SdI30 W06: *ESTYMACJA PUNKTOWA PARAMETRÓW POPULACJI*

1. **Estymacja punktowa i estymator parametru**
2. **Własności estymatorów**

   Przykład 1

   Przykład 2

   Przykład 3
3. **Metoda momentów wyznaczania estymatorów**

   Przykład 4

   Przykład 5
4. **Metoda największej wiarygodności**

   Przykład 6

   Przykład 7

   Przykład 8

Przykład 9

5. **Estymatory podstawowych charakterystyk liczbowych**
6. **Szeregi: szczegółowy, pozycyjny i rozdzielczy**
7. **Zestaw zadań**

1. Estymacja punktowa i estymator parametru

Estymacją punktową (*point estimation*) nazywamy metody statystyczne, służące do punktowego oszacowania wartości nieznanego parametru rozkładu cechy w populacji.

Niech rozkład badanej cechy X populacji zależy od nieznanego parametru θ. Parametr ten będziemy estymowali na podstawie SRS $X_1, \dots, X_n$ pobranej z badanej populacji.

Estymatorem U_n nieznanego parametru θ rozkładu badanej cechy w populacji generalnej nazywamy każdą funkcję mierzalną próby losowej $U_n = h(X_1, X_2, \dots, X_n)$ – zwaną statystyką – służącą do oszacowania tego parametru.

Estymator U_n parametru θ oznaczamy $\hat{\theta}_n$.

Estymator jest zm. l. o rozkładzie zależnym od rozkładu zm. losowych tworzących próbę oraz od postaci funkcji h.

Oceną parametru θ nazywamy wartość liczbową $q_n = h(x_1, \dots, x_n)$ estymatora, otrzymaną na podstawie realizacji próby, tj. próbki $x_1, \dots, x_n$.

Ocena parametru prawie zawsze różni się od rzeczywistej wartości parametru θ.

Miarą błędu estymacji jest ***błąd szacunku***

$$d = \hat{\theta}_n - \theta.$$

Spośród wielu estymatorów parametru θ powinniśmy wybierać estymator o „dobrych" własnościach.

2. Własności estymatorów

Statystyka $\hat{\theta}_n$ jest dobrym estymatorem nieznanego parametru θ, jeżeli ma odpowiednie własności.

Tymi własnościami są:

- nieobciążoność lub asymptotyczna nieobciążoność,
- zgodność,
- efektywność,
- dostateczność.

- **Nieobciążoność.** Estymator $\hat{\theta}_n$ nazywamy *estymatorem nieobciążonym* parametru θ, jeśli
$$\mathbb{E}\hat{\theta}_n = \theta.$$

Uwaga. Jeśli cecha X populacji ma wartość oczekiwaną $m = \mathbb{E}X$ i wariancję $\sigma^2 = \mathbb{D}^2 X$, to estymatorami nieobciążonymi tych charakterystyk liczbowych są średnia arytmetyczna i wariancja z prostej próby losowej $X_1, \dots, X_n$.

Jeśli

$$b(\hat{\theta}_n) \overset{def.}{=} \mathbb{E}\hat{\theta}_n - \theta \neq 0$$

to estymator nazywamy ***estymatorem obciążonym***.

Różnicę $b(\hat{\theta}_n)$ nazywamy ***obciążeniem estymatora***.

- **Asymptotyczna nieobciążoność**

Estymator nazywamy ***asymptotycznie nieobciążonym***, gdy

$$\lim_{n\to\infty} b(\hat{\theta}_n) = 0.$$

Przykład 1. Niech $X_1, \ldots, X_n$ będzie SRS pobraną z populacji, w której cecha X ma skończoną i różną od zera wariancję σ^2 oraz $\mathbb{E}X = m$. Zbadać, czy statystyka

$$S_*^2 = \frac{1}{n}\sum_1^n (X_i - \overline{\boldsymbol{X}})^2, \text{ gdzie } \overline{\boldsymbol{X}} = \frac{1}{n}\sum_1^n X_i$$

jest estymatorem nieobciążonym nieznanej wariancji σ^2.

