

36-705 Intermediate Statistics Fall 2016: Homework 5.

Date: 17th August 2020.

These are my attempted solutions to the homework.

Correction status: pending.

1) Chapter 9, problem 1.

Let $X_1, X_2, \dots, X_n \sim \text{Gamma}(a, b)$, where the X_i are IID. Find the method of moments estimator for a and b .

A random variable X_i with a Gamma probability density function has the following mean and variance:

$$\mathbb{E}[X_i] = ab$$

$$\mathbb{V}[X_i] = ab^2$$

Hence the second moment of the distribution is given by:

$$\mathbb{E}[X^2] = \text{Var}[X_i] + \mathbb{E}[X_i]^2 = ab^2 + a^2b^2 = ab^2(1 + a) = ab(b + ab)$$

To find the method of moments estimators of the parameters a and b , that is \hat{a}_{MoM} and \hat{b}_{MoM} , we equate the 1st and 2nd sample moments with the theoretical moments to get 2 equations.

Denoting the k th sample moment as $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$, we have that:

$$m_1 = \mathbb{E}[X_i] \implies \frac{1}{n} \sum_{i=1}^n X_i = ab$$

$$m_2 = \mathbb{E}[X_i^2] \implies \frac{1}{n} \sum_{i=1}^n X_i^2 = ab^2(1 + a)$$

Rearranging the 1st equation yields $b = \frac{1}{a} \bar{X}_n$, which we substitute into the 2nd equation to yield:

$$\begin{aligned}
m_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 = \bar{X}_n \left(\frac{1}{a} \bar{X}_n + \bar{X}_n \right) \\
\implies \frac{1}{n} \sum_{i=1}^n X_i^2 &= \frac{1}{a} \bar{X}_n^2 + \bar{X}_n^2 \\
\implies \frac{1}{a} \bar{X}_n^2 &= \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2
\end{aligned}$$

Hence we have that:

$$\begin{aligned}
\hat{a}_{\text{MoM}} &= \frac{\bar{X}_n^2}{\left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2} = \frac{m_1^2}{m_2 - m_1^2} \\
\hat{b}_{\text{MoM}} &= \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2}{\bar{X}_n} = \frac{m_2 - m_1^2}{m_1}
\end{aligned}$$

2) Chapter 9, problem 5.

Let the $X_1, \dots, X_n \sim Po(\lambda)$, i.e. have a Poisson distribution parametrised by λ .

Find the:

- a) Method of moments estimator.
- b) Maximum likelihood estimator.
- c) Fisher information.

2a)

Equating the 1st sample moment with the 1st theoretical moment, i.e. the mean, we have:

$$\frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X_i] = \lambda$$

Hence,

$$\hat{\lambda}_{\text{MoM}} = \bar{X}_n$$

2b)

The Poisson distribution has probability mass function:

$$f_{X_i}(x_i; \lambda) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad x_i \geq 0$$

The likelihood function $L(\lambda)$ is a function of the parameter λ and is given by the product of individual Poisson probability mass functions, as the data is IID.

The value of the parameter that maximises the log-likelihood function $l(\lambda)$ evaluated at the values of the data (x_1, x_2, \dots, x_n) is the maximum likelihood estimator, $\hat{\lambda}_{\text{ML}}$:

$$\hat{\lambda}_{\text{ML}} = \underset{\lambda}{\operatorname{argmax}} \prod_{i=1}^n \log f_{X_i}(x_i; \lambda)$$

The likelihood function $L(\lambda)$ is:

$$L(\lambda) = \prod_{i=1}^n f_{X_i}(x_i; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i!)}$$

The log-likelihood function $l(\lambda)$ is:

$$l(\lambda) = \log \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i!)} = -n\lambda + \sum_{i=1}^n x_i \log \lambda - \sum_{i=1}^n \log(x_i!)$$

Taking derivatives with respect to λ and setting to 0, we have that:

$$\frac{d}{d\lambda} = -n + \sum_{i=1}^n x_i \left(\frac{1}{\lambda} \right) = 0$$

Hence we have that $\lambda_{\text{ML}} = \bar{X}_n$, i.e. the maximum likelihood estimator of the Poisson parameter is the sample mean, which coincides with that of the method of moments estimator.

