HW8

Zhuodiao Kuang

2023-11-23

Problem 1

Since $T_i \sim LN(\mu, \sigma^2)$, we have the p.d.f. of the log-normal:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)}{2\sigma^2}\right)$$

The log-likelihood of the sample $\{t_1, t_2, \ldots, t_n\}$ is:

$$LL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln t_i - \mu)^2 - \sum_{i=1}^{n}\ln t_i$$

Taking the F.O.C. with respect to μ and σ, and set them to zero, we can get:

$$\frac{\partial LL}{\partial \mu} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}\ln t_i - \mu\right) = 0$$
$$\frac{\partial LL}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(\ln t_i - \mu)^2}{2\sigma^4} = 0$$

We can solve them MLE $\hat{\mu}, \hat{\sigma}^2$ from the above equations:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln t_i$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln t_i - \hat{\mu})^2$$

Therefore, MLEs have a closed form.

Problem 2

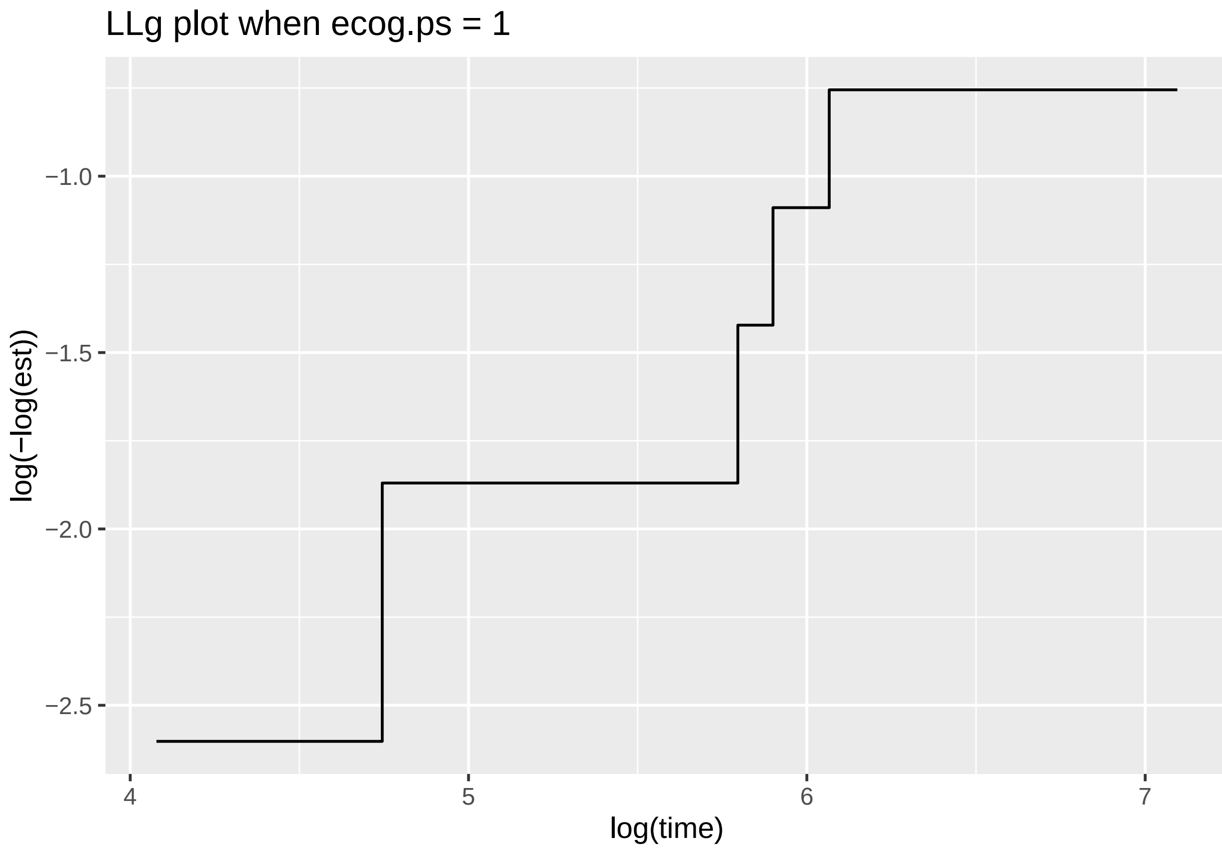
```
##
## Call:
## survreg(formula = Surv(futime, fustat) ~ ecog.ps, data = ovarian,
##     dist = "exponential")
##              Value Std. Error     z       p
## (Intercept)  7.893      0.971  8.13 4.3e-16
## ecog.ps     -0.475      0.586 -0.81    0.42
##
```

