Irregular Maximum Likelihood Estimation

We look at several examples in which a MLE is not found by taking derivatives and the usual large-sample theory for MLEs is not applicable.

Example 1: $X_1, X_2, \ldots, X_n$ is a Random Sample from UNIF($-\theta$, θ).

$$L(\theta) = (1/2\theta)^n, \qquad \text{for } -\theta \le x_1, \ldots, x_n \le \theta$$
$$= 0, \qquad \text{otherwise}$$

Look at the condition $-\theta \le x_1, \ldots, x_n \le \theta$ to see what it says about restrictions on θ in terms of data.

$-\theta \le x_1, \ldots, x_n \le \theta \;\Rightarrow\; x_{(1)} \ge -\theta$ and $x_{(n)} \le \theta \;\Rightarrow\; -x_{(1)} \le \theta$ and $x_{(n)} \le \theta$

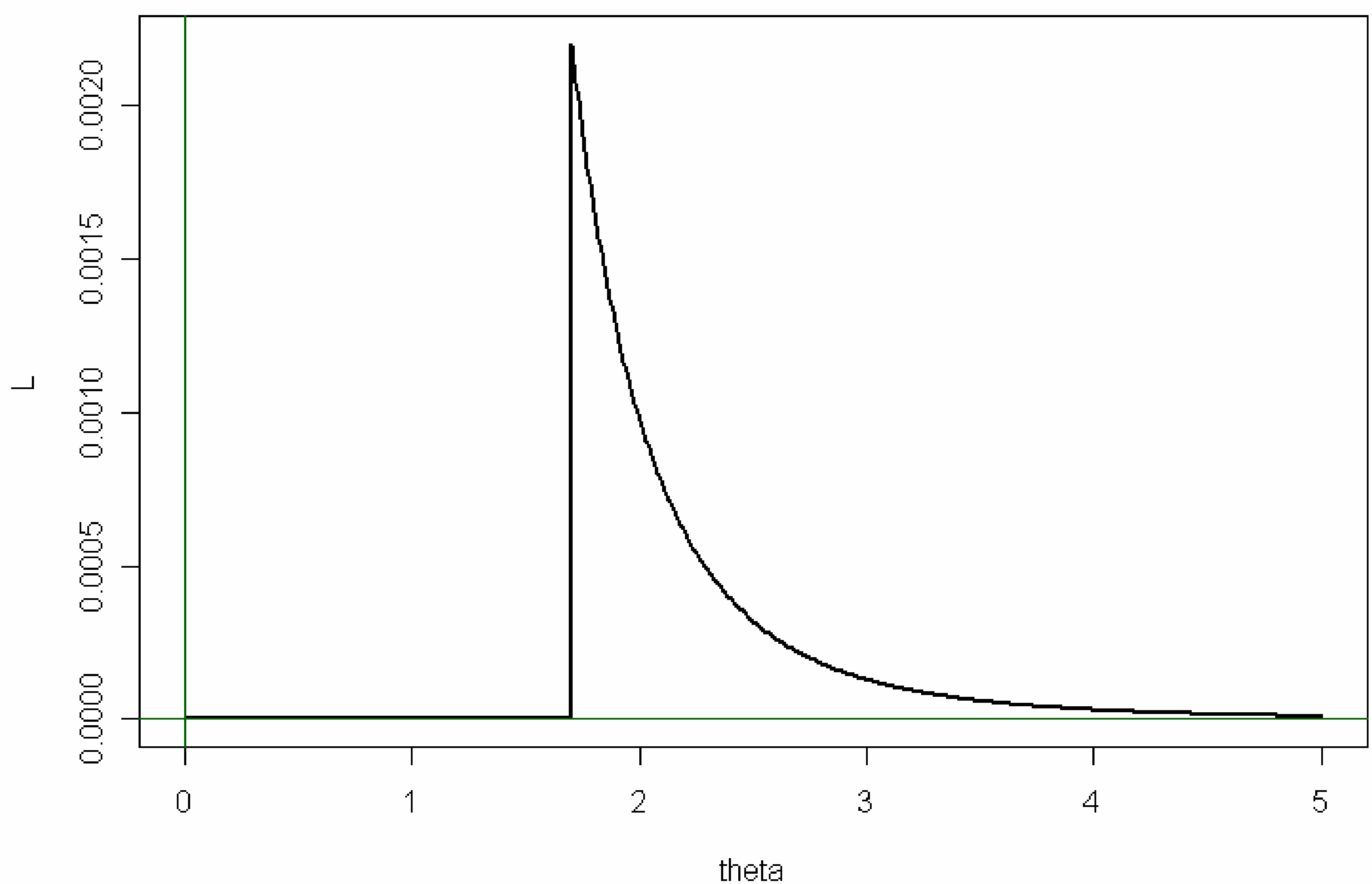
Denote $Y = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$. Then the condition is equivalent to $\theta \ge Y$.

$$L(\theta) = (1/2\theta)^n, \qquad \text{for } \theta \ge Y,$$
$$= 0, \qquad \text{otherwise}$$

Maximum value of $L(\theta)$ occurs where θ is as small as possible. So MLE of θ is Y.

