Chi-square distribution

Mingwei Fei

December 12, 2022

1 Normal Distribution

Generally, the normal distribution encountered with σ^2 known, so the sufficient statistics for μ is $\sum_{i=1}^{n} y_i$. However, if σ^2 is unknown, the sufficient statistics for μ depends on σ^2, and we need to pay attention to that.

The second we need to pay attention is that, the normal distribution has different μ_i for each patient. And we usually write in the matrix form of the likelihood function.

The projection operator generally applies to normal distribution, so we will usually write in matrix form for multivariate normal distribution. Because the likelihood function could be considered as MVN, as we are estimating β, σ^2 using all the y_i simultaneously.

2 Chi-square distribution

If $Z_1, ..., Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = Z_i^2 \sim \chi^2(k)$$
$$p(k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} exp(-\frac{x}{2})$$

3 Non-Central Chi-square distribution

The non-central chi-square distribution: Let $(X_1, X_2, \ldots, X_i, \ldots, X_k)$ be k independent, normally distributed random variables with means μ_i and unit variances. Then the random variable

$$Q = \sum_{i=1}^{k} X_i^2 \sim \chi^2(k, \lambda), \qquad \lambda = \sum_{i=1}^{k} \mu_i^2$$

where the degrees of freedom is k.

The sample mean of n i.i.d. chi-squared variables of degree k is distributed according to a gamma distribution with shape α and scale θ parameters:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim Gamma(\alpha = nk/2, \theta = 2/n)$$

3.1 Lemma

Let $Q_i \sim \chi^2_{k_i}(\lambda_i)$ for $i = 1, \ldots, n$, be independent. Then, $Q = \sum_{i=1}^{n} Q_i$ is a noncentral $\chi^2_k(\lambda)$, where $k = \sum_{i=1}^{n} k_i$ and $\lambda = \sum_{i=1}^{n} \lambda_i$.

Proof:

The distribution transformation use moment generating function.

3.1.1 Moment Generating Function

We can get MGF from $E[x^2 t]$

$$\begin{aligned}
M_i(t) = E[x^2 t] &= \frac{1}{\sqrt{2\pi}} \int exp(x^2 t) exp\left(-\frac{(x-\mu)^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int exp\left((t - \frac{1}{2})x^2 + \mu x - \frac{\mu^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int exp\left(-\frac{1}{2}(1-2t)\{x^2 - \frac{2\mu x}{(1-2t)} + \frac{\mu^2}{(1-2t)^2}\} + \frac{\mu^2}{2(1-2t)} - \frac{\mu^2}{2}\right) dx \\
&= \frac{1}{\sqrt{(1-2t)}} \int \frac{(1-2t)}{\sqrt{2\pi}} exp\left(-\frac{(x - \frac{\mu}{1-2t})^2}{2(1-2t)^{-1}}\right) dx \left[exp\left(\frac{\mu^2 t}{1-2t}\right)\right] \\
&= \frac{1}{\sqrt{(1-2t)}} exp\left(\frac{\mu^2 t}{1-2t}\right), \qquad \lambda = \mu^2 \\
&= \frac{1}{\sqrt{(1-2t)}} exp\left(\frac{\lambda t}{1-2t}\right)
\end{aligned}$$

Then the MGF for $Q_i \sim \chi^2_{k_i}(\lambda_i)$

$$\begin{aligned}
M(t) = E[\sum_{i=1}^{k} x_i^2 t] &= \prod_{i=1}^{k} M_i(t) \\
&= \left(\frac{1}{\sqrt{(1-2t)}}\right)^k exp\left(\frac{\sum_{i=1}^{k} \lambda_i t}{1-2t}\right) \\
&= \left(\frac{1}{\sqrt{(1-2t)}}\right)^k exp\left(\frac{\lambda t}{1-2t}\right) \\
&= (1-2t)^{-k/2} exp\left(\frac{\lambda t}{1-2t}\right), \qquad \text{i.i.d}
\end{aligned}$$

The general case of a linear combination of independent $\chi^2_{k_i}(\lambda_i)$

$$Q = \sum_{i=1}^{k} a_i Q_i$$

We also can prove using MGF.

3.1.2 Linear Combination of Chi-Square Distribution

The linear combination of chi-square distribution Y_j. Let us denote by $X \sim \Gamma(r, \lambda)$ the fact that the r.v. X has a Gamma distribution with shape parameter r and rate parameter λ

$$f_X(x) = \frac{\lambda^x}{\Gamma(r)} exp(-\lambda x) x^{r-1}, \qquad (r, \lambda > 0, x > 0)$$

Then we have, for $j = 1, ..p$,

$$Y_j \sim \Gamma(\frac{k_j}{2}, \frac{1}{2}) \rightarrow Z_j = w_j Y_j \sim \Gamma(\frac{k_j}{2}, \frac{1}{2w_j})$$