2c)

For n observations, the Fisher information, $I_n(\lambda)$, is the variance of the score function, and the latter is the partial derivative of the log-likelihood. Suppressing the dependence of the score function on the data, we have that:

$$I_n(\lambda) = \operatorname{Var}_{\lambda}(s_n(\lambda)) = \operatorname{Var}_{\lambda}[l'(\lambda)]$$

$$s_n(\lambda) = l'(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i$$

$$\operatorname{Var}_{\lambda}[s_n(\lambda)] = \operatorname{Var} \left[-n + \frac{1}{\lambda} \sum_{i=1}^n x_i \right] = \frac{1}{\lambda^2} \left(\sum_{i=1}^n x_i \right)$$

As the data X_i are IID, we have that $\operatorname{Cov}[X_i, X_j] = 0$, meaning that:

$$\operatorname{Var}_{\lambda}[s_n(\lambda)] = \operatorname{Var} \left[\frac{1}{\lambda} \sum_{i=1}^n x_i \right] = \frac{1}{\lambda^2} \sum_{i=1}^n \operatorname{Var}[x_i] = \frac{1}{\lambda^2} n\lambda = \frac{n}{\lambda}$$

Hence the Fisher information $I_n(\lambda) = \frac{n}{\lambda}$.

3) Chapter 9, problem 7 (a, c, d, e)

Let $X_1, \dots, X_n \sim N(\theta, 1)$, where the X_i are IID.

Define:

$$Y_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \leq 0 \end{cases}$$

Let $\psi = P(Y_i = 1)$.

a) Find the maximum likelihood estimator $\hat{\psi}$ of ψ .

b) Define $\tilde{\psi} = \frac{1}{n} \sum_{i=1}^n Y_i$, and show that $\tilde{\psi}$ is a consistent estimator of ψ .

d) Compute the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$ and use the Delta method to get the standard error of maximum likelihood estimator. Then compute the standard error of $\tilde{\psi}$.

e) Suppose that the data is not Normal. Show that the $\hat{\psi}$ is not consistent. What, if anything, does $\hat{\psi}$ converge to?

3a)

The random variables Y_i have Bernoulli distribution with parameter ψ .

We have that:

$$\psi = P(Y_i = 1) = P(X_i > 0) = 1 - P(X_i < 0) = 1 - P\left(z < \frac{-\theta}{1}\right) = 1 - \Phi(-\theta) = \Phi(\theta)$$

Therefore we also have that $1 - \psi = \Phi(-\theta) = 1 - \Phi(\theta)$

As the Y_i have a Bernoulli distribution, the mean and variance of the Y_i are given by $\mathbb{E}[Y_i] = \psi$ and $\text{Var}[Y_i] = \psi(1 - \psi)$.

The maximum likelihood estimator $\hat{\theta}$ of the parameter θ is the sample mean, $\hat{\theta} = \bar{X}_n = \sum_{i=1}^n X_i$.

Noting that our parameter $\psi = g(\theta) = \Phi(\theta)$, we can use the equivariance property of the maximum likelihood estimator to yield:

$$\hat{\psi} = g(\hat{\theta}) = \Phi(\hat{\theta}) = \Phi(\bar{X}_n)$$

- Is this result contingent on regularity conditions, or is that more specifically for asymptotic results on MLE such as consistency and asymptotic Normality. Seems like consistency is a statistical concern, whereas the setting above is probabilistic.

3b)

The proposed estimator $\tilde{\psi}$ is the sample mean of the Bernoulli random variables Y_i , and the parameter ψ is the population mean of the Bernoulli distribution, $\psi = \mathbb{E}[Y_i]$.

As the X_i are IID, and the Bernoulli random variables Y_i have been defined as indicator functions dependent on the positivity of the X_i , the former are also IID.

The weak law of large numbers (WLLN) for IID random variables, the sequence of sample means converges in probability to the population mean. We have that for all $\epsilon > 0$, as $n \rightarrow \infty$:

$$P(|\bar{Y}_n - \mathbb{E}[Y_i]| > \epsilon) \rightarrow 0 \implies P(|\tilde{\psi} - \psi| > \epsilon) \rightarrow 0 \implies \tilde{\psi} \xrightarrow{p} \psi$$

Hence $\tilde{\psi}$ is a consistent estimator the parameter ψ .

