Final test contents:

Problem 1: Confidence interval for an unknown population proportion; sample size calculation

Problem 2: CI for population mean; sample size calculation

Problem 3: Bivariate discrete distribution
- Marginal distribution
- Marginal distribution mean
- Marginal distribution variance
- Covariance
- Correlation
- Independence

Problem 4: Expected value of $C_1X_1 + C_2X_2$
Variance of $C_1X_1 + C_2X_2$

Problem 5: Distribution of $\overline{X}$ (random sample mean) using central limit theorem
$E(\overline{X})$
$Var(\overline{X})$
Percentiles of Normal distribution

Problem 6: Maximum likelihood estimator derivation
Bias. Mean squared error (Bonus)
Central limit theorem.

Exercise set for chapter 5-8

1 Chapter 5

Problem 1. Two random variables have a joint probability mass function in Table 1. Show that the correlation between X and Y is zero but X and Y are not independent.

Table 1: Joint probability mass function

x	-1	0	0	1
y	0	-1	1	0
$f_{X,Y}(x,y)$	1/4	1/4	1/4	1/4

① Marginal distribution of X

x	-1	0	1
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Marginal distribution of Y

y	-1	0	1
$f(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

② $E(X) = (-1)\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} = 0$

$E(Y) = 0$

$Var(X) = E[X^2] - 0^2 = (-1)^2\cdot\frac{1}{4} + 0^2\cdot\frac{1}{2} + 1^2\cdot\frac{1}{4} = \frac{1}{2}$

$Var(Y) = \frac{1}{2}$

③ $E(XY) = (-1)\cdot 0\cdot\frac{1}{4} + 0\cdot(-1)\cdot\frac{1}{4} + 0\cdot 1\cdot\frac{1}{4} + 1\cdot 0\cdot\frac{1}{4} = 0$

④ $Cov(X,Y) = E(XY) - E(X)\cdot E(Y) = 0$

⑤ $Corr(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} = 0$

$f_{X,Y}(1,1) = 0$

$f_X(1) = \frac{1}{4}$

$f_Y(1) = \frac{1}{4}$

$f_{X,Y}(1,1) \neq f_X(1)\,f_Y(1)$

Independence of X and Y $\iff$ $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y.

X and Y are not independent

Problem 2. A random variable X has the probability distribution $f_X(x) = x/18$, $0 \le x \le 6$. Determine the following:

(a) The expected value of X. $= \int_0^6 x\cdot\frac{x}{18}\,dx = \frac{x^3}{54}\Big|_0^6 = \frac{36}{9} = 4$

(b) The variance of X. $= E[X^2] - (E[X])^2 = \int_0^6 x^2\cdot\frac{x}{18}\,dx - 6^2$

$= \frac{x^4}{72}\Big|_0^6 - 16 = 18 - 16 = 2$

(c) The cumulative distribution of X.

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x u/18\,du = \frac{u^2}{36}\Big|_0^x = \frac{x^2}{36} & 0 \le x \le 6 \\ 1 & x > 6 \end{cases}$$

① $E(C_1 X_1 + C_2 X_2) = C_1 \cdot E(X_1) + C_2 E(X_2)$

② $Var(C_1 X_1 + C_2 X_2) = C_1^2 \cdot Var(X_1) + C_2^2 Var(X_2) + 2 \cdot C_1 \cdot C_2 \, Cov(X_1, X_2)$

Problem 3. X and Y are independent, normal random variables with $E(X) = 2$, $Var(X) = 5$, $E(Y) = 6$, $Var(Y) = 8$.

(a) $E(3X + 2Y)$. $= 3 \cdot E(X) + 2 \cdot E(Y) = 3 \cdot 2 + 2 \cdot 6 = 18$

(b) $Var(3X + 2Y)$. $= 3^2 \cdot Var(X) + 2^2 \cdot Var(Y) = 3^2 \cdot 5 + 2^2 \cdot 8 = 45 + 32 = 77$

(c) Note that the sum of two normal random variables is still a normal random variable. What is the distribution of $3X + 2Y$? Please specify its mean and variance.

$3X + 2Y \sim N(18, 77)$

2 Chapter 6

range for sample1 : 4

Problem 4. Consider the following two samples: Sample 1: 10, 9, 8, 7, 8, 6, 10, 6; Sample 2: 10, 6, 10, 6, 8, 10, 8, 6.

range for sample 2 : 4

sd_1 : 1.6

sd_2 : 1.85

(a) Calculate the sample range for both samples. Would you conclude that both samples exhibit the same variability? Explain.

(b) Calculate the sample standard deviations for both samples. Do these quantities indicate that both samples have the same variability? Explain.

3 Chapter 7

Problem 5. Describe the population/distribution parameter that the following numerical summaries trying to estimate.

(a) Sample mean $\xrightarrow{\text{Estimate}}$ Population mean

(b) Sample variance $\xrightarrow{\text{Estimate}}$ Population variance

(c) Sample proportion $\xrightarrow{\text{Estimate}}$ Population proportion

$\hat{\lambda} = \bar{x} \xrightarrow{\text{Estimate}} \lambda$

Problem 6. A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size $n_2 = 9$ is taken from another normal population with mean 70 and standard deviation 12. Let $\bar{X}_1$ and $\bar{X}_2$ be the two sample means. Note the difference of two normal random variables is still a normal random variable. Find:

(a) The probability that $\bar{X}_1 - \bar{X}_2$ exceeds 4.

(b) The 95th percentile of the distribution for $\bar{X}_1 - \bar{X}_2$.

$\bar{X}_1$ follows a Normal distribution $N(75, \frac{8^2}{16})$

$\bar{X}_2$ follows a Normal distribution $N(70, \frac{12^2}{9})$

$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = 5$

(Assume the two samples are drawn independently) $Var(\bar{X}_1 - \bar{X}_2) = 4 + 16 = 20$

$\bar{X}_1 - \bar{X}_2 \sim N(5, 20)$

a) $P(\bar{X}_1 - \bar{X}_2 > 4) = 1 - P(\bar{X}_1 - \bar{X}_2 \leq 4) = 1 - \Phi_{0,1}\left(\frac{4-5}{\sqrt{20}}\right)$

$= 1 - \Phi_{0,1}(0.224)$

$= 1 - 0.412$

$= 0.588$

b) $\mu + z_{0.95} \cdot \sigma$

$z_{0.95} \rightarrow \Phi^{-1}_{0,1}(0.95)$

$z_{0.95} \rightarrow$ 95th percentile of standard Normal distribution

$= 5 + 1.64 \cdot \sqrt{20}$

$= 12.37$

Problem 7. Suppose we have a random sample of size $2n$ from a population denoted by X, and $E(X) = \mu$, $Var(X) = \sigma^2$. Let $\bar{X}_1 = \frac{1}{2n}\sum_{i=1}^{2n} X_i$ and $\bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$ be two estimators of μ. Which is the better estimator of μ? Hint, compare the bias and the MSE of the two estimators.

Bias for $\bar{X}_1$: $E(\bar{X}_1) - \mu = \mu - \mu = 0$

Bias for $\bar{X}_2$: $E(\bar{X}_2) - \mu = \mu - \mu = 0$

$E\left(\frac{1}{2n}\sum_{i=1}^{2n} x_i\right) = \frac{1}{2n} \cdot \sum_{i=1}^{2n} E(x_i) = \frac{1}{2n} \cdot E(X_1) + \frac{1}{2n} E(X_2) + \cdots \frac{1}{2n} E(X_{2n})$

$= \frac{1}{2n} \cdot 2n \cdot \mu = \mu$

MSE for $\bar{X}_1$: $E(\bar{X}_1 - \mu)^2 = Var(\bar{X}_1) = \frac{\sigma^2}{2n}$

MSE for $\bar{X}_2$: $E(\bar{X}_2 - \mu)^2 = Var(\bar{X}_2) = \frac{\sigma^2}{n}$

$\bar{X}_1$ is better than $\bar{X}_2$ in terms of MSE.

4 Chapter 8

An article in the Journal of the American Statistical Association ["Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling" (1990, Vol. 85, pp. 972–985)] measured the weight of 30 rats under experiment controls. Suppose that 12 were underweight rats.

(a) Calculate a 95% two-sided confidence interval on the true proportion of rats that would show underweight from the experiment.

$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\hat{p} = \frac{12}{30} = 0.4, \quad n = 30$

$0.4 \pm 1.96\sqrt{\frac{0.4 \cdot 0.6}{30}}$

$\rightarrow [0.225, 0.575]$

(b) Using the point estimate of p obtained from the preliminary sample, what sample size is needed to be 95% confident that the error in estimating the true value of p is less than 0.02?

(c) How large must the sample be if you wish to be at least 95% confident that the error in estimating p is less than 0.02, regardless of the true value of p?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right) \cdot \underline{p(1-p)}$$

$z_{\alpha/2} \rightarrow \Phi^{-1}_{0,1}(1-\frac{\alpha}{2})$

$E \rightarrow$ Margin of error

$p(1-p)$ reaches maximum at $p=0.5$

b) $\hat{p} = 0.4$

$$n \geq \left(\frac{1.96}{0.02}\right)^2 \cdot 0.4 \cdot 0.6$$

≥ 2304.96

c) $$n \geq \left(\frac{1.96}{0.02}\right)^2 \cdot 0.5 \cdot 0.5$$

≥ 2401