1 Błąd średniokwadratowy

W tej części zadań zajmiemy się błędem średniokwadratowym.

1. (Eg 48/9) Zmienne losowe $X_1, X_2, ..., X_n$, $n > 2$ są niezależne i $\mathbf{E}X_i = m$ oraz $\mathbf{Var}X_i = \frac{m^2}{i}$, $i = 1, 2, ..., n$, gdzie m jest nieznanym parametrem rzeczywistym. Niech $\bar{m}$ będzie estymatorem parametru m minimalizującym błąd średniokwadratowy w klasie estymatorów postaci

$$\hat{m} = \sum_{i=1}^{n} a_i X_i,$$

gdzie a_i, $i = 1, 2, ...n$, są liczbami rzeczywistymi. Wtedy współczynniki a_i są równe
Odp: D-> $a_i = \frac{2i}{n^2+n+2}$, $i = 1, 2, ..., n$.

Rozwiązanie. Obliczamy błąd średniokwadratowy

$$\mathbf{E}(m - \hat{m})^2 = (\mathbf{E}\hat{m} - m)^2 + \mathbf{Var}\hat{m} = m^2[(\sum_{i=1}^{n} a_i - 1)^2 + \sum_{i=1}^{n} \frac{a_i^2}{i}].$$

Najlepsze a_i wyznaczamy z warunku na pochodne

$$(\sum_{i=1}^{n} a_i - 1) = \frac{a_i}{i}, \quad ß = 1, 2, ..., n.$$

Stąd również

$$\frac{n(n+1)}{2}(\sum_{i=1}^{n} a_i - 1) = (\sum_{i=1}^{n} a_i),$$

co dalej oznacza

$$\sum_{i=1}^{n} a_i = \frac{n(n+1)}{n^2+n-2}, \quad \sum_{i=1}^{n} a_i - 1 = \frac{2}{n^2+n-2}.$$

Otrzymujemy

$$a_i = \frac{2i}{n^2+n-2}, \quad i = 1, 2, ..., n.$$

■

2. (Eg 49/8) Niech $X_1, X_2, ..., X_n$, $n \geqslant 2$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym o wartości oczekiwanej 1 i nieznanej wariancji σ^2. Rozważamy rodzinę estymatorów parametru σ postaci $S_a = a\sum_{i=1}^{n} |X_i - 1|$, przy czym a jest liczbą dodatnią. Wyznaczyć a^*, tak aby estymator S_{a^*} był estymatorem o najmniejszym błędzie średniokwadratowym wśród estymatorów postaci S_a.
Odp: D-> $a^* = \frac{\sqrt{2\pi}}{2n+\pi-2}$.

Rozwiązanie. Niech X ma rozkład $\mathcal{N}(1, \sigma^2)$. Należy wyznaczyć błąd średniokwadratowy estymatorów S_a, to znaczy

$$f(a) = \mathbf{E}(S_a - \sigma)^2 = \mathbf{E}(a\sum_{i=1}^{n}(|X_i - 1| - \mathbf{E}|X_i - 1|) + an\mathbf{E}|X - 1| - \sigma)^2 = a^2 n\mathbf{Var}|X - 1| + (an\mathbf{E}|X - 1| - \sigma)^2.$$

Oczywiście $\mathbf{E}|X - 1| = \frac{\sqrt{2}}{\sqrt{\pi}}\sigma$ oraz

$$\mathbf{Var}|X - 1| = \mathbf{E}(X - 1)^2 - (\mathbf{E}|X - 1|)^2 = \sigma^2 - \frac{2}{\pi}\sigma^2.$$

Zatem

$$f(a) = a^2 n(1 - \frac{2}{\pi})\sigma^2 + (an\frac{\sqrt{2}}{\sqrt{\pi}} - 1)^2\sigma^2.$$

Znajdujemy punkt minimum tej funkcji, czyli a^* z równania $f'(a) = 0$ czyli

$$f'(a) = 2an(1 - \frac{2}{\pi})\sigma^2 + 2(an\frac{\sqrt{2}}{\sqrt{\pi}} - 1)n\frac{\sqrt{2}}{\sqrt{\pi}}\sigma^2 = 0$$

co jest równoważne

$$(1 + (n-1)\frac{2}{\pi})a = \frac{\sqrt{2}}{\sqrt{\pi}}.$$

Zatem $a = \frac{\sqrt{2\pi}}{2n-2+\pi}$.

■

3. (Eg 50/4) Rozpatrzmy następujący model regresji liniowej bez wyrazu wolnego:

$$Y_i = \beta \cdot x_i + \varepsilon_i, \ \ (i = 1, 2, ..., 16),$$

gdzie $x_i > 0$ są znanymi liczbami, β jest nieznanym parametrem, zaś ε_i są błędami losowymi. Zakładamy, że ε_i są niezależnymi zmiennymi losowymi o rozkładach normalnych i

$$\mathbf{E}[\varepsilon_i] = 0 \ \text{ i } \ \mathbf{Var}[\varepsilon_i] = x_i^2, \ \ (i = 1, 2, ..., 16).$$

Niech $\bar{\beta}$ będzie estymatorem parametru β o następujących własnościach:

- $\bar{\beta}$ jest liniową funkcją obserwacji, tzn. jest postaci $\bar{\beta} = \sum_{i=1}^{16} c_i Y_i$
- $\bar{\beta}$ jest nieobciążony,
- $\bar{\beta}$ ma najmniejszą wariancję spośród estymatorów liniowych i nieobciążonych.

