PRINCIPLES OF STATISTICS – EXAMPLES 4/4

Part II, Michaelmas 2022, Po-Ling Loh (email: pll28@cam.ac.uk)
Questions by courtesy of Richard Nickl

1. Consider classifying an observation of a random vector X in $\mathbb{R}^p$ into either a $N(\mu_1, \Sigma)$ or a $N(\mu_2, \Sigma)$ population, where Σ is a known nonsingular covariance matrix and $\mu_1 \neq \mu_2$ are two distinct known mean vectors.

(a) For a prior π assigning probability q to μ_1 and $1-q$ to μ_2, show that the Bayes classifier is unique and assigns X to $N(\mu_1, \Sigma)$ whenever

$$U \equiv D - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$$

exceeds $\log((1-q)/q)$, where $D = X^T \Sigma^{-1}(\mu_1 - \mu_2)$ is the discriminant function.

(b) Show that $U \sim N(\Delta^2/2, \Delta^2)$ whenever $X \sim N(\mu_1, \Sigma)$, and $U \sim N(-\Delta^2/2, \Delta^2)$ whenever $X \sim N(\mu_2, \Sigma)$, where Δ is the *Mahalanobis distance* between μ_1 and μ_2, given by

$$\Delta^2 = (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2).$$

(c) Show that a minimax classifier is obtained from selecting $N(\mu_1, \Sigma)$ whenever $U \geq 0$.

2. Consider classifying an observation X into a population described by a probability density equal to either f_1 or f_2. Assume $P_{f_i}(f_1(X)/f_2(X) = k) = 0$ for all $k \in [0, \infty]$, where $i \in \{1, 2\}$. Show that any admissible classification rule is a Bayes classification rule for some prior π.

3. For $F : \mathbb{R} \to [0, 1]$ a probability distribution function, define its generalized inverse $F^-(u) = \inf\{x : F(x) \geq u\}$, for $u \in [0, 1]$. If U is a uniform $U[0, 1]$ random variable, show that the random variable $F^-(U)$ has distribution function F.

4. Let $f, g : \mathbb{R} \to [0, \infty)$ be bounded probability density functions such that $f(x) \leq Mg(x)$ for all $x \in \mathbb{R}$ and some constant $M > 0$. Suppose you can simulate a random variable X of density g, as well as a random variable U from a uniform $U[0, 1]$ distribution. Consider the following "accept-reject" algorithm:

`Step 1.  Draw` $X \sim g$ `and` $U \sim U[0, 1]$`.`
`Step 2.  Accept` $Y = X$ `if` $U \leq f(X)/(Mg(X))$`, and return to Step 1 otherwise.`

Show that Y has density f.

5. Let U_1 and U_2 be i.i.d. uniform $U[0, 1]$, and define

$$X_1 = \sqrt{-2\log(U_1)}\cos(2\pi U_2), \text{ and } X_2 = \sqrt{-2\log(U_1)}\sin(2\pi U_2).$$

Show that X_1, X_2 are i.i.d. $N(0, 1)$.

6. Consider observations $X_1, \ldots, X_n$ from a statistical model $\{f(\cdot, \theta) : \theta \in \Theta\}$, where $\Theta = \mathbb{R}^p$ and $p \in \mathbb{N}$, and denote by $\pi(\cdot|X_1, \ldots, X_n)$ the posterior distribution arising from a $N(0, I_p)$ prior π on Θ. Suppose the Markov chain $(\vartheta_m : m \in \mathbb{N})$ is started at an arbitrary value $\vartheta_0 \in \mathbb{R}^p$ and evolves as follows:

`Step 1.  For` $m \in \mathbb{N} \cup \{0\}$ `and` $\delta > 0$`, and given` ϑ_m`, generate` $\xi \sim \pi = N(0, I_p)$ `and set`

$$s_m = \sqrt{1 - 2\delta}\vartheta_m + \sqrt{2\delta}\xi.$$

`Step 2.  Define`

$$\vartheta_{m+1} = \begin{cases} s_m, & \texttt{with probability } \rho(\vartheta_m, s_m) \\ \vartheta_m, & \texttt{with probability } 1 - \rho(\vartheta_m, s_m), \end{cases}$$

`where the acceptance probabilities are given by`

$$\rho(\vartheta_m, s_m) = \min\left\{e^{\ell(s_m) - \ell(\vartheta_m)}, 1\right\}, \qquad \ell(\theta) = \sum_{i=1}^{n} \log f(X_i, \theta).$$

`Step 3.  Repeat the above with` $m \mapsto m+1$.

