Homework problems to be turned in

- Note. The total of this assignment is 60 points, so you only need to complete a part of the problems correctly to receive full points.

1. (10 pts) Prove Theorem 3.3 (Cramér-Rao lower bound) for the multivariate case ($k \geq 2$). Hint: find

$$\min_{c \in R^k} E\left(T(X) - g(\theta) - c^T \frac{\partial}{\partial \gamma} \log f_\gamma(X)\Big|_{\gamma=\theta} \right)^2.$$

2. (25 pts) Assume the conditions in Theorem 4.17. Suppose that $\tilde{\theta}_n$ is a consistent estimator of θ and $a_n(\tilde{\theta}_n - \theta) = O_p(1)$, where $\{a_n\}_{n=1}^\infty$ is a known sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$.

   (a) (10 pts) Construct a sequence $\{\delta_n\}_{n=1}^\infty$ so that $\delta_n > 0$ for all n,

   $$\lim_{n\to\infty} \delta_n = 0$$

   and for $c > 0$, for $\varepsilon > 0$, there exists n_1 such that

   $$n > n_1 \Rightarrow P(B_n(c) \subset \{\gamma : \|\gamma - \tilde{\theta}_n\| \leq \delta_n\}) \geq 1 - \varepsilon,$$

   where $B_n(c)$ is defined in the proof of Theorem 4.17.

   (b) (15 pts) Define

   $$\Theta_n = \{\gamma : \|\gamma - \tilde{\theta}_n\| \leq \delta_n\} \cap \{\gamma : s_n(\gamma) = 0\}$$

   for $n \geq 1$, where $\{\delta_n\}_{n=1}^\infty$ is constructed in Part (a) and s_n is the score function. Let

   $$\hat{\theta}_n = \begin{cases} \text{some } \theta^* \in \Theta_n & \text{if } \Theta_n \neq \emptyset; \\ \tilde{\theta}_n & \text{if } \Theta_n = \emptyset. \end{cases}$$

   Show that $\hat{\theta}_n$ is a consistent estimator of θ and is asymptotically efficient.

3. (25 pts) Suppose that X is a sample of size n and P_X (the distribution of X) belongs to a family $\{P_\theta : \theta \in \Theta\}$. Suppose that there exists a σ-finite measure ν such that P_θ has a density f_θ with respect to ν for all $\theta \in \Theta$. Suppose that

(*) there exist $c_i > 0$ and $\theta_i \in \Theta$ for $i \in \{1, 2, \ldots\}$ such that $P_\theta \ll P_{X_0}$ for all $\theta \in \Theta$, where X_0 is a random vector with PDF $\sum_{i=1}^{\infty} c_i f_{\theta_i}$ with respect to ν.

(a) (10 pts) Let $S_0 = \{x : \sum_{i=1}^{\infty} c_i f_{\theta_i}(x) > 0\}$. Show that $P_\theta(S_0^c) = 0$ for all $\theta \in \Theta$ and

$$\frac{dP_\theta}{dP_{X_0}}(x) = \frac{f_\theta(x)}{\sum_{i=1}^{\infty} c_i f_{\theta_i}(x)} = \frac{f_\theta(x)}{\sum_{i=1}^{\infty} c_i f_{\theta_i}(x)} I_{S_0}(x)$$

P_{X_0}-a.e. for $x \in R^n$.

(b) (15 pts) Show that if there exist nonnegative measurable functions g and h such that

$$f_\theta(x) = g(\theta, T(x))h(x)$$

for $x \in R^n$, then $T(X)$ is a sufficient statistic for θ.

Note that the existence of c_is and θ_is in (*) is guaranteed by Lemma 2.1 in the text.

4. (10 pts) Suppose that G is a CDF on R and the function G^{-1} is defined by

$$G^{-1}(t) = \inf\{x : G(x) \geq t\}$$

for $t \in (0, 1)$. Show that for $t \in (0, 1)$, $G(x) \geq t$ if and only if $x \geq G^{-1}(t)$.

