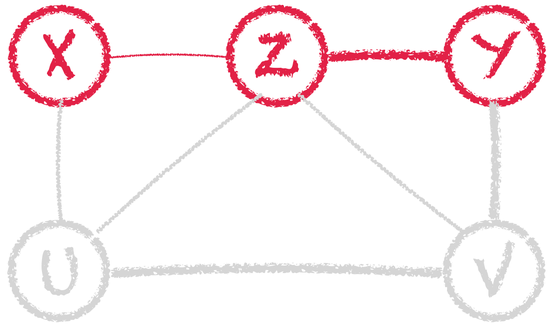
divergence tests of goodness of fit

testing conditional independence
of random variable X, Y and Z with r_X, r_Y and r_Z outcomes

$p =$ model based on empirical distribution $p(x, y, z)$ with $d(p) = r_X r_Y r_Z - 1$

$q = X \perp Y | Z$ such that $p(x, z)p(y, z)/p(z)$ with $d(q) = r_Z - 1 + r_Z(r_X - 1 + r_Y - 1)$



☑ log likelihood ratio test statistic

$$\begin{aligned}\chi^2((r_X - 1)(r_Y - 1)r_Z) &= 2nD(p, q) \\ &= 2n[H(X, Z) + H(Y, Z) - H(Z) - H(X, Y)] \\ &= 2nEJ(X, Y | Z)\end{aligned}$$

☑ independence is rejected if

$$\chi^2((r_X - 1)(r_Y - 1)r_Z) \geq (r_X - 1)(r_Y - 1)r_Z + \sqrt{8(r_X - 1)(r_Y - 1)r_Z}$$

or if the empirical expected joint entropy $J(X, Y)$ is larger than

$$[(r_X - 1)(r_Y - 1)r_Z + \sqrt{8(r_X - 1)(r_Y - 1)r_Z}]/2n$$

divergence tests of goodness of fit

testing nested model specifications

example: five dimensional data (X, Y, Z, U, V) with r_X, r_Y, r_Z, r_U, r_V outcomes

$p =$ model based on empirical distribution $p(x, y, z, u, v)$ with $d(p) = r_X r_Y r_Z r_U r_V - 1$

$q =$ model with listed imposed independence and conditional independence assumptions

examples: $q_1 = X \perp (Y, Z, U, V)$ and $U \perp (Y, Z, V)$

$q_2 = X \perp (Y, Z, U, V)$ and $U \perp (Y, Z, V)$ and $Z \perp V \,|\, Y$