Tutorial 7

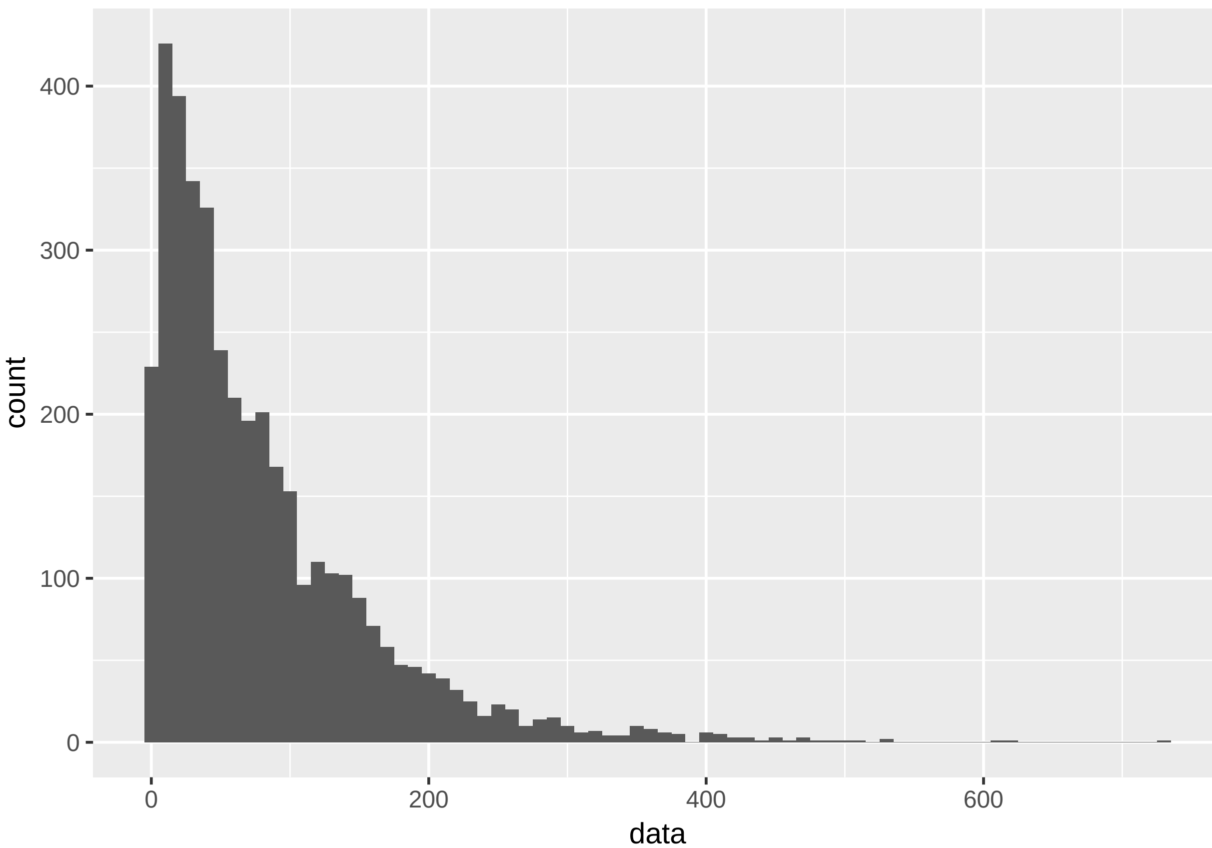
Rui Qiu

2018-04-22

Q1

a

```r
library(ggplot2)
data <- read.csv('gamma-arrivals.txt',header = FALSE)
qplot(data, geom="histogram", binwidth = 10)
```



b

$$
\begin{aligned}
E[X_i] &= ab = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \\
V[X_i] &= E[X_i - \mu]^2 = ab^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
a &= \frac{\bar{X}}{b} \\
ab^2 &= \left(\frac{\bar{X}}{b}\right) b^2 = \bar{X}b = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
\hat{b}_{MM} &= \frac{1}{n\bar{X}}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
\hat{a}_{MM} &= \frac{\bar{X}}{\hat{b}_{MM}}
\end{aligned}
$$

```
x <- unlist(data, use.names=FALSE)
xbar <- mean(x)
N <- length(x)
ss <- sum((x-mean(x))^2)
(b.hat <- ss / N / xbar)
```

```
## [1] 78.95989
```

```
(a.hat <- xbar / b.hat)
```

```
## [1] 1.012352
```

