# 10720 EECS 303003 Probability Quiz #5 Ans

## Section A (25%)

1. **(G1) Suppose we extract an individual at random from a population whose members have an average income of \$40,000, with a standard deviation of \$20,000. What is the probability of extracting an individual whose income is either less than \$10,000 or greater than \$70,000?**

Ans:

we can use Chebyshev's inequality to compute an upper bound to it. If $X$ denotes income, then $X$ is less than \$10,000 or greater than \$70,000 if and only if

$$|X-\mu| \geq k$$

where $\mu = 40,000$ and $k = 30,000$. The probability that this happens is:

$$\mathrm{P}(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2} = \frac{400,000,000}{900,000,000} = \frac{4}{9}$$

Therefore, the probability of extracting an individual outside the income range \$10,000-\$70,000 is less than $4/9$.

Another answer could be get from using z table, however there is no saying that this random variable follows a normal distribution, so we should be really careful when considering the case.

2. **(G2) Before starting to play the roulette in a casino, you want to look for biases that you can exploit. You therefore watch 100 rounds that result in a number between 1 and 36, and count the number of rounds for which the result is odd. If the count exceeds 55, you decide that the roulette is not fair. Find an approximation for the probability that you will decide the roulette is not fair.**

Ans:

**Solution to Problem 5.8.** Let $S$ be the number of times that the result was odd, which is a binomial random variable, with parameters $n = 100$ and $p = 0.5$, so that $\mathbf{E}[X] = 100 \cdot 0.5 = 50$ and $\sigma_S = \sqrt{100 \cdot 0.5 \cdot 0.5} = \sqrt{25} = 5$. Using the normal approximation to the binomial, we find

$$\mathbf{P}(S > 55) = \mathbf{P}\left(\frac{S-50}{5} > \frac{55-50}{5}\right) \approx 1 - \Phi(1) = 1 - 0.8413 = 0.1587.$$

A better approximation can be obtained by using the de Moivre-Laplace approximation, which yields

$$\mathbf{P}(S > 55) = \mathbf{P}(S \geq 55.5) = \mathbf{P}\left(\frac{S-50}{5} > \frac{55.5-50}{5}\right)$$
$$\approx 1 - \Phi(1.1) = 1 - 0.8643 = 0.1357.$$

3. **(G3) Assume knowing that the number of products manufactured from the factory in one week is a random variable, and its expected value of the products is 50 pieces.**

**(a) What is the tightest upper bound we can say from the number of products manufactured this week being greater than 75 pieces by Markov's inequality?**

**(b) If we know the variance of products manufactured this week is 25. What can we say about the probability of the products' amount manufactured between 40 pieces to 60 pieces?**

Ans:

(a) From Markov's inequality

$$P\{ X > 75 \} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) From Chebyshev's inequality

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

So

$$P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

The probability of the products' amount manufactured between 40 pieces to 60 pieces this week is at least 0.75.

**4. (G4) Let $X_i, i = 1, 2, 3, ..,$ be independent random variables all distributed according to the PDF $f_X(x) = x/8$ for $0 \leq x \leq 4$. Let $S = \frac{1}{100}\sum_{i=1}^{100} X_i$. Then, what's the value of P(S>3)?**

Ans:

**Solution:** Let $S = \frac{1}{100}\sum_{i=1}^{100} Y_i$ where $Y_i$ is the random variable given by $Y_i = X_1/100$. Since $Y_i$ are *iid*, the distribution of $S$ is approximately normal with mean $\mathbf{E}[S]$ and variance $\mathrm{var}(S)$. Thus, $\mathbf{P}(S > 3) = 1 - \mathbf{P}(S \leq 3) \approx 1 - \Phi\left(\frac{3 - \mathbf{E}(S)}{\sqrt{\mathrm{var}(S)}}\right)$. Now,

$$\begin{aligned}
\mathbf{E}[X_i] &= \int_0^4 x\frac{x}{8}\,dx = \left.\frac{x^3}{24}\right|_0^4 = \frac{8}{3} \\
\mathrm{var}(X_i) &= \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \int_0^4 x^2\frac{x}{8}\,dx - \left(\frac{8}{3}\right)^2 = \left.\frac{x^4}{32}\right|_0^4 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbf{E}[S] &= \frac{1}{100}\mathbf{E}[X_i] + \ldots + \frac{1}{100}\mathbf{E}[X_{100}] = 8/3. \\
\mathrm{var}(S) &= \frac{1}{100^2}\mathrm{var}(X_i) + \ldots + \frac{1}{100^2}\mathrm{var}(X_i) = \frac{8}{9} \times \frac{1}{100}.
\end{aligned}$$

and

$$\mathbf{P}(S > 3) \approx 1 - \Phi\left(\frac{3 - 8/3}{\sqrt{\frac{8}{9} \times \frac{1}{100}}}\right) = 1 - \Phi\left(\frac{5}{\sqrt{2}}\right).$$