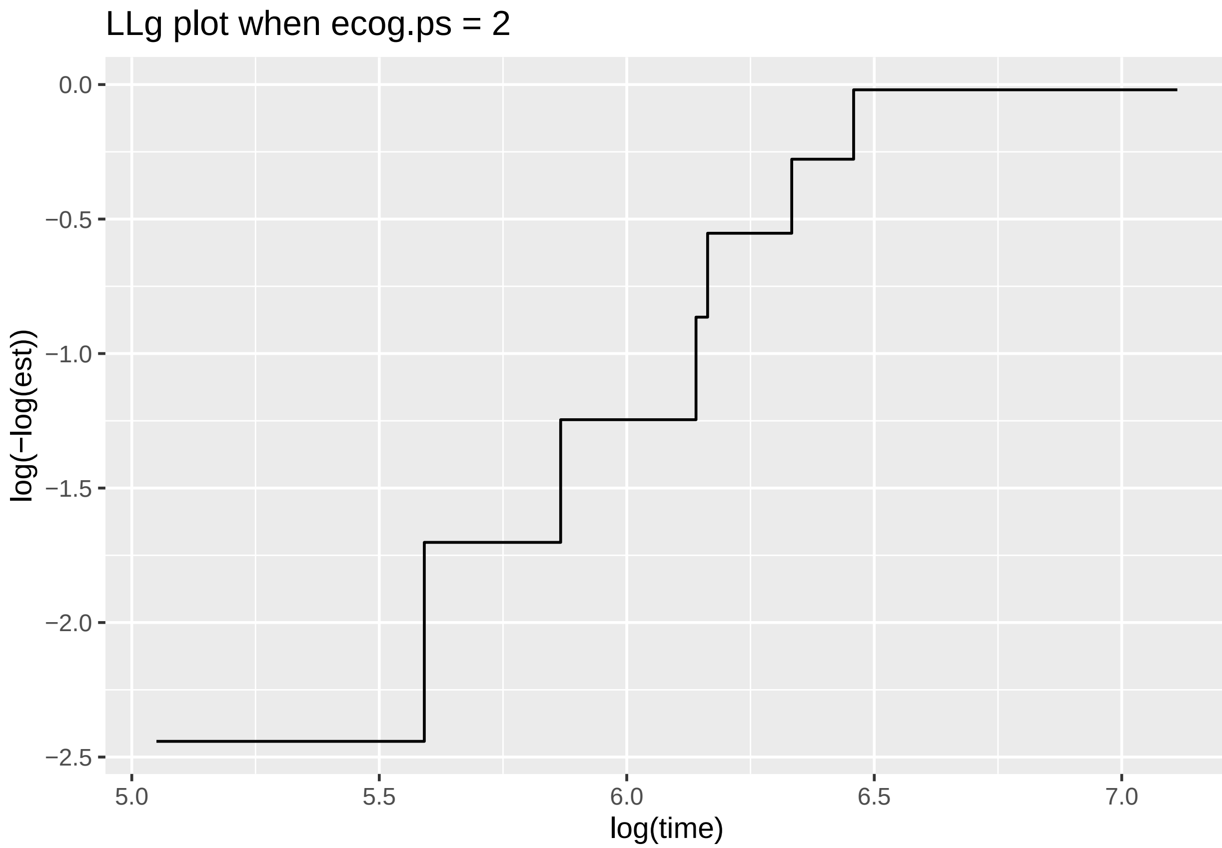
```
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -97.7   Loglik(intercept only)= -98
##  Chisq= 0.67 on 1 degrees of freedom, p= 0.41
## Number of Newton-Raphson Iterations: 4
## n= 26
```

```
##
## Call:
## survreg(formula = Surv(futime, fustat) ~ ecog.ps, data = ovarian,
##     dist = "weibull")
##              Value Std. Error     z      p
## (Intercept)  7.766      0.915  8.49 <2e-16
## ecog.ps     -0.431      0.534 -0.81   0.42
## Log(scale)  -0.108      0.254 -0.43   0.67
##
## Scale= 0.897
##
## Weibull distribution
## Loglik(model)= -97.6   Loglik(intercept only)= -98
##  Chisq= 0.69 on 1 degrees of freedom, p= 0.41
## Number of Newton-Raphson Iterations: 5
## n= 26
```

