Maximum of a set of IID Random Variables

Luca Ballan

April 16, 2018

Uniformly Distributed Random Variables

Continuous Case Given

$$\overline{X} \sim \mathcal{U}(a, b)$$

the variance of $\overline{X}$ is

$$\begin{aligned} E\left[\overline{X}\right] &= \frac{b+a}{2} \\ var\left(\overline{X}\right) &= \frac{1}{12}(b-a)^2 \end{aligned}$$

Discrete Case Given

$$\overline{X} \sim \mathcal{U}^D(a, b)$$

the variance of $\overline{X}$ is

$$\begin{aligned} E\left[\overline{X}\right] &= \frac{b+a}{2} \\ var\left(\overline{X}\right) &= \frac{1}{12}(b-a)(b-a+2) \end{aligned}$$

and when $a = 0$, $var\left(\overline{X}\right) = b(b+2)/12$.

Proof:
for continuous variables

$$\begin{aligned}
var\left(\overline{X}\right) &= E\left[\left(\overline{X} - E\left[\overline{X}\right]\right)^2\right] \\
&= E\left[\left(\overline{X} - \frac{b+a}{2}\right)^2\right] \\
&= \int_a^b \left(x - \frac{b+a}{2}\right)^2 \ p\left(\overline{X} = x\right) \ dx \\
&= \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2}\right)^2 \ dx \\
&= \frac{1}{b-a} \int_{(a-b)/2}^{(b-a)/2} y^2 \ dy \\
&= \frac{1}{b-a} \frac{1}{3} \left[y^3\right]_{(a-b)/2}^{(b-a)/2} \\
&= \frac{1}{12} (b-a)^2
\end{aligned}$$

for discrete variables

$$\begin{aligned}
var\left(\overline{X}\right) &= E\left[\left(\overline{X} - E\left[\overline{X}\right]\right)^2\right] \\
&= \sum_{i=a}^{b} \left(i - \frac{b+a}{2}\right)^2 \ p\left(\overline{X} = i\right) \\
&= \frac{1}{b-a+1} \sum_{i=a}^{b} \left[i^2 + \left(\frac{b+a}{2}\right)^2 - 2i\left(\frac{b+a}{2}\right)\right] \\
&= \frac{1}{b-a+1} \left((b-a+1)\left(\frac{b+a}{2}\right)^2 - 2\left(\frac{b+a}{2}\right)\sum_{i=a}^{b} i + \sum_{i=a}^{b} i^2\right) \\
&= \frac{1}{4}(b+a)^2 + \frac{1}{b-a+1}\left(-2\left(\frac{b+a}{2}\right)\left(\frac{1}{2}(a+b)(b-a+1)\right) + \sum_{i=a}^{b} i^2\right) \\
&= \frac{1}{4}(b+a)^2 - \frac{1}{2}(b+a)^2 + \frac{1}{b-a+1}\left(\sum_{i=a}^{b} i^2\right) \\
&= -\frac{1}{4}(a+b)^2 + \frac{1}{6}\left(2a^2 + 2b^2 + 2ab - a + b\right) \\
&= \frac{1}{12}(b-a)(b-a+2)
\end{aligned}$$

since

$$\begin{aligned}
\sum_{i=a}^{b} i &= \frac{b(b+1)}{2} - \frac{(a-1)a}{2} = \frac{1}{2}(a+b)(b-a+1) \\
\sum_{i=a}^{b} i^2 &= \frac{b(b+1)(2b+1)}{6} - \frac{a(a-1)(2a-1)}{6}
\end{aligned}$$

Maximum of n Uniform Random Variables (continuous case)

Given a set of n i.i.d. uniform random variables $\overline{X}_i$,

$$\overline{X}_i \sim \mathcal{U}\left(a, b\right)$$

let $\overline{Y}$ be the random variable representing the maximum of this set

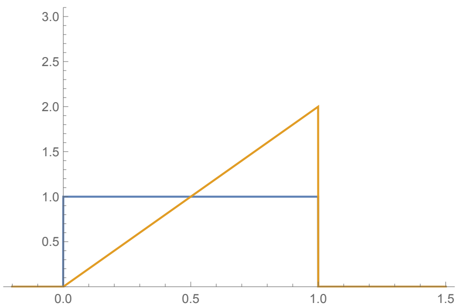
$$\overline{Y} = \max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$$

then

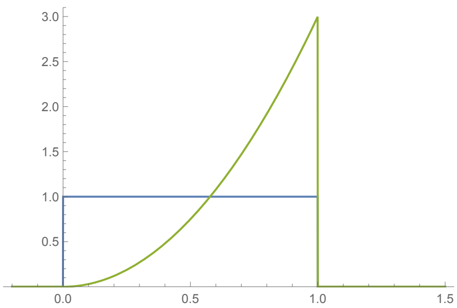
$$E\left[\overline{Y}\right] = \frac{nb + a}{n+1}$$

and

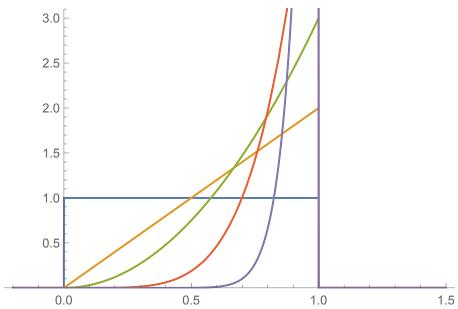
$$p\left(\overline{Y} = y\right) = \begin{cases} 0 & y < a \\ n\left(\frac{1}{b-a}\right)^n \left(y-a\right)^{n-1} & a \leqslant y \leqslant b \\ 0 & b < y \end{cases}$$



