Statistics 582, Problem Set 1 Solutions

Wellner; 1/14/2009

1. Suppose that $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. with bivariate normal distribution $N_2(\mu, \Sigma)$ where $\mu \in R^2$ and

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{pmatrix}$$

where $\sigma^2 > 0$, $\tau^2 > 0$, and $\rho \in (-1, 1)$.
(a) If we assume that $\mu_1 = \mu_2 \equiv \theta$ and Σ is known, what is the MLE of θ?
(b) If we assume that μ is known and $\sigma^2 = \tau^2 \equiv \theta$, what is the MLE of θ?
(c) What is the asymptotic distribution of the estimator you found in (b)?
(d) Under the same assumption as in (b), what is the MLE of ρ?
(e) What is the asymptotic distribution of the estimator you found in (d)?

Solution: (a) When $\mu_1 = \mu_2 = \theta$ and Σ is known, then the log-likelihood for one observation is (relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\log p(x; \theta) = -\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\theta)^2}{\sigma^2} - 2\rho \frac{(x-\theta)}{\sigma} \frac{(y-\theta)}{\tau} + \frac{(y-\theta)^2}{\tau^2} \right\} + constant.$$

Hence the score for θ for one observation is

$$\begin{aligned} \dot{l}_\theta(x, y) &= \frac{1}{1-\rho^2} \left\{ \frac{(x-\theta)}{\sigma^2} + \frac{(y-\theta)}{\tau^2} - \frac{\rho}{\tau} \frac{(x-\theta)}{\sigma} - \frac{\rho}{\sigma} \frac{(y-\theta)}{\tau} \right\} \\ &= \frac{1}{1-\rho^2} \left\{ \frac{(x-\theta)}{\sigma} \left(\frac{1}{\sigma} - \frac{\rho}{\tau} \right) + \frac{(y-\theta)}{\tau} \left(\frac{1}{\tau} - \frac{\rho}{\sigma} \right) \right\}. \end{aligned}$$

Thus the score equation for θ is

$$0 = \dot{l}_{n\theta}(\theta) = \frac{n}{1-\rho^2} \left\{ \frac{(\overline{X}_n - \theta)}{\sigma} \left(\frac{1}{\sigma} - \frac{\rho}{\tau} \right) + \frac{(\overline{Y}_n - \theta)}{\tau} \left(\frac{1}{\tau} - \frac{\rho}{\sigma} \right) \right\},$$

and hence

$$\begin{aligned} \widehat{\theta}_n &= \frac{\overline{X}_n \left(\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau} \right) + \overline{Y}_n \left(\frac{1}{\tau^2} - \frac{\rho}{\sigma\tau} \right)}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}} \\ &= a\overline{X}_n + (1-a)\overline{Y}_n \end{aligned}$$

where

$$a = \frac{\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau}}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}}.$$

Note that this yields

$$Var(\widehat{\theta}) = \frac{1}{n} \left\{ a^2\sigma^2 + 2a(1-a)\rho\sigma\tau + (1-a)^2\tau^2 \right\}.$$

(b) and (d) If $\sigma^2 = \tau^2 = \theta$ and μ is known, then the log-likelihood for one observation is (again relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\begin{aligned}\log p(x;\theta,\rho) &= -\log\theta - \frac{1}{2}\log(1-\rho^2) \\ &\qquad - \frac{1}{2(1-\rho^2)\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} + \text{constant}.\end{aligned}$$

Thus the scores for θ and ρ are given by

$$\dot{l}_\theta(x,y) = -\frac{1}{\theta} + \frac{1}{2(1-\rho^2)\theta^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\},$$

$$\begin{aligned}\dot{l}_\rho(x,y) &= \frac{\rho}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\qquad + \frac{1}{\theta(1-\rho^2)}(x-\mu)(y-\nu).\end{aligned}$$

