Homework 1

José Miguel Ponciano

August 24, 2015

The following homework is due Tuesday Sept 15th. Please send me your answers as a single pdf document. For the questions that require writing math, you may scan your pencil and paper answers if you don't know a word processor capable of including math.

1. Read the "Appendix 1: a basic probability review" of the class notes (pages 55 to 80). Do exercises 1.2 (p. 61), 1.3 (p. 64), 1.4 (p. 65), 1.5 (p. 68), 1.6 (p. 70), 1.8 (p.73), 1.9 (p 78).

2. Consider the following joint distribution of X and Y:

X =	Y=1	2	Total
1	0.3	0.1	0.4
2	0.4	0.2	0.6
Total	0.7	0.3	1

   Compute:

   a) $E(X)$, $E(Y)$

   b) $Var(X)$, $Var(Y)$

   c) $E(X^2)$, $E(Y^2)$

   d) $Cov(X,Y)$, $Corr(X,Y)$

3. While modeling demographic stochasticity and survival in Mark-Recapture analysis, the chance of survival of 1 individual is modeled with a simple Bernoulli distribution where $P(X=x) = \begin{cases} p \text{ if } x=1 \\ 1-p \text{ if } x=0 \end{cases}$

   When $X = x = 1$ the individual survives, and when $X = x = 0$ the individual dies. Suppose we are interested in modelling the annual survival for a population of n animals. The number of survivors from one year to the next can be modelled using a Bernoulli distribution with probability p. Then, the total number of survivors for the next year is a random variable denoted by S. Let X_i be 1 if individual i survives and 0 if it does not for next year ($i = 0, 1, 2, \cdots, n$). Assume that each individual survives/dies independently from each other.

   a) Compute the expected number of survivors for the next generation, $E(S)$.

   b) Compute the covariance of the number of survivors, $Var(S)$.

   c) If environmental conditions induce a correlation in survival such that $Cov(X_i, X_j) = \tau \neq 0$, $i \neg j$ and $i, j = 1, 2, \cdots, n$, compute $Var(S)$.

4. Examining the properties of the of the sample mean and sample variance:

   Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random samples from a population of interest (e.g. a quantitative character, the heights of adult women in Florida), that can be modelled with a Normal distribution $N(\mu, \sigma^2)$. The estimate of the variance, σ^2 is denoted as $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Note that S is a random variable (each time you get a single $x_1, x_2, \cdots, x_n$, you get a different value s^2 of the sample covariance). A well known property of S^2 is that it is an unbiased estimator, that is $E(S^2 - \sigma^2) = 0$. Prove that this property is true. (Hint: $\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \mu + \overline{X} - \overline{X})^2$ is equivalent to $\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left((X_i - \overline{X}) + (\overline{X} - \mu)\right)^2$. Develop the right hand side as $(a+b)^2 = a^2 + 2ab + b^2$ and take the expectation (the expected value) in both sides of the equation.)

5. The Normal Distribution:

   The pdf of normal distribution with mean μ and variance σ^2 is given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

   (a) Using R draw the pdf with the the $\mu = 2$, 4, and 6 all with $\sigma^2 = 1.5$

   (b) Using R find the proportions of ACT scores between 20 and 30. ACT scores can be modeled with a Normal distribution with $\mu = 21.37$ and standard deviation $\sqrt{\sigma^2} = 4.55$.

6. Simulation experiment: For each of the following sample sizes from a normal distribution with $\mu = 100$ and variance $\sigma^2 = 30$, and sample sizes $n_i = (5, 10, 50, 100, 1000)$.

   (a) Simulate 500 random samples for each size n_i. For each of the 500 random samples above, compute the mean (using the R function `mean()`).

   (b) Plot the histogram of the 500 sampled means, for each of the $n_i's$.

   (c) Compare the 5 histograms of the sampling distributions of the sample mean and comment (reassert the Law of Large numbers and the Central Limit Theorem).

   (d) What is the theoretical mean of the distribution of the sample mean? What is the pdf of the distribution of the sample mean?

7. Read the "Rintro.pdf" notes and do: exercises 1,2, and 3 p. 3; 1-4 p. 5; 1-3 p. 7; 1 p. 9; 1-3 p. 11; 1 and 2 p. 12; 1 p. 13; 1-3 p. 17. Bonus: ex. 1 p. 18 and ex. 1 p. 19.