Wyznaczyć stałą c taką, że spełniony jest warunek

$$\mathbf{P}(|\bar{\beta} - \beta| < c) = 0,95.$$

Odp: A-> $c = 0,49$.

Rozwiązanie. Zmienne Y_i są niezależne o rozkładach $\mathcal{N}(\beta \cdot x_i, x_i^2)$. Z postaci gęstości (albo po prostu ze wzoru na postać dla rozkładów wykładniczych) obliczamy statystyki dostateczne i zupełne $\sum_{i=1}^n \frac{Y_i}{x_i}$, $\sum_{i=1}^n \frac{Y_i^2}{x_i^2}$. Nietrudno zauważyć, że $n^{-1}\sum_{i=1}^n \frac{Y_i}{x_i}$ jest nieobciążonym estymatorem β opartym na statystyce zupełnej i dostatecznej. Na mocy twierdzenia Rao-Blackwell'a spełnia on wszystkie postulaty wymienione w zadaniu. Obliczamy dla $n = 16$

$$\mathbf{P}(|\sum_{i=1}^{16}(\frac{Y_i}{x_i} - \beta)| < 16c) = \mathbf{P}(|\sum_{i=1}^{n}\frac{\varepsilon_i}{x_i}| < 16c).$$

Oczywiście $\sum_{i=1}^n \frac{\varepsilon_i}{x_i}$ ma rozkład taki sam jak $4Z$, gdzie $Z \sim \mathcal{N}(0,1)$. Stąd

$$\mathbf{P}(|\sum_{i=1}^{n}\frac{\varepsilon_i}{x_i}| < 16c) = \mathbf{P}(|Z| < 4c)$$

Obliczamy z tablic $4c = 1,96$, czyli $c = 0,49$.

■

4. (Eg 51/9) Niech $X_1, X_2, ...X_n$, $n > 1$ będzie próbką z rozkładu Poissona z nieznanym parametrem λ (parametr jest wartością oczekiwaną pojedynczej obserwacji, $\lambda = \mathbf{E}_\lambda X_i > 0$. Interesuje nas drugi moment obserwacji, czyli wielkość $m_2(\lambda) = \mathbf{E}_\lambda(X_i^2)$. Estymator nieobciążony o minimalnej wariancji funkcji $m_2(\lambda)$ jest równy
Odp: E-> $\frac{1}{n^2}((\sum_{i=1}^n X_i)^2 + (n-1)\sum_{i=1}^n X_i)$.

Rozwiązanie. Ponownie poszukujemy statystyki dostatecznej i zupełnej. którą jest w tym przypadku $T = \sum_{i=1}^n X_i$. Estymator ENMW dla $m_2(\lambda)$ jest funkcją T. Możemy go albo zwyczajnie zgadnąć (w przypadku tego zadania zauważając, że będzie to funkcja kwadratowa) albo obliczając ze wzoru $\mathbf{E}(X_i^2|T)$. Zachodzi

$$m_2(\lambda) = \lambda^2 + \lambda, \;\; \mathbf{E}T = n\lambda, \;\; \mathbf{E}T^2 = (n\lambda)^2 + n\lambda.$$

Z tych danych znajdujemy, że właściwy estymator ma postać

$$\hat{m}_2(\lambda) = \frac{T^2}{n^2} + (n-1)\frac{T}{n^2}.$$

■

5. (Eg 52/4) Zakładamy, że zależność czynnika Y od czynnika x (nielosowego) opisuje model regresji liniowej $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Obserwujemy $2n$ elementową próbkę, w której $x_1 = x_2 = ... = x_n = -1$ i $x_{n+1} = x_{n+2} = ... = x_{2n} = 1$. Zmienne losowe $Y_1, Y_2, ..., Y_{2n}$ są niezależne i błędy mają rozkłady normalne o wartości oczekiwanej 0, przy czym $\mathbf{Var}\varepsilon_i = \sigma^2$, gdy $i = 1, 2, ..., n$ i $\mathbf{Var}\varepsilon_i = 9\sigma^2$, gdy $i = n+1, ..., 2n$. Wyznaczono estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$ parametrów β_0 i β_1 wykorzystując ważoną metodę najmniejszych kwadratów, to znaczy minimalizując sumę $\sum_{i=1}^{2n} \frac{(Y_i - \beta_0 - \beta_1 x_i)^2}{\mathbf{Var}\varepsilon_i}$. Wyznacz stałą z_1 tak aby

$$\mathbf{P}(|\bar{\beta}_1 - \beta_1|\sqrt{n} < z_1\sigma) = 0,95.$$

Spośród podanych odpowiedzi wybierz odpowiedź będącą najlepszym przybliżeniem.
Odp: E-> $z_1 = 3,099$.

