# leukemia

Shengjia Tu

2024-05-02

## (a)

Please see the proof at the end of this report.

## (b)

To find the maximum likelihood estimate, we construct the negative log likelihood function. This allows for maximizing the log likelihood by minimizing the negative log likelihood using the R optim() function.

```r
f <- function(par,y,x,u,n) {
  beta0 <- par[1]
  beta1 <- par[2]
  beta2 <- par[3]
  delta <- par[4]
   n <- 33
  # i <- 1:n
  z <- (y-beta0-beta1*x-beta2*u)/delta
  -((-1)*n*log(delta)+sum(z-exp(z)))
}# the output is negative log-likelihood
library(MASS)
x <- log(leuk$wbc/10000)
y <- log(leuk$time)
u <- as.numeric(leuk$ag) - 1# change AG to 0 and 1
```

```r
# this optim() minimizes negative log-likelihood which maximizes log-likelihood to calculate MLE
out.opt <- optim(par = c(1,1,1,1), fn = f, x = x, y = y, u = u, n = 33, hessian = T)
```

The output of the maximum likelihood estimate of $\beta_0$, $\beta_1$, $\beta_2$ and $\delta$ is:

```r
out.opt$par
```

```
## [1]  2.9940090 -0.3100765  1.0213056  1.0403512
```

The corresponding standard errors are obtained from the diagonal elements of the asymptotic variance which is the inverse of the negative of the second derivative of the log-likelihood. The negative of the second derivative of the log-likelihood is the Hessian matrix because optim() minimizes the negative log-likelihood.

```r
sqrt(diag(solve(out.opt$hessian)))
```

```
## [1] 0.2837543 0.1313000 0.3780126 0.1447655
```