$\Gamma(1.012352, 78.95989)$ ### c

Can use `fitdistr()` function in `MASS` package to hack this problem.

```
library(MASS)
(mle.est <- fitdistr(x, "gamma", start=list(shape=1, rate=1))$estimate)
```

```
##      shape       rate
## 1.02717388 0.01285847
```

```
1/mle.est[2]
```

```
##     rate
## 77.76975
```

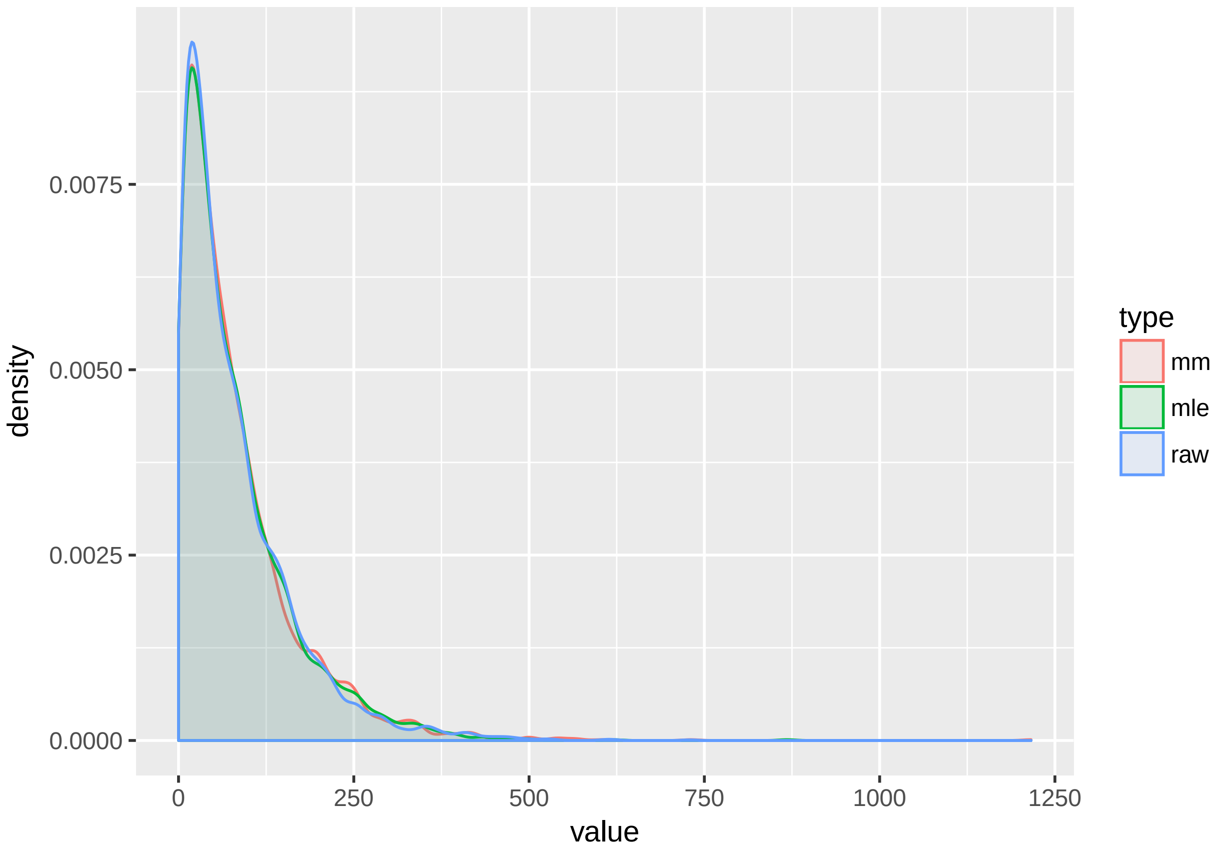
$\Gamma(1.02717388, 77.76975)$

d

Instead of plotting histogram of raw data, we'd like to plot the density of raw data instead. In this way, the comparison is more evident.

```
mm.den <- rgamma(3000, shape=1.012352, scale=78.95989)
mle.den <- rgamma(3000, shape=1.02717388, scale=77.76975)
mm.table <- data.frame(type = rep("mm", 3000), value = mm.den)
mle.table <- data.frame(type = rep("mle", 3000), value = mle.den)
raw.table <- data.frame(type = rep("raw", N), value = x)
```

```r
df <- data.frame(rbind(mm.table, mle.table, raw.table))
ggplot(df, aes(value, colour=type, fill=type)) + geom_density(alpha=0.1)
```



Q2

a

$$
\begin{aligned}
L(u,v,w;p,q) &= \frac{(u+v+w)!}{u!v!w!} \cdot (pq + (1-p)q^2)^u \\
&= \cdot (p(1-q) + (1-p)(1-q)^2)^v \cdot ((1-p)2q(1-q))^w \\
l(u,v,w;p,q) &= \log(u+v+w)! - \log u! - \log v! - \log w! \\
&+ u\log(pq + (1-p)q^2) + v\log(p(1-q) + (1-p)(1-q)^2) \\
&+ w\log((1-p)2q(1-q))
\end{aligned}
$$

b

E-step

Compute

$$
Q(p,q;p^{(t)},q^{(t)}) = E_{p^{(t)},q^{(t)}}[l(p,q;u,v,w)|u_{\text{obs}}, v_{\text{obs}}, w_{\text{obs}}]
$$

M-step

Find $p^{(t+1)}, q^{(t+1)}$ such that

