IES 302: Engineering Statistics **2011/2**

HW Solution 6 — March 28

Lecturer: Prapun Suksompong, Ph.D.

Instructions

(a) ONE part of a question will be graded (5 pt). Of course, you do not know which part will be selected; so you should work on all of them.

(b) It is important that you try to solve all problems. (5 pt)

(c) Late submission will be heavily penalized.

(d) Use second part of Table III Cumulative Standard Normal Distribution from [Montgomery and Runger, 2010] to evaluate the Φ function and all probabilities associated with Gaussian distributions. Hereafter, this table will be referred to as the Φ table.

Problem 1. Let X be a uniform random variable on the interval $[0,1]$. Set

$$A = \left[0, \frac{1}{2}\right), \quad B = \left[0, \frac{1}{4}\right) \cup \left[\frac{1}{2}, \frac{3}{4}\right), \quad \text{and } C = \left[0, \frac{1}{8}\right) \cup \left[\frac{1}{4}, \frac{3}{8}\right) \cup \left[\frac{1}{2}, \frac{5}{8}\right) \cup \left[\frac{3}{4}, \frac{7}{8}\right).$$

Are the events $[X \in A], [X \in B]$, and $[X \in C]$ independent?

Solution: Note that

$$P[X \in A] = \int_0^{\frac{1}{2}} dx = \frac{1}{2},$$

$$P[X \in B] = \int_0^{\frac{1}{4}} dx + \int_{\frac{1}{2}}^{\frac{3}{4}} dx = \frac{1}{2}, \text{ and}$$

$$P[X \in C] = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{4}}^{\frac{3}{8}} dx + \int_{\frac{1}{2}}^{\frac{5}{8}} dx + \int_{\frac{3}{4}}^{\frac{7}{8}} dx = \frac{1}{2}.$$

Now, for pairs of events, we have

$$P([X \in A] \cap [X \in B]) = \int_0^{\frac{1}{4}} dx = \frac{1}{4} = P[X \in A] \times P[X \in B], \tag{6.1}$$

$$P([X \in A] \cap [X \in C]) = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{4}}^{\frac{3}{8}} dx = \frac{1}{4} = P[X \in A] \times P[X \in C], \text{ and} \tag{6.2}$$

$$P([X \in B] \cap [X \in C]) = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{2}}^{\frac{5}{8}} dx = \frac{1}{4} = P[X \in B] \times P[X \in C]. \tag{6.3}$$

Finally,

$$P([X \in A] \cap [X \in B] \cap [X \in C]) = \int_0^{\frac{1}{8}} dx = \frac{1}{8} = P[X \in A]\, P[X \in B]\, P[X \in C]. \tag{6.4}$$

From (6.1), (6.2), (6.3) and (6.4), we can conclude that the events $[X \in A], [X \in B]$, and $[X \in C]$ are $\boxed{\text{independent}}$.

Problem 2. Suppose Z is a standard Gaussian random variable.

(a) Use the Φ table to find the following probabilities:

(i) $P[Z < 1.52]$

(ii) $P[Z < -1.52]$

(iii) $P[Z > 1.52]$

(iv) $P[Z > -1.52]$

(v) $P[-1.36 < Z < 1.52]$

(b) Use the Φ table to find the value of c that satisfies each of the following relation.

(i) $P[Z > c] = 0.14$

(ii) $P[-c < Z < c] = 0.95$

Solution:

(a)

(i) $P[Z < 1.52] = \Phi(1.52) = \boxed{0.935744.}$

(ii) $P[Z < -1.52] = \Phi(-1.52) = 1 - \Phi(1.52) = 1 - 0.935744 = \boxed{0.064256.}$

(iii) $P[Z > 1.52] = 1 - P[Z < 1.52] = 1 - 0.935744 = \boxed{0.064256.}$

(iv) It is straightforward to see that the area of $P[Z > -1.52]$ is the same as $P[Z < 1.52] = \Phi(1.52)$. So, $P[Z > -1.52] = \boxed{0.935744.}$

Alternatively, $P[Z > -1.52] = 1 - P[Z \le -1.52] = 1 - \Phi(-1.52) = 1 - (1 - \Phi(1.52)) = \Phi(1.52)$.

(v) $P[-1.36 < Z < 1.52] = \Phi(1.52) - \Phi(-1.36) = \Phi(1.52) - (1 - \Phi(1.36)) = \Phi(1.52) + \Phi(1.36) - 1 = 0.935744 + 0.913085 - 1 = \boxed{0.848829.}$

(b)

(i) $P[Z > c] = 1 - P[Z \le c] = 1 - \Phi(c)$. So, we need $1 - \Phi(c) = 0.14$ or $c = 1 - 0.14 = 0.86$. From the Φ table, we have $c \approx \boxed{1.08.}$

(ii) $P[-c < Z < c] = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1$. So, we need $2\Phi(c) - 1 = 0.95$ or $\Phi(c) = 0.975$. From the Φ table, we have $c \approx \boxed{1.96.}$

Problem 3. The peak temperature T, as measured in degrees Fahrenheit, on a July day in New Jersey is a $\mathcal{N}(85, 100)$ random variable.