Make a graph for $n = 5$. Use data from UNIF(-2, 2) for an illustration.

```
n = 5;  x = runif(n, -2, 2);  y = max(abs(x))
theta = seq(0, 5, by=.01);  L = (2*theta)^-n;  L[theta < y] = 0
plot(theta, L, type="l", lwd=2)
abline(v = 0, col="darkgreen"); abline(h = 0, col="darkgreen")
x;  abs(x);  y
> x;  abs(x);  y
[1]  0.2994767  1.1443827 -1.6977629  0.2554094 -0.9491573
[1] 0.2994767 1.1443827 1.6977629 0.2554094 0.9491573
                                [1] 1.697763  # MLE
```



Properties of this MLE. Use simulation to see how good the MLE is; compare MLE with MME.

Facts: MLE = Y is biased (always too small). That is, $E(Y) < \theta$. But $U = [(n + 1)/n]Y$ is unbiased. Based on second moments, MME is $Q = [3 \Sigma X_i^2/n]^{1/2}$. Also biased, because of taking square root.

Among ***unbiased*** estimators it is reasonable to use the variance as a criterion of 'goodness,' picking the estimator with smallest variance as best. (In simulation we use SD to preserve units/scale.)

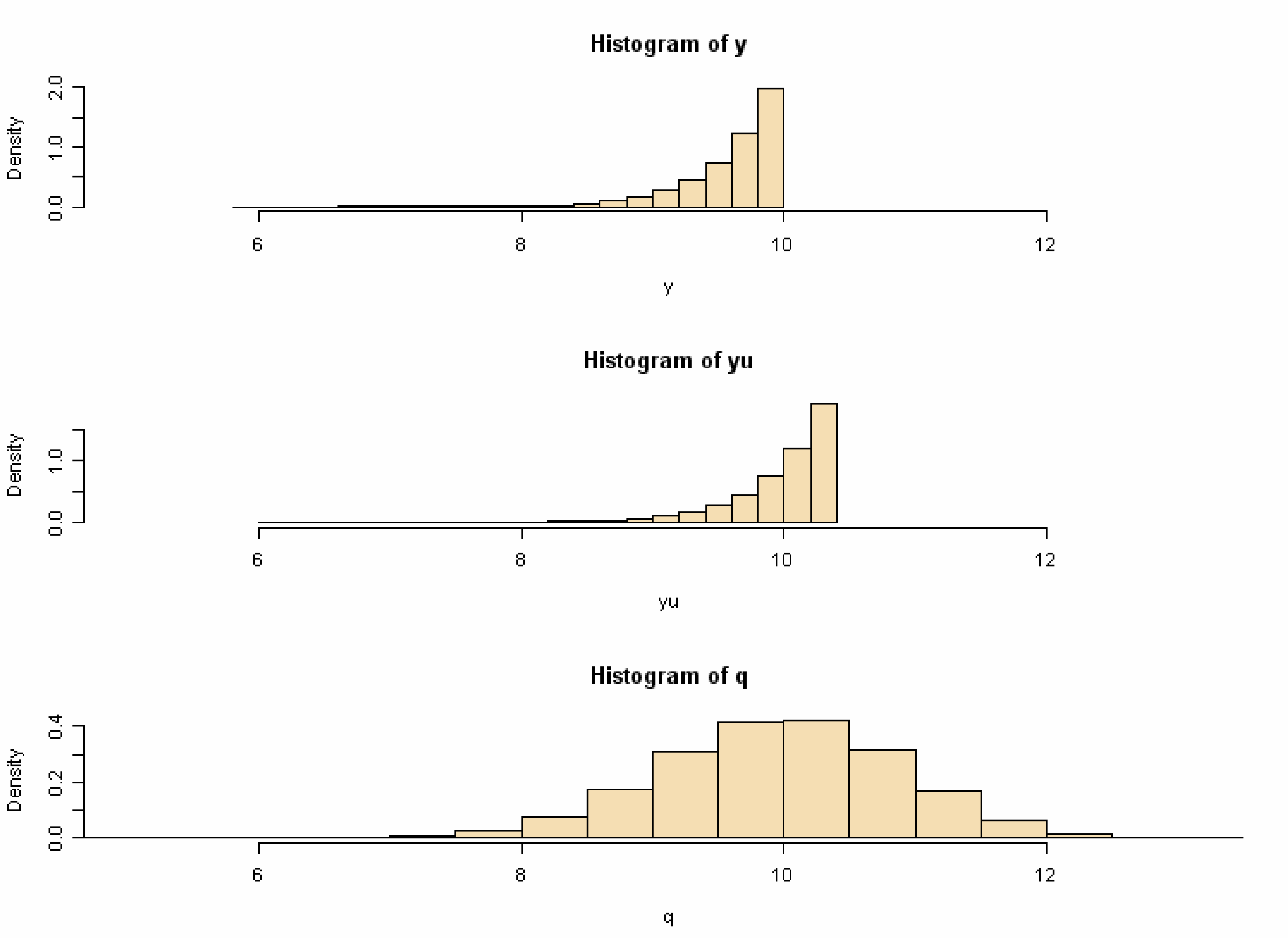
A reasonable way to compare ***biased*** estimators is 'mean square error': for example, $\text{MSE}_Y = E[(Y - \theta)^2]$. MSE = Variance + $(\text{Bias})^2$, so for unbiased estimators MSE and Variance are the same.