The MGF for linear combinations $Z_j = w_j Y_j$

$$\begin{aligned} M(t) &= E[exp(Y_j t)] = (1-2t)^{-k/2} exp\left(\frac{\lambda t}{1-2t}\right) \\ M_{Z_j}(t) &= E[exp(w_j Y_j t)] = E[exp(Y_j(w_j t))] \\ &= (1-2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1-2w_j t}\right) \end{aligned}$$

$$\begin{aligned} M_Y(t) &= E[exp(Yt)] = E[exp(t[w_1Y_1 + w_2Y_2 + w_3Y_3 + ..w_nY_n])] \\ &= E[exp(w_1tY_1)]E[exp(w_2tY_2)]...E[exp(w_ntY_n)] \\ &= M_{X_1}(w_1t)M_{X_2}(w_2t)M_{X_3}(w_3t)..M_{X_n}(w_nt) \\ &= \prod_{i=1}^{n} M_{X_i}(w_it) \end{aligned}$$

The third equation comes from the properties of exponents, as wells as from the expectation of the product of functions of independent random variables.

I need to pay attention that, only under independent and identical situation, we can write

$$M_Y(t) = M_X(t)^n$$

Other than that, we can not further simplify that. So back to the non-central chi-square distribution, we have the MGF of Y

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(w_i t) \\ &= \prod_{i=1}^{n} (1-2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1-2w_j t}\right) \end{aligned}$$

Then we can see that the shape parameter is $\frac{1}{2w_i}$.

3.2 b

Consider the following

(a) For an arbitrary model, consider the conditional score statistic

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi}$$

Show that the conditional score statistic for any model can be written as

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

The conditional score statistic is the derivative of the conditional distribution

$$\begin{aligned} U_\psi(\xi) &= \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi} \\ p(\mathbf{Y}|\xi) &= p(\mathbf{Y}|s_\lambda(\psi_0), \xi) p(s_\lambda(\psi_0)|\xi), \qquad p(\mathbf{Y}|s_\lambda(\psi_0), \xi) = \frac{p(\mathbf{Y}|\xi)}{p(s_\lambda(\psi_0)|\xi)} \\ l_c(\xi, \psi_0) &= logp(\mathbf{Y}|s_\lambda(\psi_0), \xi) = logp(\mathbf{Y}|\xi) - logp(s_\lambda(\psi_0)|\xi) \end{aligned}$$

Then we need to prove

$$\begin{aligned} U_\psi(\xi) &= \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi} = \partial_\psi logp(\mathbf{Y}|\xi) - \partial_\psi logp(s_\lambda(\psi_0)|\xi) \\ \partial_\psi logp(s_\lambda(\psi_0)|\xi) &= E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi} \end{aligned}$$

We can write

$$\begin{aligned} logp(\mathbf{Y}|\xi) &= logp(\mathbf{Y}|s_\lambda(\psi_0), \xi) + logp(s_\lambda(\psi_0)|\xi) \\ E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) &= E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) + E\left(\partial_\psi[logp(s_\lambda(\psi_0), \xi)|s_\lambda]\right) \end{aligned}$$

in which, the integral and expectation can switch, then we have

$$E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|\xi)]\right) = 0$$

So,

$$E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) = \partial_\psi logp(s_\lambda(\psi_0), \xi)$$

Then we show

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

(b) Suppose that $y_1; ..y_n$ are independent and y_i follows a Poisson distribution with mean $exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2})$, where $(x_{i1}; x_{i2})$ are covariates, $\lambda = (\lambda_0; \lambda_1)$ is the nuisance parameter vector and ψ is the parameter of interest. Derive the conditional likelihood of ψ and show that this conditional likelihood is free of λ.
The joint distribution of $(y_1, , y_n)$ is given by

$$P(Y|\lambda,\psi) = exp\left(\sum_{i=1}^n y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)$$

Thus, $S_0 = \sum_{i=1}^n y_i$ is the sufficient and complete statistics for λ_0, and $S_1 = \sum_{i=1}^n y_i x_{i1}$ is the sufficient and complete statistics for λ_1.
The conditional distribution of ψ given S_0, S_1 is given by

$$\begin{aligned}
p(\mathbf{Y},\psi|S=(S_0,S_1)) &= \frac{exp\left(\sum_{i=1}^n y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)}{\sum_{y'\in S} exp\left(\sum_{i=1}^n y'_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy\right.} \\
&= \frac{exp\left(S_1\lambda_0 + S_2\lambda_1 + S_3\psi\right) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!)}{\sum_{y'\in S} exp\left(S'_1\lambda_0 + S'_2\lambda_1 + S'_3\psi\right) - \sum_{i=1}^n exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy'_i!)} \\
&= \frac{exp\left(S_3\psi - logy_i!\right)}{\sum_{y'\in S} exp\left(S'_3\psi - logy'_i!\right)}, \qquad S_3 = \sum_{i=1}^n y_i x_{i2}, S'_3 = \sum_{i=1}^n y'_i x_{i2}
\end{aligned}$$

which is independent of λ.