Rozwiązanie. Przekształcając S_*^2 otrzymujemy

$$S_*^2 = \frac{1}{n}\sum_1^n \left((X_i - m) + (m - \overline{\boldsymbol{X}})\right)^2$$

$$= \frac{1}{n}\sum_1^n (X_i - m)^2 - (\overline{\boldsymbol{X}} - m)^2$$

Ponieważ X_i są niezależnymi zmiennymi losowymi o tym samym rozkładzie co badana cecha X populacji, więc

$$\mathbb{E}(X_i - m)^2 = \mathbb{E}(X - m)^2 = \sigma^2 \text{ dla } i = 1, \ldots, n,$$

a na podstawie własności wariancji

$$\mathbb{E}(\bar{\boldsymbol{X}} - m)^2 = \mathbb{D}^2\bar{\boldsymbol{X}} = \mathbb{D}^2\left(\frac{1}{n}\sum_1^n X_i\right) = \frac{1}{n^2}\mathbb{D}^2\sum_1^n X_i = \frac{\sigma^2}{n}$$

Zatem

$$\mathbb{E}S_*^2 = \frac{1}{n}n\sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Zatem statystyka ta jest obciążona, ale nie asymptotycznie.

Ponieważ $\mathbb{E}S^2 = \sigma^2$, gdzie $S^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{\boldsymbol{X}})^2$, więc S^2 przyjmujemy za estymator nieznanej wariancji σ^2.

Zgodność. Estymator $\hat{\theta}_n$ nazywamy *estymatorem zgodnym* parametru θ, jeśli jest stochastycznie zbieżny do szacowanego parametru, tj. dla każdego $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) = 1$$

Jeśli rośnie liczebność próby, to rośnie prawd., przyjęcia przez estymator wartości coraz bliższych szacowanemu parametrowi. Tym samym zwiększając liczebność próby, zmniejszamy ryzyko popełnienia błędu.

Uwaga.

1. Z prawa wielkich liczb Czebyszewa wynika, że średnia arytmetyczna z próby jest zgodnym estymatorem wartości oczekiwanej w populacji generalnej, tzn.:
$$\lim_{n\to\infty} \mathrm{P}(|\bar{\boldsymbol{X}}_n - \mathbb{E}X| < \varepsilon) = 1$$
2. Jeśli estymator $\hat{\theta}_n$ parametru θ jest zgodny, to jest asymptotycznie nieobciążony. Tw. odwrotne nie jest prawdziwe.
3. Jeśli estymator $\hat{\theta}_n$ parametru θ jest nieobciążony (lub asymptotycznie nieobciążony) oraz jeśli jego wariancja spełnia warunek
$$\lim_{n\to\infty} \mathbb{D}^2\hat{\theta}_n = 0,$$
to $\hat{\theta}_n$ jest estymatorem zgodnym.

- **Efektywność**

Spośród wszystkich nieobciążonych estymatorów $U_{1,n}, U_{2,n}, \dots, U_{r,n}$ parametru θ, estymator o najmniejszej wariancji nazywamy ***estymatorem najefektywniejszym***.

Do wyznaczenia najefektywniejszego estymatora potrzebna jest znajomość wariancji wszystkich estymatorów nieobciążonych danego parametru.

W praktyce korzystamy z ***nierówności Rao-Cramera***.

http://pl.wikipedia.org/wiki/Nier%C3%B3wno%C5%9B%C4%87_Rao-Cram%C3%A9ra

Przykład 2. Zbadać, który z nieobciążonych estymatorów wartości oczekiwanej w populacji generalnej o dowolnym rozkładzie: średnia arytmetyczna, czy i-ta obserwacja X_i jest efektywniejszym estymatorem.