5. (10 pts) Suppose that $\{(X_{n,1}, \ldots, X_{n,k})\}_{n=1}^{\infty}$ and $\{(p_{n,1}, \ldots, p_{n,k})\}_{n=1}^{\infty}$ are two sequences of random vectors and $p_1, \ldots, p_k$ are positive constants such that for $j = 1, \ldots, k$, $X_{n,j}/n$ converges to p_j in probability as $n \to \infty$ and $p_{n,j}$ converges to p_j in probability as $n \to \infty$. Suppose that

$$\sum_{j=1}^{k} \frac{(X_{n,j} - np_{n,j})^2}{X_{n,j}}$$

converges in distribution to $\chi^2(k-1)$ as $n \to \infty$. Show that

$$\sum_{j=1}^{k} \frac{(X_{n,j} - np_{n,j})^2}{np_{n,j}}$$

converges to $\chi^2(k-1)$ in distribution as $n \to \infty$.

6. (10 pts) Suppose that X_1, ..., X_n are IID random variables and X_1 takes values in $\{1, \ldots, k\}$. Let $p_j = P(X_1 = j)$ for $j \in \{1, \ldots, k\}$. For $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, k\}$, let

$$U_{i,j} = \begin{cases} 1 & \text{if } X_i = j; \\ 0 & \text{otherwise.} \end{cases}$$

For $j \in \{1, \ldots, k\}$, let $W_{n,j} = (\sum_{i=1}^n U_{i,j} - np_j)/\sqrt{n}$, then it follows from the central limit theorem that $(W_{n,1}, \ldots, W_{n,k})$ converges in distribution to the normal distribution of mean $(0, \ldots, 0)$ and covariance Σ, where Σ is the covariance matrix of $(U_{1,1}, \ldots, U_{1,k})$.

(a) (5 pts) Verify that the (j, ℓ)-th element of Σ is

$$\begin{cases} p_j(1 - p_j) & \text{if } j = \ell; \\ -p_j p_\ell & \text{otherwise.} \end{cases}$$

for j, $\ell \in \{1, \ldots, k\}$.

(b) (5 pts) Let $v = (\sqrt{p_1}, \ldots, \sqrt{p_k})^T$ and let D be the diagonal matrix whose j-th diagonal element is $p_j^{-1/2}$ for $j \in \{1, \ldots, k\}$. Verify that

$$D\Sigma D = I_k - vv^T. \tag{1}$$

You may verify (1) analytically or verify it numerically by computing

$$\|D\Sigma D - (I_k - vv^T)\|^2,$$

based on simulated (p_1, ..., p_k), where for a matrix A, $\|A\|^2$ is the sum of squares of all elements of A. If you choose to verify (1) numerically, you need to provide a R function with input (p_1, ..., p_k) and output

$$\|D\Sigma D - (I_k - vv^T)\|^2,$$

desribe how you generated (p_1, ..., p_k) and give a summary of your simulation results.

7. (25 pts) Suppose that X and Y are discrete random variables, X has m possible values x_1, ..., x_m and Y has k possible values y_1, ..., y_k. Suppose that (X_1, Y_1), ..., (X_n, Y_n) are IID observations and (X_1, Y_1) has the same distribution as (X, Y). Let

$$p_{\ell,j} = P(X = x_\ell \text{ and } Y = y_j)$$

for $\ell \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, k\}$. Let

$$p_{\ell,\cdot} = \sum_{j=1}^{k} p_{\ell,j}$$

for $\ell \in \{1, \ldots, m\}$ and let

$$p_{\cdot,j} = \sum_{\ell=1}^{m} p_{\ell,j}$$

for $j \in \{1, \ldots, k\}$. Then X and Y are indepdendent if and only if

$$p_{\ell,j} = (p_{\ell,\cdot})(p_{\cdot,j})$$

for $\ell \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, k\}$. Consider testing

$$H_0: \; X \text{ and } Y \text{ are indepdendent}$$

versus

$$H_1: \; X \text{ and } Y \text{ are not indepdendent}$$

based on (X_1, Y_1), ..., (X_n, Y_n). Let λ_n be the likelihood ratio statistic. It can be shown that the conditions in Theorem 4.16 hold except that we have IID bivariate data instead of univariate data. Thus by Theorem 6.6, under H_0, $-2\log(\lambda_n)$ converges in distribution to $\chi^2(r)$ as $n \to \infty$. You do not have to verify the conditions in Theorem 4.16 for this problem but be sure that you know how to do so.

