Fall 2013

1. For $p \in (0,1)$ unknown, let $X_0, X_1, \ldots$ be independent identically distributed random variables taking values in $\{0,1\}$ with distribution

$$P(X_i = 1) = 1 - P(X_i = 0) = p,$$

and suppose that

$$T_n = \sum_{i=0}^{n-1} I(X_i = 1, X_{i+1} = 1) \tag{1}$$

is observed.

a) Calculate the mean and variance of T_n.
Solution. Writing $Y_i = I(X_i = 1, X_{i+1} = 1)$, we see that

$$\mathbb{E}T_n = \sum_{i=0}^{n-1} P(Y_i) = np^2,$$

and

$$\begin{aligned}\mathrm{Var}(T_n) &= \sum_{i=0}^{n-1} \mathrm{Var}(Y_i) + \sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j) \\ &= \sum_{i=0}^{n-2} (\mathbb{E}Y_i^2 - (\mathbb{E}Y_i)^2) + 2\sum_{i=0}^{n-1} (\mathbb{E}Y_i Y_{i+1} - \mathbb{E}Y_i \mathbb{E}Y_{i+1}) \\ &= n(p^2 - p^4) + 2(n-1)(p^3 - p^4).\end{aligned}$$

b) Find a consistent method of moments $\hat{p}_n = g_n(T_n)$ estimator for the unknown p as a function g_n of T_n that may depend on n, and prove that your estimate is consistent for p.
Solution. Since $\mathbb{E}T_n = np^2$, we see that

$$\hat{p}_n = \sqrt{\frac{T_n}{n}} = g_n(T_n)$$

is a method of moments estimator for p. Note that

$$\mathbb{E}\left|\frac{T_n}{n} - p^2\right|^2 = \mathrm{Var}\left(\frac{T_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(T_n) = O\left(\frac{1}{n}\right) \to 0$$

as $n \to \infty$. Since L^2 convergence implies convergence in probability, we have that

$$\frac{T_n}{n} \to p^2$$

in probability, and so taking $g(x) = \sqrt{x}$, by the continuous mapping theorem, we have that

$$\hat{p} = \sqrt{\frac{T_n}{n}} \to p$$

in probability. Thus, our moment estimate $\hat{p}$ is consistent for p. □

c) Show that T_n is not the sum of independent, identically distributed random variables. Nevertheless, determine the non-trivial limiting distribution of $\hat{p}_n$, after an appropriate centering and scaling, as if (1) was the sum of iid variables and has the same mean and variance as the one computed in part (a).

Solution. Since $P(Y_0 = 1, Y_1 = 1) = P(X_0 = 1, X_1 = 1, X_2 = 1) = p^3$ and $P(Y_0 = 1)P(Y_1 = 1) = p^4$, we have that Y_i's are not independent. Hence, T_n is not the sum of iid random variables.

Now, if we assumed that T_n is the sum of iid variables, then since $\mathbb{E}Y_i = p^2$ and $\mathrm{Var}(Y_i) = \mathbb{E}Y_i^2 - (\mathbb{E}Y_i)^2 = p^2 - p^4$, and so, by the Central Limit Theorem, we have that

$$\sqrt{n}\left(\frac{T_n}{n} - p^2\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - p^2\right) \Rightarrow \mathcal{N}\left(0, p^2 - p^4\right).$$

Using the Delta method with $g(x) = \sqrt{x}$, we have that

$$\sqrt{n}\,(\hat{p} - p) = \sqrt{n}\left(g\left(\frac{T_n}{n}\right) - g(p^2)\right) \Rightarrow \mathcal{N}\left(0, (p^2 - p^4)\frac{1}{2p}\right) = \mathcal{N}\left(0, \frac{p - p^3}{2}\right).$$

□

d) Explain why you would, or would not, expect $\hat{p}_n$ to have the same limiting distribution as the one determined in part (c).

2. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with density given by

$$f_\beta(x) = \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp(-x/\beta), \text{ for } x > 0,$$

where $\alpha > 0$ and is known. Suppose it is desired to estimate β^3.

a) Find the Cramer-Rao lower bound for the variance of an unbiased estimator of β^3.
Solution. The likelihood function is

$$\mathcal{L}(\beta^3; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{x_i}{\beta}\right) = \frac{\left(\prod_{i=1}^n x_i\right)^{\alpha-1}}{\beta^{\alpha n}\Gamma(\alpha)^n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\beta}\right),$$

and so, the log-likelihood function is

$$\log \mathcal{L}(\beta^3; \boldsymbol{x}) = (\alpha - 1)\sum_{i=1}^{n} \log x_i - \alpha n \log \beta - n \log \Gamma(\alpha) - \frac{1}{\beta}\sum_{i=1}^{n} x_i.$$

Hence,

$$\frac{d}{d\beta^3} \log \mathcal{L} = -\frac{\alpha n}{3\beta^3} + \frac{1}{3}\frac{\sum x_i}{\beta^4},$$

and so,

$$\frac{d^2}{d(\beta^3)^2} \log \mathcal{L} = \frac{\alpha n}{3\beta^6} - \frac{4}{9}\frac{\sum x_i}{\beta^7}.$$

Thus, the Fisher information for β^3 is

$$\mathcal{I}(\beta^3) = -\mathbb{E}\left[\frac{\alpha n}{3\beta^6} - \frac{4}{9}\frac{\sum X_i}{\beta^7}\right] = -\frac{\alpha n}{3\beta^6} + \frac{4}{9}\frac{n\alpha\beta}{\beta^7} = \frac{1}{9}\frac{n\alpha}{\beta^6},$$

and so, the Cramer-Rao lower bound is

$$\mathcal{I}(\beta^3)^{-1} = \boxed{\frac{9\beta^6}{\alpha n}}.$$

b) Find a complete and sufficient statistic for β. Then, compute its kth moment, where k is a positive integer.
Solution. In part (a), we saw that the likelihood function was

$$\mathcal{L}(\beta; \boldsymbol{x}) = \frac{\left(\prod_{i=1}^n x_i\right)^{\alpha-1}}{\beta^{\alpha n}\Gamma(\alpha)^n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\beta}\right),$$

and so, by the Neyman-Fisher factorization theorem, we see that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for β. Also, since X_i's are chosen iid from an exponential family and the set $\{1/\beta : \beta \in \mathbb{R}^+\}$ contains an open set in $\mathbb{R}$, we see that $T(\boldsymbol{X})$ is also complete. Also, since the X_i's are iid, we have that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i \sim \Gamma(n\alpha, \beta)$. Hence, the moment generating function of $T(\boldsymbol{X})$ is

$$M_T(t) = (1 - \beta t)^{-n\alpha},$$

for $t < \beta^{-1}$, and taking multiple derivatives gives us the kth moments of T:

$$\mathbb{E}T^k = \boxed{n\alpha\beta^k(n\alpha + 1)(n\alpha + 2)\cdots(n\alpha + k - 1)}.$$

c) If a UMVUE exists, find its variance and compare it to the bound in part (a).

Solution. From part (b), we see that

$$\mathbb{E}T^3 = n\alpha\beta^3(n\alpha+1)(n\alpha+2).$$

Hence,

$$S(\boldsymbol{X}) = \frac{1}{n\alpha(n\alpha+1)(n\alpha+2)}T^3(\boldsymbol{X})$$

is an UMVUE of β^3 by Lehmann-Scheffe. As for the variance of $S(\boldsymbol{X})$, we have

$$\begin{aligned}
\mathrm{Var}(S(\boldsymbol{X})) &= \frac{1}{n^2\alpha^2(n\alpha+1)^2(n\alpha+2)^2}\left(\mathbb{E}T^6 - (\mathbb{E}T^3)^2\right) \\
&= \frac{\beta^6 n\alpha(n\alpha+1)\cdots(n\alpha+5) - \left(\beta^3 n\alpha(n\alpha+1)(n\alpha+2)\right)^2}{n^2\alpha^2(n\alpha+1)^2(n\alpha+2)^2} \\
&= \beta^6\frac{(n\alpha+3)(n\alpha+4)(n\alpha+5) - n\alpha(n\alpha+1)(n\alpha+2)}{n\alpha(n\alpha+1)(n\alpha+2)} \\
&= \beta^6\frac{9n^2\alpha^2+45n\alpha+60}{n^3\alpha^3+3n^2\alpha^2+2n\alpha} \\
&= \frac{\beta^6}{n\alpha}\frac{9n^2\alpha^2+45n\alpha+60}{n^2\alpha^2+3n\alpha+2} \\
&> \frac{\beta^6}{n\alpha}\frac{9n^2\alpha^2+27n\alpha+18}{n^2\alpha^2+3n\alpha+2} = 9\frac{\beta^6}{n\alpha},
\end{aligned}$$

and so, we see that our UMVUE does not achieve the Cramer-Rao lower bound. □