ISQS 5347 Final exam Spring 2012.

Closed book, notes and no electronic devices. Points (out of 200) are in parentheses.

1. (10) "Model produces data, model has unknown parameters" is central to the course. Recall the commute time example in class about how long it takes you all to get to class (in minutes). Give a good model for how these data are produced, and explain why it is a good model.

   Solution: If the times are $Y_1, Y_2, \ldots$, then the model $Y_1, Y_2, \ldots \sim_{\text{iid}} p(y)$ is a good model. The model is good if the data produced by the model look like the actual data; here, the actual data certainly look as if produced at random from some distribution. Independence is possibly questionable if, for example, roommates or married couples ride together, but other than that, the data look as if produced as iid from some pdf $p(y)$. Since no particular form is assumed for $p(y)$, it can be anything, and therefore the data will not contradict the model, as they would if the assumed form was normal, exponential, etc.

2. (10) Refer to the commute time example of problem 1. Use the definition of independence to explain why commute time (Y) and distance in miles from the university (X) are not independent. Draw graphs of pdfs, labeling axes, that support your answer.

   Solution: Commute time is independent of distance if the distribution of commute time is the same, no matter what the distance is. Of course this is silly. Say distance = 10 miles. Then the mean commute time will probably be around 20 minutes. Say distance = 1 mile. Then the mean commute time will probably be around 5 minutes. So draw two pdfs, one with mean around 20 minutes, the other with mean around 5 minutes. (It will be more realistic if these two pdfs are both positively skewed.) Since these distributions are different, the variables are not independent.

3. (10) The distribution of a random variable Y is continuous and right-skewed (i.e., the skewness is greater than zero). Draw a graph that shows why there is a 90% chance that Y will be between the 0.05 quantile and the 0.95 quantile of the distribution, and explain the graph.

   Solution. Draw a graph with a tail on the right htat is longer than the one on the left. Then mark the 0.05 quantile on the horizontal axis, to the left of the mean, and the 0.95 quantile to the right of the mean. Shade the area under the curve between these two points. There is 0.95 area to the left of the 0.95 quantile, and this includes the 0.05 area to the left of the 0.05 quantile. So the area between the two quantiles is 0.95 – 0.05 = 0.90.

4. (10) The estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ. Show that $2\hat{\theta}_1 - \hat{\theta}_2$ is also an unbiased estimator of θ. Justify every step carefully; in other words, provide the reasons.

Solution: E($2\hat{\theta}_1 - \hat{\theta}_2$) = E($2\hat{\theta}_1 + (-1)\hat{\theta}_2$) (By algebra)
= E($2\hat{\theta}_1$) + E($(-1)\hat{\theta}_2$) (By the additivity property of expectation)
= 2E($\hat{\theta}_1$) + (–1)E($\hat{\theta}_2$) (By the linearity property of expectation)
= $2\theta + (-1)\theta$ (Since $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ)
= θ (By algebra).

Since E($2\hat{\theta}_1 - \hat{\theta}_2$) = θ, the estimator $2\hat{\theta}_1 - \hat{\theta}_2$ is unbiased, by definition of unbiasedness.

5. (10) The sample proportion $\hat{\pi}$ has variance $\pi(1-\pi)/n$. Explain why, justifying every step carefully; in other words, provide the reasons. (Hint: $\hat{\pi}$ is an average of iid Bernoulli random variables.)

Solution: Write $\hat{\pi} = (1/n)\sum Y_i$, where the Y_i are iid Bernoulli(π). Then
Var($\hat{\pi}$) = Var($(1/n)\sum Y_i$) (by substitution)
= $(1/n)^2$ Var($\sum Y_i$) (by the linearity property of variance)
= $(1/n)^2$ (Var(Y_1) + Var(Y_2) + … + Var(Y_n)) (by the additivity property of variance, assuming independence).

Now, each Y_i is sampled from the Bernoulli distribution:

y	$p(y)$
0	$1-\pi$
1	π

So, by definition of expectation of a discrete random variable, E(Y_i) = $\Sigma\, y\, p(y)$ = $0(1-\pi) + 1(\pi) = \pi$. Further, by definition of variance of a discrete random variable, Var(Y_i) = $\Sigma\,(y-\mu)^2 p(y) = (0-\pi)^2(1-\pi) + (1-\pi)^2(\pi) = \pi(1-\pi)$.

Hence, by substitution and algebra, Var($\hat{\pi}$) = $(1/n)^2$ (Var(Y_1) + Var(Y_2) + … + Var(Y_n))
= $(1/n)^2\,(\pi(1-\pi) + \pi(1-\pi) + \ldots + \pi(1-\pi))$
= $(1/n)^2\, n\pi(1-\pi)$
= $\pi(1-\pi)/n$.

