ISQS 5347 Final, Fall 2011. Closed book, closed notes, no electronic devices. Points (out of 200) are given. All questions are either multiple choice, short answer, or involve drawing a graph.

Problems 1-16 are ten points each.

1. Explain, using an example from the book or class (but *not* one involving dice, coin tosses, or casino gambling), why probabilistic models are more realistic than deterministic models. Use the definition of a "good" model in your answer.

Solution: Consider the case where Y = housing expense and X = Income. A deterministic model would say that if you know X, you know Y. In other words, everyone who has the same income spends exactly the same amount on their housing. Clearly, the data produced by this model do not look like real housing expense data, because people spend different amounts on their housing, depending on their preferences. A better model is $Y|X=x \sim p(y|x)$, which states that the data on housing expense for fixed income x are produced by a probability distribution. Since probability models produce variable data, this model is much more realistic, in that the data produced by the model (variable) look like the actual housing expenses that we observe (also variable), and is therefore a good model. Recall: a model is good if the data produced by the model look like the data that you actually observe.

2. Suppose Y is produced by the distribution $p(y) = \frac{1}{\sqrt{2\pi}10}\exp\frac{-(y-70)^2}{2\times 10^2}$, for $-\infty<y<\infty$. Explain briefly how you would use random number generation in Microsoft Excel to estimate $\int_{70}^{90} p(y)dy$.

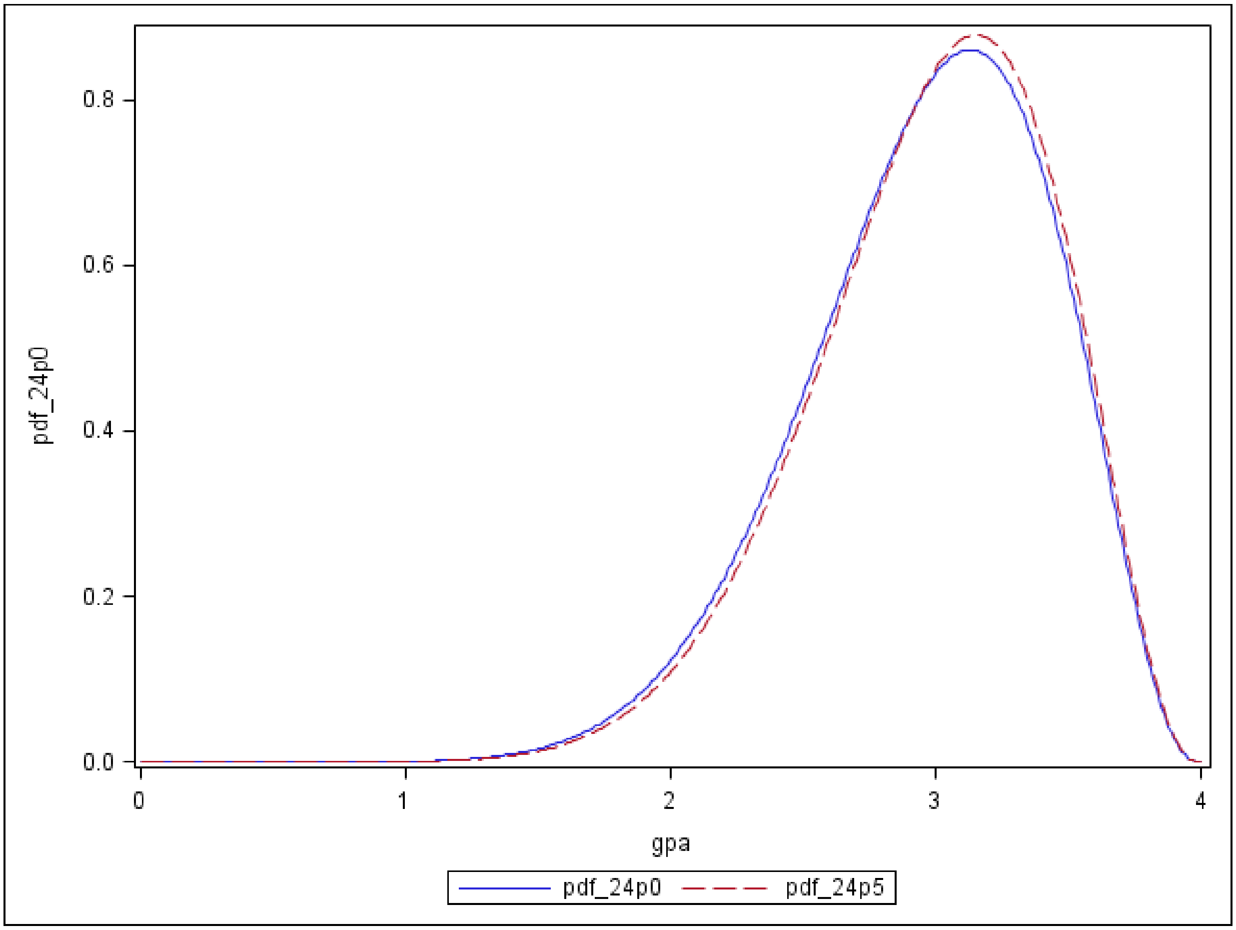
Solution: This integral is the probability that a normally distributed random variable with mean 70 and standard deviation 10 will lie between 70 and 90. So I would generate many, say 10,000, random numbers from the normal random number generator with mean 70 and standard deviation 10. Then I would count how many of them are between 70 and 90, and divide by 10,000 to estimate the probability.

