Let $X_1 \ldots X_n$ be iid r.v. with continuous cdf F.

$H_0: F = F_0 \qquad H_1: F \neq F_0$

where F_0 is known cdf.

ξ - r.v. with continuous cdf F_ξ

$F_\xi(x) = P\{\xi \le x\}$

$\Rightarrow \quad \eta = F_\xi(\xi) \sim \mathcal{U}(0,1)$

$F_\eta(x) = P\{\eta \le x\} = \left/ \begin{matrix} 0, & x \le 0 \\ 1, & x > 1 \end{matrix} \right/ =$

$= P\{F_\xi(\xi) \le x\} = P\{\xi \le F_\xi^{-1}(x)\} =$

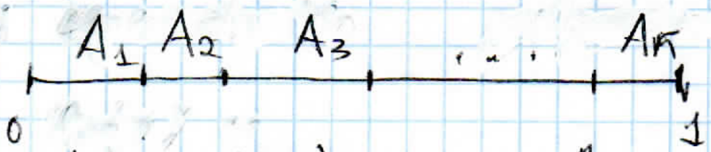
$= F_\xi(F_\xi^{-1}(x)) = x \quad , \quad 0 < x \le 1.$

$F_\eta = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1 \end{cases} \Rightarrow \eta \sim \mathcal{U}(0,1)$

$U_1 = F_0(x_1) \ldots \quad U_n = F_0(X_n)$

$H_0: U \sim \mathcal{U}(0,1); \qquad H_1: U \nsim \mathcal{U}(0,1)$

<u>Pearson's</u> <u>chi-square</u> <u>test</u>



$\bigcup_j A_j = (0,1) \qquad A_j \cap A_\ell = \emptyset, \; j \neq \ell.$

$N_j = \#\{U_i \in A_j\}$

$p_j = P_0(U_i \in A_j) = |A_j|$

The statistics:

$$P_K = \sum_{j=1}^{K} \frac{(N_j - np_j)^2}{np_j}$$

$P_K \sim \chi^2(k-1)$ (under H_0, asymptotically)

$P_K > qchisq(1-\alpha, df = k-1) \Rightarrow$ reject H_0.

Neyman's test

Legendre's polynomials in $L_2([-1,1], du)$

$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}\left((x^2-1)^n\right)$, $n = 0, 1, \ldots$

Ortonormal system in $L_2([-1,1], du)$:

$\{\sqrt{n+\frac{1}{2}}\, P_n(x) : n \geq 0\}$

$\tilde{P}_n(x) = P_n(2x-1) \Rightarrow$

$\tilde{P}_n(x) = \frac{1}{n!} \frac{d^n}{dx^n}\left[(x^2-x)^n\right]$, $n \geq 0$

$\Rightarrow$ Ortonormal system in $L_2([0,1], du)$:

$\{\sqrt{2n+1}\, \tilde{P}_n(x), n \geq 0\}$

$\tilde{P}_0(x) = 1$;
$\tilde{P}_1(x) = 2x-1$,
$\tilde{P}_2(x) = 6x^2-6x+1$;
$\tilde{P}_3(x) = 20x^3-30x^2+12x-1$;
$\tilde{P}_4(x) = 70x^4-140x^3+90x^2-20x+1$; ...

$b_j(x) := \sqrt{2j+1}\, \tilde{P}_j(x)$

The statistics

$$N_K = \sum_{j=1}^{K} \left(\frac{1}{\sqrt{n}} \sum_{l=1}^{n} b_j(u_l)\right)^2$$

$N_K \sim \chi^2(k)$ (under H_0, asymptotically)

$N_K > qchisq(1-\alpha, df = k) \Rightarrow$ reject H_0.

Kolmogorov - Smirnov test:

$$KS = \sqrt{n} \sup_{u \in (0,1)} |\tilde{G}_n(u) - u|,$$

where $\tilde{G}_n(u)$ - empirical cdf in the sample $U_1 \ldots U_n$

KS ~ Kolmogorov's distribution (asymptotically under H_0).

(i) P_4 and P_8
(ii) N_1, N_4, N_8
(iii) KS

$\alpha = 0.05$.