6. (5) How are the Law of Large Numbers and Central Limit Theorem *similar*?

   Solution: Both refer to the behavior of averages as the sample size, n, gets larger.

7. (5) How are the Law of Large Numbers and Central Limit Theorem *different*?

   Solution: The LLN states that the average gets closer to the mean μ as n gets larger. It says nothing about the how the distribution of the sample average looks, for any n. On the other hand, the CLT says that the *distribution* of the average becomes better approximated by a normal distribution as n gets larger.

8. (5) Give an example of an estimator that is unbiased but inconsistent.

   Solution: Take an iid sample $Y_1, Y_2, \ldots, Y_n$ from a distribution $p(y)$ whose mean is μ. Let the estimator of μ be the last observation, Y_n. Since the data are all produced by the same process, $E(Y_n) = \mu$, and hence the estimator is unbiased. But Y_n does not get closer to μ as n grows: its variance is always σ^2.

9. (5) Give an example of an estimator that is biased but consistent.

   Solution: As discussed in the book, the ordinary standard deviation estimator is biased but consistent.

10. (10) How many people are on the Board of Directors for Microsoft Corporation? Draw your *prior distribution*, labeling axes carefully, and explain why you drew your prior the way you did.

    Solution: Not knowing much about this except that the number of people is discrete, I would put a little probability on 0, on 1, on 2, etc. Maybe Microsoft has no board? In that case 0 is a prior possibility. I'd spread it out a lot, and make the probabilities pretty constant, since I really don't know whether a board has 3 people of 20 people. I am guessing that boards probably don't have 1,000 people, so my prior probabilities are less and less for values in the 1,000's and higher. But I don't want to be dogmatic and completely rule out such values, because I just don't know, so I'll leave a miniscule bit of probability for the extremes.

11. (10) Explain what the symbol "$\propto$" means in the formula $p(\theta\,|\,\text{data}) \propto L(\theta\,|\,\text{data})\, p(\theta)$. Also explain what specific changes are needed to make the "$\propto$" symbol an "=" symbol.

Solution: The term "$\propto$" means "proportional to", and specifically means that $p(\theta\,|\,\text{data}) = c\, L(\theta\,|\,\text{data})\, p(\theta)$ for some constant of proportionality c. Since this is a pdf, the area under the curve has to be 1.0, so $c = 1/\int L(\theta\,|\,\text{data})\, p(\theta)\, d\theta$.

12. (10) Life expectancy (Y, in years) for a restaurant that has been in business for one year has probability density function $p(y) = 1/y^2$ for $y \geq 1$, and $p(y) = 0$ for $y < 1$. Find the median life expectancy for such a restaurant. (Hint: $\int y^n dy = y^{n+1}/(n+1)$).

Solution: The median is the 0.50 quantile, and the quantile is the inverse cdf. So let's find the cdf first: $P(y) = \int_1^y 1/t^2 dt = \int_1^y t^{-2} dt = t^{-2+1}/(-2+1)\Big|_1^y = -t^{-1}\Big|_1^y = -y^{-1} - (-1) = 1 - 1/y$. The median is the y for which $P(y) = 0.5$, or for which $1 - 1/y = 0.5$. Solving, you get $y = 2.0$ is the median.

13. (10) I made a big deal about the *population* definition of the probability model $p(y)$. What *is* the population definition of the probability model $p(y)$? Give a real example. Be sure that N is part of your answer.

Solution: Suppose the population is the population of inventory valuations. These numbers are dollar amounts, say $y_1, y_2, \ldots, y_N$, where N is the number of inventory items owned by the company. The pdf is then

y	$p(y)$
y_1	$1/N$
y_2	$1/N$
…	…
y_N	$1/N$

with collation in case of repeat inventory dollar values.

14. (10) You can check the assumption of *independence* in your observable data $Y_1, Y_2, \ldots, Y_n$ by constructing and comparing *two* histograms. If independence is a reasonable model, then these *two* histograms will be similar in shape, location, and spread. *Which* two histograms are these?

Solution: Match each current data value with the previous value: $(Y_2, Y_1), (Y_3, Y_2), \ldots, (Y_n, Y_{n-1})$. Now, create two groups of current values, one where the previous value is Low (say, less than the median), and the other where the previous value is High (say, more than the median). One histogram is that of the current values in the first group, the other is that of the current values in the second group.

The next problems are all 5 points each. Select only one answer for each problem.

