**Machine Learning HW 2**

**Ang Li, anl125@pitt.edu**

**Problem 1. Mean estimates and the effect of the sample size**In this problem we study the influence of the sample size on the estimate of the mean.The data for this experiment are in file *mean study data.txt* in the homework assignmentfolder. The data were generated from the normal distribution with mean=15 and standarddeviation=5.

**(Part 1)** Load the data in the mean study data.txt Calculate and report the meanand standard deviation of the data. Compare them to the true mean and std above.

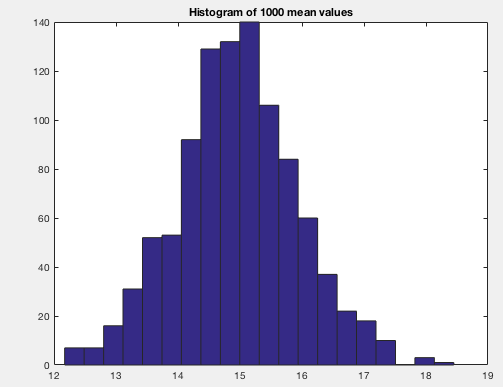
*mean = 15.0415*

*standard deviation = 5.0279*

**(Part 2)**Write (and submit) a function [newdata] = subsample(data, k) that randomly selects k instances from the data in the *mean study data.txt*

**(Part 3)** Use the above function to randomly generate 1000 subsamples of the data ofsize 25. For each subsample calculate its mean and save the results in the vector of1000 means. Plot a histogram of 1000 mean values using 20 bins.

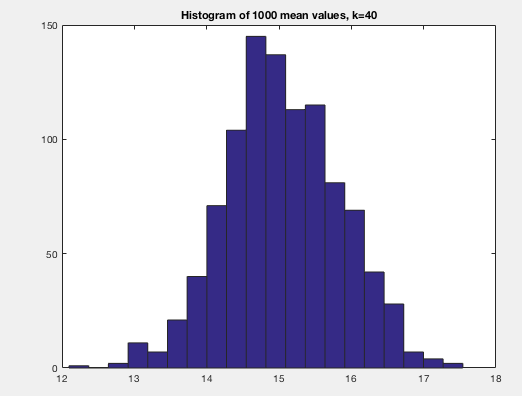
**(Part 4)** Include the histogram in your report. Analyze and the means calculated on1000 subsamples of size 25 and compare them to the true mean and the mean thatwas calculated in step 1 on all examples in the dataset. Report your observations.



*Mean of these 1000 subsampled (k=25) means is: 14.9584.*

*This number is close to both true mean and the mean reported in step 1*

**(Part 5)** Repeat steps form part 3 but now generate 1000 subsamples of size 40.Include the histogram in the report and compare it to the histogram generated inpart 4 for subsamples of size 25, and to the mean of the original data. What are thedifferences? What conclusions can you make by comparing the means for subsamplesof size 25 and 40.



*Mean of these 1000 subsampled (k=40) means is: 15.0511.*

*This number is closer to original mean (15.0415) reported in step 1 than use the sample size k=25*

*In addition, compare with two histogram, the variation from k=40 is less than the variance from k=25, which indicate a better estimation of original sample mean.*

*The conclusion is larger sample size the better to estimate the original sample mean.*

**(6)** Take first 25 examples from the original data in the mean study data.txt and calculatetheir mean. Use the function t-test to calculate and report the 0.95 confidenceinterval for the mean estimates. Does the true mean value fall into the 0.95 confidenceinterval?

*mean(sample25) = 14.5652 with confident interval: ci = [12.6329, 16.4974]*

*yes, true mean fall into the confident interval*

**Problem 2. Train-test splitting using k-fold cross validation**When testing the performance of a learning algorithm using a simple holdout method theresults may be biased by the training/testing data split. To alleviate the problem various random resampling schemes, such as k-fold cross-validation, random subsampling or bootstrap (see lecture notes) can be applied to estimate the statistics of interest by averaging the results across multiple train/test splits. Please do the following tasks:

**(Part 1)** Please write and submit the function: *[train test] = kfold­\_crossvalidation(data, k,m)* that takes the data, **k (the number of folds)** and **m (the target fold)** as inputs, and returns the training and testing data sets, such that the testing set corresponds to m-th fold under the k-th fold cross-validation scheme. To implement the procedure, please place the folds over indexes of the data, by assuring that each fold has equal number of entries that do not overlap. If this is not possible, the fold sizes (number of instances in each fold) should differ by at most one. The file should be named *kfold\_crossvalidation.m.*