Remark: Do not forget that, for our class, the second parameter in $\mathcal{N}(\cdot, \cdot)$ is the variance (not the standard deviation).

(a) Express the cdf of T in terms of the Φ function

Hint: Recall that the cdf of a random variable T is given by $F_T(t) = P[T \le t]$.

(b) Express each of the following probabilities in terms of the Φ function(s). Make sure that the arguments of the Φ functions are positive. (Positivity is required so that we can use only the second part of Table III Cumulative Standard Normal Distribution from [Montgomery and Runger, 2010] to evaluate the probabilities.)

(i) $P[T > 100]$

(ii) $P[T < 60]$

(iii) $P[70 \le T \le 100]$

(c) Evaluate each of the probabilities in part (b) using the Φ table.

Solution:

(a) When $X \sim \mathcal{N}(m, \sigma^2)$, we know that its standardized version $\frac{X-m}{\sigma}$ is $\mathcal{N}(0,1)$. So,

$$P\left[\frac{X-m}{\sigma} \le c\right] = \Phi(c).$$

The cdf of X is $F_X(x) = P[X \le x]$. Note that $X \le x$ if and only if $\frac{X-m}{\sigma} \le \frac{x-m}{\sigma}$. Therefore,

$$F_X(x) = P\left[\frac{X-m}{\sigma} \le \frac{x-m}{\sigma}\right] = \Phi\left(\frac{x-m}{\sigma}\right).$$

Here, $T \sim \mathcal{N}(85, 10^2)$. Therefore, $F_T(t) = \boxed{\Phi\left(\frac{t-85}{10}\right).}$

(b)

(i) $P[T > 100] = 1 - P[T \le 100] = 1 - F_T(100) = 1 - \Phi\left(\frac{100-85}{10}\right) = 1 - \Phi(1.5)$

(ii) $P[T < 60] = P[T \le 60]$ because T is a continuous random variable and hence $P[T = 60] = 0$. Now, $P[T \le 60] = F_T(60) = \Phi\left(\frac{60-85}{10}\right) = \Phi(-2.5) = \boxed{1 - \Phi(2.5).}$ Note that, for the last equality, we use the fact that $\Phi(-x) = 1 - \Phi(x)$.

(iii)

$$P[70 \le T \le 100] = F_T(100) - F_T(70) = \Phi\left(\frac{100-85}{10}\right) - \Phi\left(\frac{70-85}{10}\right)$$
$$= \Phi(1.5) - \Phi(-1.5) = \Phi(1.5) - (1 - \Phi(1.5)) = \boxed{2\Phi(1.5) - 1.}$$

(c)

(i) $1 - \Phi(1.5) = 1 - 0.9332 = \boxed{0.0668.}$

(ii) $1 - \Phi(2.5) = 1 - 0.99379 = \boxed{0.0062.}$

(iii) $2\Phi(1.5) - 1 = 2(0.9332) - 1 = \boxed{0.8664.}$

Problem 4. In a large class, suppose your instructor tells you that you need to obtain a grade in the top 10% of your class to get an A on a particular exam. From past experience she is able to estimate that the mean and standard deviation on this exam will be 72 and 13, respectively. What will be the minimum grade needed to obtain an A? (Assume that the grades will be approximately normally distributed.)

Solution: Let X be the grade of a randomly picked student. We are given that $X \sim \mathcal{N}(72, 13^2)$. To get an A, need a minimum grade c such that

$$P[X > c] = 10\% = 0.1.$$

Equivalently, need

$$F_X(c) = P[X \le c] = 90\% = 0.9.$$

Now, $F_X(c) = \Phi\left(\frac{c-72}{13}\right)$. From the Φ table, we need $\frac{c-72}{13} = 1.28$. Therefore,

$$c = 1.28(13) + 72 = \boxed{88.64.}$$

Thus, if a student receive an 89 or higher, he or she can expect to be in the top 10% (which means he or she gets an A).

Problem 5. The incomes of junior executives in a large corporation are approximately normally distributed. A pending cutback will not discharge those junior executives with earnings within \$4900 of the mean. If this represents the middle 80% of the incomes, what is the standard deviation for the salaries of this group of junior executives?

Solution: Denote the income by X. It is given that $X \sim \mathcal{N}(m, \sigma^2)$ for some m and σ and that

$$P[m - \$4900 \le X \le m + \$4900] = 0.8.$$

Let $Z = \frac{X-m}{\sigma}$. Then, $Z \sim \mathcal{N}(0,1)$ and

$$P[m - \$4900 \le X \le m + \$4900] = P\left[\frac{-\$4900}{\sigma} \le Z \le \frac{\$4900}{\sigma}\right].$$

Now, observe that for any positive number c

$$P[-c \le Z \le c] = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1.$$

In this case, we want to find c such that $P[-c \le Z \le c] = 0.8$ which is the same as $\Phi(c) = \frac{0.8+1}{2} = 0.9$. From the Φ table, we get $c = 1.28$. The value of σ that would make $\frac{\$4900}{\sigma} = c$ is given by

$$\sigma = \frac{\$4900}{c} \approx \boxed{\$3,828}$$

That is, the current standard deviation for the salaries of junior executives is approximately \$3,828.

