第一卷

PhD QR exam 96.09 <u>MathStat</u>

1. (25 points) Let $(X_i, Y_i), i = 1, ..., n$ be iid according to the bivariate normal distribution with parameter $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. Let R denote the sample correlation coefficient, $\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$, and consider the hypothesis testing problem $H_0 : \rho = 0$ v.s. $H_1 : \rho \neq 0$.
(a) (15 points) Suppose that $\rho = 0$. First show that the conditional distribution of $T = \sqrt{n-2}R/\sqrt{1-R^2}$ given $x_1, ..., x_n$ has the t-distribution with $n-2$ degrees of freedom provided $\sum(x_i - \bar{x})^2 > 0$. Then argue that this is therefore also the unconditional distribution of T.
(b) (10 points) Show that the UMP unbiased test rejects H_0 when

$$|R|/\sqrt{(1-R^2)/(n-2)} > K_0,$$

and determine K_0 if the significance level is α.

2. (25 points) If X is a nonnegative random variable, then Markov's Inequality says that for any $a > 0$, $P(X \geq a) \leq E(X)/a$.
(a) (10 points) Let X be a random variable with moment generating function $M(t) = E(e^{tX})$. Use the above inequality to prove the following: for $a > 0$,

$$\begin{aligned} P(X \geq a) &\leq e^{-ta}M(t) \text{ for all } t > 0 \\ P(X \leq a) &\leq e^{-ta}M(t) \text{ for all } t < 0. \end{aligned}$$

(b) (10 points) If X is $Poi(\lambda)$, then $M(t) = e^{\lambda(e^t-1)}$. Find the best bound on $P(X \geq j)$ based on part (a).
(c) (5 points) Compare the bound you obtained in (b) with that using Markov's Inequality.

3. (35 points) Consider a hypothesis testing problem H_0: P_0 is true v.s. H_1: P_1 is true, where P_0 and P_1 are two probability distributions. Suppose that we sample $x_1, x_2, ...$ until

$$N = \inf\{n : \ell_n \notin (A, B)\}, \tag{1}$$

where ℓ_n is the loglikelihood function based on n i.i.d. observations and $0 < A < B < \infty$ are two positive constants. Then, we reject H_0 if $\ell_N \geq B$ and not reject H_0 if $\ell_N \leq A$. It is easily seen that the type I and type II errors are $\alpha = P_0(\ell_N \geq B)$ and $\beta = 1 - P_1(\ell_N \geq B)$.

For given A and B, we can do computer simulation and estimate α by

$$\hat{\alpha} = \frac{1}{t}\sum_{k=1}^{t} 1[\ell_{N_k} \geq B],$$

where $1[\cdot]$ is the indicator function, each N_k is the random sample size satisfying (1), and the experiments are conducted under P_0.
(a) (10 points) Find the mean and variance of $\hat{\alpha}$.
(b) (5 points) If one wants to estimate α to within 10% of its value with probability

95%, how large t do we need?

(c) (5 points) Show that

$$P_0(\ell_N \geq B) = \int_{\ell_N \geq B} \frac{1}{\ell_N} dP_1.$$

(d) (10 points) Based on the equation above, can you propose an unbiased estimator of α under P_1 using computer simulation?

(e) (5 points) Denote the estimator you give in (d) as $\tilde{\alpha}$. What can you say about the variance of $\tilde{\alpha}$?

4. (15 points) Suppose that $X_1, ..., X_n$ is a random sample from $Poi(\lambda)$ distribution. Find a variance stabilizing transformation of $\bar{X}_n$.