Alternatively, as the proposed estimator $\tilde{\psi}$ is a sample mean, it is an unbiased estimator of the parameter ψ , i.e. $B = 0$. Hence the mean squared error of the estimator will equal to the variance of the estimator:

$$\text{MSE} = \mathbb{E}[(\tilde{\psi} - \psi)^2] = B^2 + V = V = \text{Var}_\psi[\tilde{\psi}] = \frac{\psi(1 - \psi)}{n}$$

As $n \rightarrow \infty$, we have that $\text{MSE} \rightarrow 0$, and we have convergence in quadratic mean of the estimator to the parameter, $\tilde{\psi} \xrightarrow{\text{q.m.}} \psi$, which implies convergence in probability, $\tilde{\psi} \xrightarrow{p} \psi$.

- Do we have regularity conditions to think about here?
- Consistency of the MLE, vs WLLN.

3c)

The asymptotic relative efficiency (ARE) of $\tilde{\psi}$ to $\hat{\psi}$, is the following ratio of their asymptotic variances:

$$\text{ARE}(\tilde{\psi}, \hat{\psi}) = \frac{\sigma_{\hat{\psi}}^2}{\sigma_{\tilde{\psi}}^2}$$

This comparison valid only if the two estimators are asymptotically Normal. We begin by computing the asymptotic variance of $\hat{\psi}$ followed by that of $\tilde{\psi}$.

Recall that $\hat{\psi} = g(\hat{\theta})$, i.e. a function of a maximum likelihood estimator. Asymptotic normality can be established by the delta method, which under appropriate conditions, states that:

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N\left(0, \frac{g'(\theta)^2}{I_n(\theta)}\right) \implies (g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N\left(0, \frac{g'(\theta)^2}{I_n(\theta)}\right) \implies g(\hat{\theta}) \xrightarrow{d} N\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right)$$

Where $I_n(\theta)$ is the Fisher information on the entire dataset and $I(\theta)$ is the Fisher information of a single datapoint. As the data are IID, we have that $I_n(\theta) = nI(\theta)$.

Hence we have that the estimator $\hat{\psi}$ is asymptotically Normal with mean $g(\theta)$ and (asymptotic) variance:

$$\sigma_{\hat{\psi}}^2 = \frac{g'(\theta)^2}{I_n(\theta)}$$

As $g(\theta) = \Phi(\theta)$, which is the standard Normal cumulative distribution function, we have that $g'(\theta)^2 = \phi^2(\theta)$, where $\phi(\cdot)$ is the standard Normal probability density function.

To compute the Fisher information over the entire dataset, $I_n(\theta)$, we use the result that it is equal to the variance of the score function over the entire dataset, $s_n(\theta, x_1, \dots, x_n)$:

$$I_n(\theta) = \text{Var}_\theta[s_n(\theta)]$$

Where we have suppressed the dependence of the score on the data for notational brevity.

As the score function is equal to the derivative of the log-likelihood, $l'(\theta)$, denoting the PDF of the datapoint X_i as f_{X_i} we have that:

$$\begin{aligned} s_n(\theta) = l'(\theta) &= \frac{\partial}{\partial \theta} \log \prod_{i=1}^n f_{X_i}(x_i; \theta) \\ &= \frac{\partial}{\partial \theta} \log \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \theta)^2\right) \\ &= \frac{\partial}{\partial \theta} \sum_{i=1}^n -\frac{1}{2} \log(2\pi) - \frac{1}{2}(x_i - \theta)^2 \\ &= -\frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial \theta} (\log(2\pi) + (x_i - \theta)^2) \\ &= -\frac{1}{2} \sum_{i=1}^n -2x_i + 2\theta \\ &= \left(\sum_{i=1}^n x_i\right) - n\theta \\ &= n(\bar{X}_n - \theta) \end{aligned}$$

As the mean of the score function is 0, that is $\mathbb{E}_\theta[s_n(\theta)] = 0$, the variance of the score function is equal to its second moment:

$$\begin{aligned} \text{Var}_\theta[s_n(\theta)] &= \mathbb{E}_\theta[s_n^2(\theta)] = \mathbb{E}_\theta[n^2(\bar{X}_n - \theta)^2] \\ &= n^2 (\mathbb{E}[\bar{X}_n^2] - 2\theta \mathbb{E}[\bar{X}_n] + \theta^2) \\ &= n^2 \left[\left(\frac{1}{n} + \theta^2\right) - 2\theta^2 + \theta^2 \right] \\ &= n^2 \left(\frac{1}{n}\right) \\ &= n \end{aligned}$$

Where we have used the fact that \bar{X}_n is a sample mean with mean $\mathbb{E}[\bar{X}_n] = \theta$, variance $\text{Var}[\bar{X}_n] = \frac{1}{n}$ and 2nd moment $\mathbb{E}[\bar{X}_n^2] = \text{Var}[\bar{X}_n] + \mathbb{E}[\bar{X}_n]^2 = \frac{1}{n} + \theta$.

