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

AMS 132: Discussion Section 4

Let's dig a bit deeper this week into data analysis with the Exponential distribution.

In a consulting project that one of my Ph.D. students and I worked on at the University of Bath in England before I came to Santa Cruz, a researcher from the Department of Electronic and Electrical Engineering (EEE) at Bath wanted help in analyzing some data on failure times for a particular kind of metal wire (in this problem, failure time was defined to be the number of times the wire could be mechanically stressed by a machine at a given point along the wire before it broke). The $n = 14$ raw data values y_i in one part of his experiment, arranged in ascending order, were

495 541 1461 1555 1603 2201 2750 3468 3516 4319 6622 7728 13161 21194

To save you the trouble of typing the data vector in, I've stored it for you on the course web page in a file called `wire-failure-data-r.txt` .

As we've begun to see in earlier Discussion Sections, probably the simplest model for failure time data is the *Exponential* distribution. Something you have to watch out for with this distribution is that it's commonly *parameterized* in two different ways: earlier we used the parameterization

$$p(y_i \,|\, \beta) = \left\{ \begin{array}{cc} \beta\, e^{-\beta\, y_i} & y_i \geq 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{1}$$

(in which $\beta > 0$), but (about equally often) people instead use the parameterization

$$p(y_i \,|\, \lambda) = \left\{ \begin{array}{cc} \frac{1}{\lambda} \exp(-\frac{y_i}{\lambda}) & y_i \geq 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2}$$

for some $\lambda > 0$. You can see that the two parameterizations are related by the simple expressions $\lambda = \frac{1}{\beta}$ and $\beta = \frac{1}{\lambda}$. In this Discussion Section we'll work with the λ parameterization, using the sampling model $(Y_i \,|\, \lambda) \stackrel{\text{IID}}{\sim} \text{Exponential}(\lambda)$ for $i = 1, \ldots, n$, by which I mean that the marginal sampling distribution for Y_i given λ is specified by equation (2).

(a) Write out the likelihood and log likelihood functions for this sampling model.

(i) Identify a (minimal) sufficient statistic for λ.

(ii) Show that the maximum-likelihood estimate of λ is $\hat{\lambda}_{MLE} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$, which on this data set takes the value 5,043.9.

(iii) In class we found that the MLE for β using the other parameterization was $\hat{\beta}_{MLE} = \frac{1}{\bar{y}}$. Comparing this with your result in (ii), what conjecture about the general behavior of maximum-likelihood estimates does this suggest? (We'll explore this conjecture in class soon.)

(b) To see if this model fits the dataset above, you can make an *Exponential probability plot*, analogous to a Gaussian quantile-quantile (qq) plot to check for normality (we'll look at Gaussian qqplots in class soon). In fact the idea works for more or less any distribution: you plot

$$y_{(i)} \quad \text{versus} \quad F^{-1}\left(\frac{i-0.5}{n}\right), \tag{3}$$

where $y_{(i)}$ are the y values sorted from smallest to largest and F is the CDF of the distribution (the 0.5 is in there to avoid problems at the edges of the data). In so doing it turns out that you're graphing the data values against an approximation of *what you would have expected the data values to look like if the CDF of the y_i really had been F*, so the plot should resemble the 45° line if the fit is good.

(i) Show that the inverse CDF of the Exponential(λ) distribution (parameterized as in equation (2)) is given by

$$F_Y(y|\lambda) = p \iff y = F^{-1}(p) = -\lambda \, \log(1-p)\,. \tag{4}$$

(ii) Use the result in (i) in R to make an Exponential probability plot of the 14 data values in this problem, thereby (informally) demonstrating that the exponential model does provide a reasonably good fit to the data. For the value of λ in this plot you can use the MLE $\bar{y}$, since it's a good estimate of λ.

(c) Show that the conjugate family for the Exponential(λ) likelihood (parameterized as in (2))) is the set of *inverse gamma* distributions $\Gamma^{-1}(\alpha,\beta)$ for $\alpha > 0, \beta > 0$:

$$\lambda \sim \Gamma^{-1}(\alpha,\beta) \quad \iff \quad p(\lambda) = \left\{ \begin{array}{cc} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} \exp\left(-\frac{\beta}{\lambda}\right) & \lambda > 0 \\ 0 & \text{otherwise} \end{array} \right\}. \tag{5}$$

(d) By directly using Bayes' Theorem (and ignoring constants), show that the prior-to-posterior updating rule in this model is

$$\left\{ \begin{array}{ccc} \lambda & \sim & \Gamma^{-1}(\alpha,\beta) \\ (Y_i|\lambda) & \stackrel{\text{IID}}{\sim} & \text{Exponential}(\lambda) \end{array} \right\} \Longrightarrow (\lambda|\boldsymbol{y}) \sim \Gamma^{-1}(\alpha+n, \beta + n\,\bar{y})\,, \tag{6}$$

in which $i = 1, \ldots, n$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$.

(e) It turns out that the mean and variance of the $\Gamma^{-1}(\alpha,\beta)$ distribution are $\frac{\beta}{\alpha-1}$ (provided that $\alpha > 1$) and $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, respectively (as long as $\alpha > 2$). Use this to write down an explicit formula showing that the posterior mean is a weighted average of the prior and sample means. Note also from the formula for the likelihood in this problem that, when thought of as a distribution in λ, it's equivalent to a constant times the $\Gamma^{-1}(n-1, n\,\bar{y})$ distribution.

(f) The guy from EEE has prior information from another experiment he judges to be comparable to this one: from this other experiment the prior for λ should have a mean of about $\mu_0 = 4500$ and an SD of about $\sigma_0 = 1800$. Show that this corresponds to a $\Gamma^{-1}(\alpha_0,\beta_0)$ prior with $(\alpha_0,\beta_0) = (8.25, 32{,}625)$. Plot the prior, likelihood and posterior densities on the same graph, labeling each one, and briefly describe how the posterior is a compromise between the prior information and the data information.