Stat 5102 Final Exam **May 12, 2012**

Name ______________________________ Student ID __________________

The exam is closed book and closed notes. You may use three $8\frac{1}{2} \times 11$ sheets of paper with formulas, etc. You may also use the handouts on "brand name distributions" and Greek letters. Put all of your work on this test form (use the back if necessary). Show your work or give an explanation of your answer. No credit for numbers with no indication of where they came from.

The points for the questions total to 200. There are 10 pages and 8 problems.

1. [25 pts.] This problem is about the Beta$(\theta, 2)$ distribution. (Hint: Use the gamma function recursion formula to eliminate gamma functions from the PDF.)

   (a) Find the Jeffreys prior.

   (b) Say whether it is proper or improper.

2. [25 pts.] The following Rweb output fits a linear model and does several tests of model comparison about it.

```
Rweb:>  out <- aov(y ~ trt * blk)
Rweb:>  summary(out)
            Df Sum Sq Mean Sq F value   Pr(>F)
trt          3 120.78   40.26  39.052 6.29e-13 ***
blk          5  41.48    8.30   8.046 1.42e-05 ***
trt:blk     15  43.99    2.93   2.844  0.00306 **
Residuals   48  49.49    1.03
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) Explain what models are being compared (what is the null, what is the alternative) in each of the three tests that goes with each of the three P-values in the rightmost column of the table.

(b) Explain where the degrees of freedom come from for each of the degrees of freedom listed in the leftmost column of the table. How many levels must the factor `trt` have had, and how many levels must the factor `blk` have had? How many data values were there in all, that is, what was the length of the response vector `y`?

(c) If one has to choose among the models, which does one choose on grounds of simplicity and statistical significance? Explain.

3. [25 pts.] The is Rweb output fitting a generalized linear model

```
Rweb:>  out <- glm(y ~ x + I(x^2) + I(x^3), family = poisson)
Rweb:>  summary(out)

Call:
glm(formula = y ~ x + I(x^2) + I(x^3), family = poisson)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.09193  -0.58356   0.02008   0.61513   1.67098

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.56932    0.05892  26.634  < 2e-16 ***
x            0.21240    0.07191   2.954  0.00314 **
I(x^2)       0.18256    0.04720   3.868  0.00011 ***
I(x^3)      -0.03526    0.02511  -1.404  0.16017
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 142.380  on 99  degrees of freedom
Residual deviance:  76.776  on 96  degrees of freedom
AIC: 438.74

Number of Fisher Scoring iterations: 4
```

(a) Find a 95% confidence interval for the true unknown regression coefficient for the predictor `I(x^2)`. (Hint: The 0.95 quantile of the standard normal distribution is 1.645, and the 0.975 quantile of the standard normal distribution is 1.96.)

(b) Perform a hypothesis test of whether this same regression coefficient is zero (the null hypothesis) versus nonzero (the alternative hypothesis), reporting and interpreting the P-value.

(c) Perform a hypothesis test of whether this same regression coefficient is zero (the null hypothesis) versus greater than zero (the alternative hypothesis), reporting and interpreting the P-value.

(d) Repeat part (b) except for the true unknown regression coefficient coefficient for the predictor `I(x^3)`.

4. [25 pts.] Suppose (X_i, Y_i), $i = 1, \ldots, n$, are IID from the distribution with PDF

$$f_\theta(x, y) = \exp\left(-x\theta - \frac{y}{\theta}\right), \qquad 0 < x < \infty, 0 < y < \infty.$$

Show that $(\overline{X}_n, \overline{Y}_n)$ is a two-dimensional sufficient statistic for this model.

5. [25 pts.] Suppose X_1, ..., X_n are IID from the distribution having PDF

$$f_\theta(x) = \frac{\theta}{(1 + x)^{\theta+1}}, \qquad x > 0,$$

where θ is an unknown parameter satisfying $\theta > 0$. Show that this statistical model is an exponential family. Identify the natural parameter and natural statistic.

6. [25 pts.] Suppose X_1, …, X_n are IID from the distribution with PDF

$$f(x \mid \theta) = \theta x^{\theta - 1}, \qquad 0 < x < 1,$$

where θ is an unknown parameter satisfying $0 < \theta < \infty$. And suppose we use the improper unnormalized prior

$$g(\theta) = \frac{1}{\theta}, \qquad 0 < \theta < \infty.$$

Find the posterior distribution for θ. (Hint: the posterior is a brand name distribution.)

7. [25 pts.] The t distribution with five degrees of freedom has PDF

$$g(z) = \frac{8}{3\pi\sqrt{5}(1 + \frac{z^2}{5})^3}, \qquad -\infty < z < \infty$$

(this agrees with the formula in the brand name distributions handout, we have just simplified the gamma functions so you do not have to). The location-scale family of distributions having g as its standard distribution has PDF's

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} \cdot g\left(\frac{x - \mu}{\sigma}\right) \qquad -\infty < x < \infty,$$

where μ is the location parameter and σ is the scale parameter, and these satisfy $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. (Caution: σ^2 is not the variance of the distribution having PDF $f_{\mu,\sigma}$.)

Suppose X_1, X_2, ... are IID from the distribution with density $f_{\mu,\sigma}$. Find the asymptotic relative efficiency (ARE) of the sample mean and sample median of an IID sample from this distribution, both considered as estimators of μ. Also state which estimator is better.

8. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID $\text{Gam}(\alpha, \lambda)$, where $\alpha > 0$ is a known constant and $\lambda > 0$ is an unknown parameter.

   (a) Find a method of moments estimator of λ.

   (b) Find the asymptotic normal distribution of your estimator $\hat{\lambda}_n$.