4.) Diskretna porazdelitev na $m+1$ točkah: $\eta_0, \eta_1, \dots, \eta_m$ z verjetnostmi $(p_0, p_1, \dots, p_m)$

$H_0: (p_0, p_1, p_2, p_3, p_4) = (\frac{1}{10}, \frac{1}{10}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5})$, $\alpha = 0.05$, $m = 4$

za $n = 30, 50, 70, 90$ izračunajte

a) eksaktno velikost preizkusa domneve H_0 na podlagi razmerja verjetij

Privzemimo, da vzorčimo (NEP) $X \sim \begin{pmatrix} \eta_0 & \eta_1 & \eta_2 & \eta_3 & \eta_4 \\ p_0 & p_1 & p_2 & p_3 & p_4 \end{pmatrix}$

Parametrični prostor je $\Theta = \{(p_0, p_1, p_2, p_3, p_4) \mid p_j \in (0,1), \sum_{j=0}^{m} p_j = 1\}$ $\quad m = 4$

Preizkušamo $H_0: (p_0, p_1, p_2, p_3, p_4) = (\frac{1}{10}, \frac{1}{10}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5})$, kar je primerno za Wilksov izrek

Enostavno vidimo tudi, da je $\dim_H = 0$

Spomnimo se: CNV je podana kot $\hat{p} = (\frac{T_0}{n}, \frac{T_1}{n}, \frac{T_2}{n}, \frac{T_3}{n}, \frac{T_4}{n})$, kjer je T_j frekvenca pojavitev vrednosti η_j v vzorcu $(x_1, \dots, x_n)$. Vemo tudi: $L(x; p) = \prod_{i=1}^{m} p_j^{T_j(x)}$

Delamo preizkus z razmerjem verjetij: $\longrightarrow$ porazdeljeno multinomsko

$$P((T_0, T_1, T_2, T_3, T_4) = (t_0, t_1, t_2, t_3, t_4)) = \binom{n}{t_0, t_1, t_2, t_3, t_4} p_0^{t_0} p_1^{t_1} p_2^{t_2} p_3^{t_3} p_4^{t_4}$$

Označimo: $H := \{(\Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4)\}$ $\rightarrow \dim H_0 = 0$

Sledi
$$\lambda(x) = \frac{L(x; \hat{\vartheta}_H(x))}{L(x; \hat{\vartheta}(x))} = \frac{\prod_{i=1}^{m} \Pi_j^{T_j(x)}}{\prod_{i=1}^{m} \left(\frac{T_j(x)}{n}\right)^{T_j(x)}} = \prod_{j=0}^{m} \left(\frac{n\Pi_j}{T_j(x)}\right)^{T_j(x)}$$

Domnevo zavrnemo, če je $-2\ln\lambda > \chi^2_{m;\alpha}$ $\longrightarrow$ $\begin{cases} H \text{ zavrnemo}; & -2\ln\lambda > \chi^2_{m;\alpha} \\ H \text{ ne zavrnemo}; & -2\ln\lambda \le \chi^2_{m;\alpha} \end{cases}$

Zanima nas eksaktna velikost preizkusa = velikost testa = verjetnost, da pride do napake 1. vrste. Zanima nas torej $P(\text{test zavrne } H_0 \mid H_0)$

Kako se lotimo zadeve za posamezen n:

- za dan n zapišemo vse možne izide ~~...~~ naše multinomske porazdelitve $(m=4)$
  ↳ ~~...~~ vse možne delitve n bombonov med $m+1$ ljudi $\begin{pmatrix} T_0 & T_1 & T_2 & T_3 & T_4 \\ 30 & 0 & 0 & 0 & 0 \\ 29 & 1 & 0 & 0 & 0 \\ 23 & 0 & 5 & 1 & 1 \\ & & \vdots & & \end{pmatrix}$
- za vsak ~~...~~ izid izračunamo verjetnost, da se pri naših podanih vrednostih realizira
  ↳ v R: dmultinom $((T_{0_i}, T_{1_i}, T_{2_i}, T_{3_i}, T_{4_i}), (p_0, p_1, p_2, p_3, p_4))$ (za i-to realizacijo)
- za vsako ~~...~~ možno realizacijo poračunamo vrednost $\lambda = \prod_{j=0}^{4} \left(\frac{n\Pi_j}{T_j(x)}\right)^{T_j(x)}$ $\quad (\Pi_j = p_j)$
- Izračunamo testne statistike: $-2\ln\lambda$ (za vsak izid)
- Za vsak izid naredimo test: $-2\ln\lambda$ primerjamo s $\chi^2_{4;0.05}$

Vzamemo vse izide, kjer test hipotezo zavrne – tam bo veljalo $\{\text{test zavrne } H_0 \mid H_0\}$, saj smo za računanje testne statistike uporabili H_0. Da dobimo eksaktno velikost preizkusa le še seštejemo verjetnosti iz druge alineje (pri izidih, kjer je test zavrnil H_0).

Program implementiramo v R in dobimo eksaktne velikosti preizkusa za H_0 na podlagi razmerja verjetij:

$n = 30$: 0.06627206

$n = 50$: 0.05697982

$n = 70$: 0.05516617

$n = 90$: 0.05327839

b) Eksaktna velikost preizkusa domneve H_0 na podlagi statistike: $\sum_j \frac{(T_j - n\pi_j)^2}{n\pi_j}$

Pri tej točki postopamo popolnoma enako kot v a), le da ~~na~~ namesto testne statistike $-2\ln\lambda$ uporabimo Pearsonovo statistiko $\sum_j \frac{(T_j - n\pi_j)^2}{n\pi_j}$

Iz predavanj vemo tudi, da velja $\sum_{j=0}^{m} \frac{(T_j - n\pi_j)^2}{n\pi_j} \xrightarrow[n\pi_j]{D} \chi^2_m$, torej tudi test ostane enak. Še vedno je seveda $\pi_j = p_j$ in T_j ~~opažena~~ frekvenca z_j.

Implementiramo v R in dobimo eksaktne velikosti preizkusa na podlagi Pearsonove statistike:

$n = 30$: 0.05058951

$n = 50$: 0.04788236

$n = 70$: 0.04916341

$n = 90$: 0.04841445

Rezultati obeh točk delujejo primerno.