Statistical inference - Tutorial 3

Estevão B. Prado

April 12, 2021

1 Exercise 1

1) We know that $\{X_i\}_{i=1}^n \sim$ Negative Binomial(r, p), where r is known. We have that

$$f(x; r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}. \tag{1}$$

1.1 Part a

To solve this exercise, we'll use the definition of Exponential family distributions that's given in section 4.3 of **Sufficiency and the Rao-Blackwell Theorem** (weeks 6 and 7).

Definition. A k-parameter member of the exponential family has a density or frequency function of the form

$$f(\underline{x}; \theta) = \exp\left[\sum_1^k c_i(\theta) T_i(\underline{x}) + d(\theta) + S(\underline{x})\right]$$

for $\underline{x} \in A$ where the set A does not depend on θ.

That is, we need to show that $f(x; r, p)$ has the same form of $f(x; \theta)$. For instance, we can rewrite $f(.)$ as follows.

$$
\begin{aligned}
f(x;r,p) &= \binom{x-1}{r-1} p^r (1-p)^{x-r}, && (2)\\
&= \exp\left[\log\left[\binom{x-1}{r-1} p^r (1-p)^{x-r}\right]\right], && (3)\\
&= \exp\left[\log\binom{x-1}{r-1} + r\log(p) + (x-r)\log(1-p)\right], \text{by rule of log}, && (4)\\
&= \exp\left[S(x) + d(p) + c(p)\times T(x)\right], && (5)
\end{aligned}
$$

where $S(x) = \log\binom{x-1}{r-1}$, $d(p) = r\log(p)$, $c(p) = log(1-p)$ and $T(x) = x - r$, where r is known.

1.2 Part b

We recall the Factorisation Theorem, which is introduced in section 4.1 (Sufficiency) on page 2.

> **Theorem.** A necessary and sufficient condition for $T(\underline{X})$ to be sufficient for θ is that the joint density (or probability mass function) factors in the form
>
> $$f(\underline{x};\theta) = g(T(\underline{x});\theta)h(\underline{x})$$
>
> i.e. f denends on θ through T only.

$$
\begin{aligned}
\prod_{i=1}^{n} f(x_i;r,p) &= \prod_{i=1}^{n} \binom{x_i-1}{r-1} p^r (1-p)^{x_i-r}, && (6)\\
&= \prod_{i=1}^{n}\left[\binom{x_i-1}{r-1}\right] p^{nr}(1-p)^{\sum_{i=1}^{n}(x_i-r)}, && (7)\\
&= p^{nr}(1-p)^{\sum_{i=1}^{n}(x_i-r)} \prod_{i=1}^{n}\left[\binom{x_i-1}{r-1}\right], && (8)\\
&= g(T(\mathbf{x});p)h(\mathbf{x}), && (9)
\end{aligned}
$$

where $T(\mathbf{x}) = \sum_{i=1}^{n}(x_i - r)$ is the sufficient statistics for p.

Exercise 3

We know that $\{X_i\}_{i=1}^n \sim$ Negative Binomial(r, p), where r is known. We have that

$$f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}. \tag{10}$$

Part a

It's possible to answer this using two approaches: Factorisation Theorem or via a corollary of the Factorisation theorem, which uses the MLE. First, we use the FT as follows.

> **Theorem.** A necessary and sufficient condition for $T(\underline{X})$ to be sufficient for θ is that the joint density (or probability mass function) factors in the form
>
> $$f(\underline{x}; \theta) = g(T(\underline{x}); \theta) h(\underline{x})$$
>
> i.e. f denends on θ through T only.

$$\begin{aligned}
\prod_{i=1}^n f(x_i; k, p) &= \prod_{i=1}^n \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}, &(11)\\
&= \prod_{i=1}^n \left[\binom{k}{x_i}\right] p^{\sum_{i=1}^n x_i} (1-p)^{\sum_{i=1}^n (k-x_i)}, &(12)\\
&= p^{\sum_{i=1}^n x_i} (1-p)^{nk - \sum_{i=1}^n x_i} \prod_{i=1}^n \left[\binom{k}{x_i}\right], &(13)\\
&= g(T(\mathbf{x}); p) h(\mathbf{x}), &(14)
\end{aligned}$$

where $T(\mathbf{x}) = \sum_{i=1}^n x_i - r$ is the sufficient statistics for p.

Another possibility is to use a corollary of the Factorisation theorem. To use it, we first need to find the MLE of the parameter of interest (p).

Corollary of Factorisation theorem

If T is sufficient for θ, the MLE for θ is a function of T.

Here, we know that the MLE is $\hat{p} = \sum_{i=1}^{n} x_i/nk$. Hence, as the sufficient statistics T is a function of $\hat{p}$, we conclude that $T = \sum_{i=1}^{n} x_i$, as n and k are known.