$n = 1, 2$



$n = 1, 3$



$n = 1, 2, 3, 6, 15$

Proof: In general the cumulative probability distribution of the max of a sequence of i.i.d. random variables has the form

$$\begin{aligned}
P\left(\overline{Y} \leqslant y\right) &= P\left(\max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\} \leqslant y\right) \\
&= P\left(\left(\overline{X}_1 \leqslant y\right) \wedge \ \ldots \ \wedge \left(\overline{X}_n \leqslant y\right)\right) \\
&= \prod P\left(\overline{X}_i \leqslant y\right)
\end{aligned}$$

if $\overline{X}_i$ is uniform $\mathcal{U}(a, b)$, i.e.

$$P\left(\overline{X}_i \leqslant y\right) = \begin{cases} 0 & y < a \\ \frac{1}{b-a}(y-a) & a \leqslant y \leqslant b \\ 1 & b < y \end{cases}$$

then $P\left(\overline{Y} \leqslant y\right)$ is

$$P\left(\overline{Y} \leqslant y\right) = \begin{cases} 0 & y < a \\ \left[\frac{1}{b-a}(y-a)\right]^n & a \leqslant y \leqslant b \\ 1 & b < y \end{cases}$$

The probability density of $\overline{Y}$ is equal to the derivative of $P\left(\overline{Y} \leqslant y\right)$ with respect to y, which is therefore

$$\begin{aligned}
p\left(\overline{Y} = y\right) &= \frac{\partial}{\partial y} P\left(\overline{Y} \leqslant y\right) \\
&= \begin{cases} 0 & y < a \\ n\left(\frac{1}{b-a}\right)^n (y-a)^{n-1} & a \leqslant y \leqslant b \\ 0 & b < y \end{cases}
\end{aligned}$$

and the expected value is

$$\begin{aligned}
E\left[\overline{Y}\right] &= \int_a^b y \ p\left(\overline{Y} = y\right) \ dy \\
&= \frac{n}{(b-a)^n} \int_a^b y \ (y-a)^{n-1} \ dy \\
&= \frac{n}{(b-a)^n} \left[\frac{(y-a)^n (ny+a)}{n(n+1)}\right]_a^b \\
&= \frac{nb+a}{n+1}
\end{aligned}$$

because:

$$\begin{aligned}
\int y \ (y-a)^{n-1} &= \int (x+a) \ x^{n-1} \\
&= \int x^n + a \ x^{n-1} \\
&= \frac{nx^{n+1} + (n+1) a \ x^n}{n(n+1)} \\
&= \frac{(y-a)^n (ny+a)}{n(n+1)}
\end{aligned}$$

Maximum of n Uniform Random Variables (discrete case)

Given a set of n i.i.d. uniform discrete random variables $\overline{X}_i$,

$$\overline{X}_i \sim \mathcal{U}^D(a,b)$$

let $\overline{Y}$ be the random variable representing the maximum of this set

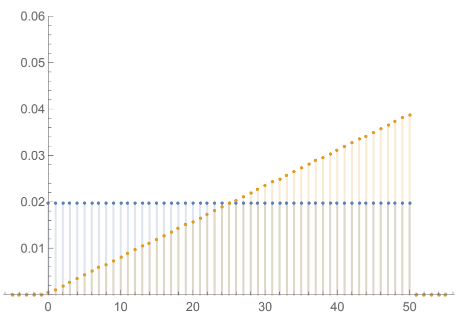
$$\overline{Y} = \max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$$

then

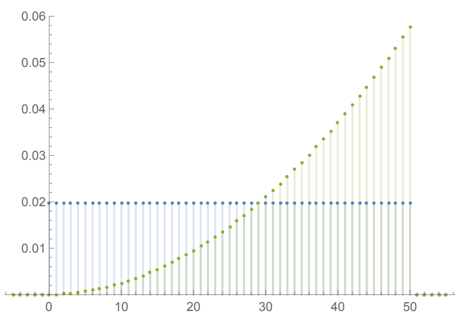
$$E\left[\overline{Y}\right] = \frac{1}{(b-a+1)^n}\left[\sum_{y=a}^{b} y\,(y-a+1)^n - \sum_{y=a}^{b} y\,(y-a)^n\right]$$

and

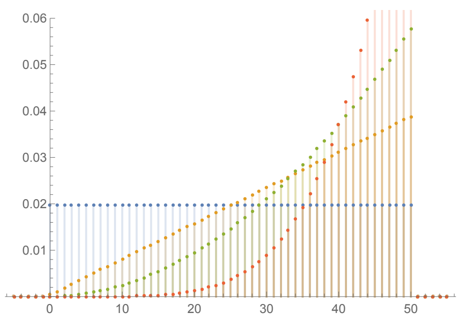
$$p\left(\overline{Y}=y\right) = \begin{cases} 0 & y < a \\ \frac{1}{(b-a+1)^n}\left[(y-a+1)^n - (y-a)^n\right] & a \leqslant y \leqslant b \\ 0 & b < y \end{cases}$$



