Co to jest plan losowania? Zapisz dowolny plan losowania i podaj jego nazwę, przedstaw jego schemat losowania (np. lpbz)

Planem losowania próby S nazywamy rozkład prawdopodobieństwa P(s) określony na przestrzeni próby P, który dla każdej próby S∈P spełnia warunki P(s)≥0 i $\sum_P P(s) = 1$

Def. Plan bezzwrotnego losowania próby prostej (**lpbz**): $P_{bz}(s) = \binom{N}{n}^{-1}$

Schemat **lpzz**: $p(k) = \frac{1}{N}$, $k = 1, 2, \ldots, N$

Omówić plan losowania Bernoulliego. Omówić plan losowania Poissona.

$$\boldsymbol{P_\pi(s) = \prod_{k=1}^{N} \pi^{s_k} (1-\pi)^{1-s_k} \quad , gdzie\ s \in Q\ i\ \pi \in [0, 1]}$$

Parametr π jest prawdopodobieństwem wylosowania każdego elementu populacji do próby, czyli $\pi_k=E(S_k)= \pi$ dla k=1,..,N.

Próby Bernoulliego nie mają powtarzających się elementów populacji, a ich rozmiar jest zmienną losową, której wartość spełnia nierówności: 0≤n(s)≤N.

Poisson

$P_P(s) = \prod_{k=1}^{N} \pi_k^{s_k} (1-\pi_k)^{1-s_k}$,gdzie s∈Q i π∈[0,1]

Element k-ty z populacji jest losowany do próby z prawdopodobieństwem π_k, k=1,…,N, które jest jednocześnie prawdopodobieństwem inkluzji rzędu pierwszego oraz $E(S_k)= \pi_k$. Elementy populacji są losowane do próby w sposób niezależny. Zatem prawd. Inkluzji II rzędu wynosi $\pi_{k,l}= \pi_k \pi_l$ dla k≠l, k=1,..,N, l=1,..,N. Wartość oczekiwana liczebności próby wynosi $E(n(S))=\sum_{k=1}^{N} \pi_k$

Podać definicje prawdopodobieństw inkluzji pierwszego i drugiego rzędu. Podać nazwę dowolnego planu losowania - ile wynoszą prawdopodobieństwa inkluzji pierwszego i drugiego rzędu dla tego planu losowania.

Prawdopodobieństwo doboru do próby o ustalonej liczebności k-tego elementu populacji (k=1,..,N) określa wyrażenie: $\pi_k=\sum_{s \in A(k)} P(s)$, A(k)={s: k∈s}

Prawdopodobieństwo inkluzji rzędu drugiego, czyli, że jednocześnie k-ty oraz t-ty elementy populacji należą do próby (k=1,..,N oraz k≠t), są postaci $\pi_k=\sum_{s \in A(k,t)} P(s)$, A(k)={s: k∈s , t ∈s, k≠t}

Przykładowy plan losowania z inkluzjami - Poisson

Def. Plan bezzwrotnego losowania próby prostej (**lpbz**): $P_{bz}(s) = \binom{N}{n}^{-1}$. Tu: $\pi_k = \frac{n}{N}$, $\pi_{kl} = \frac{n(n-1)}{N(N-1)}$,

Na czym polega losowanie warstwowe? Zapisać plan losowania losowania warstwowego.

W losowaniu warstwowym przed przystąpieniem do losowania populację dzieli się na pewną liczbę rozłącznych i wewnętrznie jednorodnych warstw (np. klasy miejscowości, rodzaje działalności). Po ustaleniu ile jednostek losuje się z każdej warstwy dokonuje się losowania niezależnie z każdej warstwy. Próbę stanowią jednostki wylosowane ze wszystkich warstw.

Plan losowania warstwowego (próby losowane z warstw ze zwracaniem) $P_w(s) = \prod_{h=1}^{H} N_h^{-n_h}$

Plan losowania warstwowego (próby losowane z warstw bezzwracania) $P_w(s) = \prod_{h=1}^{H} \binom{N_h}{n_h}^{-1}$

Co oznacza alokacja proporcjonalna próby w warstwach – zapisać wzorem, wyjaśnić wszystkie użyte oznaczenia.

Jeżeli próba losowana jest w sposób proporcjonalny, tzn. liczebności próby w h-tej warstwie wynoszą
$n_h = \frac{N_h}{N} * n = W_h * n$

gdzie n-liczebność całej próby,
N_h - określa liczeb h-tej warstwy,
n_h - liczebn próby w h-tej warstwie,
W_h - wagi dla h-tej warstwy;iloraz liczebn h-tej przez liczeb całej popul

Proporcjonalna alokacja próby - polega na losowaniu podprób z poszczególnych warstw tak, że stosunek liczebności każdej podpróby do liczebności całej próby jest równy stosunkowi liczebności danej warstwy w stosunku do liczebności całej populacji generalnej. Dzięki takiemu podejściu otrzymana próba jest automatycznie próbą wyważoną.

