Math 281A Homework 2 Solution

1. Let $X_1, \ldots, X_n$ be i.i.d. from $N(0,1)$, show that $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are independent.

   Solution: It's easy to see that $(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})$ is multivariate normal, so showing the desired independence is equivalent as checking $\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = 0$, for $i = 1, 2, \ldots, n$. Now,

   $$\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = \mathrm{Cov}(\bar{X}, X_i) - \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X_i)}{n} - \mathrm{Var}(\bar{X}) = \frac{1}{n} - \frac{1}{n} = 0.$$

2. Suppose that random vector (X, Y) has probability density function

   $$\frac{1}{\pi} e^{-\frac{x^2+y^2}{2}} I(xy > 0).$$

   Does (X, Y) possess a multivariate normal distribution? Find the marginal distributions.

   Solution: No, the density function is only defined on two quadrants. Marginally, when $x > 0$,

   $$f_X(x) = \int_0^\infty \frac{1}{\pi} e^{-\frac{x^2+y^2}{2}} \mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

   and when $x < 0$,

   $$f_X(x) = \int_{-\infty}^0 \frac{1}{\pi} e^{-\frac{x^2+y^2}{2}} \mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

   and $f_X(0) = 0$.

3. Suppose T_n and S_n are sequences of estimators such that

   $$\sqrt{n}(T_n - \theta) \xrightarrow{d} N_k(0, \Sigma), \text{ and } S_n \xrightarrow{P} \Sigma,$$

   for a certain vector θ and a nonsingular matrix Σ. Show that

   (a) S_n is nonsingular with probability tending to one;

   Solution 1: First notice that $S_n \xrightarrow{P} \Sigma$ implies $|S_n| \xrightarrow{P} |\Sigma|$, so for any $\epsilon > 0$,

   $$\mathbb{P}(|S_n| > |\Sigma| - \epsilon) \to 1.$$

   Since $|\Sigma| > 0$, taking $\epsilon = |\Sigma|/2$ gives us

   $$\mathbb{P}(|S_n| > |\Sigma|/2) \to 1,$$

   which implies $\mathbb{P}(S_n \text{ is nonsingular}) \to 1$.

   Solution 2: Denote $D_n = S_n - \Sigma$, and let $\bar{\sigma}(A)$ and $\underline{\sigma}(A)$ be the largest and smallest singular values of matrix A respectively. $S_n \xrightarrow{P} \Sigma$ gives us for any $\epsilon > 0$,

   $$\mathbb{P}(\bar{\sigma}(D_n) < \epsilon) \to 1. \tag{1}$$

   Since $S_n = D_n + \Sigma$, we have the following inequality derived from triangle inequality,

   $$\underline{\sigma}(S_n) \geq \underline{\sigma}(\Sigma) - \bar{\sigma}(D_n).$$

   Then, because Σ is nonsingular, $\underline{\sigma}(\Sigma) > 0$, we consider

   $$\begin{aligned}\mathbb{P}(\underline{\sigma}(S_n) > \underline{\sigma}(\Sigma)/2) &\geq \mathbb{P}(\underline{\sigma}(\Sigma) - \bar{\sigma}(D_n) > \underline{\sigma}(\Sigma)/2) \\ &= \mathbb{P}(\bar{\sigma}(D_n) < \underline{\sigma}(\Sigma)/2) \to 1,\end{aligned}$$

   where the last step comes from (1). This implies

   $$\mathbb{P}(S_n \text{ is nonsingular}) \to 1.$$

(b) $\{\theta : n(T_n - \theta)^\top S_n^{-1}(T_n - \theta) \le \chi^2_{k,\alpha}\}$ is a confidence ellipsoid of asymptotic confidence level $1 - \alpha$.

Solution: The set $\{\theta : n(T_n - \theta)^\top S_n^{-1}(T_n - \theta) \le \chi^2_{k,\alpha}\}$ is only defined when S_n^{-1} exists, and from (a) we know that $\mathbb{P}(S_n^{-1} \text{ exists}) \to 1$. Now conditional on the event that S_n^{-1} exists, $S_n \xrightarrow{P} \Sigma$ implies $n(T_n - \theta)S_n^{-1}(T_n - \theta) - n(T_n - \theta)\Sigma^{-1}(T_n - \theta) \xrightarrow{P} 0$, and $\sqrt{n}(T_n - \theta) \xrightarrow{d} N_k(0, \Sigma)$ implies $n(T_n - \theta)\Sigma^{-1}(T_n - \theta) \xrightarrow{d} \chi^2_k$. It follows from Slutsky's theorem that

$$n(T_n - \theta)S_n^{-1}(T_n - \theta) \xrightarrow{d} \chi^2_k,$$

which completes the proof.

4. Suppose that $X_m \sim \text{Binomial}(m, p_1), Y_n \sim \text{Binomial}(n, p_2)$ and they are independent. To test $H_0 : p_1 = p_2 = a$, we consider the test statistic

$$C^2_{m,n} = \frac{(X_m - ma)^2}{ma(1-a)} + \frac{(Y_n - na)^2}{na(1-a)}.$$

(a) Find the limit distribution of $C^2_{m,n}$ as $m, n \to \infty$;

Solution: We can write X_m and Y_n as

$$X_m = \sum_{i=1}^m \tilde{X}_i, \text{ and } Y_n = \sum_{i=1}^n \tilde{Y}_i,$$

where $\tilde{X}_i \sim \text{Bernoulli}(p_1)$ and $\tilde{Y}_i \sim \text{Bernoulli}(p_2)$. Under null hypothesis $p_1 = p_2 = a$, we have

$$\frac{X_m - ma}{\sqrt{ma(1-a)}} \xrightarrow{d} N(0,1), \text{ and } \frac{Y_n - na}{\sqrt{na(1-a)}} \xrightarrow{d} N(0,1)$$

from central limit theorem. This leads us to the conclusion

$$C^2_{m,n} \xrightarrow{d} \chi^2_2.$$

(b) How would you modify the test statistic if a were unknown? What's the limit distribution after modification? You don't need to rigorously prove this question.

Solution: If a were unknown, then we replace it with its MLE

$$\hat{a} = \frac{X_m + Y_n}{m + n}$$

in the test statistic. By doing this, we add one more restriction so that one degree of freedom is sacrificed, and the limit distribution becomes χ^2_1.