MAST30027: Modern Applied Statistics

Akira Wang, Student ID: 913391
Assignment 3, 2019
Tutorial: Qiuyi Li, Mon 12pm-1pm

September 30, 2019

All R code and working out is attached at the end of the assignment.

1. (a) The likelihood of Negative Binomial is:

$$\begin{aligned}\ell(r,p) &= \sum_{i=1}^{n} \ln\left(\frac{\Gamma(k_i+r)}{\Gamma(k_i+1))\Gamma(r)}\right) + \sum_{i=1}^{n} r\ln(1-p) + \sum_{i=1}^{n} k_i \ln(p) \\ &= nr\ln(1-p) + \sum_{i=1}^{n} k_i \ln(p) + c.\end{aligned}$$

Then, we solve for the MLE:

$$\begin{aligned}\frac{\partial}{\partial p}\ell(r,p) &= \frac{-nr}{1-p} + \frac{\sum_{i=1}^{n} k_i}{p} = 0 \\ \frac{nr}{1-p} &= \frac{\sum_{i=1}^{n} k_i}{p} \\ \hat{p} &= \frac{\sum_{i=1}^{n} k_i}{nr + \sum_{i=1}^{n} k_i}.\end{aligned}$$

Since k_i is known and $r = 1.5$,

$$\hat{p} = 0.9165.$$

(b) We are given $p \sim \beta(1/2, 1/2)$ and require $p|K$, where K is the set of data points.

$$\begin{aligned}p|K &\propto f(K|p)f(p) \\ &\propto p^{a-1}(1-p)^{b-1}\prod_{i=1}^{n}\frac{\Gamma(k_i+r)}{\Gamma(k_i+1))\Gamma(r)}(1-p)^r p^{k_i} \\ &\propto p^{a-1+\sum_{i=1}^{n} k_i}(1-p)^{b-1+nr} \\ &\propto \beta(a + \sum_{i=1}^{n} k_i, b + nr)\end{aligned}$$

Since k_i is known and $r = 1.5$, $p|K \sim \beta(2403.5, 219.5)$.

(c) We have from above:

$$\hat{p} == \frac{\sum_{i=1}^n k_i}{nr + \sum_{i=1}^n k_i},$$

$$E(p|K) = \frac{a + \sum_{i=1}^n k_i}{a + \sum_{i=1}^n k_i + b + nr}.$$

Let Δ be the difference between the MLE and posterior mean is:

$$\begin{aligned}
\Delta &= \frac{\sum_{i=1}^n k_i}{nr + \sum_{i=1}^n k_i} - \frac{a + \sum_{i=1}^n k_i}{a + \sum_{i=1}^n k_i + b + nr} \\
&= \frac{\sum_{i=1}^n k_i(a + \sum_{i=1}^n k_i + b + nr) - (nr + \sum_{i=1}^n k_i)(a + \sum_{i=1}^n k_i)}{(nr + \sum_{i=1}^n k_i)(a + \sum_{i=1}^n k_i + b + nr)} \\
&= \frac{b \sum_{i=1}^n -nra}{(nr + \sum_{i=1}^n k_i)(a + \sum_{i=1}^n k_i + b + nr)}.
\end{aligned}$$

By definition, $(nr + \sum_{i=1}^n k_i)(a + \sum_{i=1}^n k_i + b + nr) > 0$, so:

$$\text{sign}(\Delta) = \text{sign}(b \sum_{i=1}^n -nra).$$

If $n \to \infty$ or $a \to \infty$, then $nra > b \sum_{i=1}^n k_i$.
This implies the posterior mean will not always be smaller than the MLE estimate.

