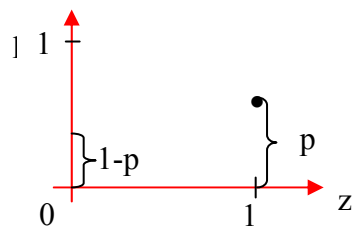
To put a little historical perspective behind the names used with these distributions, [James Bernoulli](#) (1654-1705) was a Swiss mathematician whose book *Ars Conjectandi* (published posthumously in 1713) was the first significant book on probability; it gathered together the ideas on counting, and among other things provided a proof of the binomial theorem. [Siméon-Denis Poisson](#) (1781-1840) was a professor of mathematics at the Faculté des Sciences whose 1837 text *Recherchés sur la probabilité des jugements en matière criminelle et en matière civile* introduced the discrete distribution now called the Poisson distribution. Keep in mind that scholars such as these evolved their theories with the objective of providing sophisticated abstract models of real-world phenomena (an effort which, among other things, gave birth to the calculus as a major modeling tool).

### I. Bernoulli Distribution

A [Bernoulli event](#) is one for which the probability the event occurs is  $p$  and the probability the event does not occur is  $1-p$ ; i.e., the event has two possible outcomes (usually viewed as success or failure) occurring with probability  $p$  and  $1-p$ , respectively. A [Bernoulli trial](#) is an instantiation of a Bernoulli event. So long as the probability of success or failure remains the same from trial to trial (i.e., each trial is independent of the others), a sequence of Bernoulli trials is called a [Bernoulli process](#). Among other conclusions that could be reached, this means that for  $n$  trials, the probability of  $n$  successes is  $p^n$ .

A [Bernoulli distribution](#) is the pair of probabilities of a Bernoulli event, which is too simple to be interesting. However, it is implicitly used in “yes-no” decision processes where the choice occurs with the same probability from trial to trial (e.g., the customer chooses to go down aisle 1 with probability  $p$ ) and can be cast in the same kind of mathematical notation used to describe more complex distributions:

$$p(z) = \begin{cases} p^z(1-p)^{1-z} & \text{for } z = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$



The [expected value](#) of the distribution is given by

$$E(X) = (1-p) \cdot 0 + p \cdot 1 = p$$

The [standard deviation](#) is given by

$$\sigma = \sqrt{(1-p)(0-p)^2 + p(1-p)^2} = \sqrt{p \cdot (1-p)}$$

While this is notational overkill for such a simple distribution, it's construction in this form will be useful for understanding other distributions.

Sampling from a discrete distribution, requires a function that corresponds to the distribution function of a continuous distribution  $f$  given by

$$F(x) = \int_{-\infty}^x f(z) dz$$

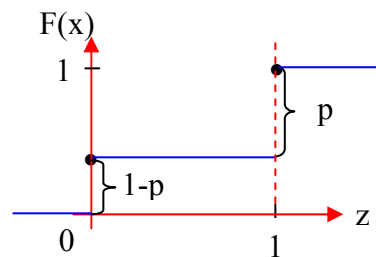
This is given by the mass function  $F(x)$  of the distribution, which is the step function obtained from the cumulative (discrete) distribution given by the sequence of partial sums

$$\sum_0^x p(z)$$

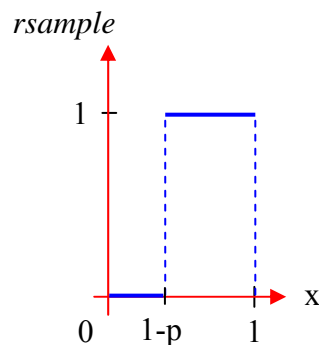
For the Bernoulli distribution,  $F(x)$  has the construction

$$F(x) = \begin{cases} 0 & \text{for } -\infty \leq x < 0 \\ 1-p & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

which is an increasing function (and so can be inverted in the same manner as for continuous distributions). Graphically,  $F(x)$  looks like



which inverted yields the sampling function



In other words, for random value  $x$  drawn from  $[0,1)$ ,

$$rsample = \begin{cases} 0 & \text{if } 0 \leq x < 1-p \\ 1 & \text{if } 1-p \leq x < 1 \end{cases}$$

In essence, this demonstrates that sampling from a discrete distribution, even one as simple as the Bernoulli distribution, can be viewed in the same manner as for continuous distributions.