3. Hans claims that the temperature outside (Y) is independent of % cloud cover (X). If Hans is correct, then what can you say about temperature and cloud cover? Use the correct definition of independence in your answer.

Solution: Hans claims that distribution of temperature is the same, regardless of cloud cover. For example, he claims that the distribution of possible temperatures when cloud cover =100% is the same as the distribution of possible temperatures when cloud cover = 0%. The means of those distributions are identical, as are their variances, skewnesses, kurtoses, etc. Hans is silly.

4. Draw graphs of the conditional distributions of graduate student's grade point average (Y=GPA) when student's ages are X=24.0 years and when student's ages are X=24.5 years. Put both distributions on the same axes (i.e., on the same graph), label both axes, and put numbers on both axes. The concept of "morphing" should be apparent.

Solution. There really shouldn't be much difference. Maybe older students tend to be slightly better.



Code: (not requested, but here for completeness)

```
data gpa;
   do gpa = 0 to 4 by .01;
      pdf_24p0 = pdf('beta', gpa, 10,3.5,0,4);
       pdf_24p5 = pdf('beta', gpa, 10.3,3.5,0,4);

          output;
        end;
run;

proc sgplot;
   series y = pdf_24p0 x = gpa;
   series y = pdf_24p5 x = gpa;
run;
```

5. Using an example discussed in the book or in class, explain why the "population" interpretation of statistical models is misleading when you consider conditional distributions.

Solution: The inventory example is good. Maybe in a population of inventory, there is exactly one furniture item costing \$310.12, and no electronics items costing \$310.12. So, using the population interpretation, the distribution of (Furniture/Electronics) given price = \$310.12 is

y	$p(y)$
furn	1.0
elec	0.0

This model is misleading in that it suggests that it is impossible for an electronics item to be valued at \$310.12. Of course this is silly. There is no physical constraint on pricing of electronic items that precludes \$310.12 as a possible value.

6. When might the Law of Large Numbers fail? What happens when the LLN fails?

Solution: The LLN can fail when the mean of the distribution that produces the data is infinite. When this happens, the sample average $\bar{Y}_n$ will increase towards infinity as n increases, rather than converging to a particular value as happens when the mean in finite.

7. Suppose $Y_1,Y_2,\ldots,Y_n \sim_{\text{i.i.d.}} p(y)$, where $\int y\, p(y)\, dy = \mu$. Show, step by step and with all justifications, that $E\{(1/n)(Y_1+Y_2+\ldots+Y_n)\} = \mu$.