Rozwiązanie. Ponieważ

$$\mathbb{D}^2\overline{\boldsymbol{X}}_n = \frac{\mathbb{D}^2 X}{n} < \mathbb{D}^2 X_i = \mathbb{D}^2 X,$$

więc średnia arytmetyczna $\overline{\boldsymbol{X}}_n$ jest efektywniejszym estymatorem wartości oczekiwanej niż i-ta zmienna X_i z próby.

<u>Przykład 3.</u> Zbadać zgodność i efektywność empirycznego wskaźnika struktury $\bar{P}_n$, jako estymatora parametru p w rozkładzie Bernoulliego, $X \sim B(p)$.

Rozwiązanie. Niech $X_1, \dots, X_n$ będzie prostą próbą z populacji $X \sim B(p)$.

Ponieważ

$$\mathbb{E}\bar{P}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i = p$$

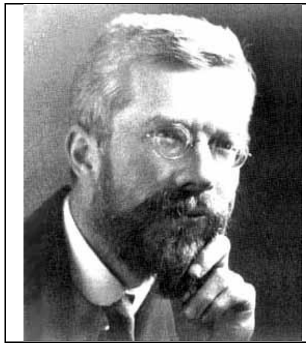
oraz

$$\mathbb{D}^2\bar{P}_n = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{D}^2X_i = \frac{p(1-p)}{n}$$

i $\bar{P}_n$ jest estymatorem o minimalnej wariancji, więc jest estymatorem zgodnym i najefektywniejszym dla parametru p.

- **Dostateczność**

Pojęcie dostateczności (wystarczalności) estymatora wprowadził Fisher[1]. Estymator dostateczny parametru θ to taki estymator, który skupia w sobie wszystkie informacje o tym parametrze, tzn. żaden inny estymator nie zawiera w sobie więcej informacji o parametrze θ wyciągniętej z próby losowej.



[1] **Ronald Aylmer Fisher** (1890-1962) – genetyk i statystyk brytyjski. Twórca podstaw współczesnej statystyki. Stworzył m.in. statystyczną metodę największej wiarygodności (ang. *maximum likelihood*), analizę wariancji (*ANOVA*) oraz liniową analizę dyskryminacyjną.

3. Metoda momentów wyznaczania estymatorów

Wprowadzona około roku 1900 przez K. Pearsona.

Polega na przyjmowaniu momentów z próby jako estymatorów odpowiednich momentów rozkładu cechy w populacji ogólnej.

Momenty są zazwyczaj funkcjami parametrów θ_i rozkładu.

Z otrzymanego układu równań wyznacza się estymatory parametrów.

Niech $X_1, \dots, X_n$ będzie SRS z populacji, w której badana cecha X ma pmf lub pdf $f(x)$.

Dla $k = 1, 2, 3, \dots$, wielkość $\mathbb{E}X^k$ jest k-tym momentem populacji, natomiast $\frac{1}{n}\sum_{i=1}^{n} X_i^k$ jest k-tym momentem z próby.

Stąd pierwszy moment populacji $\mathbb{E}X = m$, a pierwszy moment z próby $\bar{X}$.

Drugimi momentami populacji i próby są $\mathbb{E}X^2$ i $\frac{1}{n}\sum_{i=1}^{n} X_i^2$, odpowiednio.

Momenty populacji są funkcjami nieznanych parametrów.

Niech $X_1, \dots, X_n$ będzie SRS z populacji, w której badana cecha X ma pmf lub pdf $f(x; \theta_1, \dots, \theta_m)$, gdzie $\theta_1, \dots, \theta_m$ są parametrami, których wartości są nieznane.

Estymatory tych parametrów metodą momentów (MM) uzyskuje się poprzez porównanie pierwszych m momentów próby z odpowiadającymi im momentami populacji.

Przykład 4. Różnica wskazań dowolnych dwóch przyrządów pomiarowych jest zmienną losową o rozkładzie jednostajnym w przedziale (a, b). Oszacować metodą momentów końce przedziału.