## II. Binomial Distribution

The Bernoulli distribution represents the success or failure of a single Bernoulli trial. The *Binomial Distribution* represents the number of successes and failures in  $n$  independent Bernoulli trials *for some given value of  $n$* . For example, if a manufactured item is defective with probability  $p$ , then the binomial distribution represents the number of successes and failures in a lot of  $n$  items. In particular, sampling from this distribution gives a count of the number of defective items in a sample lot. Another example is the number of heads obtained in tossing a coin  $n$  times.

The binomial distribution gets its name from the binomial theorem which states that the binomial

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \text{ where } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

It is worth pointing out that if  $a = b = 1$ , this becomes

$$(1 + 1)^n = 2^n = \sum_{k=0}^n \binom{n}{k}$$

Yet another viewpoint is that if  $S$  is a set of size  $n$ , the number of  $k$  element subsets of  $S$  is given by  $\frac{n!}{k!(n-k)!} = \binom{n}{k}$

This formula is the result of a simple counting analysis: there are

$$n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

ordered ways to select  $k$  elements from  $n$  ( $n$  ways to choose the 1<sup>st</sup> item,  $(n-1)$  the 2<sup>nd</sup>, and so on). Any given selection is a permutation of its  $k$  elements, so the underlying subset is counted  $k!$  times. Dividing by  $k!$  eliminates the duplicates.

Note that the expression for  $2^n$  counts the total number of subsets of an  $n$ -element set.

For  $n$  independent Bernoulli trials the pdf of the binomial distribution is given by

$$p(z) = \begin{cases} \binom{n}{z} p^z (1-p)^{n-z} & \text{for } z = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Note that

by the binomial theorem,  $\sum_{z=0}^n p(z) = (p + (1-p))^n = 1$ , verifying that  $p(z)$  is a pdf.

When choosing  $z$  items from among  $n$  items, the term

$$\binom{n}{z} p^z (1-p)^{n-z}$$

represents the probability that  $z$  are defective (and concomitantly that  $(n-z)$  are not defective).

The binomial theorem is also the key for determining the expected value  $E(X)$  for the random variable  $X$  for the distribution.  $E(X)$  is given by

$$E(X) = \sum_1^n p(z_i) \cdot z_i$$

(the expected value is just the sum of the discrete items weighted by their probabilities, which corresponds to a sample's mean value; this is an extension of the simple average value obtained by dividing by  $n$ , which corresponds to a weighted sum with each item having probability  $1/n$ ).

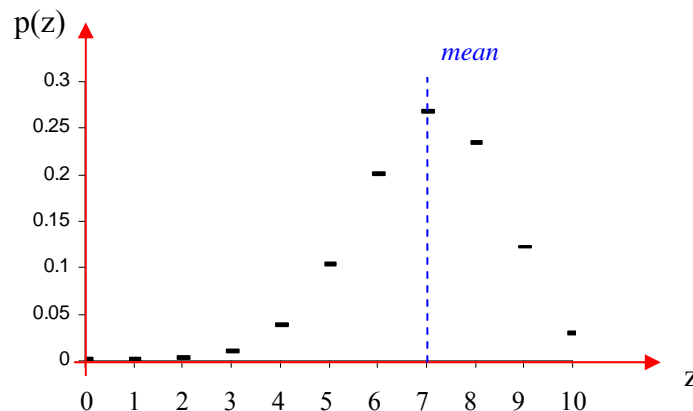
For the binomial distribution the calculation of  $E(X)$  is accomplished by

$$\begin{aligned} E(X) &= \sum_0^n \binom{n}{z} p^z (1-p)^{n-z} \cdot z = \sum_1^n \frac{n!}{z!(n-z)!} p^z (1-p)^{n-z} \cdot z \quad \text{term in common} \\ &= \sum_1^n \binom{n-1}{z-1} p^{z-1} (1-p)^{n-z} \cdot np = np \cdot \sum_1^n \binom{n-1}{z-1} p^{z-1} (1-p)^{n-z} \\ &\quad \text{present in every summand} \quad \text{apply the binomial theorem to this} \\ &\quad \text{(note that } n-z = (n-1) - (z-1) \text{)} \end{aligned}$$

