Homework

2023-03-03

Import Data

```
BallBearing <- c(17.88, 28.92, 33, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96,
                 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12,
                 98.64, 105.12, 105.84, 127.92, 128.04, 173.40)
```

Q1

Exponential distribution

```
theta.hat <- mean(BallBearing) ### theoretical estimates
theta.hat
```

```
## [1] 72.22087
```

```
exp_loglik <- function(theta, data){
  n <- length(data)
  l <- -n * log(theta) - (1 / theta * sum(data))
  return(-l)
}

exp.par <- optim(100, exp_loglik, data = BallBearing,
                 method = "Brent", lower = 0, upper = 200)
exp.par$par
```

```
## [1] 72.22087
```

Normal distribution

```
mu.hat <- mean(BallBearing) ### theoretical estimates
sig2.hat <- sum((BallBearing - mu.hat) ^ 2) / length(BallBearing)
c(mu.hat, sig2.hat)
```

```
## [1]   72.22087 1344.46343
```

```
norm_loglik <- function(pars, data){
  mu <- pars[1]
  sig2 <- pars[2]
  if(sig2 > 0){
    n <- length(data)
    l <- -n / 2 * log(2 * pi * sig2) - sum((data - mu) ^ 2 / (2 * sig2))
    return(-l)
  }else{
    return(10^10)
  }
```

```
}

norm.par <- optim(c(72, 1300), norm_loglik, data = BallBearing)
norm.par$par
```

```
## [1]   72.21673 1344.33714
```

F(t) = 1 - exp(-t ^ theta)

```
cdf_loglik <- function(theta, data){
  n <- length(data)
  l <- n * log(theta) + (theta - 1) * sum(log(data)) - sum(data ^ theta)
  return(-l)
}

cdf.par <- optim(100, cdf_loglik, data = BallBearing,
                 method = "Brent", lower = 0, upper = 200)
cdf.par$par
```

```
## [1] 0.1918497
```

Two-parameter exponential distribution

```
TwoParExp_loglik <- function(pars, data) {
  n <- length(data)
  theta <- pars[1]
  gamma <- pars[2]
  # life < gamma
  if (min(data) < gamma) {
    return(10^10)
  } else{
    l <- -(n * log(theta)) - sum(data-gamma) / theta
    return(-l)
  }
}

TwoParExp.par <- optim(c(100, 0), TwoParExp_loglik, data = BallBearing)
TwoParExp.par$par
```

```
## [1] 54.38196 17.88000
```

Q2

```
exp.par$par
```

```
## [1] 72.22087
```

```
norm.par$par
```

```
## [1]   72.21673 1344.33714
```

```
cdf.par$par
```

```
## [1] 0.1918497
```

```r
TwoParExp.par$par
```

```
## [1] 54.38196 17.88000
```

Q3

Choose a better model by likelihood values or AIC.

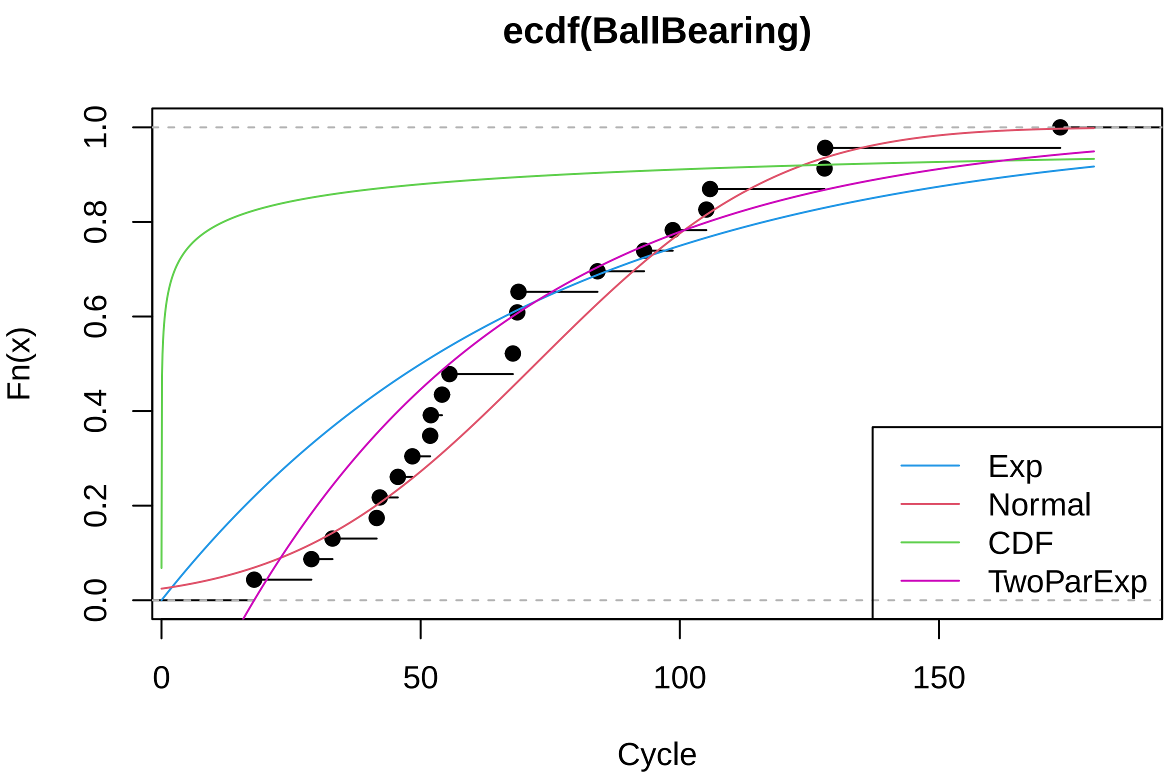
```r
results <- data.frame(Distribution = c("Exp", "Normal", "F(t)", "TwoParExp"),
                      Likelihood = c(-exp.par$value, -norm.par$value,
                                     -cdf.par$value, -TwoParExp.par$value),
                      AIC = c(2*exp.par$value+2*1, 2*norm.par$value+2*2,
                              2*cdf.par$value+2*1, 2*TwoParExp.par$value+2*2))
results
```

```
##   Distribution Likelihood      AIC
## 1          Exp  -121.4338 244.8675
## 2       Normal  -115.4787 234.9574
## 3         F(t)  -166.3686 334.7372
## 4    TwoParExp  -114.8914 233.7827
```