Hence the score equations for estimation of θ and ρ are given by

$$0 = \dot{l}_{n\theta}(\theta) = \sum_{i=1}^n \dot{l}_\theta(X_i,Y_i) = -\frac{n}{\theta} + \frac{n}{\theta^2 2(1-\rho^2)}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\},$$

and

$$0 = \dot{l}_{n\rho}(\rho) = \sum_{i=1}^n \dot{l}_\rho(X_i,Y_i) = \frac{n\rho}{1-\rho^2} - \frac{n\rho}{\theta(1-\rho^2)^2}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + \frac{n}{\theta(1-\rho^2)}S_{XY}$$

where

$$S_{XX} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)^2, \quad S_{XY} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)(Y_i-\nu), \quad S_{YY} \equiv n^{-1}\sum_{i=1}^n (Y_i-\nu)^2.$$

Solving the first of these for $\widehat{\theta}$ yields

$$\widehat{\theta} = \frac{1}{2(1-\widehat{\rho}^2)}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\};$$

Rewriting the score equation for ρ with a common denominator of $\theta(1-\rho^2)^2$ yields

$$\theta\rho(1-\rho^2) - \rho\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + (1-\rho^2)S_{XY} = 0;$$

and then plugging in the estimator $\widehat{\theta}$ of θ yields the equation

$$(1-\widehat{\rho}^2)S_{XY} = \frac{1}{2}\widehat{\rho}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\}.$$

This has the solution

$$\widehat{\rho} = \frac{2S_{XY}}{S_{XX}+S_{YY}};$$

plugging this (or more precisely the last form of the equation for $\widehat{\rho}$) into the expression for θ yields $\widehat{\theta} = (S_{XX} + S_{YY})/2$.
(c) and (e) To find the asymptotic distributions of $\widehat{\theta}$ and $\widehat{\rho}$ we could either (i) proceed directly from first principles (central limit theorms and the delta method), or (b) use theorem 4.1.5 concerning the asymptotic behavior of maximum likelikhood estimators. I'll take the second route here. The first step in this direction is to compute the information matrix for (θ, ρ). Now

$$\ddot{l}_{\theta\theta}(x,y) = \frac{1}{\theta^2} - \frac{1}{(1-\rho^2)\theta^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} ,$$

$$\begin{aligned}\ddot{l}_{\theta\rho}(x,y) &= \frac{2\rho}{2\theta^2(1-\rho^2)^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad - \frac{1}{2\theta^2(1-\rho^2)^2}2(x-\mu)(y-\nu) ,\end{aligned}$$

and

$$\begin{aligned}\ddot{l}_{\rho\rho}(x,y) &= \frac{1}{1-\rho^2} + \frac{2\rho^2}{(1-\rho^2)^2} \\ &\quad - \left\{\frac{1}{(1-\rho^2)^2} + \frac{4\rho^2}{(1-\rho^2)^3}\right\}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{\rho}{\theta(1-\rho^2)^2}2(x-\mu)(y-\nu) + \frac{2\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu) \\ &= \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{1-3\rho^2}{\theta(1-\rho^2)^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{4\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu) .\end{aligned}$$

Here

$$E\left\{(X-\mu)^2 - 2\rho(X-\mu)(Y-\nu) + (Y-\nu)^2\right\} = 2\theta(1-\rho^2)$$

and

$$E(X-\mu)(Y-\nu) = \rho\theta.$$

Thus we find that

$$I_{\theta\theta} = E(-\ddot{l}_{\theta\theta}(X,Y)) = \theta^{-2} ,$$

$$I_{\theta\rho} = E(-\ddot{l}_{\theta\rho}(X,Y)) = \frac{\rho}{\theta(1-\rho^2)} ,$$

and

$$I_{\rho\rho} = E(-\ddot{l}_{\rho\rho}(X,Y)) = \frac{1+\rho^2}{(1-\rho^2)^2} .$$

This yields

$$I_{\theta\theta\cdot\rho} = I_{\theta\theta} - I_{\theta\rho}I_{\rho\rho}^{-1}I_{\rho\theta} = \frac{1}{\theta^2}\frac{1}{1+\rho^2}$$

and

$$I_{\rho\rho\cdot\theta} = I_{\rho\rho} - I_{\rho\theta}I_{\theta\theta}^{-1}I_{\theta\rho} = (1-\rho^2)^{-2} .$$

Hence it follows from theorem 4.1.5 that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow_d N(0, I^{-1}_{\theta\theta\cdot\rho}) = N(0, \theta^2(1+\rho^2))$$

while

$$\sqrt{n}(\widehat{\rho}_n - \rho) \rightarrow_d N(0, I^{-1}_{\rho\rho\cdot\theta}) = N(0, (1-\rho^2)^2)\,.$$

