Practice Problems for Final, Probability and Statistics, Fall 2009

1. let X be the value you get when tossing a single fair die.
   (a) write down the probability mass function for X.
   (b) compute the mean of X, that is $E(X)$.
   (c) compute $E(X^2)$.
   (d) compute the variance of X.

2. Data suggests that a certain switchboard receives an average of 6 phone calls per minute. These calls arrive at random. What is the probability that it receives exactly 5 calls in a given minute?

3. Suppose that you are looking for 42 Olin students to survey for your project. By performing a difficult computation involving busyness ratios and helpfulness ratios, you know that the probability that any randomly selected Olin student will agree to take the survey is .90. What is the probability that you need to ask exactly 48 people in order to get exactly 42 people to agree to take the survey?

4. Suppose you have a random variable X that is known to have a normal distribution with mean 30 and variance 36.
   (a) Sketch the graph of the probability density function for X. Describe and label the graph with as much info as possible.
   (b) Now suppose that you take random samples of size 25 from the normal data, and for each size 25 sample, you form a sample mean $\bar{X}$. Sketch the graph of the probability density function for $\bar{X}$. Describe and label the graph with as much info as possible. Explain your reasoning.

5. If you wanted to test a claim about the mean of some distribution, what are two point estimators that might be reasonable to investigate? Describe two criteria that you could use to decide which estimator is the better estimator. (You need not apply the criteria to your estimators, simply describe the criteria.)

6. A manufacturer of fuses claims that with a 20% overload, the fuses will blow in 12.40 minutes on the average. To test this claim, a sample of 20 of the fuses was subjected to a 20% overload, and the times it took them to blow had a sample mean of 11.30 minutes and a sample standard deviation of 2.48 minutes. Assume that the data constitutes a random sample from a normal population. You want to test the null hypothesis H_0 that $\mu = 12.40$ at an $\alpha = .05$ level of significance.
   (a) If the alternative hypothesis H_a is that $\mu \neq 12.40$, does the above data give evidence to reject or fail to reject the null hypothesis?
   (b) If the alternative hypothesis H_a is that $\mu < 12.40$, does the above data give evidence to reject or fail to reject the null hypothesis?

7. A grand jury in Los Angeles has 40 members. If the jurors are selected at random from the adult population, of which 55% are female, what is the mean and standard deviation of the number of women jurors?

8. The continuous uniform distribution over the interval $[a, b]$ has the cumulative probability distribution function given by

$$\Phi(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x \end{cases}$$

   Calculate the mean, median, and variance of this uniform distribution.

9. Estimate the probability that, if 100 fair coins are tossed, between 40 and 60 heads will be observed. Describe your assumptions.

10. A hand of five cards is dealt at random from a standard deck. Given that the hand has three red cards and two black cards, what is the probability that the hand has three hearts ($\heartsuit$) and two clubs ($\clubsuit$)?

11. Consider the following gambling game, called the Petersburg Game: Toss a fair coin repeatedly until the first tail is observed. If the tail occurs on the first toss, you do not win anything; if the first tail occurs on the second toss, you win 2 dollars; if the first tail occurs on the third toss, you win 4 dollars. Thus, if the first tail occurs on toss $k+1$, you win 2^k dollars, $k \geq 1$. Compute the average duration of the game. Now compute the expected winnings of the player. Are you surprised by your results?

12. Five *distinguishable* flags are randomly placed onto 12 distinct flag poles, where any pole could hold up to 7 flags. What is the probability that all of the flags are on different poles? Would your answer change if the flags were all identical?

13. The average score for the SAT mathematics IC test in the year 2004 was 588 with a standard deviation of 98, and is approximately normally distributed. The average score of first-year students at a certain Boston Metro-West selective technical school on this test was 760. What percentile does this average correspond to?

14. As discussed in class, the time until failure of an incandescent light bulb follows an exponential probability density with expected value 1000 hours. Calculate the probability that a light bulb will last more than 2000 hours.

15. Larry Bird's lifetime free-throw percentage was about 88.6%; we will take this statistic to mean that, on any given free throw, Larry the Legend had a chance of $q = 0.114$ of missing the shot, and a chance of $p = 0.886$ of making it. For a game in which Larry took 20 free-throws, how unusual would it be for him to have made 15 or fewer of those shots?

16. Grant Fuhr played professional hockey for twenty years as a goalie with several National Hockey League teams (including the famous Edmonton Oilers teams that won the Stanley Cup with Wayne Gretzky). In the 1995–96 season, he allowed 209 goals in 218 periods of play. Argue why the number of goals allowed by Fuhr, per period, ought to have a Poisson distribution. On that assumption, compute the chance that Fuhr would allow more than two goals in a given period.

