Statistics 581, Problem Set 3 Solutions

Wellner; 10/13/2017

1. Ferguson, ACILST, page 34, problem 1(a) (modified slightly)
   Suppose that $X_1, X_2, \ldots$ are i.i.d. in R^2 with distribution giving probability θ_1 to $(1,0)'$, probability θ_2 to $(0,1)'$, θ_3 to $(0,0)'$ and θ_4 to $(-1,-1)'$ where $\theta_j \ge 0$ for $j = 1,2,3,4$ and $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 1$.
   (a) Find $\mu = E(X_1)$.
   (b) Compute $E(X_1 X_1^T)$ and $\Sigma = E(X_1 - \mu)(X_1 - \mu)^T$.
   (c) Find the limiting distribution of $\sqrt{n}(\overline{X}_n - \mu)$ and describe the resulting approximation to the distribution of $\overline{X}_n$.
   (d) Find values of $(\theta_1, \ldots, \theta_4)$ such that Σ has rank 1 and $\det(\Sigma) = 0$.

   Solution: (a) The mean of X_1 is given by

   $$E(X_1) = \theta_1(1,0)' + \theta_2(0,1)' + \theta_3(0,0)' + \theta_4(-1,-1)' = (\theta_1 - \theta_4, \theta_2 - \theta_4)\,.$$

   (b) Now

   $$\begin{aligned} E(XX') &= \theta_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \theta_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \theta_3 \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \theta_4 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \theta_1 + \theta_4 & \theta_4 \\ \theta_4 & \theta_2 + \theta_4 \end{pmatrix}. \end{aligned}$$

   Hence it follows that

   $$\begin{aligned} \Sigma &= E(XX') - E(X)E(X') \\ &= \begin{pmatrix} \theta_1 + \theta_4 & \theta_4 \\ \theta_4 & \theta_2 + \theta_4 \end{pmatrix} - \begin{pmatrix} (\theta_1 - \theta_4)^2 & (\theta_1 - \theta_4)(\theta_2 - \theta_4) \\ (\theta_1 - \theta_4)(\theta_2 - \theta_4) & (\theta_2 - \theta_4)^2 \end{pmatrix} \\ &= \begin{pmatrix} \theta_1 + \theta_4 - (\theta_1 - \theta_4)^2 & \theta_4 - (\theta_1 - \theta_4)(\theta_2 - \theta_4) \\ \theta_4 - (\theta_1 - \theta_4)(\theta_2 - \theta_4) & \theta_2 + \theta_4 - (\theta_2 - \theta_4)^2 \end{pmatrix}. \end{aligned}$$

   (c) By the multivariate CLT it follows that

   $$\sqrt{n}(\overline{X}_n - E(X_1)) \to_d N_2(0, \Sigma)\,.$$

   For example, if $\theta_j = 1/4$ for $j = 1,2,3,4$, then $E(X_1) = 0$,

   $$\Sigma = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}$$

   so the variances are both $1/2$ and the correlation is $(1/4)/\sqrt{(1/2)(1/2)} = 1/2$; moreover $\det(\Sigma) = 1/4 - 1/16 = 3/16 > 0$. In this case the resulting normal

approximation to the distribution of $\overline{X}_n$ is centered at 0 with a variance-covariance matrix $n^{-1}\Sigma$ with Σ as in the last display.
(d) Note that Σ does not depend explicitly on θ_3, but only through the constraint that $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 1$. When $\theta_3 = 1$, then $\theta_1 = \theta_2 = \theta_4 = 0$, $P(X_1 = 0) = 1$ and $\Sigma = 0$ has rank 0. Thus we reduce to the case $\theta_3 = 0$. If $\theta_2 = \theta_3 = 0$ and $\theta_1 = a = 1 - \theta_4$, then

$$\Sigma = \begin{pmatrix} 4a(1-a) & 2a(1-a) \\ 2a(1-a) & a(1-a) \end{pmatrix},$$

and $\det(\Sigma) = 0$. Thus for $\underline{\theta} = (a, 0, 0, 1-a)$ with $0 < a < 1$, Σ has rank 1. Similarly Σ has rank 1 for $\theta = (0, a, 0, 1-a)$ with $0 < a < 1$.