2. Problem 1, page 117, Ferguson, ACILST. What happens if $\Theta = [1, \infty)$ or $(0, \infty)$?

Solution: (i) Ferguson's problem: here is a verification of the five conditions (a)-(e) when $\Theta = [1, 2]$:
(a) $\Theta = [1, 2]$ is closed and bounded, hence compact.
(b) For fixed $x \le 1$, $p(x, \theta) = 1/\theta$ is continuous; for $1 < x \le 2$, $p(x, \theta) = \theta^{-1}1\{\theta \ge x$, which is upper semi-continuous. For $x > 2$, $p(x, \theta) = 0$ which is a continuous function of θ. Since log is a continuous function on $(0, \infty)$, these (semi-)continuities carry over to $f(x, \theta) = \log p(x, \theta) - \log p(x, \theta_0)$.
(c) Fix $\theta_0 \in \Theta$. Then

$$\frac{p(x,\theta)}{p(x,\theta_0)} = \frac{\theta_0}{\theta}\frac{1_{[0,\theta]}(x)}{1_{[0,\theta_0]}(x)} = \begin{cases} (\theta_0/\theta), & x \le 1 \\ (\theta_0/\theta)1_{[x,\infty)}(\theta), & 1 < x \le \theta_0 \\ \infty, & x > \theta_0, \end{cases}$$

so

$$\sup_{\theta\in\Theta}\frac{p(x,\theta)}{p(x,\theta_0)} \equiv K(x) = \theta_0 1\{x \le 1\} + \frac{\theta_0}{x}1_{(1,\theta_0]}(x) + \infty 1_{(\theta_0,\infty)}(x),$$

and

$$\sup_{\theta\in\Theta}\log\frac{p(x,\theta)}{p(x,\theta_0)} \equiv F(x) = (\log\theta_0)1\{x \le 1\} + \log(\theta_0/x)1_{(1,\theta_0]}(x) + \infty 1_{(\theta_0,\infty)}(x),$$

satisfies $E_{\theta_0}F(X) < \infty$.
(d) The function $\varphi(x, \theta, \rho)$ is given by

$$\varphi(x,\theta,\rho) = \sup_{\theta':|\theta'-\theta|<\rho}\frac{1}{\theta'}1_{[0,\theta']}(x) = \begin{cases} 1/(\theta-\rho), & x < \theta-\rho \\ 1/x, & |x-\theta| \le \rho \\ 0, & x > \theta+\rho, \end{cases}$$

which is clearly measurable.
(e) If $p(x, \theta') = p(x, \theta)$ a.e. Lebesgue, then

$$0 = \frac{1}{2}\int |p(x,\theta) - p(x,\theta')|dx = d_{TV}(P_\theta, P_{\theta'}) = 1 - \eta(P_\theta, P_{\theta'}) = 1 - \frac{\theta'\wedge\theta}{\theta'\vee\theta}$$

where the last equality follows by a computation as in the final exam for stat 581, Fall '05. This yields $\theta' \vee \theta = \theta' \wedge \theta$, which implies $\theta = \theta'$.
(ii) When $\Theta = [1, \infty)$ or $\Theta = (0, \infty)$, then Θ is no longer compact, and Wald's theorem does not apply directly. One way to remedy the problem is to compactify the set Θ by some appropriate identification of points in Θ with points on the unit half-circle (as shown in class on 1/9/06). Another way to proceed is to show that the MLE is in a compact set eventually with probability 1; see e.g. van der Vaart's

re-working of Wald's theorem. In this case the MLE is $\widehat{\theta}_n = \max_{1\le i\le n} X_i = X_{(n)}$, and it follows that for any $\theta_0 > \delta > 0$

$$P_{\theta_0}(\theta_0 - \delta > \widehat{\theta}_n) = \left(\frac{\theta_0 - \delta}{\theta_0}\right)^n,$$

which has a finite sum on n, and hence by the Borel-Cantelli lemma

$$P_{\theta_0}(\widehat{\theta}_n < \theta_0 - \delta \ \text{ infinitely often}) = 0,$$

or, equivalently,

$$P_{\theta_0}(\theta_0 \ge \widehat{\theta}_n \ge \theta_0 - \delta \ \text{ almost always}) = 1.$$

Note that this argument already yields almost sure consistency of the MLE in this case since δ can be chosen to be arbitrarily small.