Rozwiązanie. Wyznaczamy $\hat{\beta}_0$ i $\hat{\beta}_1$

$$\begin{cases} \sum_{i=1}^{2n} \frac{Y_i - \beta_0 - \beta_1 x_i}{\mathbf{Var}\varepsilon_i} &= 0 \\ \sum_{i=1}^{2n} \frac{x_i(Y_i - \beta_0 - \beta_1 x_i)}{\mathbf{Var}\varepsilon_i} &= 0 \end{cases}$$

Zatem

$$\hat{\beta}_0 = \frac{1}{2n}(\sum_{i=1}^{2n} Y_i), \;\; \hat{\beta}_1 = \frac{1}{2n}(-\sum_{i=1}^{n} Y_i + \sum_{i=n+1}^{2n} Y_i).$$

Znajdujemy rozkład $(\bar{\beta}_1 - \beta_1)/\sqrt{n}$ jako $\frac{\sqrt{10}\sigma}{2}Z$, gdzie Z ma rozkład normalny $\mathcal{N}(0,1)$. Stąd

$$\mathbf{P}(|\bar{\beta}_1 - \beta_1|\sqrt{n} < z_1\sigma) = \mathbf{P}(|Z| < \frac{2z_1}{\sqrt{10}}).$$

Czyli $\frac{2z_1}{\sqrt{10}} = 1,96 \simeq 3,099$.

■

6. (Eg 53/8) W pewnej populacji prawdopodobieństwo tego, że osobnik przeżyje pierwszy rok jest równe $(1 - \theta^2)$. Jeżeli osobnik przeżył pierwszy rok, to prawdopodobieństwo warunkowe tego, że przeżyje następny rok jest równe $\frac{2\theta}{1+\theta}$. W próbce losowej liczącej n osobników z tej populacji zanotowano:

- n_0 przypadków, kiedy osobnik nie przeżył pierwszego roku
- n_1 przypadków, kiedy osobnik przeżył pierwszy rok, ale nie przeżył drugiego roku,
- n_2 przypadków, kiedy osobnik przeżył dwa lata.

Błąd średniokwadratowy estymatora największej wiarogodności parametru θ wyraża się wzorem:
Odp: C-> $\frac{\theta(1-\theta)}{2n}$.

Rozwiązanie. Obliczamy rozkłady. Dla $i = 1, ..., 10$, zmienna $X_i - \bar{X}$ ma rozkład $\mathcal{N}(\frac{(\mu_1-\mu_2)}{3}, \frac{41}{45}\sigma^2)$ oraz $\mathcal{N}(-\frac{2(\mu_1-\mu_2)}{3}, \frac{122}{45})$. Nadto $\bar{X}_1$ ma rozkład $\mathcal{N}(\mu_1, \frac{\sigma^2}{10})$, a $\bar{X}_2$ rozkład $\mathcal{N}(\mu_2, \frac{3\sigma^2}{5})$ stąd też $\bar{X}_1 - \bar{X}_2$ ma rozkład $\mathcal{N}(\mu_1 - \mu_2, \frac{7}{10}\sigma^2)$. Zatem

$$\mathbf{E}\bar{\sigma}^2 = 10a(\frac{1}{9}(\mu_1 - \mu_2)^2 + \frac{44}{45}\sigma^2) + 5a(\frac{4}{9}(\mu_1 - \mu_2)^2 + \frac{122}{45}\sigma^2) +$$
$$+ b((\mu_1 - \mu_2)^+ \frac{9}{10}\sigma^2).$$

Co oznacza, że $b = -\frac{10}{3}a$, nadto $\frac{70}{3}a - \frac{7}{10}b = 1$, czyli $a = \frac{3}{63} = \frac{1}{21}$.

■

7. (Eg 54/5) Przeprowadzamy wśród wylosowanych osób ankietę na delikatny temat. Ankietowana osoba rzuca kostką do gry, i w zależności od wyniku rzutu kostką (wyniku tego nie zna ankieter) podaje odpowiednio zakodowaną odpowiedź na pytanie:

'Czy zdarzyło się Panu/Pani w roku 2009 dać łapówką w klasycznej formie pieniężnej przekraczającą 100 zł'

- $X = 1$ jeśli odpowiedź brzmi 'TAK',
- $X = 0$ jeśli odpowiedź brzmi 'NIE',

Pierwszych 200 **osób** udziela odpowiedzi $Z_1, ..., Z_{200}$ zgodnie z regułą:

- jeśli wyniku rzutu kostką to liczba oczek równa $1, 2, 3$ lub 4, to:

$$Z_i = X_i$$

- jeśli wynik rzutu kostką to liczba oczek równa 5 lub 6, to:

$$Z_i = 1 - X_i$$

Następnych 200 **osób** udziela odpowiedzi $Z_{201}, ..., Z_{400}$ zgodnie z regułą:

- jeśli wyniku rzutu kostką to liczba oczek równa 1 lub 2, to:

$$Z_i = X_i$$

- jeśli wynik rzutu kostką to liczba oczek równa $3, 4, 5$ lub 6, to:

$$Z_i = 1 - X_i$$

Dla uproszczenia zakładamy, że 400 ankietowanych osób to próba prosta z (hipotetycznej) populacji o nieskończonej liczebności, a podział na podpróby jest całkowicie losowy. Interesujący nas parametr tej populacji to oczywiście $q = \mathbf{P}(X = 1)$. Niech

$$\bar{Z}_1 = \frac{1}{200}\sum_{i=1}^{200} Z_i, \quad \bar{Z}_2 = \frac{1}{200}\sum_{i=201}^{400} Z_i.$$

Estymator parametru q uzyskany metodą największej wiarygodności jest równy
Odp: D-> $\frac{1}{2} + \frac{3}{2}\bar{Z}_1 - \frac{3}{2}\bar{Z}_2$.

Rozwiązanie. Oczywiście główne zadanie to ustalić rozkład Z_i dla $i = 1, 2, ..., 400$. Zachodzi

$$\begin{array}{rcl} \mathbf{P}(Z_i = 1) & = & \frac{2q}{3} + \frac{(1-q)}{3} \\ \mathbf{P}(Z_i = 0) & = & \frac{2(1-q)}{3} + \frac{q}{3} \end{array} , \text{ dla } i = 1, 2, ..., 200$$

oraz

$$\begin{array}{rcl} \mathbf{P}(Z_i = 1) & = & \frac{q}{3} + \frac{2(1-q)}{3} \\ \mathbf{P}(Z_i = 0) & = & \frac{(1-q)}{3} + \frac{2q}{3} \end{array} , \text{ dla } i = 201, 202, ..., 400.$$

Możemy obliczyć wiarygodność

$$L(q, k) = \left(\frac{(1+q)}{3}\right)^{200(1+\bar{k}_1-\bar{k_2})} \left(\frac{2-q}{3}\right)^{200(1-\bar{k}_1+\bar{k}_2)},$$

dla $\bar{k}_1 = \frac{1}{200}\sum_{i=1}^{200} k_i$, $\bar{k}_2 = \frac{1}{200}\sum_{i=201}^{400} k_i$. Obliczamy pochodną funkcji $f(q) = \log L(q, k)$

$$f'(q) = 200(1 + \bar{k}_1 - \bar{k_2}) \cdot \frac{1}{1+q} - 200(1 - \bar{k}_1 + \bar{k}_2) \cdot \frac{1}{2-q}.$$

Z warunku $f'(q) = 0$ odczytujemy $q = \frac{1}{2} + \frac{3\bar{k}_1}{2} - \frac{3\bar{k}_2}{2}$ stąd ENW parametru q ma postać $\frac{1}{2} + \frac{3}{2}\bar{Z}_1 - \frac{3}{2}\bar{Z}_2$.

■

8. (Eg 55/10) Niech X oznacza zmienną losową równą liczbie sukcesów w n ($n \geqslant 2$) niezależnych próbach Bernoulliego. Prawdopodobieństwo sukcesu θ, ($\theta \in (0, 1)$) jest nieznane. Rozważamy estymator parametru θ postaci $\bar{\theta} = aX + b$, o wartościach nieujemnych, którego błąd średniokwadratowy jest stały niezależny od wartości parametru θ. Błąd średniokwadratowy tego estymatora jest równy
Odp: D-> $\frac{1}{4(\sqrt{n}+1)^2}$.

Rozwiązanie. Obliczamy

$$\mathbf{E}(\theta - aX - b)^2 = \mathbf{Var}(aX) + ((an-1)\theta + b)^2 = na^2\theta(1-\theta) + (an-1)^2\theta^2 + 2(an-1)b\theta + b^2.$$

Zatem $na^2 = (1-an)^2$, $na^2 = -2(an-1)b$, czyli $-\sqrt{n}a = an - 1$, $a = \frac{1}{n+\sqrt{n}}$, $b = \frac{1}{2(\sqrt{n}+1)}$. Stąd

$$\mathbf{E}(\theta - aX - b)^2 = \frac{1}{4(\sqrt{n}+1)^2}.$$

■

9. (Eg 56/1) Zakładamy, że $X_1, X_2, ..., X_{10}, X_{11}, X_{12}..., X_{15}$ są niezależnymi zmiennymi losowymi o rozkładach normalnych, przy czym $\mathbf{E}X_i = \mu_1$ i $\mathbf{Var}X_i = \sigma^2$ dla $i = 1, 2, ..., 10$, oraz $\mathbf{E}X_i = \mu_2$ i $\mathbf{Var}X_i = 3\sigma^2$ dla $i = 11, 12, ..., 15$. Parametry μ_1, μ_2 i σ są nieznane. Niech $\bar{X}_1 = \frac{1}{10}\sum_{i=1}^{10} X_i$, $\bar{X}_2 = \frac{1}{5}\sum_{i=11}^{15} X_i$, $\bar{X} = \frac{1}{15}\sum_{i=1}^{15} X_i$. Dobrać stałe a i b tak, aby statystyka

$$\bar{\sigma}^2 = a\sum_{i=1}^{15}(X_i - \bar{X})^2 + b(\bar{X}_1 - \bar{X}_2)^2$$

była estymatorem nieobciążonym parametru σ^2.
Odp: A-> $a = \frac{1}{21}$, $b = -\frac{10}{63}$.