### (c)

To check if the estimate calculated above is the MLE, we calculate the log-likelihood for a range of values in the neighborhood of the MLE calculated above (the range is from -1 to 1).

```r
d <- as.matrix(
  rbind(cbind(seq(-1, 1, 0.1), 0, 0, 0),
        cbind(0, seq(-1, 1, 0.1), 0, 0),
        cbind(0, 0, seq(-1, 1, 0.1), 0),
        cbind(0, 0, 0, seq(-1, 1, 0.1)))
)
m <- matrix(rep(out.opt$par, each = 21*4), ncol = 4) + d
d <- cbind(d, -apply(m, MARGIN = 1, FUN = f, y = y, x = x, u = u, n = n))
d <- cbind(d, d[,5] - (-f(out.opt$par, y = y, x = x, u = u, n = n)))
d[,5] <- round(d[,5], 2)
d[,6] <- round(d[,6], 2)
colnames(d) <- c("diff_beta0", "diff_beta1", "diff_beta2", "diff_delta", "likelihood", "difference")
d
```

```
##       diff_beta0 diff_beta1 diff_beta2 diff_delta    likelihood    difference
##  [1,]       -1.0        0.0        0.0        0.0 -7.468000e+01 -2.156000e+01
##  [2,]       -0.9        0.0        0.0        0.0 -6.995000e+01 -1.683000e+01
##  [3,]       -0.8        0.0        0.0        0.0 -6.594000e+01 -1.282000e+01
##  [4,]       -0.7        0.0        0.0        0.0 -6.258000e+01 -9.470000e+00
##  [5,]       -0.6        0.0        0.0        0.0 -5.983000e+01 -6.710000e+00
##  [6,]       -0.5        0.0        0.0        0.0 -5.762000e+01 -4.500000e+00
##  [7,]       -0.4        0.0        0.0        0.0 -5.590000e+01 -2.780000e+00
##  [8,]       -0.3        0.0        0.0        0.0 -5.463000e+01 -1.510000e+00
##  [9,]       -0.2        0.0        0.0        0.0 -5.377000e+01 -6.500000e-01
## [10,]       -0.1        0.0        0.0        0.0 -5.327000e+01 -1.600000e-01
## [11,]        0.0        0.0        0.0        0.0 -5.312000e+01  0.000000e+00
## [12,]        0.1        0.0        0.0        0.0 -5.327000e+01 -1.500000e-01
## [13,]        0.2        0.0        0.0        0.0 -5.369000e+01 -5.700000e-01
## [14,]        0.3        0.0        0.0        0.0 -5.437000e+01 -1.250000e+00
## [15,]        0.4        0.0        0.0        0.0 -5.527000e+01 -2.160000e+00
## [16,]        0.5        0.0        0.0        0.0 -5.639000e+01 -3.270000e+00
## [17,]        0.6        0.0        0.0        0.0 -5.769000e+01 -4.570000e+00
## [18,]        0.7        0.0        0.0        0.0 -5.916000e+01 -6.040000e+00
## [19,]        0.8        0.0        0.0        0.0 -6.079000e+01 -7.670000e+00
## [20,]        0.9        0.0        0.0        0.0 -6.256000e+01 -9.440000e+00
## [21,]        1.0        0.0        0.0        0.0 -6.446000e+01 -1.134000e+01
## [22,]        0.0       -1.0        0.0        0.0 -1.150400e+02 -6.192000e+01
## [23,]        0.0       -0.9        0.0        0.0 -9.929000e+01 -4.617000e+01
## [24,]        0.0       -0.8        0.0        0.0 -8.686000e+01 -3.374000e+01
## [25,]        0.0       -0.7        0.0        0.0 -7.713000e+01 -2.401000e+01
## [26,]        0.0       -0.6        0.0        0.0 -6.959000e+01 -1.647000e+01
## [27,]        0.0       -0.5        0.0        0.0 -6.385000e+01 -1.073000e+01
## [28,]        0.0       -0.4        0.0        0.0 -5.960000e+01 -6.480000e+00
```

```
## [29,]          0.0        -0.3         0.0         0.0 -5.657000e+01 -3.460000e+00
## [30,]          0.0        -0.2         0.0         0.0 -5.458000e+01 -1.460000e+00
## [31,]          0.0        -0.1         0.0         0.0 -5.347000e+01 -3.500000e-01
## [32,]          0.0         0.0         0.0         0.0 -5.312000e+01  0.000000e+00
## [33,]          0.0         0.1         0.0         0.0 -5.345000e+01 -3.300000e-01
## [34,]          0.0         0.2         0.0         0.0 -5.440000e+01 -1.280000e+00
## [35,]          0.0         0.3         0.0         0.0 -5.594000e+01 -2.830000e+00
## [36,]          0.0         0.4         0.0         0.0 -5.808000e+01 -4.960000e+00
## [37,]          0.0         0.5         0.0         0.0 -6.081000e+01 -7.690000e+00
## [38,]          0.0         0.6         0.0         0.0 -6.419000e+01 -1.108000e+01
## [39,]          0.0         0.7         0.0         0.0 -6.829000e+01 -1.518000e+01
## [40,]          0.0         0.8         0.0         0.0 -7.321000e+01 -2.009000e+01
## [41,]          0.0         0.9         0.0         0.0 -7.908000e+01 -2.596000e+01
## [42,]          0.0         1.0         0.0         0.0 -8.609000e+01 -3.297000e+01
## [43,]          0.0         0.0        -1.0         0.0 -6.422000e+01 -1.110000e+01
## [44,]          0.0         0.0        -0.9         0.0 -6.178000e+01 -8.660000e+00
## [45,]          0.0         0.0        -0.8         0.0 -5.972000e+01 -6.600000e+00
## [46,]          0.0         0.0        -0.7         0.0 -5.799000e+01 -4.870000e+00
## [47,]          0.0         0.0        -0.6         0.0 -5.657000e+01 -3.450000e+00
## [48,]          0.0         0.0        -0.5         0.0 -5.543000e+01 -2.320000e+00
## [49,]          0.0         0.0        -0.4         0.0 -5.455000e+01 -1.430000e+00
## [50,]          0.0         0.0        -0.3         0.0 -5.390000e+01 -7.800000e-01
## [51,]          0.0         0.0        -0.2         0.0 -5.345000e+01 -3.300000e-01
## [52,]          0.0         0.0        -0.1         0.0 -5.320000e+01 -8.000000e-02
## [53,]          0.0         0.0         0.0         0.0 -5.312000e+01  0.000000e+00
## [54,]          0.0         0.0         0.1         0.0 -5.319000e+01 -8.000000e-02
## [55,]          0.0         0.0         0.2         0.0 -5.341000e+01 -3.000000e-01
## [56,]          0.0         0.0         0.3         0.0 -5.376000e+01 -6.500000e-01
## [57,]          0.0         0.0         0.4         0.0 -5.423000e+01 -1.110000e+00
## [58,]          0.0         0.0         0.5         0.0 -5.480000e+01 -1.690000e+00
## [59,]          0.0         0.0         0.6         0.0 -5.547000e+01 -2.360000e+00
## [60,]          0.0         0.0         0.7         0.0 -5.623000e+01 -3.120000e+00
## [61,]          0.0         0.0         0.8         0.0 -5.707000e+01 -3.960000e+00
## [62,]          0.0         0.0         0.9         0.0 -5.799000e+01 -4.870000e+00
## [63,]          0.0         0.0         1.0         0.0 -5.896000e+01 -5.850000e+00
## [64,]          0.0         0.0         0.0        -1.0 -8.742300e+15 -8.742300e+15
## [65,]          0.0         0.0         0.0        -0.9 -4.202173e+04 -4.196862e+04
## [66,]          0.0         0.0         0.0        -0.8 -7.709100e+02 -7.177900e+02
## [67,]          0.0         0.0         0.0        -0.7 -1.963200e+02 -1.432000e+02
## [68,]          0.0         0.0         0.0        -0.6 -1.052600e+02 -5.214000e+01
## [69,]          0.0         0.0         0.0        -0.5 -7.599000e+01 -2.287000e+01
## [70,]          0.0         0.0         0.0        -0.4 -6.353000e+01 -1.041000e+01
## [71,]          0.0         0.0         0.0        -0.3 -5.758000e+01 -4.470000e+00
## [72,]          0.0         0.0         0.0        -0.2 -5.470000e+01 -1.580000e+00
## [73,]          0.0         0.0         0.0        -0.1 -5.344000e+01 -3.300000e-01
## [74,]          0.0         0.0         0.0         0.0 -5.312000e+01  0.000000e+00
## [75,]          0.0         0.0         0.0         0.1 -5.335000e+01 -2.300000e-01
## [76,]          0.0         0.0         0.0         0.2 -5.393000e+01 -8.200000e-01
## [77,]          0.0         0.0         0.0         0.3 -5.474000e+01 -1.620000e+00
## [78,]          0.0         0.0         0.0         0.4 -5.568000e+01 -2.560000e+00
## [79,]          0.0         0.0         0.0         0.5 -5.671000e+01 -3.590000e+00
## [80,]          0.0         0.0         0.0         0.6 -5.779000e+01 -4.680000e+00
## [81,]          0.0         0.0         0.0         0.7 -5.891000e+01 -5.790000e+00
## [82,]          0.0         0.0         0.0         0.8 -6.003000e+01 -6.910000e+00
```

```
## [83,]         0.0         0.0         0.0         0.9 -6.116000e+01 -8.040000e+00
## [84,]         0.0         0.0         0.0         1.0 -6.228000e+01 -9.160000e+00
```

