CE93 Review 2

Solutions to Quizzes 3 and 4

Quiz 3

Question 1 (20 points)

The time between arrivals of vehicles at an intersection is an exponential random variable with mean 10 sec and variance 100 sec^2.

(a) (10 points) You were feeling bored so you recorded 100 measurements of the time between arrivals. What is the distribution of the average of the 100 measurements and its parameter(s)? Explain and show your work.

(b) (10 points) What is the probability that the average you obtain in part (a) is less than 11? Used the attached normal table at the end of the exam.

Question 1

Question 1 (20 points)

The time between arrivals of vehicles at an intersection is an exponential random variable with mean 10 sec and variance 100 sec^2.

(a) (10 points) You were feeling bored so you recorded 100 measurements of the time between arrivals. What is the distribution of the average of the 100 measurements and its parameter(s)? Explain and show your work.

From CLT, the sum of 100 independent samples from the same distribution is approximately normal. The average of these 100 measurements will have an expected value equal to the population mean – 10. The variance to the average is the population variance divided by 100 – 1. Thus: $\bar{X} \sim \mathrm{N}(\mu = 1, \sigma^2 = 1)$

b) (10 points) What is the probability that the average you obtain in part (a) is less than 11? Use the attached normal table at the end of the exam.

$$Z = \frac{11 - \mu}{\sigma} = 1 \quad P(\bar{X} \leq 11) = P(Z \leq 1) = F(Z) = .8413$$

Question 2 (15 points)

Answer the following True or False questions and briefly explain your answer in both cases.

(a) (3 points) If $Z = X_1+\ldots+X_n$ and $X_1, \ldots, X_n$ are independent, then $\sigma_Z^2 = \sigma_{X1}^2 + \ldots + \sigma_{Xn}^2$.

□ **True** □ **False**

Explain using a few words or equations.

(b) (3 points) If W and L are independent random variables, then $Cov(L, W) = 0$.

□ **True** □ **False**

Explain using a few words or equations.

(c) (3 points) Let X and Y be independent lognormal random variables. The product XY also has a lognormal distribution.

□ **True** □ **False**

Explain using a few words or equations.

(d) (3 points) For any random variable, $E(X^2) \geq [E(X)]^2$.

□ **True** □ **False**

Explain using a few words or equations.

Question 2 (15 points)

Answer the following True or False questions and briefly explain your answer in both cases.

(a) (3 points) If $Z = X_1 + \ldots + X_n$ and $X_1, \ldots, X_n$ are independent, then $\sigma_Z^2 = \sigma_{X1}^2 + \ldots + \sigma_{Xn}^2$.

X True ☐ False

Explain using a few words or equations.

Variance of the sum is the sum of the variances.

(b) (3 points) If W and L are independent random variables, then $Cov(L, W) = 0$.

X True ☐ False

Explain using a few words or equations.

Independent random variables have 0 covariance.

(c) (3 points) Let X and Y be independent lognormal random variables. The product XY also has a lognormal distribution.

X True ☐ False

Explain using a few words or equations.

log(XY)=log(X)+log(Y). log(X) and log(Y) are normally distributed and the sum of normally distributed variables if also normally distributed. So log(XY) is normally distributed and XY is lognormally distributed.

(d) (3 points) For any random variable, $E(X^2) \geq [E(X)]^2$.

X True ☐ False

Explain using a few words or equations.

$$var(X) = E(X^2) - [E(X)]^2 \geq 0$$

Question 3 (10 points)

You have two coins. One is a fair coin so that a coin flip has a 50% probability of being a head or a tail. The other is an unfair coin for which a coin flip has a 90% chance of being a head. You are not sure which coin is fair and which is unfair. You pick one of the coins and toss it 3 times, obtaining 2 heads and 1 tail. Using the concept of maximum likelihood, determine whether the coin you tossed is more likely to be the fair one or the unfair one. Show your calculations. *Hint: Do not estimate the probability of a head from the data. Instead compare the likelihood of the observed data if the coin is fair and if the coin is unfair.*

Question 3 (10 points)

You have two coins. One is a fair coin so that a coin flip has a 50% probability of being a head or a tail. The other is an unfair coin for which a coin flip has a 90% chance of being a head. You are not sure which coin is fair and which is unfair. You pick one of the coins and toss it 3 times, obtaining 2 heads and 1 tail. Using the concept of maximum likelihood, determine whether the coin you tossed is more likely to be the fair one or the unfair one. Show your calculations. *Hint: Do not estimate the probability of a head from the data. Instead compare the likelihood of the observed data if the coin is fair and if the coin is unfair.*

