AAEC/ECON 5126 FINAL EXAM: SOLUTIONS

SPRING 2015 / INSTRUCTOR: KLAUS MOELTNER

This exam is open-book, open-notes, but please work strictly on your own. Please make sure your name is on every sheet you're handing in. You have 120 minutes to complete this exam. You can collect a maximum of 50 points. Each question is scored as indicated below. Vectors are given in lower-case boldface. Matrices are written in upper-case boldface.

Question I (20 points): Estimating a Population Proportion via Maximum Likelihood

You are researching groundwater contamination in a community of homes, all of which receive their water supply through a private well. You have a sample of n wells, y of which are contaminated ($y \leq n$, of course). Your parameter of interest is θ, the proportion of contaminated wells in the entire community.

You believe the sample likelihood follows a Binomial distribution, given as

$$\begin{aligned} L(\theta) &= \binom{n}{y} \theta^y (1-\theta)^{n-y}, \quad \theta \in [0,1], \quad y = 1, 2, \ldots, n \quad \text{where} \\ \binom{n}{y} &= \frac{n!}{(n-y)!y!} \end{aligned} \tag{1}$$

In this notation, n is the "number of trials" (the number of wells in your sample), and y is the number of "successes" (the number of contaminated wells in your sample). The expectation and variance of the binomial are given, respectively, as $E(y|\theta, n) = n\theta$, and $V(y|\theta, n) = n\theta(1-\theta)$.

Part (a) *14 points*
Derive the following analytical constructs:

(1) The log-likelihood function $lnL(\theta)$ (2 points)
(2) The gradient $g(\theta)$ (2 points)
(3) The Hessian $H(\theta)$ (2 points)
(4) The Information matrix $I(\theta)$ (2 points)
(5) Show that the score identity holds. (2 points)
(6) Show that the information identity holds. (4 points)

Date: May 12, 2015.

Solution:

$$
\begin{aligned}
&lnL(\theta) = ln\binom{n}{y} + y * ln(\theta) + (n-y)\, ln(1-\theta) \\
&g(\theta) = y\theta^{-1} - (n-y)(1-\theta)^{-1} = (\theta(1-\theta))^{-1} y - n(1-\theta)^{-1} \\
&H(\theta) = -y\theta^{-2} - (n-y)(1-\theta)^{-2} \\
&I(\theta) = -E(H(\theta)) = \frac{n}{\theta(1-\theta)} \quad (\text{use } E(y|\theta,n) = n*\theta \text{ for } y \text{ in } H(\theta)) \\
&E(g(\theta)) = 0 \quad (\text{use } E(y|\theta,n) = n*\theta \text{ for } y \text{ in } g(\theta)) \\
&V(g(\theta)) = (\theta(1-\theta))^{-2} V(y|\theta,n) = (\theta(1-\theta))^{-2} n\theta(1-\theta) = \frac{n}{\theta(1-\theta)} = I(\theta), \\
&\text{using } V(y|\theta,n) = n\theta(1-\theta) \text{ in the second expression of } g(\theta)).
\end{aligned}
\tag{2}
$$

Part (b) *6 points*
Assume you have a sample of 100 wells, 20 of which are contaminated. Find the MLE estimate for θ and its standard error.

Solution:
Setting the gradient from above to zero and solving for θ we obtain $\hat{\theta} = \frac{y}{n}$, which, for this sample, equals 0.2. The variance of θ is the inverse of the information matrix: $V(\theta) = \frac{\theta(1-\theta)}{n}$. So in this case we obtain $\hat{V}\left(\hat{\theta}\right) = 0.0016$. The square root of this gives the standard error of 0.04.

Question II (30 points): Estimating a Population Proportion via Bayesian Methods

Continuing with the problem from above, you have obtained information from other communities that are located on the same groundwater aquifer. For these locations, the average proportion of contaminated wells is 0.25 with a standard deviation of 0.14.

You use this information to specify a prior distribution for θ as a *Beta density* with shape parameters $\alpha_0 = 2$ and $\beta_0 = 6$. This density is given as:

$$p(\theta) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \theta^{\alpha_0 - 1} (1-\theta)^{\beta_0 - 1} \quad \theta \in [0,1], \tag{3}$$

where $\Gamma(.)$ denotes the mathematical gamma function. The expectation and variance of the Beta are given, respectively, as $E(\theta) = \frac{\alpha_0}{\alpha_0+\beta_0}$, and $V(\theta) = \frac{\alpha_0\beta_0}{(\alpha_0+\beta_0)^2(\alpha_0+\beta_0+1)}$.