Problem 6. Hours That Children Watch Television: A. C. Neilsen reported that children between the ages of 2 and 5 watch an average of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 26.3 hours. [Source: Michael D. Shook and Robert L. Shook, The Book of Odds.]

Solution: Since the variable is approximately normally distributed, the distribution of sample means will be approximately normal, with a mean of 25. The standard deviation of the sample means is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{20}} = 0.6708.$$

Therefore,

$$P\left[\overline{X} > 26.3\right] = 1 - P\left[\overline{X} \leq 26.3\right] = 1 - \Phi\left(\frac{26.3-25}{0.6708}\right) = 1 - \Phi(1.9379)$$
$$\approx 1 - 0.973810 = \boxed{0.02619.}$$

Problem 7. Meat Consumption: The average number of pounds of meat that a person consumes per year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal. [Source: Michael D. Shook and Robert L. Shook, The Book of Odds.]

(a) Find the probability that a person selected at random consumes less than 224 pounds per year.

(b) If a sample of 40 individuals is selected, find the probability that the mean of the sample will be less than 224 pounds per year.

Solution:

(a) $P[X < 224] = \Phi\left(\frac{224-218.4}{25}\right) \approx \Phi(0.22) \approx \boxed{0.587064.}$

(b) $P\left[\overline{X} < 224\right] = \Phi\left(\frac{224-218.4}{25/\sqrt{40}}\right) \approx \Phi(1.42) \approx \boxed{0.922196.}$

Comparing the two probabilities, you can see that the probability of selecting an individual who consumes less than 224 pounds of meat per year is 58.71%, but the probability of selecting a sample of 40 people with a mean consumption of meat that is less than 224 pounds per year is 92.22%. This rather large difference is due to the fact that the distribution of sample means is much less variable than the distribution of individual data values. (Note: An individual person is the equivalent of saying $n = 1$.)

Problem 8. The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that the mean of their age is between 90 and 100 months. [Source: Harpers Index.]

Solution: Since the sample is 30 or larger, the normality assumption is not necessary. The sample mean will be approximately gaussian by the CLT.

$$
\begin{aligned}
P\left[90<\overline{X}<100\right] &= \Phi\left(\frac{100-96}{16/\sqrt{36}}\right)-\Phi\left(\frac{90-96}{16/\sqrt{36}}\right) \\
&= \Phi(1.5)-\Phi(-2.25)=\Phi(1.5)-(1-\Phi(2.25)) \\
&= \Phi(1.5)+\Phi(2.25)-1 \\
&\approx 0.933193+0.987776-1=\boxed{0.920969.}
\end{aligned}
$$

Problem 9. Consider Bernoulli trials whose success probability for each trial is p. Let X_i be the result (1 = success; 0 = failure) of the ith trial. Then, the X_i are Bernoulli random variables. Note that the number of successes in n trials can be found by adding the values of $X_1, X_2, \ldots, X_n$. Recall that this number is a binomial(n,p) random variable. Regarding it as a summation of n Bernoulli random variables suggests that it can be approximated by Gaussian random variable with the same expected value and variance. Use this approximation to solve the following problem:

A student has passed a final exam by supplying correct answers for 26 out of 50 multiple-choice questions. For each question, there was a choice of three possible answers, of which only one was correct. The student claims not to have learned anything in the course and not to have studied for the exam, and says that his correct answers are the product of guesswork. Use the Φ table to determine whether you should believe him.

Hint:

(a) Calculate the probability of identifying 26 or more correct answers through guesswork. If this probability is small, you judge that the student is bluffing.

(b) Recall that the expected value and variance of Binomial(n,p) random variable are np and $np(1-p)$, respectively.

Solution: If all the answers are guessed at, then the number of correct answers can be seen as the number of successes in $n = 50$ independent trials of a Bernoulli experiment having a success probability of $p = 1/3$. The binomial probability model is thus applicable.

A generally useful method of determining whether 26 correct answers is exceptional is based on finding out how many standard deviations lie between the observed number of correct answers achieved and the expected number. To do so, a quick approach is to approximate the binomial distribution with parameters $n = 50$ and $p = 1/3$ by a normal distribution with expected value $np = 50/3 \approx 16.67$ and standard deviation $\sqrt{np(1-p)} = 10/3 \approx 3.33$. So, the observed value of 26 correct answers lies

$$\frac{26-\frac{50}{3}}{\frac{10}{3}} = 2.8$$

standard deviations above the expected value.

Note that it is useful to remember as a rule of thumb that the probability of a normally distributed random variable taking on a value lying three or more standard deviations above the expected value is very small (the probability is 0.00135). To be more precise, we get $1 - \Phi(2.8) \approx 1 - 0.99744 = 0.0026$.

From the analysis above, we see that the probability of such a deviation occurring is quite small. There is very good reason, therefore, to suppose that the student is bluffing, and that he in fact did prepare for the exam.