$likelihood(P) = 3P^2(1-P)$

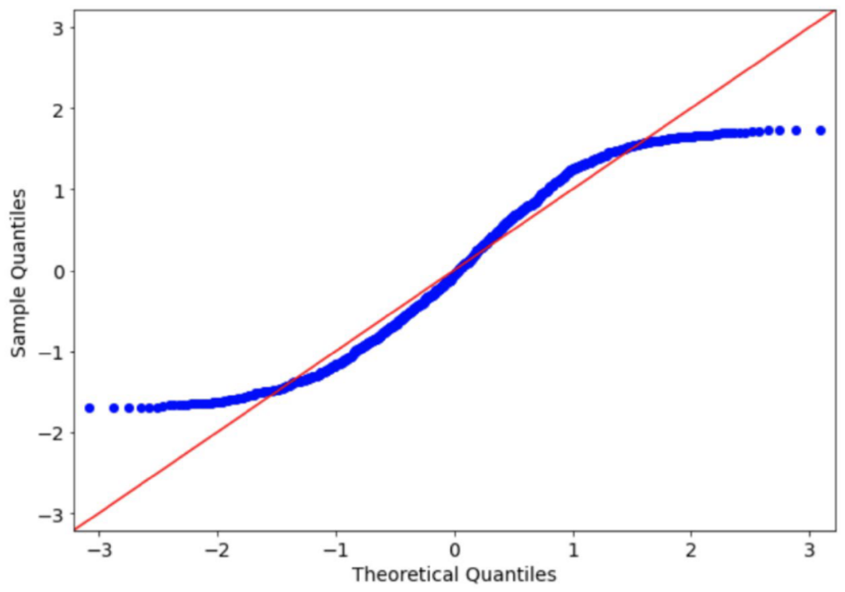
$likelihood(.5) = 3 \cdot 0.5^2\ (1-0.5)$=$\frac{3}{8}$=.375

$likelihood(.9) = 3 \cdot 0.9^2\ (1-0.9)$=$3 \cdot 0.81 = .243$

More likely to be fair coin.

Question 4 (5 points)

You think some data may come from a normal distribution and construct a Q-Q plot to see if this is true. The Q-Q plot you obtain is shown below. Based on the plot, is it likely that the data comes from a normal distribution? Explain.



If that data follows a normal distribution than the sample quantiles should roughly equal the theoretical quantiles. In this case the sample quantiles diverge from the theoretical quantiles. The data is unlikely to come from a normal distribution.

Quiz 4

Question 1

Daily food intake, in calories, of two randomly selected Berkeley graduate students and two randomly selected Berkeley undergraduate students is measured. Calories for the graduate students are 2300 and 2500, while for undergraduates the results are 2200 and 2400.

(a) Assume that the calorie intake of both student populations is known to have a standard deviation of 100.

   i. What is the 90% confidence interval for the food intake difference between graduates and undergraduates?

   ii. Test the null hypothesis that graduates and undergraduates have the same average food intake, against the alternative hypothesis that they are not the same, at the 0.2 level of significance.

   iii. What is the p-value for the test in (a)ii.?

   iv. What is the p-value for the test that graduates and undergraduates have the same average food intake, against the alternative hypothesis that graduates eat more?

(b) Assume that the calorie intake standard deviation of both student populations is unknown, but has the same value.

   i. What is the 90% confidence interval for the food intake difference between graduates and undergraduates?

   ii. Test the null hypothesis that graduates and undergraduates have the same average food intake.

   iii. What is the p-value for the test in (b)ii.?

(c) What is the 95% confidence interval for the variance of graduate student food intake?

Daily food intake, in calories, of two randomly selected Berkeley graduate students and two randomly selected Berkeley undergraduate students is measured. Calories for the graduate students are 2300 and 2500, while for undergraduates the results are 2200 and 2400.

(a) Assume that the calorie intake of both student populations is known to have a standard deviation of 100.

i. What is the 90% confidence interval for the food intake difference between graduates and undergraduates?

