A note on comparison of conditional means.


Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. The problem is to estimate \(\theta = E(X_1|X_1 > 0)/E(Y_1|Y_1 > 0)\). This problem arises in comparing the fishing ability of two different ships, in which \((X_i, Y_i)\) denotes the catch of ships \(X\) and \(Y\) at site \(i\) or time \(i\). The author proposes the estimate \(\hat{\theta} = \tilde{X}/\tilde{Y}\), where \(\tilde{X} = \sum X_i/(\text{no. of nonzero } X_i)\) and \(\tilde{Y} = \sum Y_i/(\text{no. of nonzero } Y_i)\), and investigates its properties.

Specifically, if it is assumed that \(\rho_0 < \infty\), where \(\rho_s = E|X_1/x|^s + E|Y_1/y|^s\), for some \(x\) and \(y\), and that \(P(0 < Y < \delta) = 0\) for some \(\delta\), it can be shown that (1) \(E(\hat{\theta}) = \theta + (a/n) + O(n^{-2})\), (2) \(nE(\hat{\theta} - \theta)^2 = b^2 + O(n^{-1})\), and (3) \(\sup_x |P(\sqrt{n}(\hat{\theta} - \theta) \leq xb) - \Phi(x)| \ll n^{-1/2} \log n\), where \(a\) and \(b\) are constants that can be calculated or estimated, and \(\Phi\) is the standard normal distribution function.

Thomas W. Sager (Austin, TX)