15. Suppose $Y_1, Y_2 \sim_{\text{iid}} \mathrm{N}(0, 1)$. What is the distribution of $Y_1 + Y_2$?
A. T_2 B. $F_{2,2}$ C. $F_{1,2}$ D. N(0, 1) **E. N(0, 2)** F. χ_2^2 G. $\chi_2^2/2$ H. $\sqrt{\chi_2^2/2}$

16. Suppose $Y_1, Y_2 \sim_{\text{iid}} \mathrm{N}(0, 1)$. What is the distribution of $Y_1^2 + Y_2^2$?
A. T_2 B. $F_{2,2}$ C. $F_{1,2}$ D. N(0, 2) E. N(2, 4) **F.** χ_2^2 G. $\chi_2^2/2$ H. $\sqrt{\chi_2^2/2}$

17. Suppose $Y \sim \mathrm{N}(\mu, \sigma^2)$. What is the distribution of $(Y - \mu)/\sigma$?
A. T_1 B. $F_{1,2}$ C. $F_{1,1}$ **D. N(0, 1)** E. $\mathrm{N}(2\mu, 4\sigma^2)$ F. χ_2^2 G. $\chi_2^2/2$ H. $\sqrt{\chi_1^2/1}$

18. Nature favors
A. continuity over discontinuity B. discreteness over continuity
C. normal distributions over non-normal distributions D. linearity over curvature

19. If the mean is 10 and the standard deviation is 2, then
A. 75% of the data are between 6 and 14.
B. 95% of the data are between 6 and 14.
C. At least 75% of the data are between 6 and 14.
D. At least 95% of the data are between 6 and 14.

20. Why is it better to simulate 100,000 values from the posterior distribution than 1,000 values?
    A. Because a larger sample size is needed for the Central Limit Theorem to take hold
    B. Because you will get a more accurate estimate the posterior distribution
    C. Because Bayesian analysis is only valid when sample size > 1,000
    D. Because your credible interval will have a higher confidence level

21. What is the likelihood function for an iid sample?
    A. The average of the likelihoods for the individual observations
    B. The sum of the likelihoods for the individual observations
    C. The product of the likelihoods for the individual observations
    D. The logarithm of the likelihoods for the individual observations

22. If $\bar{Y}$ is calculated from an iid sample, then E{ $(\bar{Y})^2$ } is
    A. $= \mu^2$ **B. $> \mu^2$**   C. $< \mu^2$

23. The likelihood function is $\pi^2(1-\pi)^8$ and the prior is uniform. Then the kernel of the posterior distribution is
    A. $\boldsymbol{\pi^2(1-\pi)^8}$ B. π^2 C. $(1-\pi)^8$ D. 1.0   E. $e^{-\pi}$ F. Beta(2, 8)

24. Where does prior information come from?
    A. **Other similar, historical studies**   B. From your current data set
    C. From the normal distribution   D. All of the above

25. What is in the data set work.try?

```
data try;
   do y = 1 to 3 by 1;
   output;
   end;
run;
```

A. $y_1\ y_2\ y_3$ **B. 1 2 3**   C. 2.0   D. nothing at all

26. What does the following code output?

```
proc means lclm uclm alpha=.05;
   var y;
run;
```

A. 95% confidence limits for μ

B. 90% confidence limits for μ

C. 10% confidence limits for μ

D. 5% confidence limits for μ

27. The null distribution of the likelihood ratio test statistic is

A. chi-squared B. normal **C. approximately chi-squared** D. approximately normal

28. The unrestricted (full) model is

$$Y \mid (X_1, X_2, X_3) = (x_1, x_2, x_3) \sim \mathrm{N}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \sigma^2).$$

The restricted (null) model is

$$Y \mid (X_1, X_2, X_3) = (x_1, x_2, x_3) \sim \mathrm{N}(\beta_0, \sigma^2).$$

The degrees of freedom for the likelihood ratio test statistic are *df*. What is *df*?

A. 1 B. 2 **C. 3** D. 4 E. 5 F. 6 G. 7 H. 8 I. 9 J. 10

29. The ***inverse cdf method*** is used to

A. Calculate mean values of continuous distributions

B. Calculate mean values of discrete distributions

C. Simulate data from continuous distributions

D. Simulate data from discrete distributions

30. What happens when you select a larger sample size *n* for your study? (Choose one).

A. Your data *Y* become closer to normally distributed

B. Your noncentrality parameter increases

C. Your 95% confidence interval becomes closer to a 100% confidence interval

D. Var(*Y*) becomes smaller