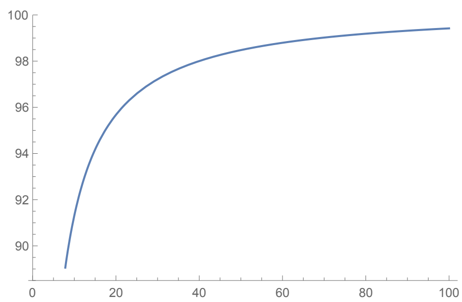
$n = 1, 2$



$n = 1, 3$



$n = 1, 2, 3, 6$



$E\left[\overline{Y}\right]$ for $a = 0,\ b = 100$ and varying n

Proof: in this case, $P\left(\overline{X}_i \leqslant y\right)$ is

$$P\left(\overline{X}_i \leqslant y\right) = \begin{cases} 0 & y < a \\ \frac{1}{b-a+1}\left(y-a+1\right) & a \leqslant y \leqslant b \\ 1 & b < y \end{cases}$$

therefore $P\left(\overline{Y} \leqslant y\right)$ is

$$P\left(\overline{Y} \leqslant y\right) = \begin{cases} 0 & y < a \\ \left[\frac{1}{b-a+1}\left(y-a+1\right)\right]^n & a \leqslant y \leqslant b \\ 1 & b < y \end{cases}$$

The probability density of $\overline{Y}$ is in this case equal to

$$\begin{aligned} p\left(\overline{Y} = y\right) &= P\left(\overline{Y} \leqslant y\right) - P\left(\overline{Y} \leqslant y-1\right) \\ &= \begin{cases} 0 & y < a \\ \frac{1}{(b-a+1)^n}\left[(y-a+1)^n - (y-a)^n\right] & a \leqslant y \leqslant b \\ 0 & b < y \end{cases} \end{aligned}$$

and the expected value is

$$\begin{aligned} E\left[\overline{Y}\right] &= \sum y\, p\left(\overline{Y} = y\right) \\ &= \frac{1}{(b-a+1)^n}\left[\sum_{y=a}^{b} y\left(y-a+1\right)^n - \sum_{y=a}^{b} y\left(y-a\right)^n\right] \end{aligned}$$

Unfortunately, this series is not easy to simplify to an analytical form.
Numerically $E\left[\overline{Y}\right]$,

$$\begin{array}{lcl} a=0,\ b=100,\ n=2 & \longrightarrow & \sim 66.8 \\ a=0,\ b=100,\ n=3 & \longrightarrow & \sim 75.2 \\ a=0,\ b=100,\ n=5 & \longrightarrow & \sim 83.7 \\ a=0,\ b=100,\ n=10 & \longrightarrow & \sim 91.3 \\ a=0,\ b=100,\ n=20 & \longrightarrow & \sim 95.7 \end{array}$$

Maximum of n Uniform Random Variables (discrete with no repetitions)

Given a set of n uniform discrete random variables $\overline{X}_i$,

$$\overline{X}_i \sim \mathcal{U}^D(a, b)$$

such as

$$P\left(\overline{X}_i = \overline{X}_j\right) = 0 \qquad , \forall\ i \neq j$$

Let $\overline{Y}$ be the random variable representing the maximum of this set

$$\overline{Y} = \max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$$

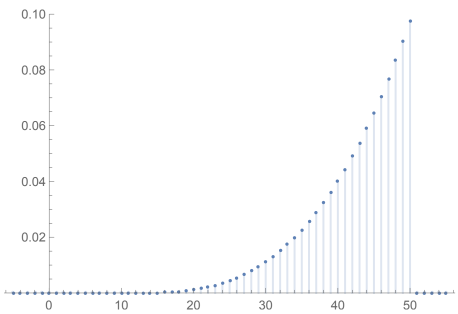
then

$$E\left[\overline{Y}\right] = \frac{n(b+1) + a - 1}{n+1}$$

$$var\left[\overline{Y}\right] = n\frac{(a-b-2)(a-b+n-1)}{(n+2)(n+1)^2}$$

and

$$P\left(\overline{Y} = y\right) = \frac{\binom{y-a}{n-1}}{\binom{b-a+1}{n}} \qquad \text{when } n + a - 1 \leqslant y \leqslant b$$



$a = 10, b = 50, n = 4$

Proof: Differently than before $\overline{X}_i \not\leq \overline{X}_j$, because each $\overline{X}_i$ must be different than the other. The probability of $\overline{X} = \left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$ is the probability of choosing n numbers from the set $[a, \ldots, b]$, i.e.,

$$P\left(\overline{X} = x\right) = \frac{1}{\binom{b-a+1}{n}}$$

The case $\overline{Y} = y$ happens when one of the $\overline{X}_i$ is equal to y, while the other $\overline{X}_i$ are all less than y. They cannot be equal to y since there is no repetition. Therefore the probability of $\overline{Y} = y$ is equal to the probability of choosing $n-1$ numbers from the set $[a, \ldots, y-1]$, i.e.,

$$P\left(\overline{Y} = y\right) = \frac{\binom{y-1-a+1}{n-1}}{\binom{b-a+1}{n}} \qquad \text{when } n+a-1 \leqslant y \leqslant b$$

while $P\left(\overline{Y} = y\right) = 0$ if $y < n+a-1$ or $y > b$.

$$\begin{aligned} E\left[\overline{Y}\right] &= \sum_{y=n+a-1}^{b} yP\left(\overline{Y} = y\right) \\ &= \frac{(b-a+1-n)!n!}{(b-a+1)!\,(n-1)!} \sum_{y=n+a-1}^{b} y \frac{(y-a)!}{(y-a-n+1)!} \\ &= \frac{n\,(b+1)+a-1}{n+1} \end{aligned}$$

The variance of $\overline{Y}$ is

$$\begin{aligned} E\left[\left(\overline{Y} - E\left[\overline{Y}\right]\right)^2\right] &= \sum_{y=n+a-1}^{b} \left(y - \frac{n\,(b+1)+a-1}{n+1}\right)^2 P\left(\overline{Y} = y\right) \\ &= \frac{(b-a+1-n)!n!}{(b-a+1)!\,(n-1)!} \sum_{y=n+a-1}^{b} \left(y - \frac{n\,(b+1)+a-1}{n+1}\right)^2 \frac{(y-a)!}{(y-a-n+1)!} \\ &= n\frac{(a-b-2)(a-b+n-1)}{(n+2)(n+1)^2} \end{aligned}$$