Wyjaśnić, co oznacza optymalna alokacja próby w warstwach uzyskana poprzez warunkową minimalizację wariancji – zapisać postać zadania optymalizacyjnego, wyjaśnić wszystkie użyte oznaczenia, krótko opisać.

a) Warunkowa minimalizacja wariancji (podręcznik s. 107-107)

$$\begin{cases} D^2(\overline{y}_{wS}) = minimum \\ k(\mathrm{n}_1,\ldots,\mathrm{n}_H) \le K,\ \ 2 \le n_h \le N_h, h = 1,2,\ldots,H \end{cases}$$, gdzie K – maksymalny akceptowalny koszt

Optymalne liczebności prób w warstwach: $\underline{n}_h = c_o \mathrm{w}_\mathrm{h}\sqrt{\frac{\mathrm{v}_{*\mathrm{h}}}{\mathrm{k}_\mathrm{h}}}$, gdzie $c_o = \frac{K}{\sum_{h=1}^{H} w_h\sqrt{v_{*h}k_h}}$ przy założeniu:

$N_h \to \infty$ lub $N_h \ge \frac{1}{k_0}\left(K - 2\sum_{t=1}^{H} k_h + k_0\right)$ gdzie $k_0 = \min_h\{k_h\}$

b) Warunkowa minimalizacja kosztów <u>(podręcznik s. 110-111)</u>

$$\begin{cases} k(n_1,\ldots,n_H) = minimum \\ D^2(\overline{y}_{wS}) \le f_0, 2 \le n_h \le N_h, h = 1,2,\ldots,H \end{cases}$$, gdzie f_0 - maksymalna dopuszczalna wartość wariancji

Optymalne liczebności prób w warstwach $\underline{n}_h = c_I w_h \sqrt{\frac{v_{*h}}{k_h}}$, gdzie $c_I = \frac{\sum_{h=1}^{H} w_h \sqrt{k_h v_{*h}}}{f_0 + \sum_{h=1}^{H} \frac{w_h}{N_h}\sqrt{v_{*h}}}$ przy założeniu:

$N_h \to \infty$ lub $N_h \ge \frac{1}{k_0}\left(K - 2\sum_{t=1}^{H} k_h + k_0\right)$ gdzie $k_0 = \min_h\{k_h\}$

Na czym polega losowanie grupowe? Zapisać plan losowania losowania grupowego.

Populacja Ω o liczebności N dzieli się na G rozłącznych i niepustych zbiorów Ω_p (gdzie p=1,2,...,G) nazywanych grupami lub zespołami każdy o liczebności M (zakładamy, że są równoliczne), tak, że $\Omega = \bigcup_{p=1}^{G} \Omega_p$ i $N = GM$. Losowana jest g elementowa bezzwrotna próba grup spośród G grup, w każdej z g grup badane są wszystkie elementy populacji.

Plan losowania: $P(s) = \frac{1}{\binom{G}{g}}$

Jakie wartości przyjmuje współczynnik korelacji wewnątrzgrupowej i jakie ma znaczenie w przypadku losowania grupowego?

$r^{(w)} = 1 - \frac{v_w}{v} \in \left[-\frac{1}{M-1}; 1\right]$ to wsp. korelacji wewnątrzgrupowej, gdzie $v_w = \frac{1}{G}\sum_{p=1}^{G} v_{*p}$ to wariancja wewnątrzgrupowa (natomiast $v_{*p} = \frac{1}{M-1}\sum_{k\in\Omega_p}(y_k - \overline{y}_p)^2$ a $\overline{y}_p$ jest średnią w p-tej grupie).

Losowanie grupowe może być bardziej efektywne niż lpbz jeśli $r^{(w)}$ jest ujemne. Stąd, v_w powinno być duże co oznacza, że populacja powinna być dzielona na grupy tak, aby zróżnicowanie wewnątrz grup było jak najsilniejsze.

M – liczebności grup

Na czym polega losowanie dwustopniowe? Zapisać plan losowania losowania dwustopniowego.

Populacja Ω o liczebności N dzieli się na G rozłącznych i niepustych zbiorów Ω_h (gdzie h=1,2,...,G) nazywanych grupami, o liczebnościach N_h, gdzie $\bar{N} = \frac{1}{G}\sum_{h=1}^{G} N_h$.

Pierwszy stopień losowania – losujemy g spośród G grup (lpbz).

Drugi stopień losowania – z każdej grupy Ω_h losujemy próbę s_h o liczbności n_h.

Plan losowania: $P(s) = \binom{G}{g}^{-1} \prod_{h=1}^{g} \binom{N_h}{n_h}^{-1}$

Jak należy dobierać cechę dodatkową, gdy wykorzystujemy estymator ilorazowy. Przedstawić wzór, wyjaśnić wszystkie użyte oznaczenia.

Jeśli $\rho(x,y) > 0$ (współczynnik korelacji w populacji pomiędzy zmiennymi x i y jest dodatni) to:

$\frac{|\mathrm{o}_I|}{D(\bar{y}_{IS})} \le z(x)$, gdzie $z(x)$ to współczynnik zmienności zmiennej x w populacji

Wniosek: Jeśli chcemy korzystać z estymatora ilorazowego to zmienną x należy tak wybrać, aby była dodatnio skorelowana ze zmienną y i miała możliwie mały współczynnik zmienności

Porównać precyzję estymatora regresyjnego dla lpb z próbą prostą dla lpbz. Przedstawić wzór, wyjaśnić wszystkie użyte oznaczenia.

Porównanie precyzji estymatora regresyjnego ze średnią z próby prostej (lpbz):

$$def(\bar{y}_{rS}) = \frac{D^2(\bar{y}_{rS})}{D^2(\bar{y}_S)} \approx \frac{\frac{N-n}{nN} v_*(y)(1-\rho^2(\mathrm{x},\mathrm{y}))}{\frac{N-n}{nN} v_*(y)} = 1-\rho^2(\mathrm{x},\mathrm{y})$$

Estymator regresyjny jest nie mniej efektywny niż średnia z próby prostej, a jego precyzja jest tym większa im większa jest siła liniowej zależności pomiędzy zmiennymi x i y.