3. Consider the model introduced in Ferguson, ACILST, problem 17.2, page 117. Show that Theorem 4.3, page 28, of the Chapter 4 notes (or Theorem 17, Ferguson, ACILST, page 114) applies to the MLE of θ in this model.

   Solution: We proceed to verify the conditions (a) - (e) of Theorem 4.3 of the course notes. First (a) holds since $[0,1]$ is compact. (b) holds since for $0 < x < 1$ the function $\theta \mapsto p(x,\theta)$ is continuous (and hence upper-semicontinuous), while for $x = 0$ the function $\theta \mapsto p(0,\theta) = 1_{\{0\}}(\theta)$ is upper-semicontinuous, and similarly for $x = 1$ the function $\theta \mapsto p(1,\theta) = 1_{\{1\}}(\theta)$ is also upper-semicontinuous. To see that (c) holds, consider the two cases $\theta > \theta_0$ and $\theta \le \theta_0$. When $\theta > \theta_0$,

$$\begin{aligned}
f(x,\theta) &= \log p(x,\theta) - \log p(x,\theta_0) \\
&= \begin{cases} \log(\theta_0/\theta) & 0 \le x \le \theta_0 < \theta \\ \log(x/(1-x)) - \log(\theta/(1-\theta_0)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-\theta)) & \theta < x \le 1 \end{cases} \\
&\le \begin{cases} \log(\theta_0/x) & 0 \le x \le \theta_0 < \theta \\ \log((1-\theta_0)/(1-x)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-x)) & \theta < x \le 1 \end{cases} \\
&= 1_{[0,\theta_0]}(x)\log(\theta_0/x) + 1_{(\theta_0,1]}(x)\log((1-\theta_0)/(1-x)) \\
&\equiv F(x).
\end{aligned}$$

   Here for the middle term we used

$$\begin{aligned}
\log x - \log(1-x) + \log(1-\theta_0) + \log(1/\theta) &\le \log x - \log(1-x) + \log(1-\theta_0) + \log(1/x) \\
&= \log((1-\theta_0)/(1-x))
\end{aligned}$$

   since $1/\theta \le 1/x$ on the set $\theta_0 < x \le \theta$. Note that

$$E_{\theta_0}F(X) = 2\int_0^{\theta_0} \frac{x}{\theta_0}\log\left(\frac{\theta_0}{x}\right)dx + 2\int_{\theta_0}^1 \frac{1-x}{1-\theta_0}\log\left(\frac{1-\theta_0}{1-x}\right)dx < \infty.$$

Similarly, when $\theta \le \theta_0$,

$$\begin{aligned}
f(x,\theta) &= \log p(x,\theta) - \log p(x,\theta_0) \\
&= \begin{cases} \log(\theta_0/\theta) & 0 \le x \le \theta \le \theta_0 \\ \log((1-x)/x) - \log((1-\theta)/\theta_0) & \theta < x \le \theta_0 \\ \log((1-\theta_0)/(1-\theta)) & \theta_0 < x \le 1 \end{cases} \\
&\le \begin{cases} \log(\theta_0/x) & 0 \le x \le \theta_0 < \theta \\ \log((1-\theta_0)/(1-x)) & \theta_0 < x \le \theta \\ \log((1-\theta_0)/(1-x)) & \theta \le \theta_0 < x \le 1 \end{cases} \\
&= 1_{[0,\theta_0]}(x)\log(\theta_0/x) + 1_{(\theta_0,1]}(x)\log((1-\theta_0)/(1-x)) \\
&= F(x),
\end{aligned}$$

so the same envelope function works in this case. To verify (d) note that $\theta \mapsto p(x,\theta)$ is a continuous function of θ for $0 < x < 1$, and two indicator functions of θ when $x \in \{0,1\}$, and hence the supremum involved in the condition is measurable. Finally, the identifiability condition (e) holds easily: note that

$$\begin{aligned}
\rho(P_{\theta_0}, P_\theta) &= \int_0^1 \sqrt{p_{\theta_0}(x)p_\theta(x)}dx \\
&= \frac{\theta^2}{\sqrt{\theta\theta_0}} + \frac{2}{\sqrt{\theta_0(1-\theta)}}\int_\theta^{\theta_0}\sqrt{x(1-x)}dx + \frac{(1-\theta_0)^2}{\sqrt{(1-\theta_0)(1-\theta)}} \\
&= 1 \quad \text{ if and only if } \quad \theta = \theta_0.
\end{aligned}$$