17. Three different airlines, called AM, UN and SW, fly out of Providence. AM airline has 60 flights per day, of which 15% are late departures. UN airline has 40 flights per day, of which 12% are late, and SW has 50 flights per day, of which 10% are late. You randomly hear someone at the airport complaining about their late flight, but do not hear them say which airline. What is the probability that they were traveling on SW airline?

18. While entering computer code in a long program, the probability per line of code of an error is $p = 0.01$. Estimate the probability that 1000 lines of computer code will have no more than 5 errors.

19. In clinical nursing care facilities, there are about twice as many women as men, since the average lifespan of women is considerably longer than for men. A medical research project is underway to study geriatric problems. If a group of 100 patients is selected at random for the study, estimate the probability that 75 or more of the group would be female? What would be the smallest group of patients that should be selected in order to make the chance less that 0.05 of having more than 70% of them female?

20. In machines for which a failure of a critical component would be catastrophic, engineers often design "redundant" components, so a total system failure in a given time interval would require independent failure of more than one component. A good example of this design are the navigational computers on the space shuttle. Suppose that the time until failure of a computer is exponentially distributed with a mean time of 96 hours. Of the five computers on the space shuttle, the mission can continue if at least three of them are functioning; if two or fewer are functioning, the mission is ended early. What is the probability that an eight-day shuttle mission will be terminated early? What would that probability be if the shuttle had only one computer?

21. Select a sample of size $n = 10$ from a normal population with true mean $\mu = 52$; the sample standard deviation is $s = 3$, and the sample average is $\bar{x} = 50$. What is the probability of observing a sample average from this population as, or smaller than, the one that is observed?

22. A random variable X has a double exponential density if its PDF is

$$f(x) = \frac{\theta}{2} e^{-\theta|x|}.$$

Determine the maximum likelihood estimator for the parameter α.

23. The duration of Alzheimer's disease ranges from 3 to 20 years. The average is 8 years and the standard deviation is 4 years. For a clinical study 30 patients are randomly selected that have been determined to be at the very beginning stage of the disease.

    (a) What is the probability that the average duration of the disease among the sampled patients will be less than 7 years? Discuss your assumptions.
    (b) What is the probability that the average duration of the disease among the sampled patients will be between 7 and 9 years?
    (c) Discuss any assumptions that you have made to do this problem.

24. The performance between two sections of physics on an examination were compared. The following data was collected:

Group	n	$\bar{x}$	s
Section 1	54	58	15
Section 2	30	53	13

Test the hypothesis that these two means are the same.

    (a) List the assumptions that you must make to perform the test.
    (b) Compute the relevant test statistic and its p-value for the two-sided test of the hypothesis that the difference of means is zero:
    (c) Do you accept or reject the hypothesis at the 95% significance level?

25. A random sample of $n = 12$ observations from a normal population produced the following estimates: $\bar{x} = 47.1$ and $s^2 = 4.7$.

    (a) Test the hypothesis $H_0 : \mu = 48$ versus $H_a : \mu \neq 48$ with $\alpha = 0.1$.
    (b) What is the p-value for your test?
    (c) Find a 90% confidence interval for the mean, and interpret this interval.

26. The variability in the amount of impurities present in a batch of chemicals used for a particular process depends on the length of time that the process is in operation. Suppose a sample of size 25 is drawn from the usual process which is to be compared to a sample of a new process that has been developed to reduce the variability of impurities. The results of the data for the variation in impurities is shown below:

	Sample 1	Sample 2
n	25	25
s^2	1.04	0.51

Test the null hypothesis: $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_a : \sigma_1^2 > \sigma_2^2$.

27. A doctor is called to see a sick child. The doctor knows (prior to the visit) that 80% of the sick children in that neighborhood are sick with the flu (F) while 20% are sick with the measles (M). For simplicity, assume that M and F are complementary events. A well-known symptom of measles is a rash, denoted by R. The probability of having a rash for a child sick with the measles is 0.9. However, occasionally children with the flu also develop a rash, with probability 0.05. Upon examining the child, the doctor finds a rash. What is the probability that the child has the flu?

28. Consider the following gambling game, called the Petersburg Game: Toss a fair coin repeatedly until the first tail is observed. If the tail occurs on the first toss, you do not win anything; if the first tail occurs on the second toss, you win 2 dollars; if the first tail occurs on the third toss, you win 4 dollars. Thus, if the first tail occurs on toss $k+1$, you win 2^k dollars, $k \geq 1$. Compute the average duration of the game. Now compute the expected winnings of the player. Are you surprised by your results?