Average of n Uniform Random Variables (continuous, discrete, discrete with no repetitions)

Given a set of n uniform random variables $\overline{X}_i$, either

$$\begin{aligned} \overline{X}_i &\sim \mathcal{U}(a,b) \\ \overline{X}_i &\sim \mathcal{U}^D(a,b) \\ \overline{X}_i &\sim \mathcal{U}^D(a,b) \qquad\qquad \text{and } P\left(\overline{X}_i = \overline{X}_j\right) = 0 \end{aligned}$$

Let $\overline{Y}$ be the random variable representing the average of this set

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}\overline{X}_i$$

then, in any case,

$$E\left[\overline{Y}\right] = \frac{b+a}{2}$$

Proof: In the discrete case, since $E\left[\cdot\right]$ is linear

$$E\left[\overline{Y}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\overline{X}_i\right]$$

This is valid independently of the fact that the $\overline{X}_i$ are not independent.

$$\begin{aligned}
E\left[\overline{X}_i\right] &= \sum_{y=a}^{b} y \; P\left(\overline{X}_i = y\right) \\
&= \frac{1}{(b-a+1)} \sum_{y=a}^{b} y \\
&= \frac{1}{(b-a+1)} \left[\sum_{y=1}^{b} y - \sum_{y=1}^{a-1} y\right] \\
&= \frac{1}{(b-a+1)} \left[\frac{b^2 - a^2 + a + b}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}$$

In case of continuous variables

$$\begin{aligned}
E\left[\overline{X}_i\right] &= \int_a^b y \; p\left(\overline{X}_i = y\right) \; dy \\
&= \frac{1}{b-a} \int_a^b y \; dy \\
&= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\
&= \frac{b+a}{2}
\end{aligned}$$

Estimating the Maximum of a Population

Given a uniformly distributed random variables $\overline{X} \sim \mathcal{U}(a, b)$, observed n times using independent measurements. Let $x_1, \dots x_n$ be the observed values. Estimate b assuming a and n known in the case of continuous random variable and in the case of discrete random variable with or without repetitions.

Solution: We propose three estimators

o_m	based on mean
o_M	based on max
o_{MC}	based on max corrected
o_{MCNR}	based on max corrected for the no repetitions case

Estimator based on Max o_M (continuous case)

Let o_M be

$$o_M\left(\overline{X}\right) = \max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$$

and let $\overline{\tau}\left(b\right)$ be the function generating a set of n random variables $\overline{X}_1, \ldots, \overline{X}_n$ i.i.d. with uniform probability $\mathcal{U}\left(a, b\right)$.
The error of o_M in estimating b,

$$\overline{\varepsilon}\left(b\right) = o_M\left(\overline{\tau}\left(b\right)\right) - b$$

has distribution

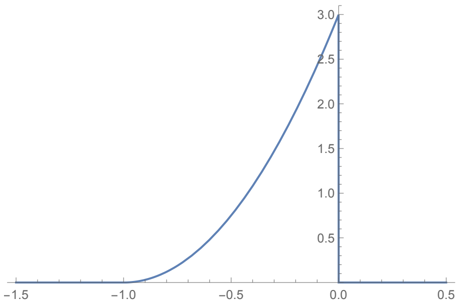
$$p\left(\overline{\varepsilon}\left(b\right) = y\right) = \begin{cases} 0 & y < a - b \\ n\left(\frac{1}{b-a}\right)^n\left(y + b - a\right)^{n-1} & a - b \leqslant y \leqslant 0 \\ 0 & 0 < y \end{cases}$$

The bias of o_M is

$$bias\left(o_M, b\right) = \frac{a-b}{n+1}$$

The variance of o_M is

$$var\left(o_M, b\right) = \frac{n\left(b-a\right)^2}{(n+1)^2(n+2)}$$



$p\left(\overline{\varepsilon}\left(b\right) = y\right)$

Proof: Using maximum i.i.d. theorem, we know that

$$p\left(o_M\left(\overline{\tau}\left(b\right)\right)=y\right)=\begin{cases}0 & y<a\\ n\left(\frac{1}{b-a}\right)^n\left(y-a\right)^{n-1} & a\leqslant y\leqslant b\\ 0 & b<y\end{cases}$$

Therefore $p\left(o_M\left(\overline{\tau}\left(b\right)\right)-b=y\right)$ is the translated version of $p\left(o_M\left(\overline{\tau}\left(b\right)\right)=y\right)$.
The bias of o_M is

$$\begin{aligned}bias\left(o_M,b\right) &= E\left[\overline{\varepsilon}\left(b\right)\right]\\ &= E\left[o_M\left(\overline{\tau}\left(b\right)\right)\right]-b\\ &= \frac{nb+a}{n+1}-b\\ &= \frac{a-b}{n+1}\end{aligned}$$

since

$$E\left[o_M\left(\overline{\tau}\left(b\right)\right)\right]=\frac{nb+a}{n+1}$$

The variance of o_M is

$$var\left(o_M,b\right)=E\left[\overline{\varepsilon}\left(b\right)^2\right]-bias\left(o_M,b\right)^2$$

