MAST90105: Lab and Workshop Problems for Week 9

The Lab and Workshop this week covers problems arising from Module 6. In the lab this week, you will enter some commands from this sheet into R-studio and see the results.

1 Lab

Introduction

Data: In this lab we will estimate distributions for the Tasmanian Rainfall data that are the maximum daily rainfall in each year from 1995 to 2014 recorded by 34 weather stations in Tasmania. Data source: Australian Bureau of Meteorology. (`http://www.bom.gov.au/`)

Goals: (i) Estimate parameters using maximum likelihood; (ii) estimate parameters using method of moments (iii) estimate relative error in maximum likelihood and method of moments estimators by simulation.

Task 1 — Reading the data in

From last week, you know that in R can import the data with the command:
`tasmania=read.csv("EditedRainfall.csv")`

If the file is located in the working directory, you do not need to provide a full path to this file. You can create variables with the following commands:

```
s1 = tasmania[, 2]  # create a vector for station 1 (Burnie)
s2 = tasmania[, 3]  # create a vector for station 2 (Cape Grim)
```

Task 2 — Maximum likelihood estimation

In this task, we will look at maximum likelihood estimation. In simple cases the MLE is available in closed form. For example, if observations were iid from $N(\mu, \sigma^2)$, then we already know that the MLE for μ and σ^2 are $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = (n-1)S^2/n$. Thus, the ML estimates for the extreme rainfall data can be quickly computed as

```
mu.hat = mean(s1)
mu.hat
n = length(s1)
sigma.hat = sqrt((n - 1) * var(s1)/n)
sigma.hat
```

Probability theory says that a better model for maxima is the extreme value (EV) distribution, or Gumbel distribution, with pdf

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma}\right\} \exp\left[-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right], \quad x \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \sigma > 0.$$

The log-likelihood function for this model is

$$\ell(\mu, \sigma) = -n \log \sigma - \sum_i \frac{x_i - \mu}{\sigma} - \sum_i \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}.$$

Unfortunately, as it is often the case in practice, there is no closed form solution for the MLE and maximisation has to be carried out numerically.

There are number of routines available to maximize the log-likelihood function, including `nlm` and `optim` functions in R.

```r
gumbel.fit1 = function(x, start = c(50, 10)) {
    n = length(x)
    likf = function(pr) {
        mu = pr[1]
        si = pr[2]
        if (si <= 0)
            return(1e+100)
        sx = (x - mu)/si
        # minimize negative log-likelihood
        nloglik = n * log(si) + sum(sx) + sum(exp(-sx))
        nloglik
    }
    mle = nlm(likf, p = start)
    mle
}
out1 = gumbel.fit1(s1, start = c(50, 10))
out1
```

```r
gumbel.fit2 = function(x, start = c(50, 10)) {
    n = length(x)
    likf = function(pr) {
        mu = pr[1]
        si = pr[2]
        sx = (x - mu)/si
        # minimize negative log-likelihood
        nloglik = n * log(si) + sum(sx) + sum(exp(-sx))
        nloglik
    }
    mle = optim(start, likf, method = "L-BFGS-B", lower = c(-50,
        1e-04))
    mle
}
out2 = gumbel.fit2(s1, start = c(50, 10))
out2
```

The arguments in `optim` are the vector of starting values and the function to be minimised. A number of other tuning parameters can be used depending on the optimisation method specified in optim. The function returns a number of useful outputs, including final parameter estimates and value of the minimised negative log-likelihood, and information about convergence (to see them type `fit`). A detailed explanation can be found in `Value` in `help(optim)`. Various classic optimisation algorithms are available within this function (see `Details` and references in `help(optim)`). When the parameters' are known to be in some interval, optimisation can be carried out using the method `method="L-BFGS-B"` for `optim`, or the function `nlminb`, which allow us to specify upper and lower bound for each parameter.

```r
library(MASS)
# you may need to
# install.packages('evd',repos='https://cloud.r-project.org')
# evd contains functions pgumbel, qgumbel, dgumbel,
# rgumbel for the cdf, quantiles, pdf and random
# numbers for the Gumbel distribution the argument
# names are loc and scale
library(evd)
```

Next, carry out a visual check on the fitted models. The following plots the fitted normal and Gumbel models over the histogram of the data.

```r
# Write fitted normal and Gumbel pdfs
mu.hat2 = out1$estimate[1]
sigma.hat2 = out1$estimate[2]
pdf1 = function(x) {
    dnorm(x, mean = mu.hat, sd = sigma.hat)
}
pdf2 = function(x) {
    dgumbel(x, loc = mu.hat2, scale = sigma.hat2)
}

# Plot data and fitted models
hist(s1, freq = FALSE, col = "gray", main = NULL, xlab = "x",
    xlim = c(0, 100))
curve(pdf1, from = 0, to = 100, col = 2, lty = 2, lwd = 2,
    add = TRUE)
curve(pdf2, from = 0, to = 100, col = 1, lty = 1, lwd = 2,
    add = TRUE)
```

From the plots, which distribution appears to be a better fit?

