Statistics 582, Problem Set 6

Wellner; 2/07/2018

Reading: Chapter 6, sections 1-2.
Due: Wednesday, February 15, 2015.
Reminder: Midterm exam, Monday, February 12;

1. Suppose that $X_1, \ldots, X_n$ are i.i.d. Exponential(θ), so the X's have density $p_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$. with respect to Lebesgue measure on R, and that $\theta \sim \Gamma(\alpha, \beta)$:
$$\lambda(\theta) = \beta \frac{(\beta\theta)^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta) 1_{[0,\infty)}(\theta).$$
(a) Find the Bayes rule $d_B(\underline{X})$ for estimation of θ with squared error loss $L(\theta, a) = |\theta - a|^2$. Find the Bayes rule $d_{Bw}(\underline{X})$ for estimation of θ with weighted squared error loss $L(\theta, a) = (\theta - a)^2/\theta$. Is the maximum likelihood estimator among either of these families of Bayes estimators?
(b) Are the Bayes estimators d_B and d_{Bw} consistent? What are the limit distributions of d_B and d_{Bw}? Compare them with the maximum likelihood estimator.
(c) Suppose that instead of the Gamma prior distribution, θ has the Pareto(θ_0, α) distribution with density λ given by
$$\lambda(\theta) = \left(\frac{\alpha}{\theta_0}\right) \left(\frac{\theta_0}{\theta}\right)^{\alpha+1} 1_{(\theta_0,\infty)}(\theta);$$
here $E(\theta) = \frac{\alpha}{\alpha-1}\theta_0$ where $\alpha > 1$ and $\theta_0 > 0$ are known. What can you say about the Bayes estimator for squared error loss with this prior? For what values of θ_0 is the Bayes rule consistent?

2. Specialize the decision rule in Theorem 5.2 of the course notes to the case when P_i is the normal distribution $N_d(\mu_i, I)$, $i = 1, \ldots, k$ where $\mu_1, \ldots, \mu_k$ are distinct vectors in $\mathbb{R}^d$, $\mu_i \neq \mu_j$ for $i \neq j$. What happens if we replace I by Σ?

3. Lehmann and Casella, TPE, Problem 5.17, page 293, parts (a) and (c) (Also note Problems 5.18, 5.19, 5.20, page 293.) The original proof of Theorem 5.7 (Lehmann and Casella page 260), used Rényi's entropy functions (Rényi, 1961)
$$R_\alpha(f, g) = \frac{1}{\alpha - 1} \log \int f^\alpha(x) g^{1-\alpha}(x) d\mu(x)$$
where f and g are densities, μ is a dominating measure, and $\alpha \neq 1$ is a constant.
(a) Show that $R_\alpha(f, g)$ satisfies $R_\alpha(f, g) \geq 0$ and $R_\alpha(f, f) = 0$.
(c) Show that $\lim_{\alpha \to 1} R_\alpha(f, g) = K(f, g)$.

4. Problem 3.9, Lehmann and Casella, TPE, page 286. For the natural exponential family $p_\eta(x)$ of (3.3.7) and the conjugate prior $\pi(\eta|k,\mu)$ of (3.3.19) establish that:
(a) $E(X) = A'(\eta)$ and $Var(X) = A''(\eta)$ where the expectation is with respect to the sampling density $p_\eta(x)$.
(b) $EA'(\eta) = \mu$ and $Var(A(\eta)) = (1/k)EA''(\eta)$, where the expectation is with respect to the prior distribution.

5. **Optional bonus problem 1:** Part (b) of Problem 5.17, Lehmann and Casella, page 293: show that Theorem 5.7 holds if $R_\alpha(f,g)$ is used instead of $K(f,g)$.

6. **Optional bonus problem 2:** (Birgé). Let $X = (X_0, X_1, \ldots, X_k)$ be a $(k+1)-$dimensional vector, and assume that $X \sim N_{k+1}(\theta, I_{k+1})$ where I_{k+1} denotes the $(k+1)\times(k+1)$ identity matrix. For any vector $\theta \in \mathbb{R}^{k+1}$, let θ' denote the projection of θ onto the $k-$dimensional linear space spanned by the $k-$last coordinates. Consider the subset Θ_0 of $\Theta = \mathbb{R}^{k+1}$ given by

$$\Theta_0 = \{\theta \in \mathbb{R}^{k+1} : |\theta_0| \le k^{1/4} \ \text{ and } \ \|\theta'\| \le 2(1 - k^{-1/4}|\theta_0|)\}.$$

(a) Show that the MLE of θ over Θ_0 is given by $\widehat{\theta}_0 = 0$ and $\widehat{\theta}' = 2X'/\|X\|$ *on the event*

$$\Omega_0 \equiv \{\|X'\|^2 > 3k/4 \ \text{ and } \ |X_0| < k^{1/4} + 1.21\}.$$

(b) Show that $P_\theta(\Omega_0) \ge 3/4$ for all $\theta \in \Theta_0$.
(c) Let $\widetilde{\theta} = (X_0, \underline{0})$. Show that for $k \ge 128$ we have

$$\sup_{\theta\in\Theta_0} E_\theta\|\theta - \hat{\theta}\|^2 \ge (3/4)\sqrt{k} + 3, \quad \text{and}$$
$$\sup_{\theta\in\Theta_0} E_\theta\|\theta - \widetilde{\theta}\|^2 \le 5.$$

Thus the maximal risk of the MLE may be much larger than the minimax risk when k is large.
Hint: A non-central χ^2_k distribution is stochastically larger than a central χ^2_k distribution; then use Lemma 1 of Laurent and Massart (2000) [Laurent, B.; Massart, P. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** (2000), 13021338].