Part b

First, we know that $\mathbb{E}(X_i) = kp$ and $\mathbb{V}(X_i) = kp(1-p)$. Again, there are two ways to answer this. For instance, we could show that $\hat{p}$ is an unbiased estimator (i.e., $\mathbb{E}(\hat{p}) = p$) and that $\mathbb{V}(\hat{p})$ attains the Cramer-Rao Lower bound. I guess this is the easiest way to show what we need.

The second alternative is to use the Lehmann-Scheffe Theorem.

Theorem. (Lehmann-Scheffe) If T is a complete sufficient statistic and $\hat{\theta}$ is an unbiased estimator of θ, then $\hat{\theta}_1 = \mathrm{E}(\hat{\theta}|T)$ is the unique best unbiased estimator of θ.

Completeness is tricky to prove, but it holds for most distributions we consider (Binomial, Exponential, Normal....)

We know that $\hat{p} = \sum_{i=1}^{n} x_i/nk$ is unbiased because

$$\begin{aligned}
\mathbb{E}(\hat{p}) &= \mathbb{E}\left(\frac{\sum_{i=1}^{n} X_i}{nk}\right), && (15)\\
&= \frac{\sum_{i=1}^{n} \mathbb{E}(X_i)}{nk}, (X_i's \text{ are i.i.d}) && (16)\\
&= \frac{n\mathbb{E}(X_i)}{nk}, (X_i's \text{ are i.i.d}) && (17)\\
&= \frac{\mathbb{E}(X_i)}{k}, && (18)\\
&= \frac{kp}{k}, && (19)\\
&= p. && (20)
\end{aligned}$$

In addition, recall that in a) we found that $T = \sum_{i=1}^{n} x_i$ is a sufficient statistics for p. However, by the Lehmann-Scheffe Theorem T is a **complete sufficient statistics** for p because the completeness of sufficient statistics holds for all exponential distributions (see section 4.3, page 6), which is the case of the Binomial distribution. Consequently, if $\mathbb{E}(\hat{p})$ is unbiased estimator and T is a **complete sufficient statistics**, then $\hat{p}$ is BUE.

Part c

To answer this, we need to get back all the way to weeks 2 and 3 (MLE, page 9) and refresh our memory about the Invariance property of the MLE.

Invariance property of the MLE

> **Definition.** If $\hat{\theta}$ is the mle of a parameter θ then for any function g $g(\hat{\theta})$ is the mle of $g(\theta)$.

That is, the MLE for $g(p) = kp(1-p)^{k-1}$ is $g(\hat{p}) = k\hat{p}(1-\hat{p})^{k-1}$, where $\hat{p} = \sum_{i=1}^{n} x_i/nk$.

Part d

First, we'll take the hint and we'll consider an unbiased estimator of $g(p)$ in the form of

$$\tilde{g(p)} = \begin{cases} 1, \text{if } X_1 = 1 \\ 0, \text{otherwise}. \end{cases} \tag{21}$$

Hence, we can check that

$$\begin{aligned} \mathbb{E}(\tilde{g(p)}) &= 1 \times p(X_1 = 1) + 0 \times p(X_1 \neq 1), && (22) \\ &= p(X_1 = 1), && (23) \\ &= g(p). && (24) \end{aligned}$$

In addition, we that $T = \sum_{i=1}^{n} x_i$ and that consequently $T \sim$ Binomial(nk, p). Now, let's recall the Rao-Blackwell Theorem.

Theorem. (Rao-Blackwell) Let $\hat{\theta}$ be an estimator for θ and let T be a sufficient statistic. Let $\hat{\theta}_1$ be another estimator for θ constructed as $\hat{\theta}_1 = \mathrm{E}(\hat{\theta}|T)$. Then $MSE(\hat{\theta}_1) \leq MSE(\hat{\theta})$. The inequality is strict unless $\hat{\theta}_1 = \hat{\theta}$.

Based on the theorem above, we have an estimator for $g(p)$, which $\widetilde{g(p)}$, and we have a sufficient statistics $T = \sum_{i=1}^{n} x_i$. Still according to the theorem, a natural estimator for $g(p)$ is given by $\mathbb{E}(\widetilde{g(p)}|T)$. We know that

$$\begin{aligned}
\mathbb{E}(\widetilde{g(p)}|T) &= p(X_1 = 1|T = t), && (25)\\
&= \frac{p(T = t|X_1 = 1)p(X_1 = 1)}{P(T = t)}, \text{due to Bayes' theorem} && (26)\\
&= \frac{p(T = t-1)p(X_1 = 1)}{P(T = t)}, && (27)\\
&= \frac{\text{Binomial}(t-1; n-1, p)kp(1-p)^{k-1}}{\text{Binomial}(t; n, p)} && (28)\\
&= \frac{\binom{(n-1)k}{t-1}k}{\binom{nk}{t}}. && (29)
\end{aligned}$$

Along with the Lehmann-Scheffe Theorem, we have that $\mathbb{E}(\widetilde{g(p)}|T)$ is BUE for $g(p)$.