Hence we have the Fisher information over the entire dataset $I_n(\theta) = n$, and together with $g'(\theta) = \phi^2(\theta)$, this yields the following asymptotic variance on $\hat{\psi}$:

$$\sigma_{\hat{\psi}}^2 = \frac{\phi^2(\theta)}{n}$$

And therefore the estimator $\hat{\psi}$ has the following asymptotically Normal distribution:

$$\hat{\psi} \approx \left(\Phi(\theta), \frac{\phi^2(\theta)}{n} \right)$$

We now compute the asymptotic variance of the estimator $\tilde{\psi} = \sum_{i=1}^n Y_i$. As this estimator is the maximum likelihood estimator of the Bernoulli parameter ψ , and also the sample mean of the data Y_i , we can use the asymptotic Normality of the maximum likelihood estimator or the Central Limit Theorem on the sample mean to yield:

$$\frac{\sqrt{n}(\tilde{Y}_n - \mathbb{E}[Y_i])}{\text{Var}[Y_i]} \xrightarrow{d} N(0, 1) \implies \tilde{\psi} \xrightarrow{d} N\left(\psi, \frac{\psi(1-\psi)}{n}\right)$$

Hence the estimator $\tilde{\psi}$ has the following asymptotically Normal distribution:

$$\tilde{\psi} \approx N\left(\Phi(\theta), \frac{\Phi(\theta)(1-\Phi(\theta))}{n}\right)$$

And its asymptotic variance is given by:

$$\sigma_{\tilde{\psi}}^2 = \frac{\Phi(\theta)(1-\Phi(\theta))}{n}$$

And hence we have that the asymptotic relative efficiency of the two estimators is:

$$\text{ARE}(\tilde{\psi}, \hat{\psi}) = \frac{\sigma_{\hat{\psi}}^2}{\sigma_{\tilde{\psi}}^2} = \frac{\phi^2(\theta)}{\Phi(\theta)(1-\Phi(\theta))}$$

The asymptotic standard errors of the two estimators are given by:

$$\sigma_{\hat{\psi}} = \frac{\phi(\theta)}{\sqrt{n}} \quad , \quad \sigma_{\tilde{\psi}} = \sqrt{\frac{\Phi(\theta)(1-\Phi(\theta))}{n}}$$

Plugging in the maximum likelihood estimate, $\hat{\theta}$ of the parameter θ yields the following estimated standard errors:

$$\widehat{\text{se}}(\hat{\psi}) = \frac{\phi(\hat{\theta})}{\sqrt{n}} = \frac{\phi(\bar{X}_n)}{\sqrt{n}} \quad , \quad \widehat{\text{se}}(\tilde{\psi}) = \sqrt{\frac{\Phi(\hat{\theta})(1-\Phi(\hat{\theta}))}{n}} = \sqrt{\frac{\Phi(\bar{X}_n)(1-\Phi(\bar{X}_n))}{n}}$$

These delta method also allows for the use of these estimated standard errors to yield the following asymptotic distributions:

$$\hat{\psi} \xrightarrow{d} N\left(\Phi(\theta), \frac{\phi^2(\bar{X}_n)}{n}\right)$$

And

$$\tilde{\psi} \xrightarrow{d} N\left(\Phi(\theta), \frac{\Phi(\bar{X}_n)(1 - \Phi(\bar{X}_n))}{n}\right)$$

4) Chapter 9, problem 7 (a, b, c).

Comparing two treatments, we have that n_1 people are given treatment 1 and n_2 people are given treatment 2. Let X_1 be the number of people treatment 1 who respond favourably to the treatment, and let X_2 be the number of people on treatment 2 who respond favourably.

Assume that $X_1 \sim \text{Binomial}(n_1, p_1)$, and $X_2 \sim \text{Binomial}(n_2, p_2)$; and let $\psi = p_1 - p_2$.

- a) Find the maximum likelihood estimator $\hat{\psi}$ for ψ .
- b) Find the Fisher information matrix $I(p_1, p_2)$.
- c) Use the multiparameter delta method to find the asymptotic standard error of $\hat{\psi}$.