Since $\overline{\varepsilon}\left(b\right)$ is always negative

$$\begin{aligned}P\left(\overline{\varepsilon}\left(b\right)^2\leqslant y\right) &= P\left(\overline{\varepsilon}\left(b\right)\leqslant\sqrt{y}\wedge\overline{\varepsilon}\left(b\right)\geqslant-\sqrt{y}\right)\\ &= P\left(\overline{\varepsilon}\left(b\right)\geqslant-\sqrt{y}\right)\\ &= 1-P\left(\overline{\varepsilon}\left(b\right)\leqslant-\sqrt{y}\right)\\ &= 1-\int_{a-b}^{-\sqrt{y}}n\frac{1}{\left(b-a\right)^n}\left(y+b-a\right)^{n-1}dy\\ &= 1-\left[\frac{1}{\left(b-a\right)^n}\left(y+b-a\right)^n\right]_{a-b}^{-\sqrt{y}}\\ &= 1-\frac{1}{\left(b-a\right)^n}\left(-\sqrt{y}+b-a\right)^n\\ p\left(\overline{\varepsilon}\left(b\right)^2=y\right) &= \frac{\partial}{\partial y}P\left(\overline{\varepsilon}\left(b\right)^2\leqslant y\right)\\ &= \begin{cases}0 & y<0\\ -\frac{\partial}{\partial y}\frac{1}{\left(b-a\right)^n}\left(-\sqrt{y}+b-a\right)^n & 0\leqslant y\leqslant\left(b-a\right)^2\\ 0 & y>\left(b-a\right)^2\end{cases}\\ &= \begin{cases}0 & y<0\\ \frac{n}{\left(b-a\right)^n}\left(b-a-\sqrt{y}\right)^{n-1}\frac{1}{2\sqrt{y}} & 0\leqslant y\leqslant\left(b-a\right)^2\\ 0 & y>\left(b-a\right)^2\end{cases}\end{aligned}$$

Important: in case of continuous distribution

$$p\left(\overline{\varepsilon}\left(b\right)^2=y\right)\neq p\left(\overline{\varepsilon}\left(b\right)=-\sqrt{y}\right)$$

the only way to compute $p\left(\overline{\varepsilon}\left(b\right)^2=y\right)$ is to pass through $P\left(\overline{\varepsilon}\left(b\right)^2\leqslant y\right)$ as we did before.

The expected value

$$\begin{aligned}
E\left[\overline{\varepsilon}\left(b\right)^{2}\right] &= \int_{0}^{(b-a)^{2}} y\ p\left(\overline{\varepsilon}\left(b\right)^{2}=y\right)\ dy \\
&= \frac{n}{(b-a)^{n}} \int_{0}^{(b-a)^{2}} y\ \left(b-a-\sqrt{y}\right)^{n-1} \frac{1}{2\sqrt{y}}\ dy \\
&= \frac{2\left(b-a\right)^{2}}{(n+1)(n+2)}
\end{aligned}$$

Therefore

$$\begin{aligned}
var\left(o_{M},b\right) &= \frac{2\left(b-a\right)^{2}}{(n+1)(n+2)} - \frac{\left(a-b\right)^{2}}{\left(n+1\right)^{2}} \\
&= \frac{2\left(b-a\right)^{2}\left(n+1\right) - \left(a-b\right)^{2}\left(n+2\right)}{(n+1)^{2}(n+2)} \\
&= \frac{n\left(b-a\right)^{2}}{(n+1)^{2}(n+2)}
\end{aligned}$$

Estimator based on Max o_M (discrete case)

Let o_M be

$$o_M\left(\overline{X}\right) = \max\left\{\overline{X}_1, \ldots, \overline{X}_n\right\}$$

The bias of o_M is

$$bias\left(o_M, b\right) = \frac{1}{\left(b-a+1\right)^n}\left[\sum_{y=a}^{b} y\left(y-a+1\right)^n - \sum_{y=a}^{b} y\left(y-a\right)^n\right] - b$$

and

$$\lim_{n \longrightarrow \infty} bias\left(o_M, b\right) = 0$$

Proof: The bias of o_M is

$$\begin{aligned} bias\left(o_M, b\right) &= E\left[\overline{\varepsilon}\left(b\right)\right] \\ &= E\left[o_M\left(\overline{\tau}\left(b\right)\right)\right] - b \end{aligned}$$

where

$$E\left[o_M\left(\overline{\tau}\left(b\right)\right)\right] = \frac{1}{\left(b-a+1\right)^n}\left[\sum_{y=a}^{b} y\left(y-a+1\right)^n - \sum_{y=a}^{b} y\left(y-a\right)^n\right]$$

Estimator based on the Mean o_m (continuous and discrete case)

Let o_m be

$$o_m\left(\overline{X}\right) = -a + \frac{2}{n}\sum_{i=1}^{n}\overline{X}_i$$

The bias of o_m is

$$bias\left(o_m, b\right) = 0$$

The variance of o_m is

$$\begin{array}{ll} var\left(o_m, b\right) = \frac{1}{3n}\left(b-a\right)^2 & \text{for continuous } \overline{X}_i \\ var\left(o_m, b\right) = \frac{1}{3n}\left(b-a\right)\left(b-a+2\right) & \text{for discrete } \overline{X}_i \end{array}$$