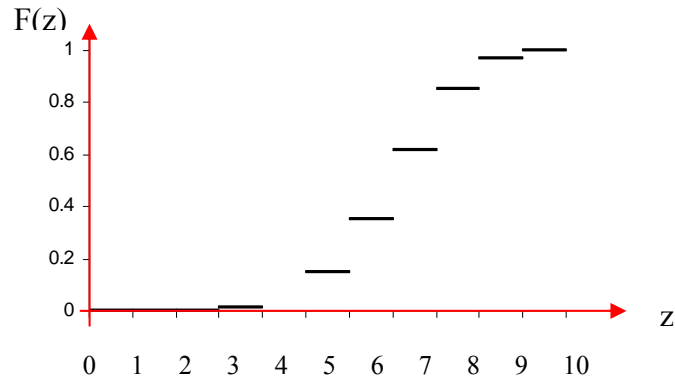
This gives the result that  $E(X) = np$  for a binomial distribution on  $n$  items where probability of success is  $p$ . It can be similarly shown that the standard deviation is

$$\sqrt{np \cdot (1-p)}$$

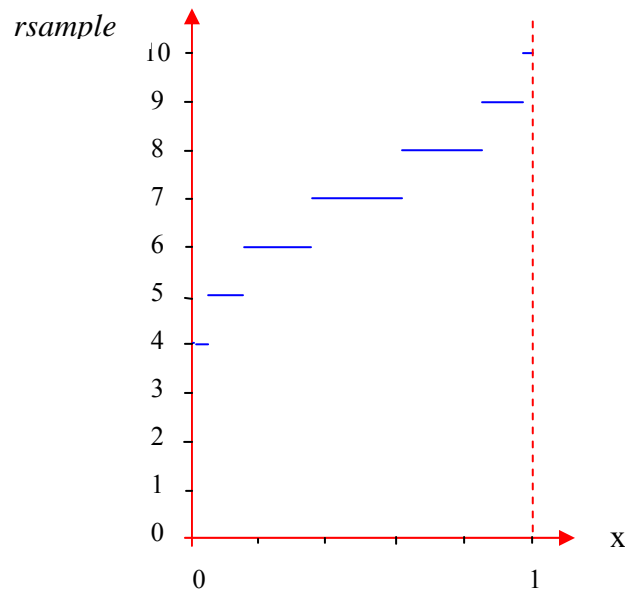
The binomial distribution with  $n=10$  and  $p=0.7$  appears as follows:



Its corresponding mass function  $F(x)$  is given by



which provides the sampling function



A typical sampling tactic is to accumulate the sum

$$\sum_0^{rsample} p(z)$$

increasing *rsample* until the sum's value exceeds the random value between 0 and 1 drawn for  $x$ . The final *rsample* summation limit is the sample value. Note that in contrast to a continuous pdf described by some formula, the function for a finite discrete pdf has to be given in its relational form by a table of pairs, which in turn mandates the kind of "search" algorithm approach used above to obtain *rsample*.

### III. Poisson Distribution (values $z = 0, 1, 2, \dots$ )

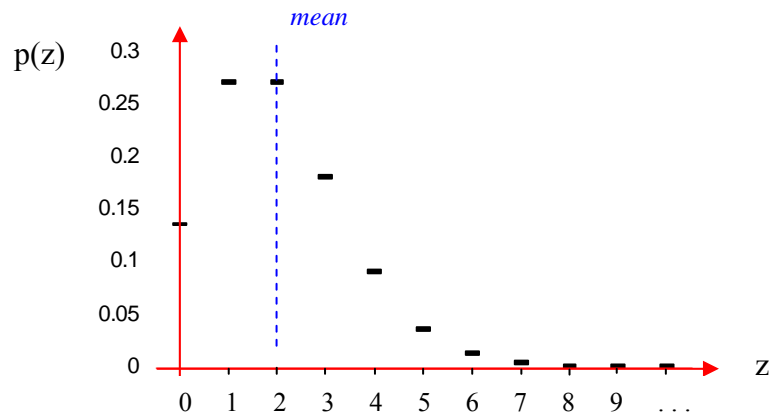
The Poisson distribution is the limiting case of the binomial distribution where  $p = \lambda/n \rightarrow 0$  and  $n \rightarrow \infty$ .