We explore, using simulation with $n = 12$ and $\theta = 10$.

```
m = 100000; n = 25;  th = 10
DTA = matrix(runif(m*n, -th, th), nrow=m)     # m samples, each with n obs.
y = apply(abs(DTA), 1, max);  yu = ((n+1)/n)*y  # y is vector of m MLEs
q = sqrt(3*rowMeans(DTA^2))                   # q is vector on m MMEs
y.desc = c(mean(y),  sd(y),  sqrt(mean((y-th)^2)))
yu.desc = c(mean(yu),  sd(yu),  sqrt(mean((yu-th)^2)))  # Sim. E, SD √MSE
q.desc = c(mean(q),  sd(q),  sqrt(mean((q-th)^2)))
round(rbind(y.desc, yu.desc, q.desc), 3)

lw = min(y, yu, qu);  up = max(y, yu, qu)     # to put hist's in same int.
par(mfrow=c(3,1))                             # 3 graphs on a 'page'
  hist(y, prob=T, xlim=c(lw, up), col="wheat")    # dist'n of MLEs
  hist(yu, prob=T, xlim=c(lw, up), col="wheat")   # dist'n of unb. MLEs
  hist(q, prob=T, xlim=c(lw, up), col="wheat")    # dist'n of MMEs
par(mfrow=c(1,1))
```

```
> round(rbind(y.desc, yu.desc, q.desc), 3)
              E     SD   √MSE
y.desc    9.613 0.371 0.535
yu.desc   9.998 0.385 0.385  # lowest sqrt(mse)  Var(Yunb) = MSE(Yunb)
q.desc    9.958 0.903 0.904  # bias is very small
```



Distribution of Y does not converge to normal with increasing n.

Example 2: X1, X2, ..., Xn is a Random Sample from UNIF($\mu - \theta_0$, $\mu + \theta_0$), where θ_0 is known.

$$L(\mu) = (1/2\theta_0)^n, \quad \text{for } \mu - \theta_0 < x_1, \ldots, x_n < \mu + \theta_0$$
$$= 0, \quad \text{otherwise}$$

Look at $\mu - \theta_0 < x_1, \ldots, x_n < \mu + \theta_0$ to see what it says about restrictions on θ in terms of data.

$\mu - \theta_0 < x_1, \ldots, x_n < \mu + \theta_0 \;\Rightarrow\; x_{(1)} > \mu - \theta_0$ and $x_{(n)} < \mu + \theta_0 \;\Rightarrow\; x_{(n)} - \theta_0 < \mu < x_{(n)} + \theta_0$

Then the likelihood function becomes a ***constant*** (no μ) over the interval $(x_{(n)} - \theta_0,\ x_{(n)} + \theta_0)$.