5. (G5) A factory produces $X_n$ gadgets on day n, where the $X_n$ are independent and identically distributed random variables, with mean 5 and variance 9. Use Central Limit Theorem and the attached Gaussian table to solve the following problems.
   (a) Find an approximation to the probability that the total number of gadgets produced in 100 days is less than 440.
   (b) Find (approximately) the largest value of n such that $P(X_1 + X_2 + \cdots + X_n \geq 200 + 5n) \leq 0.05$.
   (c) Let N be the first day on which the total number of gadgets produced exceeds 1000. Calculate an approximation to the probability that $N \geq 220$.

Ans:

(a) Let $W_N = \sum_{n=1}^{N} X_n$. Then
$P[\sum_{n=1}^{100} X_n \leq 440] = P[W_{100} \leq 440] \simeq \Phi(\frac{440-100*5}{\sqrt{9*100}}) = \Phi(-2) = 1 - \Phi(2) \simeq 0.02275$

(b) $P[W_n \geq 200 + 5n] = 1 - \Phi(\frac{200+5n-5n}{3\sqrt{n}})$
$1 - \Phi(\frac{200}{3\sqrt{n}}) \leq 0.05$
$\Phi(\frac{200}{3\sqrt{n}}) \geq 0.95$
$\frac{200}{3\sqrt{n}} \geq 1.65$
$n \leq 1632$

(c) $P[N \geq 220] = P[\sum_{n=1}^{219} X_n < 1000] = \Phi(\frac{1000-219*5}{3\sqrt{219}}) = \Phi(-2.1398) \simeq 0.016$

## Section B (35%)

1. (G2) Let $\Theta$ be a positive random variable, with known mean $\mu$ and variance $\sigma^2$, to be estimated on the basis of a measurement X of the form $X = \sqrt{\Theta}W$. W is assumed independent of $\Theta$ with zero mean, unit variance, and known fourth moment $E[W^4]$. Thus, the conditional mean and variance of X given $\Theta$ are 0 and $\Theta$, respectively, trying to estimate the variance of X given an observed value. Find the linear LMS estimator of $\Theta$ based on $X^2$.

Ans:

$$\text{cov}(\Theta, X) = \mathbf{E}[\Theta^{3/2}W] - \mathbf{E}[\Theta]\,\mathbf{E}[X] = \mathbf{E}[\Theta^{/2}]\,\mathbf{E}[W] - \mathbf{E}[\Theta]\,\mathbf{E}[\Theta] = 0,$$

so the linear LMS estimator of $\Theta$ is simply $\hat{\Theta} = \mu$, and does not make use of the available observation.

Let us now consider the transformed observation $Y = X^2 = \Theta W^2$, and linear estimators of the form $\hat{\Theta} = aY + b$. We have

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[\Theta W^2] = \mathbf{E}[\Theta]\,\mathbf{E}[W^2] = \mu, \\ \mathbf{E}[\Theta Y] &= \mathbf{E}[\Theta^2 W^2] = \mathbf{E}[\Theta^2]\,\mathbf{E}[W^2] = \sigma^2 + \mu^2, \\ \text{cov}(\Theta, Y) &= \mathbf{E}[\Theta Y] - \mathbf{E}[\Theta]\,\mathbf{E}[Y] = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2, \\ \text{var}(Y) &= \mathbf{E}[\Theta^2 W^4] - \big(\mathbf{E}[Y]\big)^2 = (\sigma^2 + \mu^2)\mathbf{E}[W^4] - \mu^2. \end{aligned}$$

