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

AMS 132: Discussion Section 3

1. In Discussion Section 1 we looked at the Exponential distribution as a model for waiting times. The Exponential turns out to be a special case of several other waiting-time distributions, one of which is the *Weibull* distribution: to say that an IID random sample $\boldsymbol{Y} \triangleq (Y_1, \ldots, Y_n)$ of random variables follows the Weibull distribution with parameters $k > 0$ and $\beta > 0$ — written Weibull(k, β) — is to say that the marginal sampling distribution for each Y_i is given by the following:

$$\begin{array}{c} (Y_i \,|\, k, \beta) \stackrel{\text{IID}}{\sim} \text{Weibull}(k, \beta) \\ (i = 1, \ldots, n) \end{array} \quad \text{iff} \quad p(y_i \,|\, k, \beta) = \left\{ \begin{array}{cc} \beta \, k \, (\beta \, y_i)^{k-1} \exp\left[-(\beta \, y_i)^k \right] & \text{if } y_i > 0 \\ 0 & \text{otherwise} \end{array} \right\} \quad (1)$$

(here, as will be usual for the rest of the course, I'm using $p(y_i \,|\, k, \beta)$ to implicitly define which random variable's density is being discussed: $p(y_i \,|\, k, \beta)$ is shorthand for what would have been written $p_{Y_i}(y_i \,|\, k, \beta)$ in AMS 131).

(a) Verify that the Exponential distribution with rate parameter β is a special case of the Weibull by finding a value of k that yields Exponential(β).

(b) k is called the *shape* parameter of the Weibull distribution.

(i) Figure out the sense in which this is a good name for k by writing and running some `R` code to superimpose the following densities on the same plot: $\{$Weibull$(1, 1)$, Weibull$(2, 1)$, Weibull$(4, 1)$, Weibull$(8, 1)\}$. *Hint 1:* You can either write your own function to evaluate the Weibull density or use the `dweibull` built-in `R` function, but if you use `dweibull` you'll need to issue the command `help( dweibull )` and study what `R` says about its parameterization of what is called Weibull(k, β) in equation (1) above (i.e., `R` uses a different parameterization). *Hint 2:* Your code for doing this should look a lot like the code in the file called `weibull-plotting-r.txt` on the course web page, reproduced for convenience below in Table 1.

(ii) What happens to the shape of the Weibull(k, β) distribution as k increases? Explain briefly.

Inference about the parameters (k, β) in the Weibull(k, β) sampling model is more difficult if k is unknown; we'll return to this problem later in the course, when we have stronger tools with which to tackle it. So for the rest of the problem let's adopt the sampling model

$$(Y_i \,|\, k, \beta) \stackrel{\text{IID}}{\sim} \text{Weibull}(k, \beta) \quad (i = 1, \ldots, n), \quad k > 0 \text{ known} \,. \qquad (2)$$

(c) With the observed data vector given by $\boldsymbol{y} \triangleq (y_1, \ldots, y_n)$, show that the likelihood function $\ell(\beta \,|\, \boldsymbol{y})$ arising from model (2) is

$$\ell(\beta \,|\, \boldsymbol{y}) = \beta^{n\,k} \left(\prod_{i=1}^n y_i \right)^{k-1} \exp\left(-\beta^k \sum_{i=1}^n y_i^k \right) , \qquad (3)$$

Table 1: *R code template for problem 1 (b) (i).*

```
y.low <-                # you fill this in and adjust as needed
                        # to create a plot in which the horizontal
                        # and vertical limits frame the curves
                        # to be plotted without cutting them off
                        # and without a lot of extraneous space
                        # in which the curve isn't interesting;
                        # i suggest starting with y.low = 0

y.high <-               # you fill this in and adjust as needed;
                        # i suggest starting with y.high = 4

y.grid <- seq( y.low, y.high, length = 500 )

max.density <-          # you fill this in and adjust as needed;
                        # i suggest starting with max.density = 1

plot( y.grid, dweibull( y.grid, 1, 1 ), xlab = 'y', ylab = 'Density',
  type = 'l', lwd = 2, ylim = c( 0, max.density ) )

lines( y.grid, dweibull( y.grid, 2, 1 ), lwd = 2, lty = 2, col = 'red' )

lines( y.grid, dweibull( y.grid, 4, 1 ), lwd = 2, lty = 3, col = 'blue' )

# and so on
```

and that therefore the log likelihood function in this model (ignoring irrelevant terms that are constant in β) is

$$\ell\ell(\beta \,|\, \boldsymbol{y}) = n\,k\,\log\beta \;-\; \beta^k \sum_{i=1}^{n} y_i^k\,. \tag{4}$$

