Def. 1 (Dystrybuanta empiryczna) $\hat{F_n}(t) = \frac{card\{x_i : x_i \leqslant t\}}{n}$

$$\sup_t |F(t) - \hat{F_n}(t)| \underset{n \to \infty}{\longrightarrow} 0$$

Def. 2 (Statystyka k-wymiarowa) $T : R^n \to R^k$ $T(X_1, \ldots, X_n) \wedge X_1, \ldots, X_n \sim iid\, F$

k-ta statystyka pozycyjna $X_{k:n}$ k-ta liczba w uporządkowanym $\geqslant$ ciągu

Def. 3 (Statystyka dostateczna) *T, że rozk. warunkowy $P_\theta\{\cdot|T = t\}$ nie zależy od θ*

Tw. 1 (O faktoryzacji) *T jest dostateczna $\iff$*

$$f_\theta(\mathbf{x_n}) = g_\theta(T(\mathbf{x_n}))h(\mathbf{x_n})$$

g_θ - zależy od θ, zależy od $\mathbf{x_n}$ tylko przez T
h - nie zależy od θ, zależy od $\mathbf{x_n}$

Def. 4 (Mediana)

$$M_e = \begin{cases} X_{\frac{n+1}{2}:n} \iff n \equiv (1 mod 2) \\ \frac{1}{2}(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}) \iff 2 \mid n \end{cases}$$

z tw. Cantora $\overline{X} \underset{(-)}{+} \overline{Y} \sim N(m_1 \underset{(-)}{+} m_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

Tw. 2 (Fisher) $X_1, \ldots, X_n \sim iid\, N(m, \sigma^2)$

$$\overline{X}, S^2 \text{ są niezależne} \wedge \overline{X} \sim (m, \frac{\sigma^2}{n}) \wedge \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

Tw. 3 $X_1, \ldots, X_k \sim iid\, N(0, 1)$

1. $Y = \sum X_i^2 \sim \chi^2_k$
2. $EY = E(\sum X_i^2) = \sum_{i=1}^k EX_i^2 = k$

Tw. 4 $Y_1, \ldots, Y_m\, nzal \wedge Y_i \sim \chi^2_{v_i}$

$$\sum Y_i \sim \chi^2_{\sum v_i}$$

Tw. 5 $X_1, \ldots, X_{n1} \sim iid\, N(m_1, \sigma_1^2) \wedge Y1, \ldots, Y_{n2} \sim iid\, N(m_2, \sigma_2^2)$

$$\frac{n_1, S_1^2}{\sigma_1^2} + \frac{n_2, S_2^2}{\sigma_2^2} \sim \chi^2_{n_1+n_2-2}$$

Tw. 6 (Gosset) $z : X, Y\, nzal \wedge X \sim N(0, 1) \wedge Y \sim \chi^2_v$

1. $T = \frac{X}{\sqrt{\frac{Y}{v}}} \sim t_v$
2. $T = \frac{\sqrt{v}\frac{\overline{X}-m}{\sigma^2}}{\sqrt{\frac{\frac{vS^2}{\sigma^2}}{v-1}}} = \frac{\overline{X}-m}{S}\sqrt{v-1} \sim t_{v-1}$

Tw. 7 $X, Y\, nzal \wedge X \sim \chi^2_{v_1} \wedge Y \sim \chi^2_{v_2}$

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}} \sim F_{v_1, v_2}$$

Tw. 8 $X_1, \ldots, X_n\, iid \sim N(m_1, \sigma_1^2) \wedge S_1^2 = \frac{1}{n_1}\sum(X_i - \overline{X})^2 \wedge Y_1, \ldots, Y_n\, iid \sim N(m_2, \sigma_2^2) \wedge S_2^2 = \frac{1}{n_2}\sum(Y_i - \overline{Y})^2$

$$F = \frac{\frac{\frac{n_1 S_1^2}{\sigma_1}}{n_1-1}}{\frac{\frac{n_2 S_2^2}{\sigma_2}}{n_2-1}} \sim F(n_1 - 1, n_2 - 1)$$