2. Suppose that $X_1, X_2, \ldots$ are i.i.d. (μ, σ^2) with $\mu_4 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample mean and sample variance respectively.
(a) Show that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$\begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}.$$

(b) Suppose $\mu \neq 0$. Use (a) to find the limiting distribution of the sample *coefficient of variation* $C_n \equiv S_n/\overline{X}_n$; i.e. show that $\sqrt{n}(C_n - c) \to_d N(0, V^2)$ with $c \equiv \sigma/\mu$ and find V^2.

Solution: (a) Since $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \mu)^2 + o_p(1/\sqrt{n})$, we have

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix} + o_p(1) \\ &\to_d \underline{Z} \sim N_2(0, \Sigma)\end{aligned}$$

by the multivariate CLT where Σ is as given above. Note that with $\underline{Y}_i \equiv (X_i - \mu, (X_i - \mu)^2 - \sigma^2)^T$, to apply the multivariate CLT we only need to verify that $E\|\underline{Y}_1\|^2 < \infty$ and compute

$$\Sigma_Y = E(\underline{Y}_1 \underline{Y}_1^T) = \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}.$$

(b) The function $g(u, v) = \sqrt{v}/u$ is differentiable at points (u, v) with $u \neq 0$, and the derivative is

$$\nabla g(u, v) = (-\sqrt{v}u^{-2}, (1/2)v^{-1/2}u^{-1}) = -\frac{\sqrt{v}u}{(}\frac{1}{u}, -\frac{1}{2v}).$$

so that $g(\mu, \sigma^2) = -c(\mu^{-1}, -(2\sigma^2)^{-1})$. Hence it follows from the delta method (g' theorem) that

$$\begin{aligned} \sqrt{n}(C_n - c) &= \sqrt{n}(g(\overline{X}_n, S_n^2) - g(\mu, \sigma^2)) \\ &\to_d \nabla g \cdot \underline{Z} \sim N(0, \nabla g^T \Sigma \nabla g) \end{aligned}$$

and it is easy to calculate that

$$\nabla g^T \Sigma \nabla g = c^4 \left(1 - c^{-1}\gamma_1 + \frac{1}{4} c^{-2}(2 + \gamma_2)\right)$$

where $\gamma_1 \equiv \mu_3/\sigma^3$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$. Note that when the X_i's are normal (so $\gamma_1 = \gamma_2 = 0$), this reduces to $c^4(1 + c^{-2}/4) = c^4 + c^2/4$. Thus under normality we have

$$\sqrt{n}(g(C_n) - g(c)) \to_d N(0, 1)$$

if $g(x) = \sqrt{2}\left(\log(x) - \log(1 + \sqrt{1 + 2x^2})\right)$.

3. Ferguson, ACILST, page 34, problem 1(b) (modified slightly)
Suppose that $X_1, \ldots, X_n$ is a sample from the Poisson distribution with parameter $\lambda > 0$: $P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \ldots$. Let $Z_n = (1/n)\sum_{i=1}^n 1_{[X_i = 0]}$.
(a) What is the joint asymptotic distribution of

$$\sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, e^{-\lambda})')?$$

(b) Let $p_0(\lambda) \equiv P_\lambda(X_1 = 0)$. What is the asymptotic distribution of $\hat{p}_0 \equiv p_0(\hat{\lambda}_n)$ where $\hat{\lambda}_n = \overline{X}_n$?
(c) What is the joint asymptotic distribution of $(Z_n, \hat{p}_0)$ (after centering and rescaling)?
(d) Compute the ratio of the asymptotic variances of the two estimators Z_n and $\hat{p}_0$ of $p_0(\lambda)$. Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

Solution: (a). Let $W_i \equiv (X_i, Y_i) \equiv (X_i, 1_{[X_i = 1]})$. Then the W_i's are i.i.d. with mean $E(W_1) = (\lambda, e^{-\lambda})'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & e^{-\lambda}(1 - e^{-\lambda}) \end{pmatrix}. \tag{1}$$

Hence the multivariate CLT implies that

$$\sqrt{n}(\overline{W} - E(W_1)) = \sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, e^{-\lambda})) \to_d T \sim N_2(0, \Sigma) \tag{2}$$

where Σ is given in (1).