(c) Derive the conditional score statistic for part (b) and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ based on $U_\psi(\xi)$.
The log likelihood of the conditional distribution is

$$l_c(\psi) = S_3\psi - logy_i! - log\left[\sum_{y'\in S} exp\left(S'_3\psi - logy'_i!\right)\right], \qquad S_3 = \sum_{i=1}^n y_i x_{i2}, S'_3 = \sum_{i=1}^n y'_i x_{i2}$$

The score function and observed fisher information is

$$\begin{aligned}
U_\psi(\xi) &= \frac{\partial l_c(\xi,\psi_0)}{\partial\psi}|_{\psi_0=\psi} \\
&= \psi - \frac{\sum_{y'\in S} S'_3 exp\left(S'_3\psi - logy'_i!\right)}{\sum_{y'\in S} exp\left(S'_3\psi - logy'_i!\right)}
\end{aligned}$$

$$\frac{\partial^2 l_c(\xi,\psi_0)}{\partial\psi^2} = \left[\frac{\sum_{y'\in S} S'_3 exp\left(S'_3\psi - logy'_i!\right)}{\sum_{y'\in S} exp\left(S'_3\psi - logy'_i!\right)}\right]^2 - \frac{\sum_{y'\in S} S'^2_3 exp\left(S'_3\psi - logy'_i!\right)}{\sum_{y'\in S} exp\left(S'_3\psi - logy'_i!\right)}$$

The newton-Raphson algorithm

$$\psi^{k+1} = \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial \psi^2}\right]^{-1} U_\psi(\psi^k)$$

where $\frac{\partial^2 l_c(\psi^k)}{\partial \psi^2}, U_\psi(\psi^k)$ are from above equations.

(d) Now suppose that we only have two random variables $y_1 \sim Poisson(\mu_1)$ and $y_2 \sim Poisson(\mu_2)$, where y_1 and y_2 are independent. We are interested in making inferences on the ratio $\psi = \mu_1/\mu_2$. Let $\xi = (\psi, \lambda)$, where λ represents the nuisance parameter.

(i) Show that the log-likelihood function of ξ can be written as

$$l(\xi) = (y_1 + y_2)\lambda + y_1 log(\psi) - exp(\lambda)(1 + \psi)$$

where λ is a function of μ_2. Explicitly state what λ is.
Write the joint distribution of y_1, y_2

$$\begin{aligned}
P(y_1, y_2) &= \frac{\mu_1^{y_1} e^{-\mu_1}}{y_1!} \frac{\mu_2^{y_2} e^{-\mu_2}}{y_2!} \\
logP(y_1, y_2) &= y_1 log\mu_1 - \mu_1 + y_2 \log \mu_2 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + y_1 log\mu_2 + y_2 log\mu_2 - \mu_1 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + (y_1 + y_2) log\mu_2 - \mu_2(\mu_1/\mu_2 + 1) - logy_1! - logy_2!
\end{aligned}$$

where

$$\begin{aligned}
\psi &= log\frac{\mu_1}{\mu_2} \\
\lambda &= log\mu_2
\end{aligned}$$

(ii) Derive the conditional likelihood of ψ and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ .
From part (a), we see $y_1 + y_2$ is the sufficient statistics for λ, while $y_1 + y_2 \sim Poission(\mu_1 + \mu_2)$ then we have conditional distribution of ψ condition on $S = y_1 + y_2$.

$$\begin{aligned}
Y(\psi | S = y_1 + y_2, \lambda) &= \frac{exp\left[y_1\psi + (y_1 + y_2)\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[(y_1 + y_2) log(\mu_1 + \mu_2) - (\mu_1 + \mu_2) - log(y_1 + y_2)!\right]} \\
&= \frac{exp\left[y_1\psi + S\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[S(\lambda + log(\psi + 1)) - exp(\lambda)(\psi + 1) - logS!\right]} \\
&= \frac{exp\left[y_1\psi - logy_1! - logy_2!\right]}{exp\left[(y_1 + S - y_1) log(\psi + 1)) - logS!\right]} \\
&= \binom{S}{y_1} \left(\frac{\psi}{1 + \psi}\right)^{y_1} \left(\frac{1}{1 + \psi}\right)^{S - y_1}
\end{aligned}$$

The conditional distribution is a binomial, $B(S, \psi/(1+\psi))$.
The score function and observed fisher information

$$
\begin{aligned}
logY(\psi|S,\lambda) &= y_1 log\psi - Slog(1+\psi) + log\binom{S}{y_1} \\
\partial_\psi logY(\psi|S,\lambda) &= \frac{y_1}{\psi} - \frac{S}{1+\psi} = 0, \qquad \hat{\psi} = y_1/(S-y_1) \\
\partial_\psi^2 logY(\psi|S,\lambda) &= -\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}
\end{aligned}
$$

The $CMLE = \hat{\psi} = y_1/(S-y_1)$. And the newton-Raphson equation

$$
\begin{aligned}
\psi^{k+1} &= \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial\psi^2}\right]^{-1} U_\psi(\psi^k) \\
&= \psi^k - \left[-\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}\right]^{-1} \left[\frac{y_1}{\psi} - \frac{S}{1+\psi}\right] |_{\psi=\psi^k} \\
&= \psi^k + \frac{y_1/\psi^k - S/(1+\psi^k)}{y_1/{\psi^k}^2 - S/(1+\psi^k)^2}
\end{aligned}
$$