The expected value  $E(X) = \lambda$ . The standard deviation is  $\sqrt{\lambda}$ . The pdf is given by

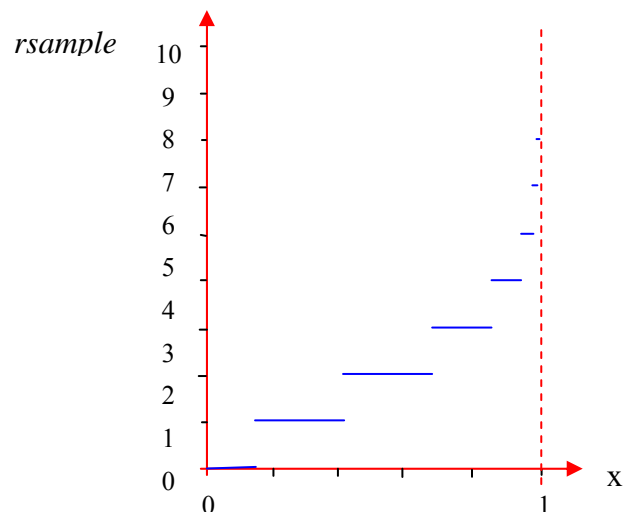
$$p(z) = \frac{\lambda^z \cdot e^{-\lambda}}{z!}$$

This distribution dates back to Poisson's 1837 text regarding civil and criminal matters, in effect scotching the tale that its first use was for modeling deaths from the kicks of horses in the Prussian army. In addition to modeling the number of arrivals over some interval of time (recall the relationship to the exponential distribution; a *Poisson process* has exponentially distributed interarrival times), the distribution has also been used to model the number of defects on a manufactured article. In general the Poisson distribution is used for situations where the probability of an event occurring is very small, but the number of trials is very large (so the event is expected to actually occur a few times).

Graphically, with  $\lambda = 2$ , it appears as:



The sampling function looks like:



## IV. Geometric Distribution

The geometric distribution gets its name from the geometric series:

$$\text{for } r < 1, \sum_0^{\infty} r^n = \frac{1}{1-r}, \sum_0^{\infty} n \cdot r^n = \frac{r}{(1-r)^2}, \sum_0^{\infty} (n+1) \cdot r^n = \frac{1}{(1-r)^2}$$

*various flavors of the geometric series*

The pdf for the geometric distribution is given by

$$p(z) = \begin{cases} (1-p)^{z-1} \cdot p & \text{for } z=1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

The geometric distribution is the discrete analog of the exponential distribution. Like the exponential distribution, it is "memoryless"; i.e.,

$$P(X > a+b \mid X > a) = P(X > b)$$

(the geometric distribution is the only discrete distribution with this property just as the exponential distribution is the only continuous one behaving in this manner).

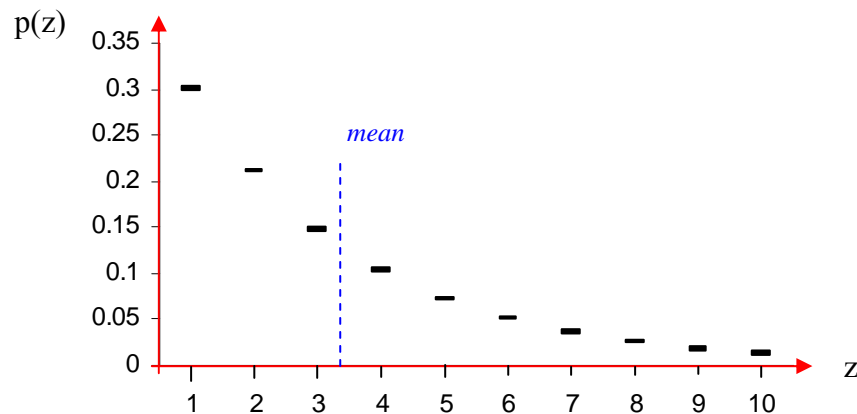
Its **expected value** is given by

$$E(X) = \sum_1^{\infty} z(1-p)^{z-1} \cdot p = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

(by applying the 3<sup>rd</sup> form of the geometric series).

The **standard deviation** is given by  $\frac{\sqrt{1-p}}{p}$ .

A plot of the geometric distribution with  $p = 0.3$  is given by





A typical use of the geometric distribution is for modeling the number of failures before the first success in a sequence of independent Bernoulli trials. This is a typical scenario for sales. Suppose that the probability of making a sale is 0.3. Then

$p(1) = 0.3$  is the probability of success on the 1<sup>st</sup> try,

$p(2) = (1-p)p = 0.7 \times 0.3 = 0.21$

which is the probability of failing on the 1<sup>st</sup> try (with probability  $1-p$ ) and succeeding on the 2<sup>nd</sup> (with probability  $p$ ).

$p(3) = (1-p)(1-p)p = 0.15$  is the probability that the sale takes 3 tries, and so forth.

A random sample from the distribution represents the number of attempts needed to make the sale.