Midterm practice problem set

1

Probabilities of events

Problem 1.1

Note: A' is notation for 'not A'.

A digital scale is used to provide weight measurements to the nearest gram. Let X denote the measurement of a sample observation using this scale. Three sets or intervals of measurement values are defined:

A: weight exceeds 20 grams

B: weight is less than or equal to 15 grams

C: weight is greater than 15 grams and less than 24 grams

Based on passed experience, the following probabilities have been determined:

$$P(X \in A) = 0.5$$
$$P(X \in B) = 0.3$$
$$P(X \in C) = 0.6$$

(a) Are A and B mutually exclusive? Are B and C? Are A and C?

(b) Describe A' and find $P(X \in A')$

(c) Describe C' and find $P(X \in C')$

(d) Find $P(15 < X \leq 20)$

Problem 1.2

The probability that a can of soft drink contains at least 12 fluid ounces is 0.90. Suppose that 10 cans are measured and that they are independent. Determine the following probabilities.

(a) All 10 cans contain at least 12 fluid ounces.

(b) No cans contain at least 12 fluid ounces.

Problem 1.3

The cycle time required to manufacture a complex engine part can vary depending on many factors, including tool wear and temperature of part. Three sets or intervals of time are defined:

A: Less than 8 hours

B: More than 7 hours but less than or equal to 7.5 hours

C: More than 7.5 hours

Production records indicate that 85% of the parts required less than 8 hours to machine, 25% required more than 7 hours but less than or equal to 7.5 hours and 35% required more than 7.5 hours.

(a) Are *A* and *B* mutually exclusive? Are *B* and *C*? Are *A* and *C*?

(b) What is the probability that a part requires 8 or more hours to machine?

(c) What is the probability that a part requires 7.5 hours or less?

(d) Suppose that the time interval of set *B* is determined to be optimum; what percentage of parts do not conform to this optimum cycle time?

Problem 1.4

Let X denote the life of a semiconductor laser (in hours) with the following probabilities:

$$P(X \leq 5000) = 0.05$$
$$P(5000 < X \leq 7000) = 0.5$$
$$P(X > 7000) = 0.45$$

(a) What is the probability that the life is less than or equal to 7000 hours?

(b) What is the probability that the life is greater than 5000 hours?

Problem 1.5

Suppose that women obtain 54% of all bachelor degrees in a particular country and that 20% of all bachelor degrees are in business.

Also, 8% of all bachelor degrees go to women majoring in business.

Are the events “The bachelor degree holder is a woman” and “The bachelor degree is in business” statistically independent?

Problem 1.6

Staff Inc., a management consulting company, is surveying the personnel of Acme Ltd. It determined that 35% of the analysts have an MBA and that 40% of all analysts are over age 35. Further, of those who have an MBA, 30% are over age 35.

a. What is the probability that a randomly chosen analyst both has an MBA and also is over age 35?
b. What is the probability that a randomly chosen analyst who is over age 35 has an MBA?
c. What is the probability that a randomly chosen analyst has an MBA or is over age 35?
d. What is the probability that a randomly chosen analyst who is over age 35 does not have an MBA?
e. e. Are the events MBA and over age 35 independent?

Problem 1.7

It is known that 20% of all farms in a state exceed 160 acres and that 60% of all farms in that state are owned by persons over 50 years old. Of all farms in the state exceeding 160 acres, 55% are owned by persons over 50 years old.

a. What is the probability that a randomly chosen farm in this state both exceeds 160 acres and is owned by a person over 50 years old?
b. What is the probability that a farm in this state either is bigger than 160 acres or is owned by a person over 50 years old (or both) ?
c. What is the probability that a farm in this state, owned by a person over 50 years old, exceeds 160 acres?
d. Are size of farm and age of owner in this state statistically independent?

Problem 1.8

A record store owner assesses customers entering the store as high school age, college age, or older, and finds that of all customers 30%, 50%, and 20% respectively, fall into these categories. The owner also found that purchases were made by 20% of high school age customers, by 60% of college age customers, and by 80% of older customers.

a. What is the probability that a randomly chosen customer entering the store will make a purchase?

