
HOMEWORK 1 : STAT 5113

A1: Find the MLE of the unknown parameter $\theta \in (0, 1)$ based on one observation of $X \sim \text{Bin}(n, \theta)$. Are there particular values of the data x for which the MLE does not exist?

Answer:

As we are considering only one observation, the pmf for this observation is the likelihood function $L(\theta)$ itself

$$L(\theta) = P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

if the log-likelihood function is denoted as $l(\theta)$ then

$$l(\theta) = \ln L(\theta) = x \ln \theta + (n - x) \ln(1 - \theta) + \ln \binom{n}{x}$$

$$\frac{\partial l}{\partial \theta} = \frac{x}{\theta} + \frac{n-x}{1-\theta}(-1) + 0 = \frac{x-n\theta}{\theta(1-\theta)}$$

for maximum likelihood

$$\frac{\partial l}{\partial \theta} = \frac{x-n\theta}{\theta(1-\theta)} = 0$$

which yields the estimate of the parameter $\hat{\theta}$ as

$$\hat{\theta} = \frac{x}{n}$$

as $\theta \in (0, 1)$ so, $x \neq 0$ and $x \neq n$ i.e. $0 < x < n$

A2: Find the MLE of the unknown positive parameter θ from a random sample of size n from the continuous uniform distribution on $[0, \theta]$.

Answer:

For a random sample of size n i.e. x_1, \dots, x_n from a uniform distribution on $[0, \theta]$ the pdf of x_i is

$$f(x_i; \theta) = \frac{1}{\theta}$$

the likelihood function $L(\theta)$ can be defined

$$L(\theta) = \frac{1}{\theta^n} = \theta^{-n}, \quad \theta > 0$$

if the log-likelihood function is denoted as $l(\theta)$ then

$$l(\theta) = -n \ln \theta$$

after taking the derivative

$$\frac{\partial l}{\partial \theta} = -\frac{n}{\theta}$$

this likelihood function is a decreasing function and it has the maximum value if

$$\theta = \max(x_1, \dots, x_n)$$

if $\hat{\theta} = \max(x_1, \dots, x_n)$ then it is the MLE estimate itself

A3: Consider independent random variables with:

$$X_i \sim N(\mu_1, \sigma^2), \quad i = 1, \dots, n_1, \quad Y_j \sim N(\mu_2, \sigma^2), \quad j = 1, \dots, n_2,$$

Formulate the statistical model and derive the MLE of the unknown parameters.

Answer:

The pdf of any observation x_i and y_j from random variables X_i and Y_j respectively is

$$f(x_i; \mu_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_i - \mu_1)^2}{2\sigma^2}$$

$$g(y_j; \mu_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y_j - \mu_2)^2}{2\sigma^2}$$

so, the likelihood function $L(\mu_1, \mu_2, \sigma^2)$ is

$$L(\mu_1, \mu_2, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n_1+n_2}{2}} \exp^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (x_i - \mu_1)^2 + \sum_{j=1}^{n_2} (y_j - \mu_2)^2 \right]}$$

if the log-likelihood function is $l(\mu_1, \mu_2, \sigma^2)$ then

$$l(\mu_1, \mu_2, \sigma^2) = \ln L(\mu_1, \mu_2, \sigma^2) = -\frac{n_1 + n_2}{2} \ln(2\pi) - \frac{n_1 + n_2}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (x_i - \mu_1)^2 + \sum_{j=1}^{n_2} (y_j - \mu_2)^2 \right]$$

taking the derivative of $l(\mu_1, \mu_2, \sigma^2)$ w.r.t μ_1 and setting it equal to 0 we get

$$\frac{\partial l}{\partial \mu_1} = 0 + 0 - \frac{1}{2\sigma^2} \left[2(-1) \sum_{i=1}^{n_1} (x_i - \mu_1) \right] = \frac{\sum_{i=1}^{n_1} x_i - n_1 \mu_1}{\sigma^2} = 0$$

