Ostatni raz wygenerowano: 13 czerwca 2023. Kontakt: thasiow(at)onet.pl

1 Egzamin z 26 października 1996

Zadanie 1.1 Mamy dwoch strzelców. Prawdopodobieństwo trafienia w cel pojedynczym strzałem dla lepszego z nich 0.8, a dla gorszego 0.4 Nie wiemy, który z nich jest gorszy, a który lepszy. Testujemy strzelców poddając ich ciągowi prób, z których kazda polega na oddaniu jednego strzału przez kazdego z nich. Test przerywamy po pierwszej takiej próbie, w wyniki której jeden ze strzelców trafił, a drugi spudłował. Nastepnie strzelec, który w ostatniej próbie trafił, oddaje jeszcze jeden strzał. Jakie jest prawdopodobieństwo, iż tym razem trafił w cel?

Rozwiązanie 1.1 *To zadanie na prawdopodobieństwo warunkowe. Niech zdarzenie A oznacza zdarzenie w którym strzelec, który trafił w 1 próbie trafi jeszcze raz, zdarzenie B natomiast niech oznacza, że trafił dokładnie jeden strzelec*

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

Prawdopodobieństwo, że trafił jeden strzelec wynosi:

Przypadek 1: trafia lepszy, nie trafia gorszy: $0.8 \cdot 0.6 = 0.48$
Przypadek 2: trafia gorszy, nie trafia lepszy: $0.4 \cdot 0.2 = 0.08$

Prawdopodobieństwo zdarzenia B wynosi więc $P(B) = 0.48 + 0.08 = 0.56$

Prawdopodobieństwo zdarzenia $A \wedge B$*:*

Przypadek 1 wraz z tym, że lepszy trafi jeszcze raz: $0.48 \cdot 0.8 = 0.384$
Przypadek 2 wraz z tym, że gorszy trafi jeszcze raz $0.08 \cdot 0.4 = 0.032$

Czyli

$$P(A|B) = \frac{0.384 + 0.032}{0.56} = \frac{26}{25}$$

Zadanie 1.2 *Tarczę strzelniczą umieszczamy na płaszczyźnie ze środkiem tarczy..*

Rozwiązanie 1.2 *Do rozwiązania.*

Zadanie 1.3 Oblicz $Pr(\min(k_1, k_2, k_3) = 3)$, jesli k_1, k_2, k_3 to liczby oczek uzyskane w wyniku rzutu trzema (uczciwymi) kostkami.

Rozwiązanie 1.3 *Rozpatrujemy wszystkie przypadki: *3 możliwości ułożenia wyników dla kostek*

	Kostka 1	Kostka 2	Kostka 3	Liczba możliwości
Przypadek1	3	(4,5,6)	(4,5,6)	$3 \cdot 3 \cdot 3$ *
Przypadek 2	3	3	(4,5,6)	$3 \cdot 3$
Przypadek 3	3	3	3	1

Razem wszystkich możliwośic jest $6 \cdot 6 \cdot 6$*, stąd szukane prawdopodobieństwo jest równe*

$$Pr(\min(k_1, k_2, k_3) = 3) = \frac{3 \cdot 3 \cdot 3 + 3 \cdot 3 + 1}{6 \cdot 6 \cdot 6} = \frac{37}{216}$$

Zadanie 1.4 Funckja gęstości dana jest wzorem:

$$f_{X,Y}(x,y) = \begin{cases} x + y & \text{dla } (x,y) \in (0,1)x(0,1), \\ 0 & \text{poza tym} \end{cases}$$

$E(X|Y = \frac{1}{2})$ wynosi: ?.

Rozwiązanie 1.4 *Rysujemy rysunek pomocniczy. Mamy:*

$$E(X|Y = 1/2) = \frac{\int_0^1 x(x+1/2)dx}{\int_0^1 (x+1/2)dx} = \frac{7}{12}$$

Zadanie 1.5 *Do zrobienia.*

Zadanie 1.6 *Do zrobienia.*

Zadanie 1.7 *Do zrobienia.*

Zadanie 1.8 Niech X ma funkcję gęstości

$$f_X(x) = \begin{cases} 0.5 \cdot x + 0.5 & \text{dla } -1 \leqslant x \leqslant 1, \\ 0 & \text{poza tym} \end{cases}$$

gęstość $f_Y(y)$ zmiennej losowej $Y = X^2$ dana jest dla $y \in (0,1)$ wzorem:?

Zadanie 1.9 *Tu nie możemy skorzystać ze wzoru wykorzystującego Jakobian (pochodna się zeruje w zerze). Więc zastosujemy podejście wykorzystujące dystrybuantę. Mamy:*

$$P(X^2 < t) = P(\sqrt{t} \leqslant X \leqslant \sqrt{t}) = \int\int_{-\sqrt{t}} \sqrt{t} 0.5x + 0.5 = \sqrt{t}$$

Gęstość:

$$f(t) = \frac{\partial \sqrt{t}}{\partial t} = \frac{1}{2\sqrt{t}}$$

2 Egzamin z 11 października 2004

Zadanie 2.1 Obserwujemy działanie pewnego urządzenia w kolejnych chwilach $t = 0, 1, 2, \ldots$. Działanie tego urządzenia zależy od pracy dwóch podzespołów A i B. Każdy z nich moze ulec awarii w jednostce czasu z prawdopodobieństwem 0.1 niezależnie od drugiego. Jeżeli jeden z podzespołów ulega awarii, to urządzenie nie jest naprawiane i działa dalej wykorzystując drugi podzespół. Jeżeli oba podzespoły są niesprawne w chwili t, to następuje ich naprawa i w chwili $t + 1$ oba są sprawne. Prawdopodobieństwo, że podzespół B jest sprawny w chwili t dąży, przy t dążącym do nieskończoności, do następującej liczby (z dokładnością do 0.001): ?.

Rozwiązanie 2.1 *Mamy 4 stany:*
A, B *- oba urządzenie działają*
$\sim A, B$ *- urządzenie A uległo awarii*
$A, \sim B$ *- urządzenie B uległo awarii*
$\sim A, \sim B$ *- oba urządzenia nie działają*

Macierz prawdopodobieństw przejścia stanów w jednym kroku: zauważmy, że

	A, B	$\sim A, B$	$A, \sim B$	$\sim A, \sim B$
A, B	0.9^2	0.09	0.09	0.01
$\sim A, B$	0	0.9	0	0.1
$A, \sim B$	0	0	0.9	0.1
$\sim A, \sim B$	1	0	0	0

wiersze zawsze sumują się do 1. Gybyśmy chcieli uzyskać prawdopobieństwa przejścia w 2 kroków to musielibyśmy zrobić iloczyn powyższej macierzy. W przypadku 3 kroków byłaby to macierz do potęgi 3-ciej. Definiujemy prawdopodobieństwa, że po nieskończenie długim czasie jesteśmy w danym stanie (na egzaminach te prawdopodobieństwa zawsze istnieją, w teorii nie muszą).

Dla stanu A, B *definiujemy* P_1
Dla stanu $\sim A, B$ *definiujemy* P_2
Dla stanu $A, \sim B$ *definiujemy* P_3
Dla stanu $\sim A, \sim B$ *definiujemy* P_4

Teraz zapisujemy wykorzystując informacje z kolumn (do zapamietania: odwrotnie niż sumowanie do jedynki):

$$P_1 = 0.81 \cdot P_1 + 0 \cdot P_2 + 0 \cdot P_3 + 1 \cdot P_4$$

$$P_2 = 0.09 \cdot P_1 + 0.9 \cdot P_2$$

$$P_3 = 0.09 \cdot P_1 + 0.9 \cdot P_3$$

$$P_4 = 0.01 \cdot P_1 + 0.1 \cdot P_2 + \cdot P_3$$

oraz

$$P_1 + P_2 + P_3 + P_4 = 1$$

szukamy prawdopodobieństwa, że urządzenie B jest sprawne, czyli

$$P_1 + P_2 = ?$$

Rozwiązując układ równań mamy:

$$P_1 + P_2 = 0.6354$$

Zadanie 2.2 Niech $X_1, X_2, \ldots, X_n, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym o gęstości

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{gdy } x > 0 \\ 0 & \text{gdy } x \leqslant 0 \end{cases}$$

gdzie $\alpha > 0$ jest ustalonym parametrem.
Niech N będzie zmienną losową, niezależną od $X_1, X_2, \ldots, X_n, \ldots$, o rozkladzie ujemnym dwumianowym

$$P(N = n) = \binom{n + r - 1}{n} p^r (1 - p)^n$$

dla $n = 0, 1, 2, \ldots$ gdzie $r > 0$ i $p \in (0; 1)$ są ustalonymi parametrami. Niech

$$Z_N = \begin{cases} \min(X_1, X_2, \ldots, X_N) & \text{gdy } N > 0 \\ 0 & \text{gdy } N = 0 \end{cases}$$

Oblicz $E(NZ_N)$ i $Var(NZ_N)$.

Rozwiązanie 2.2 *Korzystając z własności wartości oczekiwanej (Iteracyjność)*

$$E(E(Y|X)) = EY$$

mamy:

$$E(NZ_N) = E[E(NZ_N | N = n)]$$

W 1 kroku liczymy rozkład zmiennej Z_n, mamy:

$$F(x) = \int_0^x \alpha e^{-\alpha s} ds = 1 - e^{-\alpha x}$$

$$P(\min(X_1, \ldots, X_N) < t) = 1 - P(\min(X_1, \ldots, X_N) > t) = 1 - P(X_1 > t)^n =$$
$$= 1 - (1 - (1 - e^{-\alpha x}))^n = 1 - e^{-\alpha x n}$$
$$f_{Z_N}(z) = \alpha \cdot n \cdot e^{-\alpha x n}$$

czyli Z_N ma rozkład wykładniczy z $\beta = \alpha n$, jego wartość oczekiwana to $\frac{1}{\alpha n}$. Stąd mamy:

$$E(N \cdot Z_N) = \begin{cases} \frac{1}{\alpha} & z\ p\text{-}p\ \sum_{n=1}^{\infty} \binom{n+r-1}{n} p^r (1-p)^n = 1 - p^r \\ 0 & dla\ N = 0 \end{cases}$$

w powyższym należy zauważyć, że w $\frac{1}{\alpha n}$ skróciło się n. Powyższa równość wynika z faktu, że w rozkładzie ujemnym dwumianowym mamy:

Rozkład ujemny dwumianowy

$$p_k = \frac{\Gamma(r+k)}{\Gamma(r)k!} p^r (1-p)^k$$

dla $k = 0, 1, \ldots$

$$E(X) = \frac{r(1-p)}{p}$$

$$VAR(X) = \frac{r(1-p)}{p^2}$$

gdzie

$$\frac{\Gamma(r+k)}{\Gamma(r)k!} = \frac{(r+k-1)!}{(r-1)!k!} = \binom{r+k-1}{k}$$

stąd suma po p_k od $k = 0$ daje 1, dla $k = 0$ mamy p^r stąd cała suma daje $1 - p^r$.

Ostatecznie mamy:

$$E(N \cdot Z_N) = \frac{1}{\alpha}(1 - p^r) + 0 \cdot p^r = \frac{1}{\alpha}(1 - p^r)$$

drugą częścią zadania jest obliczenie wariancji.

$$E(N^2 \cdot Z_N^2) = ?$$

Wiemy, że Z_N ma rozkład wykładniczy z $\beta = \alpha n$

Rozkład wykładniczy

$$f(x) = \beta e^{-\beta x}$$

$$E(X) = \frac{1}{\beta}$$

$$VAR(X) = \frac{1}{\beta^2}$$

stąd

$$Var(Z_N) = \frac{1}{(\alpha n)^2}$$

czyli

$$E(Z_N^2) = \frac{1}{(\alpha n)^2} + \frac{1}{(\alpha n)^2} = \frac{2}{(\alpha n)^2}$$

$$E(N^2 \cdot Z_N^2) = \begin{cases} \frac{2}{\alpha^2} & z\ p\text{-}p\ \sum_{n=1}^{\infty} \binom{n+r-1}{n} p^r (1-p)^n = 1 - p^r \\ 0 & dla\ N = 0 \end{cases}$$

stąd

$$E(N^2 Z_n^2) = \frac{2}{\alpha^2}(1 - p^r)$$

$$Var(NZ_N) = \frac{2}{\alpha^2}(1-p^r) - \frac{1}{\alpha^2}(1-p^r)^2 = \frac{2(1-p^r) - (1-p^r)^2}{\alpha^2} = $$
$$= \frac{1 - p^{2r}}{\alpha^2}$$

Zadanie 2.3 Niech (X, Y) będzie dwumiarową zmienną losową o funkcji gęstości

$$f(x, y) = \begin{cases} e^{-x} & \text{gdy } x > 0 \text{ i } y \in (0; 1) \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

Niech $Z = X + 2Y$. Wtedy łączny rozkład zmiennych Z, X jest taki, że:?

Rozwiązanie 2.3 *W ogólności: Niech X - zmienna o gęstości f. Żeby obliczyć gęstość zmiennej Y zdefiniowanej jako:*

$$Y = g(x)$$

używamy następującego wzoru:

$$g(y) = f(h(y))|h'(y)|$$

gdzie h - funkcja odwrotna do g (tzn. $h(g(t)) = t$).

Robimy zamianę zmiennych żeby uzyskać łaczny rozkład $g(z, x)$.

Krok 1: funkcja/e odwrotna/e

$$\begin{cases} Z = X + 2Y \\ V = X \end{cases} \quad \rightarrow \quad Y = \tfrac{1}{2}Z - \tfrac{1}{2}V$$

Krok 2: Jakobian i wyznacznik (Licząc jakobian: w pierwszym wierszu pochodne starej zmiennej po nowych zmiennych, w drugim wierszu pochodne drugiej starej zmiennej po nowych zmiennych). Mamy:

$$J = \left\| \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial Z} & \frac{\partial Y}{\partial V} \end{vmatrix} \right\| = \left\| \begin{vmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} \right\| = \left| 0 - \frac{1}{2} \right| = \frac{1}{2}$$

Krok 3:

$$g(y) = f(h(y))|h'(y)|$$

czyli

$$g(z,x) = e^{-v} \cdot \frac{1}{2} = \frac{1}{2} \cdot e^{-x}$$

dla $x > 0$ *i* $Y \in (0,1)$ *stąd dla* $z \in (x, x+2)$.

Zadanie 2.4 Dysponujemy $N+1$, $(N > 1)$ identycznymi urnami. Każda z nich zawiera N kul białych i czarnych. Liczba kul białych w i-tej urnie jest równa $i-1$, gdzie $i = 1, 2, \ldots, N+1$.
Losujemy urnę, a następnie ciągniemy z niej jedną kulę i okazuje się, że otrzymana kula jest biała. Oblicz prawdopodobieństwo, że ciągnąc drugą kulę z tej samej urny (bez zwracania pierwszej) również otrzymamy kulę białą.

Rozwiązanie 2.4 *rozpatrujemy pierwsze zdanie: losujemy urnę i ciągniemy kulę, kula jest biała. Tu mamy prawdopodobieństwo warunkowe*

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

gdzie A *oznacza, że wylosowaliśmy urnę z* n *kulami białami a* B *oznacza prawdopodobieństwo, że wylosowaliśmy białą za pierwszym razem. Mamy (p-p wylosowania urny:* $\frac{1}{N+1}$*):*

$$P(B) = \frac{1}{N+1} \cdot \frac{0}{N} + \frac{1}{N+1}\frac{1}{N} + \ldots + \frac{1}{N+1} \cdot \frac{N}{N} =$$

$$= \frac{1}{N+1} \cdot \frac{1}{N} \cdot \frac{0+N}{2} \cdot (N+1) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{N+1} \cdot \frac{n}{N}$$

Stąd

$$P(A|B) = \frac{2n}{N(N+1)}$$

Teraz rozpatrzmy co dzieje się w drugim losowaniu: gdy z tej samej urny chcemy wyslosować również białą mamy

$$\frac{2n}{N(N+1)} \cdot \frac{n-1}{N-1}$$

musimy wysumować powyższe dla wszyskich możliwych liczb kól białych tj.:

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

$$= \frac{2(n-1)N(N+1)}{3N(N+1)(N-1)} = \frac{2}{3}$$

3 Egzamin z 17 czerwca 2013

Zadanie 3.1 Zakładając, że zmienne losowe $X_1, X_2, \ldots, X_{16}$ są niezależne i mają rozkłady normalne $X_i \sim N(m\sqrt{i}, i)$, $i = 1, 2, \ldots, 16$, zbudowano test jednostajnie najmocniejsz dla weryfikacji hipotezy $H_0 : m = 0$ przy alternatywie $H_1 : m > 0$ na poziomie istotności 0.05. W rzeczywistości okazało się, że wektor $(X_1, X_2, \ldots, X_{16})$ ma rozkład normalny taki, że $EX_i = m\sqrt{i}$, $VarX_i = i$, $i = 1, 2, \ldots, 16$, oraz współczynnik korelacji

$$\rho(X_i, X_j) = \begin{cases} 0.5 & \text{gdy } |i-j| = 1 \\ 1 & \text{poza tym } i = j \\ 0 & \text{w pp} \end{cases}$$

Wyznaczyć rzeczywisty rozmiar testu.

Rozwiązanie 3.1 *Dla przypomnienia poziom istotności (rozmiar testu):*

$$P_\theta(K) \leqslant \alpha$$

gdzie $K = \{T(x) > c\}$. Test jednostajnie najmocniejszy → najmniejsze prawdopodobieństwo błedu drugiego rodzaju (nie odrzucenie H_0 gdy jest fałszywa). Buduje się się go przez iloraz funkcji wiarygodności w nastąpujący sposób:

Przy H_0 mamy

$$L_0 = \prod_{i=1}^{16} f(x_i, 0, i)$$

Przy H_1 mamy

$$L_1 = \prod_{i=1}^{16} f(x_i, m\sqrt{i}, i)$$

wiemy, że

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$N(0, i) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{i}} e^{-\frac{x^2}{2i}}$$

$$N(m\sqrt{i}, i) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{i}} e^{-\frac{(x-m\sqrt{i})^2}{2i}}$$

wtedy mamy

$$\frac{L_1}{L_0} \to \textit{Jeżeli to jest zbyt duże to odrzucamy } H_0$$

stąd

$$\frac{L_1}{L_0} = \frac{\prod_{i=1}^{16} \frac{1}{\sqrt{2\pi} \cdot \sqrt{i}} e^{-\frac{(x-m\sqrt{i})^2}{2i}}}{\prod_{i=1}^{16} \frac{1}{\sqrt{2\pi} \cdot \sqrt{i}} e^{-\frac{x^2}{2i}}} = e^{\sum_{i=1}^{16} -\frac{(x_i - m\sqrt{i})^2}{2i} + \frac{x_i^2}{2i}} =$$

$$= e^{\sum_{i=1}^{16} \frac{2x_i m\sqrt{i} - m^2 i}{2i}}$$

H_0 odrzucamy, gdy

$$e^{\sum_{i=1}^{16} \frac{2x_i m\sqrt{i} - m^2 i}{2i}} > t$$

$$\sum_{i=1}^{16} \frac{2x_i m\sqrt{i} - m^2 i}{2i} > t' \rightarrow \sum_{i=1}^{16} \frac{2x_i m\sqrt{i} - m^2 i}{i} > t'$$

Skoro test na poziomie istotności 0.05 to prawdopodobieństwo odrzucenia H_0 przy jej prawdziwości

$$P\left(\sum_{i=1}^{16} \frac{2x_i m\sqrt{i} - m^2 i}{i} > t'\right) = 0.05$$

Dalej

$$P\left(\sum_{i=1}^{16} \frac{x_i\sqrt{i}}{i} > t'\right) = 0.05$$

Z własności rozkładu normalnego:

$$aX + b \sim N(a\mu + b, (a\sigma)^2)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma^2 + \sigma^2)$$

mamy:

$$T = \sum_{i=1}^{16} \frac{x_i}{\sqrt{i}} \sim N(0, 16)$$

bo

$$\frac{x_i}{\sqrt{i}} \sim N(0, \frac{i}{i}) = N(0, 1)$$

musimy unormować sumę bo ma rozkład $N(0, 16)$. mamy

$$Y = \frac{T - 0}{\sigma} = \frac{T}{\sqrt{16}} \sim N(0, 1)$$

stąd

$$P\left(\frac{T}{\sqrt{16}} > t\right) = 0.05 \rightarrow P\left(\frac{T}{\sqrt{16}} < t\right) = 0.95$$

z tablic

$$t = 1.645 \rightarrow t'' = 1.645 \cdot \sqrt{16} = 6.58$$

stąd, jeżeli $\sum_{i=1}^{n} \frac{x_i}{\sqrt{i}} > 6.58$ to odrzucam H_0. Poziom istotności testu = rozmiar testu. Zrobiłem test przy złym poziomie istotności, czyli mam prawdopodobieństwo odrzucenia H_0 jeżeli jest prawdziwa. Wiemy, że $X_1, X_2, \ldots, X_{16}$ są skorelowane. Rzeczywisty rozmiar testu będzie obliczony poprzez:

$$P\left(\sum_{i=1}^{16} \frac{x_i}{\sqrt{i}} > 6.58\right)$$

przy H_0, gdzie $\sum_{i=1}^{16} \frac{x_i}{\sqrt{i}} = V$, a x_i są skorelowane. Wiemy, że $V \sim N(0, ?)$, czyli musimy obliczyć

$$Var(\sum_{i=1}^{16} \frac{x_i}{\sqrt{i}}) = ?$$

Z własności wariancji pamiętamy, że

$$Var(X+Y+Z) = Var(X)+Var(Y)+Var(Z)+2Cov(X,Y)+2Cov(X,Z)+2Cov(Y,Z)$$

współczynnik korelacji

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

podobnie dla 16 zmiennych (uwzględniając przypadki zerowej kowariancji):

$$Var\left(\sum_{i=1}^{16} \frac{x_i}{\sqrt{i}}\right) = \sum_{i=1}^{16} \frac{x_i}{\sqrt{i}} + 2\sum_{i=1}^{15} Cov\left(\frac{x_i}{\sqrt{i}}, \frac{x_{i+1}}{\sqrt{i+1}}\right) = 16 + 2 \cdot 15 \cdot 0.5 = 31$$

Mamy:

$$V \sim N(0, 31) \rightarrow S = \frac{V-0}{\sqrt{31}}$$

stąd

$$P(V > 6.58) = P\left(\frac{V}{\sqrt{31}} > \frac{6.58}{\sqrt{31}}\right) = 1 - 0.88136 \approx 0.12$$

Zadanie 3.2 Niech X będzie pojedynczą obserwacją z rozkładu o gęstości

$$p_\theta(x) = \begin{cases} \frac{1}{\theta^2}(\theta - |x|) & \text{gdy } x \in [-\theta, \theta] \\ 0 & \text{gdy } x \notin [-\theta, \theta] \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Weryfikujemy hipotezę $H_0 : \theta = 1$ przy alternatywie $H_1 : \theta \neq 1$ za pomocą testu opartego na ilorazie wiarygodności na poziomie istotności 0.20. Moc tego testu przy alternatywie $\theta = 4$ jest równa ?.

Rozwiązanie 3.2 *Dla przypomnienia:*

Błąd I rodzaju: ***odrzucenie H_0, gdy jest prawdziwa***

Błąd II rodzaju: ***nie odrzucenie H_0, gdy jest fałszywa***

Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia H_1 przy jej prawdziwości (odrzucenie H_0 przy jej fałszywości).

Test jest oparty na ilorazie wiarygodności więc potrzebujemy dwóch funkcji wiarygodności (i ich pochodnych do policzenia supremum):

$$L_0(x, \theta = 1) = \frac{1}{1}(1 - |x|) = 1 - |x|$$

$$L_1(x, \theta \neq 1) = \frac{1}{\theta^2}(\theta - |x|)$$

bo test oparty na ilorazie wiarygodności:

$$\frac{\sup_{\theta \in \Theta_1} L(\theta, x)}{\sup_{\theta \in \theta_0} L(\theta, x)}$$

*Test oparty na ilorazie wiarygodności: gdy wynik L_1/L_0 jest większy od pewnej liczby to **odrzucamy** hipotezę zerową. Tym samym, gdy L_0/L_1 mniejsze od tej liczy to również odrzucamy.*

$$\frac{L_1}{L_0} = \frac{\frac{1}{\theta^2}(\theta - |x|)}{1 - |x|}$$

Do testu potrzebujemy supremum

$$\sup_{\theta \in \Theta_0} L_0 = 1 - |x|$$

$$\sup_{\theta \in \Theta_1} L_1 = ?$$

$$\frac{\partial}{\partial \theta}\left(\frac{1}{\theta^2}(\theta - |x|)\right) = -\frac{1}{\theta^2} + 2\frac{|x|}{\theta^3} = 0 \to \theta = 2|x|$$

wstawiamy to do ilorazu wiarygodności i mamy:

$$\frac{L_1}{L_0} = \frac{1}{4(|x| - |x|^2)}$$

H_0 odrzucimy gdy

$$\frac{1}{4(|x| - |x|^2)} > \gamma$$

czyli

$$4(|x| - |x|^2) < \gamma$$

Przy H_0 mamy:

$$P(4|x| - |x|^2 < \gamma|\theta = 1) = 0.2$$

sprawdźmy dla jakich γ powyższe zachodzi

$$|x| - |x|^2 < \gamma/4 \to |x| - |x|^2 - \gamma/4 < 0$$

$$|x_1| = 1/2 - 1/2\sqrt{1 - \gamma}$$

$$|x_2| = 1/2 + 1/2\sqrt{1 - \gamma}$$

to oznacza, że pomiędzy tymi dwoma punktami funkcja kwadratowa jest wieksza od zera co oznacza, że w tym obszarze przyjmujemy H_0. *Test jest dwustronny (mamy wartości bezwzględne), więc o ile dla całego obszaru* $P(x \in D) = 0.8$ *to dla powyższego jest to połowa* $P(x \in D) = 0.4$. *Stąd gęstość:*

$$\frac{1}{\theta^2}(\theta - |x|)$$

przy $H_0 : \theta = 1$ *jest* $1 - |x|$, *w połowie obszaru* $1 - x$

$$\int_{1/2-1/2\sqrt{1-\gamma}}^{1/2+1/2\sqrt{1-\gamma}} (1-x)dx = 0.4 \to \sqrt{1-\gamma} = 0.8$$

Moc testu to odrzucenie H_0 *przy jej fałszywości czyli, 1 - przyjęcie* H_0 *przy* $\theta = 4$ *czyli*

$$\int_{1/2-1/2\sqrt{1-\gamma}}^{1/2+1/2\sqrt{1-\gamma}} \frac{1}{4^2}(4 - |x|)dx = 0.35$$

$$ODP = 1 - 0.35 = 0.65$$

4 Egzamin z 10 marca 2014

Zadanie 4.1 Załóżmy, że zmienne losowe $X_1, \ldots, X_5, X_6, \ldots, X_20$ są niezależne, o jednakowym rozkładzie normalnym o wartości oczekiwanej 1 i wariancji 4, oraz przyjmijmy oznaczenia:

$$S_5 = X_1 + \ldots + X_5$$

$$S_{20} = X_1 + \ldots + X_{20}$$

Wtedy $E(S_5^2|S_{20} = 16)$ jest równa ?.

Rozwiązanie 4.1 *Liczenie bezpośrednie nie jest dobrym pomysłem.*

$$E(S_5^2|S_{20}) =?$$

Tutaj jest ciekawy pomysł, żeby znaleźć rozbicie S_5 na takie zmienne, żeby uzyskać niezależność od S_{20}. Ustalmy, $S_{15} = \sum_{i=6}$

5 Egzamin z 26 maja 2014

Zadanie 5.1 ***Podobne do zadania z 1.10.2012***

Niech $X_1, X_2, \ldots, X_n$, $n > 2$, będą niezależnymi zmiennymi losowymi z rozkładu jednostajnego na przedziale $(0, \theta_1)$, a $Y_1, Y_2, \ldots, Y_n$, niezależnymi zmiennymi losowymi z rozkładu jednostajnego na przedziale $(0, \theta_2)$, gdzie θ_1, θ_2 są niezależnymi prametrami dodatnimi. Wszystkie zmienne są niezależne. Niech $\hat{\theta}_1$, $\hat{\theta}_2$ będą estymatorami największej wiarogodności parametrów θ_1 i θ_2 w oparciu o próby $X_1, X_2, \ldots, X_n$ i $Y_1, Y_2, \ldots, Y_n$ odpowiednio. Weryfikujemy hipotezę $H_0 = \theta_1 = \theta_2$ przy alternatywie $H_1 : \theta_1 = 2\theta_2$ testem o obszarze krytycznym

$$K = \left\{ \frac{\hat{\theta}_1}{\hat{\theta}_2} > c \right\}$$

, gdzie c jest stałą dobraną tak, aby test miał rozmiar 0.1. Najmniejsze n, przy którym moc tego testu jest nie mniejsza niż 0.9 jest równe: ?

Rozwiązanie 5.1 *Należy zauważyć, że estymator największej wiarogodności dla rozkładu jednostajnego na przedziale $[0, \theta]$ to maksimum z zmiennych losowych tj.*

$$\hat{\theta} = \max\{X_1, \ldots, X_n\}$$

Wytłumaczenie:
First note that $f(x|\theta) = \frac{1}{\theta}$, for $0 \leqslant x \leqslant \theta$ and 0 elsewhere. Let $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ be the order statistics. Then it is easy to see that the likelihood function is given by

$$L(\theta|x) = \prod_{i=1}^{n} \frac{1}{\theta} = \theta^{-n} \quad (*)$$

for $0 \leqslant x_{(1)}$ and $\theta \geqslant x_{(n)}$ and 0 elsewhere Now taking the derivative of the log Likelihood wrt θ gives:

$$\frac{d \ln L(\theta|x)}{d\theta} = -\frac{n}{\theta} < 0$$

So we can say that $L(\theta|x) = \theta^{-n}$ is a decreasing function for $\theta \geqslant x_{(n)}$. Using this information and (*) we see that $L(\theta|x)$ is maximized at $\theta = x_{(n)}$. Hence the maximum likelihood estimator for θ is given by

$$\hat{\theta} = x_{(n)}$$

Mamy:

$$K = \left\{ \frac{\max\{X_1, \ldots, X_n\}}{\max\{Y_1, \ldots, Y_n\}} > c \right\}$$

Potrzebujemy wyznaczyć rozkład $\max\{Y_1, \ldots, Y_n\}$. *Wiemy, że*

Rozkład jednostajny ciągły na przedziale $[a, b]$

$$f(x) = \frac{1}{b-a}$$

$$F(x) = \frac{x-a}{b-a}$$
$$E(X) = \frac{a+b}{2}$$
$$VAR(X) = \frac{(b-a)^2}{12}$$

czyli dla przedziału $[0, \theta]$ *jest*

$$F(x) = \frac{x}{\theta}$$

mamy

$$P(\underbrace{\max\{X_1, \ldots, X_n\}}_{Z} < x) = P(X_1 < x) \cdot \ldots \cdot P(X_n < x) = \left(\frac{x}{\theta}\right)^n$$

Żeby obliczyć gęstość liczymy pochodną:

$$f(z) = n\frac{x^{n-1}}{\theta^n}$$

przy H_0 *mamy:*

$$P_0\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = P_0\left(X_{6:6} > cY_{6:6}\right) = P_0\left(Y_{6:6} < \frac{1}{c}X_{6:6}\right) =$$
$$= \int_0^\theta \int_0^{\frac{x}{c}} f_{X_{6:6}}(x) f_{Y_{(6:6)}}(y) dydx =$$
$$= \int_0^\theta \int_0^{\frac{x}{c}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{\theta^n} dydx = n^2 \frac{1}{\theta^{2n}} \int_0^\theta x^{n-1} \frac{(\frac{x}{c})^n}{n} dx =$$
$$= n^2 \frac{1}{\theta^{2n}} \frac{1}{c^n} \frac{\theta^{2n}}{n \cdot 2n} = 0.1$$

Równe 0.1 *bo poziom istotności=rozmiar testu równy* 0.1, *stąd*

$$c^n = \frac{1}{0.2} = 5$$

Przy $H_1 : \theta_1 = 2\theta_2$, *stąd* $\theta_2 = \frac{\theta_1}{2}$ *mamy:*

$$P_0\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = P_0\left(Y_{6:6} < \frac{1}{c}X_{6:6}\right) = \int_0^? \int_0^? n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{\left(\frac{\theta}{2}\right)^n} dydx$$

żeby określić przedziały całkowania dobrze jest zrobić rysunek pomocniczy. Wiemy z odpowiedzi, że n *to co najmniej 4, stąd* c *jest mniejsze od* $5^{1/4} = 1.495$. *Rysujemy* X *i* Y, *zaznaczamy* θ, $\theta/2$ *oraz wykreslamy* x/c , *gdzie* c *mniejsze od* 1.495. *Widać, że w pewnym punkcie dochodzimy do granicy* $\theta/2$ *(bo* y *maksymalnie może być* $\theta/2$, *ten punkt to* $\frac{x}{c} = \frac{\theta}{2} \Rightarrow x = \frac{\theta c}{2}$. *Stąd możemy całkować dla*

x w przedziale od 0 do $\frac{\theta c}{2}$ a dla y od 0 do $\frac{x}{c}$, natomiast kolejna całka jest dla x od $\frac{\theta c}{2}$ do θ i dla y od 0 do $\frac{\theta}{2}$

$$P_1\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = P_1\left(Y_{6:6} < \frac{1}{c}X_{6:6}\right) =$$

$$= \int_0^{\frac{\theta c}{2}} \int_0^{\frac{x}{c}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{\left(\frac{\theta}{2}\right)^n} dydx + \int_{\frac{\theta c}{2}}^{\theta} \int_0^{\frac{\theta}{2}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{\left(\frac{\theta}{2}\right)^n} dydx =$$

całki są dość żmudne

$$= \frac{c^n}{2^n \cdot 2} + 1 - \frac{c^n}{2^n} = \frac{2^{n+1} - c^n}{2^{n+1}} = \frac{2^{n+1} - 5}{2^{n+1}}$$

wynik to jest prawdopodobieństwo odrzucenia H_0 przy jej fałszywości. Moc testu statystycznego to:

Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia H_1 przy jej prawdziwości (odrzucenie H_0 przy jej fałszywości).

Żeby moc była większa od 0.9 to n musi być conajmniej równe 5

Rozwiązanie przepisane przez przypadek drugi raz:

Podobne do zadania z 1.10.2012

***Rozkład jednostajny ciągły na przedziale** $[a, b]$*

$$f(x) = \frac{1}{b-a}$$

$$F(x) = \frac{x-a}{b-a}$$

$$E(X) = \frac{a+b}{2}$$

$$VAR(X) = \frac{(b-a)^2}{12}$$

***Dla rozkładu jednostajnego na przedziale** $[0, \theta]$ **estymator największej wiarygodności to:** $\hat{\theta} = \max\{x_1, \ldots, x_n\}$*

Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia H_1 przy jej prawdziwości (odrzucenie H_0 przy jej fałszywości).

W zadaniu mamy:

$$K = \left\{\frac{\max(X_1, \ldots, X_n)}{\max(Y_1, \ldots, Y_n)} > c\right\} = \left\{\frac{X_{6:6}}{Y_{6:6}} > c\right\}$$

Jaki rozkład ma maksimum?

$$P(\max(X_1, \ldots, X_N) < t) = P(X_1 < t) \cdot \ldots \cdot P(X_n < t) = \left(\frac{t}{\theta}\right)^n$$

gęstość:

$$f_{X_{6:6}}(t) = n\frac{t^{n-1}}{\theta}$$

Teraz wykorzystując informację o rozmiarze testu policzymy c (przy prawdziwości H_0:

$$P_0(K) = 0.1$$

$$P_0\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = P_0\left(Y_{6:6} < \frac{1}{c}X_{6:6}\right) = (*)$$

można mały rysunek pomocniczy i mamy:

$$(*) = \int_0^\theta \int_0^{\frac{x}{c}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{\theta^n} dydx = 0.1$$

stąd

$$c^n = \frac{1}{0.2} = 5$$

Teraz przy prawdziwości H_1 mamy:

$$P_1\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = ?$$

Tu należy wykonać rysunek pomocniczy i mamy

$$P_1\left(\frac{X_{6:6}}{Y_{6:6}} > c\right) = \int_0^{\frac{\theta c}{2}} \int_0^{\frac{x}{c}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{(\theta/2)^n} dydx + \int_{\frac{\theta c}{2}}^{\theta} \int_0^{\frac{\theta}{2}} n\frac{x^{n-1}}{\theta^n} n\frac{y^{n-1}}{(\theta/2)^n} dydx =$$

$$= \frac{c^n}{2^n \cdot 2} + \frac{2^n - c^n}{2^n} = \frac{2^{n+1} - c^n}{2^{n+1}}$$

$$\frac{2^{n+1} - \overbrace{c^n}^{=5}}{2^{n+1}} \geqslant 0.9 \to 2^n \geqslant \frac{2.5}{0.1} = 25$$

$$n \geqslant \frac{\ln(25)}{\ln(2)} = 4.64$$

stąd najmniejsze n równe 5.

Zadanie 5.2 Rozważamy model regresji liniowej postaci $Y_i = a + bx_i + \epsilon_i$, $i = 1, 2, 3, 4, 5$, gdzie a, b są nieznanymi parametrami rzeczywistymi $x_1 = x_2 = 1$, $x_3 = 3$, $x_4 = x_5 = 5$, a ϵ_i są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym o wartości oczekiwanej 0 i wariancji 9.
Hipotezę $H_0 : b = 0$, przy alternatywie $H_1 : b \neq 0$ weryfikujemy testem o obszarze krytycznym postaci $\{|\hat{b} > c|\}$, gdzie $\hat{b}$ jest estymatorem największej wiarogodności parametrów b, a stała c dobrana jest tak, aby test miał rozmiar 0.05. Stała c jest równa: ?

Rozwiązanie 5.2 *PODOBNE 28.05.2012*

Dla przypomnienia w rozkładzie normalnym:

$$aX + b \sim N(a\mu + b, (a\sigma)^2)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

W rozkładzie normalnym:

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

W regresji liniowej funkcja wiarygodności jest postaci:

$$L = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-a-bx_i)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5 e^{\sum_{i=1}^{5} -\frac{(y_i-a-bx_i)^2}{2\sigma^2}}$$

Bierzemy logarytm z funkcji wiarygodności:

$$\ln L = -5\ln\sqrt{2\pi} - 5\ln\sigma +$$

$$-\frac{(y_1-a-b)^2 + (y_2-a-b)^2 + (y_3-a-3b)^2 + (y_4-a-5b)^2 + (y_5-a-5b)^2}{2\sigma^2}$$

Pochodna po b:

$$\frac{\partial \ln L}{\partial b} =$$

$$-\frac{-2(y_1-a-b) - 2(y_2-a-b) - 2\cdot 3(y_3-a-b) - 2\cdot 5(y_4-a-5b) + 2\cdot 5(y_5-a-5b)}{2\sigma^2} = 0$$

stąd

$$61b = y_1 + y_2 + 3y_3 + 5y_4 + 5y_5 - 15a$$

Analogiczna pochodna po a daje:

$$y_1 + y_2 + y_3 + y_4 + y_5 - 15b = 5a$$

Podstawiamy do równiania z b i mamy:

$$16b = -2y_1 - 2y_2 + 2y_4 + 2y_5 \to b = \frac{-y_1 - y_2 + y_4 + y_5}{8}$$

stąd uwzględniając rozkład ϵ_i mamy:

$$b \sim N\left(4 \cdot \frac{9}{8^2}\right) = N\left(0, \frac{9\cdot 4}{64}\right)$$

Test ma rozmiar 0.05, ale to test dwustronny (rozkład normalny) więc

$$P(b > c) = 0.025$$

$$P(b < c) = 0.975$$

Musimy unormować b:

$$\frac{b}{\frac{\sqrt{9 \cdot 2}}{\sqrt{64}}} < \frac{c}{\frac{\sqrt{9 \cdot 2}}{\sqrt{64}}}$$

Kwantyl stopnia 0.975 wynosi 1.96 *stąd*

$$\frac{c}{\frac{\sqrt{9 \cdot 2}}{\sqrt{64}}} = 1.96 \rightarrow c = 1.47$$

Rozwiązanie przez przypadek przepisane drugi raz:

Dla przypomnienia

$$aX + b \sim N(a\mu + b, (a\sigma)^2)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Rozkład normalny

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Rozwiązując zadanie z regresją: reszty mają rozkład normalny. Stąd funkcja wiarygodności:

$$L = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - a - bx_i)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5 e^{\sum_{i=1}^{5}} - \frac{(y_i - a - bx_i)^2}{2\sigma^2}$$

bierzemy logarytm z funkcji wiarygodności:

$$\ln L = -5 \ln \sqrt{2\pi} - 5 \ln \sigma - \sum_{i=1}^{5} \frac{(y_i - a - bx_i)^2}{2\sigma^2} =$$

$$= -5 \ln \sqrt{2\pi} - 5 \ln \sigma -$$

$$+ \frac{(y_1 - a - b)^2 + (y_2 - a - b)^2 + (y_3 - a - 3b)^2 + (y_4 - a - 5b)^2 + (y_5 - a - 5b)^2}{2\sigma^2}$$

Pochodna po b:

$$\frac{\partial L}{\partial b} =$$

$$-\frac{-2(y_1 - a - b) - 2(y_2 - a - b) - 6(y_3 - a - 3b) + -10(y_4 - a - 5b) - 10(y_5 - a - 5b)}{2\sigma^2} = 0$$

stąd

$$61b = y_1 + y_2 + 3y_3 + 5y_4 + 5y_5 - 15a$$

liczymy analogicznie pochodną po a i mamy:

$$y_1 + y_2 + y_3 + y_4 + y_5 - 15b = 5a$$

podstawiamy i mamy:

$$\hat{b} = \frac{-y_1 - y_2 + y_4 + y_5}{8}$$

biorąc pod uwagę $H_0 : b = 0$ oraz rozkład ϵ_i,

$$\hat{b} \sim N(0, 4 \cdot \frac{9}{8^2}) = N\left(0, \frac{9 \cdot 4}{64}\right)$$

Mamy test dwustronny, więc:

$$P(\hat{b} > c) = 0.025$$

$$P(b < c) = 0.975$$

żeby użyć tablic rozkładu normalnego musimy znormalizować $\hat{b}$

$$\frac{\hat{b}}{\frac{\sqrt{92}}{\sqrt{64}}} < \frac{c}{\frac{\sqrt{92}}{\sqrt{64}}} \to \frac{c}{\frac{\sqrt{92}}{\sqrt{64}}} = 1.96 \to c = 1.47$$

Zadanie 5.3 Załóżmy, że niezależne zmienne losowe $X_1, X_2, \ldots, X_n$, $n > 1$, mają rozkłady Pareto, zmienna X_i rozkład o gęstosci

$$f_i(x) \begin{cases} \frac{2i}{(1+x)^{2i+1}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}$$

$i = 1, 2, \ldots, n$.
Wtedy prawdopodobieństwo $P(X_1 = \min\{X_1, \ldots, X_n\})$ jest równe?

Rozwiązanie 5.3 *Analogiczne do zadania 4 z 04.10.2010.*

Dla przypomnienienia:

$$P(X \leqslant Y) = \int_0^\infty \int_0^y f(x)f(y)dxdy$$

Mamy:

$$P(X_1 = \min\{X_1, \ldots, X_n\}) = P(X_1 \leqslant X_1, X_1 \leqslant X_2, \ldots, X_1 \leqslant X_n) =$$

$$P(X_1 \leqslant X_2, \ldots X_1 \leqslant X_n) = P(X_1 \leqslant \min\{X_2, \ldots, X_n\})$$

Jaki rozkład ma minimum?

$$P(\underbrace{\min\{X_2, \ldots, X_n\}}_{=Z} \leqslant t) = 1 - P(\min\{X_2, \ldots, X_n\} > t) = 1 - \prod_{i=2}^{n} P(X_i > t) = (*)$$

liczymy dystrybuantę

$$P(X_i < t) = \int_0^x \frac{2i}{(1+x)^{2i+1}} dx = 1 - (1+x)^{-2i}$$

$$(*) = 1 - \prod_{i=2}^{n}(1+x)^{-2i} = 1 - \frac{1}{(1+x)^{2\sum_{i=2}^{n} i}} = 1 - \frac{1}{(1+x)^{n^2+n-2}}$$

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

Obliczyliśmy dystrybuantę dla minimum. Gęstość to:

$$f(z) = (n^2+n-2)\frac{1}{(1+x)^{n^2+n-1}}$$

$$P(X_1 \leqslant Z) = \int_0^{\infty}\int_0^{z} \frac{2}{(1+x)^3}\cdot(n^2+n-2)\cdot\frac{1}{(1+z)^{n^2+n-1}}dxdz = \ldots =$$

$$= \frac{2}{n+n^2}$$

Zadanie 5.4 Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi takimi, że $EX_i = im$ i $VarX_i = 2i^2m^2$ dla $i = 1, 2, \ldots, n$, gdzie $m > 0$ jest nieznanym parametrem. W klasie estymatorów parametru m postaci $\hat{m} = \sum_{i=1}^{n} c_iX_i$ wyznaczono estymator o najmniejszym błędzie średniokwadratowym. Błąd średniokwadratowy tego estymatora jest równy:?

Rozwiązanie 5.4 *Analogiczne do zadania 9 z 3.10.2011*

Błąd średniokwadratowy estymatora

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}((\hat{\theta}-\theta)^2)$$

W zadaniu błąd średniokwadratowy to

$$E((m-\hat{m})^2)$$

odwróciliśmy kolejność...bo tak. Nie wpływa to istotnie na dalsze obliczenia.

1. Obliczmy błąd średniokwadratowy:

$$E((m-\hat{m})^2) =?$$

Pamiętamy, że $VarX = EX^2 - (EX)^2$ *stąd* $EX^2 = VarX + (EX)^2$

$$E((m-\hat{m})^2) = (E(m-\hat{m}))^2 + Var(m+\hat{m}) = (E(m-\hat{m}))^2 + Var(\hat{m}) = (*)$$

m znika wariancji bo to wartość liczbowa $m > 0$

$$(*) = (E(\hat{m}))^2 - 2E(\hat{m})\cdot m + m^2 + Var(\hat{m}) = (*)$$

Liczymy:

$$E(\hat{m}) = \sum_{i=1}^{n} c_iE(X_i) = \sum_{i=1}^{n} c_i\cdot i\cdot m$$

$$Var(\hat{m}) = Var\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 Var(X_i) = \sum_{i=1}^{n} 2 \cdot c_i^2 \cdot i^2 \cdot m^2$$

$$(*) = (\sum_{i=1}^{n} c_i \cdot i \cdot m)^2 - 2m \cdot \sum_{i=1}^{n} c_i \cdot i \cdot m + m^2 + \sum_{i=1}^{n} 2 \cdot c_i^2 \cdot i^2 \cdot m^2$$

W zadaniu mamy całą klasę estymatorów, które wyznaczane są przez c_i. Niech błąd będzie funkcją wszystkich c_i

$$f(c_1, \ldots, c_n) = \left(\sum_{i=1}^{n} c_i \cdot i \cdot m\right)^2 - 2m \cdot \sum_{i=1}^{n} c_i \cdot i \cdot m + m^2 + \sum_{i=1}^{n} 2 \cdot c_i^2 \cdot i^2 \cdot m^2$$

2. Minimalizacja polega na policzeniu pochodnych względem c_i (pojedynczych c_i). Mamy

$$\frac{\partial f}{\partial c_i} = 2\sum_{i=1}^{n} (c_i \cdot i \cdot m)^2 \cdot i \cdot m - 2 \cdot m^2 \cdot i + 2 \cdot 2 \cdot c_i \cdot i^2 m^2 = 0$$

dochodzimy do

$$\sum c_i \cdot i - 1 + 2c_i \cdot i = 0$$

niech teraz

$$f(i) = \sum c_i \cdot i - 1 + 2c_i \cdot i$$

3. Teraz niech

$$\sum_{i=1}^{n} f(i) = 0$$

stąd

$$\sum f(i) = n \sum c_i \cdot i - n + 2 \sum c_i \cdot i = 0$$

$$(n+2) \sum c_i \cdot i = n$$

$$\sum c_i \cdot i = \frac{n}{n+2}$$

4. Wstawiamy wynik do $f(i) = 0$

$$\frac{n}{n+2} = 1 - 2c_i \cdot i \to c_i \cdot i = \frac{1}{n+2}$$

5. Liczymy błąd średniokwadratowy

$$E((m - \hat{m})^2) = m^2 \left(\sum c_i \cdot i\right)^2 - 2m \sum c_i \cdot i + m^2 + m^2 2 \sum (c_i \cdot i)^2 =$$

podstawiamy wyliczone wcześniej wartości

$$= \frac{2m^2}{n+2}$$

Zadanie 5.5 Mamy dwie urny: I i II. Na początku doświadczenia w każdej z urn znajdują się 2 kule białe i 2 czarne. Losujemy po jednej kuli z każdej urny - po czym kulę wylosowaną z urny I wrzucamy do urny II, a tę wylosowaną z urny II wrzucamy do unry I. Czynność powtarzamy wielkrotnie. Niech X_n oznacza zmienną losową równą liczbie kul białych w urnie I po n-tym powtórzeniu czynności. Wtedy granica:

$$\lim_{n\to\infty} E(X_n X_{n+1})$$

jest równa ?.

Rozwiązanie 5.5 *Identyczne jak 4 z 25.03.2013. Wykorzystamy łańcuchy Markova i macierz prawdopodobieństw przejść pomiędzy stanami. Macierz p-p przejść przedstawia się następująco: mamy wobec tego prawdopodobieństwa ze po wielu*

	0	1	2	3	4
0	0	1	0	0	0
1	1/16	6/16	9/16	0	0
2	0	4/16	8/16	4/16	0
3	0	0	9/16	6/16	1/16
4	0	0	0	1	0

próbach będziemy przebywali w danym stanie (tip: zapisujemy w odwrotnej kolejności niż sumowanie się do jedynki (czyli kolumnami w naszym przypadku))

$$P_0 = 1/16P_1$$

$$P_1 = P_0 + 6/16P_1 + 4/16P_2$$

$$P_2 = 9/16P_1 + 8/16P_2 + 9/16P_3$$

$$P_3 = 9/16P_2 + 6/16P_3 + 1/16P_4$$

$$P_4 = 1/16P_3$$

$$P_0 + P_1 + P_2 + P_3 + P_4 = 1$$

stąd $P_0 = 1/70$, $P_1 = 16/70$, $P_2 = 36/70$, $P_3 = 16/70$, $P_4 = 1/70$ *macierz wartości* $X_n X_{n+1}$

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	6	8
3	0	3	6	9	12
4	0	4	8	12	16

$$\lim_{n\to\infty} E(X_n \cdot X_{n+1}) = 1 \cdot P_1 \cdot 6/16 + 2 \cdot P_1 \cdot 9/16 +$$

$$+2 \cdot P_2 \cdot 4/16 + 4 \cdot P_2 \cdot 9/16 + 6 \cdot P_2 \cdot 4/16+$$
$$6 \cdot P_3 \cdot 9/16 + 9 \cdot P_3 \cdot 6/16 + 12 \cdot P_3 \cdot 1/16+$$
$$+12 \cdot P_4 \cdot 1 = 30/7$$

Zadanie 5.6 O zmiennych losowych $X_1, X_2, \ldots, X_{12}$ o tej samej wartości oczekiwanej równej 2 oraz tej samej wariancji równej 1, zakładamy, iż:

$$COV(X_i, X_j) = \frac{1}{3}$$

dla $i \neq j$.

Zmienne losowe $\epsilon_1, \ldots \epsilon_{12}$ są nawzajem niezależne oraz niezależne od zmiennych losowych $X_1, \ldots, X_{12}$ i mają rozkłady prawdopodobieństwa postaci:

$$P(\epsilon_i = 1) = P(\epsilon_i = 1/2) = P(\epsilon_i = 0) = \frac{1}{3}$$

Wariancja zmiennej losowej $S = \sum_{i=1}^{12} \epsilon_i \cdot X_i$ jest równa ?.

Rozwiązanie 5.6 *Z definicji kowariancji:*

$$Cov(X, Y) = E(X \cdot Y) - EX \cdot EY$$

$$Var(S) = Var\left(\sum_{i=1}^{12} \epsilon_i \cdot X_i\right) = E\left(\left(\sum_{i=1}^{12} \epsilon_i \cdot X_i\right)^2\right) - \left(E\left(\sum_{i=1}^{12} \epsilon_i \cdot X_i\right)\right)^2$$

$$E\left(\left(\sum_{i=1}^{12} \epsilon_i \cdot X_i\right)^2\right) =$$

rozpisujemy powyższe i wymnażamy

$$= 12E(\epsilon_1^2) \cdot E(X_i^2) + 12(12-1)E(\epsilon_i)E(\epsilon_j) + E(X_i X_j) = (*)$$

ostatni czynnik zawiera zmienne skorelowane

$$E(\epsilon_i) = 1/2$$
$$E(\epsilon_i^2) = 1^2 \cdot 1/3 + (1/2)^2 \cdot 1/3 = 5/12$$
$$E(X_i^2) = Var(X_i) + (EX_i)^2 = 5$$
$$E(\sum \epsilon_i X_i) = 12E(\epsilon_i X_i) = 12$$
$$Cov(X_i, X_j) = 1/3 = E(X_i X_j) - EX_i \cdot EX_j \rightarrow E(X_i X_j) = 13/3$$
$$(*) = 12 \cdot 5/12 \cdot 5 + 12 \cdot 11 \cdot 1/2 \cdot 1/2 \cdot 13/3 = 168$$
$$Var(S) = 168 - 144 = 24$$

Zadanie 5.7 Załózmy, że $X_1, \ldots, X_5$ jest próbka z rozkładu normalnego $N(\mu, \sigma^2)$ o nieznanej wartości oczekiwanej i nieznanej wariancji, zaś X_6 jest zmienną losową z tego samego rozkładu, niezależną od próbki. Intepretujemy zmienną X_6 jako kolejną obserwację, ktora pojawi się w przyszłości, ale obecnie jest nieznana.
Zbuduj 'przedział ufności'

$$[L, U] = [L(X_1, \ldots, X_5), U(X_1, \ldots, X_5)]$$

oparty na próbce $X_1, \ldots X_5$ taki, że

$$Pr(L(X_1, \ldots, X_5) \leqslant X_6 \leqslant U(X_1, \ldots, X_5)) = 0.95$$

przy tym żądamy, żeby przedział był symetryczny tzn. $\frac{1}{2}(L+U) = \bar{X}$. Uzywamy tutaj oznaczeń:

$$\bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$$

$$S^2 = \frac{1}{4}\sum_{i=1}^{5}(X_i - \bar{X})^2$$

Rozwiązanie 5.7 *Takie samo jak zadanie 6 z 17.05.2003*

Patrzymy na odpowiedzi i widzimy, że mamy zbudować symetryczny przedział ufności postaci:

$$P(|X_6 - \bar{X}| < cS)$$

Wiemy, że

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{5})$$

$$X_6 - \bar{X} \sim N\left(0, \frac{6}{5}\sigma^2\right)$$

Dla przypomnienia parę własności:
1. Dla rozkładu normalnego S^2 jest niezależne od $\bar{X}$.

2. Jeżeli $Z \sim N(0,1)$ oraz $Y \sim \chi^2(n)$ to

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}} \sim T(n)$$

3.Gdy $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$, $\bar{X} = \frac{1}{n}\sum X_i$

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim T(n-1)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \to \frac{1}{\sigma^2}\sum(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

W zadaniu widać, że obie strony nalezy podzielić przez S a wtedy jestesmy blisko rozkładu t-studenta. Zrobimy następujące obliczenia korzystając z powyższych własności:

$$\frac{4 \cdot S^2}{\sigma^2} \sim \chi^2 4$$

zapiszmy:

$$P\left(\frac{\frac{|X_6-\bar{X}|}{\sqrt{\frac{6}{5}\sigma^2}}}{\sqrt{\frac{\frac{4S^2}{\sigma^2}}{4}}} < \frac{cS}{\sqrt{\frac{\frac{4S^2}{\sigma^2}}{4}} \cdot \sqrt{\frac{6}{5}\sigma^2}}\right) =$$

$$= P(\underbrace{|Z|}_{\sim T(4)} < c\frac{\sqrt{5}}{\sqrt{6}}) = 0.95$$

oczywiście moduł nie ma rozkładu T studenta, ale w tablicach mamy wartości krytyczne dla poziomu α i testu dwustronnego, więc

$$c\frac{\sqrt{5}}{\sqrt{6}} = 2.7764 \rightarrow c = 3.04$$

i to należało znaleźć.

Zadanie 5.8 Niech $N, X_1, X_2, \ldots$ będą niezależnymi zmiennymi losowymi przy $\Lambda = \lambda$. Zmienne X_i, $i = 1, 2, \ldots$, mają warunkowe rozklady wykladnicze o wartości oczekiwanej $\frac{1}{\lambda}$ przy $\Lambda = \lambda$. Warunkowy rozkład zmiennej losowej N przy danym $\Lambda = \lambda$ jest rozkladem Poissona o wartości oczekiwanej λ. Rozkład brzegowy zmiennej Λ jest rozkładem gamma o gęstości:

$$f(\lambda) = \begin{cases} \frac{8}{3}\lambda^3 e^{-2\lambda} & \text{gdy } \lambda > 0, \\ 0 & \text{gdy } \lambda \leqslant 0. \end{cases}$$

niech

$$S = \begin{cases} \sum_{i=1}^{N} X_i & \text{gdy } N > 0, \\ 0 & \text{gdy } N = 0. \end{cases}$$

i

$$T = \begin{cases} \sum_{i=1}^{N} Y_i & \text{gdy } N > 0, \\ 0 & \text{gdy } N = 0. \end{cases}$$

gdzie $Y_i = \min\{X_i, 2\}$

Oblicz współczynnik kowariancji $Cov(S, T)$.

Rozwiązanie 5.8 *Analogiczne do zadania 3 z 30.09.2013. Zadnie jest średnio rozwiązywalne na egzaminie (za dużo całek)*

Z definicji

$$Cov(S, T) = E(S \cdot T) - E(S) \cdot E(T) \cdot E(S)$$

$$E(S) = E(E(S|N))$$
$$E(S|N) = N \cdot E(X_1) = \frac{N}{\lambda}$$
$$E(N/\lambda) = \frac{1}{\lambda} \cdot \lambda = 1$$

oczywiście powyższe również jest warunkowe przy znanym λ. Dalej mamy:

$$E(T) = E\left(\sum_{i=1}^{N} \min(X_i, 2)\right) = N \cdot E(\min(X_i, 2))$$

$$E(\min(X_i, 2)) = \int_0^2 xf(x)dx + 2 \cdot Pr(X_i > 2) = (*)$$

X_i ma rozkład wykładniczy $f(x) = \lambda e^{-\lambda x}$, dystrybuanta: $F(x) = 1 - e^{-\lambda x}$

$$\int_0^2 x\lambda e^{-\lambda x} dx = -2e^{-2\lambda} + \frac{1 - e^{-\lambda 2}}{\lambda} = \frac{1 - e^{-2\lambda}(1 + 2\lambda)}{\lambda}$$

$$Pr(X_i > 2) = e^{-2\lambda}$$

wracając

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

Względem N

$$E\left(N \cdot \frac{1 - e^{-2\lambda}}{\lambda}\right) = 1 - e^{-2\lambda}$$

Względem λ

$$E(1 - e^{-2\lambda}) = 1 - \int_0^\infty e^{-2\lambda} \frac{8}{3} \lambda^3 e^{-2\lambda} d\lambda =$$
$$= 1 - \frac{8}{3} \int_0^\infty \lambda^3 e^{-4\lambda} d\lambda = (*)$$

Z rozkładu Gamma:

Rozklad Gamma

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0$$
$$E(X) = \frac{\alpha}{\beta}$$
$$Var(X) = \frac{\alpha}{\beta^2}$$

Widzimy, ze $\alpha = 4$, $\beta = 4$ brakuje składnika

$$\frac{\beta^\alpha}{\Gamma(\alpha)} = \frac{4^4}{\Gamma(4)} = \frac{256}{3!} = \frac{256}{6} = \frac{128}{3}$$

mamy:

$$(*) = 1 - \frac{8}{3}\frac{3}{128}\underbrace{\int_0^\infty \frac{128}{3}\lambda^3 e^{-4\lambda} d\lambda}_{=1} = \frac{15}{16}$$

Teraz trzeba rozważyć

$$E(S \cdot T) = NE(X_i \cdot \min(X_i, 2)) + N(N-1)E(X_i \cdot \min(X_j, 2)) =$$

$$= N \cdot E(X_i \min(X_i, 2) + N(N-1) \cdot \frac{1}{\lambda} \cdot E(\min(X_j, 2)) = (*)$$

kluczowe jest:

$$E(X_i \min(X_i, 2)) = \int_0^2 x^2 f(x)dx + \int_2^\infty 2xf(x)dx =$$

$$= \frac{2 - 2e^{-2\lambda}(2\lambda^2 + 2\lambda + 1)}{\lambda^2} + \frac{e^{-2\lambda}(4\lambda + 2)}{\lambda}$$

wracając

$$(*) = N \cdot \frac{2 - 2\lambda e^{-2\lambda} - 2e^{-2\lambda}}{\lambda^2} + N(N-1) \cdot \frac{1}{\lambda} \cdot \frac{1 - e^{-2\lambda}}{\lambda} = \Delta$$

$$E(\Delta) = \frac{2}{\lambda} - 2e^{-2\lambda} - \frac{2}{\lambda}e^{-2\lambda} + 1 - e^{-2\lambda} = \Delta_2$$

$$E(\Delta_2) = \frac{95}{48}$$

Wynik:

$$\frac{95}{48} - \frac{15}{16} = \frac{25}{24}$$

Zadanie 5.9 Urna zawiera 5 kul o numerach 0,1,2,3,4. Z urny ciągniemy kulę, zapisujemy numer i kulę wrzucamy z powrotem do urny. Czynność tę powtarzamy, aż kule z numerami 1,2,3 zostaną wyciągnięte co najmniej raz. Oblicz prawdopodobieństwo zdarzenia, że czynność powtórzymy 5 razy.

Rozwiązanie 5.9 *Mamy 5 losowań:*

$$[], [], [], [], []$$

Liczba wszystkich możliwych wyników tego losowania to

$$5^5 = 3125.$$

W 5 tym losowaniu musi zostać wyciągnięta kula o numerze 1 lub 2 lub 3 żeby zakończyć proces. Dla ustalenia uwagi przyjmijmy, że to 1. (potem odpowiednie kombinacje pomnożymy razy 3 możliwości, widać od razu, że odpowiedzi A i E odpadają bo licznik nie dzieli się przez 3). Do zadania podejdziemy w następujący sposób: gdy wiemy, jak kończy się układ losowań to pozostałych razem jest

$4^4 = 256$ *możliwości. Wśród tych możliwości znajdziemy zdarzenia niesprzyjające.*

1. Losowania w których nie ma 1,2,3:

$$[0/4], [0/4], [0/4], [0/4]$$

razem $2^4 = 16$ *sposobów.*

2. Jeżeli wylosowaliśmy dwójkę, ale nie wylosowaliśmy trójki:

$$[2], [0/4], [0/4], [0/4] \rightarrow \binom{4}{1} 2^3 = 32 \textit{ sposobów}$$

$$[2], [2], [0/4], [0/4] \rightarrow \binom{4}{2} 2^2 = 24 \textit{ sposobów}$$

$$[2], [2], [2], [0/4] \rightarrow \binom{4}{3} 2^1 = 8 \textit{ sposobów}$$

$$[2], [2], [2], [2] \rightarrow 1 \textit{ sposób}$$

razem: 65 sposób 3. Analogicznie jeżeli wylosowaliśmy trójkę, ale nie wylosowaliśmy dwójki mamy 65 sposobów.

Wobec tego niesprzyjających zdarzeń jest $65 + 65 + 16 = 146$, *odejmujemy od wszystkich możliwości w tym podzbiorze zdarzeń i mamy:* $256 - 146 = 110$ *sprzyjających sposobów. Wobec tego jeżeli uwzględnimy mnożenie przez 3 z początku mamy* 330 *sprzyjajacych zdarzeń. Stąd prawdopodobieństwo jest równe:*

$$\frac{330}{5^5}$$

Zadanie 5.10 Niech X_1, X_2, X_3, X_4 ... zadanie analogiczne do zadania 1 z 15.03.2010.

6 Egzamin z 29 września 2014

Zadanie 6.1 Identyczne jak w 13.10.2001
Niech $N_0 = N - N_1$ oraz

$$N_1 = \begin{cases} \sum_{i=1}^{N} X_i & \text{gdy } N > 0, \\ 0 & \text{gdy } N = 0. \end{cases}$$

gdzie N jest zmienną losową o rozkładzie ujemnym dwumianowym:

$$P(N = n) = (n+1)\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^n$$

gdzie $n = 0, 1, 2, \ldots$ zaś $X_1, X_2, \ldots, X_n, \ldots$ są zmiennymi losowymi niezależnymi od N i od siebie nawzajem. Zakładamy, że każda ze zmiennych X_i ma rozkład Bernoulliego: $P(X_i = 1) = \frac{3}{4}$ i $P(X_i = 0) = \frac{1}{4}$. Wtedy:

$$E\left(\frac{N_1}{N_0 + 1}\right) = ?$$

Rozwiązanie 6.1 *Mamy:* $E\left(\frac{N_1}{N_0+1}\right)$, N_1 *oraz* N_0 *zależne jest od wartości* N. *Dla przypomnienia:*

Law of total expectation:

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X \mid A_i)\, \mathrm{P}(A_i).$$

Z własności warunkowej wartości oczekiwanej $E(X) = E(E(X|Y))$.*Przykład:* $E(T) = E(E(T|N)) = \sum_{n=0}^{\infty} E(T|N = n)P(N = n)$

The following formulation of the law of iterated expectations plays an important role in many economic and finance models:

$$\mathrm{E}(X \mid I_1) = \mathrm{E}(\mathrm{E}(X \mid I_2) \mid I_1),$$

więc w pierwszym kroku:

$$E\left(\frac{N_1}{N_0+1}\right) = \sum_{k=1}^{N} E\left(\frac{N_1}{N_0+1} \mid N = n\right) \cdot P(N = n) =$$

zaczynamy od $n = 1$ *ze względu na fakt, że gdy* $N = 0$ *to mamy 0.*

$$= \sum_{n=1}^{N} E\left(\frac{N_1}{N_0+1} \mid N = n\right) \cdot (n+1)\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^n = (*)$$

Dla przypomnienia:

Rozkład dwumianowy

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

gdzie $k = 0, 1, \ldots, n$

$$E(X) = np$$

$$Var(X) = np(1-p)$$

przy ustalonym N *jest*

$$E\left(\frac{N_1}{N_0+1} \mid N = k\right) = E\left(\frac{\sum_{i=1}^n X_i}{n - \sum_{i=1}^n X_i}\right)$$

Suma zmiennych losowych o rozkładzie zero-jedynkowym na rozkład dwumianowym. Równocześnie korzystamy z prawa leniwego statystyka i mamy:

$$(*) = \left(\frac{1}{3}\right)^2 \sum_{n=1}^{\infty} \left((n+1) \left(\frac{2}{3}\right)^n \sum_{k=0}^{n} \frac{k}{n-k+1} \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k} \right) =$$

zauważmy, że dla $k = 0$ *jest zero mamy:*

$$= \left(\frac{1}{3}\right)^k \sum_{n=1}^{\infty} \left((n+1) \left(\frac{2}{3}\right)^n \sum_{k=1}^{n} \frac{k}{n-k+1} \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k} \right) =$$

Teraz będziemy starali się przesunąć wszystko tak żeby otrzymać rozkad dwumianowy i go zwinąć do jedynki (suma po funkcji rozkładu daje jeden)

$$= \left(\frac{1}{3}\right)^2 \sum_{n=1}^{\infty} \left((n+1) \left(\frac{2}{3}\right)^n \sum_{k=0}^{n} \underbrace{\frac{k}{n-k+1} \frac{n!}{k!(n-k)!}}_{= \frac{n!}{(k-1)!(n-k+1)!)}} \left(\frac{3}{4}\right)^k \underbrace{\left(\frac{1}{4}\right)^{n-k}}_{(\frac{3}{4})^{k-1} (\frac{1}{4})^{n-k+1} \cdot \frac{3}{4} \cdot 4} \right) =$$

$$= \left(\frac{1}{3}\right)^2 \sum_{n=1}^{\infty} \left((n+1) \left(\frac{2}{3}\right)^n \cdot 3 \cdot \underbrace{\sum_{k=0}^{n-1} \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k}}_{1-(3/4)^k} \right) =$$

$$= 3 \cdot \frac{1}{3^2} \left(\sum_{n=1}^{\infty} n(2/3)^n = \sum_{n=1}^{\infty} n(1/2)^n + \sum_{n=1}^{\infty} (2/3)^n - \sum_{n=1}^{\infty} (1/2)^n \right) = (*)$$

Dla przypomnienia wzory:

$$I_0(v) = \sum_{n=0}^{\infty} v^n = \frac{1}{1-v}$$

$$I_1(v) = \sum_{n=1}^{\infty} nv^n = \frac{v}{(1-v)^2}$$

$$I_2(v) = \sum_{n=1}^{\infty} n^2v^n = \frac{v(v+1}{(1-v)^3}$$

po podstawieniach:

$$(*) = \frac{5}{3}$$

Zadanie 6.2 Podobne do 3 z 5.06.2006 inne eg 49/10

Niech $X_1, \ldots, X_{10}$ będą niezależnymi zmiennymi losowymi z rozkładu prawdopodobieństwa o gęstości

$$f_{\theta_\alpha}(x) = \begin{cases} \frac{1}{\alpha}e^{-(x-\theta)/\alpha} & \text{dla } x \geqslant> \theta, \\ 0 & \text{dla } x < \theta. \end{cases}$$

Wyznaczono estymator największej wiarygodności $(\hat{\theta}, \hat{\alpha})$ parametrów (θ, α) w sytuacji, gdy oba parametry są nieznane $\alpha > 0$. A następnie zbudowano przedział ufności dla parametru α, w oparciu o estymator $\hat{\alpha}$, postaci $[c\hat{\alpha}, d\hat{\alpha}]$, taki, że:

$$P_{\theta,\alpha}(\alpha < c\hat{\alpha}) = P_{\theta,\alpha}(\alpha > d\hat{\alpha}) = 0.05.$$

Liczba c jest równa: ?.

Rozwiązanie 6.2 *Podany w zadaniu rozkład to przesunięty rozkład wykładniczy (szukać shifted exponential; two parameter exponential). Dla przypomnienia:*

Rozkład wykładniczy

$$f(x) = \beta e^{-\beta x}$$

$$E(X) = \frac{1}{\beta}$$

$$VAR(X) = \frac{1}{\beta^2}$$

Własności:

(1) gdy $X \sim Exp(\beta)$ to $kX \sim Exp(\beta/k)$.
(2) gdy $X \sim Exp(\beta)$ i $Y \sim Exp(\beta)$ to $X + Y \sim G(1+1, \beta)$

W zadaniu należy wyznaczyć estymator α. Estymator największej wiarygodności uzyskujemy przy pomocy funkcji wiarygodności:

$$L = \prod_{i=1}^{10} \frac{1}{\alpha} e^{\frac{-(x_i-\theta)}{\alpha}} = \prod_{i=1}^{10} = \left(\frac{1}{\alpha}\right)^{10} e^{-\sum_{i=1}^{10} \frac{x_i-\theta}{\alpha}}$$

logarytmujemy:

$$\ln L = \underbrace{10 \ln \frac{1}{\alpha}}_{-10 \ln \alpha} - \sum_{i=1}^{10} \frac{x_i - \theta}{\alpha}$$

liczymy pochodną po α

$$\frac{\partial L}{\partial \alpha} = -10 \frac{1}{\alpha} + \alpha^{-2} \sum_{i=1}^{10} (x_i - \theta) = 0$$

stąd

$$\alpha = \bar{X} - \theta$$

więc potrzebny nam jest również estymator θ, $\frac{\partial L}{\partial \theta}$ *nic nam nie da, ale wiemy, że* $x_i \geqslant \theta$ *oraz znamy postać funkcji wiarygodności. Przy ustalonym* α *maksymalizacja funkcji wiarygodności następuje gdy czynnik* $\frac{\theta}{\alpha}$ *jest możliwie największy. Skoro* θ *mniejsza lub równa od każdego* x_i *to estymatorem największej wiarygodności będzie* $\min(x_1, \ldots, x_{10})$*. Stąd:*

$$\hat{\alpha} = \bar{X} - \min(x_1, \ldots, x_{10})$$

Przydatne własności

Przy użyciu powyższych, gdy $X_i \sim Exp(\beta)$

$$\sum_{i=1}^{n} X_i \sim Gamma(n, \beta)$$

$$\beta \bar{X}_n \sim Gamma(n, n)$$

gdy $X \sim Gamma(a, b) \rightarrow 2 \cdot b \cdot X \sim \chi^2(n)$, $n = 2 \cdot a$ *bo gdy* $\alpha = \frac{n}{2}$ *oraz* $\beta = \frac{1}{2}$ *to mamy rozkład ch-kwadrat z n stopniami swobody*
stąd

$$2n\beta \bar{X}_n \sim \chi^2(2n)$$

Wracając do zadania. Mamy

$$P(\alpha < c\hat{\alpha}) = P\left(\frac{1}{c} < \frac{1}{\alpha}\left(\bar{X} - X_{1:10}\right)\right) = 0.05$$

musimy wyznaczyć rozkład. Należy zwrócić uwagę, że zmienne powyżej pochodzą z przesuniętego rozkładu wykładniczego. Żeby operować na zmiennych ze zwykłego rozkładu wykładniczego potrzebujemy $X' = X - \theta$ *bo korzystając z transformacji zmiennych, gdy* $Y = aX + b$ *to* $g(y) = f\left(\frac{y-b}{a}\right)\frac{1}{|a|}$*. Mamy:*

$$\bar{X} - X_{1:10} = \bar{x} - \min(X_1, \ldots, X_{10}) = \bar{X} - \theta - (\min(X_1, \ldots, X_{10}) - \theta) =$$

$$\frac{1}{10} \sum_{i=1}^{10} (X_i - \theta) - \min(X_1 - \theta, \ldots, X_{10} - \theta) =$$

$$= \frac{1}{10}\sum_{i=1}^{10} X_i' - \min\left(X_1', \ldots, x_{10}'\right)$$

Korzystając z własności rozkładów:

$$\frac{1}{10}X_i' \sim Exp(10 \cdot \beta)$$

$$\min\left(X_1', \ldots, X_{10}'\right) \sim Exp(10 \cdot \beta)$$

Dalej:

$$\frac{1}{10}\sum_{i=1}^{10} X_i' - \min\left(X_1', \ldots X_{10}'\right) \sim Gamma(9, 10\beta)$$

Z treści $\beta = \frac{1}{\alpha}$, *stąd*

$$\frac{1}{\alpha}\left(\frac{1}{10}\sum_{i=1}^{10} X_i' - \min\left(X_1', \ldots X_{10}'\right)\right) \sim Gamma(9, 10)$$

Mamy:

$$P\left(\underbrace{\frac{1}{\alpha}(\bar{X} - \min\left(X_1, \ldots, X_{10}\right))}_{\sim Gamma(9,10)} > \frac{1}{c}\right) = 0.05$$

Jako, że dostaniemy tablice chi-kwadrat a nie gamma, robimy transformację zgodnie z powyższymi wzorami.$\alpha = \frac{n}{2}$, $\beta = \frac{1}{2}$*:*

$$P\left(\underbrace{\frac{20}{\alpha}(\bar{X} - \min\left(X_1, \ldots, X_{10}\right))}_{\sim \chi^2(18)} > \frac{20}{c}\right) = 0.05$$

Odczytujemy z tablic wartość krytyczną dla 0.05 i mamy:

$$\frac{20}{c} = 28.8693 \rightarrow c \approx 0.69$$

Zadanie 6.3 Niech (X, Y) będzie dwuwymiarową zmienną losową o funkcji gęstości:

$$f(x, y) = \begin{cases} e^{-x} & \text{gdy } x > 0, y \in (0; 1), \\ 0 & \text{w przeciwnym wypadku }. \end{cases}$$

Niech $Z = X + 2Y$. Wtedy $E(X|Z = 3)$ jest równa?

Rozwiązanie 6.3 *Warunkowa wartość oczekiwana:*

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y)dx = \frac{1}{f_Y(y)}\int_{-\infty}^{\infty} x f(x, y) = \frac{\int_{-\infty}^{\infty} x f(x, y)dx}{\int_{-\infty}^{\infty} f(x, y)dx}$$

Z treści zadania:

$$E(X|X+2Y=3) = E(X|Y = \frac{3}{2} - \frac{1}{2}X)$$

Tu potrzebny jest rysunek pomocniczy, który pozwala stwierdzić, że $x \in (1,3)$ *(bo z treści* $y \in (0,1)$ *). Stąd*

$$E(X|X+2Y=3) = \frac{\int_1^3 xe^{-x}dx}{\int_1^3 e^{-x}dx} = \frac{2-4e^{-2}}{1-e^{-2}}$$

Zadanie 6.4 Podobne do z 10 z 11.10.2004

Niech X_1, X_2, X_3, X_4 będzie próbą z rozkładu jednostajnego o gęstości

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{gdy } x \in (0,\theta) \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

Zakładamy, że nieznany parametr θ jest zmienną losową o rozkładzie z funkcją gęstości daną wzorem

$$\pi(\theta) = \begin{cases} \frac{4}{3}\theta^4 e^{-2\theta} & \text{gdy } \theta > 0 \\ 0 & \text{gdy} \theta \leqslant 0. \end{cases}$$

Hipotezę $H_0 : \theta \leqslant 3$ przy alternatywnie $H_1 : \theta > 3$ odrzucamy dla tych wartości (x_1, x_2, x_3, x_4), dla których prawdopodobieństwo a posteriori zbioru $\theta : \theta > 3$ jest większe niż $\frac{1}{2}$. Rozmiar tego testu jest równy: ?.

Rozwiązanie 6.4 *Kilka uwag poczatkowych: rozmiar testu = poziom istotności,* $x \in (0,\theta)$ *oznacza, że* $\theta \geqslant \max(x_1,\ldots,x_4)$*. Bayesowski przedział ufności* $[\theta_1,\theta_2]$ *na poziomie*

$$\alpha = \int_{\theta_1}^{\theta_2} f(\theta|x)dx$$

gdzie $f(\theta|x)$ *- gęstość a posteriori dla* θ*. Rozwiązanie:*

Wyznaczamy gęstość rozkładu a posteriori parametru θ*, czyli:*

$$f_{\theta|X}(\theta|x) = \frac{f(x_1,\ldots,x_4;\theta)}{f(x_1,\ldots,x_4)} = ?$$

Żeby policzyć coś takiego najpierw liczymy odwrotny warunek:

$$f(x|\theta) = \frac{f(x,\theta)}{f(\theta)} \to f(x,\theta) = f(x|\theta)f(\theta) =$$

$$= f(x_1,\ldots,x_4|\theta)f(\theta) = \left(\prod_{i=1}^4 f(x_1|\theta)\right)\underbrace{f(\theta)}_{=\pi(\theta)} =$$

$$= \frac{1}{\theta^4} \cdot \chi_{X_{4:4}}(0,\theta) \cdot \frac{4}{3}\theta^4 e^{-2\theta} = \frac{4}{3}e^{-2\theta}\chi_{X_{4:4}}(0,\theta)$$

gdzie $\chi_{X_{4:4}}(0,\theta)$ to funkcja indykatorowa pokazująca dla jakiego θ pochodna jest niezerowa tj. θ musi być większa maksimum z próby. Wiemy, że $f(x_1,\ldots,x_4)$ to rozkład brzegowy gęstości łącznej $f(x_1,\ldots,x_4,\theta)$, stąd:

$$f(x_1,\ldots,x_4) = \int_0^\infty \frac{4}{3}e^{-2\theta}\chi_{X_{4:4}}(0,\theta)d\theta =$$

biorąc pod uwagę, że $\theta \geqslant \max(x_1,\ldots,x_4)$ to mamy:

$$\int_{X_{4:4}}^\infty \frac{4}{3}e^{-2\theta}d\theta = \frac{4}{3}\frac{e^{-2\theta}}{-2}|^\infty_{\underbrace{x_m}_{=x_{max}}} = \frac{2}{3}e^{-2x_m}$$

$$f(\theta|x) = \frac{\frac{4}{3}e^{-2\theta}\chi_{X_{4:4}}(0,\theta)}{\frac{2}{3}e^{-2x_m}} = 2e^{-2\theta}e^{2x_m}\chi_{X_{4:4}}(0,\theta) =$$

$$= 2e^{-2\theta}e^{2x_m} \quad dla\ \theta > x_m$$

Z treści, prawdopodobieństwo a posteriori (po tym gdy wiem, co się wylosowało) ma być większe od $1/2$ czyli:

$$P(\theta > 3|x_1,\ldots,x_4) > \frac{1}{2}$$

mamy

$$P(\theta > 3|x_1,\ldots,x_4) = \int_3^\infty 2e^{-2\theta}e^{2x_m}d\theta = e^{2x_m}e^{-6}$$

czyli mamy warunek

$$\underbrace{e^{2x_m}e^{-6} > \frac{1}{2}}_{obszar\ krytyczny} \to x_m > 3 - \frac{\ln 2}{2}$$

czyli

$$K = \{x_m > 3 - \frac{\ln 2}{2}\}$$

Dla przypomnienia poziom istotności (rozmiar testu): $P_\theta(K) \leqslant \alpha$

$$P\left(x_m > 3 - \frac{\ln 2}{2}\right) =?$$

$$P\left(\max(x_1,\ldots,x_4)\langle x\right) = P(x_1 < x) \cdot \ldots \cdot P(x_4 < x) = (\frac{x}{\theta})^4$$

czyli

$$f_{x_m} = 4\left(\frac{x}{\theta}\right)^3$$

$$P(x_m > 3 - \frac{\ln 2}{2}) = 1 - P(x_m < 3 - \left(\frac{\ln 2}{2}) = 1 - \frac{3 - \frac{\ln 2}{2}}{\theta}\right)^4 = (*)$$

gdy $\theta = 3$ mamy

$$(*) \approx 0.388$$

Zadanie 6.5 Podobne do z 6 z 30.11.2009
Załóżmy, że $X_1, \ldots, X_n, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym ciąglym rozkładzie prawdopodobieństwa, mającym momenty rzędu 1,2 i 3. Znamy

$$\mu = E(X_i) \qquad \sigma^2 = Var(X_i).$$

Niech $f(x)$ oznacza gęstość rozkładu pojedynczej zmiennej X_i. Wiemy, że rozkład jest symetryczny w tym sensie, że $f(\mu + x) = f(\mu - x)$ dla każdego x. Niech

$$S_N = \begin{cases} X_1 + \ldots + X_n & \text{gdy } N = n > 0 \\ 0 & \text{gdy} N = 0. \end{cases}$$

gdzie N jest zmienną o rozkładzie Poissona o wartości oczekiwanej 1. Trzeci moment $E(S_N^3)$ jest równy=?

Rozwiązanie 6.5 ***Momenty i współczynniki***

Moment zwykły rzędu k:

$$m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$$

Moment centralny rzędu k

$$\mu_k = E((X - (EX))^k)$$

Wspołczynnik asymetrii:

$$A = \frac{M_3}{\sigma^3}$$

gdzie M_3 - 3 moment centralny, s - odchylenie standardowe

Kurtoza:

$$K = \frac{M_4}{\sigma^4} - 3$$

Własności:

(1) jeżeli X i Y niezależne to 3 moment centralny ich sumy równa się sumie trzecich momentów centralnych. Czyli $E((X+Y-E(X+Y))^3) = E((X-EX)^3) + E((Y-EY)^3)$

Rozwiązanie:

Dla X_i mamy

$$\mu_3 = E((X - EX)^3) = E(X^3 - 3X^2EX + 3X(EX)^2 - (EX)^3) =$$

$$= E(X^3) - 3E(X^2) \cdot EX + 3(EX)^3 - (EX)^3 = m_3 - 3m_2 \cdot \mu + 2\mu^3$$

Z powyższego:

$$m_3 = \mu_3 + 3m_2\mu - 2\mu^3 \quad (*)$$

Wzór wyżej patrz Otto s 46. Korzystając z wzoru na współczynnik asymetrii:

$$A = \gamma_x = \frac{\mu_{3,x}}{\mu_{2,x}^{\frac{3}{2}}}$$

gdzie indeks x *oznacza wartość wskaźnika dla zmiennej* x*. Dla sumy*

$$\gamma_{S_N} = \frac{\mu_{3,S_N}}{\mu_{2,S_N}^{3/2}} = \frac{\sum_{i=1}^N \overbrace{\mu_{3,x}}^{=\gamma_i \cdot \sigma^3}}{\left(\sum_{i=1}^N \mu_{2,x}\right)^{3/2}} = \frac{\sum_{i=1}^N \gamma_i \cdot \sigma_i^3}{\left(\sum_{i=1}^N \sigma_i^2\right)^{3/2}} = 0$$

z własności (1) oraz z faktu, że $Var(X+Y) = VarX + VarY + 2Cov(X,Y)$*, a gdy zmienne niezależne to* $Cov(X,Y) = 0$ *oraz z treści zadania (rozkład jest symetryczny, moment centralny rzędu 3 równy zero). Dalej*

$$\mu_{3,S_N} = \gamma_{S_N} \cdot (\sigma_{S_N})^3 = 0$$

Korzystając z $(*)$ *mamy*

$$E(S_n^3) = \mu_{3,S_N} + 3\mu_{S_N} m_{2,S_N} - 2\mu_{S_N}^2 =$$

$$= \gamma_{S_N}\sigma_{S_N}^3 + 3\mu_{S_N} m_{2,S_N} - 2\mu_{S_N}^3 = (*)$$

pamiętamy, że $m_{2,S_N} = Var(S_N) + (E(S_N))^2$ *innymi symbolami:* $m_{2,S_N} = \sigma_{S_N}^2 + (\mu_{S_N})^2$

$$(*) = \gamma_{S_N}\sigma_{S_N}^2 + \mu_{S_N}^3(\sigma_{S_N}^2 + \mu_{S_N}^2) - 2\mu_{S_N}^3 =$$

$$= \gamma_{S_N}\sigma_{S_N}^3 + \mu_{S_N}^3 + 3\mu_{S_N} \cdot \sigma_{S_N}^2 = (*)$$

Mamy $\mu_{S_N} = N \cdot \mu$, $\sigma_{S_N}^2 = N\sigma^2$

$$(*) = \underbrace{\gamma_{S_N}\sigma_{S_N}}_{=0} + N^3\mu^3 + 3N^2\mu\sigma^2 = N^2\mu(N\mu^2 + 3\sigma^2)$$

Potrzebujemy rozkład Poissona

Rozkład Poissona

$$f(k,\lambda) = \frac{\lambda^k e^-\lambda}{k!}$$

gdzie $k = 0, 1, 2, \ldots$

$$E(X) = \lambda$$

$$Var(x) = \lambda$$

W zadaniu $\lambda = 1$*. Biorąc pod uwagę, że* N *ma rozkład Poissona mamy:*

$$E(N^3\mu^3 + 3N^2\mu\sigma^2) = \mu^3 E(N^3) + 3\mu\sigma^2 E(N^2) = (\Delta)$$

Wiemy, że $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n}$

$$E(N^3) = \sum_{n=1}^{\infty} n^3 \frac{\lambda^n e^{-\lambda}}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} n^2 \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} (n+1)^2 \frac{\lambda^n}{n!} =$$

$$= \lambda e^{-\lambda} \sum_{n=0}^{\infty} (n^2+2n+1) \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} + 2\lambda e^{-\lambda} \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} + \lambda e^{-\lambda} e^{\lambda} = (*^2)$$

gdzie $\sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!} = \lambda$ *z wartości oczekiwanej. Dalej*

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

Wracając do gwiazdki 2:

$$(*^2) = \lambda^3 + 3\lambda^2 + \lambda \underbrace{=}_{\lambda=1} = 5$$

Stąd odpowiedź

$$(\Delta) = 5\mu^3 + 6\mu\sigma^2$$

Zadanie 6.6 Pan A przeznaczył 6 zł na pewną grę. W pojedycznej kolejce gry pan A wygrywa 1 zł z prawdopodobieństwem 1/3 lub przegrywa 1 zł z prawdopodobieństwem 2/3. Pan A kończy grę, gdy wszystko przegra lub gdy będzie miał 9 zł.
Prawdopodobieństwo, że pan A wszystko przegra jest równe:?

Rozwiązanie 6.6 *Analogiczne jak 10 z 30.11.2009*

Niech p_i *oznacza prawdopodobieństwo, że startując ze stanu i zł dojdziemy do stanu 0 zł (przegramy):*

1. $p_8 = \frac{2}{3} p_7$
2. $p_7 = \frac{1}{3} p_8 + \frac{2}{3} p_6$
3. $p_6 = \frac{1}{3} p_7 + \frac{2}{3} p_5$
4. $p_5 = \frac{1}{3} p_6 + \frac{2}{3} p_4$
5. $p_4 = \frac{1}{3} p_5 + \frac{2}{3} p_3$
6. $p_3 = \frac{1}{3} p_4 + \frac{2}{3} p_2$
7. $p_2 = \frac{1}{3} p_3 + \frac{2}{3} p_1$
8. $p_1 = \frac{1}{3} p_2 + \frac{2}{3}$

Musimy wyznaczyć p_6. Należy wyznaczyć p_7 wtedy:

$$p_7 = \frac{1}{3} \cdot \frac{2}{3} \cdot p_7 + \frac{2}{3} \cdot p_6 \to p_7 = \frac{6}{7} p_6$$

Z drugiej strony

$$p_6 = \frac{1}{3} p_7 + \frac{2}{3} p_5$$

należy cofnąc się z prawodpodobieństwami tak, żeby wyznaczyć p_5, po tym kroku mamy:

$$p_5 = \frac{31}{63} p_6 + \frac{32}{63}$$

Podstawiając mamy:

$$p_6 = \frac{1}{3} \cdot \frac{6}{7} p_6 + \frac{2}{3} \left(\frac{31}{63} p_6 + \frac{32}{63} \right)$$

stąd

$$p_6 = \frac{64}{73} \approx 0.88$$

Zadanie 6.7 Rozważmy następujący schemat urnowy:

W każdej z 10 urn znajdują się 2 kule, oznaczone liczbami:

- W urnie 1 znajdują się 2 kule oznaczone liczbą 1,
- w urnie 2 znajdują się 2 kule oznaczone liczbą 2,
- ...
- w urnie 10 znajdują się 2 kule oznaczone liczbą 10.

Losujemy kulę z urny 1 i przekładamy ją do urny 2. Następnie (po wymieszaniu kul) losujemy kulę z urny 2 i przekładamy do urny 3, itd., kulę wylosowaną z urny 9 przekładamy do urny 10, wreszcie losujemy kulę z urny 10. Jakie jest prawdopodobieństwo, że ta ostatnia wylosowana kula ma numer większy niż 6?