The above table shows the changes of the estimate for each parameter, the corresponding values of the log likelihood and the differences between these values and the value of the log-likelihood at the MLE obtained from optim(). The log-likelihood evaluated at the MLE is larger than all the other values of the log-likelihood. This indicates the log-likelihood evaluated at the MLE is at least a local maximum.

Proof of (a). Consider

$$f\left(t_i|\alpha_i,\gamma\right)=\frac{\gamma}{\alpha_i}\left(\frac{t_i}{\alpha_i}\right)^{\gamma-1}\exp(-(t_i/\alpha_i)^{\gamma})$$

By taking log, we have the log-likelihood for the $i$th subject:

$$\begin{aligned}\ell\left(t_i|\alpha_i,\gamma\right) &= \log\left(\frac{\gamma}{\alpha_i}\right)+(\gamma-1)\log\left(\frac{t_i}{\alpha_i}\right)-\left(\frac{t_i}{\alpha_i}\right)^{\gamma}\\ &= \log\gamma-\log\alpha_i+(\gamma-1)\log\left(\frac{t_i}{\alpha_i}\right)-\left(\frac{t_i}{\alpha_i}\right)^{\gamma}\end{aligned}$$

Let $z_i=\frac{\log t_i-\log\alpha_i}{\delta}=\frac{\log\left(\frac{t_i}{\alpha_i}\right)}{\delta}$, $\alpha_i=\beta_0+\beta_1x_i+\beta_2u_i$ and $\delta=\frac{1}{\gamma}$. Then

$$\frac{t_i}{\alpha_i}=\exp\left(\delta z_i\right),\quad \log\left(\frac{t_i}{\alpha_i}\right)=\delta z_i,\quad \gamma\delta=1$$

Thus we can express $\ell\left(t_i|\alpha_i,\gamma\right)$ as:

$$\begin{aligned}\ell\left(t_i|\alpha_i,\gamma\right) &= \log\gamma-\log\alpha_i+(\gamma-1)\delta z_i-\left[\exp\left(\delta z_i\right)\right]^{\gamma}\\ &= \log\left(\frac{1}{\delta}\right)-\log\alpha_i+(\gamma-1)\delta z_i-\exp\left(\gamma\delta z_i\right)\\ &= -\log\delta-\log\alpha_i+(\gamma-1)\delta z_i-\exp\left(z_i\right)\\ &= -\log\delta-\log\alpha_i+z_i-\delta z_i-\exp\left(z_i\right)\\ &= -\log\delta-\log\alpha_i-\delta z_i+z_i-\exp\left(z_i\right)\end{aligned}$$

Since

$$\log\left(t_i\right)-\log\left(\alpha_i\right)=\delta z_i$$

we have:

$$\log\left(t_i\right)=\log\left(\alpha_i\right)+\delta z_i$$

Thus $\log\alpha_i+\delta z_i=\log\left(t_i\right)$ does not involve parameters and can be dropped and the resulting log likelihood is:

$$\ell\left(t_i|\alpha_i,\gamma\right)=-\log\delta+z_i-\exp\left(z_i\right)$$

For the whole sample, the log-likelihood is given by:

$$\ell\left(\alpha_i,\gamma\right)=\sum\ell\left(t_i|\alpha_i,\gamma\right)=-n\log\delta+\sum\left(z_i-\exp\left(z_i\right)\right)$$