BST 140.651
Problem Set 2

Problem 1. Using the rules of expectations prove that $\mathrm{Var}(X) = E[X^2] - E[X]^2$ where $\mathrm{Var}(X) = E[(X-\mu)^2]$.

Problem 2. Let $g(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)$ where f_1 and f_2 are densities with associated means and variances μ_1, σ_1^2, μ_2, σ_2^2, μ_3, σ_3^2, respectively. Here $\pi_1, \pi_2, \pi_3 \geq 0$ and $\sum_{i=1}^3 \pi_i = 1$. Show that g is a valid density. What is it's associated mean and variance?

Problem 3. Suppose that a density is of the form $(k+1)x^k$ for some constant $k > 1$ and $0 < x < 1$.

a. What is the mean of this distribution?

b. What is the variance?

Problem 4. Suppose that the time in days until hospital discharge for a certain patient population follows a density $f(x) = \frac{1}{3.3}\exp(-x/3.3)$ for $x > 0$.

a. Find the mean and variance of this distribution.

b. The general form of this density (the exponential density) is $f(x) = \frac{1}{\beta}\exp(-x/\beta)$ for $x > 0$ for a fixed value of β. Calculate the mean and variance of this density.

c. Plot the exponential pdf for $\beta = 0.1, 1, 10$.

Problem 5. The Gamma density is given by

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}\exp(-x/\beta) \quad \text{for} \quad x > 0$$

for fixed values of α and β.

a. Derive the mean and variance of the gamma density. You can assume the fact (proved in HW 1) that the density integrates to 1 for any $\alpha > 0$ and $\beta > 0$.

b. The Chi-squared density is the special case of the Gamma density where $\beta = 2$ and $\alpha = p/2$ for some fixed value of p (called the "degrees of freedon"). Calculate the mean and variance of the Chi-squared density.

Problem 6. The Beta density is given by

$$\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for} \quad 0 < x < 1$$

and $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

a. Derive the mean of the beta density. Note the following is useful for simplifying results: $\Gamma(c+1) = c\Gamma(c)$ for $c > 0$.

b. Derive the variance of the beta density.

Problem 7. The Poisson mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for} \quad x = 0, 1, 2, 3, \ldots$$

a. Derive the mean of this mass function.

b. Derive the variance of this mass function. Hint, consider $E[X(X-1)]$.

Problem 8. Suppose that, for a randomly drawn subject from a particular population, the proportion of a their skin that is covered in freckles follows a uniform density (constant between 0 and 1).

a. What is the expected percentage of a (randomly selected) person's body that is covered in freckles? (Show your work.)

b. What is the variance? (Show your work.)

Problem 9. You have an MP3 player with a total of 1000 songs stored on it. Suppose that songs are played randomly *with replacement*. Let X be the number of songs played until you hear a repeated song.

a. What values can X take, and with what probabilities?

b. What is the expected value for X?

c. What is the variance for X?

Problem 10. When at the free-throw line for two shots, a basketball player makes at least one free throw 90% of the time. 80% of the time, the player makes the first shot, while 70% of the time she makes both shots.

a. Does it appear that the player's second shot success is independent of the first?

b. What is the conditional probability that the player makes the second shot given that she made the first? What would it be if she missed the first?

Problem 11. Assume that an act of intercourse between an HIV infected person and a non-infected person results in a $1/500$ probability of spreading the infection. How many acts of intercourse would an uninfected person have to have with an infected persons to have a 10% probability of obtaining an infection? State the assumptions of your calculations.

Problem 12. You meet a person at the bus stop and strike up a conversation. In the conversation, it is revealed that the person is a parent of two children and that one of the two children is a girl. However, you do not know the gender of the other child, nor whether the daughter she mentioned is the older or younger sibling.

a. What is the probability that the other sibling is a girl? What assumptions are you making to perform this calculation?

b. Later in the conversation, it becomes apparent that she was discussing the older sibling. Does this change your probability that the other sibling is a girl?

Problem 13. A particularly sadistic warden has three prisoners, A, B and C. He tells prisoner C that the sentences are such that two prisoners will be executed and let one free, though he will not say who has what sentence. Prisoner C convinces the warden to tell him the identity of one of the prisoners to be executed. The warden has the following strategy, which prisoner C is aware of. If C is sentenced to be let free, the warden flips a coin to pick between A and B and tells prisoner C that person's sentence. If C is sentenced to be executed he gives the identity of whichever of A or B is also sentenced to be executed.

a. Does this new information about one of the other prisoners give prisoner C any more information about his sentence?

b. The warden offers to let prisoner C switch sentences with the other prisoner whose sentence he has not identified. Should he switch?