Problem 1.9

Manufacturer workers are working at both textile factory and assembly plant. 60% work at textile factory and 40% at assembly plant. Last year 35% of textile worker got injury and 20% of assembly worker got injury.

a. What will be the percentage of workers who didn't get injury during this period?

b. If one worker got injury. What will be the percentage that he is a worker of assembly plant?

2

Univariate pdfs

Problem 2.1

Note: You will not have to do this sort of manipulation of CDFs on the test. Still it is a good exercise to test/improve your understanding.

2. A certain river floods every year. Suppose that the low-water mark is set at 1 and the high-water mark Y has distribution function

Note: $F_Y(y)$ is the cdf $\Phi_Y(y)$

$$F_Y(y) = 1 - \frac{1}{y^2}, \quad 1 \leq y < \infty.$$

(a) Verify that $F_Y(y)$ is a probability distribution function.

(b) Find $f_Y(y)$, the pdf of Y.

(c) If the low-water mark is reset at 0 and we use a unit of measurement that is $\frac{1}{10}$ of that given previously, the high-water mark become $Z = 10(Y-1)$. Find $F_Z(z)$.

Problem 2.2

Note: There will be no proofs on the test, but this is a good exercise for testing/improving your understanding.

5. Given a random variable X with pdf $f_X = 1$ for $0 < x < 1$ and any two points a_1, a_2 in the interval $(0, 1)$, such that $a_1 < a_2$ and $a_1 + a_2 \leq 1$,

(a) show that

$$P[a_1 < X < (a_1 + a_2)] = a_2.$$

(b) In general, if $f(x)$ is uniform in the interval (a, b), and if $a \leq a_1$, $a_1 < a_2$, and $a_1 + a_2 \leq b$, show that

$$P[a_1 < X < (a_1 + a_2)] = \frac{a_2}{(b - a)}.$$

There will not be proofs in the midterm, but this is a good exercise for testing/improving your understanding.

Problem 2.3

- A filling machine is set to pour 500 g-s of cereal into a box container. Denote the actual weight of cereal filled into the container by X, and assume that $X \sim N\left(500, 20^2\right)$.
- A random sample of $n = 25$ boxes is selected (i.e. $x_1, \ldots, x_{25}$ is drawn), and the plant manager stops the process if $\overline{x} > 510$ or $\overline{x} < 490$.
- What is the probability of stopping?

Problem 2.4

The life of a hydraulic pump is exponentially distributed with a mean of 10 years.
a) What is the probability that the pump lasts more than 8 years?
b) After how many years would you expect 90% of the pumps to have failed?

3

Expected value and variance of Univariate random variables

Problem 3.1

Suppose you make a $1,000 investment in a risky venture. There is a 40% chance that the payoff from the investment will be $2,000, a 50% chance that you will just get your money back, and a 10% chance that you will receive nothing at all from your investment. Find the expected value and standard deviation of the payoff from your investment of $1,000.

Problem 3.2

An internal combustion engine contains several cylinders bored into the engine block. Let X represent the bore diameter of a cylinder, in millimeters. Assume that the probability density function of X is

$$f(x) = \begin{cases} 10 & 80.5 < x < 80.6 \\ 0 & \text{otherwise} \end{cases}$$

Let $A = \pi X^2/4$ represent the area of the bore. Find the mean of A.

4

Mutlivariate RVs:
Conditional, marginal, correlation

Problem 4.1

The following distributions of *X* and *Y* have been developed. If *X* and *Y* are independent, determine the joint probability distribution of *X* and *Y* .

x	0	1	2
P(x)	0.6	0.3	0.1

Y	1	2
P(y)	0.7	0.3

Problem 4.2

Consider the following bivariate distribution of X and Y.

		X	
		10	20
y	1	0.4	0.2
	2		0.3

a. Determine the mean, variance, and standard deviation for X and for Y
b. Compute the covariance between X and Y, and determine if X and Y are independent.
c. Compute the coefficient of correlation.

Problem 4.3

1. Consider a random variable vector (X, Y) with joint pdf

$$f(x,y) = \begin{cases} e^{-y}, & 0 < x < y < \infty \\ 0, & otherwise \end{cases}.$$

~~(a) Compute $P(X + Y \geq 1)$.~~

(b) Find the marginal pdfs f_X and f_Y.