4a)

In this case, we assume that the the number of people who respond favourably to treatment 1 and treatment 2 are independent. That is, $X_1 \perp X_2$. The setting we have differs from the usual IID setting because by modelling the number of people who respond to treatment 1 and treatment 2 with distinct Binomial distributions with their own sets of parameters means that the two random variables are not identically distributed - we are dealing with a multiparameter problem.

A discrete random variable X_i with Binomial distribution, $X_i \sim \text{Binomial}(n_i, p_i)$ has the following probability mass function:

$$f_{X_i}(x_i; n_i, p_i) = P(X_i = x_i) = \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}$$

We stack the scalar parameters and maximum likelihood estimators as vectors like so:

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \quad \hat{\mathbf{p}} = \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix}$$

We treat the number of trials parameters, $\mathbf{n} = (n_1, n_2)^T$ in the Binomial distributions as nuisance parameters, and denote their maximum likelihood estimators as $\hat{\mathbf{n}}$.

The likelihood function $L(\mathbf{n}, \mathbf{p})$ is a function of all parameters, including our nuisance parameters. We now define the profile likelihood for our parameters of interest, \mathbf{p} , which is the likelihood function maximised only with respect to our nuisance parameters:

$$L(\mathbf{p}) = \sup_{\mathbf{n}} L(\mathbf{n}, \mathbf{p}) = L(\hat{\mathbf{n}}, \mathbf{p})$$

We now consider maximising the profile log-likelihood with respect to the parameter of interest \mathbf{p} :

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmax}} l(\mathbf{p}) = \sum_{i=1}^2 \log \binom{\hat{n}_i}{x_i} + x_i \log p_i + (\hat{n}_i - x_i) \log(1 - p_i)$$

As the parameters are not coupled, we can treat this maximisation as 2 independent maximisation problems. Taking the derivative with respect to p_1 and setting to 0, we have that:

$$\frac{\partial l}{\partial p_1} = \frac{x_1}{p_1} - \frac{\hat{n}_1 - x_1}{1 - p_1} = \frac{x_1 - p_1 x_1 - \hat{n}_1 p_1 + p_1 x_1}{p_1(1 - p_1)} \implies p_1 = \frac{x_1}{\hat{n}_1}$$

And applying similar reasoning to p_2 , we have the maximum likelihood estimator:

$$\hat{\mathbf{p}} = \begin{pmatrix} \frac{x_1}{\hat{n}_1} \\ \frac{x_2}{\hat{n}_2} \end{pmatrix}$$

We now define the following scalar function of the parameter vector $\psi = g(\mathbf{p}) = p_1 - p_2$ where $g : \mathbb{R}^2 \mapsto \mathbb{R}$. Using equivariance of the maximum likelihood estimator we have that:

$$\hat{\psi} = g(\hat{\mathbf{p}}) = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{\hat{n}_1} - \frac{x_2}{\hat{n}_2}$$

Which is the maximum likelihood estimator for ψ .

4b)

The Fisher information matrix $I(p_1, p_2) \in \mathbb{R}^2$ can be computed by taking negative expectations of the Hessian matrix, \mathbf{H} . Where the Hessian matrix is with respect to the scalar-valued log-likelihood function, and with expectations being taken over the joint distribution of the data at the true value of the parameter:

$$I(p_1, p_2)_{ij} = -\mathbb{E}_{\mathbf{p}}[H_{ij}] = -\mathbb{E}_{\mathbf{p}} \left[\frac{\partial^2 l(\mathbf{p})}{\partial p_i \partial p_j} \right]$$

The Hessian of the scalar-valued log-likelihood function is equivalent to the Jacobian matrix, \mathbf{J} of the gradient of the log-likelihood $\nabla_{\mathbf{p}} l(\mathbf{p})$. And the gradient of the scalar-valued log-likelihood function is the vector-valued score function $s(\mathbf{n}, \mathbf{p})$:

$$\nabla_{\mathbf{p}} l(\mathbf{p}) = s(\mathbf{n}, \mathbf{p}) = \begin{pmatrix} \frac{1}{p_1(1-p_1)} x_1 - n_1 p_1 \\ \frac{1}{p_2(1-p_2)} x_2 - n_2 p_2 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