Statystyki pozycyjne

$$F_{X_{k:n}} = P(X_{k:n} < x) = \sum_{i=k}^{n} \frac{n!}{m!(n-m)!}(F(x))^i(1-F(x))^{n-i} =$$

$$= \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1}(1-t)^{n-k}dt$$

$$f_{X_{k:n}} = P(X_{k:n} = x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}f(x)$$

Własności estymatora

- Nieobciążoność: $E(T_n) = \theta$
- Zgodność $\lim_{n\to\infty} P\{|T_n - \theta| < \epsilon\} = 1 \wedge \epsilon > 0$
- Efektywność $MSE(T_n) = E(T_n - \theta)^2 = D^2(T_n) + B^2(T_n)$
  $MSE(T_{n1}) < MSE(T_{n2})$ to estymator jest 1 efektywniejszy; $B^2(T_n)$ - obciążenie

Def. 5 (Kres dolny wariancji) $D^2(T_n)$: $D^2(T_n^*) = \frac{1}{nE(\frac{\partial \ln P_x(x,\theta)}{\partial\theta})^2}$

Def. 6 (Efektywność estymatora) $e(T_n) = \frac{D^2(T_n^*)}{D^2(T_n)} \leqslant 1$

Tw. 9 (Czebyszew) $P(|X - EX| \geqslant \epsilon) \leqslant \frac{D^2X}{\epsilon^2} \wedge \epsilon > 0$

Metoda Największej Wiarygodności (MNW)

$$f(x_1, \ldots, x_n, \theta) = \Pi_{i=1}^n f(x_i, \theta)$$

$$\hat{\theta} : \frac{\partial \ln f(x_1, \ldots, x_n, \theta)}{\partial\theta} = 0$$

Metoda Momentów (MM)

$$x = EX = (\theta) \Rightarrow \theta = f^{-1}(x)$$

Testy

$$(\Omega, B, P) \;\; P = \{P_\theta : \theta \in \Theta\} \;\; \Theta = \Theta_0 \cup \Theta_1 \wedge \Theta_0 \cap \Theta_1 = \emptyset$$

$$H_0 : \theta_o \in \Theta_0 \;\; H_1 : \theta_0 \in \Theta_1$$

Błąd I rodzaju: odrzucenie prawdziwego H_0
Poziom istotności, rozmiar testu $\alpha = P\{\mathbf{x_n} \in W|H_0\} =$ P(I rodz.)
$\beta = P\{\mathbf{x_n} \in (X \setminus W|H_1\} =$ P(II rodz.)

Def. 7 (Moc testu) $P\{\mathbf{x_n} \in W|H_1\} = M(W) = 1 - \beta$

Def. 8 (Funkcja mocy testu) $M(\theta_1, W) = P\{\mathbf{x_n} \in W|\theta = \theta_1\}$

Def. 9 (Test nieobciążony) *Dla* $\alpha \in (0,1)$ $P\{\mathbf{x_n} \in W|H_0\} = \alpha \wedge P\{\mathbf{x_n} \in W|H_1\} > \alpha$

Def. 10 (Test zgodny) $\lim_{n\to\infty} P\{\mathbf{x_n} \in W|H_1\} = 1$

Tw. 10 (Porównanie mocy testów) *z:* $W, W^* \in X, \alpha$ $P\{\mathbf{x_n} \in W^*|H_0\} \leqslant P\{\mathbf{x_n} \in W|H_0\}$
Jeśli $M(W^*) = P\{\mathbf{x_n} \in W^*|H_1\} \geqslant P\{\mathbf{x_n} \in W|H_1\} = M(W)$ *to test oparty na* W^* *jest jednostajnie mocniejszy od testu opartego na* W

