Estimating θ with *n* = 5 Observations from UNIF(0, θ)

Reasonable choices to estimate θ might be to double the sample mean, median, or midrange, or to use 1.2 max. All are unbiased. Which is most accurate (least variable)? Can prove it is 1.2 max. (UMVUE). Exact distributions (based on BETA) are easy to derive for median (*n* odd) and max. We could use the CLT to get the approximate normal distribution for sample mean.

```
m = 50000; n = 5;  th = 10         # arbitrary value of parameter theta
x = runif(m*n, 0, th)
DTA = matrix(x, nrow=m)            # 50K rows: each a sample of size 5
A = 2 * rowMeans(DTA)              # double the mean
Mx = apply(DTA, 1, max);  Mn = apply(DTA, 1, min)
B = ((n+1)/n) * Mx                 # unbiased max
C = Mx + Mn                        # double the midrange
H = 2 * apply(DTA, 1, median)      # double the median

# Comparable histograms
lr = c(0, 2*th)                    # left and right extent of horiz scale
xx = seq(0, 2*th, length=500)      # x-values for plotting density curves
par(mfrow=c(2,2))
  hist(H, breaks=20, prob=T, xlim=lr, col="wheat")
    k = th*2                       # Midrange/th ~ BETA((n+1)/2, (n+1)/2), odd n
    if((n%%2)==1) lines(xx, dbeta(xx/k, (n+1)/2, (n+1)/2)/k, col="blue")
  hist(A, breaks=20, prob=T, xlim=lr, col="wheat")
    mu = th;  sg = 2*th/sqrt(12*n) # A aprx NORM(mu, sg)
    lines(xx, dnorm(xx, mu, sg), col="blue", lty="dashed")
  hist(C, breaks=20, prob=T, xlim=lr, col="wheat")
  hist(B, breaks=20, prob=T, xlim=lr, col="wheat")
    k = th*(n+1)/n                  # Max/th ~ BETA(n, 1)
    lines(xx, dbeta(xx/k, n, 1)/k, col="blue")
par(mfrow=c(1,1))
```

```
# Verify all estimators are unbiased
round(c(mean(H), mean(A), mean(C), mean(B)), 3)

# Which has smallest sd?
round(c(sd(H), sd(A), sd(C), sd(B)), 3)
```

```
> # Verify all estimators are unbiased
> round(c(mean(H), mean(A), mean(C), mean(B)), 3)
[1] 10.002  9.999  9.996  9.988

> # Which has smallest sd? (also smallest var, but use sd to preserve units)
> round(c(sd(H), sd(A), sd(C), sd(B)), 3)
[1]  3.793  2.594  2.198  1.698            # unbiased max is best
```

