COMP 233, Winter 2016 **CONCORDIA UNIVERSITY**

PROBABILITY AND STATISTICS FOR COMPUTER SCIENCE.
Assignment 3.
Solutions.

1. (a) [6%] A random variable X has the standard normal distribution. Use the standard normal Table to determine :

$$P(X \le -0.5), \qquad P(X \le 0.5), \qquad P(\mid X \mid \ge 0.5),$$
$$P(\mid X \mid \le 0.5), \qquad P(-1 \le X \le 1), \qquad P(-1 \le X \le 0.5) \ .$$

(b) [6%] Suppose X is normally distributed with mean $\mu = 1.5$ and standard deviation $\sigma = 2.5$. Use the standard normal Table to determine :

$$P(X \le -0.5), \quad P(X \ge 0.5), \quad P(\mid X - \mu \mid \ge 0.5), \quad P(\mid X - \mu \mid \le 0.5) \ .$$

Solution.

(a) Using the standard normal distribution Table :

- $P(X \le -0.5) \ = \ \Phi(-0.5) \ \cong \ 31 \ \%$
- $P(\ X \le 0.5) \ = \ 1 - \Phi(-0.5) \ \cong \ 69 \ \%$
- $P(\mid X \mid \ge 0.5) \ = \ 2 \cdot \Phi(-0.5) \ \cong \ 62 \ \%$
- $P(\mid X \mid \le 0.5) \ = \ 1 - P(\mid X \mid \ge 0.5) \ \cong \ 38 \ \%$
- $P(-1 \le X \le 1) \ = \ 1 - 2 \cdot \Phi(-1) \ \cong \ 1 - 2 \ \cdot \ 0.16 \ = \ 68 \ \%$
- $P(-1 \le X \le 0.5) \ = \ \Phi(0.5) - \Phi(-1) \ = \ \big[1 - \Phi(-0.5)\big] - \Phi(-1) \ \cong \ (1 - 0.31) - 0.16 \ = \ 53 \ \%$

(b)

- $P(X \le -0.5) \ = \ \Phi(\frac{-0.5-1.5}{2.5}) \ = \ \Phi(-0.8) \ \approx \ 21 \ \%$
- $P(X \ge 0.5) \ = \ 1 - P(X < 0.5) \ = \ 1 - \Phi(\frac{0.5-1.5}{2.5})$
- $\qquad = \ 1 - \Phi(-0.4) \ \approx \ 1 - 0.34 \ = \ 0.66 \ \%$
- $P(\mid X - \mu \mid \ge 0.5) \ = \ 2 \cdot P(X \le 1.0) \ = \ 2 \cdot \Phi(\frac{1.0-1.5}{2.5})$
- $\qquad = \ 2 \cdot \Phi(-0.2) \ \approx \ 2 \cdot 0.42 \ = \ 84 \ \%$
- $P(\mid X - \mu \mid \le 0.5) \ = \ 1 - P(\mid X - \mu \mid \ge 0.5)$
- $\qquad \approx \ 1 - 0.84 \ = \ 16 \ \%.$

2. (a) [6%] Use the Central Limit Theorem (CLT) to estimate

$$P(\chi^2_{32} \le 24), \qquad P(\chi^2_{32} \ge 40), \qquad P(\mid \chi^2_{32} - 32 \mid \le 8) \ .$$

(b) [6%] Suppose $X_1, \ X_2, \ \cdots \ , \ X_{12}$, are identical, independent, uniform random variables on $[0,1]$. Let $\bar{X} = \frac{1}{12}(X_1 + X_2 + \cdots + X_{12})$. Use the CLT to compute approximate values of

$$P(\bar{X} \le \frac{1}{3}), \qquad P(\bar{X} \ge \frac{2}{3}), \qquad P(\mid \bar{X} - \frac{1}{2} \mid \le \frac{1}{3}) \ .$$

(c) [6%] Suppose $X_1, \ X_2, \ \cdots, \ X_9$, are identical, independent, exponential random variables, with $f(x) = e^{-x}$. Let $\bar{X} = \frac{1}{9}(X_1 + X_2 + \cdots + X_9)$. Use the CLT to compute approximate values of

$$P(\bar{X} \le 0.4), \qquad P(\bar{X} \ge 1.6), \qquad P(\mid \bar{X} - 1 \mid \le 0.6) \ .$$

Solution.