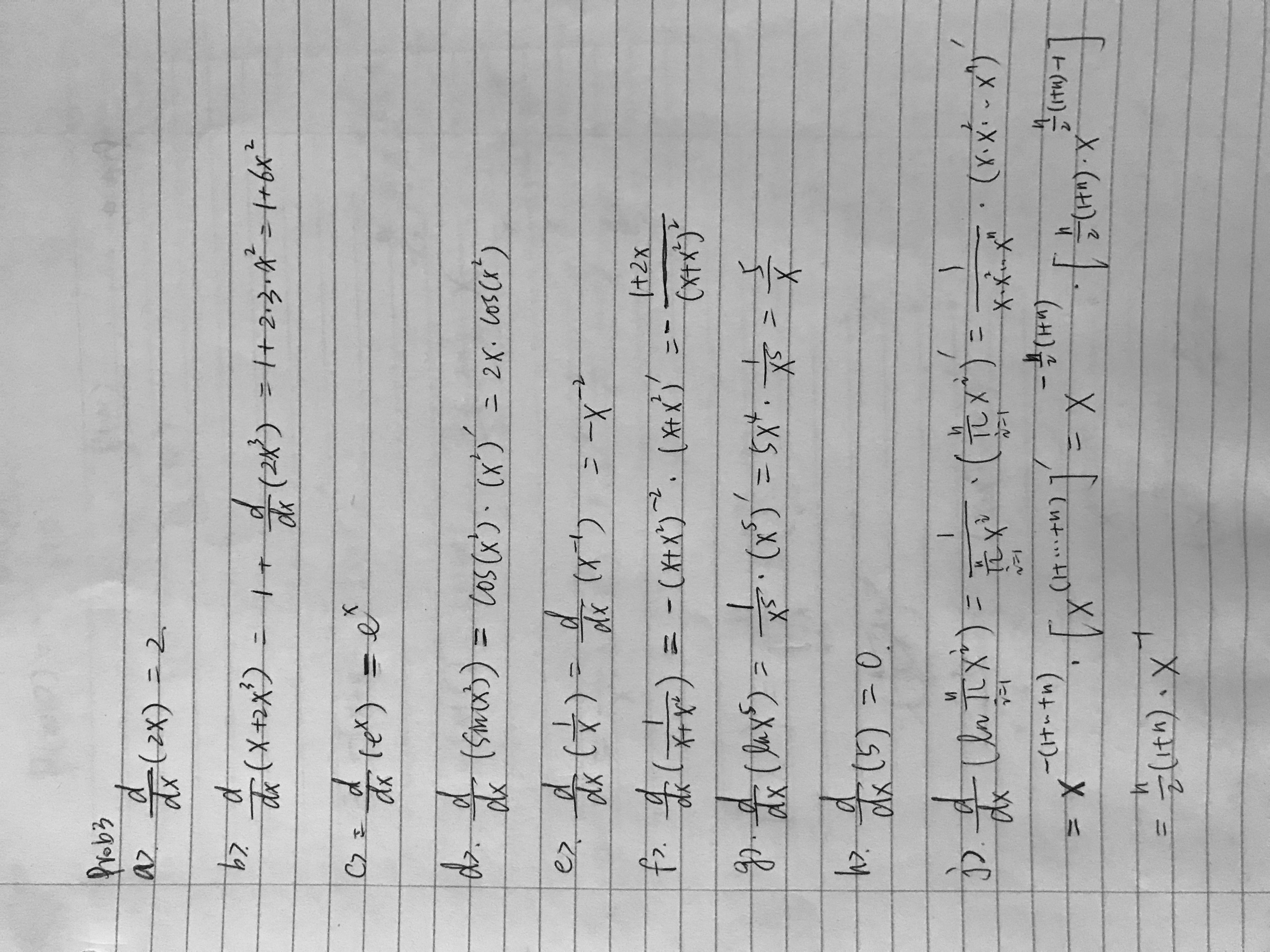
*See code: kfold\_crossvalidation.m*

**(Part 2).** Run/test your function on data in the file resampling data.txt. Morespecifically, run your *kfold\_crossvalidation* function on all data in the file by settingk (number of folds) to 10 and by varying the test fold index (parameter m) from 1,to 10. For each test data (generated for the different value of m) that were returnedby your function calculate the mean and std and report them.

*mean(mean\_tests) = 2.3271*

*std(std\_tests) = 0.5513*

**Problem 3. Function derivatives**



**Problem 4. Probabilities**Part a. Assume you have 2 fair dice. What are the probabilities associated with the differentoutcomes that are obtained by summing together the numbers on the two dice?

*There are total of 6\*6 = 36 possible outcomes*

*Sum = 2 = (1, 1); probability of sum=2 is: 1/36*

*Sum = 3 = [(1, 2); (2, 1)]; probability of sum = 3 is: 2/36*

*Sum = 4 = [(1,3); (3,1); (2,2)]; probability of sum = 4 is: 3/36*

*Sum = 5 = [(1,4); (4,1); (2,3) ; (3,2)]; probability of sum = 5 is: 4/36*

*Sum = 6 = [(1,5); (5,1); (2,4) ; (4,2) ; (3,3)]; probability of sum = 6 is: 5/36*

*Sum = 7 = [(1,6); (6,1); (2,5) ; (5,2) ; (3,4) ; (4,3)]; probability of sum = 7 is: 6/36*