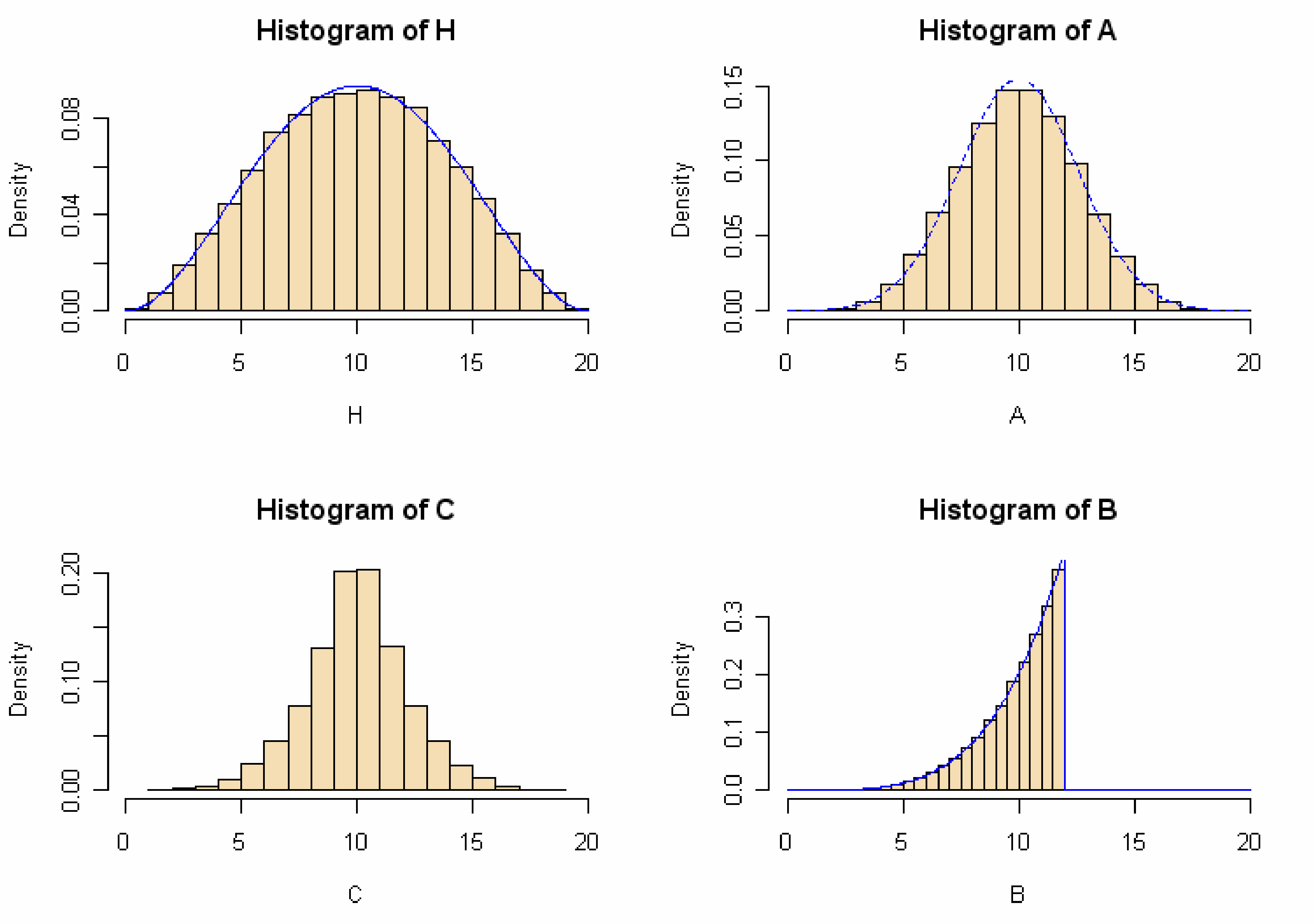
Some computations (n = 5). If $X \sim \text{BETA}(\alpha, \beta)$, then $E(X) = \alpha/(\alpha+\beta)$, $V(X) = \alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$.

Sample max: $X_{(n)}/\theta \sim \text{BETA}(n, 1)$. $\mu = n/(n+1) = 5/6$, $\sigma^2 = n/[(n+2)(n+1)^2] = 5/252$.
$E[(6/5)X_{(5)}] = \theta = 10$; $V[B] = V[(6/5)\,X_{(n)}] = \theta^2(6/5)^2\sigma^2 = 0.\,02857\theta^2$; $SD[B] = 0.1690\theta =$ **1.690**.
Let $U_1, \ldots, U_5$ iid $N(0, 1)$, $Y = \max(U_i)$, and $0 < y < 1$. We see that $Y \sim \text{BETA}(5, 1)$ as follows:
$F_Y(y) = P\{Y \le y\} = P\{U_1 \le y, \ldots U_5 \le \text{y}\} = [P\{U \le y\}]^5 = y^5$, so $f_Y(y) = F_Y'(y) = 5y^4$.

Sample median: n odd; $k = (n+1)/2$. $X_{(k)}/\theta \sim \text{BETA}(k, k)$. $\mu = 1/2$, $\sigma^2 = 1/[4(2k+1)] = 1/28$.
$E[H] = E[2X_{(3)}] = \theta = 10$; $V[H] = V[2X_{(3)}] = 4\theta^2\sigma^2 = (1/7)\theta^2$; $SD(H) =$ **3.780**.

Sample mean: By the CLT, A is only *approximately* normal, but means and variances are exact.
$E(A) = 10$, $V[A] = 2^2(1/60)\theta^2$. $SD[A] = 0.2582\theta =$ **2.582**.

[For the midrange, computation of the SD is more difficult because the min and max are correlated.]


Histogram of H
Density
0.00
0.04
0.08
0
5
10
15
20
H
Histogram of A
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
A
Histogram of C
Density
0.00
0.10
0.20
0
5
10
15
20
C
Histogram of B
Density
0.0
0.1
0.2
0.3
0
5
10
15
20
B

For n = 25: Relative standings of SDs are the same: Unbiased max still has smallest SD. Mean and median tend towards normal distributions with relatively large SD. Distributions of Midrange and unbiased max do not tend toward normal.

```
> round(c(sd(H), sd(A), sd(C), sd(B)), 3)
[1] 1.924 1.153 0.528 0.382
```