Thus, the linear LMS estimator of $\Theta$ based on $Y$ is of the form

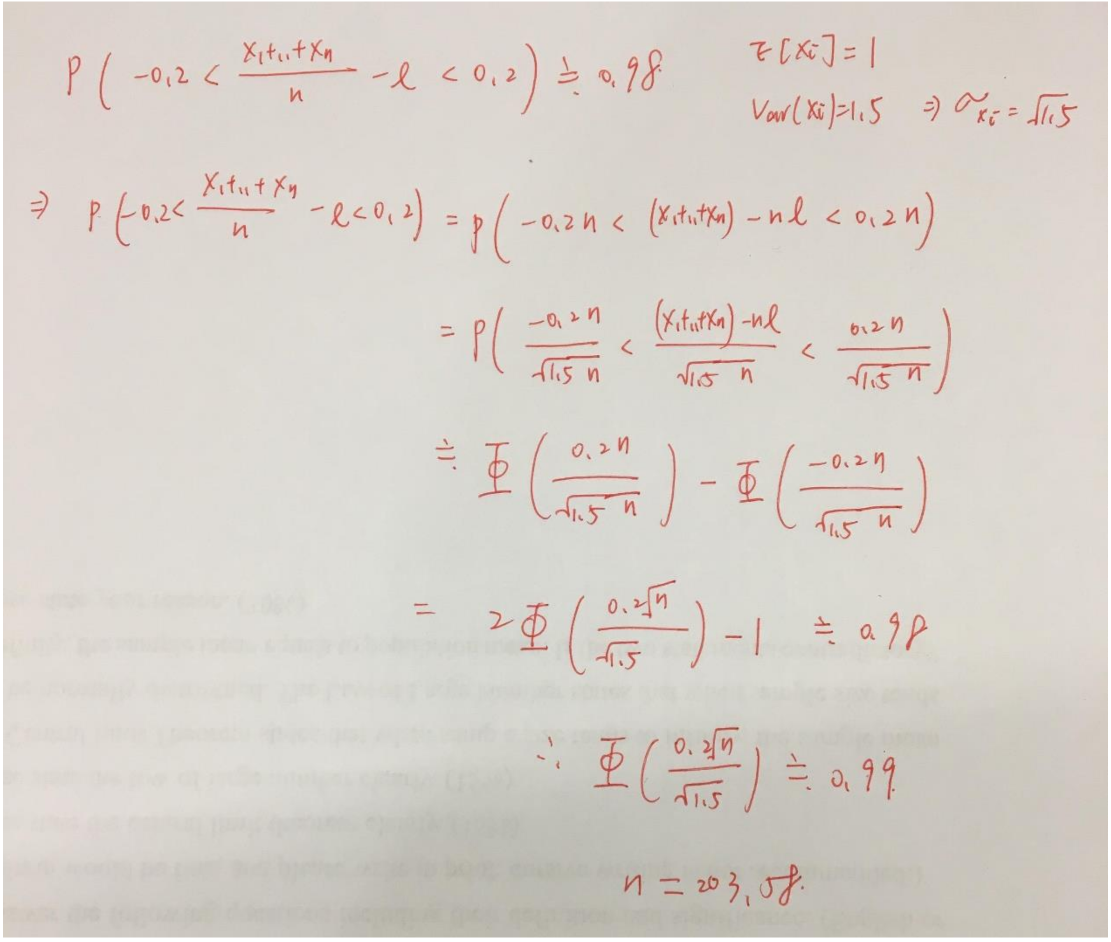
$$\hat{\Theta} = \mu + \frac{\sigma^2}{(\sigma^2 + \mu^2)\mathbf{E}[W^4] - \mu^2}\big(Y - \sigma^2\big),$$

and makes effective use of the observation: the estimate of $\Theta$, the conditional variance of $X$ becomes large whenever a large value of $X^2$ is observed.

**2. (G4) A biologist wants to estimate $\ell$, the life expectancy of a certain type of insect. To do so, he takes a simple of size $n$ and measures the lifetime from birth to death of each insect. Then he finds the average of these numbers. If he believes that the lifetimes of these insects are independent random variables with variance 1.5 days, how large a sample should he choose to be 98% sure that his average is accurate with $\pm 0.2$ ($\pm 4.8$ hours)?**

Ans:

For i=1,2,….,n, let Xi be the lifetime of the ith insect of the sample.We want to determine n, so that P(-0.2<((X1+X2+….+Xn)/n)-l<0.2)≒0.98




Therefore, the biologist should choose a sample of size 204.

Another answer could be got by applying the Chebyshev inequality, however the approximation would not be the least amount of $n$ we need. Since the problem description is not clear on specifying where the random variable will follow the normal distribution, we give credit to both the two type of answers.

