Math 494: Mathematical Statistics
Solutions to Midterm1

1. (12 points) Assuming X is a discrete random variable whose pmt is $p(x)$ supported on a finite set $\{a_1, \ldots, a_m\}$. Let the quantity of interest be $\theta_j = p(a_j)$, $j = 1, \ldots, m$. Let $X_1, \ldots, X_n$ be a random sample from the distribution with pmf $p(x)$.

   (a) (6 points) Show that the estimator $\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^n 1(X_i = a_j)$ is an unbiased estimator of θ_j. Here $1(A) = 1$ if A is true and 0 otherwise.

   (b) (6 points) Find the variance of $\hat{\theta}_j$.

Solution:

(a) [1pt] We need to show that $E(\hat{\theta}_j) = \theta_j$.

$$
\begin{aligned}
\text{[1pt]} \quad \mathbb{E}(\hat{\theta}_j) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big(1(X_i = a_j)\big) \\
\text{[2pts]} \quad &= \frac{1}{n}\sum_{i=1}^n P(X_i = a_j) \\
\text{[2pts]} \quad &= \frac{1}{n}\sum_{i=1}^n p(a_j) = \frac{1}{n}\sum_{i=1}^n \theta_j = \theta_j.
\end{aligned}
$$

(b) Since $1(X_i = a_j)$'s, $i = 1, 2, \ldots, n$, are i.i.d., we have

$$\text{[2pts]} \quad Var(\hat{\theta}_j) = \frac{1}{n^2}\sum_{i=1}^n Var\big(1(X_i = a_j)\big).$$

For each i,

$$\text{[2pts]} \quad Var\big(1(X_i = a_j)\big) = p(a_j)(1 - p(a_j)) = \theta_j(1 - \theta_j)$$

since $1(X_i = a_j)$ is a binomial random variable with $p = p(a_j)$. So

$$\text{[2pts]} \quad Var(\hat{\theta}_j) = \frac{\theta_j(1 - \theta_j)}{n}.$$

2. (10 points) Let $\bar{X}$ be the mean of a random sample of size n from the distribution $N(\mu, 4)$. Find n such that $P(\bar{X} - 1 < \mu < \bar{X} + 1) = 0.95$. Represent your answer by the notation Z_α and give the proper value of α. (Note that $P(Z > Z_\alpha) = \alpha$ for a standard normal random variable Z.)

Solution: [2pts] $\overline{X} \sim N(\mu, \frac{4}{n})$, so

$$\text{[4pts]} \qquad P\left[\left|\frac{\overline{X} - \mu}{2/\sqrt{n}}\right| < z_{0.025}\right] = 0.95.$$

This is $P\left[\overline{X} - \frac{2}{\sqrt{n}} z_{0.025} < \mu < \overline{X} + \frac{2}{\sqrt{n}} z_{0.025}\right] = 0.95$.
[4pts] Let $\frac{2}{\sqrt{n}} z_{0.025} = 1$, we find $n = 4z^2_{0.025}$.

3. (20 points) Let $Y_1 < Y_2$ be the order statistics of a random sample of size 2 from $U(0, 1)$.
   (a) (10 points) Find $E(Y_1)$.
   (b) (10 points) Find the covariance of Y_1 and Y_2.

Solution:

(a) Using (4.4.2) in textbook, the p.d.f. of Y_1 is

$$\text{[5pts]} \quad g(y_1) = \frac{2!}{0!1!}[F(y_1)]^0(1 - F(y_1))^1 f(y_1) = 2(1 - y_1),$$

where $F(\cdot)$ and $f(\cdot)$ are the c.d.f. and p.d.f. of the $U(0, 1)$ distribution. So

$$\text{[5pts]} \quad \mathbb{E}(Y_1) = \int_0^1 2(1 - y_1)y_1 dy_1 = \frac{1}{3}.$$

(b) Using (4.4.3) in textbook, the joint p.d.f. of Y_1 and Y_2 is

$$\text{[2pts]} \quad g_{12}(Y_1, Y_2) = \frac{2!}{0!0!0!}[F(y_1)]^0[F(y_2) - F(y_1)]^0[1 - F(y_2)]^0 f(y_1)f(y_2) = 2.$$

So

$$\text{[2pts]} \quad \mathbb{E}(Y_1Y_2) = \int_0^1 \int_0^{y_2} y_1y_2g_{12}(y_1, y_2)dy_1dy_2 = 2\int_0^1 \int_0^{y_2} y_1y_2dy_1dy_2$$
$$\text{[1pt]} \quad = 2\int_0^1 y_2 \int_0^{y_2} y_1dy_1dy_2 = 2\int_0^1 y_2\frac{y_2^2}{2}2dy_2 = \frac{1}{4}.$$

Similar to part (a), using (4.4.2) we have the p.d.f. of Y_2 as $g(y_2) = 2y_2$ [2pts], then

$$\text{[1pt]} \quad \mathbb{E}(Y_2) = \int_0^1 2y_2^2 dy_2 = \frac{2}{3}.$$

So

$$\text{[2pts]} \quad Cov(Y_1, Y_2) = \mathbb{E}(Y_1Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) = \frac{1}{4} - \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{36}.$$

4. (20 points) Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ denote the sample variance of a random sample $X_1, \ldots, X_n$ from $N(\mu, \sigma^2)$.