Rozwiązanie. Estymator nieobciążony oznacza, że niezależnie od μ_1 i μ_2

$$\mathbf{E}\hat{\sigma}^2 = \sigma^2.$$

Zauważmy, że $X_i = \sigma\hat{X}_i + \mu_i$, $i \in \{1, 2, ..., 15\}$. Nadto $Y_i = Z_i$ dla $i \in \{1, 2, ..., 10\}$ oraz $Y_i = \sqrt{3}Z_i$, $i \in \{11, ..., 15\}$, gdzie Z_i są niezależne o rozkładzie $\mathcal{N}(0, 1)$. Stąd

$$\sum_{i=1}^{15}(X_i - \bar{X})^2 = \sum_{i=1}^{15}(\sigma Y_i - \sigma\bar{Y} + \mu_i - \frac{2}{3}\mu_1 - \frac{1}{3}\mu_2)^2$$

co daje

$$\mathbf{E}\sum_{i=1}^{15}(X_i - \bar{X})^2 = \sum_{i=1}^{15}\mathbf{E}(Y_i - \bar{Y})^2 + 10(\frac{1}{3}(\mu_1 - \mu_2))^2 + 5(\frac{2}{3}(\mu_1 - \mu_2))^2$$

a stąd

$$\mathbf{E}\sum_{i=1}^{15}(X_i - \bar{X})^2 = 10((\frac{14}{15})^2 + \frac{24}{(15)^2}) + 5((\frac{14}{15})^2 + \frac{12}{(15)^2}) + \frac{10}{3}(\mu_1 - \mu_2)^2.$$

Z drugiej strony

$$\bar{X}_1 - \bar{X}_2 = \sigma(\bar{Z}_1 - \sqrt{3}\bar{Z}_2) - \mu_1 + \mu_2,$$

czyli

$$\mathbf{E}(\bar{X}_1 - \bar{X}_2)^2 = \sigma^2\mathbf{E}(\bar{Z}_1 - \sqrt{3}\bar{Z}_2)^2 + (\mu_1 - \mu_2)^2 = \frac{7}{10}\sigma^2 + (\mu_1 - \mu_2)^2.$$

Aby estymator $\hat{\sigma}^2$ był nieobciążony dostajemy równania

$$\begin{cases} \frac{70}{3}a + \frac{7}{10}b &= 1 \\ \frac{10}{3}a + b &= 0 \end{cases}$$

Dostajemy $a = \frac{1}{21}$, $b = -\frac{10}{63}$.

■

10. (Eg 57/10) Zakładamy, że zależność czynnika Y od czynnika x (nielosowego) opisuje model regresji liniowej $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, gdzie błędy ε_i są niezależne i mają rozkłady normalne o wartości oczekiwanej 0 i wariancji 1. Obserwujemy zmienne losowe $Y_1, Y_2, ..., Y_n$ przy danych wartościach $x_1, x_2, ..., x_n$. Test najmocniejszy dla weryfikacji hipotezy

$$H_0 : \ \beta_0 = 0 \ \text{ i } \ \beta_1 = 1$$

przy alternatywie

$$H_1 : \ \beta_0 = -1 \ \text{ i } \ \beta_1 = 2$$

na poziomie istotności $0,05$ odrzuca hipotezę H_0, gdy spełniona jest nierówność
Odp: A-> $\frac{\sum_{i=1}^{n}(Y_i - x_i)(x_i - 1)}{\sqrt{\sum_{i=1}^{n}(1 - x_i)^2}} > 1,645$.

Rozwiązanie. Korzystamy z testu Neymana-Pearsona. Iloraz funkcji wiarygodności

$$\frac{L((-1, 2, y))}{L((0, 1, y)} = \exp(\sum_{i=1}^{n} y_i(x_i - 1) - \frac{1}{2}(1 - 2x_i)^2 + \frac{1}{2}x_i^2).$$

Zatem obszar krytyczny ma postać

$$\mathcal{K} = \{y \in \mathbb{R}^n : \ \sum_{i=1}^{n} y_i(x_i - 1) > C\}$$

dla C takiego, że

$$\mathbf{P}_{0,1}(Y \in \mathcal{K}) = \mathbf{P}_{0,1}(\sum_{i=1}^{n} Y_i(x_i - 1) > C) = 0,05.$$

Łatwo zauważyć, że $\sum_{i=1}^n (x_i-1)Y_i$ ma rozkład $\mathcal{N}(\sum_{i=1}^n x_i(x_i-1), \sum_{i=1}^n (1-x_i)^2)$. Stąd ostatecznie $\frac{\sum_{i=1}^n (Y_i-x_i)(x_i-1)}{\sqrt{\sum_{i=1}^n (1-x_i)^2}}$ ma rozkład zmiennej Z o rozkładzie $\mathcal{N}(0,1)$. Zatem obszar krytyczny możemy zapisać jako

$$\mathbf{P}_{0,1}(\sum_{i=1}^n Y_i(x_i-1) > C) = \mathbf{P}_{0,1}(\frac{\sum_{i=1}^n (Y_i-x_i)(x_i-1)}{\sqrt{\sum_{i=1}^n (1-x_i)^2}} > \bar{C}) = \mathbf{P}(Z > \bar{C}) = 0,05.$$

Stąd $C \simeq 1,645$.

