Stat 5102 (Geyer) Final Exam

Problem 1

Looking at the simplest moment first,

$$E(X_i) = \frac{\alpha}{\lambda} = \frac{\theta}{\theta} = 1,$$

we see that it is not a function of the unknown parameter and hence useless for finding a method of moments estimator. Moving up to the second central moment,

$$\operatorname{var}(X_i) = \frac{\alpha}{\lambda^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

is a simple function of θ and gives the method of moments estimator

$$\hat{\theta}_n = \frac{1}{V_n}.$$

(or the same with V_n replaced by S_n^2).

Problem 2

The formula for f should be familiar. It defines the location-scale family with base density g (Sections 4.1 and 9.2 of the course notes). The variables

$$Y_i = \frac{X_i - \mu}{\sigma}$$

are i. i. d. with density g. We get the same answer for ARE whether we compare $\widetilde{X}_n$ and $\overline{X}_n$ as estimators of μ or whether we compare $\widetilde{Y}_n$ and $\overline{Y}_n$ as estimators of zero.

$\widetilde{Y}_n$ is asymptotically normal with variance

$$\frac{1}{4ng(0)^2} = \frac{4}{n}$$

(Corollary 7.28 in the notes), and $\overline{Y}_n$ is asymptotically normal with variance

$$\frac{\operatorname{var}(Y)}{n} = \frac{\pi^2}{3n}$$

Thus the ARE is either $\pi^2/12$ or $12/\pi^2$ depending on which way you form the ratio. The important point is that $\overline{X}_n$ is the better estimator since

$$\frac{\pi^2}{3} = 3.2899$$

is less than 4.

Alternate Solution

Things are only a little different if we don't realize we can give the answer for $\overline{Y}_n$ and $\widetilde{Y}_n$ instead of $\overline{X}_n$ and $\widetilde{X}_n$.

Since $X_i = \mu + \sigma Y_i$,

$$E(X_i) = \mu$$

and

$$\operatorname{var}(X_i) = \sigma^2 \operatorname{var}(Y_i) = \frac{\pi^2 \sigma^2}{3}$$

so

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\pi^2 \sigma^2}{3n}\right)$$

Note that g is symmetric about zero but f is symmetric about μ, so μ is both the population mean and the population median. And the asymptotic variance of $\widetilde{X}_n$ is

$$\frac{1}{4nf(\mu)^2} = \frac{4\sigma^2}{n}$$

and

$$\widetilde{X}_n \approx \mathcal{N}\left(\mu, \frac{4\sigma^2}{n}\right)$$

The ratio of asymptotic variances is the same as before.

Problem 3

We need to apply the delta method to the estimator $g(\overline{X}_n)$, where

$$g(x) = \frac{2x}{1-x}$$

has derivative

$$g'(x) = \frac{2}{1-x} + \frac{2x}{(1-x)^2} = \frac{2}{(1-x)^2}$$

Because $g(\overline{X}_n)$ is a method of moments estimator it must have asymptotic mean θ (you can check this if you like, but it is already part of the question statement and thus not a required calculation in the answer).

The asymptotic variance is

$$\begin{aligned}
g'(\mu)^2 \sigma^2 &= g'\left(\frac{\theta}{2+\theta}\right)^2 \frac{2\theta}{(2+\theta)^2(3+\theta)} \\
&= \left[\frac{2}{\left(1 - \frac{\theta}{2+\theta}\right)^2}\right]^2 \frac{2\theta}{(2+\theta)^2(3+\theta)} \\
&= \frac{(2+\theta)^4}{4} \frac{2\theta}{(2+\theta)^2(3+\theta)} \\
&= \frac{\theta(2+\theta)^2}{2(3+\theta)}
\end{aligned}$$

Problem 4

If anyone is wondering whether sample size one is "large," recall that a single Poisson random variable is approximately normal if the mean is large (Section F.3 in the appendices of the notes).

The obvious point estimate of $\mu_Y - \mu_X$ is

$$\hat{\mu}_Y - \hat{\mu}_X = Y - X = 410 - 320 = 90$$

which has variance

$$\operatorname{var}(\hat{\mu}_Y - \hat{\mu}_X) = \operatorname{var}(Y) + \operatorname{var}(X) = 410 + 320 = 730$$

and standard deviation

$$\operatorname{sd}(\hat{\mu}_Y - \hat{\mu}_X) = \sqrt{730} = 27.02$$

Thus the large sample confidence interval is

$$90 \pm 1.96 \times 27.02$$

or

$$90 \pm 52.96$$

or

$$(37.0, 143.0)$$

Problem 5

The density of the data is

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

The likelihood for a sample of size n is

$$L_n(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta - 1} = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta - 1} = \theta^n \exp\left((\theta - 1) \sum_{i=1}^{n} \log(x_i) \right)$$

or, if we introduce the variables $y_i = \log(x_i)$,

$$L_n(\theta) = \theta^n e^{(\theta - 1) n \bar{y}_n} = \theta^n e^{\theta n \bar{y}_n} e^{-n \bar{y}_n}$$

