

Statystyczna Analiza Danych, zestaw 2

Zad 1 Niech $X_1, \dots, X_n$ będzie próbą prostą z rozkładu o skończonej i nieznanej wariancji σ^2 . Wyznacz obciążenie estymatora wariancji zadanego wzorem

$$S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \quad \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

(Let $X_1, \dots, X_n$ be a random sample with unknown mean and unknown variance σ^2 . Compute the bias of the variance estimator

$$S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \quad \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

)

Zad 2 Niech $X_1, \dots, X_n$ będzie próbką z rozkładu $\text{Ber}(p)$, tzn. $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$.

- (a) Wyznacz estymator największej wiarygodności dla parametru p .
- (b) Sprawdź, czy powyższy estymator jest obciążony.
- (c) Wyznacz nieobciążony estymator wartości p^2 .

(Let $X_1, \dots, X_n$ be i.i.d. $\text{Ber}(p)$ (i.e. $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$)

- (a) Compute the maximum likelihood estimator of p .
- (b) Determine whether it is unbiased.
- (c) Compute an unbiased estimator for p^2 .

)

Zad 3 Niech $\text{Geom}(p)$ będzie rozkładem geometrycznym, tzn. jeśli $X \sim \text{Geom}(p)$, to $\mathbb{P}(X = k) = p(1 - p)^{k-1}$. Na podstawie próbki $X_1, \dots, X_n$ z tego rozkładu wyznacz estymator największej wiarygodności dla parametru p . ($\mathbb{P}(X = k) = p(1 - p)^{k-1}$ for $k \geq 1$. Let $X_1, \dots, X_n$ be a random sample from this distribution. Compute the maximum likelihood estimator for p .)

Zad 4 Niech $X_1, \dots, X_n$ będą czasami świecenia żarówek, modelowanymi przez niezależne zmienne losowe o rozkładzie $\text{Exp}(\lambda)$, gdzie λ jest nieznanym parametrem. Wyznacz estymator największej wiarygodności dla parametru λ .

(Let $X_1, \dots, X_n$ be a random sample from an $\text{Exp}(\lambda)$ distribution, λ unknown. Compute the maximum likelihood estimator of λ).

Zad 5 Niech $X_1, \dots, X_n \sim \text{Unif}([0, a])$, gdzie a jest nieznanym parametrem.

- (a) Wyznacz nieobciążony estymator a oparty na $\bar{X}$ (średniej z próby). Przeanalizuj jego sensowność dla różnych ciągów obserwacji. Oblicz jego wariancję.
- (b) Wyznacz estymator największej wiarygodności dla a . Zmodyfikuj go tak, aby otrzymać estymator nieobciążony. Oblicz wariancję tak otrzymanego estymatora i porównaj ją z wariancją estymatora z poprzedniego punktu.

$(X_1, \dots, X_n \text{ i.i.d. } \text{Unif}([0, a]), \text{ where } a \text{ is unknown.})$

- (a) Compute an unbiased estimator of the form $a\bar{X} + b$ (where $\bar{X}$ is the sample average). Compute its variance.
- (b) Compute the unbiased estimator of a . Modify it by a multiplicative constant to get an unbiased estimator and compute the mean squared error of both estimators.)

Zad 6 Rozważmy model statystyczny, w którym obserwacje X i Y są niezależne oraz $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 3$, a także $\text{Var}(X) = \text{Var}(Y) = \sigma^2$, gdzie σ^2 jest nieznanym parametrem. Wyznacz wartości stałych a , b , dla których $aX^2 + bY^2$ jest nieobciążonym estymatorem parametru σ^2 .

($\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 3$, $\text{Var}(X) = \text{Var}(Y) = \sigma^2$, unknown. Find constants a and b such that $aX^2 + bY^2$ is an unbiased estimator of σ^2 .)

Zad 7 Let (X, Y) , (X_1, Y_1) and (X_2, Y_2) be i.i.d.. Show that

$$\text{Cov}(X, Y) = \frac{1}{2} \mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)]$$

Let h and g be two increasing functions. Show that $\text{Cov}(h(X), g(X)) \geq 0$.

Zad 8 Liczba wypadków samochodowych zgłoszonych do towarzystwa ubezpieczeniowego w k -tym miesiącu jest zmienną losową W_k o rozkładzie Poissona z parametrem λz_k , gdzie z_k jest liczbą samochodów zgłoszonych do ubezpieczenia w tym miesiącu, zaś λ jest nieznanym parametrem. Zmienne losowe W_k są niezależne. Wyznaczyć estymator największej wiarygodności parametru λ na podstawie próby $W_1, \dots, W_{12}$.

(W_k number of accidents in time period k is Poisson, parameter λz_k , z_k number of cars declared for insurance in the time period. Compute m.l.e. of λ based on sample $W_1, \dots, W_{12}$).