We have the following derivatives of the score function:

$$\begin{aligned}
\frac{\partial s_1}{\partial p_1} &= \frac{\partial}{\partial p_1} \left(\frac{x_1 - n_1 p_1}{p_1(1 - p_1)} \right) \\
&= \frac{-n_1[p_1(1 - p_1)] - (x_1 - n_1 p_1)(1 - 2p_1)}{[p_1(1 - p_1)]^2} \\
&= \frac{-n_1 p_1 + n_1 p_1^2 - (x_1 - 2p_1 x_1 - n_1 p_1 + 2n_1 p_1^2)}{[p_1(1 - p_1)]^2} \\
&= \frac{-n_1 p_1^2 - x_1 + 2p_1 x_1}{[p_1(1 - p_1)]^2}
\end{aligned}$$

And the cross-derivatives of the log-likelihood are 0:

$$\frac{\partial s_1}{\partial p_2} = \frac{\partial}{\partial p_2} \left(\frac{x_1 - n_1 p_1}{p_1(1 - p_1)} \right) = 0$$

Hence we have the following Hessian:

$$\mathbf{H} = \begin{pmatrix} \frac{-n_1 p_1^2 - x_1 + 2p_1 x_1}{[p_1(1 - p_1)]^2} & 0 \\ 0 & \frac{-n_2 p_2^2 - x_2 + 2p_2 x_2}{[p_2(1 - p_2)]^2} \end{pmatrix}$$

Computing negative expectations on the non-zero diagonal entries, we have:

$$\begin{aligned}
-\mathbb{E}[H_{11}] &= -\frac{1}{[p_1(1 - p_1)]^2} \mathbb{E}[-n_1 p_1^2 - x_1 + 2p_1 x_1] \\
&= -\frac{1}{[p_1(1 - p_1)]^2} (-n_1 p_1^2 - \mathbb{E}[x_1] + 2p_1 \mathbb{E}[x_1]) \\
&= -\frac{1}{[p_1(1 - p_1)]^2} (-n_1 p_1^2 - n_1 p_1 + 2n_1 p_1^2) \\
&= \frac{n_1 p_1 - n_1 p_1^2}{[p_1(1 - p_1)]^2} \\
&= \frac{n_1}{p_1(1 - p_1)}
\end{aligned}$$

Yielding the following Fisher information matrix:

$$I(p_1, p_2) = \begin{pmatrix} \frac{n_1}{p_1(1 - p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1 - p_2)} \end{pmatrix}$$

4c)

In order to compute the asymptotic standard error, and then the estimated standard errors, we use the multiparameter delta method, which states that for $\psi = g(\mathbf{p})$ and $\hat{\psi} = g(\hat{\mathbf{p}})$:

$$\frac{\hat{\psi} - \psi}{\text{se}(\psi)} \xrightarrow{d} N(0, 1) \implies \hat{\psi} \xrightarrow{d} N(\psi, \text{se}^2(\psi))$$

Denoting the gradient of $g(\cdot)$ evaluated at the value of the parameter as $\nabla_{\mathbf{p}}g(\mathbf{p})$; the asymptotic standard error, $\text{se}(\psi)$ is given by:

$$\text{se}(\psi) = \sqrt{\nabla_{\mathbf{p}}^T I^{-1}(p_1, p_2) \nabla_{\mathbf{p}}}$$

We have that the gradient of $g(\cdot)$ is given by:

$$\nabla_{\mathbf{p}}g(\mathbf{p}) = \begin{pmatrix} \frac{\partial g}{\partial p_1} \\ \frac{\partial g}{\partial p_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial p_1}(p_1 - p_2) \\ \frac{\partial}{\partial p_2}(p_1 - p_2) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

And as the Fisher information matrix is a diagonal matrix, we have that:

$$I^{-1}(p_1, p_2) = \begin{pmatrix} I_{11}^{-1} & 0 \\ 0 & I_{22}^{-1} \end{pmatrix} = \begin{pmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{pmatrix}$$

Yielding the following asymptotic standard errors:

$$\text{se}(\psi) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Plugging in the maximum likelihood estimators we derived, we have the following estimated asymptotic standard errors:

$$\widehat{\text{se}}(\hat{\psi}) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

And the above asymptotic distribution on $\hat{\psi}$ also holds if we replace the asymptotic standard errors $\text{se}(\psi)$ with their estimates $\widehat{\text{se}}(\hat{\psi})$.

