

Second Question: Again assume you are given arbitrary cdfs F_X and F_Y describing random variables X and Y . Find a nontrivial *lower* bound to $\bar{d}(F_X, F_Y)$ which depends only on σ_X^2 and σ_Y^2 .

Solution: The trivial lower bound is 0, since the expected value of the square of a real random variable is nonnegative. As before we know that

$$E[(X - Y)^2] = E(X^2) + E(Y^2) - 2E(XY) = \sigma_X^2 + \sigma_Y^2 - 2E(XY)$$

so to get a lower bound to $\bar{d}(F_X, F_Y)$ we want an upper bound to the correlation $E(XY)$. One of the most important bounds in probability has exactly this form. The Cauchy-Schwartz inequality states that

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)} = \sigma_X \sigma_Y$$

so that for any joint cdf with the required marginals,

$$E[(X - Y)^2] \geq \sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y = (\sigma_X - \sigma_Y)^2$$

and hence

$$\bar{d}(F_X, F_Y) \geq (\sigma_X - \sigma_Y)^2$$

is the desired lower bound.

A few people used the equivalent fact that the correlation coefficient $E(XY)/\sigma_X \sigma_Y$ has magnitude less than 1.