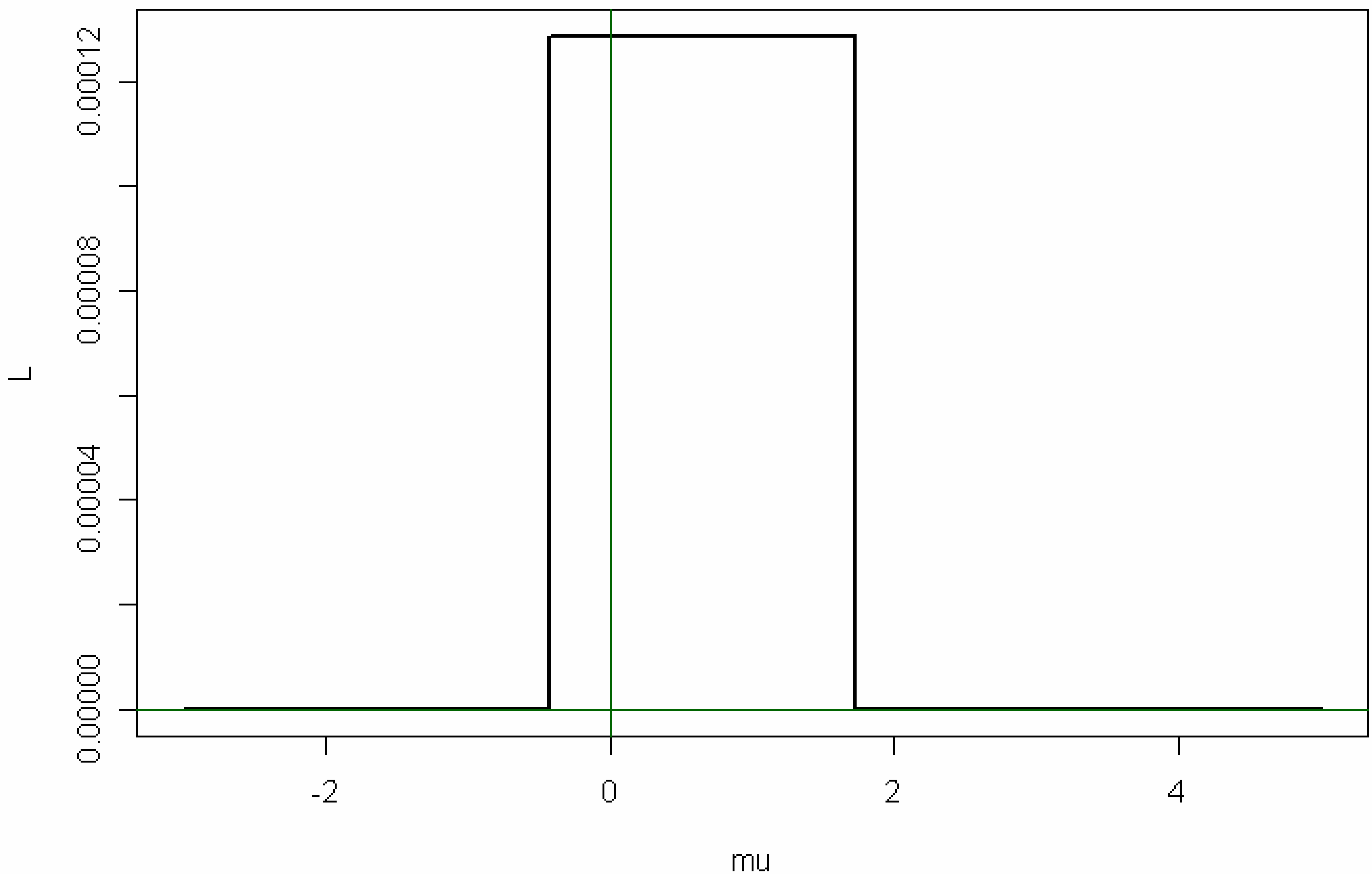
$$L(\mu) = (1/2n)^n, \quad \text{for } x_{(n)} - \theta_0 < \mu < x_{(n)} + \theta_0,$$
$$= 0, \quad \text{otherwise}$$

The MLE is not unique because $L(\mu)$ has its maximum value anywhere in $(x_{(n)} - \theta_0,\ x_{(n)} + \theta_0)$.

Make a graph for $n = 5$. Let $\theta_0 = 3$. Use data from UNIF($1 - 3 = -2$, $1 + 3 = 4$) for an illustration.

```
n = 5;  x = runif(n, -2, 4);  x.min = min(x);  x.max = max(x)
mu = seq(-3, 5, by=.01);  L = (mu^0*2*3)^-n; L  # trick so L is 'fcn' of mu
L[(mu < x.max-3)|(mu > x.min+3)] = 0  # set to 0 outside interval
plot(mu, L, type="l", lwd=2)
abline(v = 0, col="darkgreen"); abline(h = 0, col="darkgreen")
x; x.min; x.max; x.max-3; x.min+3
```

```
> x; x.min; x.max; x.max-3; x.min+3
[1]  0.6786577  1.3466360 -1.2723000  2.5639691  1.8075686
[1] -1.2723
[1] 2.563969
[1] -0.4360309  # lower end of interval of MLEs
[1] 1.7277      # upper end of interval of MLEs  Interval includes μ = 1.
```


0.00012
0.00008
0.00004
0.00000
L
-2
0
2
4
mu

Properties of this MLE. Because the MLE is not unique, we try using the midpoint of the interval of possible values. This is the average of the max and the min, usually called the ***midrange***.