You can easily verify (but you don't have to...) that for the given prior shape parameters $\alpha_0 = 2$ and $\beta_0 = 6$ we obtain a prior expectation for θ of 0.25 and a prior standard deviation of (approx.) 0.14. Along with the natural bounds of the Beta at 0 and 1, this seems indeed like a reasonable prior distribution for θ.

Part (a) *8 points*
Using the same binomial likelihood as in Q.1 (that is, the *un-logged* version), find the posterior distribution of θ, call it $p(\theta|y,n)$. Show that it is again a Beta with shape parameters α_1 and β_1, and show that these posterior parameters are a function of the prior parameters and the data.

(*Hint*: Multiply all *relevant* parts of the prior with all *relevant* parts of the likelihood to obtain the posterior kernel. Simplify, and recognize the resulting kernel as the kernel of another Beta distribution with parameters α_1 and β_1. Then show the explicit form of α_1 and β_1. This should take no more than a minute or two...)

Solution:

$$\begin{gathered} p(\theta|y,n) \propto \theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1} * \theta^{y}(1-\theta)^{n-y} = \\ \theta^{(\alpha_0 + y - 1)}(1-\theta)^{(\beta_0 + n - y - 1)} \end{gathered} \tag{4}$$

Thus, $\theta|y,n \sim \text{Beta}(\alpha_1, \beta_1)$, with $\alpha_1 = \alpha_0 + y$, and $\beta_1 = \beta_0 + n - y$.

Part (b) *6 points*
For your sample of 100 wells with 20 contamination cases, compute the posterior mean and standard deviation of θ. Please be precise to the fourth decimal.

Solution:
First note that for the given data and prior parameters, we have $\alpha_1 = \alpha_0 + y = 22$, and $\beta_1 = \beta_0 + n - y = 86$. We can now use these values in the given expressions for the mean and variance

of a Beta random variable:

$$\begin{aligned} E(\theta|y,n) &= \frac{\alpha_1}{\alpha_1+\beta_1} = 0.2037 \\ V(\theta|y,n) &= \frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)^2(\alpha_1+\beta_1+1)} = 0.0015 \end{aligned} \tag{5}$$

This yields a posterior standard deviation of $sd(\theta|y,n) = sqrt(0.0015) = 0.0386$.

Part (c) *6 points*
Answer the following questions:

(1) Compare the MLE estimate and the posterior mean of θ. Why are they (at least slightly) different? (2 points)
(2) Has the collected data brought meaningful information to add to the prior? How can you tell? (2 points)
(3) For the same collected data, how would you expect your posterior mean to change (up, down, or no change) if the prior mean of θ had been 0.4? Explain. (2 points)

Solution:

(1) The posterior mean is slightly higher due to the influence of the prior distribution, which has a higher mean of 0.25. The sample size is not large enough to completely overpower the prior.
(2) Yes - the posterior standard deviation is much smaller than the prior standard deviation. So the information from the data has "tightened up" the prior distribution.
(3) We would expect the posterior mean to increase, due to the influence of the new prior.

Part (d) *10 points*
Now consider instead a sample of only 10 wells, with 2 contaminations (and the original priors from above). Compute the new MLE estimate and its standard error, and the new posterior mean and standard deviation. (4 points)

(1) How has the posterior mean changed compared to its previous value? Why? (2 points)
(2) Why has the posterior mean changed and the MLE estimate not? (2 points)
(3) Compare the new MLE standard error and the new posterior standard deviation. Which one has increased more (in both absolute and relative terms), and why? (2 points)

Solution:
Using all the new sample information and the analytical results from above, we obtain $\hat{\theta} = 0.2$, $\hat{se}\left(\hat{\theta}\right) = 0.1265$, $\alpha_1 = 4$, $\beta_1 = 14$, $E(\theta|y,n) = 0.2222$, and $sd(\theta|n,y) = 0.0945$.

(1) The posterior mean has increased - this is because the prior influence is now stronger than before due to the smaller sample size.
(2) Same answer as above - the posterior mean is under the influence of the prior, which is completely absent for the MLE estimate.

(3) The MLE standard error has increased much more. In contrast, the precision of the posterior density has suffered less from the decreased sample size. This is due to the relatively high precision brought to the model by the prior distribution. (We have an "informed" prior).