(d) We are given $r \sim \exp(1/\lambda = 1.5)$ and $p \sim \beta(1/2, 1/2)$.
By the Law of Total Probability:

$$\begin{aligned}
f(r|\mathbf{y}) &\propto f(\mathbf{y}|r) f(r) \\
&\propto \int_0^1 f(\mathbf{y}, p|r) f(r) dp \\
&= \int_0^1 f(\mathbf{y}|r, p) f(r) f(p) dp \\
&\propto \left[\prod_{i=1}^n \frac{\Gamma(y_i + r)}{\Gamma(y_i + 1)\Gamma(r)} (1-p)^r p^{y_i}\right] p^{a-1}(1-p)^{b-1} e^{-\lambda r} \\
&\propto \int_0^1 \left[\prod_{i=1}^n \frac{\Gamma(y_i + r)}{\Gamma(y_i + 1)\Gamma(r)}\right] p^{a-1+\sum_{i=1}^n y_i}(1-p)^{b-1+nr} e^{-\lambda r} dp.
\end{aligned}$$

Solve for $1/\lambda = 1.5$ to get $\lambda = 2/3$.
Hence,

$$f(r|\mathbf{y}) \propto \left[\prod_{i=1}^n \frac{\Gamma(y_i + r)}{\Gamma(y_i + 1)\Gamma(r)}\right] \beta(y + 1/2, nr + 1/2) e^{-2/3r}.$$

2. (a) We are given $X \sim \gamma(\alpha, 1)$ which has CDF:

$$F_X = P(X < x) = \frac{1}{\Gamma(\alpha)}\gamma(\alpha, 1).$$

Let $Y = X/\lambda$, then:

$$\begin{aligned} F_Y &= P(Y < y) \\ &= P(X/\lambda < y) \\ &= P(X < \lambda y) \\ &= \frac{1}{\Gamma(\alpha)}\gamma(\alpha, \lambda). \end{aligned}$$

Therefore, $X/\lambda \sim \gamma(\alpha, \lambda)$.

(b) Let $Y = h(x) \sim \gamma(\alpha, 1)$.
The CDF is:

$$\begin{aligned} P(Y < y) &= P(h(x) < y) \\ &= P(x \leq h^{-1}(y)) \\ &= \int_0^{h^{-1}(y))} f_X(x)dx. \end{aligned}$$

Using Leibniz's Rule:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\int_0^{h^{-1}(y))} f_X(x)dx \\ &= f_X(h^{-1}(y))\frac{d}{dy}h^{-1}(y)) - f_X(0)\frac{d}{dy}(0) + 0 \\ &= f_X(h^{-1}(y))\frac{d}{dy}h^{-1}(y)) \\ &= \frac{y^{\alpha-1}e^{-y}h'h^{-1}(y))}{\Gamma(\alpha)}\frac{d}{dy}h^{-1}(y)) \\ &= \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}. \end{aligned}$$

Hence, we have $Y \sim \gamma(\alpha, 1)$.

(c) Refer to R Jupyter Notebook Output.

In [1]:

```
library(MASS)
data(quine)
```

In [2]:

```
p.hat = function(k, r=1.5) {
    return(sum(k) / (length(k) * r + sum(k)))
}
```

Q1a) Answer:

$\hat{p}$ = 0.916475972540046

In [3]:

```
p.post = function(k, r=1.5, a=0.5, b=0.5) {
    return (c(sum(k) + a, length(k) * r + b))
}
```

In [4]:

```
p.post(quine$Days)
```

2403.5 219.5

Q1b) Answer:

