

Some notes on different families of distributions:

© 2002 Michael J. Rosenfeld

1) Poisson:

Formal Definition of Poisson(λ)

Mean= λ

Variance= λ

Std Deviation = $\sqrt{\lambda}$

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

for $k=0,1,2,\dots$

Intuitive explanation:

a) The poisson distribution has one parameter, λ . This parameter describes the mean number of events. Let's say we have a hundred apple trees, all identical. During harvest season, we put a basket under each tree and we catch a few apples a day. Let's say the average is 2 apples. The poisson distribution tells us how many of the baskets will have exactly 2 apples each day, how many will have 0, 1, 2, 3, 4, and so on. As long as each tree is identical, and as long as each apple's fall into the basket is a separate and independent event (i.e. nobody shakes the branches to make a lot of apples fall at once), then the poisson distribution will do a good job of describing the distribution of apples in the baskets. A process that may more easily fit the model is the capture of neutrinos (a tiny subatomic particle) from the sun in a special detector. Every day the sun emits the same number of neutrinos, and each neutrino has a small but constant chance of hitting the detector on your desk (the real detectors are deep under ground, but let's pretend you could have a detector on your desk). If the average count is 2, the poisson distribution tells you how likely you are to catch 10 (or 2, or 1) on any given day. If a single day's neutrinos have a Poisson(2) distribution, it seems reasonable that if you checked the machine every other day, you would have a Poisson(4) distribution. So, as long as the events are independent, Poisson(A)+ Poisson (B)=Poisson(A+B).

b) See my poisson distribution [examples](#).

c) Also notice, in the examples I demonstrate above, that Poisson(15) looks a lot like a Normal distribution, but Poisson(.1) and Poisson(2) do not. Why is this? For one thing, the Poisson distribution only allows nonnegative values (the smallest number of apples or neutrinos you can get is zero), whereas the Normal distribution does not discriminate against our friends the negatives. When the mean is close to zero, the Poisson distributions are more obviously skewed. When the mean is large, say 15, the Poisson(15) distribution looks fairly symmetrical because the mean is large and values as far away as zero are implausible anyway. Another reason why Poisson(15) looks like a Normal distribution with mean 15, is because of the Central Limit Theorem. One simple version of the Central Limit Theorem says:

If variable X has mean μ and variance σ^2 , and $S_n = \sum_{i=1}^n X_i$

where the X_i are independent and identically distributed,

then $\lim_{n \rightarrow \infty}$ of S_n is $N(n\mu, n\sigma^2)$.

What this means is simply that if you take (almost) any old kind of variable X , and you take enough samples of it, and you add them up (or, more commonly, take their average), you end up with something that has a Normal distribution. The Central Limit Theorem in its various guises is of Central importance, hence the name. Think of Poisson(15) as Poisson(1)+Poisson(1), each independent,..15 times. The Central Limit Theorem suggests that Poisson(15) should start to look a lot like Normal (15,15). Of course, one could also note that Poisson(2), which looks very non-Normal is really just 15 separate and independent instances of Poisson(2/15) summed together. So why doesn't Poisson(2) look Normal? Well, the Central Limit Theorem says as n goes to infinity, S_n will be Normally distributed. It doesn't say how fast you'll get there. If you start with Poisson(2/15), you need to add together a lot more than 15 independent copies to get to a Normal distribution.

d) Based on my description above, in part (a), you can see that the poisson distribution is the summary results of many individual events at the apple or neutrino level. You can think of each apple having a fixed chance of falling into a basket, and each neutrino of having a fixed chance of falling into the detector. At the level of the apple or neutrino, we are dealing with a family of distributions called the binomials (binomial because each apple has 2 choices- in the basket or not, and each neutrino has 2 choices- hit the detector or not). To be very specific, the Poisson distribution is a special case of the Negative Binomial distribution, and poisson regression is a special case of negative binomial regression. Whereas the Poisson distribution has only one parameter that fixes both the mean and variance (similar to the Chisquare but different from the Normal), the Negative Binomial distribution has more flexibility- the mean and variance are determined by two separate parameters. Stata has a number of functions for negative binomial regression.

e) It is also worth noting, that if X is distributed as Poisson(λ), then the square root of X has an approximately Normal distribution, with constant variance. I'm not going to justify why this is, but I'll simply point out that when you want things to behave more Normally, you sometimes have to enforce a transformation. What a square root transformation does (see, again, my examples) is it reels in the high value outliers. When you have variables that take on only positive values, and have some high value outliers (income is one example, counts of events is another example), it is common to take the log or the square root of those variables in order to bring the high values down, and make the distribution less skewed.

2) Chisquare Distribution:

with n (integer) degrees of freedom,

$$f(x) = \frac{(1/2)^{n/2} x^{(n/2)-1} e^{-1/2x}}{\Gamma(n/2)}$$

for $x \geq 0$

Mean = n

Variance = 2n

Standard Deviation = $\sqrt{2n}$

a) Unlike the Poisson distribution, which can take on only integer values or zero, the chisquare distribution has a range of all positive real numbers. The greek letter Γ in the denominator is just Gamma, indicating the Gamma function. The Gamma function is nothing more than a fancy way of extending the factorial function to all the real numbers. $\Gamma(x) = (x-1)!$, when x is an integer. So if you think of Gamma as just an extended version of the factorial function, you'll see that the Chisquare distribution and the Poisson distribution have some similarities in the way their probability densities are defined. They also have similar shapes, see below.

b) $\chi^2(1)$, or the Chisquare distribution with one degree of freedom is defined as the square of a Standard Normal variable. In other words, if z has the familiar N(0,1) distribution whose cumulative distribution is the source of tables in the back of every statistics text book (i.e. Normal with mean of zero and variance of 1), and if $y = z^2$, then y has a $\chi^2(1)$ distribution. This also means that if you have a statistic expressed a value from a $\chi^2(1)$ distribution, you can take the square root and you will have the familiar z-score. When switching back and forth from $\chi^2(1)$ and N(0,1) you do have to keep in mind that the Normal distribution has two tails (in the positive and negative directions), whereas as the Chisquare distribution only has the one tail.

c) Under independence, $\chi^2(a) + \chi^2(b) = \chi^2(a+b)$. Another way to look at this is that $\chi^2(a) = \chi^2(1) + \chi^2(1) + \chi^2(1) + \dots$ a times (with each component being independent). Given what we know about the Central Limit Theorem, you would expect that $\chi^2(n)$ would look more and more like the Normal distribution, the larger n gets (since $\chi^2(n)$ is just n combinations of independent $\chi^2(1)$ variables). The [examples](#) of the chisquare distribution will verify that that $\chi^2(16)$ looks quite Normal (and in this case it approximates N(16,32)). Also note that, comparing the $\chi^2(n)$ and Poisson(n) distributions when n is the same or similar, shows that the distributions have some similarity in shape, which is not

surprising since their probability density functions have some similar elements, and since the Central Limit Theorem forces both distributions to be more Normal when n grows large. We do know that $\chi^2(n)$ has a variance of $2n$, and Poisson(n) has a variance of n , so clearly the Chisquare distribution has longer tails.

d) One property of the Chisquare distribution we've used throughout the class is that if Model 1 has goodness of fit chisquare $\chi^2(n)=V$, and Model 2 adds m additional terms and has a goodness of fit Chisquare of $\chi^2(n-m)=U$, then the comparison of Model 1 and Model 2 is $\chi^2(m)=V-U$. We have also said that this comparison only works if Model 1 is nested within Model 2 (that is, if Model 2 contains Model 1).