   (a) (10 points) Find a $(1-\alpha)$ confidence interval of the format $(0, (n-1)S^2/c)$ for σ^2 using the sampling distribution of S^2 in Student's theorem. Give the constant c as the quantile of a chi-square distribution.

   (b) (10 points) Consider testing $H_0 : \sigma^2 = \sigma_0^2$ v.s. $H_1 : \sigma^2 > \sigma_0^2$. What is the size of the test to reject H_0 if $(n-1)S^2/\sigma_0^2 > c$, where c is the constant in (a)? Explain.

Solution:

(a) [4pts] $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
[4pts] Let U be a random variable that has a χ^2-distribution with $n-1$ degrees of freedom, and ξ_α is the real number for which $P(U > \xi_\alpha) = 1-\alpha$. Then

$$P\left[\frac{(n-1)S^2}{\sigma^2} > \xi_\alpha\right] = 1 - \alpha.$$

This is

$$P\left[\sigma^2 < \frac{(n-1)S^2}{\xi_\alpha}\right] = 1 - \alpha.$$

[2pts] Hence, $c = \xi_\alpha$.

(b) [4pts] By definition, size$=P$(reject H_0 $|H_0$).
[4pts] Under H_0, $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$, so

$$\text{[2pts]} \qquad \text{size} = P\left[\frac{(n-1)S^2}{\sigma_0^2} > \xi_\alpha\right] = 1 - \alpha.$$

5. (15 points) The chi-square test statistic for testing the independence between the row and column variables in a 2×2 contingency table is the form

$$Q = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(X_{ij} - O_{ij})^2}{O_{ij}},$$

where X_{ij}'s are the observed cell counts in the table. The following table summarizes results from a study on whether attending class influences how students perform on an exam. For example, 25 students who attended classes passed the exam.

	Pass	Fail
Attended	25	6
Skipped	8	15

(a) (5 points) Suppose that we apply the chi-square test to the above table. Find the values of O_{ij}'s.

(b) (5 points) Under the null hypothesis that the row and column variables are independent, what distribution does the test statistic Q have?

(c) (5 points) The chi-square test is an asymptotic test. Do you see any pitfall in this application of the test? Explain.

Solution:

(a) [1pt]

$$\begin{aligned} N &= 25 + 6 + 8 + 15 = 54 \\ n_{1.} &= 25 + 6 = 31, \; n_{.1} = 25 + 8 = 33, \\ n_{2.} &= 8 + 15 = 23, \; n_{.2} = 6 + 15 = 21. \end{aligned}$$

[4pts]

$$\begin{aligned} O_{11} &= \frac{n_{1.}n_{.1}}{N} = \frac{31 * 33}{54} = 18.94, \\ O_{21} &= \frac{n_{2.}n_{.1}}{N} = \frac{23 * 33}{54} = 14.06, \\ O_{12} &= \frac{n_{1.}n_{.2}}{N} = \frac{31 * 21}{54} = 12.06, \\ O_{22} &= \frac{n_{2.}n_{.2}}{N} = \frac{23 * 21}{54} = 8.94. \end{aligned}$$

(b) [3pts] The test statistic has a Chi-square distribution.
[2pts] The degree of freedom is

$$df = (2 - 1) * (2 - 1) = 1.$$

(c) [2pts] The chi-square test is sensitive to small expected counts in one or more of the cells in the table.
[3pts] Since expected counts are greater than 5 in this example, there is no problem of applying the chi-square test.

6. (22 points) Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and a finite variance σ^2.

   1. (5 points) Show that $\hat{\theta} = \frac{1}{\bar{X}_n}$ is a consistent estimator of $\theta = \frac{1}{\mu}$.
   2. (10 points) Show that $\sqrt{n}(\hat{\theta} - \theta) \to N(0, \frac{\sigma^2}{\mu^4})$ in distribution.
   3. (7 points) Explain how to construct an asymptotic $(1 - \alpha)$ confidence interval of θ using the result in (b).

Solution:

(a) [3pts] From the Weak Law of Large Numbers, we know

$$\overline{X}_n \xrightarrow{P} \mu.$$

[2pts] By Theorem 5.1.4, the function $g(x) = \frac{1}{x}$ is continuous at $x = \mu$, we get

$$g(\overline{X}_n) = \frac{1}{\overline{X}_n} \xrightarrow{P} g(\mu) = \frac{1}{\mu}.$$

(b) [4pts] From the Central Limit Theorem, we have

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

[3pts] Using the $\Delta-$method, and $g'(\mu) = -\frac{1}{\mu^2}$,
[3pts]

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{d} N(0, [g'(\mu)]^2\sigma^2) \equiv N(0, \frac{\sigma^2}{\mu^4})$$

(c) [3pts] Based on the distribution in (b), we have

$$P(|\frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma/\mu^2}| < z_{\alpha/2}) = 1 - \alpha$$

As σ and μ are both unknown, this can not be used directly for constructing a confidence interval. But they can be consistently estimated by $\hat{\sigma} = s$ and $\hat{\mu} = \overline{X}$.
[4pts] By Slutsky's theorem,

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{s/\overline{X}^2} \xrightarrow{d} N(0, 1)$$

So, $\hat{\theta} \pm z_{\alpha/2}\frac{s}{\overline{X}^2}\frac{1}{\sqrt{n}}$ gives an asymptotic $(1 - \alpha)$ confidence interval of θ.