Rozwiązanie. Ponieważ $X \sim u(a, b)$, więc

$$\mathbb{E}X = \frac{1}{2}(a+b), \quad \mathbb{D}X = \frac{1}{2\sqrt{3}}(b-a),$$

Zastępując zgodnie z metodą momentów $\mathbb{E}X$ przez $\bar{\boldsymbol{X}}_n$ i $\mathbb{D}X$ przez S otrzymujemy estymatory

$$a = \bar{\boldsymbol{X}}_n - S\sqrt{3} \text{ oraz } b = \bar{\boldsymbol{X}}_n + S\sqrt{3}.$$

Przykład 5. Niech badana cecha X ma rozkład gamma o gęstości

$$f(x|\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \mathbb{I}_{(0,\infty)}(x)$$

z nieznanymi parametrami $\alpha, \beta > 0$.

Wyznaczyć metodą momentów estymatory $\hat{\alpha}, \hat{\beta}$ parametrów α, β.

Następnie zastosować te estymatory do oceny tych parametrów na podstawie danych:

152	115	109	94	88	137	152	77	160	165
125	40	128	123	136	101	62	153	83	69

Rozwiązanie. Pierwsze dwa momenty zwykłe tego rozkładu dane są wzorami:

$$m_1 = \alpha\beta, m_2 = \alpha(\alpha+1)\beta^2$$

Stąd na podstawie n-elementowej próby uzyskujemy równania

$$\alpha\beta = \bar{X},\ \alpha(\alpha+1)\beta^2 = \frac{1}{n}\sum_1^n X_i^2$$

Wyznaczając z tych równań α i β, otrzymujemy estymatory

$$\hat{\alpha} = \frac{\bar{X}^2}{\frac{1}{n}\sum X_1^2 - \bar{X}^2},$$

$$\hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - \bar{X}^2}{\bar{X}}$$

Obliczenia ocen parametrów na podstawie danych

$\bar{x} = 113{,}5$ oraz $\frac{1}{20}\sum x_i^2 = 14087{,}8$

stąd

$$\hat{\alpha} = \frac{(113{,}5)^2}{14087{,}8 - (113{,}5)^2} = 10{,}7$$

$$\hat{\beta} = \frac{14087{,}8 - (113{,}5)^2}{113{,}5} = 10{,}6$$

4. Metoda największej wiarygodności

Opracowana przez R. A. Fishera. Jest efektywniejsza od innych metod. Niech rozkład badanej cechy X zależy od k nieznanych parametrów $\theta_1, \dots, \theta_k$, które chcemy oszacować.

Krok 1. Wyznaczamy funkcję wiarygodności próby:

$$L(x_1, \dots, x_n | \theta_1, \dots, \theta_k) = \prod_{i=1}^{n} f(x_i | \theta_1, \dots, \theta_k)$$

gdzie f oznacza PDF dla rozkładu typu ciągłego lub PMF dla rozkładu typu dyskretnego.

Krok 2. Za estymatory parametrów przyjmujemy $\hat{\theta}_1, \dots, \hat{\theta}_k$, dla których L (lub $\ln L$) przyjmuje wartość największą

Wartości maksymalizujące muszą spełniać układ równań

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \quad \text{dla} \quad i = 1, \dots, k$$

Krok 3. Sprawdzamy warunek konieczny i wystarczający dla maksimum funkcji. W szczególności dla $k = 1$ oznacza to, że druga pochodna w punkcie $\hat{\theta}$ jest ujemna.

Przykład 6. Cecha X pewnej populacji ma rozkład trzypunktowy z nieznanym parametrem p

$$f = \begin{pmatrix} -1 & 0 & 1 \\ 0{,}5 - p & 0{,}5 & p \end{pmatrix}$$

Wyznaczyć estymator parametru p

a) metodą momentów,
b) metodą największej wiarygodności.