■

11. (Eg 58/9) Niech $X_1, X_2, ..., X_n, ...$ będą niezależnymi zmiennymi losowymi takimi, że $\mathbf{E}X = im$ i $\mathbf{Var}X_i = im^2$ dla $i = 1, 2, ..., n$, gdzie $m > 0$ jest nieznanym parametrem. W klasie estymatorów parametru m postaci $\bar{m} = \sum_{i=1}^n c_i X_i$ (gdzie c_i są liczbami rzeczywistymi) najmniejszy błąd średniokwadratowy ma estymator, dla którego c_i są równe
Odp: C-> $c_1 = c_2 = ... = c_n = \frac{2}{n(n+1)+2}$.

Rozwiązanie. Błąd średniokwadratowy ma postać

$$f(c) = \mathbf{E}(m-\bar{m})^2 = (1 - \sum_{i=1}^n ic_i)^2 m^2 + \sum_{i=1}^n ic_i^2 m^2.$$

Minimalizacja polega na policzeniu pochodnych względem c_i.

$$\frac{\partial f}{\partial c_i} = -2i(1 - \sum_{i=1}^n ic_i)m^2 + 2im^2 c_i.$$

Czyli

$$c_i = (1 - \sum_{i=1}^n ic_i).$$

Stąd $c_1 = c_2 = ... = c_n = c$. Nadto $c = 1 - \frac{(n+1)n}{2}c$, czyli $c = \frac{2}{n(n+1)+2}$.

■

12. (Eg 59/7) Pobieramy próbkę niezależnych realizacji zmiennych losowych o rozkładzie Poissona z wartością oczekiwaną $\lambda > 0$. Niestety sposób obserwacji uniemożliwia odnotowanie realizacji o wartości 0. Pobieranie próbki kończymy w momencie, gdy liczebność odnotowanych realizacji wynosi n. Tak więc, każda z naszych kolejnych odnotowanych realizacji $K_1, K_2, ..., K_n$ wynosi co najmniej 1 i nic nie wiemy o tym, ile w międzyczasie pojawiło się obserwacji o wartości 0. Estymujemy parametr λ za pomocą estymatora postaci

$$\bar{\lambda} = \frac{1}{n}\sum_{i=2}^{\infty} iN_i,$$

gdzie N_i jest liczbą obserwacji o wartości i. Błąd średniokwadratowy estymatora $\bar{\lambda}$ jest równy
Odp: E-> $\frac{\lambda^2-\lambda+\lambda e^\lambda}{n(e^\lambda-1)}$.

Rozwiązanie. Przy założeniach zadania przyjmujemy, że K_i mają rozkład

$$\mathbf{P}(K_i = k) = \frac{\lambda^k}{k!(e^\lambda - 1)}, \quad k = 1, 2, 3, ...$$

Nadto zauważmy, że

$$\frac{1}{n}\sum_{i=2}^{\infty} iN_i = \frac{1}{n}(\sum_{i=1}^n K_i - N_1).$$

Dalej

$$\mathbf{E}\frac{1}{n}\sum_{i=1}^{n}K_i=\frac{\lambda e^{\lambda}}{e^{\lambda}-1},\ \ \mathbf{E}\frac{1}{n}N_1=\frac{\lambda}{e^{\lambda}-1}.$$

co oznacza, że

$$\mathbf{E}\frac{1}{n}(\sum_{i=1}^{n}K_i-N_1)=\lambda.$$

czyli estymator $\bar{\lambda}$ jest nieobciążony. Zauważmy, że $N_1=\sum_{i=1}^{n}1_{K_i=1}$ mamy

$$\mathbf{E}\frac{1}{n}\sum_{i=1}^{n}(K_i-\mathbf{E}K_i)\frac{1}{n}\sum_{j=1}^{n}(1_{K_j=1}-\mathbf{E}1_{K_j=1})=\frac{1}{n^2}\mathbf{E}\sum_{i=1}^{n}(K_i-\mathbf{E}K_i)(1_{K_i=1}-\mathbf{E}1_{X_i=1})=$$

$$=\frac{1}{n}(\mathbf{E}K_1 1_{K_1=1}-\mathbf{E}K_1\mathbf{P}(K_1=1))=\frac{1}{n}\mathbf{P}(K_1=1)(1-\mathbf{E}K_1)=\frac{1}{n}\frac{\lambda}{e^{\lambda}-1}(1-\frac{\lambda e^{\lambda}}{e^{\lambda}-1}).$$

