Variance of the Estimated Efficiency

Introduction

Hoaglin and Andrews (1975) have observed that it is important to give interval estimates or measures of uncertainty to quantities estimated by simulation. Here we derive a general formula for estimating the standard error of the statistical efficiency estimated in a simulation study.

Derivation Using Taylor Series Expansion

Suppose N simulations are to be performed and N is quite large. Let X_i and Y_i, $i = 1, \ldots, N$ denote the squared error in the i-th simulation for estimators $\mathcal{X}$ and $\mathcal{Y}$ respectively. We will assume that before the simulations are run, that the bivariate random variables (X_i, Y_i) are an independent sequence of random variables with finite variance. Note that X_i and Y_i may themselves be correlated and hence not independent of each other but that it is the sequence which is assumed to be independent.

Then sample average squared errors, $(\overline{X}, \overline{Y})$, are approximately bivariate normal since by the usual central limit theorem, any linear combination of $\bar{X}$ and $\overline{Y}$ is approximately normal. The estimated relative efficiency of method $\mathcal{Y}$ vs. method $\mathcal{X}$ is $\hat{e} = \overline{X} / \overline{Y}$. Let $\mu_{\bar{x}} = \mathrm{E}\{\overline{X}\}$ and $\mu_{\bar{y}} = \mathrm{E}\{\overline{Y}\}$. Employing a Taylor series linearization of $\hat{e}$,

$$\hat{e} \doteq \frac{\mu_{\bar{x}}}{\mu_{\bar{y}}} + (\overline{X} - \mu_{\bar{x}}) \frac{1}{\mu_{\bar{y}}} - (\overline{Y} - \mu_{\bar{y}}) \frac{\mu_{\bar{x}}}{\mu_{\bar{y}}^2}$$

Hence, after some algebraic simpliciation it can be shown that,

$$\mathrm{Var}\{\hat{e}\} \doteq$$
$$\frac{1}{N} \left(\frac{\sigma_x}{\mu_y^2} + \frac{\sigma_y \, \mu_x^2}{\mu_y^4} - 2\, \sigma_{x,y} \frac{\mu_x}{\mu_y^3} \right) 2\, \sigma_{x,y} \frac{\mu_x}{\mu_y^3}$$

where $\mu_x = \mathrm{E}\{X_i\}$, $\mu_y = \mathrm{E}\{Y_i\}$, $\sigma_x = \mathrm{Var}\{X_i\}$, $\sigma_y = \mathrm{Var}\{Y_i\}$ and $\sigma_{x,y} = \mathrm{Cov}\{X_i, Y_i\}$. These values can then be estimated by the sample moments of the observed X_i's and Y_i's.

Implementation in *Mathematica*

The following *Mathematica* function evaluates the estimated relative efficiency and its standard error for method $\mathcal{Y}$ vs. method $\mathcal{X}$ given observed data X_i and Y_i, $i = 1, \ldots, N$.

```
Off[General::spell]; Off[General::spell1];
SDRelativeEfficiencyYvsX[X_, Y_] :=
  Module[{μx = Mean[X], μy = Mean[Y], σx, σy, σxy},
   σx = Mean[(X - μx)^2];
   σy = Mean[(Y - μy)^2];
   σxy = Mean[(X - μx) * (Y - μy)];
   e = μx / μy;
   Sqrt[(σx / μy^2 + (σy * μx^2) / μy^4 - (2 * σxy * μx) / μy^3) / Length[X]]
  ];
```

■ Numerical Illustration

```
<< Statistics`
```

```
s1 = RandomArray[NormalDistribution[0, 0.01], 500];
d1 = s1 + RandomArray[NormalDistribution[0, 0.0025], 500];
```

```
SDRelativeEfficiencyYvsX[(s1 - Mean[s1])^2, (d1 - Mean[d1])^2]
```

```
0.0200958
```

Comparison with Bootstrap

More generally, it is easily seen that if X_i and Y_i, $i = 1, \ldots, N$, denote any random variables with finite variances and $\hat{e} = \overline{X} / \overline{Y}$ then the standard deviation of $\hat{e}$ is given by the above formula.

The bootstrap (Efron & Tibshirani, 1993) provides an alternative but less convenient method of evaluating the standard error. Efron & Tibshirani show that in most circumstances 200 bootstrap iterations gives sufficient precision for estimating the standard error.

```
SDRelativeEfficiencyYvsXBootstrap[X_List, Y_List] := Module[
  {i},
  Sqrt[
   Variance[
    Table[
     Mean[X[[i = Table[Random[Integer, {1, Length[X]}], {Length[X]}]]]] /
      Mean[Y[[i]]], {200}]
   ]
  ]
 ]
```

```
SDRelativeEfficiencyYvsXBootstrap[(s1 - Mean[s1])^2, (d1 - Mean[d1])^2]
```

```
0.0192772
```