Rozwiązanie. Niech $x_1, x_2, \dots, x_n$ będzie realizacją próby prostej.

a) W metodzie momentów wyznaczamy wartość oczekiwaną

$$m = \mathbb{E}X = 2p - \frac{1}{2}$$

czyli $p = (m + 1/2)/2$. Wstawiając moment empiryczny otrzymujemy estymator parametru p

$$\hat{p}_n = \frac{\bar{\boldsymbol{X}}_n + \frac{1}{2}}{2}$$

b) Dla uproszczenia zapisu niech k oznacza liczbę obserwacji przyjmujących wartość -1, a l – liczbę obserwacji przyjmujących wartość 0.

Funkcja wiarygodności ma postać:

$$L(x_1, x_1, \dots, x_n|p) = \mathrm{P}(X_1 = x_1) \cdot \dots \cdot \mathrm{P}(X_n = x_n) =$$
$$L(k, l, n|p) = (0{,}5 - p)^k (0{,}5)^l p^{n-k-l}$$

L osiąga maksimum w tym samym punkcie co funkcja $\ln L$.

$$\ln L(k, l, n|p) = k \ln(0{,}5 - p) + l \ln 0{,}5 + (n - k - l) \ln p$$

Funkcja $\ln L$ jest różniczkowalna względem p

$$\frac{d \ln L(k, l, n|p)}{dp} = 0$$

$$\frac{-k}{0{,}5-p}+\frac{n-k-l}{p}=0$$

Stąd

$$p=\frac{n-k-l}{2(n-l)}$$

Ostatecznie estymator wyraża się wzorem:

$$\hat{p}=\frac{n-U_1-U_0}{2(n-U_0)}$$

gdzie U_1 i U_0 są statystykami zliczającymi wystąpienia odpowiednio wartości -1 i 0 (k i l są realizacjami tych statystyk).

Przykład 7. Wyznaczyć estymator parametru rozkładu wykładniczego

a) metodą MM;

b) metodą MNW.

Rozwiązanie b) Przyjmijmy, że dysponujemy n-elementową próbą z rozkładu wykładniczego $\mathrm{Exp}(\lambda)$.

Poszukujemy estymatora funkcji parametrycznej $h(\lambda) = \frac{1}{\lambda}$.

Gęstość rozpatrywanego rozkładu ma postać:

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \lambda > 0, x > 0$$

Kolejne kroki obliczeń są następujące:

$$L(\lambda;\ x_1, \dots, x_n) = \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda x_i} = \lambda^n \mathrm{e}^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ln L = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial \lambda} \ln L = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\lambda = \frac{1}{\bar{x}}$$

$$\left. \frac{\partial^2}{\partial \lambda^2} \ln L \right|_{\lambda = \frac{1}{\bar{x}}} = -\frac{n}{\bar{x}} < 0$$

A zatem $\hat{\lambda} = \frac{1}{\bar{X}}$ jest estymatorem największej wiarygodności dla parametru λ.

Ostatecznie otrzymujemy:

$$\widehat{h(\lambda)} = \frac{1}{\hat{\lambda}} = \bar{X}$$

Otrzymany estymator nie jest nieobciążony, gdyż

$$\mathbb{E}\left(\frac{1}{\bar{X}}\right) \neq \frac{1}{\mathbb{E}\bar{X}}$$

Przykład 8. Niech $X_1, \dots, X_n$ będzie SRS z rozkładu Weibulla o gęstości

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \mathbf{1}_{[0,\infty)}(x)$$

Zapisując funkcję wiarygodności oraz jej logarytm i rozwiązując układ równań

$$\begin{cases} \dfrac{\partial}{\partial \alpha}[\ln L] = 0 \\ \dfrac{\partial}{\partial \beta}[\ln L] = 0 \end{cases}$$

otrzymamy

$$\alpha = \left[\frac{\sum x_i^\alpha \ln x_i}{\sum x_i^\alpha} - \frac{\sum \ln x_i}{n}\right]^{-1}, \beta = \left(\frac{\sum x_i^\alpha}{n}\right)^{1/\alpha}$$

Równań tych nie można wprost rozwiązać, aby otrzymać MLE estymatory $\hat{\alpha}$ i $\hat{\beta}$. Równania muszą być rozwiązane stosując numeryczną procedurę iteracyjną.

