Notes on the Pearson and other Chi-squared tests of multinomial parameters

August 30, 2011

draft: August 30, 2011

Karl (Carl) Pearson, 1857-1936. Aka "KP".

Besides being a great statistican, KP was into a lot of things, including eugenics: improving the quality of human stock by encouraging reproduction between those of good stock (those from superior races with no defects) and discouraging reproduction by those from poor stock. Karl's friend Francis Galton is considered the father of eugenics; Hitler its most infamous proponent. Presidents Teddy Roosevelt and Woodrow Wilson supported eugenics, so did John Maynard Keynes and Winston Churchill. Eugenics is now politically incorrect.

1 Chi-squared tests of multinomial parameters

1.1 The Pearson Chi-squared tests–the KP test

The Peason Chi-squared statistic is used to test the null hypothesis that a single random variable X with a multinomial distribution has certain parameter values.[1]

Without loss of generality assume X can take C values such that the probability of X taking value c is p_c. Assume N independent draws from this multinomial where x_c is the number of draws that take the value c. The data is easily represented by a table where each column is a different outcome:

$$\begin{bmatrix} \text{outcome 1} & \text{outcome 2} & ... & \text{outcome } C \\ x_1 & x_2 & ... & x_C \end{bmatrix}$$

The multinomial density function is

$$\begin{aligned} & f(x_1, x_2, ... x_C \,| N, p_1, p_2, ... p_c) \\ = & \frac{N!}{\prod_{c=1}^{C} x_c!} \prod_{c=1}^{C} p_c^{x_c} \end{aligned}$$

where $0 < p_c < 1$ and $\sum_{c=1}^{C} p_c = 1$. Let $p_1^0, p_2^0, ...p_C^0$ be the null hypothesis. One takes a sample s with N draws from a multinomial and obtains the observations $x_1^s, x_2^s, ...x_C^s$.

One wants a statistical test of a null hypothesis. Karl Pearson, KP, developed such a test, supposedly to test whether a particular roulette table was fair: the result of each spin of roulette table is an independent draw from a multinomial with parameters $p_1^0, p_2^0, ...p_C^0$, (Stigler 1986). If the wheel is fair, all outcomes are equally likely and the null hypothesis is $p_c = \frac{1}{C+1}$.

In 1900, KP proposed the following statistic (Pearson (1900))

$$KP = \sum_{c=1}^{C} \frac{(x_c - p_c^0 N)^2}{p_c^0 N}$$

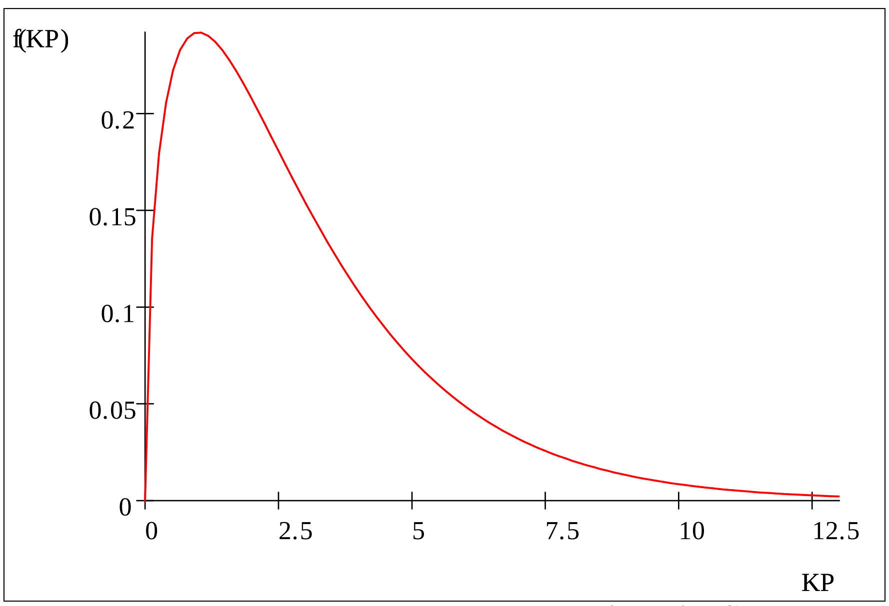
[1] Note it is "Chi-squared statistic" or "χ^2 statistic" not the "Chi-square statistic."

each term in the sum is the difference between the observed and predicted value of x_c, squared, divided by the predicted value of x_c.

Intuitively, KP will be small when the null hypothesis is correct. Alternatively, when KP is sufficiently large, one should reject the null hypothesis that the observed sample came from a multinomial with parameters $p_1^0, p_2^0, ...p_C^0$.