(b). Now $\hat{p}_0 = g(\overline{X}_n)$ where $g(v) = e^{-v}$. Hence $g'(v) = -e^{-v}$, $g'(\lambda) = -e^{-\lambda}$, and $\sqrt{n}(\overline{X}_n - \lambda) \to_d N(0, \lambda)$ by the CLT (or the first component of the convergence in distribution in part (a)). Hence it follows from the delta-method that

$$\sqrt{n}(\hat{p}_0 - p_0(\lambda)) = \sqrt{n}(g(\overline{X}_n) - g(\lambda)) \to_d g'(\lambda)N(0, \lambda) = N(0, \lambda e^{-2\lambda}).$$

(c). At this point it is a bit easier to study $(\hat{p}_0, Z_n) = g(\overline{X}_n, Z_n)$ where $g(u,v) \equiv (e^{-u}, v)$. Then in view of (2) and

$$\nabla g(\lambda, e^{-\lambda}) = \begin{pmatrix} -e^{-\lambda} & 0 \\ 0 & 1 \end{pmatrix},$$

it follows from the delta-method that

$$\sqrt{n}((\hat{p}_0, Z_n)' - e^{-\lambda}(1,1)') \to_d \nabla g(\lambda, e^{-\lambda}) T \sim N_2(0, \nabla g \Sigma (\nabla g)')$$

where

$$\nabla g \Sigma (\nabla g)' = \begin{pmatrix} \lambda e^{-2\lambda} & \lambda e^{-2\lambda} \\ \lambda e^{-2\lambda} & e^{-\lambda}(1 - e^{-\lambda}) \end{pmatrix}.$$

This is a situation in which we have two estimators of $P_\lambda(X_1 = 0) = p_0(\lambda)$, namely the MLE $\hat{p}_1 = p_0(\hat{\lambda})$ and the empirical (or "plug-in" estimator $Z_n = \#\{i \le n : X_i = 0\}/n$. Note that the ratio of the asymptotic variance of $\hat{p}_0$ to the asymptotic variance of Z_n is

$$ARE(\hat{p}_0, Z_n) \equiv \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1 - e^{-\lambda})} = \frac{\lambda e^{-\lambda}}{(1 - e^{-\lambda})} < 1$$

for all $\lambda > 0$. See the figure below

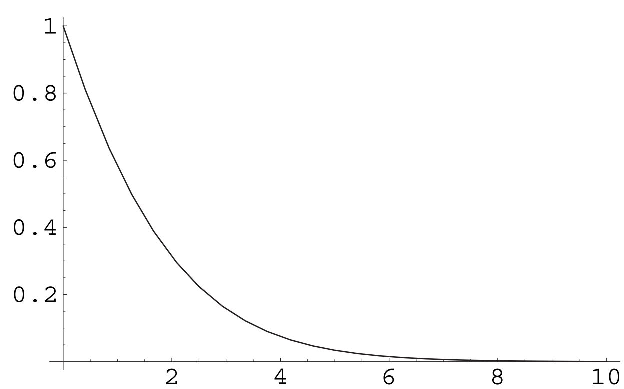

Figure 1: ARE of MLE relative to Plug-In.

4. A sequence of random variables Y_n is *bounded in probability* and we write $Y_n = O_p(1)$ if

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P(|Y_n| > \lambda) = 0;$$

i.e. for each $\epsilon > 0$ there exist λ_ϵ and N_ϵ such that $P(|Y_n| > \lambda_\epsilon) < \epsilon$ for all $n > N_\epsilon$. (a) Show that if $Y_n \to_d Y$ for some random variable Y, then Y_n is bounded in

probability. (This is Lehmann and Casella, problem 8.24, page 77.)
(b) Give an example of a sequence of random variables Y_n that is bounded in probability, but does not converge in distribution.
(c) Let $X_n \sim t_n$, the $t-$distribution with n degrees of freedom; thus $X_n \stackrel{d}{=} Z/\sqrt{Y_n/n}$ where $Y_n \sim \chi^2_n$ is independent of Z. Fix $r > 0$ (large). Compute $E|X_n|^r$ exactly as a function of n and r in terms of the Gamma function $\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt$. Use Stirling's formula, $\Gamma(r+1) \sim \sqrt{2\pi r}(r/e)^r$ to show that $\limsup_{n\to\infty} E|X_n|^r < \infty$. Use this to show that $\{|X_n|^p : \ n \geq 1\}$ is uniformly integrable for any integer $p \geq 1$, and hence that $E(X_n^p) \to E(Z^p)$.