Solution:

$E\{(1/n)(Y_1+Y_2+\ldots+Y_n)\} = (1/n)(E\{Y_1+Y_2+\ldots+Y_n)\}$	(by linearity property of expectation)
$= (1/n)\{E(Y_1)+E(Y_2)+\ldots+E(Y_n)\}$	(by additivity property of expectation)
$= (1/n)\{\mu + \mu +\ldots+\mu\}$	(since all Y's are produced by the same pdf $p(y)$, and since the mean of that pdf is given to be μ)
$= (1/n)(n\mu)$	(since there are n terms in the sum)
$= \mu$	(by algebra).

8. Explain the equation $\text{Var}(Y+5) = \text{Var}(Y)$, as if to someone who doesn't know much about statistics.

Solution: You can think of variance as a kind of average squared deviation from observable data to the mean. Thus, variance measures spread of the data. Now, if you add 5 to all the data values you shift the values higher, but you don't change the spread. For example, if the data values go from 6 to 20 at first, then the range is 20 – 6 = 14. If you add 5 to everything, the data values go from 11 to 25, with a range of 25 – 11 =14, again. The range is the same. So the spread of the data is not affected by adding a constant, like 5, to all data values.

9. The mean of a probability distribution $p(y)$ is 10 and the standard deviation is 3. What does Chebychev's theorem tell you about the data that are produced by $p(y)$? Use ±2 standard deviations in your explanation.

Solution: At least 75% of the data that are produced by $p(y)$ will be between 4 and 16.

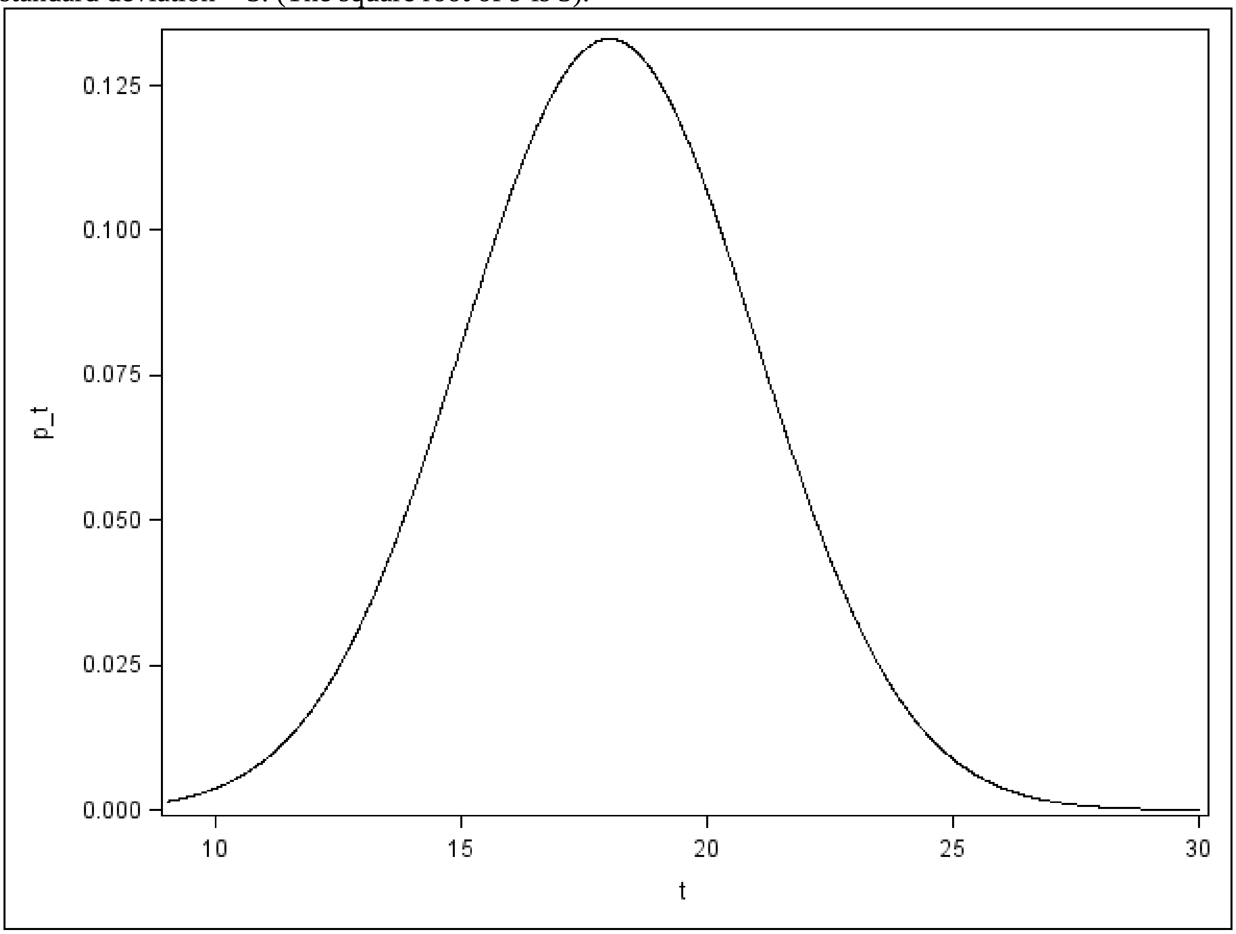
10. A data set is 43.2, 45.1, 54.5, 60.0, 60.0. Write down the bootstrap distribution $\hat{p}(y)$ in list form (this is the distribution that yields the "plug-in" estimates).