To determine the critical value of KP, one must determine the density function of KP, $f_{KP}(kp)$. For large samples, the random variable KP has, approximately, a Chi-squared distribution with parameter (degrees of freedom) $C - 1$.[2] So, to do the Pearson Chi-squared test one calculates kp for one's sample and rejects the null hypothesis if kp is greater than the chosen critical value of χ^2_{C-1}, for the chosen level of significance.

For example, if $C = 4$ and the chosen significance level is .05 (there is only a 5% probablility that $kp > \chi^2_{4-1}$), $\chi^2_{3(.05)} = \text{ChiSquareInv}(0.95; 3) = 7.8147$



When N is large, the approx distribution of KP for $C = 4$

One can use, for example, the Pearson Chi-squared test to test Mendel's theories of genetics. According to Mendel, if a redhead has a kid with a blond, maybe (I don't know) the probability of blond is $p_b = .25$, the probability of red is $p_r = .25$ and the probability of other is $p_o = .5$. Assume they have a hundred kids (one hundred independent draws from a multinomial with these

[2] One might ask why. It is explained below, in "aside 4" for the simple case of $C = 2$

probabilities). One observes the hair colors of the hundred kids and uses the Pearson Chi-squared test to see if one can reject the the Mendel probabilities.

In 1936, R.A Fisher applied Pearson's Chi-squared test to Mendel's data on the offsprings from crossing green and yellow peas; as expected, Fisher failed to reject Mendel's theory. But, the data fit the theory too well: the fit was better than one would expect if one took thousands of samples from a multinomial with Mendel's probabilities. Go figure. Politely, the conclusion was the gardener fudged the data, knowing what Mendel wanted to see, and Mendel "never knew." For details see Agresti (2002)

1.2 Asides:

1. This Pearson Chi-squared test is sometimes referred to as the "Pearson Chi-squared test of goodness of fit." I presume it is given this name because it tests how well the observed data fits the assumed multinomial parameters.

2. There are at least three types of of Chi-squared statistics used to test hypotheses about multinomial parameters, all based on properties of the multinomial ln likelihood function: *Wald statistics*, *Likelihood ratio statistics* (discussed next) and *score statistics*. The Peason Chi-squared statistic is a a *score* statistic, specifically a score test of multinomial parameters.[3]

3. Intuitively, we know the slope of the ln likelihood function is zero in each direction when evaluated at the m.l. estimates of the multinomial probabilities. So, if you are not at the the max, the slopes are not zero. The KP statistic is a measure of how the average of the slopes of the ln likelihood deviate from zero at the null values of the multinomial probabilities. The larger the statistic, the more the slopes of the likelihood function are deviating from zero at the null, and the more likely it is that the null is incorrect. (See the previous footnote.)

4. If $C = 2$, the multinomial simplifies to a binomial. In which case $KP = \sum_{c=1}^{C} \frac{(x_c - p_c^0 N)^2}{p_c^0 N}$ simplifies to $KP = \frac{(x_1 - p^0 N)^2}{p^0 N} + \frac{((N - x_1) - (1 - p^0) N)^2}{(1 - p^0) N} = \frac{(1 - p^0)(x_1 - p^0 N)^2}{(1 - p^0) p^0 N} +$

---

[3] The score of p_c is $\frac{\partial \ln L}{\partial p_c}$ where $\ln L = \sum_{c=1}^{C} x_c \ln p_c$ is the ln of the multinomial likelihood function. The score of p_c is not, itself a statistic because it is not a function of only the data, the x: the score is also a function of the unknown p parameters. This score becomes a score statistic when the unknown p are replaced by the p^0, making the result a function of only the data. It is possible to show that $KP = \sum_{c=1}^{C} \frac{\left(\frac{\partial \ln L(p^0)}{\partial p_c} \right)^2}{-E\left[\frac{\partial^2 \ln L(p^0)}{\partial p_c^2} \right]}$, so a function of the score.

$\frac{p^0((N-x_1)-(1-p^0)N)^2}{(1-p^0)p^0N} = \frac{(1-p^0)(x_1-p^0N)^2+p^0((N-x_1)-(1-p^0)N)^2}{(1-p^0)p^0N} = \frac{(x_1-Np^0)^2}{(1-p^0)p^0N} = \left(\frac{x_1-Np^0}{\sqrt{(1-p^0)p^0N}}\right)^2$ where p is the probability of outcome 1 and $(1-p)pN$ is the variance of the binomiaI. For sufficiently large samples, $\frac{x_1-Np^0}{\sqrt{(1-p^0)p^0N}}$ is a standard normal random variable, making the square of it Chi-squared. Notice that the denominator, $\sqrt{(1-p^0)p^0N}$, is the standard error of the binomial variable assuming $p = p^0$, the null hypothesis. For large samples, the distribution of $(x_1 - Np^0)$ is normal, one standardizes it by dividing it by a standard error. If one divides by the standard error assuming the null hypothesis is correct, $p = p^0$, the result is a score statistic; if one standardizes using the estimated standard error, $\sqrt{(1-\widehat{p})\widehat{p}N}$, one has a Wald statistic.[4]

