Assignment 1

Sainivedhitha Arunajatesan

Exercise 1

When the authors were in college, there were only three requirements for graduation that were the same for all students: You had to be able to tread water for 2 minutes, you had to learn a foreign language, and you had to be free of tuberculosis. For the last requirement, all freshmen had to take a TB screening test that consisted of a nurse jabbing what looked like a corn cub holder into you forearm. You were then expected to report back in 48 hours to have it checked. If you were healthy and TB-free your arm was supposed to look as though you'd never had the test.

Sometimes during the 48 hours, one of us had a reaction. When he finally saw the nurse, his arm was about 50% larger than normal and a very unhealthy red. Did he have TB? The nurse had said that the test was about 97% effective, so it seemed that the chances must be pretty high that he had TB.

Compute the probability that a freshman with a reaction (test is positive) has indeed TB. Assume that the (overall) probability of TB is 1 / 7000 (prevalence) and that the nurse is right: the test is positive in 97% of all cases of TB. Assume furthermore that the test is positive with probability 0.012 if you are free from TB.

Compute also the probability that one is free from TB given a negative test.

The confusion matrix for this situation can be filled as:

$$P(TB^+) = \frac{1}{7000} = 1.42 * 10^{-4}$$
$$P(TB^-) = 1 - \frac{1}{7000} = \frac{6999}{7000} = 0.99985$$

	positive	negative
TB^+	$0.97 * P(TB^+) = 1.3774 * 10^{-4}$	?
TB^-	0.012	?

The probability that a freshman has TB with the positive test is computed using the following conditional probability.

$$\begin{aligned}
P(TB^+|positive) &= \frac{P(TB^+ \cap positive)}{P(positive)} \\
&= \frac{P(TB^+ \cap positive)}{P(TB^+ \cap positive) + P(TB^- \cap positive)} \\
&= \frac{0.97 * P(TB^+)}{0.97 * P(TB^+) + 0.012} \\
&= \frac{0.97 * 1.42 * 10^{-4}}{0.97 * 1.42 * 10^{-4} + 0.012} = 0.0113 = 1.13\%
\end{aligned}$$

The probability that a freshman has no TB with the negative test is computed using the following conditional probability.

$$
\begin{aligned}
P(TB^-|negative) &= \frac{P(TB^- \cap negative)}{P(negative)} \\
&= \frac{P(TB^- \cap negative)}{1 - P(positive)}
\end{aligned}
$$

From the above calculation,

$$
\begin{aligned}
P(Positive) &= 0.97 * 1.42 * 10^{-4} + 0.012 = 0.0121 \\
P(TB^-|negative) &= \frac{P(TB^-) - P(TB^- \cap positive)}{0.0121} \\
&= \frac{0.99985 - 0.012 * 0.99985}{1 - 0.0121} \\
&= 0.99995 = 99.995\%
\end{aligned}
$$

Exercise 2

The survival time X of some medical equipment has a continuous distribution described by the following density function:

$f(x) = 0.2 - 0.02x; for x \epsilon (0, 10), f(x) = 0; elsewhere$

1. Compute the probabilities $P(4.4 < X < 6.6)$, $P(X > 6.0)$ and $P(X \geq 6.0)$.

$$
\begin{aligned}
P(4.4 < X < 6.6) &= \int_{4.4}^{6.6} (0.2 - 0.02x)\, dx \\
&= [(0.2(6.6) - \frac{0.02(6.6)^2}{2}) - (0.2(4.4) - \frac{0.02(4.4)^2}{2})] \\
&= 0.198
\end{aligned}
$$

$$
\begin{aligned}
P(X > 6.0) &= \int_{6.0}^{10.0} (0.2 - 0.02x)\, dx \\
&= [(0.2(10.0) - \frac{0.02(10.0)^2}{2}) - (0.2(6.0) - \frac{0.02(6.0)^2}{2})] \\
&= 0.16
\end{aligned}
$$

$$
\begin{aligned}
P(X \geq 6.0) &= \int_{6.0}^{10.0} (0.2 - 0.02x)\, dx \\
&= [(0.2(10.0) - \frac{0.02(10.0)^2}{2}) - (0.2(6.0) - \frac{0.02(6.0)^2}{2})] \\
&= 0.16
\end{aligned}
$$

2. Compute the expectation μ=E(X).

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x(0.2 - 0.02x)\, dx \\
&= [\frac{(0.2(10.0)^2}{2} - \frac{0.02(10.0)^3}{3}) - \frac{(0.2(0)^2}{2} - \frac{0.02(0)^3}{3})] \\
&= 0.16
\end{aligned}
$$