(a) (5 pts) Express r as a function of m and k.

(b) (20 pts) For $\ell \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, k\}$, let

$$N_{\ell,j} = \sum_{i=1}^{n} I(X_i = x_\ell \text{ and } Y_i = y_j),$$

where

$$I(X_i = x_\ell \text{ and } Y_i = y_j) = \begin{cases} 1 & \text{if } X_i = x_\ell \text{ and } Y_i = y_j; \\ 0 & \text{otherwise.} \end{cases}$$

Then the chi-squared test for independence is based on the statistic

$$W_n = \sum_{\ell=1}^{m} \sum_{j=1}^{k} \frac{(N_{\ell,j} - n(\hat{p}_{\ell,\cdot})(\hat{p}_{\cdot,j}))^2}{n(\hat{p}_{\ell,\cdot})(\hat{p}_{\cdot,j})},$$

where

$$\hat{p}_{\ell,\cdot} = \frac{\sum_{j=1}^{k} N_{\ell,j}}{n}$$

for $\ell \in \{1, \ldots, m\}$ and let

$$\hat{p}_{\cdot,j} = \frac{\sum_{\ell=1}^{m} N_{\ell,j}}{n}$$

for $j \in \{1, \ldots, k\}$. Derive the result that under H_0, W_n converges in distribution to $\chi^2(r)$ as $n \to \infty$ using the result that under H_0, $-2\log(\lambda_n)$ converges in distribution to $\chi^2(r)$ as $n \to \infty$.

8. (25 pts) Suppoe that $\mu = (0, 0, 0, 0)^T$ and $\Sigma = B^T B$, where

$$B^T = \begin{pmatrix} 1 & 0 & 0. & 0 \\ -0.5 & 1 & 0 & 0 \\ 0 & -0.5 & 1 & 0 \\ 0 & 0 & -0.5 & 1 \end{pmatrix}.$$

Consider generating Y_1, …, Y_N from $N(\mu, \Sigma)$ using three different methods.

- Method 1. Generate Y_1, …, Y_N independently, where for each $i \in \{1, \ldots, N\}$, generate X_1, X_2, X_3, X_4 independently from $N(0, 1)$ and take
$$Y_i = B^T \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}.$$
The Y_1, …, Y_N are IID and $Y_i \sim N(\mu, \Sigma)$.
- Method 2. Let f_1 be the continuous PDF of $N(\mu, \Sigma)$ and let f_0 be the continuous PDF of $N(\mu, I_4)$, where I_4 is the 4×4 identity matrix. Choose N_0 to be a large positive integer and take $M = N + N_0$. Generate $\{U_i\}_{i=1}^{M}$ as follows.
  (a) Generate U_1 from $N((0, 0, 0, 0)^T, I_4)$.
  (b) For each $t \in \{1, \ldots, M-1\}$, suppose that U_1, …, U_t have been generated. Carry out Steps (i)–(iv) to obatin U_{t+1}:
    i. Generate Z from $N((0, 0, 0, 0)^T, I_4)$ independently from $(U_1, \ldots, U_t)$.
    ii. Compute
    $$\beta = \min\{0, \log(f_1(Z)) - \log(f_0(Z)) - [\log(f_1(U_t)) - \log(f_0(U_t))]\}.$$

iii. Generate U^* from $U(0,1)$ independently from (U_1, ..., U_t, Z).

iv. Take

$$U_{t+1} = \begin{cases} Z & \text{if } U^* \le e^\beta; \\ U_t & \text{if } U^* > e^\beta. \end{cases}$$

Take (Y_1, ..., Y_N) = (U_{N_0+1}, ..., $U_M = U_{N_0+N}$).