The last term can be dropped, since it does not contain the parameter, giving

$$L_n(\theta) = \theta^n e^{\theta n \bar{y}_n}$$

The log likelihood is thus

$$l_n(\theta) = n \log(\theta) + n\theta \bar{y}_n$$

which has derivatives

$$\begin{aligned} l'_n(\theta) &= \frac{n}{\theta} + n\bar{y}_n \\ l''_n(\theta) &= -\frac{n}{\theta^2} \end{aligned}$$

Since the latter does not depend on the data, it is the same as it's expectation, so observed and expected Fisher information are the same

$$J_n(\theta) = I_n(\theta) = \frac{n}{\theta^2}$$

Problem 6

The likelihood is

$$L_n(p) = \prod_{i=1}^{n} p(1-p)^{x_i} = p^n (1-p)^{\sum_i x_i}$$

Note that this is the same functional form as a binomial likelihood. The only difference is the statistics. Here there are n successes and $\sum_i x_i$ failures. Since Bayesian inference depends only on the log likelihood, there is no difference between the calculations here an in Example 11.2.3 in the notes. The posterior distribution is Beta$(n+1, 1+\sum_i x_i)$ and the posterior mean is

$$E(p \mid x_1, \ldots, x_n) = \frac{n+1}{n+2+\sum_i x_i}$$

Problem 7

(a) The models in question have polynomials of degree 1, 2, 3, 4, and 5 as regression functions.

To be more precise, the regression functions are of the form

$$g(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x^i$$

for $i = 1, 2, 3, 4, 5$. Strictly speaking, the model assumptions include the assumption of i. i. d. normal errors, but almost no one said this and no deduction was made for not saying it.

(b) Because linear functions are special cases of quadratic, and so forth. You obtain the models of lower degree by setting the coefficients of higher powers of x to zero in the larger models.

(c) Starting at the bottom of the ANOVA table and reading up

- The quintic model fits no better than the quartic ($P = 0.6233$).
- The quartic model fits no better than the cubic ($P = 0.9170$).
- But the cubic model fits much better than the quadratic ($P < 2 \times 10^{-16}$).

Thus we conclude that the cubic model (Model 3) is correct, which means its supermodels (Model 4 and Model 5) must also be correct. Or to be more finicky we conclude that these data do not give any evidence that these models are incorrect. And we conclude that the quadratic model (Model 2) and its submodel (Model 1) are incorrect. The evidence for that latter conclusion is very strong (repeating what was said above, $P < 2 \times 10^{-16}$).

Many people were confused by "correct" and "incorrect." If a model is correct, then so is every supermodel. If a model is incorrect, then so is every submodel. Hence in a nested sequence of models, there is a smallest correct model (here model 3) and all the models above it are also correct, but all the models below it are incorrect.

Problem 8

$$
\begin{aligned}
f(y \mid p) &= p(1-p)^{y-1} \\
&= \exp[\log(p) + y \log(1-p) - \log(1-p)] \\
&= \exp[y \log(1-p) + \operatorname{logit}(p)]
\end{aligned}
$$

where $\operatorname{logit}(p)$ is defined by equation (12.77a) in the notes.

This clearly fits the form of equation (12.78) in the notes with

$$
\begin{aligned}
\theta &= \log(1-p) \\
\phi &= 1 \\
w &= 1 \\
b(\theta) &= -\operatorname{logit}(p)
\end{aligned}
$$

Of course, the last equation doesn't by itself define $b(\theta)$. To do that we need to know p as a function of θ, that is, we have to solve the first equation above for p giving

$$
p = 1 - e^{\theta}
$$

and plugging that in to the equation for $b(\theta)$ giving

$$
\begin{aligned}
b(\theta) &= -\operatorname{logit}(p) \\
&= -\log\left(\frac{p}{1-p}\right) \\
&= \log\left(\frac{1-p}{p}\right) \\
&= \log\left(\frac{e^{\theta}}{1-e^{\theta}}\right) \\
&= \theta - \log(1 - e^{\theta})
\end{aligned}
$$

or, if you prefer, starting just above the last line

$$\begin{aligned} &= \log\left(\frac{1}{e^{-\theta} - 1}\right) \\ &= -\log(e^{-\theta} - 1) \end{aligned}$$

Additional Stuff

This is just added for my curiosity and perhaps to go in the homework problems some future semester.

$$\begin{aligned} b'(\theta) &= \frac{e^{-\theta}}{e^{-\theta} - 1} \\ &= \frac{1}{1 - e^{\theta}} \\ &= \frac{1}{p} \end{aligned}$$

which is indeed the mean of Y given in Section B.1.8 of the appendices to the notes. Furthermore

$$\begin{aligned} b''(\theta) &= \frac{e^{\theta}}{(1 - e^{\theta})^2} \\ &= \frac{1 - p}{p^2} \end{aligned}$$

which is indeed the variance of Y given in Section B.1.8 of the appendices to the notes. So the GLM theory works and Lemma 12.20 in the notes is true (as, of course, it must be since it is proved in the notes).