Finally, the value of the maximised log-likelihood function $\ell(\hat{\theta}) = \log L(\hat{\theta})$ is a useful statistic describing the goodness-of-fit of models and can be used for model comparison (provided that the models under exam have the same number of parameters). The largest the likelihood function, the better the model fits the data at hand.

There are a number of ways to obtain this information. For example, it can be extracted from the output of `nlm` as

```
out1$minimum   #Gumbel negative log-likelihood
```

or it can be computed from scratch by evaluating directly a log-likelihood function

```
n = length(s1)
nloglik = 0.5 * n * log(2 * pi) + n * log(sigma.hat) +
    0.5 * sum((s1 - mu.hat)^2)/sigma.hat^2
nloglik  # Normal negative log-likelihood
```

The log-likelihood value for the Gumbel model is larger than that obtained from the normal model, which suggests that the Gumbel model is more appropriate for the extreme rainfall data in Burnie.

Task 3 — Method of moments estimation

In this task we compute point estimates using the method of moments (MM). Assume that the data are generated by $X \sim Gumbel(\mu, \sigma)$ with pdf given in Task 2. It can be shown that

$$E(X) = \gamma\sigma + \mu$$

$$Var(X) = \frac{\pi^2\sigma^2}{6}$$

where $\gamma = 0.577215$ is Euler's gamma constant, which is the limit as $n \to \infty$ of the difference between the harmonic series $1 + 1/2 + 1/3 + \cdots + n$ and the logarithm of n. The connection between the Euler constant and the mean can be established through differentiating the Γ function with respect to the shape parameter. In R, the constant γ is available as `-digamma(1)`, where `digamma` is the derivative of the Γ function.

Let $\overline{X} = \sum_i X_i/n$ and $\hat{\sigma}^2 = \sum_i (X_i - \bar{X})^2/n$ be the first two sample central moments. Note that the method of moments can either use central moments or non-central moments. Because the variance formula for the Gumbel distribution only involves the parameter σ, it makes sense to use central moments here.

Show that the resulting central method of moments estimators for μ and σ are

$$\widetilde{\sigma} = \frac{\sqrt{6}\hat{\sigma}}{\pi}, \quad \widetilde{\mu} = \overline{X} - \gamma\widetilde{\sigma}.$$

The estimators are then easily computed in R:

```
(sigma.tilde = sqrt(6) * sigma.hat/pi)
(mu.tilde = mu.hat + digamma(1) * sigma.tilde)
```

Compare MM and ML estimates for this model. Note that the MM estimates are quite close to the maximum likelihood estimates obtained in Task 2, but the MM estimates are cheap from a computational viewpoint and can be used as the starting point for finding the maximum likelihood estimates.

Task 4 — Error in estimation

As in lectures, the possible error in estimation can be assessed through simulation. In this case, unlike in lectures, the parameters are not known, but the estimates can be used — either the Maximum Likelihood or the Method of Moments estimators. To get reasonable accuracy, 1000 samples will be generated from the extreme value distribution with the estimated parameters from the data, the maximum likelihood and method of moments estimators found for each sample, and the means and standard deviations of the estimators obtained.

In R, only commands for the estimates starting with the maximum likelihood estimates for s1 as the simulating distribution are given. Repeat these, with the necessary changes, substituting the method of moments estimators for the parameters of simulation:

```r
# function to simulate from the Gumbel distribution
# with the same sample size as s1 and mle estimates
# for the parameters
rgum = function(x) {
    rgumbel(length(s1), loc = mu.hat2, scale = sigma.hat2)
}


# function to find the maximum likelihood
# estimators
fit.mle.gum = function(x) {
    out = gumbel.fit1(x, start = c(50, 10))
    out$estimate
}


# function to find the method of moments estimators
fit.mom.gum = function(x) {
    mu.hat = mean(x)
    sigma.hat = sqrt(var(x) * (length(x) - 1)/length(x))
    sigma.tilde = sqrt(6) * sigma.hat/pi
    mu.tilde = mu.hat + digamma(1) * sigma.tilde
    c(mu.tilde, sigma.tilde)
}


# simulate 1000 samples and obtain maximum
# likelihood estimates and method of moments
# estimates for each sample.

mle.est = matrix(nrow = 1000, ncol = 2)
mom.est = matrix(nrow = 1000, ncol = 2)

for (i in 1:1000) {
    x = rgum()
    mle.est[i, ] = fit.mle.gum(x)
```

```
    mom.est[i, ] = fit.mom.gum(x)
}

# means of estimators
mean.mle.est = apply(mle.est, 2, mean)
mean.mom.est = apply(mom.est, 2, mean)
mean.mle.est
mean.mom.est

# SDs of estimators
sd.mle.est = apply(mle.est, 2, sd)
sd.mom.est = apply(mom.est, 2, sd)
sd.mle.est
sd.mom.est
```

Comment on your results comparing the maximum likelihood estimates with the method of moments estimates, particularly on the standard deviations of the estimates.