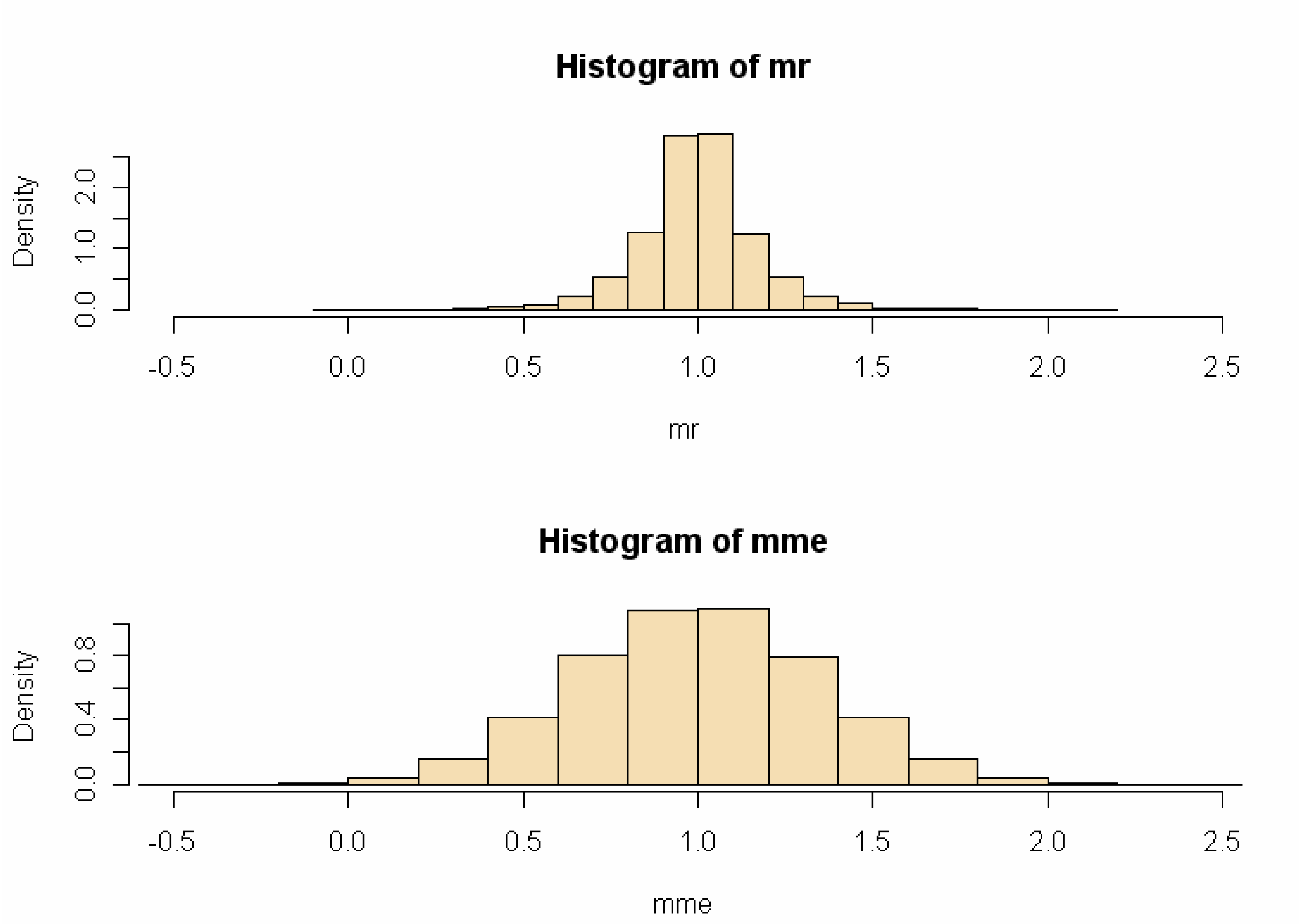
The interval of MLEs is a 100% confidence interval for μ. (You don't see many useful 100% CIs!) The program ($\theta_0 = 3$ and $n = 25$) shows that the average length of this CI is about 0.46.