Proof: The bias

$$
\begin{aligned}
bias\,(o_m, b) &= E\left[o_m\left(\overline{\tau}\,(b)\right) - b\right] \\
&= -b - a + \frac{2}{n}\sum_{i=1}^{n} E\left[\overline{X}_i\right] \\
&= -b - a + \frac{2}{n}\sum_{i=1}^{n} \frac{b+a}{2} \\
&= -a - b + b + a \\
&= 0
\end{aligned}
$$

while the variance

$$
\begin{aligned}
E\left[\left(\overline{\varepsilon}\,(b) - 0\right)^2\right] &= E\left[\left(-a - b + \frac{2}{n}\sum_{i=1}^{n}\overline{X}_i\right)^2\right] \\
&= \frac{4}{n^2}E\left[\left(\sum_{i=1}^{n}\left(\overline{X}_i - \frac{1}{n}\frac{n}{2}\,(b+a)\right)\right)^2\right] \\
&= \frac{4}{n^2}E\left[\sum_{i=1}^{n}\left(\overline{X}_i - \frac{1}{2}\,(b+a)\right)\sum_{i=1}^{n}\left(\overline{X}_j - \frac{1}{2}\,(b+a)\right)\right] \\
&= \frac{4}{n^2}E\left[\sum_{i=1}^{n}\sum_{i=1}^{n}\left(\overline{X}_i - \frac{1}{2}\,(b+a)\right)\left(\overline{X}_j - \frac{1}{2}\,(b+a)\right)\right] \\
&= \frac{4}{n^2}\sum_{i=1}^{n}\sum_{i=1}^{n}E\left[\left(\overline{X}_i - \frac{1}{2}\,(b+a)\right)\left(\overline{X}_j - \frac{1}{2}\,(b+a)\right)\right]
\end{aligned}
$$

Since $\overline{X}_i \perp \overline{X}_j$ for each $i \neq j$, we have that, for $i \neq j$,

$$
E\left[\left(\overline{X}_i - \frac{1}{2}\,(b+a)\right)\left(\overline{X}_j - \frac{1}{2}\,(b+a)\right)\right] = E\left[\overline{X}_i - \frac{1}{2}\,(b+a)\right]E\left[\overline{X}_j - \frac{1}{2}\,(b+a)\right]
$$

$$
= \left(E\left[\overline{X}_i\right] - \frac{1}{2}\,(b+a)\right)\left(E\left[\overline{X}_j\right] - \frac{1}{2}\,(b+a)\right) = 0
$$

since $E\left[\overline{X}_i\right] = (b+a)/2$.Therefore

$$
\begin{aligned}
E\left[\left(\overline{\varepsilon}\,(b)\right)^2\right] &= \frac{4}{n^2}\sum_{i=1}^{n}E\left[\left(\overline{X}_i - \frac{1}{2}\,(b+a)\right)^2\right] \\
&= \frac{4}{n^2}\sum_{i=1}^{n} var\left(\overline{X}_i\right)
\end{aligned}
$$

The variance of a uniformly distributed random variable is

$$
\begin{aligned}
var\left(\overline{X}_i\right) &= \tfrac{1}{12}\,(b-a)^2 && \text{for continuous } \overline{X}_i \\
var\left(\overline{X}_i\right) &= \tfrac{1}{12}\,(b-a)\,(b-a+2) && \text{for discrete } \overline{X}_i
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\left[\left(\overline{\varepsilon}\,(b)\right)^2\right] &= \tfrac{1}{3n}\,(b-a)^2 && \text{for continuous } \overline{X}_i \\
E\left[\left(\overline{\varepsilon}\,(b)\right)^2\right] &= \tfrac{1}{3n}\,(b-a)\,(b-a+2) && \text{for discrete } \overline{X}_i
\end{aligned}
$$

Estimator based on Max Corrected o_{MC} (continuous case)

The bias of o_M is $(a-b)/(n+1)$. This information can be used to build a second estimator o_{MC} which is unbiased.

Let o_{MC} be

$$o_{MC}\left(\overline{X}\right) = o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n}$$

The bias of o_{MC} is

$$bias\left(o_{MC}, b\right) = 0$$

The variance of o_M is

$$var\left(o_{MC}, b\right) = \frac{\left(a-b\right)^2}{n\left(n+2\right)}$$

Proof: The error of o_{MC} in estimating b is defined as

$$\begin{aligned}\overline{\varepsilon}(b) &= o_{MC}\left(\overline{\tau}(b)\right) - b \\ &= \left(o_M\left(\overline{X}\right) - b\right) + \frac{o_M\left(\overline{X}\right) - a}{n}\end{aligned}$$

therefore the bias of o_{MC} is

$$\begin{aligned}bias\left(o_{MC}, b\right) &= E\left[\overline{\varepsilon}(b)\right] \\ &= bias\left(o_M, b\right) + \frac{1}{n}\left(E\left[o_M\left(\overline{X}\right)\right] - a\right) \\ &= \frac{a-b}{n+1} + \frac{1}{n}\left(\frac{nb+a}{n+1} - a\right) \\ &= \frac{a-b}{n+1} + \frac{1}{n}\left(\frac{nb+a-an-a}{n+1}\right) \\ &= \frac{1}{n+1}\left(a-b+b-a\right) \\ &= 0\end{aligned}$$

The variance of o_M is

$$\begin{aligned}var\left(o_{MC}, b\right) &= E\left[\left(\overline{\varepsilon}(b) - 0\right)^2\right] \\ &= E\left[\left(o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n} - b\right)^2\right] \\ &= E\left[\left(o_M\left(\overline{X}\right)\left(1+\frac{1}{n}\right) - \left(\frac{a}{n} + b\right)\right)^2\right] \\ &= E\left[o_M\left(\overline{X}\right)^2\left(1+\frac{1}{n}\right)^2 + \left(\frac{a}{n}+b\right)^2 - 2o_M\left(\overline{X}\right)\left(1+\frac{1}{n}\right)\left(\frac{a}{n}+b\right)\right] \\ &= \left(1+\frac{1}{n}\right)^2 E\left[o_M\left(\overline{X}\right)^2\right] + \left(\frac{a}{n}+b\right)^2 - 2E\left[o_M\left(\overline{X}\right)\right]\left(1+\frac{1}{n}\right)\left(\frac{a}{n}+b\right) \\ &= \left(1+\frac{1}{n}\right)^2\left(\frac{2\left(b-a\right)^2}{(n+1)(n+2)} - b^2 + 2b\frac{nb+a}{n+1}\right) + \left(\frac{a}{n}+b\right)^2 - 2\frac{nb+a}{n+1}\left(1+\frac{1}{n}\right)\left(\frac{a}{n}+b\right) \\ &= \frac{\left(b-a\right)^2}{n\left(n+2\right)}\end{aligned}$$