2 Workshop

1. Find the maximum likelihood estimator, the Cramér-Rao lower bound and thus the asymptotic variance of the maximum likelihood estimator $\hat{\theta}$ for a random sample $X_1, \ldots, X_n$ taken from the following densities. Determine whether the maximum likelihood estimator is unbiased, and if so, whether the maximum likelihood estimator achieves the lower bound.

   a. $f(x; \theta) = (1/\theta^2)x \exp(-x/\theta)$, $0 < x < \infty$, $0 < \theta < \infty$

   b. $f(x; \theta) = (1/(2\theta^3))x^2 \exp(-x/\theta)$, $0 < x < \infty$, $0 < \theta < \infty$

   c. $f(x; \theta) = \frac{1}{6\theta^4}x^3 e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$

2. (Textbook 6.5-2) In some situations where the regression models are useful, it is known that the mean of Y when $X = 0$ is 0, i.e., $Y_i = \beta x_i + \epsilon_i$ where ϵ_i for $i = 1, \ldots, n$ are independent and distributed as $N(0, \sigma^2)$.

   a. Obtain the maximum likelihood estimators, $\hat{\beta}$ and $\hat{\sigma^2}$, of β and σ^2 under this model.

   b. Find the distributions of $\hat{\beta}$ and $\hat{\sigma^2}$ (You may use, without proof, the fact that $\hat{\beta}$ and $\hat{\sigma^2}$ are independent.)

   c. By looking up help in R on "formula" find out the R commands to fit this model (recall the we used "lm" in lectures for regression)

3. Let X be the number of trials needed to observe the 5th success in a sequence of independent Bernoulli trials. Then X has a negative binomial distribution.

    a. What is the mean and variance of X?

    b. Compute the likelihood of p corresponding to a single observation x on X. Call this $L(p)$.

    c. Plot the likelihood as a function of p if we observe $x = 8$.

    d. Deduce the log-likelihood if $X_1, \ldots, X_n$ are independent observations of X.

    e. Compute the derivative of the log-likelihood. Call this s.

    f. Find the maximum likelihood estimator of p based on a single observation x.

    g. Take a derivative of the score function to determine the observed information $\partial^2 \ln f(x; p)/\partial p^2$. Call this i.

    h. Compute the Fisher information.

    i. What is the Cramér-Rao lower bound of unbiased estimators of p?

4. Suppose that $X \sim \text{Binomial}(n, p)$ with n known. Find the maximum likelihood estimator of p and determine its asymptotic variance. Determine whether the maximum likelihood estimator is unbiased, and if so, whether the maximum likelihood estimator achieves the lower bound.

5. Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, where σ^2 is known.

    a. Show that $Y = (X_1 + X_2)/2$ is an unbiased estimator of θ.

    b. Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.

    c. The *efficiency* of an estimator is the ratio of the Cramér-Rao lower bound to the variance of the estimator. What is the efficiency of the estimator in (a)?

6. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \theta)$, where μ is known.

    a. Show the maximum likelihood estimator of θ is $\hat{\theta} = n^{-1} \sum_{i=1}^{n} (X_i - \mu)^2$.

    b. Determine the Cramér-Rao lower bound.

    c. What is the approximate distribution of $\hat{\theta}$ for large n?

    d. What is the exact distribution of $n\hat{\theta}/\theta$? Use your knowledge of this distribution to determine if $\hat{\theta}$ attains the Cramér-Rao lower bound?

7. Suppose $X \sim U(0, \theta)$. We take one observation on X.

    a. Find the method of moments estimator of θ.

    b. Find the maximum likelihood estimator of θ.

    c. The mean square error of an estimator is $MSE = E(\hat{\theta} - \theta)^2$. Show that

$$MSE = \text{Var}(\hat{\theta}) + \text{bias}_{\hat{\theta}}^2$$

    where $\text{bias}_{\hat{\theta}} = E(\hat{\theta}) - \theta$. Which of the estimators in (a) and (b) has the smallest MSE?

8. Let Y be the sum of the observations from a Poisson distribution with mean θ. Let the prior p.d.f. of θ be gamma with parameters α and β so that

$$f(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}, 0 \leq \theta < \infty$$

a. Find the posterior p.d.f. of θ given $Y = y$. (Hint: You should be able to recognise the form of the numerator so only consider the terms that involve θ).

b. If the loss function is $[w(y) - \theta]^2$ find the Bayesian point estimate $w(y)$ of θ.

c. Show that this $w(y)$ is a weighted average of the maximum likelihood estimate y/n and the prior mean $\alpha\beta$, with respective weights $n/(n+1/\beta)$ and $(1/\beta)(n+1/\beta)$.