The MME is the sample mean. Both the MLE and the MME are unbiased.

```
m = 100000; n = 25;  mu = 1;  th.0 = 3.
DTA = matrix(runif(m*n, mu-th.0, mu+th.0), nrow=m)
mx = apply(DTA, 1, max); mn = apply(DTA, 1, min)
mr = (mx + mn)/2;  len.int = mn+th.0 - mx+th.0    # midrange (MLE)
mme = rowMeans(DTA)                               # sample mean (MME)
mr.desc = c(mean(mr),  sd(mr),  sqrt(mean((mr-mu)^2)))
mme.desc = c(mean(mme),  sd(mme),  sqrt(mean((mme-mu)^2)))
round(rbind(mr.desc, mme.desc), 4)
mean(len.int); mean((mu < mn+th.0) & (mu > mx-th.0))

lw = min(mr,mme);  up = max(mr,mme)
par(mfrow=c(2,1))
  hist(mr, prob=T, xlim=c(lw, up), col="wheat")
  hist(mme, prob=T, xlim=c(lw, up), col="wheat")
par(mfrow=c(1,1))
```

```
> round(rbind(mr.desc, mme.desc), 4)
               E      SD    √MSE
mr.desc  0.9996 0.1599 0.1599
mme.desc 1.0001 0.3458 0.3458
> mean(len.int); mean((mu < mn+th.0) & (mu > mx-th.0))
[1] 0.4616141   # Length of MLE interval
[1] 1           # MLE interval is 100% CI
```



The distribution of the midrange does not converge to normal.

Example 3: $X_1, X_2, \ldots, X_n$ is a Random Sample the two-parameter exponential distribution EXP(θ, η).

The density function is $f(x \mid \theta, \eta) = (1/\theta) \exp[-(x - \eta)/\theta]$, for $x \ge \eta$ (and 0 otherwise).
Estimation of η is irregular, but estimation of θ is standard. Begin by estimating η for fixed $\theta = \theta_0$.
We show that the MLE of the 'delay' η is $x_{(1)}$: The likelihood function is

$$L(\theta_0, \eta) = (1/\theta_0)^n \exp[-\Sigma_i (x_i - \eta)/\theta_0], \quad \text{for } x_1, \ldots, x_n \ge \eta \text{ or } x_{(1)} \ge \eta$$
$$= 0, \quad \text{otherwise.}$$

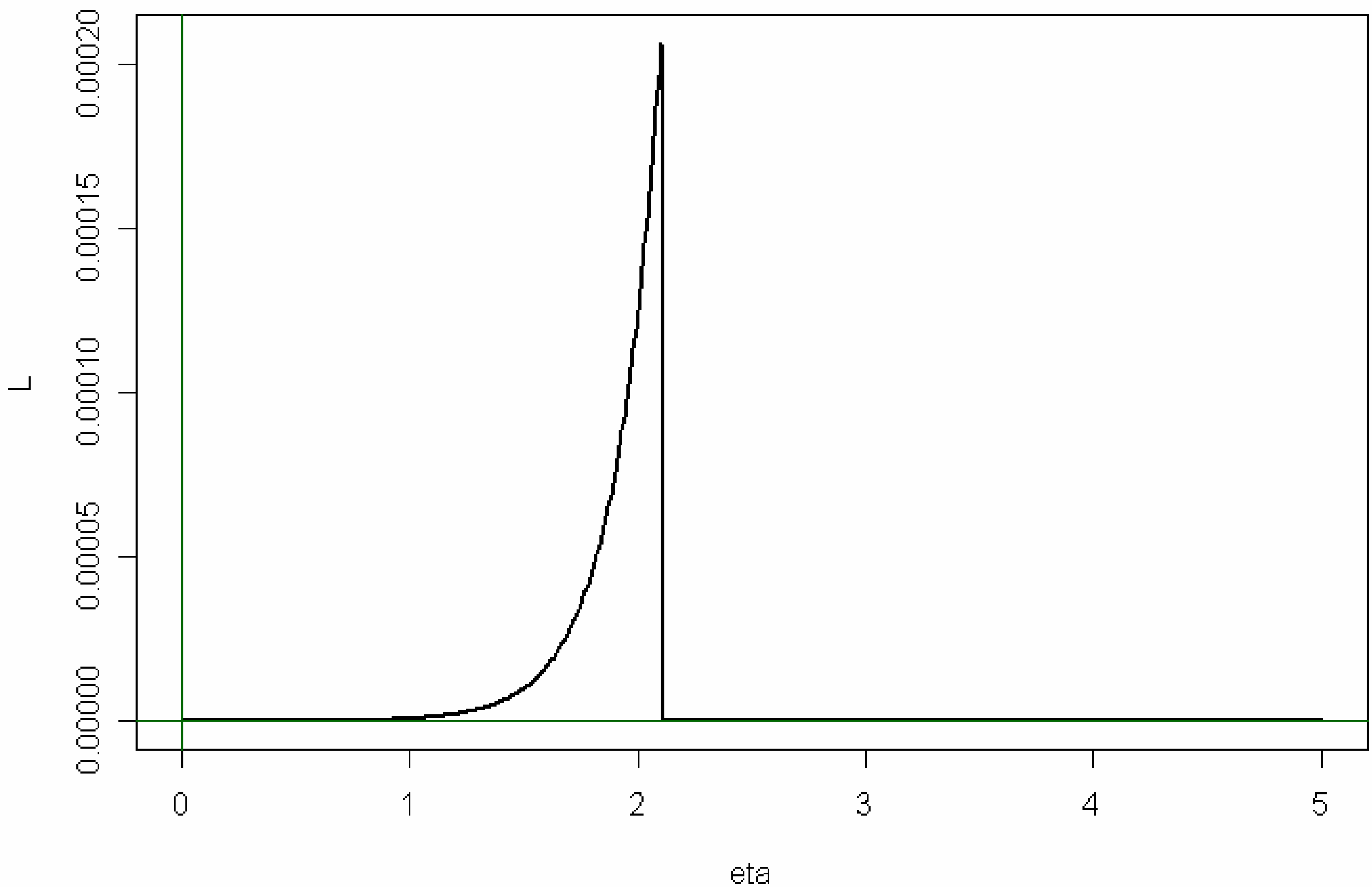
In the exponent, $\Sigma_i (x_i - \eta) = \Sigma_i x_i - n\eta$, where the sums run through $i = 1, \ldots, n$.
So $\exp[-\Sigma_i (x_i - \eta)/\theta_0] = \exp[-(\Sigma_i x_i)/\theta_0] \times \exp[(n/\theta_0)\eta]$ and