Show that the posterior distribution $\pi(\cdot|X_1, \ldots, X_n)$ is an invariant measure for $(\vartheta_m : m \in \mathbb{N})$.

7. Let $X_1, \ldots, X_n$ be i.i.d. random variables drawn from a distribution P with unknown mean μ and variance σ^2. Write $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ for the sample mean, and let $\bar{X}_n^b = (1/n)\sum_{i=1}^n X_{ni}^b$ be the mean of a bootstrap sample $(X_{ni}^b : i = 1, \ldots, n) \sim^{i.i.d.} \mathbb{P}_n$ generated from the X_i's. Choosing roots R_n such that

$$\mathbb{P}_n\left(\bar{X}_n^b - \bar{X}_n \leq \frac{R_n}{\sqrt{n}}\right) = 1 - \alpha$$

for some $0 < \alpha < 1$, let

$$C_n^b = \left\{v \in \mathbb{R} : \bar{X}_n - v \leq \frac{R_n}{\sqrt{n}}\right\}$$

be the corresponding one-sided bootstrap confidence interval. Show that R_n converges to a constant in $P^{\mathbb{N}}$-probability, and deduce further that C_n^b is an exact asymptotic level $1 - \alpha$ confidence set, i.e., as $n \to \infty$,

$$P^{\mathbb{N}}(\mu \in C_n^b) \to 1 - \alpha.$$

8. Let $X_1, \ldots, X_n$ be drawn i.i.d. from a continuous distribution function $F : \mathbb{R} \to [0,1]$, and let $F_n(t) = (1/n)\sum_{i=1}^n 1_{(-\infty,t]}(X_i)$ be the empirical distribution function. Use the Kolmogorov-Smirnov theorem to construct a confidence band for the unknown function F of the form

$$\{C_n(x) = [F_n(x) - R_n, F_n(x) + R_n] : x \in \mathbb{R}\}$$

that satisfies $P_F^{\mathbb{N}}(F(x) \in C_n(x)\ \forall x \in \mathbb{R}) \to 1 - \alpha$ as $n \to \infty$, and where $R_n = R/\sqrt{n}$ for some fixed quantile constant $R > 0$.

9. Given $X_1, \ldots, X_n$ from a regular statistical model $\{f(\cdot, \theta) : \theta \in \Theta\}$, where $\Theta = \mathbb{R}^p$, with non-singular Fisher information $I(\theta)$, consider "local" perturbations $\theta_0 + (h/\sqrt{n})$, with $h \in \mathbb{R}^p$, of the log-likelihood ratios near a "true" value θ_0. More precisely, define

$$Z_n(h) = \log \frac{\prod_{i=1}^n f(X_i, \theta_0 + h/\sqrt{n})}{\prod_{i=1}^n f(X_i, \theta_0)}, \quad X_i \sim^{i.i.d.} f(\cdot, \theta_0).$$

Next, consider a normal shift experiment given by the probability density functions $(p_h : h \in \mathbb{R}^p)$ of normal distributions $N(h, I(\theta_0)^{-1})$, and denote the corresponding likelihood ratios by

$$Z(h) = \log \frac{p_h(X)}{p_0(X)}, \quad X \sim p_0.$$

Show that for every fixed $h \in \mathbb{R}^p$, the random variables $Z_n(h)$ converges in distribution under P_{θ_0} to the law of $Z(h)$, as $n \to \infty$. [This suggests that at least in $1/\sqrt{n}$-neighborhoods of θ_0, the likelihood ratio process of any regular statistical model behaves like the one of a simple Gaussian shift experiment with mean h and covariance $I(\theta_0)^{-1}$.]