Thus $p_{\theta_0}(x) = p_\theta(x)$ a.e. Lebesgue implies that $\theta = \theta_0$. Thus the conditions (a) - (e) all hold and we conclude that the MLE $\widehat{\theta}_n$ of θ is (almost surely) consistent: $\widehat{\theta}_n \to_{a.s.} \theta$.

4. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Weibull(α_0, β_0) (if X has the Weibull(θ) distribution where $\theta = (\alpha, \beta)$, then $1 - F_\theta(x) = P_\theta(X > x) = \exp(-(x/\alpha)^\beta)$ for $x \ge 0$). Recall that the MLE $\widehat{\alpha}$ of α is given by

$$\widehat{\alpha} = \left\{\frac{1}{n}\sum_{i=1}^n X_i^{\widehat{\beta}}\right\}^{1/\widehat{\beta}}$$

where $\widehat{\beta}$ is the MLE of β. As a simpler alternative to maximum likelihood, I propose to use the alternative estimator $\overline{\beta}_n$ of β obtained from the slope of an ordinary least squares fit of a Weibull Q-Q plot, and then estimate α by

$$\overline{\alpha}_n = \left\{\frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n}\right\}^{1/\overline{\beta}_n}.$$

(a) Suppose that $\overline{\beta}_n \to_p \beta_0$ is known. Show that $\overline{\alpha}_n \to_p \alpha_0$. [Hint: use a uniform strong law of large numbers.]
(b) Show that $\overline{\alpha}_n$ is a "pseudo-MLE" in the sense that $\overline{\alpha}_n$ maximizes $l_n(\alpha, \overline{\beta}_n)$.

Solution: Fix $\delta > 0$ (small). The family of functions $\mathcal{F} = \{f(x,\beta) = x^\beta : \beta \in [\beta_0-\delta, \beta_0+\delta]\}$ are indexed by the compact set $[\beta_0-\delta, \beta_0+\delta]$, are continuous in β for every $x \ge 0$, and are bounded by

$$\sup_{\beta\in[\beta_0-\delta,\beta_0+\delta]} |f(x,\beta)| = x^{\beta_0+\delta} \vee x^{\beta_0-\delta} \le x^{\beta_0+\delta} + x^{\beta_0-\delta} \equiv F(x)$$

which satisfies $E_0 F(X) < \infty$ if $\delta < 2\beta_0$. Thus by theorem 4.4.1 (of the section 4 revision) the uniform strong law of large numbers holds for $\mathcal{F}$:

$$\sup_{\beta:|\beta-\beta_0|\le\delta} |\mathbb{P}_n f(\cdot,\beta) - P_0 f(\cdot,\beta)| \rightarrow_{a.s.} 0\,.$$

If $\overline{\beta}_n \rightarrow_{a.s.} \beta_0$, $\overline{\beta}_n \in [\beta_0-\delta, \beta_0+\delta]$, with probability 1 for n sufficiently large, and it follows from the uniform strong law law of large numbers (Theorem 1, section 4.4 revision) together with continuity of $\mu(\beta) \equiv E_0 f(X,\beta)$ that

$$\begin{aligned} \overline{\alpha}_n^{\overline{\beta}_n} &= \frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n} \\ &\rightarrow_{a.s.} E_0 f(X,\beta_0) = \alpha_0^{\beta_0}\,. \end{aligned}$$