Obliczamy

$$\mathbf{E}(\bar{\lambda}-\lambda)^2=\mathbf{Var}\frac{1}{n}\sum_{i=1}^{n}K_i+\mathbf{Var}\frac{1}{n}\sum_{i=1}^{n}1_{K_i=1}-2\mathbf{Cov}(\frac{1}{n}\sum_{i=1}^{n}K_i-\mathbf{E}K_i,\frac{1}{n}\sum_{i=1}^{n}1_{K_1=1}-\mathbf{E}1_{K_1=1})=$$

$$=\frac{1}{n}\mathbf{Var}K_1+\frac{1}{n}\mathbf{Var}1_{K_1}-\frac{2}{n}\mathbf{P}(K_1=1)(1-\mathbf{E}K_1)=\frac{1}{n}[\frac{(\lambda+\lambda^2)e^{\lambda}}{e^{\lambda}-1}-\frac{\lambda^2e^{2\lambda}}{(e^{\lambda}-1)^2}]+$$

$$+\frac{1}{n}\frac{\lambda}{e^{\lambda}-1}(1-\frac{\lambda}{e^{\lambda}-1})-\frac{2}{n}\frac{\lambda}{e^{\lambda}-1}(1-\frac{\lambda e^{\lambda}}{e^{\lambda}-1})=$$

$$=\frac{\lambda}{n(e^{\lambda}-1)}(\lambda-1+e^{\lambda}).$$

■

13. (Eg 60/3) Rozważamy model regresji liniowej postaci $Y_i=bx_i+\varepsilon_i$, $i=1,2,...,5$, gdzie b jest nieznanym parametrem rzeczywistym, $x_1=x_2=1$, $x_3=\sqrt{5}$, $x_4=x_5=3$, a ε_i są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym o wartości oczekiwanej 0 i nieznanej wariancji $\sigma^2>0$. Hipotezę $H_0: b=0$ przy alternatywie $H_1: b\neq 0$ weryfikujemy testem o obszarze krytycznym postaci $\{|\frac{\bar{b}}{\bar{\sigma}}|>c\}$, gdzie $\bar{b}$ i $\bar{\sigma}$ są estymatorami największej wiarogodności parametrów b i σ, a stała c dobrana jest tak, aby test miał rozmiar $0,05$. Stała c równa
Odp: E-> $0,62$.

Rozwiązanie. Należy wyznaczyć ENW b i σ^2. Tradycyjnie obliczając pochodne sprawdzamy, że

$$\bar{b}=\frac{\sum_{i=1}^{n}x_iY_i}{\sum_{i=1}^{n}x_i^2},\ \ \bar{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-x_i\frac{\sum_{j=1}^{n}x_jY_j}{\sum_{j=1}^{n}x_j^2})^2.$$

Stąd dla $b=0$ dostajemy

$$\bar{b}=\frac{\sum_{i=1}^{n}x_i\varepsilon_i}{\sum_{i=1}^{n}x_i^2},\ \ \bar{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i-x_i\frac{\sum_{j=1}^{n}x_j\varepsilon_j}{\sum_{j=1}^{n}x_j^2})^2.$$

Czyli dla $b=0$ zachodzi $\bar{b}=\frac{\langle x,\varepsilon\rangle}{\|x\|_2}$, $\bar{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i-\frac{x_i\langle x,\varepsilon\rangle}{\|x\|_2^2})$. Łatwo zauważyć, że $\bar{b}$ ma postać $\frac{\sigma}{\|x\|}Z$, gdzie Z ma rozkład $\mathcal{N}(0,1)$. Dalej z faktu braku korelacji wynika, że $(\varepsilon_i-\frac{x_i\langle x,\varepsilon\rangle}{\|x\|_2^2})$ jest niezależne od $\frac{\langle x,\varepsilon\rangle}{\|x\|_2}$ dla każdego $i=1,2,...,n$. Stąd niezależne są zmienne $\frac{\langle x,\varepsilon\rangle}{\|x\|_2}$ oraz

$$\sum_{i=1}^{n}(\varepsilon_i-\frac{x_i\langle x,\varepsilon\rangle}{\|x\|_2^2})^2=\|\varepsilon\|_2^2-\frac{\langle x,\varepsilon\rangle^2}{\|x\|^2}=Y.$$

Pozostaje zauważyć, że zmienna Y ma rozkład $\chi^2(n-1)$. Wynika z ogólnej reguły, że jeśli $\|\varepsilon\|_2^2$ ma rozkład $\chi^2(n)$ oraz można rozłożyć $\|\varepsilon\|_2^2$ na niezależne zmienne $= \|\varepsilon\|_2^2 - \frac{\langle x,\varepsilon\rangle^2}{\|x\|^2}$ i $\frac{\langle x,\varepsilon\rangle^2}{\|x\|^2}$, gdzie druga zmienna ma rozkład Z^2, $Z \simeq \mathcal{N}(0,1)$, to pierwsza musi mieć rozkład $\chi^2(n-1)$. Stąd

$$\mathbf{P}_{b=0}(|\frac{\bar{b}}{\bar{\sigma}}| > c) = \mathbf{P}_{b=0}(|\frac{\sqrt{n-1}Z}{\sqrt{Y}}| > c\|x\|\frac{\sqrt{n-1}}{\sqrt{n}}) = 0,05.$$

Zmienna $\frac{\sqrt{n-1}Z}{\sqrt{Y}}$ ma rozkład t-Studenta z parametrem $n-1$. Podstawiając $n=5$ oraz $\|x\|=5$ dostajemy, że $2\sqrt{5}c$ jest kwantylem symetrycznym wartości $0,05$ dla rozkładu t-Studenta z 4 stopniami swobody. Kwantyl ten wynosi $2,776$, stąd $c \simeq 0,62$.