Solution:

y	$\hat{p}(y)$
43.2	1/5
45.1	1/5
54.5	1/5
60.0	2/5
Total	1.0

11. Suppose $Y_1,\ldots,Y_9 \sim_{\text{i.i.d.}} p(y)$, where $\mathrm{E}(Y_1) = 2$ and $\mathrm{Var}(Y_1) = 1$. Let $T = Y_1+\ldots+Y_9$. Draw a graph of $p(t)$, the distribution of T. Label horizontal and vertical axes and put numbers on them. Why did you draw a bell curve?

Solution: The distribution is approximately (by the Central Limit Theorem) normal, with mean = 18 and standard deviation = 3. (The square root of 9 is 3).



Code (for completeness only)

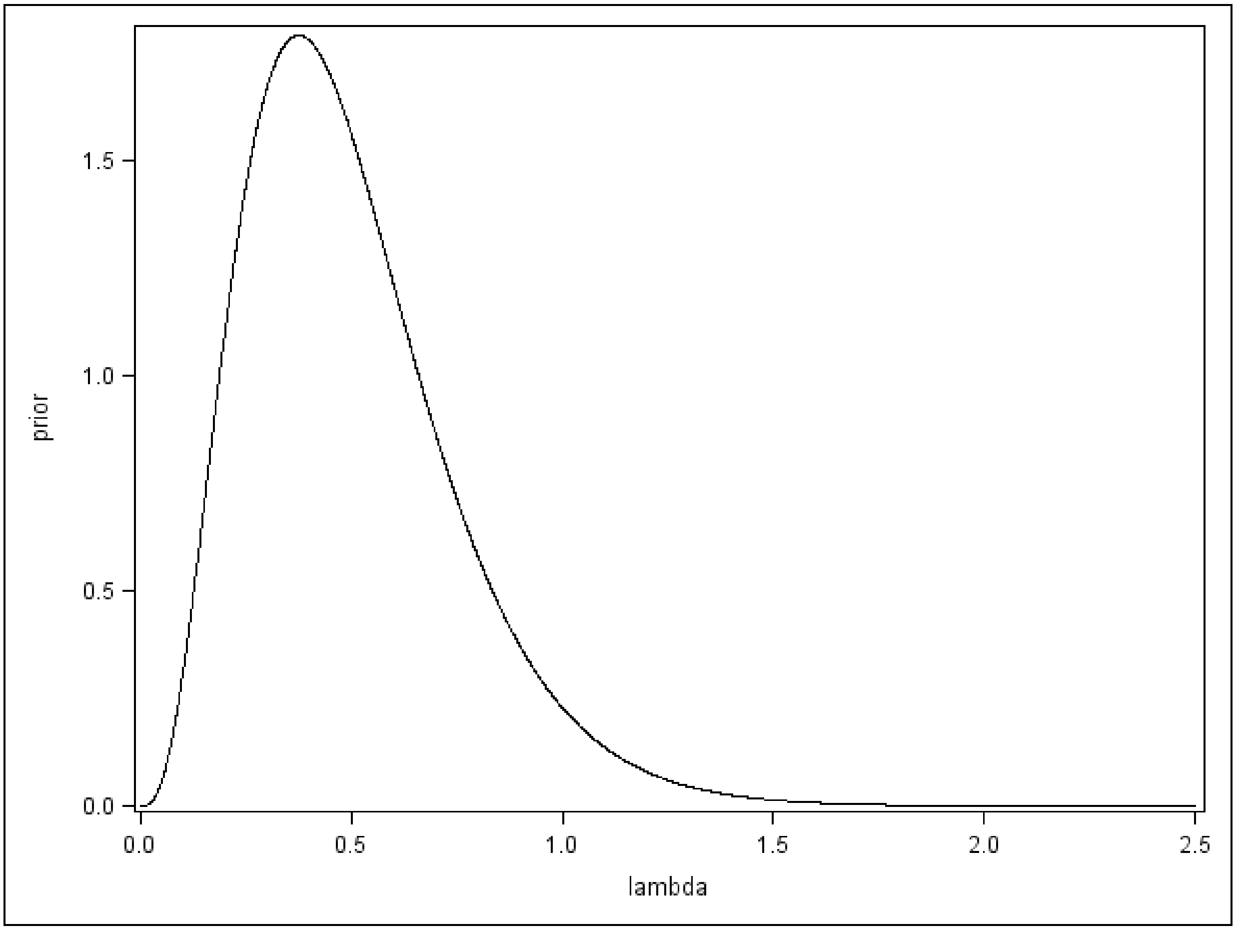
```
data gpa;
   do t = 9 to 30 by .01;
      p_t = pdf('normal', t, 18,3);

           output;
         end;
run;

proc sgplot;
   series y = p_t x = t;
run;
```

12. Observations 1,0,0,1,1 are obtained as an i.i.d. sample from the Bernoulli(π) distribution. Write down the likelihood function for π.

Solution: $L(\pi) = \pi^3(1 - \pi)^2$.

13. Draw the graph of a plausible prior distribution for λ of the Poisson distribution used in the luxury car sales example. Recall that λ is the mean number of sales per day. Label both axes and put numbers on them. Don't use a non-informative prior.

Solution: Somewhere around 0.5.



Code (not requested):

```
data gpa;
   do l = 0 to 40 by .01;
      prior = pdf('chisquare', l, 8)*16;
      lambda = l/16;
         output;
        end;
run;

proc sgplot;
   series y = prior x = lambda;
run;
```

14. In the student age example, no student reported Age = 29 years (rounded to age at last birthday). Yet we know that Pr(Age = 29) (called π_{29} in class and in the book) is not zero. What is π_{29}? Suppose that Pr(Age = 29) = 0.07. Assume that 0.07 means "7 out of 100," explain what the "7" refers to, and explain what the "100" refers to. Be specific as possible about the meaning of the "100."

Solution: There is a process that puts students into this class. For 100 students produced by this process, 7 will be 29 years old.

15. You are thinking of trying a new route to class in the morning. You decide to record times it takes to get from your home to class going your current route for a few days, and your times going the new route for a few days. There will be a difference in the average times. Is the difference explainable by chance alone? Describe a "chance-only" model that can explain the difference.

Solution: A chance-only model is that your times to class all are i.i.d. $p(y)$, where the distribution $p(y)$ is the same, regardless of which route you take. Since probabilistic models produce random data, the averages of DATA* produced by this model for the two different routes will be different, and such differences are explained by chance alone.

