Stat 5102 Final Exam **May 14, 2013**

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 distribution with probability density function (PDF)

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

where $\theta > 0$ is an unknown parameter.

(a) Find the Jeffreys prior for this distribution.

(b) Say whether it is proper or improper.

2. [25 pts.] The following is Rweb output fitting a generalized linear model

```
Rweb:>  out <- glm(y ~ x1 + x2, family = poisson)
Rweb:>  summary(out)

Call:
glm(formula = y ~ x1 + x2, family = poisson)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.99305  -0.86588   0.05479   0.86546   2.75072

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.952559   0.025932   36.73   <2e-16 ***
x1          1.002046   0.004480  223.67   <2e-16 ***
x2          0.006192   0.004423    1.40    0.161
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 74315.40  on 99  degrees of freedom
Residual deviance:   128.89  on 97  degrees of freedom
AIC: 923.72

Number of Fisher Scoring iterations: 3
```

(a) Find a 95% confidence interval for the true unknown regression coefficient for the predictor `x1`. (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 the regression coefficient for the predictor `x2` is zero (the null hypothesis) versus nonzero (the alternative hypothesis). State the value of the test statistic, the distribution of the test statistic under the null hypothesis, whether this distribution is exact or approximate (asymptotic, large n), and the P-value. Interpret the P-value. What does it say about what predictor variables should or should not be in the model?

3. [25 pts.] Suppose Y is a random r by c matrix with components Y_{ij} that form a multinomial random vector (when rearranged into a vector) that have

$$E(Y_{ij}) = np_{ij},$$

where n is the multinomial sample size and the matrix p is the matrix parameter that is a probability vector (components nonnegative and sum to one) (when rearranged into a vector). This is the setup the course slides called "two-dimensional contingency table." The submodel that has parameterization

$$p_{ij} = \alpha_i \beta_j$$

is the one the course slides called the hypothesis of independence of row and column labels (considered as random variables). Show that the row and column marginals

$$Y_{i+} = \sum_{j=1}^{c} Y_{ij}, \qquad i = 1, \ldots, r$$

$$Y_{+j} = \sum_{i=1}^{r} Y_{ij}, \qquad j = 1, \ldots, c$$

collectively constitute a sufficient statistic vector (when collected into a vector) for this model.

4. [25 pts.] Suppose $X_1, \ldots, X_n$ are independent (but not identically distributed) Poisson random variables with

$$E(X_i) = a_i \theta^{b_i},$$

where θ is an unknown parameter satisfying $\theta > 0$ and a_i and b_i are known constants with all of the a_i strictly positive. Show that this statistical model is an exponential family. Identify the natural parameter and natural statistic.

5. [25 pts.] Suppose $X_1, \ldots, X_n$ are IID from the distribution with PDF

$$f(x \mid \theta) = \frac{2 + \theta}{x^{3+\theta}}, \qquad 1 < x < \infty,$$

where $\theta > 0$ is an unknown parameter satisfying $0 < \theta < \infty$. This distribution has mean and variance

$$\begin{aligned} E_\theta(X) &= \frac{2 + \theta}{1 + \theta} \\ \operatorname{var}_\theta(X) &= \frac{2 + \theta}{\theta (1 + \theta)^2} \end{aligned}$$

(you do not have to prove these).

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

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

6. [25 pts.] Suppose X_1, ..., X_n are IID from the normal distribution with mean θ and variance θ^2, where $\theta > 0$ is the unknown parameter. Then both the sample mean $\overline{X}_n$ and the sample median $\widetilde{X}_n$ are consistent and asymptotically normal estimators of the center of symmetry θ. Neither is efficient. Neither is the maximum likelihood estimator $\hat{\theta}_n$.

(a) Find the asymptotic variance of $\overline{X}_n$.

(b) Find the asymptotic variance of $\widetilde{X}_n$.

(c) Find the asymptotic variance of $\hat{\theta}_n$. (You do not have to find an expression for $\hat{\theta}_n$. You only have to find its asymptotic variance.)

(d) Find the asymptotic relative efficiency of $\overline{X}_n$ versus $\widetilde{X}_n$. Say which is better (this may be different for different values of θ or it may not).

(e) Find the asymptotic relative efficiency of $\overline{X}_n$ versus $\hat{\theta}_n$. Say which is better (this may be different for different values of θ or it may not).

(f) Find the asymptotic relative efficiency of $\widetilde{X}_n$ versus $\hat{\theta}_n$. Say which is better (this may be different for different values of θ or it may not).

7. [25 pts.] Suppose X_1, …, X_n are IID from the distribution with PMF

$$f_\theta(x) = (1 - e^\theta)e^{\theta x}, \qquad x = 0, 1, 2, \ldots,$$

where θ is an unknown parameter satisfying $\theta < 0$.

(a) Find the maximum likelihood estimator of θ.

(b) Show that your solution to part (a) is the unique global maximizer of the likelihood if it is. If you cannot show that it is the unique global maximizer, show that it is a local maximizer.

(c) Calculate expected Fisher information for sample size n.

(d) Give an approximate, large sample 95% confidence interval for θ. (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.)

8. [25 pts.] Suppose X is $\mathrm{NegBin}(r, p)$, where $r > 0$ is a known constant and p is the unknown parameter, satisfying $0 < p < 1$. And suppose the prior distribution for p is $\mathrm{Beta}(\alpha_1, \alpha_2)$. Find the posterior distribution for p. (Hint: the posterior is a brand name distribution.)