**3. (G5) Let the probability density function of a random variable X be $f(x) = x^n e^{-x}/n!$, $x \geq 0$, show that P(0 < X < 2n+2) > n/(n+1).**

Ans:

by integration

E[X] = n+1 , E[X^2] =(n+2)(n+1)

var(X) = n+1

then , use chebyshev inequality

P(|X-(n+1)| < n+1) > 1 - ( var(x) / (n+1)^2 ) so, P(0 < X < 2n+2) > n/(n+1)

4. **(G6) You love to travel around Taiwan and spends \$50 for a day of travel on average (with a standard deviation of \$10). You are planning to go to Hua Lien for on the long holidays for a week. You plan to bring a total of \$385 for the whole trip. What is the probability that you will run out of money?**

Ans:

P (run out of money)

= P (total trip cost more than \$385)

= P (your average spending for a day in that trip is more than \$55)

Use sample mean with:

n = 7 (a week) $\mu = 50,\ \sigma = 10$

$$\sigma x = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{7}} = 3.78(rounded\ up)$$

55 – 50 = 5 above the mean

$$\frac{55-5}{3.78} = \frac{5}{3.78}$$
$$= 1.32\ (rounded\ up)$$

Then use Z-Table

P (your mean is more than 1.32 unit of standard deviation above mean)

= 1 - 0.90658

= 0.09342

Or 9.342% that you will need more money than you bring

5. **(G3) Suppose $X_1, X_2, X_3 \dots X_n$ is a sequence of independent identically distributed random variables. The random variables are all uniformly distributed over $[0,1]$. If the CDF of**

$$\frac{2(\sum_{i=1}^{n} X_i) - n}{\sqrt{n}}$$

**converges to the CDF of a normal distribution of mean 0 and variance $d$, namely $N(0,d)$. What is the value of $d$?**

Ans:

We first compute the mean and variance of the sequence $X_1, X_2, X_3 \dots X_n$.

Since they are uniformly distributed over $[0,1]$. The mean and variance are $\frac{1}{2}$ and $\frac{1}{12}$ respectively.

From central limit theorem, we know CDF of

$$\frac{(\sum_{i=1}^{n} X_i) - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

Converges to CDF of $N(0,1)$.

In other words,

$$\frac{(\sum_{i=1}^{n} X_i) - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \xrightarrow{d} N(0,1)$$

Simplify the expression,

$$\frac{2(\sum_{i=1}^{n} X_i) - n}{\sqrt{n}} = \frac{2}{\sqrt{12}}\left(\frac{(\sum_{i=1}^{n} X_i) - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right) \xrightarrow{d} \frac{2}{\sqrt{12}} N(0,1) = N\left(0, \frac{1}{3}\right)$$

Therefore, $d = \frac{1}{3}$

## Section C (40%)

**Please answer the following questions including their definition and significance. (English or Mandarin would be fine, and please write in print; cursive writing is not recommended.)**

1. **Please state the central limit theorem clearly. (15%)**
2. **Please state the low of large number clearly. (15%)**
3. **The Central limit Theorem states that when sample size tends to infinity, the sample mean will be normally distributed. The Law of Large Number states that when sample size tends to infinity, the sample mean equals to population mean. Is the two statements contradictory? Please state your reason. (10%)**

Ans:

1. It is saying that when an infinite number of successive random samples are taken from a population, the sampling distribution of the means of those samples will become approximately normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ as the

sample size, n, becomes larger, no matter what kind of shape of the population distribution. Therefore we can normalize it as $Z_n = \frac{\sum_{i=1}^{n} X_i - nE[X]}{\sqrt{n}\,\sigma_x}$ and use the Z table to approximate the probability, which denoted as $\mathrm{P}(Z_n < \mathrm{m}) = \Phi(m)$.

2. One we call the *weak law of large number* known as Bernoulli's theorem, states that if we got a sample of random variable (i.i.d), as the sample size grows larger, the probability that the sample means differs from the population mean by some amount *epsilon* is equal to 0. We can write down as $\lim_{\mathrm{n}\to\infty} P(|M_n - \mu| \geq \epsilon) \leq \lim_{n\to\infty} \frac{{\sigma_x}^2}{n\epsilon^2} = 0.$

   Another we call the *strong law of large number* states that supposing the first moment E[X] of X is finite. Then sampling mean converges almost surely to E[X], thus $\mathrm{P}\left(\lim_{\mathrm{n}\to\infty} M_n = E[X]\right) = 1.$

3. No.

   As the sample size tends to infinity, the sample mean will be normally distributed. As we keep increasing the sample size while decreasing the sample standard deviation, when n(sample size) close to infinity, the sample mean will converge to the population mean, which is just the statement of the law of large number.