16. Why is the "Student T distribution" needed? What is its main reason for being?

Solution: Using an estimated standard deviation in the normalized average gives you the T distribution. If $Y_1, Y_2, \ldots, Y_n \sim_{\text{i.i.d.}} \mathrm{N}(\mu, \sigma^2)$, then $(\bar{Y} - \mu)/(\sigma/\sqrt{n}) \sim \mathrm{N}(0,1)$, but $(\bar{Y} - \mu)/(\hat{\sigma}/\sqrt{n}) \sim \mathrm{T}_{n-1}$.

Multiple choice (4 points each)

17. You flip a coin and record Y, either Heads or Tails. Pick the best distribution for describing Y.
A. Poisson **B. Bernoulli** C. Normal D. Continuous

18. A random variable Y is produced by a distribution for which $\Pr(Y = 1) = 0.5$, $\Pr(Y=2) = 0.25$, and $\Pr(Y = 3) = 0.25$. The average of an i.i.d. sample of 10,000 such Y's will be closest to
A. 1.00 B. 1.25 C. 1.50 **D. 1.75** E. 2.00 F. 2.25 G. 2.50 H. 2.75 I. 3.00

19. Suppose $Y_1 \sim \mathrm{N}(10, 3^2)$ and $Y_2 \sim \mathrm{N}(15, 4^2)$ are independent. Then the variance of $(1/2)(Y_1+Y_2)$ is
A. 4.00 B. 5.00 **C. 6.25** D. 8.00 E. 12.50 F. 25.00

20. You are contemplating using two different estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$. Pick the best answer:

A. If $\hat{\theta}_1$ is unbiased and $\hat{\theta}_2$ is biased, then $\hat{\theta}_1$ tends to be closer to θ than $\hat{\theta}_2$.

B. If $\hat{\theta}_1$ is efficient and $\hat{\theta}_2$ is inefficient, then $\hat{\theta}_1$ tends to be closer to θ than $\hat{\theta}_2$.

C. If $\hat{\theta}_1$ is normally distributed and $\hat{\theta}_2$ is not normally distributed, then $\hat{\theta}_1$ tends to be closer to θ than $\hat{\theta}_2$.

D. If $\hat{\theta}_1$ has higher expected squared error than $\hat{\theta}_2$, then $\hat{\theta}_1$ tends to be closer to θ than $\hat{\theta}_2$.

21. What is the point of using "n – 1" in the variance estimator?

A. To obtain an unbiased estimator
B. To obtain an efficient estimator
C. To obtain a consistent estimator
D. To obtain the bootstrap estimator
E. To obtain the maximum likelihood estimator
F. To obtain the Bayes estimator

22. The posterior distribution of θ is given as
A. $p(\theta \mid y) = L(\theta \mid y)$ B. $p(\theta \mid y) \propto L(\theta \mid y)$ C. $p(\theta \mid y) = L(\theta \mid y)p(\theta)$ **D.** $\boldsymbol{p(\theta \mid y) \propto L(\theta \mid y)p(\theta)}$

23. A 95% credible interval for θ is obtained using
A. **The posterior distribution of $\boldsymbol{\theta}$** B. The Central Limit Theorem C. Repeated samples from the process D. The likelihood function for θ

24. What is a *p*-value?
A. The probability that a chance-only model explains your observed difference.
B. The probability of seeing a difference as large as your observed difference, assume the data were produced by a chance-only model.
C. The probability that the null model is true, given your observed difference.
D. The probability that the null model is false, given your observed difference.

25. In the 100(1 – α)% frequentist confidence interval for μ,
A. The interval is random and $\boldsymbol{\mu}$ is fixed.
B. The interval is fixed and μ is random.
C. $\bar{y}$ is random and μ is random.
D. α is random and the interval is random.

26. The values $Y_1, Y_2, \ldots, Y_{10}$ are i.i.d. N(0,1). What is the distribution of $T = Y_1^2 + Y_2^2 + \ldots + Y_{10}^2$?
A. N(10,20) B. N(0,20) C. χ_9^2 (**D.** χ_{10}^2) E. T_9 F. T_{10}