Stat 5102 Final Exam **May 12, 2010**

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.] Find the Jeffreys prior for the Poi(μ) distribution. It is proper or improper?

2. [25 pts.] The following Rweb output fits three linear models and does tests of model comparison between them

```
Rweb:>  out1 <- lm(y ~ x1 + x2 + x3)
Rweb:>  out2 <- lm(y ~ poly(x1, x2, x3, degree = 2, raw = TRUE))
Rweb:>  out3 <- lm(y ~ poly(x1, x2, x3, degree = 3, raw = TRUE))
Rweb:>  anova(out1, out2, out3)
Analysis of Variance Table

Model 1: y ~ x1 + x2 + x3
Model 2: y ~ poly(x1, x2, x3, degree = 2, raw = TRUE)
Model 3: y ~ poly(x1, x2, x3, degree = 3, raw = TRUE)
  Res.Df     RSS Df Sum of Sq         F Pr(>F)
1     96 4206818
2     90    9312  6   4197506 7224.2845 <2e-16 ***
3     80    7747 10      1565    1.6164 0.1169
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

(a) Explain what these models are and why they are nested models.

(b) Explain why there are 6 degrees of freedom difference between model 1 and model 2 and why there are 10 degrees of freedom difference between model 2 and model 3.

(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 following Rweb output fits a generalized linear model.

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

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

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.89479  -0.31789  -0.06147   0.37706   1.90034

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.047585   7.036637  -1.144    0.253
x            0.518776   0.758023   0.684    0.494
I(x^2)      -0.004293   0.019942  -0.215    0.830

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 40.381  on 29  degrees of freedom
Residual deviance: 17.618  on 27  degrees of freedom
AIC: 23.618

Number of Fisher Scoring iterations: 7
```

(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.

4. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID Beta(α_1, α_2), where both parameters are unknown. Show that $\prod_{i=1}^n X_i$ and $\prod_{i=1}^n (1 - X_i)$ are components of a two-dimensional sufficient statistic for this model.

5. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID $\mathcal{N}(0, \nu)$, where $\nu > 0$ is an unknown parameter.

   (a) Show that this statistical model is an exponential family. Identify the natural parameter and natural statistic.

   (b) Find the maximum likelihood estimate (MLE) for ν.

   (c) Show that this is the unique global maximizer. (Hint: use part (a)).

(d) Find the asymptotic distribution of your MLE.

6. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID $\mathcal{N}(\mu, \sigma^2)$. where μ is unknown and σ^2 is known. And suppose we use a flat prior for μ, which is improper. Find the posterior distribution for μ. (Hint: the posterior is a brand name distribution.)

7. [25 pts.] The function

$$f_\theta(x) = \frac{5}{8}[1 - (x - \theta)^4], \qquad \theta - 1 < x < \theta + 1$$

is a PDF, where the parameter θ can be any real number. 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 θ. Also state which is the estimator is better.

8. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID Beta$(\theta, 2)$. where $\theta > 0$ is an unknown parameter. The obvious method of moments estimator for θ is

$$\hat{\theta}_n = \frac{2\overline{X}_n}{1 - \overline{X}_n}$$

(you do not have to prove this). Find the asymptotic normal distribution of $\hat{\theta}_n$.