Problem 4.4

The displacement of a piston in an internal combustion engine is defined to be the volume that the top of the piston moves through from the top to the bottom of its stroke. Let X represent the diameter of the cylinder bore, in millimeters, and let Y represent the length of the piston stroke in millimeters. The displacement is given by $D = \pi X^2 Y/4$. Assume X and Y are jointly distributed with joint probability mass function

$$f(x,y) = \begin{cases} 100 & 80.5 < x < 80.6 \text{ and } 65.1 < y < 65.2 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean of D.

5

Point estimation

Problem 5.1

6.9 Shear-strength measurements for spot welds of a certain type have been found to have a standard deviation of approximately 10 psi. If 100 test welds are to be measured, find the approximate probability that the sample means will be within 1 psi of the true population mean.

Problem 5.2

This should say “95% of soils have pH in the range [5,8].”

6.11 The soil acidity is measured by a quantity called the pH, which may range from 0 to 14 for soils ranging from low to high acidity. ~~Many soils have an average pH in the 5 to 8 range.~~ A scientist wants to estimate the average pH for a large field from n randomly selected core samples by measuring the pH in each sample. If the scientist selects $n = 40$ samples, find the approximate probability that the sample mean of the 40 pH measurements will be within 0.2 unit of the true average pH for the field.

Problem 5.3

6.13 Resistors of a certain type have resistances that average 200 ohms with a standard deviation of 10 ohms. Twenty-five of these resistors are to be used in a circuit.

a Find the probability that the average resistance of the 25 resistors is between 199 and 202 ohms.

b Find the probability that the *total* resistance of the 25 resistors does not exceed 5,100 ohms. [*Hint*: Note that $P(\sum_{i=1}^{n} X_i > a) = P(n\overline{X} > a) = P(\overline{X} > a/n)$.]

c What assumptions are necessary for the answers in (a) and (b) to be good approximations?

Problem 5.4

6.15 Unaltered bitumens, as commonly found in lead-zinc deposits, have atomic hydrogen/carbon (H/C) ratios that average 1.4 with a standard deviation of 0.05. Find the probability that 25 samples of bitumen have an average H/C ratio below 1.3.

Problem 5.5

6.17 The strength of a thread is a random variable with mean 0.5 pound and standard deviation 0.2 pound. Assume the strength of a rope is the sum of the strengths of the threads in the rope.

- **a** Find the probability that a rope consisting of 100 threads will hold 45 pounds.
- **b** How many threads are needed for a rope that will hold 50 pounds with 99% assurance?

Problem 5.6

6.19 The service times for customers coming through a checkout counter in a retail store are independent random variables with a mean of 1.5 minutes and a variance of 1.0. Approximate the probability that 100 customers can be serviced in less than 2 hours of total service time by this one checkout counter.

6

Confidence intervals

Problem 6.1

7.12 An Alabama swimmer swims the 100-meter free style in times that are normally distributed, with a variance of $\sigma^2 = 0.50$ seconds. A random sample of $n = 4$ trials yields the following results: 51, 52, 50, 51 seconds.

a) Find an unbiased estimate of μ.

b) Construct a 95-percent confidence interval for μ.

Problem 6.2

7.16 A recent survey asked respondents to rate, on a scale from 0 to 100, how good a job they thought the President of the United States had done during the past six months. Assume that the population variance for this survey is known to be $\sigma^2 = 100$. Construct a 95-percent confidence interval for μ, assuming that a random sample of 256 adults yielded a mean score of 61.0. Is it necessary in this case to assume that the parent population is normal?

Problem 6.3

7.9 For a random sample of 50 measurements on the breaking strength of cotton threads, the mean breaking strength was found to be 210 grams and the standard deviation 18 grams. Obtain a confidence interval for the true mean breaking strength of cotton threads of this type, with confidence coefficient 0.90.

Problem 6.4

7.10 A random sample of 40 engineers was selected from among the large number employed by a corporation engaged in seeking new sources of petroleum. The hours worked in a particular week were determined for each engineer selected. These data had a mean of 46 hours and a standard deviation of 3 hours. For that particular week, estimate the mean hours worked for all engineers in the corporation, with a 95% confidence coefficient.