$$\overline{G} = 2400 \quad \sigma_G = 100$$

$$\overline{U} = 2300 \quad \sigma_U = 100$$

$$\overline{G} - \overline{U} \sim N\left(\mu_G - \mu_U, \sqrt{\frac{\sigma_G^2}{2} + \frac{\sigma_U^2}{2}}\right) \quad \overline{G} - \overline{U} \sim N\left(\mu_G - \mu_U, \sqrt{\frac{10000}{2} + \frac{10000}{2}}\right)$$

$$\overline{G} - \overline{U} \sim N\left(\mu_G - \mu_U, 100\right)$$

$$\frac{(\overline{G} - \overline{U}) - (\mu_G - \mu_U)}{100} \sim N(0,1) \quad \frac{100 - (\mu_G - \mu_U)}{100} \sim N(0,1)$$

$$P(-1.65 \le \frac{100 - (\mu_G - \mu_U)}{100} \le 1.65) = 0.9$$

$$P(-165 \le 100 - (\mu_G - \mu_U) \le 165) = 0.9$$

$$P(-265 \le -(\mu_G - \mu_U) \le 65) = 0.9$$

$$P(-65 \le (\mu_G - \mu_U) \le 265) = 0.9 \Rightarrow 90\% \text{ Confidence Interval for } \mu_G - \mu_U \text{ is } [-65, 265]$$

Daily food intake, in calories, of two randomly selected Berkeley graduate students and two randomly selected Berkeley undergraduate students is measured. Calories for the graduate students are 2300 and 2500, while for undergraduates the results are 2200 and 2400.

(a) Assume that the calorie intake of both student populations is known to have a standard deviation of 100.

  ii. Test the null hypothesis that graduates and undergraduates have the same average food intake, against the alternative hypothesis that they are not the same, at the 0.2 level of significance.

  iii. What is the p-value for the test in (a)ii.?

  iv. What is the p-value for the test that graduates and undergraduates have the same average food intake, against the alternative hypothesis that graduates eat more?

$$\text{ii) } H0: \mu_G - \mu_U = 0 \quad H1: \mu_G - \mu_U \neq 0$$

$$\text{Under } H0, \frac{\bar{G}-\bar{U}}{100} \sim N(0,1)$$

$$p = 0.2 \Rightarrow \text{Rejection region is } \frac{\bar{G}-\bar{U}}{100} \leq -1.28 \text{ and } \frac{\bar{G}-\bar{U}}{100} \geq 1.28$$

$$\frac{\bar{G}-\bar{U}}{100} = 1 \Rightarrow \text{fail to reject } H0$$

$$\text{iii) } p - value = 1 - \left(\Phi(1) - \Phi(-1)\right) = .3174$$

$$\text{iv) } p - value = 1 - \Phi(1) = .1587$$

Daily food intake, in calories, of two randomly selected Berkeley graduate students and two randomly selected Berkeley undergraduate students is measured. Calories for the graduate students are 2300 and 2500, while for undergraduates the results are 2200 and 2400.

b) Assume that the calorie intake standard deviation of both student populations is unknown, but has the same value.
   i. What is the 90% confidence interval for the food intake difference between graduates and undergraduates?
   ii. Test the null hypothesis that graduates and undergraduates have the same average food intake.
   iii. What is the p-value for the test in (b)ii.?

i. 90% Confidence Interval for $\mu_G - \mu_U = (\bar{G} - \bar{U}) \pm t_{4-2,0.05} \cdot s_p \sqrt{\frac{1}{2} + \frac{1}{2}}$

$$s_p = \sqrt{\frac{1 \cdot 10000 + 1 \cdot 10000}{2}} = 100 \quad t_{2,0.05} = 2.92$$

90% Confidence Interval for $\mu_G - \mu_U = 100 \pm 292$

ii. Since 90% confidence interval for for $\mu_G - \mu_U = 100 \pm 292 = 100 \pm 292$ includes 0,

we cannot reject the null hypothesis that $\mu_G - \mu_U = 0$ against $\mu_G - \mu_U \neq 0$ at 0.1 level.

iii. $t = \dfrac{\bar{G} - \bar{U}}{s_p \sqrt{\frac{1}{2} + \frac{1}{2}}} = \dfrac{100}{100} = 1 \quad t_{2,0.25} = 0.816 \quad t_{2,0.1} = 1.886 \Rightarrow$ p-value is between 0.5 and 0.2

Question 1 (30 points)

Four randomly selected Berkeley students are timed in the 100-yard dash. Their times are 11, 12, 12, and 13 seconds.

(a) (5 points each, 10 points total) Suppose you know that the variance of 100-yard dash times is 1 sec^2.
   i. What is the 90% confidence interval for the average Berkeley student 100-yard dash time?
   ii. What is the 95% upper confidence interval for this average?

(b) (5 points each, 20 points total) Suppose you don't know the variance of the Berkeley student 100-yard dash times but must instead estimate it from the data.
   i. What is the 95% confidence interval for the average Berkeley student 100-yard dash time?
   ii. What is the 95% confidence interval for the variance of the Berkelety student 100-yard dash times.
   iii. Test the null hypothesis that that the average Berkeley 100-yard dash time is greater than or equal to 12 seconds, against the alternative hypothesis that the average time is less than 12 seconds, at the significance level of .05.
   iv. Is the p-value for the hypothesis test in iii. greater than .05 or less than .0.5? Explain your reasoning.