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1}$$

similarly,

$$\hat{\mu}_2 = \frac{\sum_{j=1}^{n_2} y_j}{n_2}$$

now, taking the derivative of $l(\mu_1, \mu_2, \sigma^2)$ w.r.t σ^2 and setting it equal to 0 we get

$$\frac{\partial l}{\partial \sigma^2} = 0 - \frac{n_1 + n_2}{2\sigma^2} + \frac{1}{2\sigma^4} \left[\sum_{i=1}^{n_1} (x_i - \mu_1)^2 + \sum_{j=1}^{n_2} (y_j - \mu_2)^2 \right] = 0$$

yields

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1} (x_i - \mu_1)^2 + \sum_{j=1}^{n_2} (y_j - \mu_2)^2}{n_1 + n_2}$$

A4: The Lognormal distribution is sometimes used to model positive quantities and it has density

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad x > 0$$

Compute MLEs of μ and σ based on the leukemia survival data described in class. Compare the fit of the Lognormal model with that of the Gamma and Weibull models: which model appears to give the best fit for the survival data?

Answer:

Here is the complete code in R with explanation

```
rm(list=ls()) # Clear Workspace
cat("\014") # Clear Console
graphics.off() # Clear Plots

## Working with Leukemia Data
y = c(65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65)
hist(y, breaks = "FD", prob = TRUE, main = "Survival time (weeks)");
rug(jitter(y)) # Plot Histogram

## Lognormal Model (From Formula)
mu_hat = sum(log(y))/length(y)
mu_hat # 3.361693
sd_hat = sqrt(sum((log(y)-mu_hat)^2)/length(y))
sd_hat # 1.626578

## Lognormal Model
LN_NL = function(mu,sd){
  -sum(dlnorm(y, meanlog = mu, sdlog = sd, log = T))
}
result_lognormal <- mle(LN_NL, start = list(mu = 1, sd = 1),
  nobs = length(y), lower = c(1e-8, 1e-8))
```

```
summary(result_lognormal)
theta_lognormal <- result_lognormal@coef
theta_lognormal # mu = 3.361694 v = 1.626579
curve(dlnorm(x, meanlog = theta_lognormal["mu"],
            sdlog = theta_lognormal["sd"]), add = TRUE, lty = 2, lwd = 2)
```

this segment has output as follows

```
      mu      sd
3.361694 1.626579

### Gamma model
GM_NL <- function(alpha, beta) {
  -sum(dgamma(y, shape = alpha, rate = beta, log = TRUE))
}
result_gamma <- mle(GM_NL, start = list(alpha = 1, beta = 1),
                   nobs = length(y), lower = c(1e-8, 1e-8))
summary(result_gamma)
theta_gamma <- result_gamma@coef
theta_gamma
curve(dgamma(x, shape = theta_gamma["alpha"],
            rate = theta_gamma["beta"]), add = TRUE, lty = 3, lwd = 2)

### Weibull model
WB_NL <- function(a, b) -sum(dweibull(y, shape = a, scale = b, log = TRUE))
result_weibull <- mle(WB_NL, start = list(a = 1, b = 1),
                     nobs = length(y), lower = c(1e-8, 1e-8))
result_weibull
theta_weibull <- result_weibull@coef
theta_weibull
curve(dweibull(x, shape = theta_weibull["a"],
              scale = theta_weibull["b"]), add = TRUE, lty = 4, lwd = 2)

## Insert Legend
legend("topright", c("Lognormal", "Gamma", "Weibull"),
      lty = 2:4, lwd = 1.5, seg.len = 3, bty = 'n')

### Compare fits
logLik(result_lognormal) # lognormal
logLik(result_gamma) # gamma
logLik(result_weibull) # weibull
```

The last section has the following output

```
> ### Compare fits
> logLik(result_lognormal) # lognormal
'log Lik.' -89.54087 (df=2)

> logLik(result_gamma) # gamma
'log Lik.' -86.87377 (df=2)

> logLik(result_weibull) # weibull
'log Lik.' -87.10948 (df=2)
```

As seen from the likelihood values the Gamma distribution has the maximum value of -86.87377 so Gamma distribution gives the best fit. Also attached the graph in separate page.