7.13 In the setting of Exercise 7.10, how many engineers should be sampled if it is desired to estimate the mean number of hours worked to within 0.5 hour with confidence coefficient 0.95?

Problem 6.5

7.11 An important property of plastic clays is the percent of shrinkage on drying. For a certain type of plastic clay, 45 test specimens showed an average shrinkage percentage of 18.4 and a standard deviation of 1.2. Estimate the true average percent of shrinkage for specimens of this type in a 98% confidence interval.

7.14 Refer to Exercise 7.11. How many specimens should be tested if it is desired to estimate the percent of shrinkage to within 0.2 with confidence coefficient 0.98?

Problem 6.6

7.12 The breaking strength of threads has a standard deviation of 18 grams. How many measures on breaking strength should be used in the next experiment if the estimate of the mean breaking strength is to be within 4 grams of the true mean breaking strength, with confidence coefficient 0.90?

Problem 6.7

7.15 Upon testing 100 resistors manufactured by Company A, it is found that 12 fail to meet the tolerance specifications. Find a 95% confidence interval for the true fraction of resistors manufactured by Company A that fail to meet the tolerance specification. What assumptions are necessary for your answer to be valid?

7.16 Refer to Exercise 7.15. If it is desired to estimate the true proportion failing to meet tolerance specifications to within 0.05, with confidence coefficient 0.95, how many resistors should be tested?

Problem 6.8

7.17 Careful inspection of 70 precast concrete supports to be used in a construction project revealed 28 with hairline cracks. Estimate the true proportion of supports of this type with cracks in a 98% confidence interval.

7.18 Refer to Exercise 7.17. Suppose it is desired to estimate the true proportion of cracked supports to within 0.1, with confidence coefficient 0.98. How many supports should be sampled to achieve the desired accuracy?

Problem 6.9

7.19 In conducting an inventory and audit of parts in a certain stockroom, it was found that, for 60 items sampled, the audit value exceeded the book value on 45 items. Estimate, with confidence coefficient 0.90, the true fraction of items in the stockroom for which the audit value exceeds the book value.

Problem 6.10

7.21 The warpwise breaking strength measured on five specimens of a certain cloth gave a sample mean of 180 psi and a standard deviation of 5 psi. Estimate the true mean warpwise breaking strength for cloth of this type in a 95% confidence interval. What assumption is necessary for your answer to be valid?

Problem 6.11

7.23 Fifteen resistors were randomly selected from the output of a process supposedly producing 10-ohm resistors. The 15 resistors actually showed a sample mean of 9.8 ohms and a sample standard deviation of 0.5 ohm. Find a 95% confidence interval for the true mean resistance of the resistors produced by this process. Assume resistance measurements are approximately normally distributed.

Problem 6.12

7.29 Fatigue behavior of reinforced concrete beams in seawater was studied by T. Hodgkiess, et al. (*Materials Performance*, July 1984, pp. 27–29). The number of cycles to failure in seawater for beams subjected to certain bending and loading stress was as follows (in thousands):

774, 633, 477, 268, 407, 576, 659, 963, 193

Construct a 90% confidence interval estimate of the average number of cycles to failure for beams of this type.

7.30 Using the data given in Exercise 7.29, construct a 90% confidence interval for the variance of the number of cycles to failure for beams of this type. What assumptions are necessary for your answer to be valid?

Problem 6.13

The article “A Non-Local Approach to Model the Combined Effects of Frequency Defects and Shot-Peening on the Fatigue Strength of a Pearlitic Steel” (B. Gerin, E. Pessard, et al., *Theoretical and Applied Fracture Mechanics*, 2018:19–32) reports that in a sample of 70 steel connecting rods subject to fatigue testing, the average fatigue strength, in MPa, was 408.2 with a standard deviation of 72.9. Find a 95% confidence interval for the mean fatigue strength under these conditions.

7

Hypothesis tests

Problem 7.1

8.1 The output voltage for a certain electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean of 128.6 and a standard deviation of 2.1. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a 5% significance level.

Problem 7.2

8.7 The pH of water coming out of a certain filtration plant is specified to be 7.0. Thirty water samples independently selected from this plant show a mean pH of 6.8 and a standard deviation of 0.9. Is there any reason to doubt that the plant's specification is being maintained? Use $\alpha = 0.05$. Find the P value for this test.