(a) Use the approximation

$$P(\ \chi^2_n \le x\ ) \quad \cong \quad \Phi(\ \frac{x-n}{\sqrt{2n}}\ )\ ,$$

to compute approximate values of

$$P(\ \chi^2_{32}\ \le\ 24\ )\ \cong\ \Phi(\ \tfrac{24-32}{\sqrt{64}}\ )\ =\ \Phi(\tfrac{-8}{8})\ =\ \Phi(-1)\ \cong\ 0.16\ .$$

$$P(\ \chi^2_{32}\ \ge\ 40\ )\ \cong\ P(\ \chi^2_{32}\ \le\ 24\ )\ \cong\ 0.16\ .$$

$$P(\ \mid \chi^2_{32} - 32 \mid\ \ \le\ 8\ )\ \cong\ 1 - 2 \cdot 0.16\ =\ 0.68\ .$$

(b) $X_1, X_2, \cdots, X_{12}$, are identical, independent, uniform random variables on $[0,1]$. Each X_i has mean $\mu = \frac{1}{2}$, and standard deviation $\sigma = \frac{1}{2\sqrt{3}}$. Thus $\bar{X}$ has mean $\mu_{\bar{X}} = \frac{1}{2}$, and standard deviation $\sigma_{\bar{X}} = \frac{1}{2\sqrt{3n}}$.

$$P(\bar{X}\ \le\ \tfrac{1}{3})\quad \cong\quad \Phi(\ \tfrac{1/3-1/2}{1/(2\sqrt{3n})}\ )\quad =\quad \Phi(-2)\quad \cong\quad 2.28\ \%$$

$$P(\bar{X}\ \ge\ \tfrac{2}{3})\quad =\quad P(\bar{X}\ \le\ \tfrac{1}{3})\quad \cong\quad 2.28\ \%$$

$$P(\mid \bar{X} - \tfrac{1}{2} \mid\ \ \le\ \tfrac{1}{3})\ \cong\ 1 - 2\ \Phi(\ \tfrac{1/6-1/2}{1/(2\sqrt{3n})}\ )\ =\ 1 - 2\ \Phi(-4)\ \cong\ 100\ \%$$

(c) $X_1, X_2, \cdots, X_9$, are identical, independent, exponential random variables on $[0, \infty)$. Each X_i has mean $\mu = 1$, and standard deviation $\sigma = 1$. Thus $\bar{X}$ has mean $\mu_{\bar{X}} = 1$, and standard deviation $\sigma_{\bar{X}} = \frac{1}{\sqrt{n}}$.

$$P(\bar{X}\ \le\ 0.4)\quad \cong\quad \Phi(\ \tfrac{0.4-1}{1/\sqrt{n}}\ )\quad =\quad \Phi(-1.8)\quad \cong\quad 3.6\ \%$$

$$P(\bar{X}\ \ge\ 1.6)\quad \cong\quad P(\bar{X}\ \le\ 0.4)\quad \cong\quad 3.6\ \%$$

$$P(\mid \bar{X} - 1 \mid\ \le\ 0.6)\quad \cong\quad 1 - 2\ \Phi(-1.8)\quad \cong\quad 92.8\ \%$$

3. Mathematics test scores across a population of high school are normally distributed with mean 500 and standard deviation 100. Determine the following probabilities:

(a) [4%] A randomly chosen student scored below 600.

(b) [4%] Of five randomly chosen students, all scored below 600.

(c) [4%] Of five randomly chosen students, exactly three scored above 640.

Solution.