$$L(\theta_0, \eta) = (1/\theta_0)^n \exp[-(\Sigma_i x_i)/\theta_0]\, \mathbf{exp[(\boldsymbol{n}/\theta_0)\eta]}, \quad \text{for } \eta \le x_{(1)}$$
$$= 0, \quad \text{otherwise.}$$

For given data and fixed θ_0, the factor in blue is constant. The factor **$\exp[(n/\theta_0)\eta]$** increases with increasing η, until η reaches $x_{(1)}$, at which point it reaches its maximum.
Make a graph for $n = 5$. Use data from a population with $\theta_0 = 1$ and $\eta = 2$ for illustration.

```
n = 5;   x = rexp(n, rate=1/1)+2;   x.min = min(x);   eta = seq(0,5,by=.01)
L = (1/1)^n * exp(-sum(x)/1) * exp((5/1)*eta);  L[eta > x.min] = 0
plot(eta, L, type="l", lwd=2)
abline(v = 0, col="darkgreen"); abline(h = 0, col="darkgreen")
x; x.min; (1/1)^n * exp(-sum(x)/1)
```

```
> x; x.min; (1/1)^n * exp(-sum(x)/1)
[1] 3.881629 4.025116 2.108118 2.916504 6.053171
[1] 2.108118  # Minimum is MLE of η
[1] 5.690098e-09   # y-intercept of curve (very small here, but NOT 0)
```



Note: Once we have $x_{(1)}$ as the MLE of η, we can substitute that value into L to get $L(\theta)$. Setting the derivative $L'(\theta) = 0$, we obtain the MLE of θ.

Properties of the minimum of the data as an estimate of the delay η.

For the usual one-parameter exponential, the sample mean and standard deviation both estimate θ. The SD is not influenced by the delay. So a possible alternative estimate of η might be the sample mean minus the SD. This alternative estimate is slightly biased (because SD is).

In what follows, we take $n = 10$, $\theta_0 = 1$, $\eta = 2$.

