Prof. David Draper
Department of
   Applied Mathematics and Statistics
University of California, Santa Cruz

AMS 206: Quiz 4 *[30 points]* (with important Hint)

Name: _____

Please supply your answers to the questions below in the spaces provided. If your answers extend
to more than three pages, please ensure that each continuation answer identifies the question
it's answering on the extra page(s), and (if you're using the scanning option for submission)
make sure to scan all pages of your solution for uploading to `canvas.ucsc.edu`.

(Inference for the variance in the Gaussian model with known mean) People in Las Vegas who
are experts on the National Football League provide a *point spread* for every football game
before it occurs, as a measure of the difference in ability between the two teams (and taking
account of where the game will be played). For example, if Denver is a 3.5–point favorite to
defeat San Francisco, the implication is that betting on whether Denver's final score minus 3.5
points exceeds or falls short of San Francisco's final score is an even-money (50/50) proposition.
The top panel of Figure 1 below (based on data from Gelman et al. (2014)) presents a histogram
of the differences $d =$ (actual outcome – point spread) for a sample of $n = 672$ professional
football games in the early 1980s, with a Normal density superimposed having the same mean
$\bar{d} = 0.07$ and standard deviation (SD) $s = 13.86$ as the sample (if this distribution didn't have
a mean that's close to 0, the experts would be *uncalibrated* and you could make money by
betting against them). You can see from this figure that the model $(D_i \,|\, \sigma \,\mathcal{G} \,\mathcal{B}) \stackrel{\text{IID}}{\sim} N(0, \sigma^2)$
is reasonable for the observed differences d_i, in which $\mathcal{G}$ stands for the Gaussian sampling
distribution assumption (which we're making after looking at the data and is therefore not
part of $\mathcal{B}$). **Hint: You will find it much easier to work with the Gaussian variance
parameter σ^2 if You define $\theta = \sigma^2$ and focus all of Your calculations on θ.**

(1) Write down the likelihood and log likelihood functions for θ in this model *[4 points]*. Show
   that $\hat{\theta}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} d_i^2$, which takes the value 191.8 with the data in Figure 1, is both
   sufficient and the maximum likelihood estimator (MLE) for θ *[4 points]*. Plot the log
   likelihood function for θ in the range from 160 to 240 with these data, briefly explaining
   why it should be slightly skewed to the right. *[4 points]*

Figure 1: *Top panel: Differences d_i between observed and predicted American football scores, 1981–1984; bottom panel: prior, likelihood and posterior for θ with the data in the top panel and the improper prior.*
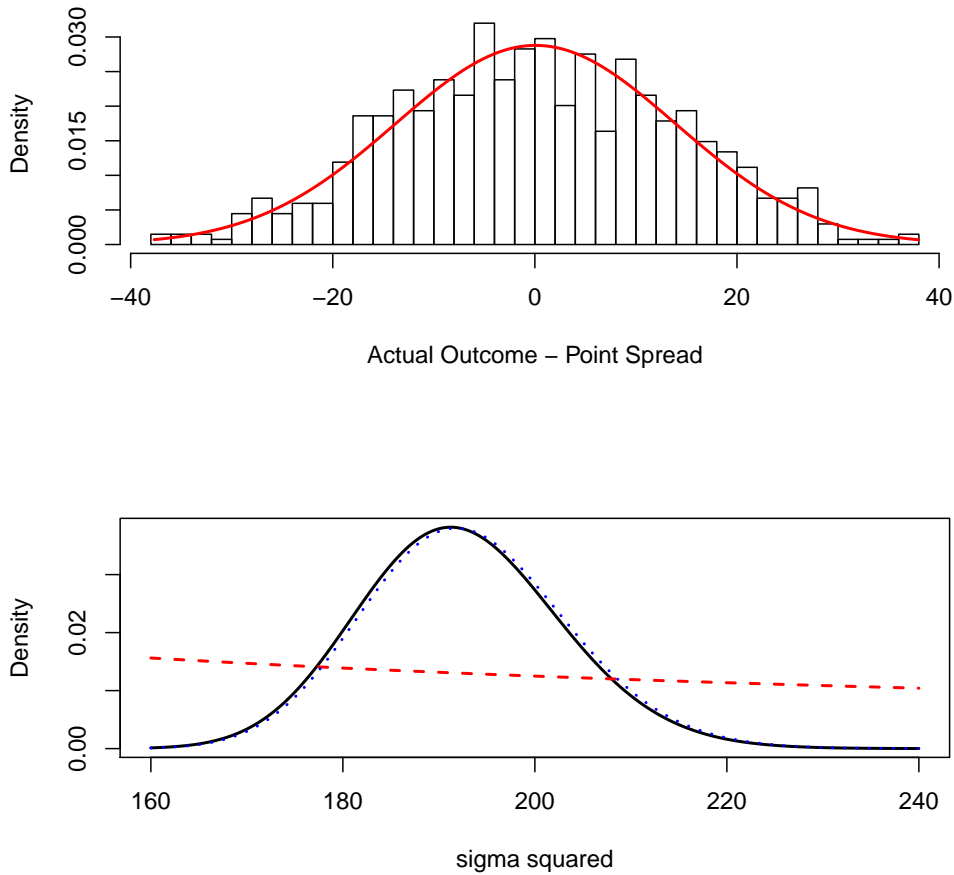
(2) The conjugate prior for θ in this model turns out to be (You don't have to show this) the *scaled inverse chi-square* distribution,