(a) A randomly chosen student has the probability of scoring below 600 of

$$P\{X < 600\} = P\{Z < \frac{600-500}{100}\} = P\{Z < 1\} = \Phi(1) = 0.8413.$$

(b) By the multiplication rule (assuming each student took the test independently of each other) the probability that of five randomly chosen students, all scored below 600 is

$$(\Phi(1))^5 = 0.4215.$$

(c) Treating each student's score as a Bernoulli experiment with the probability of scoring above 640 (i.e., a successful experiment) as

$$p = P\{X > 640\} = P\{Z > \frac{640 - 500}{100}\} = P\{Z > 1.4\} = 1 - \Phi(1.4) = 1 - 0.9292 = 0.0808.$$

we can use the Binomial distribution (with n = 5) to find

$$P\{X = 3\} = C(5, 3)p^3(1 - p)^2 = 10 \cdot 0.0808^3 \cdot 0.9292^2 = 0.0045.$$

4. [12%] According to a transportation safety board, the number of persons per car passing a certain intersection between 8:00 and 9:00am, is a random variable H with mean 4 and variance 2. For a random sample of 30 cars at this intersection during this time period, what is the probability that the average number of persons per car is at least 5?

   Solution.

   Let random variable X_i denote the number of people in the i-th car, i = 1, 2, 3,.., 30. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. We seek the probability that the average number of people/car be ≥ 5. The cars are selected randomly.

$$\begin{aligned} &P(\tfrac{1}{30}\textstyle\sum_{i=1}^{30} X_i \geq 5) \\ &= P(\overline{X}_{30} \geq 5). \end{aligned}$$

   By the Central Limit Theorem

$$\overline{X}_{30} \sim N\left(4, \frac{\sqrt{2}}{\sqrt{30}}\right),$$

   where $\sigma = \frac{\sqrt{2}}{\sqrt{30}}$. Standardizing,

$$\begin{aligned} P\left(\frac{\overline{X}_{30} - 4}{\sqrt{\frac{1}{15}}} \geq \frac{5 - 4}{\sqrt{\frac{1}{15}}}\right) &= P\left(Z \geq \frac{1}{\sqrt{\frac{1}{15}}}\right) \\ &= P(Z \geq \sqrt{15}) \\ &= P(Z \geq 3.87) \\ &= 5.44177 \cdot 10^{-5} \\ &\approx 0. \end{aligned}$$

5. [12%] A small elevator has a maximum capacity C, which is normally distributed, with mean 400 kg., and standard deviation 4 kg. The weight of the boxes being loaded into the elevator is a random variable with mean 30 kg., and standard deviation 0.3 kg. Assume that the weights of the boxes and maximum capacity are independent random variables. How many boxes may be loaded into the elevator before the probability of disaster exceeds 20%?

   Hint: Let X_i denote the weight of the i-th box being loaded into the elevator, and let P_n be probability that the weight of n loaded boxes exceeds maximum capacity C.

   Then $P_n = P(\sum_{i=1}^{n} X_i > C) = P(\sum_{i=1}^{n} X_i - C > 0)$. Now use the CLT.

   Solution.

   Let R.V. $X_i, i = 1, \ldots, n$ denote the weight of the i-th box being loaded into the elevator, P_n be probability that weight of n loaded boxes exceeds maximum capacity C. Now then

$$\begin{aligned} P_n &= P(\sum_{i=1}^{n} X_i > C) \\ &= P(\sum_{i=1}^{n} X_i - C > 0). \end{aligned}$$

By the Central Limit Theorem $X = \sum_{i=1}^{n} X_i$ is Normally Distributed with mean $30n$, and variance $0.09n$. Given that the mean of C is 400 and the variance is 16 we have

$$\begin{aligned}
E\left[\sum_{i=1}^{n} X_i - C\right] &= E[X - C] \\
&= E[X] - E[C] \\
&= 30n - 400 \\
Var(X - C) &= Var(X) + Var(C) \\
&= 0.09n + 16.
\end{aligned}$$

