ESTIMATION OF P(X < Y) FROM INDEPENDENT SAMPLES OF X AND Y

BU-160-M | D. S. Robson and G. F. Atkinson | February 1963

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ denote two independent samples from the distributions $F_X(x)$ and $F_Y(y)$ respectively. An unbiased estimator of

$$p = P(X<Y) = \int_{-\infty}^{\infty} F_X(y)dF_Y(y)$$

is given by

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} F_X^{(m)}(Y_i)$$

where

$$F_X^{(m)}(y) = \frac{1}{m} \text{ (number of observed X's less than or equal to } Y_i) \ .$$

The variance of this estimator is

$$\mathrm{var}(\hat{p}) = \frac{p(1-p)}{m} + \frac{1}{n}\left[\int_{-\infty}^{\infty} F_X^2(y)dF_Y(x) - p^2\right] + \frac{1}{mn}\left[\int_{-\infty}^{\infty} F_X(y)[1-F_X(y)]dF_Y(y) - p(1-p)\right]$$

and an unbiased estimator of the variance is

$$\widehat{\mathrm{var}}(\hat{p}) = \frac{mn}{(m-1)(n-1)}\left\{\frac{\hat{p}(1-\hat{p})}{m} + \frac{1}{n}\left[\frac{1}{n(m-1)}\sum_{i=1}^{n} F_X^{(m)}(Y_i)[mF_X^{(m)}(Y_i)-1] - \hat{p}^2\right] + \frac{1}{mn}\left[\frac{m}{n(m-1)}\sum_{i=1}^{n} F_X^{(m)}(Y_i)[1-F_X^{(m)}(Y_i)] - \hat{p}(1-\hat{p})\right]\right\} \ .$$

---

*NO. BU-160-M in the Biometrics Unit, Mimeo Series, Cornell University