Problem 14. The Chinese Mini-Mental Status Test (CMMS) is a test consisting of 114 items intended to identify people with Alzheimer's disease (AD) and dementia among people in China. An extensive clinical evaluation was performed of this instrument, whereby participants were interviewed by psychiatrists and nurses and a definitive (clinical) diagnosis of AD was made. The table below show the counts obtained on the subgroup of people with at least some formal education. Suppose a cutoff value of ≤ 20 on the test is used to identify people with AD.

	Clinical diagnosis of AD	
CMMS score	No	Yes
0-5	0	2
6-10	0	1
11-15	3	4
16-20	9	5
21-25	16	3
26-30	18	1

a. What is the sensitivity and specificity of the CMMS test using the 20 cutoff?

b. Create a plot of the sensitivity by (1 - specificity), which is the true positive rate versus the false positive rate for all of the cut-offs between 0 and 30. This is called an ROC curve.

c. Graph the positive predictive value as a function of the prevalence of AD. Do the same for the negative predictive value.

Problem 15. A web site (www.medicine.ox.ac.uk/bandolier/band64/b64-7.html) for home pregnancy tests cites the following:

> When the subjects using the test were women who collected and tested their own samples, the overall sensitivity was 75%. Specificity was also low, in the range 52% to 75%.

a. Interpret a positive and negative test result using diagnostic likelihood ratios using both extremes of the specificity.

b. A woman taking a home pregnancy test has a positive test. Draw a graph of the positive predictive value by the prior probability (prevalence) that the woman is pregnant. Assume the specificity is 63.5%

c. Repeat the previous question for a negative test and the negative predictive value.

Problem 16. Given below are the sexes of the children of 7,745 families of 4 children recorded in the archives of the Genealogical Society of the Church of Jesus Christ of Latter Day Saints in Salt Lake City, Utah. M indicates a male child and F indicates a female child.

Sequence	Freq	Sequence	Freq
MMMM	537	MFFM	526
MMMF	549	FMFM	498
MMFM	514	FFMM	490
MFMM	523	MFFF	429
FMMM	467	FMFF	451
MMFF	497	FFMF	456
MFMF	486	FFFM	441
FMMF	473	FFFF	408

a. Estimate the probability distribution of the number of male children, say X, in these families using the data below by calculating proportions.

b. Find the expected value of X.

c. Find the variance of X.

d. Find the probability distribution of $\hat{p}$, where $\hat{p}$ is the proportion of children in each family who are male. Find the expected value of $\hat{p}$ and the variance of $\hat{p}$

Problem 17. Quality control experts estimate that the time (in years) until a specific electronic part from an assembly line fails follows (a specific instance of) the **Pareto** density

$$\frac{3}{x^4} \qquad \text{for } 1 < x < \infty.$$

a. What is the average failure time for components from this density? (Show your work.)

b. What is the variance? (Show your work.)

c. The general form of the Pareto density is given by $\frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $0 < \alpha < x$ and $\beta > 0$ (for fixed α and β). Calculate the mean and variance of the general Pareto density.

Problem 18. You are playing a game with a friend where you flip a coin and if it comes up heads you give him a dollar and if it comes up tails she gives you a dollar. You play the game ten times.

a. What is the expected total earnings for you? (Show your work; state your assumptions.)

b. What is the variance of your total earnings? (Show your work; state your assumptions.)

c. Suppose that the coin is biased and you have a $.4$ chance of winning for each flip. repeat the calculations in parts a and b

Problem 19. Note that the code

```
temp <- matrix(sample(1 : 6, 1000 * 10, replace = TRUE), 1000)
xBar <- apply(temp, 1, mean)
```

In R produces $1,000$ averages of 10 die rolls. That is, it's like taking ten dice, rolling them, averaging the results and repeating this $1,000$ times.

a. Do this in R. Plot histograms of the averages.

b. Take the mean of `xBar`. What should this value be close to? (Explain your reasoning.)

c. Take the standard deviation of `xBar`. What should this value be close to? (Explain your reasoning.)

Problem 20. Note that the code

```
xBar <- apply(matrix(runif(1000 * 10), 1000), 1, mean)
```

produces $1,000$ averages of 10 uniforms.

a. Do this in R. Plot histograms of the averages.

b. Take the mean of `xBar`. What should this value be close to? (Explain your reasoning.)

c. Take the standard deviation of `xBar`. What should this value be close to? (Explain your reasoning.)