(If instead $\overline{\beta}_n \rightarrow_p \beta_0$, then for and given $\epsilon > 0$ and $n \ge N_{\epsilon,\delta}$ large, $P_{\theta_0}(\overline{\beta}_n \in [\beta_0-\delta, \beta_0+\delta]) > 1-\epsilon$ and we can simply argue on this set.) But now

$$\overline{\alpha}_n = \{\overline{\alpha}_n^{\overline{\beta}_n}\}^{1/\overline{\beta_n}} = g(\overline{\alpha}^{\overline{\beta}_n}, \overline{\beta_n})$$

where $g(u,v) \equiv u^{1/v}$ is continuous and $(\overline{\alpha}^{\overline{\beta}_n}, \overline{\beta_n}) \rightarrow_{a.s.} (\alpha_0^{\beta_0}, \beta_0)$. Hence by the continuous mapping theorem

$$\overline{\alpha}_n = g(\overline{\alpha}_n^{\overline{\beta}_n}, \overline{\beta_n}) \rightarrow_{a.s.} g(\alpha_0^{\beta_0}, \beta_0) = \alpha_0\,.$$

B. The log-likelihood is

$$l_n(\alpha,\beta) = n\log(\beta/\alpha) + (\beta-1)\sum_{i=1}^n \log(X_i/\alpha) - \sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^\beta,$$

and hence

$$\begin{aligned} l_n(\alpha,\overline{\beta}_n) &= n\log(\overline{\beta}_n/\alpha) + (\overline{\beta}_n - 1)\sum_{i=1}^n \log(X_i/\overline{\alpha}) - \sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^{\overline{\beta}_n} \\ &= -n\overline{\beta}_n \log\alpha - \frac{\sum X_i^{\overline{\beta}_n}}{\alpha^{\overline{\beta}_n}} + \text{constant in } \alpha \\ &= -n\log\eta - \frac{\sum X_i^{\overline{\beta}_n}}{\eta} + \text{constant in } \alpha \text{ and } \eta \end{aligned}$$

where $\eta \equiv \alpha^{\overline{\beta}_n}$. This is easily seen to be maximized by

$$\overline{\eta} \equiv \frac{1}{n}\sum_{i=1}^n X_i^{\overline{\beta}_n}$$

and hence

$$\overline{\alpha}_n = \left\{ \frac{1}{n} \sum_{i=1}^n X_i^{\overline{\beta}_n} \right\}^{1/\overline{\beta}_n}$$

as claimed. Thus $\overline{\alpha}_n$ is a pseudo-MLE of α.

5. (a) Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution P on R. Consider generalizing the result of the handout in class on 1/7/2009: if $V_n(r)$ is defined for $1 \le r \le 2$ by

$$V_n(r) \equiv \frac{1}{n} \sum_{i=1}^n |X_i - \overline{X}_n|^r.$$

If $E|X|^r < \infty$, show that

$$V_n(r) \to_{a.s.} v(r)$$

where $v(r) \equiv E|X_1 - \mu|^r$.
(b) Now suppose we generalize the problem considered in (a) by considering $X_1, \ldots, X_n$ i.i.d. P on R^d. Let $\| \cdot \|$ be the usual Euclidean metric in R^d, and consider

$$V_n(r) \equiv \frac{1}{n} \sum_{i=1}^n \|X_i - \overline{X}_n\|^r$$

for $1 \le r \le 2$ where $\overline{X}_n$ is the (multivariate) sample mean of the X_i's. Can the same method be used to show that $V_n(r) \to_{a.s.} v(r)$ where $v(r) \equiv E\|X_1 - \mu\|^r$ (assuming that $E\|X_1\|^r < \infty$)?

Solution: (a) Since $\overline{X}_n \to_{a.s.} \mu$ under the assumption that $E|X|^r < \infty$ for some $1 \le r \le 2$, we can consider the class of functions

$$\mathcal{F} \equiv \{f_t(x) : \; t \in [\mu - \delta, \mu + \delta]\}$$

for some $\delta > 0$ where $f_t(x) = |x - t|^r$. These functions are continuous in t for all x, $[\mu - \delta, \mu + \delta]$ is compact, and for each fixed $x \in \mathbb{R}$ we have

$$|f_t(x)| = |x - t|^r \le 2^{r-1}\{|x|^r + |t|^r\} \le 2^{r-1}\{|x|^r + (|\mu + \delta| \wedge |\mu - \delta|)^r \equiv F(x)$$