Def. 11 (Test najmocniejszy) *test, który minimalizuje* β *przy zadanym* α

Lemat 1 (Neuman - Pearson)

$$\Theta = \{\theta_0, \theta_1\} \wedge H_0 : \theta = \theta_0 \wedge H_1 : \theta = \theta_1 \wedge \alpha \in (0,1)$$

Szukamy W^* takiego, że $M \to max$

$$V(\mathbf{x_n}) = \frac{L(\mathbf{x_n}, \theta_1)}{L(\mathbf{x_n}, \theta_0)} \geqslant k \Rightarrow \mathbf{x_n} \in W^*$$

$$V(\mathbf{x_n}) = \frac{L(\mathbf{x_n}, \theta_1)}{L(\mathbf{x_n}, \theta_0)} < k \Rightarrow \mathbf{x_n} \in (X \setminus W^*)$$

$$P\{X_n \in W^* | H_0\} = \alpha \wedge M(W^*) > M(W)$$

Def. 12 (Najmocniejszy test zrandomizowany) $H_0 : \theta = \theta_0 \wedge H_1 : \theta = \theta_1$ *Jeżeli funkcja odrzucenia*

$$e(\mathbf{x_n}) = \begin{cases} 1 \iff V(\mathbf{x_n}) > k \\ \gamma \iff V(\mathbf{x_n}) = k \\ 0 \iff V(\mathbf{x_n}) < k \end{cases}$$

$$E[e(\mathbf{x_n})|H_0] = 1 \cdot P\{V(e(\mathbf{x_n}) > k|H_0\} + \gamma \cdot P\{V(e(\mathbf{x_n}) = k|H_0\} + 0 \cdot P\{V(e(\mathbf{x_n}) < k|H_0\}$$

ma tak dobrane stałe $k \wedge 0 < \gamma < 1$ że $E[e(\mathbf{x_n})|H_0] = \alpha$ to taki test nazywamy najmocniejszym testem zrandomizowanym.

Test 1 (Ilorazowy) $X, P_\theta \;\; \theta \in \Theta \subset R^r \;\; H_0 : \theta \in \Theta_0 \;\; H_1 : \theta \in (\Theta \setminus \Theta_0) = \Theta_1$

$$L(\mathbf{x_n}, \theta) = \prod_{i=1}^{n} p_x(x_i, \theta) \;\; L_{\Theta_0}(\mathbf{x_n}, \theta) = \sup_{\theta \in \Theta_0} L(\mathbf{x_n}, \theta) \;\; L_{\Theta}(\mathbf{x_n}, \theta) = \sup_{\theta \in \Theta} L(\mathbf{x_n}, \theta)$$

$$\lambda = \lambda(\mathbf{x_n}) = \frac{L_{\Theta_0}(\mathbf{x_n}, \theta)}{L_{\Theta}(\mathbf{x_n}, \theta)} \wedge 0 < \lambda \leqslant 1$$

$$W = \{\mathbf{x_n} : \lambda \leqslant k_\alpha\} \wedge P\{\lambda(\mathbf{x_n}) \leqslant k_\alpha\} = \alpha$$

Test 2 (Ilorazowy dla dużych prób) $X, P_\theta \;\; \theta \in \Theta \subset R^r \;\; H_0 : \theta \in \Theta_0 \;\; H_1 : \theta \in (\Theta \setminus \Theta_0) = \Theta_1$

$$\mathbf{x_n} = (\underbrace{x_1, \ldots, x_q}_{r-q \;\; liczb \;\; \notin H_0} , \underbrace{x_{q+1}, \ldots, x_n}_{q \;\; liczb \;\; \in H_0})$$

Jeżeli $n \to \infty \wedge H_0$ to $-2\ln\lambda \to \chi^2_{r-q,\lambda} \wedge W = \{\mathbf{x_n} : -2\ln\lambda > \chi^2_{r-q,\lambda}\} \wedge P\{\chi^2_{r-q} > \chi^2_{r-q,\alpha}\} = \alpha$