Based on the above table, the fitted two-parameter exponential distribution is better because it has largest log-likelihood value and smallest AIC.

Check if the data fit well by the estimated values.

```r
plot(ecdf(BallBearing), xlab = "Cycle")
t0 <- seq(10^-6, 180, 0.1)
# ecdf(exp)
exp.theta <- exp.par$par[1]
lines(t0, pexp(t0, (1 / exp.theta)), col = 4)
# ecdf(norm)
norm.mu <- norm.par$par[1]
norm.sigma <- norm.par$par[2] ^ 0.5
lines(t0, pnorm(t0, norm.mu, norm.sigma), col = 2)
# ecdf(cdf)
cdf.theta <- cdf.par$par
lines(t0, 1 - exp(-t0 ^ cdf.theta), col = 3)
# ecdf(TwoParExp)
TwoParExp.theta <- TwoParExp.par$par[1]
TwoParExp.gamma <- TwoParExp.par$par[2]
lines(t0, 1 - exp(-(t0 - TwoParExp.gamma) / TwoParExp.theta), col = 6)
# legend
legend("bottomright", c("Exp", "Normal", "CDF", "TwoParExp"),
       col = c(4, 2, 3, 6), lty = 1)
```



Check if the data fit well by the estimated values.

```r
# exp
ks.test(BallBearing, "pexp", 1 / exp.par$par)
```

```
## Warning in ks.test.default(BallBearing, "pexp", 1/exp.par$par): ties should not
## be present for the Kolmogorov-Smirnov test

##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  BallBearing
## D = 0.30681, p-value = 0.02634
## alternative hypothesis: two-sided
```

```r
# norm
ks.test(BallBearing, "pnorm", norm.par$par[1], norm.par$par[2]^0.5)
```

```
## Warning in ks.test.default(BallBearing, "pnorm", norm.par$par[1],
## norm.par$par[2]^0.5): ties should not be present for the Kolmogorov-Smirnov
## test

##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  BallBearing
## D = 0.18843, p-value = 0.3877
## alternative hypothesis: two-sided
```

```r
# F(t) = 1 - exp(-x ^ theta)
CDF.cdf <- function(x) {
```

```r
  p <-  1 - exp(-x ^ cdf.theta)
  return(p)
}
ks.test(BallBearing, "CDF.cdf")
```

```
## Warning in ks.test.default(BallBearing, "CDF.cdf"): ties should not be present
## for the Kolmogorov-Smirnov test
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  BallBearing
## D = 0.82428, p-value = 5.34e-14
## alternative hypothesis: two-sided
```

```r
# TwoParExp
TwoParExp.cdf <- function(x) {
  p <- 1 - exp(-(x-TwoParExp.gamma) / TwoParExp.theta)
  return(p)
}
ks.test(BallBearing, "TwoParExp.cdf")
```

```
## Warning in ks.test.default(BallBearing, "TwoParExp.cdf"): ties should not be
## present for the Kolmogorov-Smirnov test
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  BallBearing
## D = 0.22211, p-value = 0.2065
## alternative hypothesis: two-sided
```

Based on the above information, "Normal distribution" (p-value=0.3877) and "Two-parameter exponential distribution"(p-value=0.2065) are both good (because of p-value>0.05). And "Normal distribution" is better.

Conclusion

According to the results of likelihood values, AIC and K-S test, we can conclude that "Normal distribution" and "Two-parameter exponential distribution" are both good enough to fit the data.