- Method 3. Suppose that $(X_1, X_2, X_3, X_4)^T \sim N(\mu, \Sigma)$. For $i \in \{1, 2, 3, 4\}$, let X_{-i} denote the vector obtained by removing X_i from (X_1, X_2, X_3, X_4), and define

$$\mu_i(X_1, X_2, X_3, X_4) = E(X_i|X_{-i}) \tag{2}$$

and

$$\sigma_i^2 = E(X_i - E(X_i|X_{-i}))^2. \tag{3}$$

Choose N_0 to be a large positive integer and take $M = N + N_0$. Generate $\{U_i\}_{i=1}^M$ as follows.

(a) Take $U_1 = (0, 0, 0, 0)^T$.

(b) For each $t \in \{1, \ldots, M-1\}$, suppose that U_1, ..., U_t have been generated. Let $(U_{t,1}, U_{t,2}, U_{t,3}, U_{t,4})^T = U_t$. Carry out Steps (i)–(vi) to obatin U_{t+1}:

i. Generate IID $N(0, 1)$ random variables ε_1, ε_2, ε_3, ε_4 so that (ε_1, ε_2, ε_3, ε_4) is independent from (U_1, ..., U_t).

ii. Take $Z_1 = \mu_1(U_{t,1}, U_{t,2}, U_{t,3}, U_{t,4}) + \sigma_1\varepsilon_1$.

iii. Take $Z_2 = \mu_2(Z_1, U_{t,2}, U_{t,3}, U_{t,4}) + \sigma_2\varepsilon_2$.

iv. Take $Z_3 = \mu_3(Z_1, Z_2, U_{t,3}, U_{t,4}) + \sigma_3\varepsilon_3$.

v. Take $Z_4 = \mu_4(Z_1, Z_2, Z_3, U_{t,4}) + \sigma_4\varepsilon_4$.

vi. Take $U_{t+1} = (Z_1, Z_2, Z_3, Z_4)^T$.

Take (Y_1, ..., Y_N) to be (U_{N_0+1}, ..., $U_M = U_{N_0+N}$).

Note that Method 2 is based on the Metropolis algorithm and Method 3 is based on Gibbs sampling. Let

$$A = [-0.5, 0.5] \times [-1, 1] \times [-1.5, 1.5] \times [-2, 2].$$

We will estimate

$$p \equiv P(N(\mu, \Sigma) \in A)$$

by generating Y_1, ..., Y_N from $N(\mu, \Sigma)$, and then estimate p using

$$\hat{p} \equiv \frac{\sum_{t=1}^N I_A(Y_t)}{N}.$$

For $j = 1, 2, 3$, let $\hat{p}_j$ denote the estimator $\hat{p}$ when $Y_1, \ldots, Y_N$ are generated using Method j.

(a) (9 pts) For the (X_1, X_2, X_3, X_4) in Method 3, it is known that $E(X_i|X_{-i})$ is a linear function of X_{-i} for $i = 1, \ldots, 4$. Define μ_i and σ_i by (2) and (3) respectively for $i = 1, \ldots, 4$. Write a R function with input $(x_1, x_2, x_3, x_4) \in R^4$ and output $(\sigma_1^2, \mu_1(x_1, x_2, x_3, x_4))$.

(b) (8 pts) Take $N_0 = 100$ and $N = 200$. For $\alpha \in (0, 1)$, propose an approximate level α test for testing

$$H_0 : E(\hat{p}_1) = E(\hat{p}_2) \text{ v.s. } H_1 : E(\hat{p}_1) \neq E(\hat{p}_2)$$

based on a random sample of 500 $\hat{p}_1$'s and a random sample of 500 $\hat{p}_2$'s, where the two samples are independent. Generate these two random samples in R and carry out the proposed test. Report the p-value based on the generated data. Can we reject H_0 at level 0.05?

(c) (8 pts) Take $N_0 = 100$ and $N = 200$. For $\alpha \in (0, 1)$, propose an approximate level α test for testing

$$H_0 : E(\hat{p}_1) = E(\hat{p}_3) \text{ v.s. } H_1 : E(\hat{p}_1) \neq E(\hat{p}_3)$$

based on a random sample of 500 $\hat{p}_1$'s and a random sample of 500 $\hat{p}_3$'s, where the two samples are independent. Generate these two random samples in R and carry out the proposed test. Report the p-value based on the generated data. Can we reject H_0 at level 0.05?