for all $t \in [\mu - \delta, \mu + \delta]$ where $EF(X) = 2^{r-1}E|X|^r + 2^{r-1}(|\mu + \delta| \wedge |\mu - \delta|)^r < \infty$. Furthermore, the function $H(t) \equiv P|X - t|^r = \int |x - t|^r dP(x)$ is continuous at μ by the dominated convergence theorem and continuity of the functions $t \mapsto |x - t|^r$: note that $2^{r-1}\{|x|^r + |t|^r\} \le 2^{r-1}\{|x|^r + (|\mu + \delta|^r \wedge |\mu - \delta|^r)\}$ is an integrable dominating function. Thus, much as in the handout, for n so large that $\overline{X}_n \in [\mu - \delta, \mu + \delta]$,

$$\begin{aligned} |V_n(r) - v(r)| &\le |\mathbb{P}_n f_{\overline{X}_n} - P f_\mu| \\ &\le |\mathbb{P}_n f_{\overline{X}_n} - P f_{\overline{X}_n}| + |P_{\overline{X}_n} - P f_\mu| \\ &\le \sup_{t : |t - \mu| \le \delta} |\mathbb{P}_n f_t - P f_t| + |H(\overline{X}_n) - H(\mu)| \\ &\to a.s. 0 + 0 = 0 \end{aligned}$$

by using the Glivenko-Cantelli theorem 4.1 for $\mathcal{F}$ to handle the first term, and by the continuous mapping theorem and a.s. consistency of $\overline{X}_n$ for the second term.

(b) In this multivariate case we again have $\overline{X}_n \to_{a.s.} \mu$ in $\mathbb{R}^d$ since $r \geq 1$, and therefore we can reduce to consideration of the class of functions

$$\mathcal{F} = \{f_t(x) : \ \|t - \mu\| \leq \delta\}$$

for some $\delta > 0$ where $f_t(x) = \|x - t\|^r$. These functions are continuous in t for all x, the closed ball $B_\delta(\mu)$ is compact in $\mathbb{R}^d$, and (by the C_r-inequality with $r = 2$ followed by the r taken to be the present r divided by 2)

$$\begin{aligned}
|f_t(x)| = \|x - t\|^r &= \{\sum_{j=1}^{d} |x_j - t_j|^2\}^{r/2} \leq \{\sum_{j=1}^{d} 2(|x_j|^2 + |t_j|^2)\}^{r/2} \\
&= 2^{r/2}\{\sum_{j=1}^{d} 2(|x_j|^2\}^{r/2} + 2^{r/2}\{\sum_{j=1}^{d} |t_j|^2)\}^{r/2} \\
&\leq 2^{r/2}\|x\|^r + 2^{r/2} \sup_{t:\|t-\mu\|=\delta} \|t\|^r \\
&\equiv F(x)
\end{aligned}$$

for all $t \in B_\delta(\mu)$ where

$$EF(X) = 2^{r/2}E\|X\|^r + 2^{r/2} \sup_{t:\|t-\mu\|=\delta} \|t\|^r < \infty.$$

Thus the Glivenko-Cantelli theorem 4.1 holds for for class $\mathcal{F}$. Furthermore, the function $H(t) \equiv P\|X - t\|^r = \int \|x - t\|^r dP(x)$ is continuous at μ by the dominated convergence theorem and continuity of the functions $t \mapsto \|x - t\|^r$: note that $2^{r/2}\{|x|^r + \|t\|^r\} \leq 2^{r/2}\{|x|^r + \sup_{t:\ \|t-\mu\|=\delta} \|t\|^r\}$ is an integrable dominating function. Thus, much as in the handout, for n so large that $\overline{X}_n \in B_\delta(\mu)$,

$$\begin{aligned}
|V_n(r) - v(r)| &\leq |\mathbb{P}_n f_{\overline{X}_n} - P f_\mu| \\
&\leq |\mathbb{P}_n f_{\overline{X}_n} - P f_{\overline{X}_n}| + |P_{\overline{X}_n} - P f_\mu| \\
&\leq \sup_{t:\|t-\mu\|\leq\delta} |\mathbb{P}_n f_t - P f_t| + |H(\overline{X}_n) - H(\mu)| \\
&\to a.s. 0 + 0 = 0
\end{aligned}$$

by using the Glivenko-Cantelli theorem 4.1 for $\mathcal{F}$ to handle the first term, and by the continuous mapping theorem and a.s. consistency of $\overline{X}_n$ for the second term.