Standardizing

$$\begin{aligned}
P_n &= P\left(\sum_{i=1}^{n} X_i - C > 0\right) \\
&= P(X - C > 0) \\
&= P\left(\frac{((X - C) - (30n - 400))}{\sqrt{0.09n + 16}} > \frac{-(30n - 400)}{\sqrt{0.09n + 16}}\right) \\
&= P\left(Z > \frac{400 - 30n}{0.09n + 16}\right).
\end{aligned}$$

What is sought is the maximum n such that $P_n \leq 0.2$ or

$$\begin{aligned}
P\left(Z > \frac{400 - 30n}{\sqrt{0.09n + 16}}\right) &\leq 0.2. \\
P(Z \geq 0.84) &\approx 0.2 \quad \text{(refer to the Standard Normal table).}
\end{aligned}$$

Hence

$$n_{max} = \max\left\{n : \tfrac{400 - 30n}{\sqrt{0.09n+16}} \geq 0.84\right\}$$
$$n_{max} = \max\left\{n : 400 - 30n \geq 0.84\sqrt{0.09n + 16}\right\}$$
$$n_{max} = 13 \quad \text{boxes.}$$

6. [12%] The following are the percentages of ash content in 12 samples of coal:

9.2 , 14.1 , 9.8 , 12.4 , 16.0 , 12.6 , 22.7 , 18.9 , 21.0 , 14.5 , 20.4 , 16.9

Compute the sample mean. the sample median. and the sample standard deviation.

Solution.

We have

$$n = 12, \text{ and } \sum_{i=1}^{n} X_i = 188.5.$$

Hence

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{188.5}{12} \approx 15.71.$$

For the Sample Median we use the value at position $0.5(n + 1) = 0.5(13) = 6.5$. Therefore, we use the simple average of the sorted values at positions 6 and 7, which gives

$$\frac{14.5 + 16}{2} = 15.25.$$

We also have $\sum_{i=1}^{n} X_i^2 = 3173.53$, so

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n - 1} = \frac{3173.53 - 2961.65}{11} \approx 19.3.$$

Therefore, $S =\approx \sqrt{19.3} \approx 4.4$.

7. The average particulate concentration, in micrograms per cubic meter, was measured in a petrochemical complex at 36 randomly chosen times, with the following results:

`5, 18, 15, 7, 23, 220, 130, 85, 103, 25, 80, 7, 24, 6, 13, 65, 37, 25,`
`24, 65, 82, 95, 77, 15, 70, 110, 44, 28, 33, 81, 29, 14, 45, 92, 17, 53`

(a) [2%] Represent the data in a histogram.

(b) [2%] Is the histogram approximately normal?

(c) [2%] Calculate the sample mean $\overline{X}$.

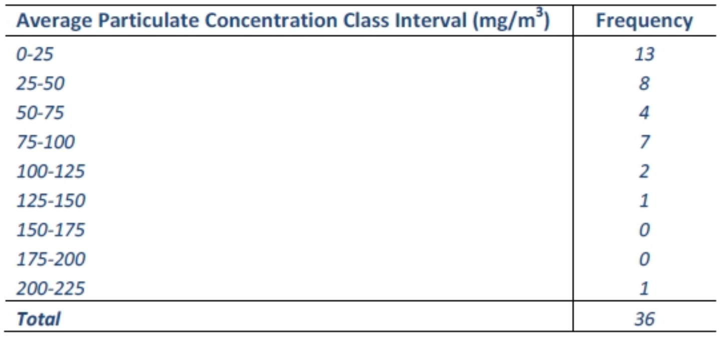
(d) [2%] Calculate the sample standard deviation S.

(e) [2%] Determine the proportion of the data values that lies within $\overline{X} \pm 1.5S$ and compare with the lower bound given by Chebyshev's inequality.