*Sum = 8 = [(2,6) ; (6,2) ; (3,5) ; (5,3); (4,4)]; probability of sum = 8 is: 5/36*

*Sum = 9 = [(3,6) ; (6,3) ; (4,5) ; (5,4)]; probability of sum = 9 is: 4/36*

*Sum = 10 = [(4,6) ; (6,4) ; (5,5)]; probability of sum = 10 is: 3/36*

*Sum = 11 = [(5,6) ; (6,5)]; probability of sum = 11 is: 2/36*

*Sum = 12 = [(6,6)]; probability of sum = 12 is: 1/36*

Part b. Calculate the expected value of the outcome for the 2 fair dice roll experiment.*E = 2\*(1/36) + 3\*(2/36) + 4\*(3/36) + 5\*(4/36) + 6\*(5/36) + 7\*(6/36) + 8\*(5/36) + 9\*(4/36) + 10\*(3/36) + 11\*(2/36) + 12\*(1/36) = 7*

Part c. Assume you play the two dice game from part a. 5 times. What is the probability,we never see the outcome of 4? What is the probability we see even-sum outcomes in all 5trials.

*Toss once the probability of not equal to 4 is: 1-3/36 = 11/12 = 91.6%*

*Toss 5 times, never see the outcome of 4: (11/12)^5 = 64.7%*

*Toss once, the probability getting even-sum: ½ =50%*

*Toss 5 times, all event outcomes: (1/2)^5 = 0.031*

**Problem 5. Bernoulli trials**Assume we have conducted a coin toss experiment with 100 coin flips. The results of theexperiment are in file ’coin.txt’ where 1 means a head and 0 means a tail. Assume that µrepresents the probability of observing a head.

(a) What is the ML estimate of µ?

*ML\_mu = sum(data)/100 = 0.6500*

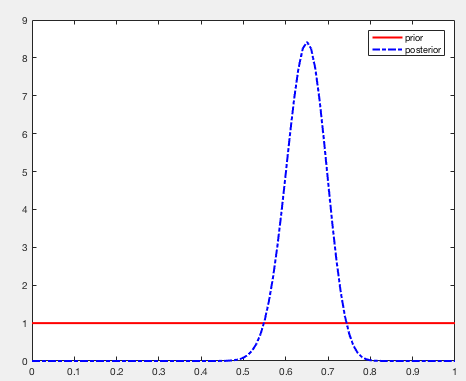
(b) Assume the prior on µ is defined by a Beta distribution Beta(µ|1, 1). Plot andreport both the prior and the posterior distributions on µ.

*Prior distribution ~ Beta(µ|α1 = 1, α2 =1)*

*N1 = Sum\_Head = sum(data)=65*

*N2 = Sum\_tail = 100-65 = 35*

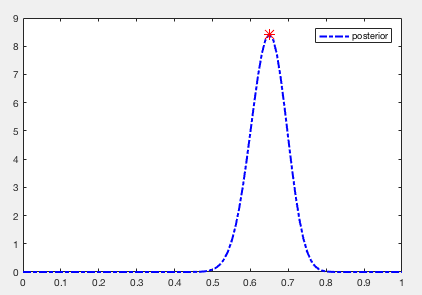
*posterior distributions ~ Beta(µ|1+65, 1+35) = Beta(µ|66, 36)*



(c) Calculate and report the MAP estimate of the µ using the prior in Part b.

Show (plot) the MAP estimate on the plot of the posterior of µ you have generated in part b.

*mu\_MAP = (N1 + α1 - 1)/(N1 + N2 + α1 + α2 - 2) = (65 + 1 - 1)/(100 + 1 + 1 -2) = 0.65*



*In this plot, I marked the MAP estimate as red star.*

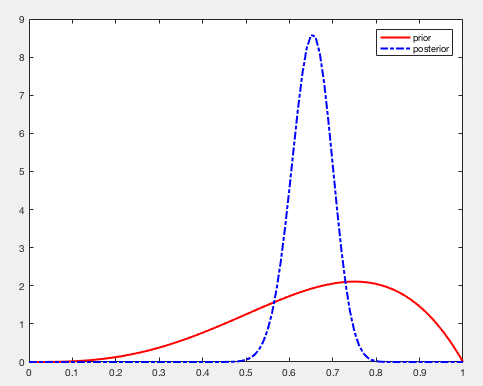
(d) Repeat part b and c by assuming that the prior on µ follows Beta(µ|4, 2).

*Prior distribution ~ Beta(µ|α1 = 4, α2 =2)*

*N1 = Sum\_Head = sum(data)=65*

*N2 = Sum\_tail = 100-65 = 35*

*posterior distributions ~ Beta(µ|4+65, 2+35) = Beta(µ|69, 37)*



*mu\_MAP = (N1 + α1 - 1)/(N1 + N2 + α1 + α2 - 2) = (65 + 4 - 1)/(100 + 4 + 2 -2) = 0.6538*