α = 2403.5, β = 219.5

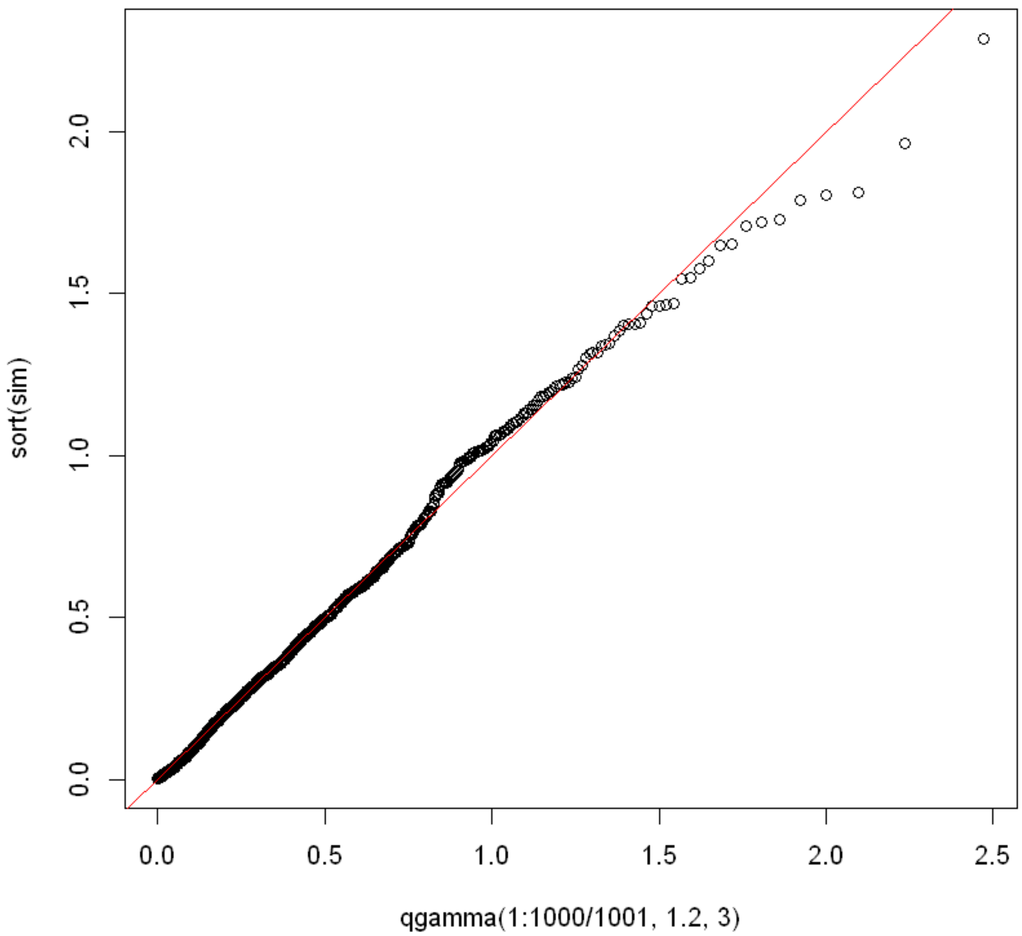
In [16]:

```
h1 = function(x, c, d) {
    return (d * (1 + c * x)^3)
}

g = function(x, c, d) {
    return (d * log((1 + c * x) ^ 3) - d * (1 + c * x)^3 + d)
}

f1 = function(x, alpha) {
    d = alpha - 1/3
    c = 1 / sqrt(9 * d)

    h1.prime = function(x, c, d) {
        return (3 * d * c * (1 + c * x)^2)
    }

    return (h1(x, c, d)^(alpha - 1) * exp(-h1(x, c, d)) * h1.prime(x, c, d))
}

f2 = function(x, alpha) {
    d = alpha - 1/3
    c = 1 / sqrt(9 * d)
    return (exp(g(x, c, d)))
}

cmp = function(alpha) {
    return (f1(1, alpha) / exp(f2(1, alpha)))
}

h2 = function(x) {
    return (exp(-x^2 / 2))
}

gamma = function(alpha = 1, beta = 1) {
    d = alpha - 1/3;
    c = 1 / sqrt(9 * d)
    r = cmp(alpha)

    while (TRUE) {
        x = rnorm(1)
        y = runif(1)

        if (x > -1/c && y < f1(x, alpha) / h2(x) * r) {
            break
        }
    }

    return (h1(x, c, d) / beta)
}

rgamma_ = function(n_iter, alpha = 1, beta = 1) {
    return (replicate(n_iter, gamma(alpha, beta)))
}
```

Q2c) Answer:

In [18]:

```
sim = rgamma_(1000, 1.2, 3)
plot(qgamma(1:1000/1001, 1.2, 3), sort(sim))
abline(0, 1, col="red")
```



In [ ]: