METODY PROBABILISTYCZNE

1. **Zdefiniuj zmienną losową, rozkład prawdopodobieństwa. Przy jakich założeniach funkcje: $F(x) = sin(x)$, $F(x) = a - e^{-x}$ mogą być dystrybuantami?**
2. **Podaj twierdzenie Lindeberga – Lévy'ego. Jakie jest jego znaczenie w estymacji wartości średniej?**
3. **Zdefiniować proces Markova. Co go w pełni charakteryzuje?**
4. **Partia pudełek zawiera 100000 pudełek zapałek. Dostawca twierdzi, że w pudełku są średnio 54 zapałki. Stwierdzić, czy hipotezę tę można odrzucić na poziomie istotności $\alpha = 0.02$. Pobrano próbę o liczności $n = 100$. Z próby tej otrzymano $m_n = 51.21$ oraz $\sigma'_n = 2.45$.**
5. **Podać definicję zmiennych losowych niezależnych dla n zmiennych $(X_1, X_2, \dots, X_n)$. Rozpatrzmy dwukrotny rzut monetą. Sprawdzić, czy zmienne losowe X i Y charakteryzujące odpowiednio pierwszy i drugi rzut są niezależne.**
6. **Zdefiniuj proces stacjonarny w szerszym i w węższym sensie. Podaj związek między nimi.**
7. **Opisać sposób wyznaczania minimalnej liczności próby zapewniającej zadaną dokładność estymacji wartości średniej cechy X w próbie generalnej.**
8. **Podać aksjomatykę Kołmogorowa, opisać jak tworzymy δ-algebrę $\mathcal{A}$. Zdefiniować prawdopodobieństwo warunkowe i udowodnić, że spełnia te aksjomaty.**
9. **Zdefiniować warunkowy rozkład prawdopodobieństwa zmiennej losowej Y pod warunkiem, że zmienna losowa X przyjęła określoną wartość oraz warunkową wartość średnią przy tym założeniu.**
10. **Sformułować słabe prawo wielkich liczb Markova i podać (krótko) jego związek z teorią estymacji.**
11. **Zdefiniować proces losowy o niezależnych przyrostach i proces Markova. Podać związek między nimi.**
12. **Zdefiniuj rozkłady brzegowe zmiennej losowej X. Podaj dystrybuantę dla typu skokowego, ciągłego i w przypadku ogólnym.**
13. **W pewnym urządzeniu trzeba przeciętnie 7 razy w roku (≈7000 h pracy) wymieniać pewien podzespół. Zakładając, że liczba wymian tego podzespołu w danym okresie ma rozkład Poissona, wyliczyć prawdopodobieństwo tego, że konieczność wymiany podzespołu wystąpi po 100 godzinach pracy.**
14. **Oblicz $E(X)$ i $D^2(X)$ rozkładu Poissona.**
15. **Podaj własności wariancji i udowodnij, że suma wariancji jest równa wariancji sumy zmiennych.**
16. **Zdefiniować proces losowy, jego wartość średnią i funkcję korelacyjną. Kiedy można wyznaczyć te charakterystyki dysponując pojedynczą realizacją procesu (warunek dostateczny dla procesów stacjonarnych w szerszym sensie + odpowiednie wzory).**
17. **Definicja próby losowej, estymatora parametru q cechy X. Dany jest rozkład $N(5000; 50)$, próba n-elementowa i średnia m. Obliczyć prawdopodobieństwo, że średnia m jest z przedziału $[5050; 5100]$.**
18. **Narysować wykres dystrybuanty oraz wyprowadzić wzór na wartość średnią i wariancję zmiennej losowej o rozkładzie dwupunktowym.**
19. **Zdefiniuj zmienną losową, rozkład prawdopodobieństwa i dystrybuantę. Wykreślić dystrybuantę dla rzutu sześcienną kostką do gry.**
20. **Obliczyć $E(X)$ i $D^2(X)$ dla rozkładu wykładniczego.**
21. **Sformułuj twierdzenie Bayesa.**
22. **Definicja estymatora.**
23. **Definicja próby losowej, sposób pobrania próby losowej, jakie postulaty musi spełniać próba losowa.**
24. **Definicja statystyki.**
25. **W jaki sposób estymujemy wartość cechy X.**
26. **W wyniku długotrwałych obserwacji znana jest wariancja pomiarów pewnej wielkości (wariancja $= 0.02$). Ilu trzeba dokonać pomiarów tej wielkości, aby z prawdopodobieństwem ≥ 0.99 średnia arytmetyczna otrzymanych wyników różniła się od wartości średniej o ≤ 0.01?**
27. **Wykazać, że wariancja sumy niezależnych zmiennych losowych X, Y jest równa sumie wariancji tych zmiennych.**
28. **Definicja procesu losowego (stacjonarnego chyba) w szerszym sensie oraz jego wartość średnia (skorzystać z ergodyczności).**
29. **Estymator najefektywniejszy wartości średniej.**
30. **Podać σ-algebrę zbiorów borelowskich.**

by Marcin R.

zad. 1

Zdefiniuj zmienną losową, rozkład prawdopodobieństwa. Przy jakich założeniach funkcje: $\boldsymbol{F(x) = sin(x)}$, $\boldsymbol{F(x) = a - e^{-x}}$ mogą być dystrybuantami?

Zmienną losową nazywamy funkcję rzeczywistą X określoną na przestrzeni zdarzeń elementarnych Ω i taką, że przeciwobrazy zbiorów borelowskich na prostej (w R^1) otrzymane za pomocą tej funkcji są zdarzeniami losowymi. Wystarczy wymagać, by przeciwobrazy przedziałów postaci $(-\infty; x)$ były zdarzeniami losowymi, gdyż każdy zbiór borelowski na prostej można otrzymać za pomocą przeliczalnej liczby działań na takich przedziałach.

Innymi słowy, zmienna losowa X to funkcja działająca z Ω w R^1 taka, że:

$$\bigwedge_x X^{-1}[(-\infty; x)] = \{\omega: X(\omega) < x\} \in \mathcal{A}$$

Widzimy zatem, że zmienna losowa to funkcja odwzorowująca Ω w prostą, mierzalna względem prawdopodobieństwa.

Rozkładem prawdopodobieństwa zmiennej losowej X nazywamy prawdopodobieństwo indukowane czyli funkcję zbioru daną wzorem:

$$P_X(S) = P[X^{-1}(S)] = P[\omega: X(\omega) \in S], \qquad S \in \mathcal{A}_X$$

Aby funkcja $F(x)$ mogła być dystrybuantą, musi spełniać warunki:

a) jest nieujemna i niemalejąca
b) $F(-\infty) = 0, \quad F(\infty) = 1$
c) jest przynajmniej lewostronnie ciągła

Aby pierwsza funkcja była dystrybuantą, definiujemy ją następująco:

$$F(x) = \begin{cases} 0, & x < 0 \\ \sin(x), & x \in \left[0; \frac{\pi}{2}\right) \\ 1, & x \geq \frac{\pi}{2} \end{cases}$$

Dla funkcji drugiej natomiast przyjmujemy $a = 1$ oraz definiujemy ją jako:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}$$

zad. 2

Podaj twierdzenie Lindeberga – Lévy'ego. Jakie jest jego znaczenie w estymacji wartości średniej?

Niech $X_1, X_2, \dots$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa, przy czym $E(X_i) = m, D^2(X_i) = \sigma^2 \neq 0, \quad i = 1,2, \dots$

Utwórzmy zmienną losową:

$$Y_n = \sum_{i=1}^{n} X_i$$

Oczywiście: $E(Y_n) = nm, \;\; D^2(Y_n) = n\sigma^2$

Unormujmy zmienną losową Y_n definiując zmienną losową Z_n wzorem:

$$Z_n = \frac{Y_n - mn}{\sigma\sqrt{n}}$$

Przy powyższych założeniach ciąg zmiennych losowych $Z_1, Z_2, \dots$ jest zbieżny wg dystrybuanty do zmiennej losowej Z o rozkładzie $N(0,1)$, czyli:

$$\bigwedge_{z} \lim_{n \to \infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} \, dz$$

gdzie $F_n(z)$ jest dystrybuantą zmiennej losowej Z_n.

Jeśli mamy więc jakiś nieznany rozkład, to dla odpowiednio dużej próby możemy przyjąć, że jest to rozkład normalny, co znacznie ułatwia np. estymację wartości średniej.

zad. 3

Zdefiniować proces Markova. Co go w pełni charakteryzuje?

Proces losowy $X(t)$ nazywamy procesem Markova jeśli dla dowolnego n, dowolnego układu chwil $t_1 < t_2 < \cdots < t_n$ oraz dla dowolnych liczb rzeczywistych $x_1, x_2, \dots, x_n$ zachodzi:

$$P[X(t_n) < x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_1) = x_1] = P[X(t_n) < x_n | X(t_{n-1}) = x_{n-1}]$$

Oznacza to, że dystrybuanta warunkowa procesu Markova w dowolnej chwili t_n pod warunkiem, że wartości tego procesu w chwilach $t_{n-1}, t_{n-2}, \dots, t_1$ są ustalone zależy tylko od ustalonej wartości tego procesu w chwili t_{n-1} (oraz wartości bieżącej). Jest to tzw. własność braku pamięci lub własność Markova.

Proces Markova jest w pełni opisany przez dystrybuantę warunkową:

F(x,y,s,t) = P[X(t)<x|X(s)=y], s<t

Innymi slowy : Proces markowa jest zatem w pelni scharakteryzowany przez laczna dystrybuante wektora losowego:

$$[X(t), X(s)]$$

wraz z tzw. dystrybuantą początkową

$$F(s, y) = P[X(s) < y].$$

Widzimy zatem, że proces Markova jest w pełni scharakteryzowany przez jego rozkłady dwuwymiarowe.

zad. 4

Partia zawiera 100000 pudełek zapałek. Dostawca twierdzi, że w pudełku są średnio 54 zapałki. Stwierdzić, czy hipotezę tę można odrzucić na poziomie istotności $\boldsymbol{\alpha = 0.02}$. Pobrano próbę o liczności $\boldsymbol{n = 100}$. Z próby tej otrzymano $\boldsymbol{m_n = 51.21}$ oraz $\boldsymbol{\sigma'_n = 2.45}$.

Badana cecha ma nieznany rozkład o nieznanej wartości średniej i odchyleniu standardowym. Weryfikujemy hipotezę $H_0\colon \mu = \mu_0$ wobec hipotezy alternatywnej $H_1\colon \mu \neq \mu_0$. Ze względu na dużą liczność próby ($n = 100$) możemy zastosować statystykę:

$$U = \frac{\bar{X} - \mu_o}{\sigma}\sqrt{n}$$

która przy założeniu prawdziwości hipotezy H_0 jest standaryzowaną zmienną losową o rozkładzie $N(0,1)$. Dodatkowo, wobec dużej wartości n jako nieznaną wartość σ przyjmujemy watrość s obliczoną z próbki.

Dane:

$\mu_0 = 54$

$\bar{x} = m_n = 51.21$

$s = \sigma'_n = 2.45$

$n = 100$

Obliczamy wartość statystyki U:

$$u_{obl} = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{51.21 - 54}{2.45}\sqrt{100} \approx -11.39$$

Test jest testem dwustronnym, więc zbiór krytyczny wyznaczamy jako:

$$\left(-\infty; -u_{\left(1-\frac{\alpha}{2}\right)}\right] \cup \left[u_{\left(1-\frac{\alpha}{2}\right)}; \infty\right)$$

Wartość $u_{\left(1-\frac{\alpha}{2}\right)}$ odczytujemy z tablic rozkładu normalnego $N(0,1)$. Jeżeli u_{obl} należy do zbioru krytycznego, to odrzucamy hipotezę H_0 na korzyść hipotezy alternatywnej. W przeciwnym wypadku nie mamy podstaw do odrzucenia weryfikowanej hipotezy na przyjętym poziomie istotności.

zad. 5

Podać definicję zmiennych losowych niezależnych dla n zmiennych $(X_1, X_2, \dots, X_n)$. Rozpatrzmy dwukrotny rzut monetą. Sprawdzić, czy zmienne losowe X i Y charakteryzujące odpowiednio pierwszy i drugi rzut są niezależne.

Zmienne losowe $X_1, X_2, \dots, X_n$ nazywamy niezależnymi jeśli dla dowolnych zbiorów borelowskich $S_1, S_2, \dots, S_n$ odpowiednio na osiach $x_1, x_2, \dots, x_n$ zdarzenia:

$$Z_i = \{\omega : X_i(\omega) \in S_i\}, \qquad i = 1,2,\dots,n$$

spełniają warunek:

$$P(Z_1 \cap Z_2 \cap \dots \cap Z_n) = P(Z_1) * P(Z_2) * \dots * P(Z_n)$$

O – wypadł orzeł, R – wypadła reszka

$x, y \in \{O, R\}$

$(x, y) \in \{(O,O), (O,R), (R,O), (R,R)\}$

$P(X = O) = P(X = R) = P(Y = O) = P(Y = R) = \frac{1}{2}$

$P(X = O, Y = O) = \frac{1}{4} = P(X = O) * P(Y = O)$

$P(X = O, Y = R) = \frac{1}{4} = P(X = O) * P(Y = R)$

$P(X = R, Y = O) = \frac{1}{4} = P(X = R) * P(Y = O)$

$P(X = R, Y = R) = \frac{1}{4} = P(X = R) * P(Y = R)$

Widać zatem, że zmienne losowe X i Y są niezależne.

zad. 6

Zdefiniuj proces stacjonarny w szerszym i w węższym sensie. Podaj związek między nimi.

Proces losowy $X(t)$ nazywamy stacjonarnym w węższym sensie jeżeli dla dowolnego n i dowolnego układu $t_1, t_2, \dots, t_n \in T$ oraz dla dowolnego Δ takiego, że $\bigwedge_i (t_i + \Delta) \in T$, zachodzi:

$$F(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = F(x_1, t_1 + \Delta; x_2, t_2 + \Delta; \dots; x_n, t_n + \Delta)$$

Proces losowy $X(t)$ dla którego istnieją: wartość średnia $m(t)$ i funkcja korelacji $R_X(t_1, t_2)$ nazywamy stacjonarnym w szerszym sensie jeśli:

$m(t) = m$

$R_X(t_1, t_2) = R_X(\tau), \tau = t_2 - t_1$

Proces losowy stacjonarny w węższym sensie, dla którego $E[X^2(t)] < \infty$ jest również stacjonarny w szerszym sensie. Dla procesów normalnych słuszne jest również twierdzenie odwrotne.

zad. 7

Opisać sposób wyznaczania minimalnej liczności próby zapewniającej zadaną dokładność estymacji wartości średniej cechy X w próbie generalnej.

Jeśli cecha X ma rozkład $N(\mu, \sigma)$, gdzie σ jest znane, to dokładność estymacji wartości średniej wyraża się wzorem:

$$d = y_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

gdzie $y_{\frac{\alpha}{2}}$ wyznaczamy z tablic rozkładu $N(0,1)$ jako spełniające równanie $F(y) = 1 - \frac{\alpha}{2}$, $p_u = 1 - \alpha$. Stąd minimalna liczność próby wyraża się wzorem:

$$n_{min} = \left(\frac{y_{\frac{\alpha}{2}} * \sigma}{d}\right)^2$$

gdzie d jest zadaną dokładnością estymacji.

Jeśli natomiast cecha X ma rozkład normalny, ale σ nie jest znane, to otrzymujemy:

$$n_{min} = \left(\frac{t_{\frac{\alpha}{2}} * {\sigma_n}'}{d}\right)^2$$

gdzie $t_{\frac{\alpha}{2}}$ wyznaczamy z tablic rozkładu t-studenta o $n-1$ stopniach swobody, jako spełniające równanie $F(t) = 1 - \frac{\alpha}{2}$, $p_u = 1 - \alpha$; natomiast d jest zadaną dokładnością estymacji.

Przy czym jeśli $n \geq 30$ to założenie o normalności rozkładu cechy X populacji generalnej jest zbędne i możemy korzystać z tablic rozkładu $N(0,1)$.

Zauważmy, że przed pobraniem próby ${\sigma_n}'$ nie jest znane. Aby je wyznaczyć, należy pobrać wstępną próbę a dopiero później próbę właściwą lub przyjąć, że estymujemy n z dokładnością $d = \frac{{\sigma_n}'}{4}$. Wówczas otrzymujemy $n_{min} = 16 * \left(t_{\frac{\alpha}{2}}\right)^2$. Prowadzi to jednak do stosunkowo dużych n_{min} .

zad. 8

Podać aksjomatykę Kołmogorowa, opisać jak tworzymy δ-algebrę $\mathcal{A}$. Zdefiniować prawdopodobieństwo warunkowe i udowodnić, że spełnia te aksjomaty.

Niech Ω będzie dowolnym zbiorem, elementy $\varpi_1, \varpi_2, \ldots \in \Omega$ nazywamy zdarzeniami elementarnymi.

Niech $\mathcal{A}$ będzie δ-algebrą podzbiorów zbioru Ω. Elementy $A_1, A_2, \ldots \in \mathcal{A}$ nazywamy zdarzeniami losowymi.

Dla każdego zdarzenia $A \in \mathcal{A}$ określona jest liczba rzeczywista $P(A)$, zwana prawdopodobieństwem zdarzenia A, spełniająca następujące aksjomaty:

A1. $0 \leq P(A) \leq 1$

A2. $P(\Omega) = 1$

A3. $\left[(A_1, A_2, \ldots \in \mathcal{A}) \wedge (A_i \cap A_j = \emptyset, \mathrm{i} \neq \mathrm{j})\right] \Rightarrow [P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)]$

Jeśli chodzi o δ-algebrę $\mathcal{A}$ to: jeśli Ω jest skończona lub przeliczalna, to jest nią zbiór wszystkich podzbiorów przestrzeni Ω. Jeśli Ω jest nieprzeliczalna, to w charakterze $\mathcal{A}$ weźmiemy δ-algebrę zbiorów borelowskich zbioru Ω.

Jeśli $P(B) > 0$, to prawdopodobieństwo warunkowe zdarzenia A pod warunkiem, że zaszło zdarzenie B definiujemy wzorem:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Spełnia ono aksjomaty A1-A3:

A1.

Wiadomo, że $P(A \cap B) \leq P(B)$. Zatem: $P(A|B) = \frac{P(A\cap B)}{P(B)} \leq 1$

Jako, że $P(A \cap B) \geq 0$ oraz $P(B) > 0$ (z definicji prawdopodobieństwa warunkowego) to:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

A2.

$$P(\Omega|\mathrm{B}) = \frac{\mathrm{P}(\Omega \cap \mathrm{B})}{\mathrm{P(B)}} = \frac{P(B)}{P(B)} = 1$$

A3.

$$P\left((\bigcup_{i=1}^{\infty} A_i)|B\right) = \frac{P((\bigcup_{i=1}^{\infty} A_i)\cap B)}{P(B)} = \frac{P(\bigcup_{i=1}^{\infty}(A_i\cap B))}{P(B)}$$

Gdy zbiory $A_1, A_2, \ldots$ są rozłączne (a takie jest założenie), to zbiory $(A_i \cap B)$ też muszą być rozłączne. Tak więc:

$$P\left((\bigcup_{i=1}^{\infty} A_i)|B\right) = \frac{P(\bigcup_{i=1}^{\infty}(A_i\cap B))}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i\cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i\cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B)$$

zad. 9

Zdefiniować warunkowy rozkład prawdopodobieństwa zmiennej losowej Y pod warunkiem, że zmienna losowa X przyjęła określoną wartość oraz warunkową wartość średnią przy tym założeniu.

Jeśli zmienna losowa (X, Y) jest typu skokowego, to dystrybuanta zmiennej losowej Y pod warunkiem, że zmienna losowa X przyjmie ustaloną wartość x_i wyraża się wzorem:

$$F(y|x_i) = \sum_{y_k<y} p(y_k|x_i),$$

gdzie $p(y_k|x_i)$ jest warunkową funkcją prawdopodobieństwa i wyraża się wzorem:

$$p(y_k|x_i) = P(Y = y_k|X = x_i) = \frac{P(Y = y_k, X = x_i)}{P(X = x_i)} = \frac{p_{ik}}{p_{i\bullet}}, \qquad k = 1,2, \dots$$

Jeśli natomiast zmienna losowa (X, Y) jest typu ciągłego, to przy pewnych założeniach dotyczących granicy dystrybuantę warunkową zmiennej losowej Y pod warunkiem, że $X = x$ definiujemy wzorem:

$$F(y|x) = \int_{-\infty}^{y} f(y|x)\, dy,$$

gdzie $f(y|x)$ jest warunkową funkcją gęstości zmiennej losowej Y:

$$f(y|x) = \frac{f(x,y)}{f_1(x)}, \qquad f_1(x) > 0$$

Warunkową wartość średnią zmiennej losowej Y pod warunkiem, że $X = x$ (lub $X = x_i$ dla zmiennej typu skokowego) definiujemy ogólnie jako:

$$E(Y|X = x) = \int_{-\infty}^{\infty} y\, dF(y|x) = \begin{cases} \sum_k y_k p(y_k|x_i), & \text{gdy } (X,Y) \text{ jest typ skokowego} \\ \int_{-\infty}^{\infty} y f(y|x)\, dy, & \text{gdy } (X,Y) \text{ jest typu ciągłego} \end{cases}$$

zad. 10

Sformułować słabe prawo wielkich liczb Markova i podać (krótko) jego związek z teorią estymacji.

Niech $X_1, X_2, \dots$ będzie ciągiem niezależnych zmiennych losowych o średnich $E(X_i) = m_i$ i wariancjach $D^2(X_i) = {\sigma_i}^2, \ \ i = 1,2, \dots$

Utwórzmy zmienną losową:

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Oczywiście:

$$E(M_n) = \frac{\sum_{i=1}^{n} m_i}{n}$$

$$D^2(M_n) = \frac{\sum_{i=1}^{n} {\sigma_i}^2}{n^2}$$

Jeśli przy powyższych założeniach $\lim_{n\to\infty} D^2(M_n) = 0 \ \ to$

$$\bigwedge_{\varepsilon>0} \lim_{n\to\infty} P(|M_n - E(M_n)| \geq \varepsilon) = 0$$

Czyli ciąg $M_1, M_2, \dots$ jest zbieżny wg prawdopodobieństwa do $E(M_n)$, czyli do zmiennej losowej o rozkładzie jednopunktowym $E(M_n)$.

DOKOŃCZYĆ!!!

zad. 11

Zdefiniować proces losowy o niezależnych przyrostach i proces Markova. Podać związek między nimi.

Proces losowy $X(t)$ nazywamy procesem o przyrostach niezależnych jeśli dla dowolnego n, dowolnego układu chwil $t_1 < t_2 < \cdots < t_n$ zmienne losowe:

$$X(t_1), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$$

są niezależne.

Proces losowy $X(t)$ nazywamy procesem Markova jeśli dla dowolnego n, dowolnego układu chwil $t_1 < t_2 < \cdots < t_n$ oraz dla dowolnych liczb rzeczywistych $x_1, x_2, \dots, x_n$ zachodzi:

$$P[X(t_n) < x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_1) = x_1] = P[X(t_n) < x_n | X(t_{n-1}) = x_{n-1}]$$

Proces losowy o przyrostach niezależnych taki, że $P[X(t_1) = c] = 1$ (c – stała) jest procesem Markova (ale nie odwrotnie).

zad. 12

Zdefiniuj rozkłady brzegowe zmiennej losowej X. Podaj dystrybuantę dla typu skokowego, ciągłego i w przypadku ogólnym.

Jeśli dla zmiennej losowej (X, Y) interesuje nas rozkład tylko jednej z jej zmiennych składowych, podczas gdy druga przyjmuje wartości dowolne, to mówimy o rozkładzie brzegowym odpowiedniej zmiennej.

Jeśli zatem zmienna losowa (X, Y) jest typu skokowego, to:

$$\bigwedge_i P(X = x_i) = P(X = x_i, Y = y_1) + P(X = x_i, Y = y_2) + \cdots = \sum_k p_{ik} = p_{i\bullet}$$

definiuje funkcję prawdopodobieństwa rozkładu brzegowego zmiennej losowej X.

Jeśli zmienna losowa (X, Y) jest typu ciągłego, to funkcja gęstości rozkładu brzegowego zmiennej losowej X wyraża się wzorem:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

Ogólnie dla zmiennej losowej (X, Y) dystrybuanta rozkładu brzegowego zmiennej losowej X dana jest wzorem:

$$F_1(x) = F(x, \infty) = P(X < x, Y < \infty)$$

W szczególności, jeśli zmienna losowa (X, Y) jest typu skokowego:

$$F_1(x) = \sum_{x_i < x} p_{i\bullet} = \sum_{x_i < x} \sum_k p_{ik}$$

Jeśli natomiast (X, Y) jest typu ciągłego:

$$F_1(x) = \int_{-\infty}^{x} f_1(x) dx = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx$$

zad. 13

W pewnym urządzeniu trzeba przeciętnie 7 razy w roku (≈7000 h pracy) wymieniać pewien podzespół. Zakładając, że liczba wymian tego podzespołu w danym okresie ma rozkład Poissona, wyliczyć prawdopodobieństwo tego, że konieczność wymiany podzespołu wystąpi po 100 godzinach pracy.

Aby obliczyć prawdopodobieństwo konieczności wymiany podzespołu po 100 h pracy, możemy obliczyć prawdopodobieństwo tego, że w ciągu 100 h będzie trzeba dokonać wymiany 0 razy.

p – prawdopodobieństwo wystąpienia konieczności wymiany

k – liczba wymian podzespołu

n – liczba godzin

$$\lambda = n * p = 100 * \frac{7}{7000} = 0.1$$

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-0.1}0.1^0}{0!} = e^{-0.1}$$

zad. 14

Oblicz $\boldsymbol{E(X)}$ i $\boldsymbol{D^2(X)}$ rozkładu Poissona.

$$E(X) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

$$D^2(X) = E(X^2) - E^2(X)$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} [k(k-1) + k] \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \left[\sum_{k=1}^{\infty} \frac{k(k-1)\lambda^k}{k!} + \sum_{k=1}^{\infty} \frac{k\lambda^k}{k!} \right]$$

$$= e^{-\lambda} \left[\lambda^2 \sum_{k=1}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] = e^{-\lambda} \left(\lambda^2 e^{\lambda} + \lambda e^{\lambda} \right) = \lambda^2 + \lambda$$

$$D^2(X) = \lambda$$

zad. 15

Podaj własności wariancji i udowodnij, że suma wariancji jest równa wariancji sumy zmiennych.

Własności wariancji $D^2(X)$:

1. $D^2(X+c) = D^2(X)$
2. $D^2(X) = E(X^2) - E^2(X)$
3. $\bigwedge_{c \neq m} D^2 < E(X-c)^2$

nierówność Czebyszewa:

4. jeśli $D^2(X) < \infty$ to $\bigwedge_{\varepsilon > 0}$ zachodzi:
   $P\{\varpi : |X-m| \geq \varepsilon\} \leq \dfrac{\delta^2}{\varepsilon^2}$

Dowód:

$$\begin{aligned} D^2(X+Y) &= E((X+Y)^2) - E^2(X+Y) = E(X^2 + 2XY + Y^2) - [E(X) + E(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E^2(X) - 2E(X)E(Y) - E^2(Y) = \end{aligned}$$

dla zmiennych niezależnych wartość średnia iloczynu jest równa iloczynowi wartości średnich

$$= [E(X^2) - E^2(X)] + [E(Y^2) - E^2(Y)] + 2E(X)E(Y) - 2E(X)E(Y) = D^2(X) + D^2(Y)$$

zad. 16

Zdefiniować proces losowy, jego wartość średnią i funkcję korelacyjną. Kiedy można wyznaczyć te charakterystyki dysponując pojedynczą realizacją procesu (warunek dostateczny dla procesów stacjonarnych w szerszym sensie + odpowiednie wzory).

Procesem losowym nazywamy rodzinę zmiennych losowych $\{X_t(\omega), t \in T\}$ określonych na przestrzeni probabilistycznej $\{\Omega, \mathcal{A}, \mathcal{P}\}$. Inaczej mówiąc, proces losowy to losowa funkcja parametru $t \in T$ czyli taka, która dla każdej wartości tego parametru jest zmienną losową.

Wartością średnią procesu losowego $X(t)$ nazywamy funkcję $m(t)$, będącą dla każdego $t \in T$ wartością średnią zmiennej losowej $X(t)$, którą proces ten jest w chwili t:

$$m(t) = E[X(t)] = \int_{-\infty}^{\infty} x \, dF(x,t)$$

Funkcją korelacyjną procesu losowego $X(t)$ definiujemy wzorem:

$$R_X(t_1, t_2) = E\{[X(t_1) - m(t_1)][X(t_2) - m(t_2)]\}$$

Analogicznie definiujemy funkcję korelacyjną rzędu $n > 2$:

$$R_X(t_1, t_2, \dots, t_n) = E\{[X(t_1) - m(t_1)][X(t_2) - m(t_2)] \dots [X(t_n) - m(t_n)]\}$$

Powyższe charakterystyki można wyznaczyć dysponując pojedynczą realizacją procesu jeżeli proces jest procesem ergodycznym. Warunek dostateczny ergodyczności procesów losowych stacjonarnych w szerszym sensie o ciągłej funkcji korelacyjnej ma postać:

$$\lim_{|\tau| \to \infty} R_X(\tau) = 0$$

Przykładowo, dla procesu stacjonarnego w szerszym sensie ergodyczność względem wartości średniej oznacza, że:

$$E[X(t)] = m \approx \frac{1}{T} \int_0^T X(t) \, dt$$

Natomiast ergodyczność tego procesu względem funkcji korelacyjnej oznacza, że:

$$R_X(\tau) = E\{[X(t) - m][(X(t+\tau) - m]\} \approx \frac{1}{T} \int_0^T [X(t) - m][(X(t+\tau) - m] \, dt$$

zad. 17

Definicja próby losowej, estymatora parametru q cechy X. Dany jest rozkład $N(5000; 50)$, próba n-elementowa i średnia m. Obliczyć prawdopodobieństwo, że średnia m jest z przedziału $[5050; 5100]$.

Próbą losową nazywamy wektor losowy $(X_1, X_2, \dots, X_n)$, którego elementy są niezależnymi zmiennymi losowymi o rozkładzie prawdopodobieństwa takim samym jak rozkład interesującej nas cechy X (zmiennej losowej).

Estymatorem parametru q cechy X w populacji generalnej nazywamy każdą statystykę

$$Q_n = f(X_1, X_2, \dots, X_n)$$

taką, że:

$$\bigvee_{c>0} \lim_{n\to\infty} P(|Q_n - q| < c) = 1$$

Wartość estymatora Q_n wyliczoną z próby odpowiadającej pojedynczemu losowaniu nazywamy oceną parametru q i oznaczamy jako: $q_n = f(x_1, x_2, \dots, x_n)$.

Wiemy, że rozkład normalny jest nieskończenie podzielny, a zatem zmienna losowa $M_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ ma rozkład $N(m, \sigma_{M_n})$, gdzie $\sigma_{M_n} = \frac{\sigma}{\sqrt{n}}$.

$$P(M_n \in [a, b]) = F(b) - F(a) = \frac{1}{\sigma_{M_n}\sqrt{2\pi}} \int_a^b exp\left[\frac{-(m_n - m)^2}{2{\sigma_{M_n}}^2}\right] dm_n$$

W celu skorzystania z tablic unormujemy zmienną losową podstawiając:

$$Y = \frac{M_n - m}{\sigma_{M_n}}$$

Zatem:

$$P(M_n \in [a, b]) = P(Y \in [y_1, y_2]) = F(y_2) - F(y_1) = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}}\, dy = p_u > 0$$

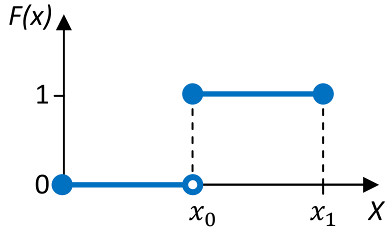
gdzie:

$$y_1 = \frac{a - m}{\sigma_{M_n}}$$

$$y_2 = \frac{b - m}{\sigma_{M_n}}$$

zad. 18

Narysować wykres dystrybuanty oraz wyprowadzić wzór na wartość średnią i wariancję zmiennej losowej o rozkładzie dwupunktowym.



$$E(X) = x_1 p + x_2(1-p) = x_1 p + x_2 - x_2 p = p(x_1 - x_2) + x_2$$

$$D^2(X) = [x_1 - E(X)]^2 p + [x_2 - E(X)]^2 (1-p) = \cdots = (x_1 - x_2)^2 (1-p)p$$

zad. 19

Zdefiniuj zmienną losową, rozkład prawdopodobieństwa i dystrybuantę. Wykreślić dystrybuantę dla rzutu sześcienną kostką do gry.

Zmienną losową nazywamy funkcję rzeczywistą X określoną na przestrzeni zdarzeń elementarnych Ω i taką, że przeciwobrazy zbiorów borelowskich na prostej (w R^1) otrzymane za pomocą tej funkcji są zdarzeniami losowymi. Wystarczy wymagać, by przeciwobrazy przedziałów postaci $(-\infty; x)$ były zdarzeniami losowymi, gdyż każdy zbiór borelowski na prostej można otrzymać za pomocą przeliczalnej liczby działań na takich przedziałach.

Innymi słowy, zmienna losowa X to funkcja działająca z Ω w R^1 taka, że:

$$\bigwedge_x X^{-1}[(-\infty; x)] = \{\omega: X(\omega) < x\} \in \mathcal{A}$$

Widzimy zatem, że zmienna losowa to funkcja odwzorowująca Ω w prostą, mierzalna względem prawdopodobieństwa.

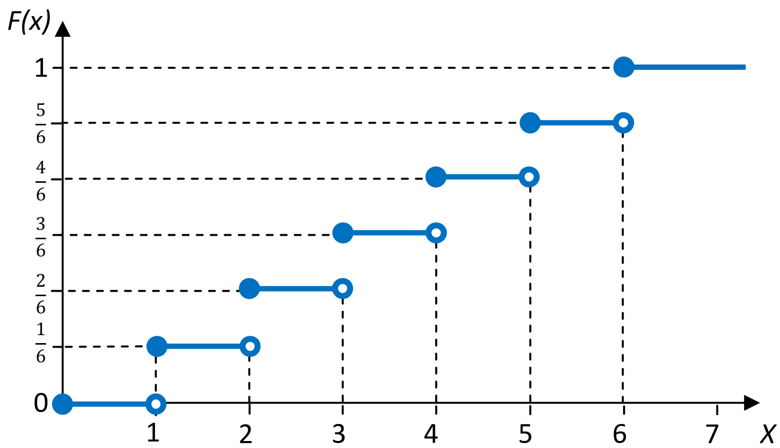
Rozkładem prawdopodobieństwa zmiennej losowej X nazywamy prawdopodobieństwo indukowane czyli funkcję zbioru daną wzorem:

$$P_X(S) = P[X^{-1}(S)] = P[\omega: X(\omega) \in S], \qquad S \in \mathcal{A}_X$$

Dystrybuantą zmiennej losowej X nazywamy funkcję rzeczywistą $F(x)$ określoną wzorem:

$$F(x) = P_X[(-\infty; x)] = P[X^{-1}(-\infty; x)] = P\{\omega: X(\omega) \in (-\infty; x)\} = P(X < x)$$

Wykres dystrybuanty dla rzutu sześcienną kostką do gry:



zad. 20

Obliczyć $E(X)$ i $D^2(X)$ dla rozkładu wykładniczego.

$$E(X) = \int_{-\infty}^{\infty} x\lambda e^{-\lambda x}\,dx = \left| \begin{matrix} u = x & u' = 1 \\ v' = \lambda e^{-\lambda x} & v = -e^{-\lambda x} \end{matrix} \right| = \left[-xe^{-\lambda x}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\lambda x}\,dx =$$

$$= \left[-xe^{\lambda x} - \frac{e^{-\lambda x}}{\lambda}\right]_{-\infty}^{\infty} = \left[-e^{\lambda x}\frac{\lambda x + 1}{\lambda}\right]_{-\infty}^{\infty}$$

dla $x \leq 0 \quad f(x) = 0$, możemy więc zmienić granice całkowania:

$$\left[-e^{\lambda x}\frac{\lambda x + 1}{\lambda}\right]_{0}^{\infty} = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}$$

$$E(X^2) = \int_{0}^{\infty} x^2\lambda e^{-\lambda x}\,dx = \left| \begin{matrix} u = x^2 & u' = 2x \\ v' = \lambda e^{-\lambda x} & v = -e^{-\lambda x} \end{matrix} \right| = \left[-x^2e^{-\lambda x}\right]_{0}^{\infty} + 2\int_{0}^{\infty} xe^{-\lambda x}\,dx = \frac{2}{\lambda}E(X) = \frac{2}{\lambda^2}$$

$$D^2(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

zad. 21

Sformułuj twierdzenie Bayesa.

Jeśli $A_1, A_2, \ldots$ tworzą podział przestrzeni $\Omega, P(A_i) > 0, i = 1,2, \ldots$ oraz B jest zdarzeniem, dla którego $P(B) > 0$, to słuszny jest następujący wzór (zwany wzorem Bayesa):

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}, \qquad i = 1,2, \ldots$$

zad. 22

Definicja estymatora.

Estymatorem parametru q cechy X w populacji generalnej nazywamy każdą statystykę

$$Q_n = f(X_1, X_2, \dots, X_n)$$

taką, że:

$$\bigvee_{c>0} \lim_{n\to\infty} P(|Q_n - q| < c) = 1$$

Wartość estymatora Q_n wyliczoną z próby odpowiadającej pojedynczemu losowaniu nazywamy oceną parametru q i oznaczamy jako: $q_n = f(x_1, x_2, \dots, x_n)$.

zad. 23

Definicja próby losowej, sposób pobrania próby losowej, jakie postulaty musi spełniać próba losowa.

Próbą losową nazywamy wektor losowy $(X_1, X_2, \dots, X_n)$, którego elementy są niezależnymi zmiennymi losowymi o rozkładzie prawdopodobieństwa takim samym jak rozkład interesującej nas cechy X (zmiennej losowej).

Próba powinna być reprezentatywna, tzn. spełniać postulaty:

a. jednakowej szansy trafienia do próby każdego elementu populacji generalnej
b. odpowiedniej liczności

Próba prosta musi spełniać dodatkowo postulat:

c. niezależności badań elementów trafiających do próby.

zad. 24

Definicja statystyki.

Statystyką nazywamy każdą funkcję próby losowej, czyli zmienną losową Y postaci:

$$Y = f(X_1, X_2, \dots, X_n), \qquad f : R^n \to R^1$$

zad. 25

W jaki sposób estymujemy wartość cechy X.

Pobierzmy próbę n-elementową $(x_1, x_2, \dots, x_n)$ i wyznaczmy dla niej wartość:

$$\sigma_n{}^2 = \overline{(x_\iota - \bar{x})^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

którą przyjmujemy jako ocenę estymatora wariancji danego wzorem:

$$S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2$$

gdzie $(X_1, X_2, \dots, X_n)$ jest próbą losową, a M_n jest estymatorem parametru m.

Żeby zbadać zgodność estymatora $S_n{}^2$ rozpatrzmy zmienne losowe $Z_i = (X_i - M_n)^2, \;\; i = 1,2, \dots, n$. Widzimy, że:

$$S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

spełnia założenia słabego prawa wielkich liczb Chińczyna, zatem ciąg $S_1{}^2, S_2{}^2, \dots$ jest stochastycznie zbieżny do $E(Z) = \sigma^2$. Estymator jest więc estymatorem zgodnym.

