Example 2: Student ID's Let $X_1, X_2, \dots, X_n$ represent a random sample of Student ID numbers from University of Calgary students.

Assume $X_i \sim Uniform(0, N)$ where N is the total number of U of C students.

$$f_{X_i}(x) = \frac{1}{N-0} = \frac{1}{N}$$

Support for x,

$$0 < x < N$$

$$F_{X_i}(x) = \int_0^x f_x(x)dx = \int_0^x \frac{1}{N}dx = \left[\frac{x}{N}\right]_{x=0}^{x=x} = \frac{x}{N}$$

(a) Is $X_{(n)}$ an unbiased estimator for N? If not, suggest an unbiased estimator for N that is a function of $X_{(n)}$.

$E[X_{(n)}] = ?$

$$f_{X_{(n)}}(x) = n\,[F_x(x)]^{n-1} f_x(x) = n\left[\frac{x}{N}\right]^{n-1}\frac{1}{N} = n\,\frac{x^{n-1}}{N^{n-1}}\,\frac{1}{N} = \frac{n}{N^n}\,x^{n-1}$$

$$E[X_{(n)}] = \int_0^N \frac{n x\, x^{n-1}}{N^n}dx = \frac{n}{N^n}\int_0^N x^n\,dx = \frac{n}{N^n}\left[\frac{x^{n+1}}{n+1}\right]_{x=0}^{x=N} = \frac{n}{N^n(n+1)}\left[N^{n+1} - 0\right]$$

$$= \frac{N^{n+1}}{N^n}\left(\frac{n}{n+1}\right) = N^{n+1-n}\left[\frac{1}{1+1/n}\right] = \frac{N}{1+\frac{1}{n}} = \frac{Nn}{n+1}$$

$B(\hat{\theta}) = \frac{Nn}{n+1} - N = \frac{-N}{n+1}$ ← $X_{(n)}$ would typically under estimate N! ← quite obvious actually!

Make unbiased! → $E[\hat{\theta}] \overset{set}{=} X_{(n)} \Rightarrow \frac{N(n)}{n+1} = X_{(n)} \Rightarrow \frac{X_{(n)}(n+1)}{n} = N$

So $\left(\frac{n+1}{n}\right) X_{(n)}$ would be an unbiased estimator for N

(b) Consider $2\bar{X}$. Is this an unbiased estimator for N?

Here $X \sim \text{Uniform}(0,N)$

$E[X] = \frac{N+0}{2} = \frac{N}{2}$

$$E[2\bar{x}] = 2E\left[\frac{\sum x_i}{n}\right] = \frac{2}{n}\left(E[X_1] + E[X_2] + \cdots + E[X_n]\right) = \frac{2}{n}\left[n\,E(X)\right]$$

$$= \frac{2}{n}\left[n\frac{N}{2}\right] = N, \quad B[2\bar{x}] = E[2\bar{x}] - N = N - N = 0$$

$2\bar{x}$ is unbiased!

(c) Assuming n is "large" which of the two unbiased estimators above a) or b) should you prefer? (Relative efficiency)

$$E[X_{(n)}^2] = \int_0^N \frac{n\,x^{n+1}}{N^n}\,dx = \cdots = \frac{N^2 n}{n+2}$$

$$V\left[\frac{(n+1)}{n}X_{(n)}\right] = \left[\frac{n+1}{n}\right]^2 V[X_{(n)}] = \left(\frac{n+1}{n}\right)^2\left[\frac{N^2 n}{n+2} - \left(\frac{Nn}{n+1}\right)^2\right] = \left(\frac{n+1}{n}\right)^2\left[\frac{N^2 n}{(n+2)(n+1)^2}\right]$$

$$= \frac{N^2}{n(n+2)} \qquad \Rightarrow MSE\left[\left(\frac{n+1}{n}\right)X_{(n)}\right] = \frac{N^2}{n(n+2)} + 0^2 = \frac{N^2}{n^2+2n}$$

$$V[2\bar{x}] = 2^2 V[\bar{x}] \underset{\text{CLT}}{\cong} 4\frac{\sigma^2}{n} = \quad \text{where } \sigma^2 = \frac{(N-0)^2}{12} = \frac{N^2}{12}$$

$$\cong 4\left[\frac{(N^2/12)}{n}\right] \cong \frac{N^2}{3n}$$

$$\Rightarrow MSE[2\bar{x}] = \frac{N^2}{3n} + 0^2 = \frac{N^2}{3n}$$

is $\frac{N^2}{n^2+2n} \overset{?}{<} \frac{N^2}{3n}$

Yes! as long as $n>1$

$X_{(n)}$ is a better estimate for N, than $2\bar{x}$ (since b/y Denominators ⇒ smaller ratio)

Alternative Rel. eff.

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

Pg 445 given two unbiased estimators $\hat{\theta}_1, \hat{\theta}_2$

$$\text{eff}\left(2\bar{x}, \left(\tfrac{n+1}{n}\right)X_{(n)}\right) = \frac{N^2/(n^2+2n)}{(N^2/3n)} = \frac{3}{n+2}. \quad \underline{\text{or}} \quad \text{eff}\left(\left(\tfrac{n+1}{n}\right)X_{(n)}, 2\bar{x}\right) = \frac{(N^2/3n)}{(N^2/(n^2+2n))} = \frac{n+2}{3}$$

if $n \geq 1$

$\text{eff}\left(2\bar{x}, \left(\tfrac{n+1}{n}\right)X_{(n)}\right) \leq 1$

better when <1

better when >1

if $n \geq 1$

$\text{eff}\left(\left(\tfrac{n+1}{n}\right)X_{(n)}, 2\bar{x}\right) \geq 1$

better since >1 when $n>1$