Practice Midterm w/ Solutions

QPM II

October 3, 2018

1. We observe iid data $X_1, \ldots, X_n$ that represents the number of faculty in a department who leave each year due to retirement, failed retentions, and tenure denials. Your dean tells you that you should model this using a Possion distribution.

$$X_i \sim \frac{\lambda^x e^{-\lambda}}{x!}$$

a. Show that this data is in the exponential family of distributions of the form $h(x)\exp\left(\eta(\theta)T(x) - A(\eta)\right)$

$$\begin{aligned} h(x) &= x^0 \\ \eta(\theta) &= \log(\lambda) \\ T(x) &= x \\ A(\eta) &= \exp(\eta) \end{aligned}$$

b. Calculate the likelihood.

$$\frac{\lambda^{\sum^n x_i} \exp(-n\lambda)}{\prod^n x_i!}$$

c. Calculate the log-likelihood.

$$\log(\lambda)\sum^n x_i - n\lambda - \sum^n x_i!$$

d. Find the MLE for λ.

$$\bar{x}$$

e. Show that the MLE is unbiased and consistent.

$$\begin{aligned} E\left[\frac{\sum x}{n}\right] &= \lambda \\ Var\left[\frac{\sum x}{n}\right] &= \frac{\lambda}{n} \end{aligned}$$

f. Prove (directly) that the MLE is a sufficient statistic for λ, given that $\text{Pois}(\alpha) + \text{Pois}(\beta) \sim \text{Pois}(\alpha + \beta)$. **Do not use the factorization theorem, or your knowledge of the exponential family form**.

$$\frac{\prod^n (\sum x_i)!}{\prod^n x_i!} n^{\sum x_i} = f(x)$$

g. Find the asymptotic distribution of the MLE for λ and show that it is asymptotically efficient.

$$N(\lambda, \frac{\lambda}{n})$$

$$CRLB = Var(\hat{\lambda})$$

h. Calculate a 95% confidence interval for the MLE.

$$\bar{x} \pm 1.96\left(\sqrt{\frac{\bar{x}}{n}}\right)$$

i. Using the delta method, find the asymptotic distribution of $2\sqrt{\lambda}$.

$$N(2\sqrt{\lambda}, \frac{1}{n})$$

2. Your dean points out that the asymptotic properties of the MLE are not relevant because the sample size is so small. Your data consists of the following:

$$\mathbf{x} = (2, 2, 1, 3, 2, 1, 0, 4)$$

a. Estimate the standard error of the asymptotic distribution using the non-parametric bootstrap. Set your number of bootstraps to 10^4.

```
## $theta_hat
## [1] 1.875
##
## $se_boot
## [1] 0.4105978
##
## $n_boot
## [1] 10000
```

b. Unrelated to the dean's request, you're interested in running a computational check for your answer to Problem 1-i. Estimate the standard error of the asymptotic distribution of $2\sqrt{\lambda}$ using the parametric bootstrap, given $\lambda = 4$. Set your number of bootstraps to 10^4. Use n = 100 for each bootstrap.

```
## $theta_hat
## [1] 4
##
## $se_boot
## [1] 0.09983653
##
## $n_boot
## [1] 10000
```

3. No luck. The dean's office remains skeptical of your ability to estimate the asymptotic distribution using such a small sample. You decide to pull out your BBG (big Bayesian guns). You decide to use a Bayesian prior with

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

a. Given $\alpha = 10$ and $\beta = 5$, find the posterior distribution of λ

$$p(\lambda|x) = \text{Gamma}(\lambda; \alpha = 25, \beta = 13)$$

b. Using this posterior, find the $E(\lambda|\mathbf{x})$ and the 95% highest posterior density. (Partial credit for the 95% credible interval.)

$$E\left[p(\lambda|x)\right] = \frac{25}{13}$$

```
## [1] "95% CI"

##     2.5%    97.5%
## 1.239392 2.735195

## [1] "95% highest posterior density"

##    lower    upper
## 1.190282 2.672481
## attr(,"credMass")
## [1] 0.95
```