FALL 2010

Stat 471: Homework Assignment 1, each question worth 12.5Due: September 24, 2010

In this assignment you are asked to produce several plots comparisoning density functions, probability mass functions, and cumulative distribution functions for a continuous and a discrete distribution. The particular distributions were assigned at random resulting in

	Name	Seed	continuous	discrete
1	Ansari,Mohammed Hidayath	2	Cauchy	Binomial
2	Ballard,Kathryn	43	Logistic	Negative Binomial
3	Bernard,Kwame Ajani	54	F	Poisson
4	Chisholm,John P	50	Cauchy	Poisson
5	Corby,Allison Fischer	11	Uniform	Binomial
6	De Carli,Lorenzo	18	F	Geometric
7	Floratou,Avrilia	15	Chi-square	Geometric
8	Fu,Rao	55	Logistic	Poisson
9	Guan,Yu	57	Normal	Poisson
10	Hao,Beini	14	Cauchy	Geometric
11	Hemken,Douglas Edward	60	Weibull	Poisson
12	Hu,Qizheng	12	Weibull	Binomial
13	Hu,Zifei	42	F	Negative Binomial
14	Kim,Junhoi	41	Exponential	Negative Binomial
15	Krco,Matthew John	45	Normal	Negative Binomial
16	Kreibich,Alex Richard	27	Chi-square	Hypergeometric
17	Lee,Ahlam	58	T	Poisson
18	Li,Ao	6	F	Binomial
19	Mater,Patrick Joseph	59	Uniform	Poisson
20	Milliman,Katelyn Elise	37	Beta	Negative Binomial
21	Raza,Syed Umair Sadaat	5	Exponential	Binomial
22	Roubert,Lisha Marie	8	Log-normal	Binomial
23	Sanchez,Fabrizzio Andrew	26	Cauchy	Hypergeometric
24	Sewankambo,Yvonne	36	Weibull	Hypergeometric
25	Shang Kuan,Long Yen	4	Gamma	Binomial
26	Song,Kaiyuan	48	Weibull	Negative Binomial
27	Stenhaug,Benjamin Alan	39	Chi-square	Negative Binomial
28	Su,Christin	9	Normal	Binomial
29	Thompson,Meredith Ellen	35	Uniform	Hypergeometric
30	Winternheimer,Julian Craig	51	Chi-square	Poisson
31	Zhang,Wenzhao	33	Normal	Hypergeometric
32	Zhang,Yiying	46	T	Negative Binomial
33	Zhao,Jinglin	7	Logistic	Binomial

Table 1: Assignment of distributions and random seeds

Most of the questions require you to provide plots. Please remember to submit your answers (plots and whatever explanations are required) as a PDF file. In this and all other homework assignments you should, in a separate text file, submit the commented R code that you used to produce the results.

To provide reproducibility (so that I can check your results if needed) set the random number seed to the value in the Seed column in the table before any question that requires you to simulate values. You can do it once only if you provide a complete R script for the assignment — otherwise set the seed before each simulation.

1. For your continuous distribution, determine appropriate values of the parameters in the distribution (you can use the values from the Wikipedia page, if you wish) and provide a comparative density plot and a comparative c.d.f. plot for at least 4 different combinations of the parameters.

2. For your discrete distribution provide a comparative plot of the probability mass function for at least four different combinations of the parameter values. (Bonus question for 10%: Provide a comparative plot of the cumulative distribution function.)

3. Simulate data from your continuous distribution and evaluate the sample mean and sample variance from the sample. Do this for at least 6 different combinations of the values of the parameters in the distribution. Find the formula for the mean and the variance for this distribution from Wikipedia and demonstrate the empirical values you calculated are close to the theoretical values. State the number of replications that you used and why you chose this value.

4. Repeat the previous question for your discrete distribution.

5. Simulate at least 10000 replications of the sample mean from samples of size 9, 16, 25 and 36 from your continuous distribution for **one** set of parameter values. Provide a comparative (empirical) density plot of the sample means. Evaluate the mean and standard deviation of each set of sample means (that is, the mean of the sample means from samples of size 9, the mean of the sample means from samples of size 16, ...). Relate the means and standard deviations to theoretical mean and the theoretical standard error for this distribution.

6. Determine 200 equally spaced (on the probability scale) quantiles from your means of samples of size 9, of size 16, etc. If your means of samples of size 9 are called `mns9` then the expression to evaluate this is

   ```
   > quant9 <- quantile(mns9, ppoints(200), names = FALSE)
   ```

   Provide a comparative normal quantile-quantile plot for each of these sets. This plot should show that larger sample sizes produce more "normal-like" behavior and less variability of the sample means.

7. Repeat the last two questions for your discrete distribution.

8. Compare the distribution of the sample mean and the sample median for a **standard normal** distribution. Simulate at least 10000 replications of each sample statistic from samples of size 9, 16, 25 and 36. For each sample size graphically compare the density of the sample mean and the sample median. Evaluate the mean and the standard deviation of the sample means and the sample medians for each sample size. Determine the interval that contains the middle 80%, 90%, 95% and 99% of the sample values for each combination of statistic and sample size. The sample mean is a minimum variance unbiased estimator of the theoretical mean. Does it have less variability than the median? The R expression to evaluate the middle 80% of, say, `mns9`, the sample means from samples of size 9, is

   ```
   > quantile(mns9, c(0.1, 0.9))
   ```