■

14. (Eg 61/3) Niech $X_1, X_2, ..., X_n$ będą niezależnymi zmiennymi losowymi z rozkładu normalnego o nieznanej wartości oczekiwanej μ i nieznanej wariancji σ^2. Niech T oznacza estymator nieobciążony o minimalnej wariancji parametru μ^2. Wtedy błąd średniokwadratowy tego estymatora, czyli

$$\mathbf{E}_{\mu,\sigma}(T-\mu^2)^2$$

jest równa
Odp: E-> $\frac{2\sigma^2}{n(n-1)} + \frac{4\mu^2\sigma^2}{n}$.

Rozwiązanie. Przypomnijmy regułę budowania estymatorów ENMW. Najpierw znajdujemy statystyki dostateczne i zupełne w tym przypadku

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; \bar{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Obliczamy

$$\mathbf{E}\bar{\mu}^2 = \mu^2 + \frac{\sigma^2}{n}, \;\; \mathbf{E}\bar{\sigma}^2 = \sigma^2.$$

To oznacza, że ENMW jest postaci $\bar{\mu}^2 - \frac{1}{n}\bar{\sigma}^2$, gdyż jest nieobciążonym estymatorem μ^2 będącym funkcją statystyk dostatecznych i zupełnych. Obliczamy jego wariancję pamiętając, że $\bar{\mu}$ i $\bar{\sigma}^2$ są niezależne oraz z faktu, że $\bar{\mu}$ ma rozkład $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ a $\bar{\sigma}^2$ ma rozkład $\sigma^2\chi^2(n-1)$

$$\mathbf{Var}(\bar{\mu}^2 - \frac{1}{n}\bar{\sigma}^2) = \mathbf{Var}(\bar{\mu}^2) + \frac{1}{n^2}\mathbf{Var}(\bar{\sigma}^2) =$$
$$= \frac{4\mu^2\sigma^2}{n} + \frac{2(n-1)\sigma^2}{n^2}.$$

■

15. (Eg 63/2) Niech X_1, X_2, X_3, X_4 będą niezależnymi zmiennymi losowymi, przy czym zmienna losowa X_i ma rozkład o wartości oczekiwanej m i wariancji im^2, $i = 1,2,3,4$, gdzie $m \neq 0$ jest nieznanym parametrem. Niech $\bar{m}$ oznacza estymator parametru m minimalizujący błąd średniokwadratowy w klasie estymatorów postaci

$$a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4,$$

gdzie współczynniki $a_i \in \mathbb{R}$, $i = 1,2,3,4$. Wtedy $\mathbf{E}(\bar{m}-m)^2$ jest równe
Odp: E-> $\frac{12}{37}m^2$.

Rozwiązanie. Mamy do wyznaczenia $\min \mathbf{E}(\sum_{i=1}^{4} a_iX_i - m)^2$, po $a_i \in \mathbb{R}$. Prowadzi to do równań na zerowanie się kolejnych pochodnych cząstkowych

$$\mathbf{E}X_i(\sum_{i=1}^{4} a_iX_i - m) = 0, \;\; i = 1,2,3,4.$$

Korzystając z faktu, że $\mathbf{E}X_i = m$ przekształcamy równania

$$\mathbf{E}(X_i - m)(\sum_{i=1}^{4} a_i(X_i - m)) = m^2(1 - \sum_{i=1}^{4} a_i), \;\; i = 1, 2, 3, 4.$$

Z niezależności wynika zatem, że

$$a_i \mathbf{Var} X_i = m^2(1 - \sum_{i=1}^{4} a_i), \;\; i = 1, 2, 3, 4.$$

Ponieważ $\mathbf{Var}X_i = im^2$ zatem a_i jest postaci c/i. Stałą c wyznaczamy z równania

$$c = 1 - c(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}), \;\; c = \frac{12}{37}.$$

Stąd $a_i = \frac{4}{9i}$, $i = 1, 2, 3, 4$. Obliczamy

$$\begin{aligned}
&\mathbf{E}(\bar{m} - m)^2 = \mathbf{E}(\sum_{i=1}^{4} a_i(X_i - m) + \frac{12m}{37})^2 = \\
&= \sum_{i=1}^{4} a_i^2 \mathbf{Var} X_i + \frac{(12)^3 m^2}{(37)^2} = \frac{(12)^3 m^2}{(37)^2}((1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}) + 1) = \\
&= \frac{12}{37} m^2.
\end{aligned}$$

■