Question 2: Maximum Likelihood

Tim Yuki Washio

11/14/2020

1.

```
load("data.RData")
```

2.

We define the log-likelihood function as the logarithm of the likelihood function:

$$\mathcal{L}_x(\vartheta) = ln(L_x(\vartheta))$$

For normal distributions we define:

$$\mathcal{L}_x(\mu, \sigma^2) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

For a given $n = 100$ we get:

$$\mathcal{L}_x(\mu, \sigma^2) = -\frac{100}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{100} (x_i - \mu)^2$$

We calculate the partial derivative of log-likelihood with respect to μ and set it equal to 0:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{100} (x_i - \mu) = 0$$

We transform our equation and get:

$$\hat{\mu} = \frac{1}{100} \sum_{i=1}^{100} x_i$$

Now we calculate the partial derivative of log-likelihood with respect to σ and set it equal to 0:

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{100}{2\sigma^2} - \left[\frac{1}{2} \sum_{i=1}^{100} (x_i - \mu)^2\right] \left(-\frac{1}{(\sigma^2)^2}\right) = -\frac{100}{2\sigma^2} + \left[\frac{1}{2} \sum_{i=1}^{100} (x_i - \mu)^2\right] \frac{1}{(\sigma^2)^2} = \frac{1}{2\sigma^2} \left[\frac{1}{\sigma^2} \sum_{i=1}^{100} (x_i - \mu)^2 - 100\right] = 0$$

We transform our equation and get:

$$\hat{\sigma}^2 = \frac{1}{100}\sum_{i=1}^{100}(x_i - \mu)^2$$

```
mu_hat = sum(data)/length(data)
sigma_hat = sqrt(sum((data-mu_hat)^2)/length(data))

cat("mu_hat: ", mu_hat, "\nsigma_hat: ", sigma_hat)
```

```
## mu_hat:  1.275528
## sigma_hat:  2.005976
```

3.

```
negative_log_likelihood <- function(param) {
  mu = param[1]
  sigma = param[2]
  n = length(data)
  log_likelihood = -n*0.5*log(2*pi*sigma^2)-(0.5/sigma^2)*sum((data-mu)^2)
  negative_log_likelihood = -(log_likelihood)
  return (negative_log_likelihood)
}

gradient <- function(params) {
  mu = params[1]
  sigma = params[2]
  n = length(data)
  result = c(-(sum(data)-n*mu)/sigma^2, (n/sigma)-(1/(sigma^3) * sum((data-mu)^2)))
  return(result)
}

cg_1 <- optim(par=c(0, 1), negative_log_likelihood, method="CG")
bfgs_1 <- optim(par=c(0, 1), fn=negative_log_likelihood, method="BFGS")
cg_2 <- optim(par=c(0, 1), fn=negative_log_likelihood, gr=gradient, method="CG")
bfgs_2 <- optim(par=c(0, 1), fn=negative_log_likelihood, gr=gradient, method="BFGS")
```

4.

Algorithm	Mean	Var	Function Counts	Gradient Counts	Gradient Specified	Value
CG	1.27553	2.00598	297.00000	45.00000	NO	211.50695
CG	1.27553	2.00598	53.00000	17.00000	YES	211.50695
BFGS	1.27553	2.00598	37.00000	15.00000	NO	211.50695
BFGS	1.27553	2.00598	39.00000	15.00000	YES	211.50695