Estimator based on Max Corrected o_{MC} (discrete case)

Let o_{MC} be

$$o_{MC}\left(\overline{X}\right) = o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n}$$

The bias of o_{MC} is

$$\lim_{n \longrightarrow \infty} bias\left(o_{MC}, b\right) = 0$$

Proof: The error of o_{MC} in estimating b is defined as

$$\begin{aligned}\overline{\varepsilon}(b) &= o_{MC}(\overline{\tau}(b)) - b \\ &= o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n} - b\end{aligned}$$

therefore the bias of o_{MC} is

$$\begin{aligned}bias\left(o_{MC}, b\right) &= E\left[\overline{\varepsilon}(b)\right] \\ &= E\left[o_M\left(\overline{X}\right)\right] + \frac{1}{n}E\left[o_M\left(\overline{X}\right)\right] - \frac{a}{n} - b \\ &= \frac{n+1}{n}E\left[o_M\left(\overline{X}\right)\right] - \frac{a}{n} - b\end{aligned}$$

where

$$E\left[o_M\left(\overline{X}\right)\right] = \frac{1}{(b-a+1)^n}\left[\sum_{y=a}^{b} y\left(y-a+1\right)^n - \sum_{y=a}^{b} y\left(y-a\right)^n\right]$$

Numerically,

$$\lim_{n\longrightarrow\infty} E\left[o_M\left(\overline{X}\right)\right] = b$$

hence,

$$\lim_{n\longrightarrow\infty} bias\left(o_{MC}, b\right) = 0$$

Estimator based on Max Corrected o_{MCNR} (discrete case with no repetitions)

Let o_{MCNR} be

$$o_{MCNR}\left(\overline{X}\right) = o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n} - 1$$

The bias of o_{MCNR} is

$$bias\left(o_{MCNR}, b\right) = -\frac{1}{n}$$

The variance of o_{MCNR} is

$$var\left(o_{MCNR}, b\right) = \frac{(b - a + 2)\,(b - a - n + 1)}{n\,(n + 2)}$$

Proof: The error of o_{MCNR} in estimating b is defined as

$$\begin{aligned}
\overline{\varepsilon}(b) &= o_{MCNR}(\overline{\tau}(b)) - b \\
&= o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n} - 1 - b
\end{aligned}$$

therefore the bias of o_{MCNR} is

$$\begin{aligned}
bias\left(o_{MCNR}, b\right) &= E\left[\overline{\varepsilon}(b)\right] \\
&= E\left[o_M\left(\overline{X}\right)\right] + \frac{1}{n}E\left[o_M\left(\overline{X}\right)\right] - \frac{a}{n} - b - 1 \\
&= \frac{n+1}{n}E\left[o_M\left(\overline{X}\right)\right] - \frac{a}{n} - b - 1 \\
&= \frac{n+1}{n}\frac{n(b+1)+a-1}{n+1} - \frac{a}{n} - b - 1 \\
&= -\frac{1}{n}
\end{aligned}$$

The variance of o_{MCNR} is

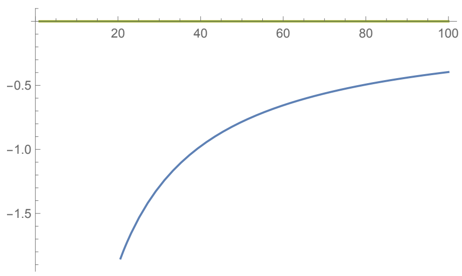
$$\begin{aligned}
var\left(o_{MCNR}, b\right) &= E\left[\left(\overline{\varepsilon}(b) + \frac{1}{n}\right)^2\right] \\
&= E\left[\left(o_M\left(\overline{X}\right) + \frac{o_M\left(\overline{X}\right) - a}{n} - 1 - b + \frac{1}{n}\right)^2\right] \\
&= E\left[\left(o_M\left(\overline{X}\right)\left(1 + \frac{1}{n}\right) + \left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)\right)^2\right] \\
&= E\left[o_M\left(\overline{X}\right)^2\left(1 + \frac{1}{n}\right)^2 + \left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)^2 + 2\left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)o_M\left(\overline{X}\right)\left(1 + \frac{1}{n}\right)\right] \\
&= \left(1 + \frac{1}{n}\right)^2 E\left[o_M\left(\overline{X}\right)^2\right] + \left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)^2 + 2\left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)\left(1 + \frac{1}{n}\right)E\left[o_M\left(\overline{X}\right)\right] \\
&= \left(\frac{n+1}{n}\right)^2 E\left[o_M\left(\overline{X}\right)^2\right] + \left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)^2 + 2\left(\frac{1}{n} - \frac{a}{n} - 1 - b\right)\frac{n(b+1)+a-1}{n} \\
&= \frac{(a-b-2)(a-b+n-1)}{n(n+2)}
\end{aligned}$$

since

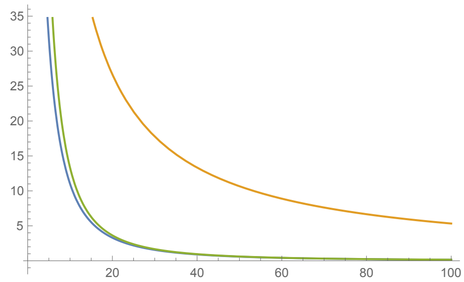
$$\begin{aligned}
E\left[o_M\left(\overline{X}\right)^2\right] &= var\left[o_M\left(\overline{X}\right)\right] + E\left[o_M\left(\overline{X}\right)\right]^2 \\
&= n\frac{(a-b-2)(a-b+n-1)}{(n+2)(n+1)^2} + \left(\frac{n(b+1)+a-1}{n}\right)^2
\end{aligned}$$