3. Compute the conditional probability $P(X > 7.0|X \geq 4.0)$.

$$\begin{aligned}
P(D|A) &= \frac{P(A \cap D)}{P(A)} \\
P(X > 7.0|X \geq 4.0) &= \frac{\int_7^{10} (0.2 - 0.02x)\,dx}{\int_4^{10} (0.2 - 0.02x)\,dx} \\
&= \frac{[(0.2(10.0) - \frac{0.02(10.0)^2}{2}) - (0.2(7) - \frac{0.02(7)^2}{2})]}{[(0.2(10.0) - \frac{0.02(10.0)^2}{2}) - (0.2(4) - \frac{0.02(4)^2}{2})]} \\
&= 0.25(or)25\%
\end{aligned}$$

Exercise 3

Consider a stochastic variable Z that has a standard normal distribution. Determine/compute the following probabilities using the standard normal table: $P(Z \leq 0.78), P(0.08 < Z \leq 1.72)$, $P(Z < -0.75)$ and $P(-0.53 < Z < 1.74)$.

According to Standard normal distribution table,

$$\begin{aligned}
P(Z \leq 0.78) &= 0.7734 \\
P(0.08 < Z \leq 1.72) &= P(Z = 1.72) - P(Z = 0.08) \\
&= 0.9573 - 0.5319 \\
&= 0.4254 \\
P(Z < -0.75) &= 1 - P(Z < 0.75) \\
&= 1 - 0.7734 \\
&= 0.2266 \\
P(-0.53 < Z < 1.74) &= P(Z = 1.74) - [1 - P(Z = 0.53)] \\
&= 0.9591 - [1 - 0.7019] \\
&= 0.661
\end{aligned}$$

Exercise 4

Consider a stochastic variable X that has a normal distribution with μ=30 and σ=7.

1. Determine/compute the following probabilities: $P(X \leq 34.50)$ and $P(21.63 < X \leq 36.71)$.

$$\begin{aligned}
P(X \leq 34.50) =&> Z = \frac{X - \mu}{\sigma} = \frac{34.50 - 30}{7} \\
Z &= 0.64 \\
P(Z = 0.64) &= 0.7389
\end{aligned}$$

By using the same formula of Z,

$$\begin{aligned}
P(21.63 < X \leq 36.71) &=> Z1 = -1.195, Z2 = 0.958 \\
P(-1.195 < Z \leq 0.958) &= 0.8315 - [1 - 0.883] \\
&= 0.7145
\end{aligned}$$

2. Determine numbers C1,C2,C3 and C4 such that: P(X<C1)=20%, $P(C1 \leq X < C2)$ =20%, $P(C2 \leq X < C3) = 20\%$, $P(C3 \leq X < C4) = 20\%$ and $P(X \geq C4) = 20\%$.

To find C1,
$$P(X < C1) = 20\% \tag{1}$$
The value at C1 is 1-0.2=0.8
The closest probability to 0.8 in the normal distribution table is 0.7995 for Z=-0.84.
Using Z=-0.84 in

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ -0.84 &= \frac{C1-30}{7} \\ C1 &= 24.12 \end{aligned}$$

To find C2,
$$P(C1 < X < C2) = 20\% \tag{2}$$
By combining equation 1 and 2,
$$P(X < C2) = 40\% \tag{3}$$
The value of C2 is 1-0.4=0.6
The closest probability to 0.6 in the normal distribution table is 0.5987 for Z=-0.25.
Using Z=-0.25 in

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ C2 &= 28.25 \end{aligned}$$

To find C3,
$$P(C2 < X < C3) = 20\% \tag{4}$$
By combining equation 3 and 4,
$$P(X < C3) = 60\%$$
The value of C3 is 0.6
The closest probability to 0.6 in the normal distribution table is 0.5987 for Z=0.25.
Using Z=0.25 in

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ C3 &= 31.75 \end{aligned}$$

To find C4, $P(X \geq C4) = 20\%$
The value at C4 is 1-0.2=0.8
The closest probability to 0.8 in the normal distribution table is 0.7995 for Z=0.84.
Using Z=0.84 in

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ 0.84 &= \frac{C4-30}{7} \\ C4 &= 35.88 \end{aligned}$$

Exercise 5

Consider a stochastic variable X that has a normal distribution with some expectation μ and some standard deviation σ

1. Verify that $P(\mu - \sigma < X < \mu + \sigma) = 68\%$ holds. Note that this reflects the meaning of the standard deviation as measure of spread.

$$
\begin{aligned}
Z &= \frac{X - \mu}{\sigma} \\
Z1 &= \frac{\mu - \sigma - \mu}{\sigma} = -1 \\
Z2 &= \frac{\mu + \sigma - \mu}{\sigma} = +1 \\
P(-1 < Z < 1) &= P(Z = 1) - [1 - P(Z = 1)] \\
&= 0.8413 - 0.1587 = 0.6826
\end{aligned}
$$

   Hence, $P(\mu - \sigma < X < \mu + \sigma) = 68\%$ is verified.