Solution: (a) Let F denote the limiting distribution, and let $Y \sim F$. Fix $\epsilon > 0$. Choose $\pm M \in C_F$ so large that $P(|Y| > M) \leq 1 - F(M) + F(-M) < \epsilon/2$. Since $\pm M \in C_F$, and $Y_n \sim F_n$ converge in distribution to $Y \sim F$, we can find an $N = N_{M,\epsilon}$ so that $|F_n(\pm M) - F(\pm M)| < \epsilon/4$. Then for $n \geq N$ it follows that

$$\begin{aligned}
P(|Y_n| > M) &= 1 - F_n(M) + F_n(-M-) \\
&\leq 1 - F(M) + F(-M) + F(M) - F_n(M) + F_n(-M) - F(-M) \\
&< \frac{1}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon;
\end{aligned}$$

i.e. Y_n is bounded in probability.
(b) Suppose that $Y_n = (-1)^n Y$ where $Y \sim$ Exponential(1). Then $|Y_n| \leq |Y| = O_p(1)$, but $Y_{2n} \to Y$ while $Y_{2n+1} \to -Y$, and therefore $Y_n \not\to_d$.
(c) By independence of Z and Y_n it follows that

$$E|X_n|^r = E|Z|^r E(n^{r/2}Y_n^{-r/2}) \tag{3}$$

where

$$\begin{aligned}
E(n^{r/2}Y_n^{-r/2}) &= \frac{n^{r/2}}{\Gamma(n/2)2^{n/2}} \int_0^\infty y^{-r/2}y^{(n/2)-1}e^{-y/2}dy \\
&= \frac{n^{r/2}}{\Gamma(n/2)2^{n/2}} 2^{(n-r)/2} \int_0^\infty w^{(n-r)/2-1}e^{-w}dw \\
&= \frac{(n/2)^{r/2}\Gamma((n-r)/2)}{\Gamma(n/2)} \qquad (4)
\end{aligned}$$

Now we proceed via Stirling's formula: $\Gamma(x+1) \sim \sqrt{2\pi x}(x/e)^x$ as $x \to \infty$. The left side of (4) is asymptotically equivalent (in the sense that the ratio converges to 1) to

$$\begin{aligned}
&\frac{(n/2)^{r/2}\sqrt{2\pi(\frac{n-r}{2}-1)}\left(\frac{\frac{n-r}{2}-1}{e}\right)^{\frac{n-r}{2}-1}}{\sqrt{2\pi(\frac{n}{2}-1)}\left(\frac{\frac{n}{2}-1}{e}\right)^{\frac{n}{2}-1}} \\
&= (n/2)^{r/2}(1+o(1))\frac{\left(\frac{\frac{n-r}{2}-1}{e}\right)^{\frac{n-r}{2}-1}}{\left(\frac{\frac{n}{2}-1}{e}\right)^{\frac{n}{2}-1}} \\
&= (1+o(1))e^{r/2}\left(\frac{n}{n-r-2}\right)^{r/2}\frac{\left(\frac{n-r}{2}-1\right)^{\frac{n}{2}-1}}{\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1}} \\
&= (1+o(1))e^{r/2}\cdot(1+o(1))\left(1-\frac{(r/2)}{(n/2)-1}\right)^{\frac{n}{2}-1} \\
&\to 1. \qquad (5)
\end{aligned}$$

Putting this together with (3) and (4) it follows that

$$\limsup_{n\to\infty} E|X_n|^r = E|Z|^r.$$

Now $X_n \to_d Z$, so it follows from Skorokhod's theorem, Proposition 2.4.3, that there exist random variable X_n^* and Z^* on a common probability space satisfying $X_n^* \stackrel{d}{=} X_n$, $Z^* \stackrel{d}{=} Z$, and $X_n^* \to_{a.s.} Z^*$. Thus it follows that $E|X_n^*|^r \to E|Z^*|^r$. By Vitali's theorem (Theorem 1.1.2), it follows that $\{|X_n^*|^r : \ n \geq 1\}$ is uniformly integrable. By Vitali's theorem again it follows that $E(X_n^p) = E((X_n^*)^p) \to E((Z^*)^p) = E(Z^p)$ for each $p \leq r$.