Comparison (continuous analysis)

In Figure o_m (orange), o_M (blue), and o_{MC} (green).



bias



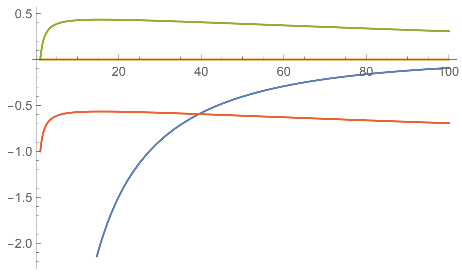
variance

- o_M have a systematic error (bias)
- o_m and o_{MC} does not have a systematic error
- o_M and o_{MC} are more precise (1/variance) than o_m
- o_M and o_{MC} have similar precision

	accuracy	precision	
o_m	perfect	low	
o_M	error	high	
o_{MC}	perfect	high	(best)

Comparison (discrete analysis with repetitions)

In Figure o_m (orange), o_M (blue), o_{MC} (green), and o_{MCNR} (red).



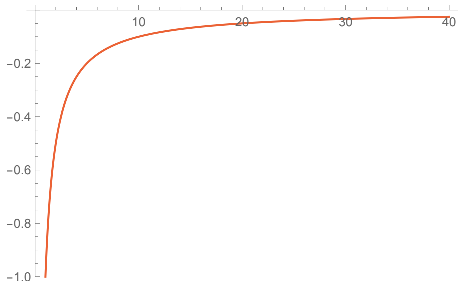
bias

- o_m does not have a systematic error (bias)
- o_M, o_{MC} and o_{MCNR} have systematic error
- o_{MC} has less systematic error than o_M when few samples are available
- o_M has less systematic error than o_{MC} when a lot of samples are available

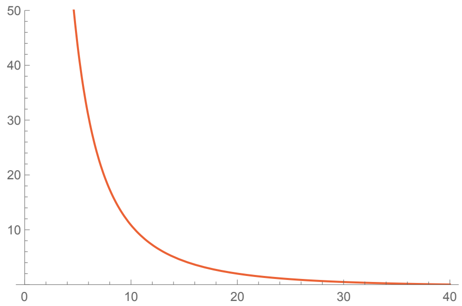
	accuracy	
o_m	perfect	
o_M	$\to 0$	good when a lot of samples are available
o_{MC}	$\to 0$	good when a few samples are available
o_{MCNR}	$\to -1$	

Comparison (discrete analysis with no repetitions)

In Figure o_{MCNR} (red). Note that n must be $< b - a$ and in this case $a = 10, b = 50$.



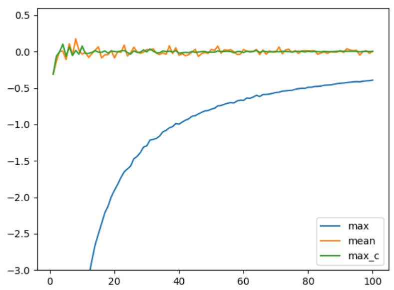
bias



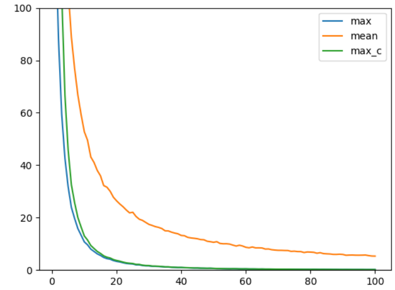
variance

Experiments

Continuous case:

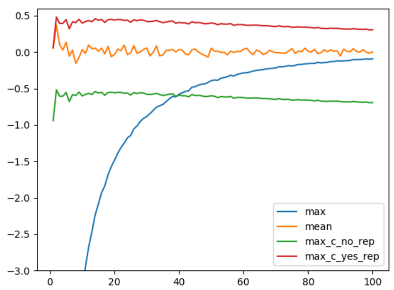


bias

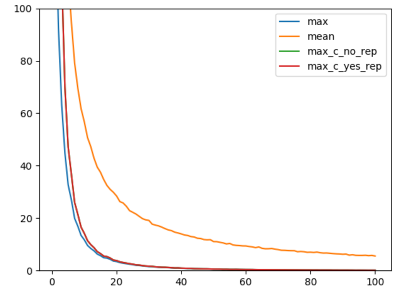


variance

Discrete case with repetitions:

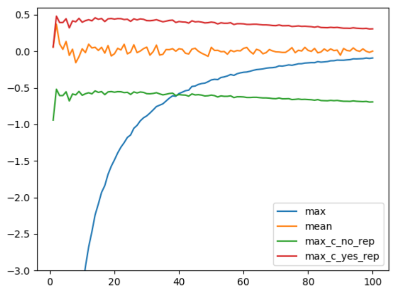


bias

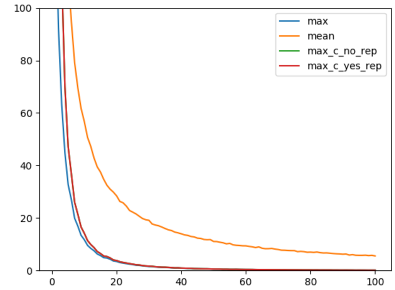


variance

Discrete case with no repetitions:



bias



variance