Żeby zbadać, czy estymator jest nieobciążony należy obliczyć:

$$E\left(S_n{}^2\right) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2\right]$$

Pamiętając, że $D^2(M_n) = \frac{\sigma^2}{n}$ otrzymujemy $E\left(S_n{}^2\right) = \frac{n-1}{n}\sigma^2$. Jest to więc estymator obciążony. Na podstawie otrzymanego wzoru widzimy natomiast, że estymator nieobciążony powinien mieć postać:

$$S_n{}^{2'} = \frac{1}{n-1} S_n{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$$

Nie jest to jednak estymator najefektywniejszy dla znanych nam rozkładów parametru σ^2.

zad. 26

W wyniku długotrwałych obserwacji znana jest wariancja pomiarów pewnej wielkości (wariancja $= \mathbf{0.02}$). Ilu trzeba dokonać pomiarów tej wielkości, aby z prawdopodobieństwem $\geq \mathbf{0.99}$ średnia arytmetyczna otrzymanych wyników różniła się od wartości średniej o $\leq \mathbf{0.01}$?

$\varepsilon = 0.1$

$\sigma^2 = 0.02$

$${\sigma_{M_n}}^2 = \frac{\sigma^2}{n}$$

$P(|M_n - E(M_n)| \geq \varepsilon) \leq 1 - 0.99 = 0.01$

Wzór Czebyszewa:

$$P(|M_n - E(M_n)| \geq \varepsilon) \leq \frac{{\sigma_{M_n}}^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} = \frac{0.02}{n * 0{,}1^2} = \frac{2}{n}$$

zatem:

$$\frac{2}{n} \leq 0.01$$

$$n \geq 200$$

Aby otrzymać podaną dokładność trzeba dokonać przynajmniej 200 pomiarów.

zad. 27

Wykazać, że wariancja sumy niezależnych zmiennych losowych X, Y jest równa sumie wariancji tych zmiennych.

$$\begin{aligned} D^2(X+Y) &= E((X+Y)^2) - E^2(X+Y) = E(X^2+2XY+Y^2) - [E(X)+E(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E^2(X) - 2E(X)E(Y) - E^2(Y) = \end{aligned}$$

dla zmiennych niezależnych wartość średnia iloczynu jest równa iloczynowi wartości średnich

$$= [E(X^2) - E^2(X)] + [E(Y^2) - E^2(Y)] + 2E(X)E(Y) - 2E(X)E(Y) = D^2(X) + D^2(Y)$$

zad. 28

Definicja procesu losowego (stacjonarnego chyba) w szerszym sensie oraz jego wartość średnia (skorzystać z ergodyczności).

Proces losowy $X(t)$ dla którego istnieją: wartość średnia $m(t)$ i funkcja korelacji $R_X(t_1, t_2)$ nazywamy stacjonarnym w szerszym sensie jeśli:

$m(t) = m$

$R_X(t_1, t_2) = R_X(\tau), \tau = t_2 - t_1$

Jeżeli proces losowy jest stacjonarny w szerszym sensie oraz:

$$\lim_{|\tau| \to \infty} R_X(\tau) = 0$$

to jest on procesem ergodycznym, Oznacza to, że:

$$E[X(t)] = m \approx \frac{1}{T} \int_0^T X(t)\, dt$$

zad. 29

Estymator najefektywniejszy wartości średniej.

Pobierzmy próbę n-elementową $(x_1, x_2, \dots, x_n)$ i wyznaczmy z niej wartość $m_n = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, którą przyjmiemy jako ocenę parametru m, czyli wartość estymatora tego parametru danego wzorem:

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

gdzie $(X_1, X_2, \dots, X_n)$ jest próbą losową.

Dla zmiennej losowej M_n spełnione są wszystkie założenia słabego prawa wielkich liczb Chińczyna, zatem estymator ten jest stochastycznie zbieżny do m (jest estymatorem zgodnym tego parametru). Jest również estymatorem nieobciążonym parametru m, ponieważ:

$$E(M_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{m * n}{n} = m, \qquad n = 1,2, \dots$$

Żeby zbadać czy jest to estymator najefektywniejszy musimy przyjąć założenie o rozkładzie cechy X populacji generalnej. Załóżmy więc, że jest to rozkład normalny $N(m, \sigma)$, czyli:

$$f(x,q) = f(x,m) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$$

zatem:

$$\ln f(x,m) = -\ln\left(\sigma\sqrt{2\pi}\right) - \frac{(x-m)^2}{2\sigma^2}$$

oraz:

$$\frac{\partial(\ln f(x,m))}{\partial m} = \frac{x-m}{\sigma^2}$$

Podstawiając do prawej strony nierówności Rao-Cramera otrzymujemy:

$$\frac{1}{n * E\left[\left(\frac{x-m}{\sigma^2}\right)^2\right]} = \frac{1}{n * E\left[\frac{(x-m)^2}{\sigma^4}\right]} = \frac{1}{n * \frac{\sigma^2}{\sigma^4}} = \frac{\sigma^2}{n}$$

Po lewej stronie nierówności mamy:

$$D^2(M_n) = D^2\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Nierówność R-C jest spełniona, zatem estymator M_n przy założeniu, że cecha X ma rozkład normalny jest estymatorem najefektywniejszym.

zad. 30

Podać σ-algebrę zbiorów borelowskich.

Algebrę $\mathcal{A}$ podzbioru przestrzeni Ω nazywamy δ-algebrą zbioru jeśli:

$(A_1, A_2, \ldots \in \mathcal{A}) \Rightarrow (\bigcup_{i=1}^{\infty} A_i \in \mathcal{A})$, z prawa de Morgana

$A_1, A_2, \ldots \in \mathcal{A}$, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

zatem δ-algebra zbioru, to klasa zbioru zamknięta ze względu na wszystkie operacje przeliczalne.

δ-algebra podzbioru przestrzeni Ω generowana przez las wszystkich zbiorów otwartych (domkniętych) nazywamy δ-algebrą zbiorów borelowskich a jej elementy zbiorami borelowskimi.

by Marcin R.