Hint: $\sqrt{2/3} = 0.82$

Question 1 (30 points)

Four randomly selected Berkeley students are timed in the 100-yard dash. Their times are 11, 12, 12, and 13 seconds.

(a) (5 points each, 10 points total) Suppose you know that the variance of 100-yard dash times is 1 sec^2.

i. What is the 90% confidence interval for the average Berkeley student 100-yard dash time?

ii. What is the 95% upper confidence interval for this average?

i) 90% confidence interval is $\overline{X} \pm z_{0.05}\sigma_{\overline{X}}$

$\sigma_{\overline{X}}=\frac{1}{\sqrt{4}}=\frac{1}{2}$ $\overline{X}$ =12 $z_{0.05}$=1.65

$\overline{X} \pm z_{0.05}\sigma_{\overline{X}}$=12$\pm$0.825

ii) 95% upper confidence interval is $[\overline{X} - z_{0.05}\sigma_{\overline{X}},\infty]$ = [12-0.825,∞]

b) (5 points each, 20 points total) Suppose you don't know the variance of the Berkeley student 100-yard dash times but must instead estimate it from the data.

i. What is the 95% confidence interval for the average Berkeley student 100-yard dash time?

ii. What is the 95% confidence interval for the variance of the Berkeley student 100-yard dash times.

iii. Test the null hypothesis that that the average Berkeley 100-yard dash time is greater than or equal to 12 seconds, against the alternative hypothesis that the average time is less than 12 seconds, at the significance level of .05.

iv. Is the p-value for the hypothesis test in iii. greater than .05 or less than .05? Explain your reasoning.

i) 95% confidence interval is $\overline{X} \pm t_{3,0.025}\frac{s_x}{4}$

$t_{3,0.025}$=3.182 $\frac{s_x}{4}=\frac{\sqrt{\frac{2}{3}}}{4}=\sqrt{\frac{1}{24}}$

95% confidence interval is $12 \pm \frac{3.182}{\sqrt{24}}$

ii) 95% confidence interval is $\left[\frac{(n-1)s^2}{\chi^2_{n-1,0.025}}, \frac{(n-1)s^2}{\chi^2_{n-1,0.975}}\right] = \left[\frac{3\cdot\frac{2}{3}}{9.348}, \frac{3\cdot\frac{2}{3}}{0.216}\right]$

iii) Test statistic is $\frac{\overline{X}-12}{s}$=0. Rejection region is <-1.65. Do not reject.

iv) Do not reject at .05 means p-value is greater than 0.05.

Question 2 (20 points)

You toss a coin 400 times and get a head 240 times.

(a) (4 points) What is the MLE for the probability that a coin toss will yield a head?

(b) (4 points) What is the 95% confidence interval for the probability that the coin toss will yield a head?

(c) (4 points) Test, at a .05, significance level, the null hypothesis that the coin has a 0.5 probability of yielding a head.

(d) (4 points) Repeat (b) if you tossed the coin 100 times and got a head 60 times.

(e) (4 points) Repeat (c) if you tossed the coin 25 times and got a head 15 times.

Question 2 (20 points)

You toss a coin 400 times and get a head 240 times.

(a) (4 points) What is the MLE for the probability that a coin toss will yield a head?
(b) (4 points) What is the 95% confidence interval for the probability that the coin toss will yield a head?
(c) (4 points) Test, at a .05, significance level, the null hypothesis that the coin has a 0.5 probability of yielding a head.
(d) (4 points) Repeat (b) if you tossed the coin 100 times and got a head 60 times.
(e) (4 points) Repeat (c) if you tossed the coin 25 times and got a head 15 times.

a) $\hat{p}=\frac{k}{n}=\frac{240}{400}=0.6$

b) $\tilde{n}=404$ $\tilde{p}=\frac{242}{404}=.599$ $z_{0.025}=1.96$ 95%CI is $\tilde{p}\pm 1.96\sqrt{.599(1-.599)}$

c) Test statistic is $\frac{0.6-0.5}{\sqrt{0.5(1-0.5)/400}}$ which has standard normal distribution. Rejection region is >1.96 and <-1.96.

d) Test statistic is $\frac{0.6-0.5}{\sqrt{0.5(1-0.5)/100}}$ which has standard normal distribution. Rejection region is >1.96 and <-1.96.

e) Test statistic is $\frac{0.6-0.5}{\sqrt{0.5(1-0.5)/25}}$ which has standard normal distribution. Rejection region is >1.96 and <-1.96.