(f) [2%] Determine the proportion of the data values that lies within $\overline{X} \pm 2S$ and compare with the lower bound given by Chebyshev's inequality.

Solution.

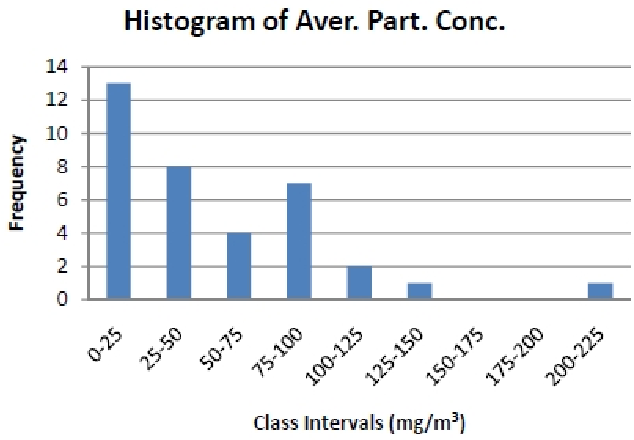
(a) Class intervals for data:



Average Particulate Concentration Class Interval (mg/m^3)	Frequency
0-25	*13*
25-50	*8*
50-75	*4*
75-100	*7*
100-125	*2*
125-150	*1*
150-175	*0*
175-200	*0*
200-225	*1*
Total	*36*

Different choices for the class intervals will lead to slightly different results.

Different choices for the class intervals will lead to slightly different results.
Resulting Histogram:



(b) The histogram is obviously **not** normally distributed (not even skewed normal).

(c) We have the following information

$$n = 36 \text{ and } \sum_{i=1}^{n} X_i = 1862$$

so

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1862}{36} \approx 51.72.$$

(d) We also have $\sum_{i=1}^{n} X_i^2 = 167058$, so

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{167058 - 96306}{35} \approx 2021.4.$$

Therefore, $S = \sqrt{2021.4} \approx 44.96$.

(e) $\bar{X} \pm 1.5S = 51.72 \pm (1.5)44.96 = (-15.72, 119.16)$. There are 34 data values in this range (that is, 34/36 or 94.4%) as compared to Chebyshev's lower bound of

$$(1 - 1/(1.5)^2) = (1 - 4/9) = 5/9 \text{ or } 55.5\%.$$

The Chebyshev's lower bound can be derived as follows:

$$P\left\{|X - \bar{X}| \geq k\sigma\right\} \leq \frac{1}{k^2}$$

thus

$$1 - P\left\{|X - \bar{X}| \geq k\sigma\right\} = P\left\{|X - \bar{X}| < k\sigma\right\} \geq 1 - \frac{1}{k^2}.$$

We obtain the result above by taking $k = 1.5$.

(f) $\bar{X} \pm 2S = 51.72 \pm (2)44.96 = (-38.20, 141.64)$. There are 35 data values in this range (that is, 35/36 or 97.2%) as compared to Chebyshev's lower bound of

$$(1 - 1/(2)^2) = (1 - 1/4) = 3/4 \text{ or } 75\%.$$

8. [12%] If 10 pairs of fair dice are rolled, approximate the probability that the sum of the values obtained (which ranges from 20 to 120) is between 30 and 40 inclusive.

Solution.

For the 10 pairs of fair dice, then $n = 20$, we want, letting X be the sum of the face values of the n case dice,

$$P\{30 < X < 40\}.$$

By the Central Limit Theorem, X is normally distributed with mean $n\mu$ and standard deviation $\sqrt{n}\sigma$. For the single roll of a die, $\mu = 3.5$, and the variance is $\sigma^2 = 35/12$, so that the standard deviation is $\sigma = \sqrt{35/12} = 1.708$.