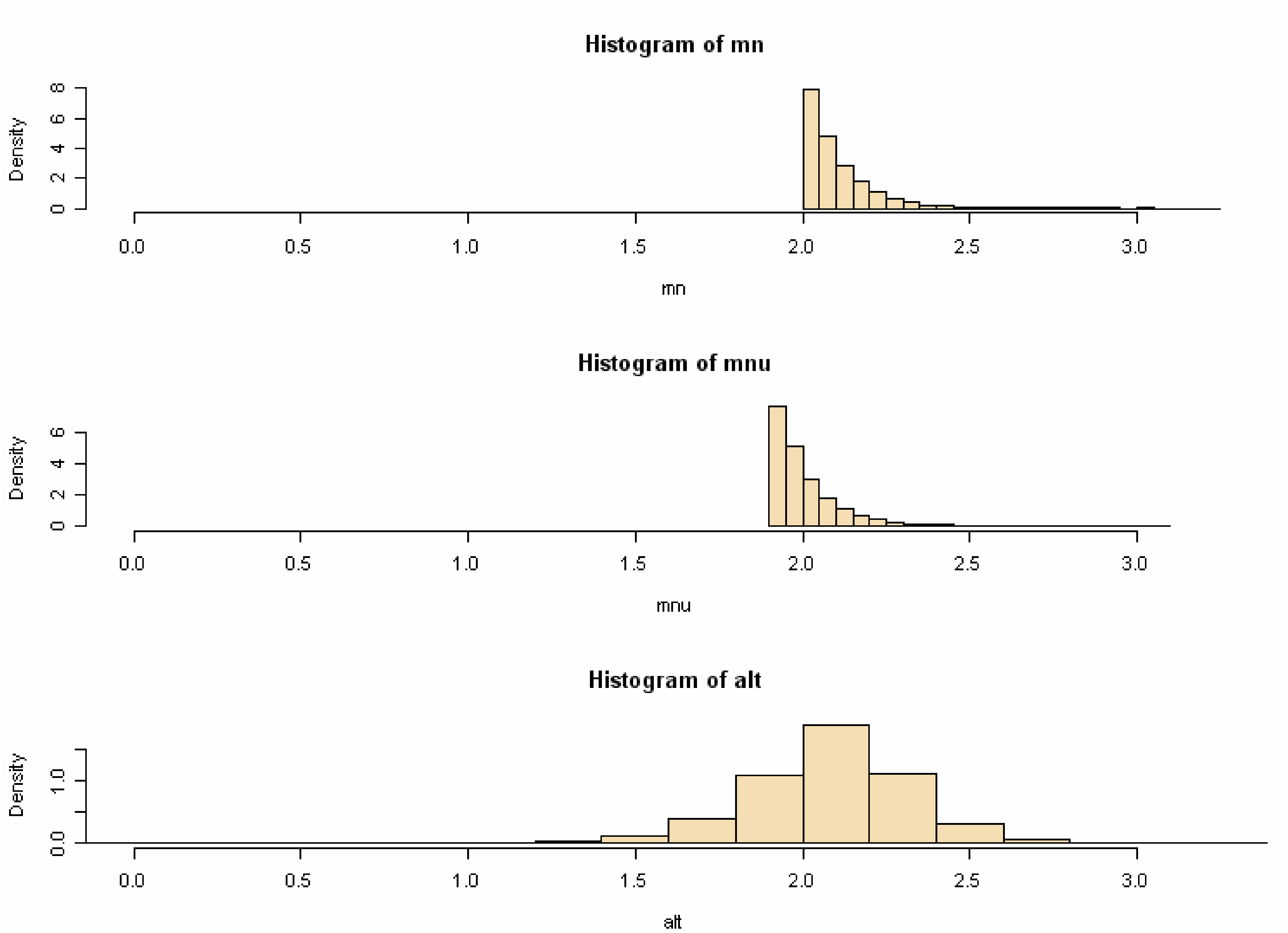
```
m = 100000; n = 10;  eta = 2;  th.0 = 1.
DTA = matrix(rexp(m*n, 1)+2, nrow=m)                # I didn't do analysis.
mn = apply(DTA, 1, min);  mnu = mn * 2*n/(2*n+1) #  I guess it's correct
x.bar = rowMeans(DTA)                               #  factor to unbias MLE
alt = x.bar - apply(DTA, 1, sd)
mn.desc = c(mean(mn),  sd(mn),  sqrt(mean((mn-eta)^2)))
mnu.desc = c(mean(mnu),  sd(mnu),  sqrt(mean((mnu-eta)^2)))
alt.desc = c(mean(alt),  sd(alt),  sqrt(mean((alt-eta)^2)))
round(rbind(mn.desc, mnu.desc, alt.desc), 4)

lw = min(mn, mnu,alt);  up = max(mn, mnu ,alt)
par(mfrow=c(3,1))
  hist(mn, prob=T, xlim=c(lw, up), col="wheat")
  hist(mnu, prob=T, xlim=c(lw, up), col="wheat")
  hist(alt, prob=T, xlim=c(lw, up), col="wheat")
par(mfrow=c(1,1))

> round(rbind(mn.desc, mnu.desc, alt.desc), 4)
```

	E	SD	√MSE
mn.desc	2.0997	0.0999	0.1411
mnu.desc	1.9997	0.0952	0.0952
alt.desc	2.0757	0.2500	0.2612

Suppose we could 'unbias' the alternative estimate. Then its SD and √MSE would both be about 0.25, still much larger than 0.0952.



Note: With $n = 10$: very long left tail for alternate method (occasional negative values).