$$\theta \sim \chi^{-2}(\nu_0, \sigma_0^2), \quad \text{i.e.,} \quad p(\theta) = c\,\theta^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(-\frac{\nu_0\,\sigma_0^2}{2\,\theta}\right), \tag{1}$$

in which ν_0 is the prior sample size and σ_0^2 is a prior estimate of θ. In an attempt to be "non-informative" people sometimes work with a version of equation (1) obtained by letting $\nu_0 \to 0$, namely $p(\theta) = c_0\,\theta^{-1}$. The resulting prior is *improper* (in the usual sense: that it integrates to ∞), but it turns out that posterior inferences will be sensible nonetheless (even with sample sizes as small as $n = 1$). Show that with the general $\chi^{-2}(\nu_0, \sigma_0^2)$ prior in equation (1), the posterior for θ is $\chi^{-2}\left(\nu_0 + n, \frac{\nu_0\,\sigma_0^2 + n\,\hat{\theta}}{\nu_0 + n}\right)$, and conclude therefore that with the improper prior the posterior distribution is $\chi^{-2}(n, \hat{\theta})$ *[4 points]*.

(3) The bottom panel of Figure 1 plots the prior, likelihood, and posterior densities on the same graph using the data in the top panel of Figure 1 and taking $c_0 = 2.5$ for convenience in the plot. Get R (or some equivalent environment) to reproduce this figure (You'll need to identify which member of the χ^{-2} family the likelihood density is); include Your code as part of Your answer *[4 points]*. Explicitly identify the three curves, and briefly discuss what this plot implies about the updating of information from prior to posterior in this case study *[4 points]*.

> **Programming notes:** If You're working with R, You can use the `scaled.inverse.chisq.density` function I wrote for you, which is posted on the course web page. If You're writing code in a language other than R, note that — because the data values in this example lead to astoundingly large and small numbers on the original scale — it's necessary to do all possible computations on the log scale and wait to transform back to the original scale until the last possible moment (look at my R function to see what I mean). You'll need to be careful to use the correct normalizing constant c in equation (1), which can be found in Appendix A of the Gelman et al. (2014) book, and You'll need a function that computes $\log \Gamma(x)$ (such a function should be built into Your programming environment; *don't* compute $\Gamma(x)$ and take its logarithm [You'll get an overflow error with the data in this problem]).

(4) It can be shown (You don't need to show this) that (as long as $\nu > 4$)

$$(\theta \,|\, \mathcal{B}) \sim \chi^{-2}(\nu, \sigma^2) \quad \rightarrow \quad E(\theta) = \left(\frac{\nu}{\nu - 2}\right) \sigma^2 \quad \text{and} \quad V(\theta) = \left[\frac{2\,\nu^2}{(\nu - 2)^2\,(\nu - 4)}\right] \sigma^4. \quad (2)$$

Use this to compute the posterior mean and SD for θ with this data set (using the improper prior), and compare the posterior mean with the MLE, briefly explaining why they're so similar in this case *[6 points]*.