B1: Consider a random sample from the family of distributions defined by the pdf

$$f(x; \alpha) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}, \quad x > x_0$$

where x_0 is a known positive value and α is an unknown positive parameter. Find the MLE of α .

Answer:

This is a random variable with *Pareto distribution*. If we gather n samples namely x_1, \dots, x_n from this distribution then the likelihood function $L(\alpha)$ becomes

$$L(\alpha) = \prod_{i=1}^n f(x_i; \alpha) = \prod_{i=1}^n \frac{\alpha x_0^\alpha}{x_i^{\alpha+1}} = \frac{\alpha^n x_0^{\alpha n}}{(\prod_{i=1}^n x_i)^{\alpha+1}}$$

then the log-likelihood function $l(\alpha)$ becomes

$$l(\alpha) = \ln L(\alpha) = n \ln \alpha + n \alpha \ln x_0 - (\alpha + 1) \ln \left(\prod_{i=1}^n x_i \right) = n \ln \alpha + n \alpha \ln x_0 - (\alpha + 1) \sum_{i=1}^n \ln x_i$$

For MLE estimate

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_0 - \sum_{i=1}^n \ln x_i = 0$$

which yields the estimate of α to be

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln x_i - n \ln x_0}$$

B2: In R, get hold of the speed-of-light measurements by typing: `sp = morley[["Speed"]]`. Use the method of maximum likelihood to fit both a normal and a Gamma model to the data. Which model gives the best fit?

Answer:

Here is the R code for this problem

```
rm(list=ls()) # Clear Workspace
cat("\014") # Clear Console
graphics.off() # Clear Plots
## Data Input ##
sp = morley[["Speed"]]
## These Two Lines Remove Two Outliers
# remove = c(max(sp), min(sp))
# sp = sp[!sp %in% remove]
hist(sp, breaks = "FD", prob = TRUE, main = "Speed of Light"); rug(jitter(sp))

## Normal Model ##
nor_negL = function(mu_nor, sd_nor){
  -sum(dnorm(sp, mean = mu_nor, sd = sd_nor, log = T))
}
result_normal = mle(nor_negL, start = list(mu_nor = 1, sd_nor = 1),
  nobs = length(sp), lower = c(1e-8, 1e-8))
summary(result_normal)
theta_normal = result_normal$coef
theta_normal
curve(dnorm(x, mean = theta_normal["mu_nor"],
  sd = theta_normal["sd_nor"]), add = TRUE, lty = 3, lwd = 2)
```

Output:

```
> theta_normal
  mu_nor  sd_nor
852.4000 78.6145

## Gamma Model ##
gamma_negL = function(alpha, beta){
  -sum(dgamma(sp, shape = alpha, rate = beta, log = T))
}
result_gamma = mle(gamma_negL, start = list(alpha = 1, beta = 1),
  nobs = length(sp), lower = c(1e-8, 1e-8))
summary(result_gamma)
```

```
theta_gamma = result_gamma@coef
theta_gamma
curve(dgamma(x, shape = theta_gamma["alpha"],
           rate = theta_gamma["beta"]), add = TRUE, lty = 2, lwd = 1.5)
```

Output:

```
> theta_gamma
      alpha      beta
116.2998738  0.1364396

## Compare Fit ##
logLik(result_normal) # normal
logLik(result_gamma)  # gamma
```

Output:

```
> logLik(result_normal) # normal
'log Lik.' -578.3495 (df=2)

> logLik(result_gamma) # gamma
'log Lik.' -578.774 (df=2)
```

Normal distribution has better fit compared to Gamma distribution. If we remove maximum and minimum values from the data (using the code snippet in initial section) we get

```
> logLik(result_normal) # normal
'log Lik.' -558.9947 (df=2)

> logLik(result_gamma) # gamma
'log Lik.' -559.106 (df=2)
```

So, the normal distribution has better fit. Graph attached in a separate page.