29. A random variable X has a *double exponential* distribution with parameter $\alpha > 0$ if its PDF is

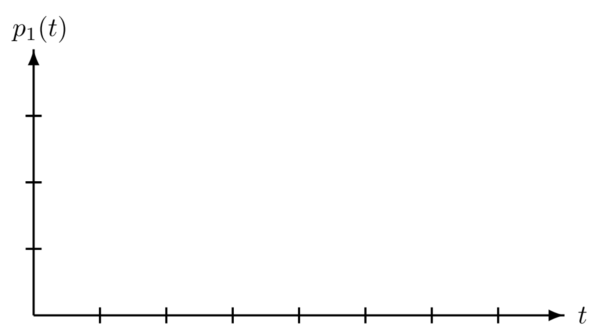
$$f(x) = \frac{1}{2}\alpha e^{-\alpha|x|}, \quad -\infty < x < \infty.$$

   (a) Determine the CDF (denoted $F(x)$) for X, and sketch its graph.

   (b) Determine the expectation, variance and standard deviation for X. You may use the fact that the standard deviation of a random variable that has an exponential distribution with parameter β is $\frac{1}{\beta}$.

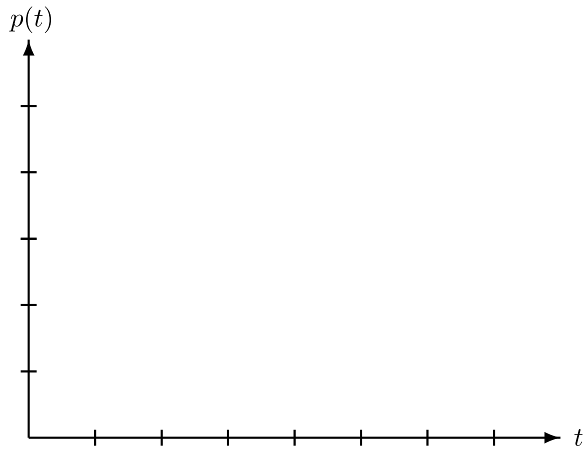
30. The service life T_1 of a certain electrical device follows an exponential distribution with mean $\lambda = 10$ units of time.

   (a) Record PDF and the CDF for T_1, and sketch a qualitatively accurate graph of the PDF. Lastly, determine the probability P that the device functions for at least 10 units of time.



   (b) The electrical devices described in the previous part of this problem are shipped in lots of 100. Let $\bar{T}$ represent the average service life of the devices in a lot[1]. Describe, as accurately as you can and with a discussion of your assumptions, the probability density $p(t)$ for $\bar{T}$. Make a reasonably accurate qualitative sketch the of density; be sure to give some indication as to the scale of the axes of your graph.

[1]Thus, if $T_1, \ldots, T_{100}$ are the service lives of the 100 parts in a randomly chosen lot, then $\bar{T} = \frac{1}{n}(T_1 + T_2 + \cdots + T_{100})$ This average will vary randomly between lots—and that is the quantity of interest



31. A political poll during the 2000 presidential election asked a random sample of 500 male and 500 female voters which candidate they preferred. The data are summarized in the following table:

Group	n	Prefer Bush	Prefer Gore
Male	500	337	163
Female	500	215	285

Test the hypothesis that the proportions of males and females preferring candidate Bush are the same.

(a) List the assumptions that you made to perform the test and assess their plausibility.

(b) Compute the relevant test statistic and its p-value for the two-sided test of the hypothesis that the difference of proportions is zero, against the alternative that the proportions are different:

(c) Find a 95% confidence interval for the difference in proportions, and interpret this interval.

(d) Do you reject or fail to reject the hypothesis at the 95% significance level?

32. A manufacturer of concrete claims that his product has a reasonably stable compressive strength, measured in units of kilograms per square centimeter. He reports that the range of strength observed over many batches of concrete is 40 kg/cm^2. Suppose that you are considering purchasing concrete from this manufacturer, but you must be convinced that the variability in strength is no greater than what is claimed. Your engineers take a random sample of $n = 10$ batches of the material and test their strength, giving sample estimates of $\bar{x} = 312$ and $s^2 = 195$. If you assume that the range of a normal distribution is about 4σ, test the hypothesis that the variance in strength of the sampled concrete is consistent with the manufacturer's claims. Use $\alpha = 0.05$ and a one-sided test.

(a) What is the appropriate test statistic to use, with how many degrees of freedom?

Test statistic = $\boxed{}$, df = $\boxed{}$

(b) Your null hypothesis is H_0 : $\boxed{}$ versus H_a : $\boxed{}$

(c) What is the value of your test statistic = $\boxed{}$

(d) Critical value of test statistic ($\alpha = 0.05$) = $\boxed{}$

(e) Find a 95% confidence interval for the variance, and interpret this interval.

$$\boxed{} < \sigma^2 < \boxed{}$$

(f) Do you reject the hypothesis, or fail to reject it?