Problem 7.3

8.11 Certain rockets are manufactured with a range of 2,500 meters. It is theorized that the range will be reduced after the rockets are in storage for some time. Six of these rockets are stored for a certain period of time and then tested. The ranges found in the tests are as follows: 2,490, 2,510, 2,360, 2,410, 2,300, and 2,440. Does the range appear to be shorter after storage? Test at the 1% significance level.

Problem 7.4

8.13 The stress resistance of a certain plastic is specified to be 30 psi. The results from ten specimens of this plastic show a mean of 27.4 psi and a standard deviation of 1.1 psi. Is there sufficient evidence to doubt the specification at the 5% significance level? What assumption are you making?

Problem 7.5

8.15 The widths of contact windows in certain CMOS circuit chips have a design specification of 3.5 μm. (See M. S. Phadke, et al., *The Bell System Technical Journal*, 62, no. 5, 1983, pp. 1273–1309 for details.) Postetch window widths of test specimens were as follows:

$$3.21, 2.49, 2.94, 4.38, 4.02, 3.82, 3.30, 2.85, 3.34, 3.91$$

(These data were also used in Exercise 7.81.) Can we reject the hypothesis that the design specification is being met, at the 5% significance level? What assumptions are necessary for this test to be valid?

Problem 7.6

8.17 The Florida Poll of February–March 1984 interviewed 871 adults from around the state. On one question, 53% of the respondents favored strong support of Israel. Would you conclude that a majority of adults in Florida favor strong support of Israel? (Source: *Gainesville Sun*, 1 April 1984.)

Problem 7.7

8.19 Wire used for wrapping concrete pipe should have an ultimate tensile strength of 300,000 pounds, according to a design engineer. Forty tests of such wire used on a certain pipe showed a mean ultimate tensile strength of 295,000 pounds and a standard deviation of 10,000. Is there sufficient evidence to suggest that the wire used on the tested pipe does not meet the design specification at the 10% significance level?

Problem 7.8

8.21 Soil pH is an important variable in the design of structures that will contact the soil. The pH at a potential construction site was said to average 6.5. Nine test samples of soil from the site gave readings of

7.3, 6.5, 6.4, 6.1, 6.0, 6.5, 6.2, 5.8, 6.7

Do these readings cast doubt upon the claimed average? (Test at the 5% significance level.) What assumptions are necessary for this test to be valid?

Problem 7.9

8.5 In a study of family income levels, nine households in a certain rural county are selected at random. The incomes in these households are known to be normally distributed with $\sigma = 2400$. An average income of \$10,000 is reported from this sample survey. Use a statistical test to decide if the null hypothesis, $\mu =$ \$12,000, can be rejected in favor of H_a, which states that the true average income is less than \$12,000. Use an 0.05 level of significance.

Problem 7.10

8.17 A firm that packages deluxe ornamental matches for fireplace use designed a process to place 18 matches in each box. The process was started and allowed to produce 400 boxes. A sample of 16 boxes was then drawn. On the basis of this sample, the number of matches per box averaged 17, while the standard deviation calculated was 2. Would a one-sided test indicate acceptance of the null hypothesis with a mean of 18 if alpha were set at 0.05? Assume that the population is normal.

Problem 7.11

In courses on surveying, field work is an important part of the curriculum. The article "Enhancing Civil Engineering Surveying Learning through Workshops" (J. Awange, A. Faisal Anwar, et al., *Journal of Surveying Engineering*, 2017, online) reports that in a sample of 67 students studying surveying, 45 said that field work improved their ability to handle unforeseen problems. Can we conclude that more than 65% of students find that field work improves their ability to handle unforeseen problems?

Problem 7.12

The article "Geographically Weighted Regression-Based Methods for Merging Satellite and Gauge Precipitation" (L. Chao, K. Zhang, et al., *Journal of Hydrology*, 2018:275–289) describes a method of measuring precipitation by combining satellite measurements with measurements made on the ground. Errors were recorded monthly for 16 months, and averaged 0.393 mm with a standard deviation of 0.368 mm. Can we conclude that the mean error is less than 0.6 mm?