1.3 A Chi-squared likelihood ratio test of the null hypothesis

Instead of doing the Pearson Chi-squared test, one can alternatively do a likelihood ratio test–one can almost always do a likelihood-ratio test. Given the sample, $x_1, x_2, ...x_C$, one finds the the maximum value of the ln likelihood function assuming no constraints on the p other than $0 < p_c < 1$ and $\sum_{c=1}^{C} p_c = 1$. The maximum value of the likelihood function is

$$\ln L^{ml} \equiv \sum_{c=1}^{C} x_c \ln p_c^{ml}$$

where the p_c^{ml} are the p_c that maximimize $\ln L = \sum_{c=1}^{C} x_c \ln p_c$, the maximimum likelihood estimates of the multinomial probabilities. One can easily show that $p_c^{ml} = \frac{x_c}{N}$, so $\ln L^{ml} = \sum_{c=1}^{C} x_c \ln p_c^{ml} = \sum_{c=1}^{C} x_c \ln(\frac{x_c}{N})$

Then find the value of the likelihood function assuming the null p: $p_1^0, p_2^0, ...p_C^0$. Denote this amount $\ln L^0 \equiv \sum_{c=1}^{C} x_c \ln p_c^0$

For large samples, the statistic $LRS = -2(L^0 - L^{ml})$, has, approximately, a Chi-squared distribution with $C - 1$ degrees of freedom (Wilks (1935, 1938).

[4] Note that $\widehat{p}$ is the maximum-likelihood estimate of p.

So, one calcuates lrs for one's sample and rejects the null hypothesis if if lrs is greater than the chosen critical value of χ^2_{C-1}, for the chosen level of significance.

The Pearson Chi-squared test and the likelihood ratio Chi-squared test will not always agree on whether the null should be rejected , but converge to the same test as N approaches infinity.

1.4 References

A. Agresti, 2002, Categorical data analysis: second edition, Wiley-Interscience

K. Pearson, 1900, On the criterion that a given set of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philos. Mag. Ser.* 5 (50), 157-175. (reprinted in Karl Pearson's early statistical papers, ed. E.S. Pearson, Cambridge University Press, 1948)

S. Stigler, 1986, The history of statistics: the measurement of uncertainty before 1900, Harvard University Press

S.S. Wilks, The likelihood test of independence in contingency tables, *Annals of Mathematical Statistics* 6: 190-196. (1935)

S.S. Wilks, The large sample distribution of the likelihood ratio for testing composite hypoheses, *Annals of Mathematical Statistics* 9: 60-62. (1935)

Some comments wrt student answers to asking them to present two chi-squared tests of the null hypothesis that the multinomial has certain specific parameter values:

The KP statistics does not require M.L. estimation. Some said it did.

Everything in the world is not a least-squares problem.

A "F" test is a F test, not a chi-squared test.

Some of you suggested a Wald test, or wrote down either a Wald statistics or something close to it. This was fine as long as you did it correctly and specifically for the multinomial problem at hand. Both the KP statistic and the Wald statistic, for our multinomial hypothesis, can be written as $\sum_{c=1}^{C} \frac{(x_c - p_c^0 N)^2}{var(x_c)}$. Whereas the the KP statistic calculates the $var(x_c)$ of x_c assuming the null hypothesis is correct ($var(x_c) = p^o N$), the Wald statistic calculates it assuming the estimated probabilities are correct. While the KP statistic, as noted above, is judging the null in terms of the slope(s) of the ln likelihood function evaluated at the null values of the probabilities, the Wald statistic is looking at the "average distance" between the null probabilities and the estimated probabilities, if they are far enough apart, reject the null.

2 Some links to other presentations and discussions of Pearson Chi-squared tests:

At Math World: http://mathworld.wolfram.com/Chi-SquaredTest.html

At Wiki: http://en.wikipedia.org/wiki/Pearson_chi-square_test .
At Wiki: http://en.wikipedia.org/wiki/Chi-square_goodness-of-fit_test

At Wiki: http://en.wikipedia.org/wiki/Chi-square_significance_test
Handbook of Biological Statistics: http://udel.edu/~mcdonald/statchigof.html