Do oceny parametrów na podstawie próby za pomocą MNW w języku $\mathcal{R}$ służy funkcja **fitdistr** z pakietu MASS. Działa ona dla prawie wszystkich opisanych wcześniej rozkładów (poza: jednostajnym, dwumianowym i hipergeometrycznym).

Przykład 9. Dla danych zawartych w zbiorze precip, zawierających informacje na temat ilości opadów dla 70 miast USA (1975 rok) wyznaczyć oceny parametrów rozkładu normalnego.

Rozwiązanie w języku $\mathcal{R}$.

fitdistr(precip,’normal’)#oceny parametrów oraz ich błędy standardowe

```
-----------------
mean           sd
34.885714      13.608393
(1.626514)     (1.150119)
--------------------
```

Znacznie większe możliwości ma pakiet **fitdistrplus**, który pozwala estymować parametry za pomocą MNW oraz MM (funkcja **fitdist**). Dla otrzymanego obiektu można użyć przeciążonych funkcji **plot** oraz **summary**, dostając w zależności od typu rozkładu (ciągły, dyskretny) wykresy diagnostyczne oraz wyniki testów statystycznych. Najciekawsza wydaje się funkcja **descdist**, która graficznie przedstawia dopasowanie rozkładu, oceniając skośność oraz kurtozę. Str. 197

5. Estymatory podstawowych charakterystyk liczbowych

A. Estymator wartości oczekiwanej. Średnia arytmetyczna jest estymatorem nieobciążonym i jednocześnie estymatorem największej wiarygodności wartości oczekiwanej zm. l. X przy spełnieniu przynajmniej jednego z poniższych założeń:

- liczba obserwacji n jest dostatecznie duża (zob. CTG),
- rozkład zmiennej X jest normalny.

B. Estymator wariancji. Jeżeli wartość oczekiwana m_X populacji X jest nieznana, to estymatorem nieobciążonym nieznanej wariancji w populacji X jest wariancja z próby, tj.

$$\hat{\sigma}_X^2 = S_n^2.$$

Jeżeli wartość oczekiwana m_X populacji X jest znana, to estymatorem nieobciążonym nieznanej wariancji dla danych szczegółowych jest statystyka S_n^2 określona wzorem:

$$S_n^2(\mathbf{X}, m_X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_X)^2$$

C. Estymator wskaźnika struktury. ***Wskaźnikiem struktury*** w populacji $X \sim B(p)$ nazywamy prawd. p zaobserwowania wyróżnionej cechy w populacji. Estymatorem wskaźnika p jest częstość w próbie, tj. $\hat{p} = \bar{P}_n$, gdzie

$$\bar{P}_n = \frac{1}{n}\sum_1^n X_i$$

$X_i \sim B(p)$, n jest licznością próby.

6. Szeregi: szczegółowy, pozycyjny i rozdzielczy

Niech $x_1, x_2, \dots, x_n$ będzie n-elementową próbą. Liczbę n jednostek wybranych do próby nazywamy ***liczebnością próby***.

Jeżeli $n \leq 30$, to próbę nazywamy ***małą próbą***.

Dane uporządkowane w ciąg niemalejący

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

tworzą szereg pozycyjny.

Jeżeli $n > 30$, to w celu ułatwienia analizy, dane grupuje się w klasy, tzn. przedziały, najczęściej jednakowej długości, przyjmując upraszczające założenie, że wszystkie wartości znajdujące się w danej klasie są reprezentowane przez środek klasy.

