Pearson's Chi-Square

Guy Lebanon

January 29, 2009

In this note we describe Pearson's chi-square technique and provide a short proof based on [1] the asymptotic distribution of the chi-square expression.

Consider n RVs $X_1, \ldots, X_n$ generated iid from a multinomial distribution $\text{Mult}(1, \theta), \theta \in \mathbb{R}^k$. That is we have n iid draws from an experiment with k possible outcomes - each with probability θ_i. We consider each RV x_i as a binary vector whose value $[X_i]_j$ is 1 if the i draw resulted in outcome j.

Due to the random nature of the experiments the expected value $\mathsf{E} \sum X_i = n\theta$ should be similar but perhaps not identical to the actual value of $\sum X_i$. Pearson suggested that a test based on this discrepancy could be used to determine θ. More specifically, he showed that the following expression has a chi-squared distribution with $k-1$ degrees of freedom

$$\begin{aligned}\mathcal{Z}(\theta, X) &= \sum_{j=1}^{k} \frac{(\text{observed}(j) - \text{expected}(j))^2}{\text{expected}(j)} = \sum_{j=1}^{k} \frac{\left(\sum_i^n [X_i]_j - n\theta_j\right)^2}{n\theta_j} = n \sum_{j=1}^{k} \frac{\left(\sum_i^n [X_i]_j / n - \theta_j\right)^2}{\theta_j} \\ &= n(\overline{X}_n - \theta)^\top (\text{diag}(\theta))^{-1} (\overline{X}_n - \theta) \sim \chi^2_{k-1}\end{aligned}$$

where $\overline{X}_n = \frac{1}{n} \sum X_i$. This observation can be used to develop tests for θ and confidence intervals for θ. For example, for the test $H_0 : \theta = \theta_0$ we have a χ^2_{k-1} distribution under the null hypothesis for $\sum_{j=1}^{k} \frac{\sum_i^n [X_i]_j - n[\theta_0]_j}{n[\theta_0]_j}$. It is also possible to computationally invert $P(\mathcal{Z}(\theta, X) \leq c_\alpha) = 1 - \alpha$ in order to obtain a confidence interval of the form $\theta \in A(X)$ with probability $1 - \alpha$. Alternative methods such as the z-test (see note on confidence intervals) may be used to perform tests and obtain confidence intervals for the multinomial distribution. However, Pearson's chi-square is generally considered the standard method for the multinomial case and is often more accurate than the alternatives.

Definition 1. A projection or idempotent matrix A satisfy $A = A^2$.

Proposition 1. *For $X \sim N(0, \Sigma), X^\top X \sim \chi^2_r$ iff Σ is a symmetric projection matrix of rank r.*

Proof. [1] For symmetric matrices Σ, the eigendecomposition states that $D = Q\Sigma Q^\top$ where Q is an orthogonal matrix and D is a diagonal matrix containing the eigenvalues of Σ.

$\Sigma^2 = \Sigma$ and Σ has rank r iff $D^2 = D$ and D has rank r (since Q is orthogonal and the number non-zero $\text{diag}(D)$ elements is the rank of Σ) which happens iff r of the diagonal elements of D are 1 and the rest 0 (if D_{ii} is not 0 or 1 we have a contradiction to $D^2 = D$).

Since X is Gaussian, so is its linear transformation $Y = QX \sim N(0, D)$. The characteristic function of $\sum_i Y_i^2 = Y^\top Y = X^\top Q^\top Q X = X^\top X$ is a product of the characteristic functions of Y_i^2 which are χ^2_1 RVs: $\prod_i (1 - 2i\mathsf{Var}(Y_i)t)^{-1/2} = \prod_i (1 - 2iD_{ii}t)^{-1/2}$ (see note on sampling distributions) which equals the characteristic function of χ^2_r, $(1 - 2it)^{-r/2})$ iff r of the diagonal elements of D are 1 and the rest 0. Combining this with the previous paragraph proves the proposition. □

A useful corollary that we will need later is that $D_{ii} \in \{0, 1\}$ are the eigenvalues of Σ and so for any symmetric projection matrix Σ, $\text{rank}(\Sigma) = \text{trace}(\Sigma)$ (since both are a sum of the eigenvalues).

Proposition 2.

$$n(\overline{X}_n - \theta)^\top (diag(\theta))^{-1} (\overline{X}_n - \theta) \sim \chi^2_{k-1}$$

Proof. [1] By the CLT $\sqrt{n}(\overline{X}_n - \theta) \rightsquigarrow Y \sim N(0, \Sigma)$ with Σ being the multinomial covariance matrix $\Sigma = \text{Cov}(X_1) = \text{diag}(\theta) - \theta\theta^\top$. By Slutsky's theorem

$$n(\overline{X}_n - \theta)^\top (\text{diag}(\theta))^{-1} (\overline{X}_n - \theta) \rightsquigarrow Y^\top (\text{diag}(\theta))^{-1} Y$$

which is χ^2_{k-1} iff the covariance matrix of $(\text{diag}(\theta))^{-1/2} Y$ is a symmetric projection matrix of rank $k-1$ (by Proposition 1). Being a covariance matrix it is obvious symmetric. To show that it is projection and of rank $k-1$ we use the fact that $\mathsf{Cov}\,(AX) = A\mathsf{Cov}\,(X)A^\top$ to obtain

$$\begin{aligned}(\text{diag}(\theta))^{-1/2}\text{Cov}(Y)(\text{diag}(\theta))^{-1/2} &= (\text{diag}(\theta))^{-1/2}(\text{diag}(\theta) - \theta\theta^\top)(\text{diag}(\theta))^{-1/2} \\ &= I - (\text{diag}(\theta))^{-1/2}\theta\theta^\top(\text{diag}(\theta))^{-1/2}\end{aligned}$$

which is a projection matrix if $R = (\text{diag}(\theta))^{-1/2}\theta\theta^\top(\text{diag}(\theta))^{-1/2}$ is a projection matrix. This can be verified by comparing R_{ij} with R_{ij}^2. We show that $I - (\text{diag}(\theta))^{-1/2}\theta\theta^\top(\text{diag}(\theta))^{-1/2}$ has rank $k-1$ by noting that since it is a projection matrix, its rank is the same as its trace and

$$\begin{aligned}\text{rank}(I - (\text{diag}(\theta))^{-1/2}\theta\theta^\top(\text{diag}(\theta))^{-1/2}) &= \text{trace}(I) - \text{trace}((\text{diag}(\theta))^{-1/2}\theta\theta^\top(\text{diag}(\theta))^{-1/2}) \\ &= k - \sum_{i=1}^{k} \frac{\theta_i^2}{\theta_i} = k - 1.\end{aligned}$$

□

References

[1] T. S. Ferguson. *A Course in Large Sample Theory.* Chapman & Hall, 1996.