Test 3 (Batrletta na równość wariancji) $X_1, \ldots, X_k \wedge X_i \sim N(m_i, \sigma_i^2)$ *z i-tej zmiennej pobiera się próbę $n_i \wedge n = \sum n_i$*
$H_0 : \forall_{i,j} \sigma_i^2 = \sigma_j^2 \;\; H_0 : \exists_{i,j} \sigma_i^2 \neq \sigma_j^2$
$\overline{x_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \;\; s^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_i})^2$

$$L(m_1, \ldots, m_k, \sigma_1^2, \ldots, \sigma_k^2) = \frac{1}{(\sqrt{2\pi})^n \prod_{i=1}^{k} \sigma_i} \exp(-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\frac{x_{ij} - m_i}{\sigma_i^2})^2)$$

Metoda reprezentacyjna N- liczebność populacji, n - liczebność próby

- **lpzz** - losowanie proste ze zwracaniem
- **lpbz** - losowanie proste bez zwracania
- **lppxzz** - losowanie proste z prawd. ze zwracanie

- **ls** - losowanie systematyczne
  interwał losowania $q = \lfloor \frac{N}{n} \rfloor$ Losowy początek $R \in N : 1 \leqslant R \leqslant q$
  próba: $R, R+q, R+2q, R+3q, \ldots$

- **lppxbz** - losowanie proste z prawd. ze zwracanie (wg. schematu Hartleya-Rao)

j	X_j	$X_J^{(sk)}$	p_j	$p_j^{(sk)}$	π_j	$\pi_j^{(sk)}$	$X_{j'}$	$X_{j'}^{(sk)}$	$p_{j'}$	$p_{j'}^{(sk)}$	$\pi_{j'}$	$\pi_{j'}^{(sk)}$
1	X_1	$X_1^{(sk)}$	p_1	$p_1^{(sk)}$	π_1	$1_j^{(sk)}$	$X_{1'}$	$X_{1'}^{(sk)}$	$p_{1'}$	$p_{1'}^{(sk)}$	$\pi_{1'}$	$\pi_{1'}^{(sk)}$
2	X_2	$X_2^{(sk)}$	p_2	$p_2^{(sk)}$	π_2	$1_j^{(sk)}$	$X_{2'}$	$X_{2'}^{(sk)}$	$p_{2'}$	$p_{2'}^{(sk)}$	$\pi_{2'}$	$\pi_{2'}^{(sk)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$X_{n'}$	$X_{n'}^{(sk)}$	$p_{n'}$	1	$\pi_{n'}$	n'
n	X_n	$X_n^{(sk)}$	p_n	1	π_n	n						

$\pi_j = n \cdot p_j = n \frac{X_j}{\sum X_j}$
Jeżeli istnieją $X_j : \pi_j \geqslant 1$ to włączamy je do próby.
Z trójek $(u_{j'}, j', \pi_{j'})$ tworzymy ciąg monotoniczny wzgl. $u_{j'}$ i przeliczamy wszystko od nowa. Do próby włączamy $X_{j'} : \pi_{j'-1}^{(sk)} < u_{(0,1)} + (i-1) \leqslant \pi_{j'}^{(sk)} \wedge i = 0, 1, 2, \cdots, n'-1$
Jeżeli $\frac{n}{N} < 0,2$ to $\pi_{j,k} \cong \frac{n-1}{n}\pi_j\pi_k(A + \frac{1}{n}(\pi_j + \pi_k)) \wedge A = 1 - \frac{1}{n^2}\sum \pi_j^2$
Estymatory $T = \overline{Y} = \frac{q}{N}\sum Y_i$ (S - próba)

- **Horvitza - Thomsona** $\overline{y}_{HT} = \frac{1}{N}\sum_{j\in S}\frac{Y_j}{np_j}$

- **Hansena - Hurwitza** $\overline{y}_{HH} = \frac{1}{nN}\sum_{j\in S}\frac{Y_{Sj}}{p_{Sj}}$