Ustalenie liczby klas k zależy od liczby obserwacji n.

W literaturze podaje się różne sposoby ustalania liczby klas, np. dowolna liczba k spełniająca warunek:

$$\max\left\{5, \frac{\sqrt{n}}{2}\right\} \leq k \leq \min\{30, \sqrt{n}\}$$

Liczbę wartości próby zawartych w i-tej klasie nazywamy liczebnością i-tej klasy i oznaczamy n_i.

Reprezentant klasy $\dot{x}_i$ oraz liczebność n_i dla $i = 1,2, \dots, m$ tworzą ciąg par

$$(\dot{x}_i, n_i), i = 1, 2, \dots, k$$

zwany **szeregiem rozdzielczym**.

Wielkość

$$n_i^* = n_1 + n_2 + \cdots + n_i$$

nazywa się **liczebnością skumulowaną** i-tej klasy.

Wielkość $\frac{n_i}{n}$ nazywa się **częstością** i-tej klasy, a $\frac{n_i^*}{n}$ **częstością skumulowaną** i-tej klasy.

Pary $(\dot{x}_i, n_i^*), i = 1, 2, \dots, k,$ tj. środki kolejnych klas $\dot{x}_i$ oraz ich liczebności skumulowane nazywamy **szeregiem rozdzielczym skumulowanym**.

Jeśli cecha jest typu dyskretnego, a liczba możliwych wartości jest bardzo duża, wtedy możemy postąpić podobnie jak w przypadku cechy typu ciągłego.

Średnią dla danych w postaci szeregu rozdzielczego nazywamy ***średnią ważoną*** i wyznaczamy ze wzoru:

$$\overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{i=1}^{k} \dot{x}_i n_i$$

gdzie $\dot{x}_i$ to liczba reprezentująca i-tą klasę, zaś n_i to liczebność i-tej klasy.

Wariancję dla danych w postaci szeregu rozdzielczego nazywamy wariancją ważoną i wyznaczamy ze wzoru:

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i(\dot{x}_i - \overline{\boldsymbol{x}}_n)^2$$

7. Zestaw zadań W06

1. Wyznaczyć estymator parametru p w rozkładzie *Bernoulliego*.

2. Wyznaczyć MM oraz MNW estymatory parametrów rozkładu normalnego.

3. Wyznaczyć MNW estymator parametru rozkładu Poissona.

4. Celem sprawdzenia dokładności wskazań pewnego przyrządu pomiarowego dokonano 10 pomiarów tej samej wielkości fizycznej X i otrzymano następujące wyniki:

9,01; 9,00; 9,02; 8,99; 8,98; 9,00; 9,00; 9,01; 8,99; 9,00.

Dokonać przekształcenia pomiarów według wzoru:

$$Y = 100(X - 9).$$

Dla wielkości X i Y oszacować ich wartości oczekiwane i wariancje.

5. Wygenerować 50 elementową próbę prostą z populacji, w której cecha X ma rozkład o gęstości

$$f(x) = \frac{x}{8}\mathbf{1}_{(0;\ 4)}(x)$$

a) Sporządzić histogram.

b) Wyznaczyć wartość oczekiwaną i wariancję oraz ich oceny na podstawie wygenerowanej próby.

6. Korzystając z dostępnego oprogramowania wygenerować 100 elementową próbę według rozkładu

a) $\mathrm{bin}(20;\ 0{,}8)$,

b) $\mathrm{nbin}(3; 0{,}1)$,

c) Poisson(5).

Sporządzić histogram i dokonać ocenę punktową parametrów.

7. Wygenerować 100 elementową próbę według rozkładu logarytmiczno-normalnego z parametrami $\mu = 2{,}3$ i $\sigma = 0{,}5$.

a) Sporządzić histogram.

b) Dokonać estymacji parametrów, ocenić wartość oczekiwaną i wariancję oraz porównać te wartości z wartościami teoretycznymi.