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 \( Y \equiv (Y_1, \ldots, Y_n) \) of random variables follows the Weibull distribution with parameters \( k > 0 \) and \( \beta > 0 \) — written \( \text{Weibull}(k, \beta) \) — is to say that the marginal sampling distribution for each \( Y_i \) is given by the following:

\[
(Y_i \mid k, \beta) \overset{\text{IID}}{\sim} \text{Weibull}(k, \beta) \quad \text{iff} \quad p(y_i \mid k, \beta) = \begin{cases} 
\beta k (\beta y_i)^{k-1} \exp[-(\beta y_i)^k] & \text{if } y_i > 0 \\
0 & \text{otherwise}
\end{cases}
\]

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

(a) Verify that the Exponential distribution with rate parameter \( \beta \) is a special case of the Weibull by finding a value of \( k \) that yields \( \text{Exponential}(\beta) \).

(b) \( k \) is called the \textit{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 \texttt{R} code to superimpose the following densities on the same plot: \{\text{Weibull}(1,1), \text{Weibull}(2,1), \text{Weibull}(4,1), \text{Weibull}(8,1)\}. \textit{Hint 1}: You can either write your own function to evaluate the Weibull density or use the \texttt{dweibull} built-in \texttt{R} function, but if you use \texttt{dweibull} you’ll need to issue the command \texttt{help(dweibull)} and study what \texttt{R} says about its parameterization of what is called \texttt{Weibull}(k, \beta) in equation (1) above (i.e., \texttt{R} uses a different parameterization). \textit{Hint 2}: Your code for doing this should look a lot like the code in the file called \texttt{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\((k, \beta)\) distribution as \( k \) increases? Explain briefly.

Inference about the parameters \((k, \beta)\) in the \( \text{Weibull}(k, \beta) \) 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

\[
(Y_i \mid k, \beta) \overset{\text{IID}}{\sim} \text{Weibull}(k, \beta) \quad (i = 1, \ldots, n), \quad k > 0 \text{ known}.
\]

(c) With the observed data vector given by \( y \equiv (y_1, \ldots, y_n) \), show that the likelihood function \( \ell(\beta \mid y) \) arising from model (2) is

\[
\ell(\beta \mid y) = \beta^n k \left( \prod_{i=1}^{n} y_i \right)^{k-1} \exp\left(-\beta k \sum_{i=1}^{n} y_i^k \right),
\]

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

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

\[
y.\text{high} <- \text{# you fill this in and adjust as needed;}\ 
  \text{# i suggest starting with y.\text{high} = 4}\n\]

\[
y.\text{grid} <- \text{seq( y.\text{low}, y.\text{high}, length = 500 )}\n\]

\[
\text{max.density} <- \text{# you fill this in and adjust as needed;}\ 
  \text{# i suggest starting with max.density = 1}\n\]

\[
\text{plot( y.\text{grid}, dweibull( y.\text{grid}, 1, 1 ), xlab = ’y’, ylab = ’Density’,}\ 
  \text{type = ’l’, lwd = 2, ylim = c( 0, max.density ) )}\n\]

\[
\text{lines( y.\text{grid}, dweibull( y.\text{grid}, 2, 1 ), lwd = 2, lty = 2, col = ’red’ )}\n\]

\[
\text{lines( y.\text{grid}, dweibull( y.\text{grid}, 4, 1 ), lwd = 2, lty = 3, col = ’blue’ )}\n\]

\[
\text{# and so on}\n\]

and that therefore the log likelihood function in this model (ignoring irrelevant terms that are constant in } \beta \text{) is}

\[
\ell(\beta \mid y) = nk \log \beta - \beta^k \sum_{i=1}^{n} y_i^k. \quad (4)
\]

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

\[
\hat{\beta} = \left( \frac{n}{\sum_{i=1}^{n} y_i^k} \right)^{\frac{1}{k}}. \quad (5)
\]

Describe in simple terms how to compute this estimator, given the data vector } y. \text{(d) Show that the Fisher information for } \beta \text{ in this model can be expressed as}

\[
\hat{I}(\beta) = \frac{nk}{\beta^2}, \quad (6)
\]

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

\[
\hat{\beta} \pm \Phi^{-1}(1 - \alpha/2) \frac{\beta}{\sqrt{nk}}. \quad (7)
\]
(e) Look again at equation (3): the term \( \prod_{i=1}^{n} y_i^{k-1} \) is constant in \( \beta \), so we can just call it \( c > 0 \) and write

\[
\ell(\beta | y) = c \beta^{nk} \exp \left( -\beta^k \sum_{i=1}^{n} y_i^k \right) = c \left( \beta^k \right)^n \exp \left( -\beta^k \sum_{i=1}^{n} y_i^k \right),
\]

(8)

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

\[
\hat{\beta}^k = \frac{n}{\sum_{i=1}^{n} y_i^k}.
\]

(9)

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

\[
\ell(\theta | y) = c \theta^{n} \exp \left[ -\left( \sum_{i=1}^{n} y_i^k \right) \theta \right].
\]

(10)

As a Bayesian I want to think about \( \ell(\theta | y) \) as an un-normalized probability density in \( \theta \), 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 \( \theta > 0 \):

\[
\theta \sim \Gamma(\alpha, \lambda) \quad \text{iff} \quad p(\theta) = \begin{cases} c \theta^{\alpha-1} \exp(-\lambda \theta) & \text{if } \theta > 0 \\ 0 & \text{otherwise} \end{cases}.
\]

(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(\( k, \beta \)) likelihood in which the unknown parameter is defined to be \( \theta = \beta^k \). Write out the conjugate updating rule for \( \theta \) in this model in the form

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

Briefly give details on how you could work out the posterior distribution for \( \beta \) from the posterior distribution for \( \theta \).