Tutorial 2: Answers

Q1

$$\sum_{j=1}^n (X_j - \bar{X})^2 = \sum_{j=1}^n (X_j - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{j=1}^n (X_j - \mu) = \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X} - \mu)^2$$

so that, using $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, therefore:

$$\mathbb{E}[S^2] = \sigma^2 \left(1 - \frac{1}{n}\right)$$

$$\text{so Bias}(S^2) = \mathbb{E}[S^2] - \sigma^2 = -\frac{\sigma^2}{n}$$

Q2

(a) $L(p; x_1, \dots, x_n) = p^k (1-p)^{n-k}$ where $k = \sum_j x_j$ so that

$$\log L = k \log p + (n-k) \log(1-p)$$

MLE satisfies:

$$\frac{k}{p} - \frac{n-k}{1-p} = 0 \implies \hat{p} = \frac{k}{n}$$

(b) k from $\text{Binom}(n, p)$ so $\mathbb{E}[\hat{p}] = \frac{np}{n} = p$ so unbiased.

(c) Note that $\mathbb{E}[\hat{p}^2] = \frac{1}{n^2} (\sum_{j,k} \mathbb{E}[X_j X_k]) = \frac{p}{n} + (1 - \frac{1}{n})p^2$

so that

$$\hat{p}^2 := \frac{n}{n-1} \left((\hat{p})^2 - \frac{\hat{p}}{n} \right)$$

gives an unbiased estimator.

Q3

$$L(p; x_1, \dots, x_n) = (1-p)^{S-n} p^n$$

$$\log L = (S-n) \log(1-p) + n \log p$$

$$\frac{n}{p} - \frac{S-n}{1-p} = 0 \implies \hat{p} = \frac{n}{S} \quad S = \sum_{j=1}^n x_j$$

Q4

$$L(\lambda; x_1, \dots, x_n) = \lambda^n \exp\left\{-\lambda \sum_{j=1}^n x_j\right\}$$

$$\log L = n \log \lambda - \lambda \sum_{j=1}^n x_j$$

$$\hat{\lambda} = \frac{n}{\sum_{j=1}^n x_j}$$

Q5

(a) $\mathbb{E}[X] = \frac{a}{2}$ so $\widehat{a} = 2\overline{X}$, $\mathbb{E}[\widehat{a}] = a$

(b)

$$L(a; x_1, \dots, x_n) = \frac{1}{a^n} \mathbf{1}_{[0, \max_j x_j]}(a)$$

$$\widehat{a}_{ML} = \max_j x_j$$

$$\mathbb{P}(\widehat{a}_{ML} \leq x) = \mathbb{P}(X \leq x)^n = \left(\frac{x}{a}\right)^n$$

$$g(x) = \frac{nx^{n-1}}{a^n}$$

$$\mathbb{E}[\widehat{a}_{ML}] = \frac{n}{a^n} \int_0^a x^n dx = \frac{n}{n+1}a$$

$$\widetilde{a} = \frac{n+1}{n} \max_j x_j$$

$$\mathbb{E}[\widehat{a}_{ML}^2] = \frac{n}{n+2}a^2$$

$$\text{Var}(\widehat{a}_{ML}) = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2} \right) a^2 = \frac{n}{(n+1)^2(n+2)} a^2.$$

Hence:

$$\text{Var}(\widetilde{a}) = \frac{1}{n(n+2)} a^2$$

$$\text{Bias}(\widehat{a}_{ML}) + \text{Var}(\widehat{a}_{ML}) = \frac{2a^2}{(n+1)(n+2)}$$

Q6 $\mathbb{E}[X^2] = 1 + \sigma^2$, $\mathbb{E}[Y^2] = 9 + \sigma^2$ $\mathbb{E}[aX^2 + bY^2] = (a + 9b) + \sigma^2(a + b)$ $a = 1 - b$
 $1 + 8b = 0$ $b = -1/8$ $a = 9/8$.

Q7

$$\mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)] = \mathbb{E}[(X_1 - \mu_X + \mu_X - X_2)(Y_1 - \mu_Y + \mu_Y - Y_2)] = 2\text{Cov}(X, Y)$$

since $\mathbb{E}[(X_1 - \mu_X)(Y_2 - \mu_Y)] = \mathbb{E}[X_1 - \mu_X]\mathbb{E}[Y_2 - \mu_Y] = 0$

Therefore

$$\text{Cov}(h(X), g(X)) = \frac{1}{2} \mathbb{E}[(h(X_1) - h(X_2))(g(X_1) - g(X_2))]$$

On $X_1 > X_2$ the integrand is positive; on $X_1 < X_2$ the integrand is positive.

Q8

$$L(\lambda : W_1, \dots, W_{12}) = \frac{\lambda^{\sum W_j} \prod_{j=1}^{12} z_j^{W_j}}{\prod_{j=1}^{12} w_j!} e^{-\lambda \sum_{j=1}^{12} z_j}$$

$$\log L = \text{const} + (\sum W_j) \log \lambda - \lambda \sum_{j=1}^{12} z_j$$

$$0 = \frac{\sum W_j}{\lambda} - \sum z_j \Rightarrow \hat{\lambda} = \frac{\sum_{j=1}^{12} W_j}{\sum_{j=1}^{12} z_j}$$