Use this to show that the maximum-likelihood estimate $\hat{\beta}_{MLE} \triangleq \hat{\beta}$ has the expression

$$\hat{\beta} = \left(\frac{n}{\sum_{i=1}^{n} y_i^k} \right)^{\frac{1}{k}}\,. \tag{5}$$

Describe in simple terms how to compute this estimator, given the data vector $\boldsymbol{y}$.

(d) Show that the Fisher information for β in this model can be expressed as

$$\hat{I}\left(\hat{\beta}\right) = \frac{n\,k}{\hat{\beta}^2}\,, \tag{6}$$

and therefore demonstrate that a large-sample $100(1-\alpha)\%$ confidence interval for β based on maximum likelihood has the simple expression

$$\hat{\beta} \,\pm\, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\hat{\beta}}{\sqrt{n\,k}}\,. \tag{7}$$

(e) Look again at equation (3): the term $\left(\prod_{i=1}^n y_i\right)^{k-1}$ is constant in β, so we can just call it $c > 0$ and write

$$\ell(\beta \mid \boldsymbol{y}) = c\, \beta^{n\,k} \exp\left(-\beta^k \sum_{i=1}^n y_i^k \right) = c\, \left(\beta^k\right)^n \exp\left(-\beta^k \sum_{i=1}^n y_i^k \right) , \tag{8}$$

There appears to be something more fundamental about β^k in this model than β itself, because β enters into equation (8) only through β^k. Further evidence for this point of view can be found by rewriting the maximum-likelihood estimator in equation (5) as

$$\hat{\beta}^k = \frac{n}{\sum_{i=1}^n y_i^k} . \tag{9}$$

So let's define a new parameter $\theta \triangleq \beta^k$; from equation (8) the likelihood function for θ (obtainable by simple substitution) is

$$\ell(\theta \mid \boldsymbol{y}) = c\, \theta^n \exp\left[-\left(\sum_{i=1}^n y_i^k \right) \theta \right] . \tag{10}$$

As a Bayesian I want to think about $\ell(\theta \mid \boldsymbol{y})$ as an un-normalized probability density in θ, and (as in the Kaiser ICU case study, with a Bernoulli sampling model) I'm wondering if there's a conjugate prior for this likelihood, because that would make the calculations easier. It turns out there *is* such a conjugate prior here: it's called the *Gamma*$(\alpha, \lambda) \triangleq \Gamma(\alpha, \lambda)$ family, with $\alpha > 0$ and $\lambda > 0$, defined for $\theta > 0$:

$$\theta \sim \Gamma(\alpha, \lambda) \quad \text{iff} \quad p(\theta) = \left\{ \begin{array}{cc} c\, \theta^{\alpha-1} \exp(-\lambda\, \theta) & \text{if } \theta > 0 \\ 0 & \text{otherwise} \end{array} \right\} . \tag{11}$$

Verify, by direct inspection, that the product of the likelihood density in equation (10) and the prior density in equation (11) is another member of the $\Gamma(\alpha, \lambda)$ family; in so doing you've just proven that the $\Gamma(\alpha, \lambda)$ distribution is conjugate to a version of the Weibull(k, β) likelihood in which the unknown parameter is defined to be $\theta = \beta^k$. Write out the conjugate updating rule for θ in this model in the form

> If your prior distribution for $\theta = \beta^k$ is $\Gamma(\alpha, \lambda)$ and your sampling distribution for $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ is Weibull(k, β) for known $k > 0$, then your posterior distribution for θ given $\boldsymbol{y} = (y_1, \dots, y_n)$ is Gamma with parameters $\underline{\quad}_1$ and $\underline{\quad}_2$ (your job is to fill in $\underline{\quad}_1$ and $\underline{\quad}_2$).

Briefly give details on how you could work out the posterior distribution for β from the posterior distribution for θ.