So to convert the above probability calculation to standard normal form, we use

$$\begin{aligned}
&P\left\{\frac{30 - n\mu}{\sqrt{n}\sigma} < \frac{X - n\mu}{\sqrt{n}\sigma} < \frac{40 - n\mu}{\sqrt{n}\sigma}\right\} = P\left\{\frac{30 - 20(3.5)}{\sqrt{20}(1.708)} < Z < \frac{40 - 20(3.5)}{\sqrt{20}(1.708)}\right\} \\
&= P\left\{\frac{30 - 70}{7.64} < Z < \frac{40 - 70}{7.64}\right\} = P\left\{\frac{-40}{7.64} < Z < \frac{-30}{7.64}\right\} \\
&= P\{-5.24 < Z < -3.93\} = \Phi(-3.93) - \Phi(-5.24) \\
&= 1 - \Phi(3.93) - (1 - \Phi(5.24)) = \Phi(5.24) - \Phi(3.93) \\
&\approx 0.9999999 - 0.9999575 \approx 4.24 \cdot 10^{-5}.
\end{aligned}$$

The probability is essentially zero.

9. The age at which the fan assembly in a laptop fails is normally distributed with variance σ^2. Let S^2 be the corresponding sample variance for seven laptops. Determine

(a) [6%] $P(S^2/\sigma^2 \leq 1.5)$ (b) [6%] $P(0.8 \leq S^2/\sigma^2 \leq 1.1)$

Solution.

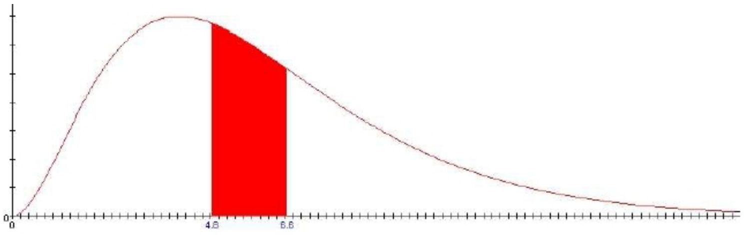
(a)

$$P\{S^2/\sigma^2 \leq 1.5\} = P\{(n-1)S^2/\sigma^2 \leq (n-1)1.5\} = P\{(6-1)S^2/\sigma^2 \leq (6-1)1.5\} = P\{\chi_6^2 \leq 9\} \approx 0.8266.$$

(b)

$$P\{4.8 \leq \chi_6^2 \leq 6.6\} = P\{\chi_6^2 \leq 6.6\} - P\{\chi_6^2 \leq 4.8\} \approx 0.6408 - 0.4304 = 0.2104.$$

The χ^2 distribution of 6 degrees of freedom has the following shape



10. Consider the following data on the diameter of a rivet. Assume a normal population.

```
6.68 6.66 6.62 6.72 6.76 6.67 6.70 6.72
6.78 6.66 6.76 6.72 6.76 6.70 6.76 6.76
6.74 6.74 6.81 6.66 6.64 6.79 6.72 6.82
```

(a) [6%] Estimate the population variance σ^2.

(b) [6%] Compute a 95 percent two-sided confidence interval for σ^2.

Solution. From the question, we gather the following information

$$n = 24\ ,\ \sum_{i=1}^{n} X_i = 161.35\ ,\ \bar{X} = 6.7229\ ,\ \sum_{i=1}^{n} X_i^2 = 1084.81$$

so

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{1084.81 - 24\ 6.722916^2}{23} \approx 0.0029.$$

Therefore, $S = \sqrt{0.0029} \approx 0.054$.

(a) The point estimate for the population variance σ^2 is

$$\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n} = \frac{1084.81 - 24\ 6.722916^2}{24} \approx 0.0028.$$

(b) We have $1-\alpha = 0.95$, or $\alpha = 0.05$. Therefore, from χ^2 distribution table, we find

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,23} = 38.076,$$

and

$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,23} = 11.689,$$

Then a 95% two-sided confidence interval on σ^2 is

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = \left(\frac{(23)0.0029}{38.076}, \frac{(23)0.0029}{11.689}\right) \approx (0.0017, 0.0057).$$