This problem can also be solved in a way that does not lead to quite as precise a result as is given in (5). In this solution we simply choose and fix some very large value of r, say $r = \lceil 2p \rceil$. Then we simply show that

$$\begin{aligned}
\frac{(n/2)^{r/2}\sqrt{2\pi(\frac{n-r}{2}-1)}\left(\frac{\frac{n-r}{2}-1}{e}\right)^{\frac{n-r}{2}-1}}{\sqrt{2\pi(\frac{n}{2}-1)}\left(\frac{\frac{n}{2}-1}{e}\right)^{\frac{n}{2}-1}} &= (n/2)^{r/2}(1+o(1))\frac{\left(\frac{\frac{n-r}{2}-1}{e}\right)^{\frac{n-r}{2}-1}}{\left(\frac{\frac{n}{2}-1}{e}\right)^{\frac{n}{2}-1}} \\
&= (1+o(1))e^{r/2}\frac{\left(\frac{n}{2}\right)^{r/2}\left(\frac{n-r}{2}-1\right)^{\frac{n-r}{2}-1}}{\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1}} \\
&\leq (1+o(1))e^{r/2}
\end{aligned}$$

since

$$
\begin{aligned}
\frac{\left(\frac{n}{2}\right)^{r/2}\left(\frac{n-r}{2}-1\right)^{\frac{n-r}{2}-1}}{\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1}} &\le \frac{\left(\frac{n}{2}\right)^{r/2}\left(\frac{n}{2}-1\right)^{\frac{n-r}{2}-1}}{\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1}} \\
&= \frac{\left(\frac{n}{2}\right)^{r/2}}{\left(\frac{n}{2}-1\right)^{r/2}} \cdot \frac{\left(\frac{n}{2}-1\right)^{r/2}\cdot\left(\frac{n}{2}-1\right)^{\frac{n-r}{2}-1}}{\left(\frac{n}{2}-1\right)^{\frac{n}{2}-1}} \\
&= 1+o(1).
\end{aligned}
$$

Combining this with (3) and (4) it follows that $\limsup_{n\to\infty} E|X_n|^r \le e^{r/2}$. Thus it follows that $\{|X_n|^s : \; n \ge 1\}$ is uniformly integrable for each $p \le s < r$. The rest of the proof is completed in the same way as before.

5. Suppose that X is a random variable with finite fourth moment; $E|X|^4 < \infty$. Then $\mu_4 = E(X-\mu)^4$ is the fourth central moment of X. The ratio $\mu_4/\sigma^4 \equiv \kappa$ is the *kurtosis* of X (or of the distribution function F of X), and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*; note that for any $N(\mu,\sigma^2)$ random variable, $\gamma_2 = 0$. Investigate the value of γ_2 for various classical distributions (t_r, uniform, bernoulli, Poission(λ), ...). How big can γ_2 be? How small can γ_2 be?

   Solution: Note that $\mu_4^{1/4} = \{E(X-\mu)^4\}^{1/4} \ge \{E(X-\mu)^2\}^{1/2} = \sigma$ by Liapunov's inequality. Thus $\mu_4/\sigma^4 \ge 1$ always, or $\gamma_2 \equiv \mu_4/\sigma^4 \ge -2$ with equality if $X = \pm 1$ with probability 1/2 each: then $\mu = 0$, $\sigma^2 = 1$, $\mu_4 = 1$, and $\gamma_2 = -2$.
   For $X \sim N(0,1)$, $\gamma_2 = 0$ since $EX^4 = 3$.
   For $X \sim t_r$, $r > 4$, $\gamma_2 = 6/(r-4) \nearrow \infty$ as $r \searrow 4$; $\gamma_2 \searrow 0$ as $r \nearrow \infty$.
   For $X \sim \text{Gamma}(\alpha,\beta)$, $\gamma_2 = 6/\alpha \nearrow \infty$ as $\alpha \searrow 0$.
   For $X \sim \text{Poisson}(\lambda)$, $\gamma_2 = 1/\lambda \nearrow \infty$ as $\lambda \searrow 0$.
   For $X \sim \text{Bernoulli}(p)$, $\gamma_2 = (1-p)^2/p + p^2/(1-p) - 3$ which $= -2$ when $p = 1/2$, and $\nearrow \infty$ when $p \to 0, 1$.