Statistics for Final Year Engineering Students
SOLUTION TO: Intermediate Class Test, Monday May 9, 2011.
Lecturer: Dr. Yoni Nazarathy.

Last Updated May 23,2011

1) A school bus has capacity for 50 students. The mean weight of a student is 55 Kg and the standard deviation is 10 Kg.

   a. What is the probability that the total weight of students on a full school bus exceeds 2800 Kg?

Result: 0.2389

$$X \sim Normal(50 * 55 = 2750,\ 50 * 10^2 = 70.71^2)$$

$$P(X > 2800) = P\left(Z > \frac{2800 - 2750}{70.71}\right) = 1 - \phi(.71) = 1 - .7611 = .2389$$

   b. State the statistical assumptions required for your answer in (a) to hold.

Assumptions: The total weight is asymptotically normally distributed. The assumption is that this “asymptote” is reached.
Another assumption is that the weights of students are independent.

2) You use a computer to generate 10 independent uniform random variables in the range [0,1]. You then throw away any random variable that is greater than 4/5, keeping only the random variables that are less than or equal to 4/5. What is the probability that you are left with 8 or less numbers? Calculate a precise numeric answer.

Result: 0.6242

$$X \sim Bin(10, \frac{4}{5})$$

$$P(X \leq 8) = 1 - P(X = 10) - P(X = 9) = 1 - \left(\frac{4}{5}\right)^{10} - 10\left(\frac{4}{5}\right)^{9}\frac{1}{5} = 0.6242$$

3) A mini-school-bus contains 7 seats numbered 1 to 7. On the way to school it stops at 4 stations, picking up one student per station. Each student that enters the bus picks a random empty seat. What is the probability that upon arrival to the school there is exactly one empty odd numbered seat?

Result: 12/35=0.3429

$$\textit{number of odd seats taken} = X \sim HG(4\ \textit{stations}, 4\ \textit{odd seats}, 3\ \textit{even seats})$$

$$P(\textit{only one empty odd seat}) = P(X=3) = \frac{\binom{4}{3}\binom{3}{4-3}}{\binom{4+3}{4}} = \frac{4*3}{35} = \frac{12}{35}$$

4) Let $c > 0$ and consider a random variable X that has the following density function:

$$f(x) = \begin{cases} c\,x & 0 \le x \le 2 \\ 0 & otherwise \end{cases}.$$

a) What is the value of c?

Result: c=1/2

Area of triangle of base 2 and height 2c is 2*2c/2. Need 2*2c/2=1. So c= 1/2.

b) What is the expected value of X?

Result: 4/3

$$E[X] = \int_0^2 \frac{1}{2}x\,x\ dx = \frac{1}{6}x^3|_{x=0}^{x=2} = \frac{4}{3}$$

c) Write the CDF (Cumulative Distribution Function) of X (it should be defined for all real x).

Result:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \le x \le 2 \\ 1 & 2 \le x \end{cases}$$

5) Let $X_1, X_2, \ldots$ be a sequence of numbers. As you know, this is the formula for the sample variance:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Write a formula representing S_{n+1}^2 as a function of $\sum_{i=1}^{n} X_i$, $\sum_{i=1}^{n} X_i^2$, X_{n+1} and n.
That is, find and expression for f() where,

$$S_{n+1}^2 = f\left(\ \sum_{i=1}^{n} X_i,\ \sum_{i=1}^{n} X_i^2,\ X_{n+1},\ n\ \right).$$

Result: $\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 + X_{n+1}^2 - \frac{1}{n+1}\left(\sum_{i=1}^{n} X_i + X_{n+1}\right)^2\right)$

$$S_{n+1}^2 = \frac{1}{n}\sum_{i=1}^{n+1}(X_i - \bar{X})^2 = \frac{1}{n}\left(\sum_{i=1}^{n+1} X_i^2 - (n+1)\left(\frac{\sum_{i=1}^{n+1} X_i}{n+1}\right)^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 + X_{n+1}^2 - \frac{1}{n+1}\left(\sum_{i=1}^{n} X_i + X_{n+1}\right)^2\right)$$

6) A random sample of 1000 red apples in a super market chain contains 12 rotten apples. An apprentice statistician at the super market chain uses the following formula for confidence intervals:

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}},$$

with $\alpha = 0.05$. Note that $z_{.975} = 1.96$.

The statistician reports to management that there is a 95% chance that the proportion of rotten apples in the whole super market chain is in the range [-0.008, 0.032].
Which of the following statements is correct?
(circle the correct answers – there may be more than one).

a) The statistician has made a calculation mistake in the above formula. TRUE
The correct range is [0.005,0.019]

b) The confidence interval contains negative values because it uses the normal approximation to the binomial distribution and $\hat{p}$ is small while n is not "big enough". Would be true if confidence interval would contain negative values) – answer here does not affect grading.

c) Increasing the confidence level to a high enough level will ensure that the confidence interval is strictly positive. FALSE. Yet answer here does not affect grading.

d) Decreasing the confidence level to a low enough level will ensure that the confidence interval is strictly positive. TRUE (but no points deducted if left out).

e) The confidence interval contains negative values because $\hat{p}$ is a biased estimator. FALSE.

7) Weights of items shipped by a mail-order company are normally distributed with a mean of 20Kg and a variance of 4 Kg. Items weighing more than 22 Kilograms incur an additional cost to

the company of 15$ per item. The company is shipping 100 items. What is the mean value of additional costs?

Result: 238.05 Brief Explanation: p=P(Weight>22)=P(Z>1)=1-$\phi(1) = 1 - .8413 = 0.1587$. Number of items with additional cost = X$\sim Bin(100, p)$ E[cost] = E[15 X] = 15 E[X] = 15 * 100 * 0.1587=238.05.

8) Let X_1 and X_2 be independent random variables each taking values in the sets $\{1, \dots, \mathrm{n}\}$ with equal probability (e.g. If n=6 these are the results of independent die throws).
Find a simple formula in terms of n for the mean of $\mathrm{Y} = \mathrm{X}_1 + \mathrm{X}_2$. Show your working.

Final Result: n+1 Working: $for\ j = 1{,}2,\ \ E[X_j] = \sum_{i=1}^{n} \frac{1}{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$. $E[Y] = E[X_1 + X_2] = E[X_1] + E[X_2] = n + 1$

(Reminder: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$).

Good Luck.