Tw. 11 *Jeżeli* $\forall_j \pi_j > 0$ *to* $\overline{y}_{HT}$ *jest nieobciążonym estymatorem* $\overline{Y}$ *a jego wariancja*

$$D^2(\overline{y}_{HT}) = \frac{1}{N^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(\pi_j\pi_k - \pi_{jk})(\frac{Y_j}{\pi_j} - \frac{Y_k}{\pi_k})^2$$

Estymatory wariancji

Tw. 12 *Jeżeli* $\pi_j = \frac{n}{N} \wedge \pi_{jk} = \frac{n(n-1)}{N(N-1)}$ *[lpbz] to* $\overline{;iney_{HT} = \frac{1}{n}\sum_{i=1}^{n} y_i}$ *czyli zwykła średnia z próby i* $D^2(\overline{y}) = (1 - \frac{n}{N})\frac{S^2}{s(?))}$

Tw. 13 *Jeżeli* $\pi_{jk} > 0$ *to nieobciążonym estymatorem* D^2 *jest estymator Yales'a - Grandy'ego*

$$\hat{D}^2_{YG}(\overline{y}_{HT}(S)) = \frac{1}{N^2}\sum_{J<K\in S}\sum_{K\in S}(\frac{\pi_J\pi_K - \pi_{JK}}{\pi_{JK}})(\frac{Y_J}{\pi_J} - \frac{Y_K}{\pi_K})^2$$

[lpbz]: $\hat{D}^2_{YG}(\overline{y}_{HT}(S)) = (1 - \frac{n}{N})\frac{s^2}{n} \wedge s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$

Tw. 14 *Jeżeli [lppxzz] to* $\overline{y}_{HH}(S)$ *jest niebciążonym estymatorem* $\overline{Y}$ *a jego wariancja* $D^2(\overline{y}_{HH}(S)) = \frac{1}{n}\sum_{j=1}^{N} p_j(\frac{Y_j}{Np_j} - \overline{Y})^2$

Tw. 15 *Jeżeli* $\forall_j p_j = \frac{1}{N}$ *[lpzz] to* $\overline{y}_{HH} = \sum_{i=1}^{n} y_i \wedge D^2(\overline{y}_{HH}(S)) = \frac{\sigma^2}{n}$

Tw. 16 *Nieobciążony estymator* $\hat{D}^2(\overline{y}_{HH}(S)) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\frac{y_i}{Np_i} - \overline{y}_{HH})^2$ *[lpzz]:* $\hat{D}^2(\overline{y}) = \frac{s^2}{n}$

$q(j, i|s_{i-1})$ j-ciągnienie, i-jednostka, s_{i-2}-próba z poprzedniego ciągnienia
$p_i(j)$ - prawd. i-tego losu po raz 1 w i-tym ciągnieniu
$\pi_j = \sum_{i=1}^{n} p_i(j)$ - prawd. włączenia do próby
$p_j = \frac{X_J}{\sum X_J} = \frac{X_J}{N\overline{X}}$

Parametr	lpzz	lpbz	lppxzz
$q(j, i\|s_0)$	$\frac{1}{N}$	$\frac{1}{N}$	
$q(j, i\|s_{i-1})$	$\frac{1}{N}$	$\frac{1}{N-(i-1)}$	
$p_i(j)$	$\frac{1}{N}(1 - \frac{1}{N})^{i-1}$	$\frac{1}{N}$	$p_i(j) = p_j(1 - p_i)^{i-1}$??
π_j	$1 - (1 - \frac{1}{N})^n$	$\frac{n}{N}$	$1 - (1 - p_j)^n$
$\pi_{j,k}$	$1 - 2(1 - \frac{1}{N})^{n?} + (1 - \frac{2}{N})^{n?}$	$\frac{n(n-1)}{N(N-1)}$	$1 - (1 - p_j)^n - (1 - p_k)^n(1 - p_j - p_k)^n$