Because both tests of two distributions generate relatively high p-values, so the survival data, by ecog status, doesn't follow the exponential distribution or the Weibull distribution.

Besides we can take $log{-}logS(t)$ and plot against log t.


LLg plot when ecog.ps = 1
−1.0
−1.5
−2.0
−2.5
log(−log(est))
4
5
6
7
log(time)



Because the slop of a straight line is not 1, it may by considered as a Weibull distribution. But the size of data is limited and the plot is not so smooth, with the p-values calculated above, we are not sure whether the survival data follows the exponential distribution or the Weibull distribution.

Problem 3

The likelihood of Weibull distribution is:

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} h\left(T_i \mid Z_i\right)^{\Delta_i} S\left(T_i\right) \\
&= \prod_{i=1}^{n} \left(h_0(T_i) e^{\beta Z_i}\right)^{\Delta_i} e^{-\lambda T_i^{\alpha}} \\
&= \prod_{i=1}^{n} \left(\lambda \alpha T_i^{\alpha-1} e^{\beta Z_i}\right)^{\Delta_i} e^{-\lambda T_i^{\alpha}} \\
&= \left(\lambda\alpha \times 16^{\alpha-1} e^{\beta}\right)^{1} e^{-\lambda \times 16^{\alpha}} \times e^{-\lambda \times 20^{\alpha}} \\
&\times \left(\lambda\alpha \times 12^{\alpha-1}\right)^{1} e^{-\lambda \times 12^{\alpha}} \times e^{-\lambda \times 14^{\alpha}} \\
&\times \left(\lambda\alpha \times 11^{\alpha-1}\right)^{1} e^{-\lambda \times 11^{\alpha}} \\
&\times \left(\lambda\alpha \times 9^{\alpha-1} e^{\beta}\right)^{1} e^{-\lambda \times 9^{\alpha}} \\
&= (\lambda\alpha)^4 (16^{\alpha-1} 12^{\alpha-1} 11^{\alpha-1} 9^{\alpha-1}) e^{2\beta - \lambda(16^{\alpha}+20^{\alpha}+12^{\alpha}+14^{\alpha}+11^{\alpha}+9^{\alpha})} \\
&= (\lambda\alpha)^4 19008^{\alpha-1} e^{2\beta - \lambda(16^{\alpha}+20^{\alpha}+12^{\alpha}+14^{\alpha}+11^{\alpha}+9^{\alpha})}
\end{aligned}
$$

Problem 4

```
##
## Call:
## survreg(formula = Surv(time, event == 1) ~ trt, data = leu_dat,
##     dist = "loglogistic")
##              Value Std. Error     z       p
## (Intercept)  1.893      0.208  9.12 < 2e-16
## trt6-MP      1.265      0.326  3.89   1e-04
## Log(scale)  -0.604      0.150 -4.02 5.7e-05
##
## Scale= 0.547
##
## Log logistic distribution
## Loglik(model)= -107.7   Loglik(intercept only)= -115.4
##  Chisq= 15.38 on 1 degrees of freedom, p= 8.8e-05
## Number of Newton-Raphson Iterations: 4
## n= 42
```

From the AFT model summary, we can see that the coefficient corresponds to the treatment effect is significant, which means that the treatment 6-MP has a significant effect on the survival of patients with Acute Myelogenous Leukemia.

The corresponding model is:

$$\log T_i = 1.8927 + 1.2655 \times I(\text{trt}_i = \text{ 6-MP})$$