$$Q(p^{(t+1)}, q^{(t+1)}; p^{(t)}, q^{(t)}) \geq Q(p, q; p^{(t)}, q^{(t)})$$

Repeat E-step and M-step until $L(p^{(t+1)}, q^{(t+1)}) - L(p^{(t)}, q^{(t)}) \leq \delta$ where δ is a small amount as a threshold.

c

I don't know how to implement this in R.

Q3

a

$$\begin{aligned}\frac{L(\theta_0; x)}{L(\theta_1; x)} &= \frac{\theta_0^x \cdot e^{-\theta_0}/x!}{\theta_1^x \cdot e^{-\theta_1}/x!} \\ &= \left(\frac{\theta_0}{\theta_1}\right)^x \cdot \exp\left(\theta_1 - \theta_0\right) \leq A\end{aligned}$$

where A is a constant.

Also, as $\theta_1 > \theta_0$, both parameters are greater than 1, then

$$x \leq \log_{\theta_0/\theta_1}(A \cdot \exp\left(\theta_0 - \theta_1\right)) = A^*$$

By definition,

$$\alpha = P(X < A^* | \theta = \theta_0) = \int_0^{A^*} \frac{\theta_0^x \cdot e^{-\theta_0}}{x!} dx = 0.05$$

Then we solve for x.

b

$$\begin{aligned}\frac{L(\theta_0; x)}{L(\theta_1; x)} &= \frac{\frac{1}{\theta_0} e^{-x/\theta_0}}{\frac{1}{\theta_1} e^{-x/\theta_1}} = \frac{\theta_1}{\theta_0} e^{x/\theta_1 - x/\theta_0} \leq A \\ e^{x/\theta_1 - x/\theta_0} &\leq A \cdot \frac{\theta_0}{\theta_1} \\ x/\theta_1 - x/\theta_0 &\leq \ln(A \cdot \theta_0/\theta_1) \\ \frac{\theta_0 - \theta_1}{\theta_0\theta_1} x &\leq \ln(A \cdot \theta_0/\theta_1) \\ x &\geq \ln(A \cdot \frac{\theta_0}{\theta_1}) \cdot \frac{\theta_0\theta_1}{\theta_0 - \theta_1} = A^*\end{aligned}$$

Also we have

$$\alpha = P(X < A^* | \theta = \theta_0) = \int_0^{A^*} \frac{1}{\theta_0} e^{-x/\theta_0} dx = 0.05$$

We solve for x.