divergence tests of goodness of fit

testing uniform distribution

of random variable X with n observations on r_X outcomes

p = model based on empirical distribution $p(x) = n(x)/n$ (the relative frequencies) with $d(p) = r_X - 1$

$q = X$ is uniformly distributed on r_X outcomes with $d(q) = 0$

☑ log likelihood ratio test statistic

$$\chi^2(r_X - 1) = 2nD(p, q)$$
$$= 2n[\log r_X - H(X)]$$

where $H(X)$ is the empirical entropy of X

☑ uniformity is rejected if

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

or if $H(X)$ deviates from its maximum values $\log(r_X)$ by more than $[r_X - 1 + \sqrt{8(r_X - 1)}]/2n$

divergence tests of goodness of fit

testing pairwise independence

of random variable X and Y with r_X and r_Y outcomes

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

$q = X \perp Y$ such that $p(x) \cdot p(y)$ with $d(q) = (r_X - 1) + (r_Y - 1)$