5)

Let θ be a scalar parameter.

a) Let $\psi = g(\theta)$ where g is a smooth, invertible function. Hence, $\theta = h(\psi)$, where $h = g^{-1}$. Let $I(\theta)$ denote the Fisher information for θ and let $I(\psi)$ denote the Fisher information for ψ . Show that $I(\psi) = I(\theta)(h'(\psi))^2$.

b) Let $X \sim N(\theta, 1)$ and let $\psi = e^\theta$. Find the Fisher information for ψ .

5a)

I am not sure how to approach this question in its generality. In particular, it seems that this may involve the chain rule and the use of derivatives of inverse functions.

5b)

Without using the above result as I have not proved it, we find the Fisher information $I(\psi)$ for ψ in this particular instance. We will also find the Fisher information $I(\theta)$ for θ , with the hope that this example will shed some light on how to generalise the result to question 5a).

To put a name to what is going on, because $\psi = g(\theta) = e^\theta$ where θ parametrises a $N(\theta, 1)$ density, we can view the density $f'_X(x; \psi)$ as a *reparametrisation* by ψ .

Hence we have that

$$X \sim N(\log \psi, 1)$$

And the reparametrised density is:

$$f'_X(x; \psi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \log \psi)^2\right)$$

The log-likelihood as a function of ψ is therefore:

$$l(\psi) = -\log \sqrt{2\pi} - \frac{1}{2}(x - \log \psi)^2$$

The derivative of the log-likelihood, or score function, is:

$$s(x; \psi) = l'(\psi) = \frac{d}{d\psi} l(\psi) = 2 \cdot -\frac{1}{2}(x - \log \psi) \cdot -\frac{1}{\psi} = \frac{x - \log \psi}{\psi}$$

The 2nd derivative of the log-likelihood, or derivative of the score function, is:

$$l''(\psi) = \frac{d}{d\psi} \left(x\psi^{-1} - \psi^{-1} \log \psi \right) = -x\psi^{-2} - \left(-\psi^{-2} \log \psi + \psi^{-1} \cdot \frac{1}{\psi} \right) = \frac{\log \psi - x - 1}{\psi^2}$$

The Fisher information in ψ is the mean of the negative derivative of the score function, where the expectation is with respect to the reparametrised density $f'_X(x, \psi)$:

$$I(\psi) = \mathbb{E}_{f'_X(x; \psi)} \left[-\frac{d}{d\psi} s(x, \psi) \right] = \mathbb{E}_{f'_X(x; \psi)} [-l''(\psi)]$$

Considering expectations with respect to f'_X , but suppressing this in the notation, we have that $I(\psi)$ is:

$$\begin{aligned}
I(\psi) &= -\mathbb{E} \left[\frac{\log \psi - x - 1}{\psi^2} \right] \\
&= - \left(\frac{\log \psi - 1}{\psi^2} - \frac{\mathbb{E}[x]}{\psi^2} \right) \\
&= \frac{1}{\psi^2}
\end{aligned}$$

Where we have used the fact that $\mathbb{E}[X] = \log \psi$.

We now compute the Fisher information $I(\theta)$ using the original density $f_X(x; \theta)$. The log-likelihood is:

$$l(\theta) = -\log \sqrt{2\pi} - \frac{1}{2}(x - \theta)^2$$

The derivative of the log-likelihood, or score function is:

$$s(x, \theta) = l'(\theta) = (x - \theta)$$

The Fisher information is also the variance of the score function, and because mean of the score function is 0, where both operators are with respect to $f_X(x, \theta)$, we have that:

$$I(\theta) = \text{Var}_{f_X(x; \theta)}[s(x, \theta)] = \mathbb{E}[s^2(x, \theta)] - \mathbb{E}[s(x, \theta)]^2 = \mathbb{E}[s^2(x, \theta)]$$

Evaluating this we have that:

$$\begin{aligned}
I(\theta) &= \mathbb{E}[(x - \theta)^2] \\
&= \mathbb{E}[x^2 - 2\theta x + \theta^2] \\
&= \mathbb{E}[x^2] - 2\theta \mathbb{E}[x] + \theta^2 \\
&= (\theta^2 + 1) - 2\theta^2 + \theta^2 \\
&= 1
\end{aligned}$$

Note that in this particular example, we have that $I(\psi) = 1/\psi^2$, $I(\theta) = 1$, and $h'(\psi)^2 = 1/\psi^2$, so this can be viewed as an instance of the general result that $I(\psi) = I(\theta)[h'(\psi)]^2$ which we are at this stage unable to prove.