An Ordered Sample Mean that's a bit like Simpson's Rule

D Williams

Abstract: A "Simpson's Rule"-like ordered sample mean is compared with the standard version

=================

Consider the following :

1. Take a sample $X_1, X_2, X_3, \dots, X_{3n}$ of size 3n.
2. Order the sample by swapping terms such that $X_1 \leq X_2 \leq X_3 \leq \dots \leq X_{3n}$
3. Calculate the "Simpson's" Ordered Sample Mean:

$$\bar{X}_{alt(3n)} = \frac{1}{8n}(3(X_1 + X_3 + X_4 + X_6 + \dots + X_{3n}) + 2(X_2 + X_5 + \dots + X_{3n-1}))$$

For example:

1. Take a sample of size 6 from exp(ran#) where ran# is a random number between 0 and 1.
2. Order them (line 2 in table below).
3. Calculate ordered sample mean and compare with standard sample mean (lines 3 and 4 below)

random numbers ordered	0.032316372	0.170959367	0.5464619095	0.621280459	0.7192639355	0.8565404457
exp(ran#)	1.0328442166	1.1864425393	1.7271314476	1.8613098475	2.0529215716	2.3549993364
sample mean	1.7026081599					
ordered sample mean	1.7129739229					

Notice the ordered mean is closer to the population mean (1.71828...=(e-1)=integral of exp(x) from 0 to1) than the standard mean.

Is this more often the case than not? Heuristic arguments suggest it might be.

The Ordered mean above is constructed by using the expression

$$\bar{X}_{alt(3)} = \frac{1}{8}(3(min + max) + 2(median))$$

(where min=minimum, max=maximum, etc on a sub-sample of size 3)

n times on n sub-intervals of the interval (0,1).

See "A Better Type of Sample Mean?" and "An Alternative Model of Probability Theory" at

vixra.org/author/d_williams for more details.

An integral approximation (for non-decreasing f on (0,1))

$$\int_0^1 f(x)dx \approx \frac{1}{8n}(3(f(\frac{1}{2n}) + f(\frac{5}{2n}) + f(\frac{7}{2n}) + f(\frac{11}{2n}) + ...f(\frac{2n-1}{2n})) + 2(f(\frac{3}{2n}) + f(\frac{9}{2n}) + ...))$$

is related to the first expression. This can be generalised to other finite intervals.

Testing this expression on 6 “random” functions with n=6 gave the following results:

	Integral value	Mid Point Rule	Simpson-like Rule
2x^3+7x+3	7	6.9930555556	7
x^(1/3)	0.75	0.7548623802	0.7530626264
10sin(pi*x/2)	6.3661977237	6.3844146463	6.365864624
10sin(pi*x/4)	4.4127120031	4.40368436	4.4122130267
exp(x)	1.7182818285	1.7162946864	1.7182674326
ln(1+x)	0.3862943611	0.3868714395	0.3863079107

Notice the “Simpson-like Rule” is better than the Mid Point Rule in each case.

Other types of sample means based on other types of integral approximation are possible and should be worth exploring.

It may be the case that a particular ordered sample mean better approximates the population mean more often than another if the integral approximation derived from its f(ran#) function for that particular n is closer to the population mean. A proof or refutation of this would be nice.