2. Compute the probabilities $P(\mu - 2\sigma < X < \mu + 2\sigma)$ and $P(\mu - 3\sigma < X < \mu + 3\sigma)$ as well.

$$
\begin{aligned}
Z &= \frac{X - \mu}{\sigma} \\
Z1 &= \frac{\mu - 2\sigma - \mu}{\sigma} = -2 \\
Z2 &= \frac{\mu + 2\sigma - \mu}{\sigma} = +2 \\
P(-2 < Z < 2) &= P(Z = 2) - [1 - P(Z = 2)] \\
&= 0.9772 - 0.0228 = 0.9544
\end{aligned}
$$

   $P(\mu - 2\sigma < X < \mu + 2\sigma) = 95.44\%$

$$
\begin{aligned}
Z &= \frac{X - \mu}{\sigma} \\
Z1 &= \frac{\mu - 3\sigma - \mu}{\sigma} = -3 \\
Z2 &= \frac{\mu + 3\sigma - \mu}{\sigma} = +3 \\
P(-3 < Z < 3) &= P(Z = 3) - [1 - P(Z = 3)] \\
&= 0.9987 - 0.0013 = 0.9974
\end{aligned}
$$

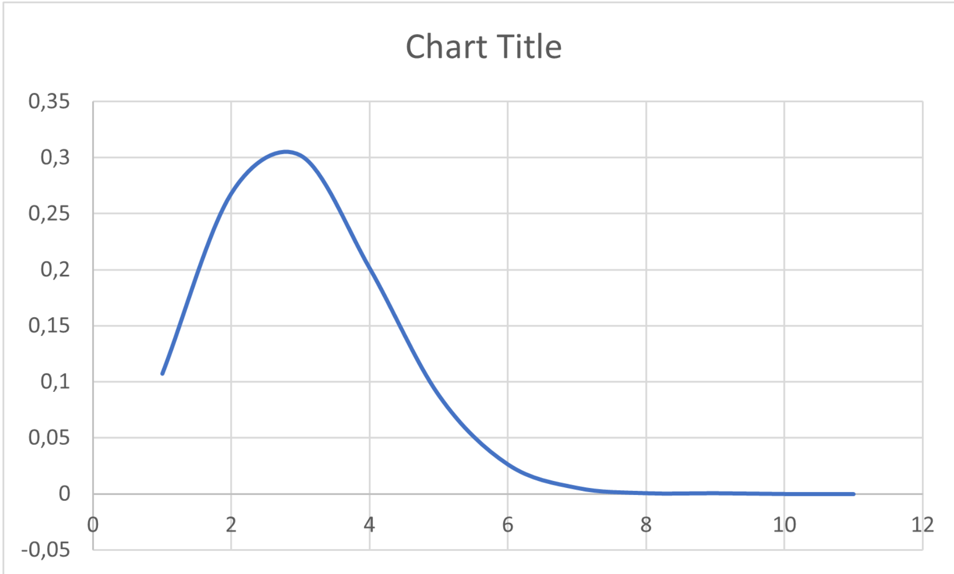
   $P(\mu - 3\sigma < X < \mu + 3\sigma) = 99.74\%$

Exercise 6

Using the program Excel one can calculate the probabilities of a binomial distribution. Within Excel go to FORMULAS and INSERT FUNCTIONS, choose BINOM.DIST from the statistical functions and select FALSE for CUMULATIVE. Show with plots using Excel (other tool is OK as well) that

1. The binomial distribution is a skew distribution for small p and small n

   Assuming the value of p=0.2 and changing the values of n between 1 and 10 in the BINOM.DIST function of EXCEL, it is proved that binomial distribution is a skew distribution for small p and n values.



2. This skewness disappears (the shape of the distribution becomes symmetrical) if we increase n. This phenomenon is the basis for the normal approximation of the binomial distribution of lecture 2a.

   Assuming the value of p=0.7 and changing the values of n between 50 and 95 at the interval of 5 in the BINOM.DIST function of EXCEL, it is shown that distribution is symmetrical. It was also found that this holds true only if the value of p is closest to 0.5.