Rozwiązanie 6.7 *Zadanie można rozpocząć od 7 urny, która będzie miała w środku 1 kulę większą od 6 i dwie mniejsze. Dalej to typowe drzewko:*

1. *z p-p 1/3 kula mniejsza ląduje w kolejnych urnach z p-p 2/3 kula większa i wtedy koniec (bo dalej już się losuje tylko z większych*
2. *dalej drzewko...*

sumujemy ścieżki które doprowadzją nast do wyniku:

$$\frac{2}{3} + \frac{1}{3}\frac{2}{3} + \left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} + \left(\frac{1}{3}\right)^3 \cdot \frac{2}{3} = \frac{80}{81}$$

Zadanie 6.8 Niech $Z_1, Z_2, \ldots, Z_n$ bedą niezależnymi zmiennymi losowymi z rozkładu jednostajnego na przedziale $(0, 1)$. Niech $Z_{1:n} = \min\{Z_1, Z_2, \ldots, Z_n\}$. Wtedy $E(Z_1 + Z_2 + \ldots + Z_n | Z_{1:n} = 0.5)$ jest równa ?.

Rozwiązanie 6.8 *Skoro* $\min\{Z_1, \ldots, Z_n\} = 0.5$ *to znaczy, że któraś ze zmiennych* $Z_i = 0.5$

$$E(Z_1 + Z_2 + \ldots + Z_n | Z_{1:n} = 0.5) = (n-1)E(Z_i | Z_i \geqslant 0.5) + 0.5$$

Z wzoru na warunkową wartość oczekiwaną:

$$E(Z_i | Z_i \geqslant 0.5) = \frac{\int_{0.5}^{1} x \cdot 1dx}{\int_{0.5}^{1} 1 \cdot dx} = \frac{3}{4}$$

czyli mamy:

$$(n-1)\frac{3}{4} + \frac{2}{4} = \frac{3n-1}{4}$$

7 Egzamin z 8 grudnia 2014

Zadanie 7.1 Identyczne jak w 13.10.2001

Mamy 5 niezależnych próbek z tego samego rozkładu normalnego $N(\mu, \sigma^2)$ z nieznaną wartością oczekiwaną μ i znaną wariancją σ^2, przy tym każda z tych próbek ma tą samą liczebność n. Dla każdej z 5 próbek oddzielnie wyznaczamy w standardowy sposób przedział ufności. Niech

$$\left[\bar{X}_i - 0.8416\frac{\sigma}{\sqrt{n}}, \bar{X}_i + 0.8416\frac{\sigma}{\sqrt{n}}\right]$$

będzie przedziałem obliczonym na podstawie i-tej próbki.
Następnie, przedział ufności oparty na wszystkich 5n obserwacjach wyznaczamy w sposób niestandardowy: za środek przedziału wybieramy medianę

$$m = med(\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4, \bar{X}_5)$$

Oblicz

$$c = Pr\left(m - 0.8416\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant m + 0.8416\frac{\sigma}{\sqrt{n}}\right)$$

Rozwiązanie 7.1 ***Przedział ufności dla wartości oczekiwanej***

Jeśli X jest próbą prostą z rozkładu normalnego z nieznanym parametrem μ i znanym σ , to przedział ufności dla μ na poziomie $1-\alpha$ ma postać:

$$\bar{X}_n \pm u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

czyli

$$P\left(\overline{X}_n - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

gdzie u_p oznacza kwantyl rzędu p w rozkładzie normalnym $N(0,1)$.

Z tablic wiemy, że $u_{0.8} = 0.8416$ stąd $1 - \frac{\alpha}{2} = 0.8 \Rightarrow \alpha = 0.4$. Co daje poziom ufności $1-\alpha = 0.6$ Przedział jest symetryczny więc

$$Pr\left(\mu < \bar{X}_i - 0.8416\frac{\sigma}{\sqrt{n}}\right) = Pr\left(\mu > \bar{X}_i + 0.8416\frac{\sigma}{\sqrt{n}}\right) = 1 - 0.8 = 0.2$$

Statystyki pozycyjne

Niech $X_{1:n} \leqslant X_{2:n} \leqslant \ldots \leqslant X_{n:n}$ będzie ciągiem zmiennych losowych powstałych z $X_1, \ldots, X_n$, po ich uporządkowaniu w ciąg niemalejący. Zmienną $X_{k:n}$, $k = 1, \ldots, n$, nazywamy k-tą statystyką pozycyjną. W szczególności

$$X_{1:n} = \min(X_1, \ldots, X_n)$$

$$X_{n:n} = \max(X_1, \ldots, X_n)$$

k-ta statystyka pozycyjna ma dystrybuantę:

$$F_{k:n}(x) = P(X_{k:n} \leqslant x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1 - F(x))^{n-i}$$

W naszym przypadku mediana jest 3 statystyką pozycyjną. Wobec tego mamy:

$$Pr\left(m - 0.8416\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant m + 0.8416\frac{\sigma}{\sqrt{n}}\right) =$$

$$Pr\left(m \leqslant \mu + 0.8416\frac{\sigma}{\sqrt{n}} \wedge m \geqslant \mu - 0.8416\frac{\sigma}{\sqrt{n}}\right) =$$

$$Pr\left(m \in (\mu - 0.8416\frac{\sigma}{\sqrt{n}}, \mu + 0.8416\frac{\sigma}{\sqrt{n}})\right) =$$

$$Pr\left(X_{3:5} \leqslant \mu + 0.8416\frac{\sigma}{\sqrt{n}}\right) - Pr\left(X_{3:5} \leqslant \mu - 0.8416\frac{\sigma}{\sqrt{n}}\right)$$

gdzie $X_{3:5}$ to 3cia statystyka pozycjna dla $N(\mu, \frac{\sigma^2}{n})$ (bo oparta na średnich). Dla przypomnienia:

$$X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, (a\sigma)^2)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Korzystając z początku zadania:

$$Pr\left(\bar{X}_i \leqslant \mu + 0.8416\frac{\sigma}{\sqrt{n}}\right) = 0.8$$

oraz

$$Pr\left(\bar{X}_i \leqslant \mu - 0.8416\frac{\sigma}{\sqrt{n}}\right) = 0.2$$

stąd

$$F_{3:5}\left(\mu + 0.8416\frac{\sigma}{\sqrt{n}}\right) - F_{3:5}\left(\mu - 0.8416\frac{\sigma}{\sqrt{n}}\right) =$$

$$= \sum_{i=3}^{5} \binom{5}{i} (0.8)^i (0.2)^{5-i} - \sum_{i=3}^{5} \binom{5}{i} (0.2)^i (0.8)^{5-i} = 0.88416$$

Zadanie 7.2 Niech $X_1, X_2, \ldots, X_n, \ldots$ będą niezależnymi zmiennymi losowymi z rozkładu o gęstości

$$f(x) = \begin{cases} 3x^2 & \text{gdy } (x) \in (0,1) \\ 0 & \text{gdy } x \neq (0,1) \end{cases}$$

Niech $T_n = \prod_{i=1}^{n} X_i^{\frac{3}{n}}$. Które z poniższych stwierdzeń jest prawdziwe? ...

(prawdziwa jest odpowiedź $\lim_{n\to\infty} P(|T_n - e_1|\sqrt{n} > 2e^{-1}) = 0.046$

Rozwiązanie 7.2 *Widzimy, że T_n to iloczyn, więc rozpatrzmy zmienną $\ln T_n$*

$$\ln T_n = \sum_{i=1}^{n} \frac{3}{n} \ln X_i$$

Jaki rozkład ma $\ln X_i$? Robimy transformację zmiennych losowych

$$Y = \ln X \to X = e^Y$$

$$g(y) = 3e^{2y} \cdot e^y = 3e^{3y}$$

widać, że to NIE jest rozkład wykładniczy dlatego musimy rozpatrzyć

$$\ln T_n = -\frac{3}{n} \sum \underbrace{-\ln X_i}_{Y}$$

wtedy wg wzoru $g(y) = f(h(y))|h'(y)|$

$$Y = -\ln X \to X = e^{-Y}$$

$$g(y) = 3e^{-2y} \cdot e^{-y} = 3e^{-3y}$$

czyli rozkład wykładniczy z $\beta = 3$ i $E(Y) = \frac{1}{3}$, $Var(Y) = \frac{1}{9}$ Korzystając z centralnego twierdzenia granicznego wiemy, że przy $n \to \infty$ suma n zmiennych losowych dązy do rozkładu normalnego, czyli

$$\sum_{i=1}^{n} -\ln X_i \sim N\left(n \cdot \frac{1}{3}, n \cdot \frac{1}{9}\right)$$

Z własności rozkładu normalnego:

$$\ln T_n = -\frac{3}{n} \sum_{i=1}^{n} -\ln X_i \sim N\left(-\frac{3}{n} \cdot n \cdot \frac{1}{3}, \left(\frac{3^2}{n^2}\right) \cdot n\frac{1}{9}\right) \sim N\left(-1, \frac{1}{n}\right)$$

Odpowiedzi są postaci:

$$P((T_n - a)\sqrt{n} > b) = x$$

oczywiście trzeba uważać na wartość bezwzględną (żeby ją zdjąć trzeba podzielić x przez 2). Wewnątrz prawdopodbieństwa:

$$\ln T_n > \ln\left(\frac{b + a\sqrt{n}}{\sqrt{n}}\right)$$

musimy wystadaryzować $\ln T_n$, mamy:

$$\underbrace{(\ln T_n + 1)\sqrt{n}}_{\sim N(0,1)} > \ln\left(a + \frac{b}{\sqrt{n}}\right)^{\sqrt{n}} + 1 \cdot \sqrt{n}$$

Wiemy, że

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$$

korzystamy z tej własności i mamy

$$(\ln T_n + 1)\sqrt{n} > \ln a^{\sqrt{n}} + \frac{b}{a} + \sqrt{n}$$

$$(\ln T_n + 1)\sqrt{n} > \sqrt{n}\cdot \ln a + \frac{b}{a} + \sqrt{n}$$

Po prawej stronie widać, że żeby znikło $\sqrt{n}$ a musi być e^{-1}

Przy odpowiedzi D (dystrybuanta równa 1-0.046/2=0.977) kwantyl wynosi 2, w związku z tym b musi być równe $2e^{-1}$

Zadanie 7.3 Załóżmy, że dysponujemy pojedynczą obserwacją X z rozkładu $Laplace'a$ $L(\mu,\lambda)$ o gęstości

$$f_{\mu,\lambda}(x) = \frac{1}{2\lambda}\exp\left(-\frac{x-\mu}{\lambda}\right)$$

gdzie $\mu \in R, \lambda > 0$ są nieznanymi parametrami.
Rozważmy zadanie testowania hipotezy:

$$H_0 : \mu = 1 \text{ i } \lambda = 2$$

$$H_1 : \mu = 2 \text{ i } \lambda = 1$$

Najmocniejszy test na pewnym poziomie istotności jest postaci:

Odrzuć H_0, gdy $x \in \left(\frac{5}{3}, b\right)$. Moc tego testu jest równa: ?

Rozwiązanie 7.3 *Moc testu: prawdopodobieństwo odrzucenia hipotezy zerowej przy jej fałszywości.*

Zbudujemy test najmocniejszy poprzez zbadanie ilorazu wiarygodności, gdy:

$$\frac{L_1}{L_0} > c$$

to odrzucamy H_0. Stąd

$$\frac{L_1}{L_0} = \frac{\frac{1}{2\cdot 1}\exp\left(-\frac{|x-2|}{1}\right)}{\frac{1}{2\cdot 2}\exp\left(-\frac{|x-1|}{2}\right)} > c$$

$$e^{-\frac{|x-2|}{1}+\frac{|x-1|}{2}} > c$$

$$|x-1| - 2|x-2| > c$$

Rozwiążmy to względem x 1. dla $x \geqslant 2$

$$x - 1 - 2x + 4 > c$$

$$x < 3 - c$$

2. dla $x < 2$ *i* $x \geqslant 2$

$$x - 1 + 2(x - 2) > c$$

$$x \geqslant \frac{5}{3} + \frac{1}{3}c$$

3. dla $x < 1$

$$-x + 1 + 2x - 4 > c$$

$$x > 3 + c$$

Z punktu 2 i tresci zadania $x \in (5/3, b)$ *z punktu drugiego widzimy, że* $x \geqslant 5/3 + 1/3c$ *stąd prosty wniosek, że* $c = 0$, *z powyższych równań wynika również, że dla* $x \geqslant 2$ *mamy* $x < 3 - 0$, *czyli* b *musi równac się* 3. *Moc testu to prawdopodobieństwo odrzucenia* H_0 *gdy jest fałszywa, czyli:*

$$\int_{5/3}^{3} f_{\mu_1,\lambda_1}(x)dx = \int_{5/3}^{3} \frac{1}{2}e^{-\frac{|x-2|}{1}}dx$$

widać, że trzeba rozdzielić na dwie całki:

$$\int_{5/3}^{2} \frac{1}{2}e^{x-2}dx + \int_{2}^{3} e^{-(x-2)}dx = 0.458$$

Zadanie 7.4 Niech $X_1, X_2, \ldots, X_6, Y_{1,2}, \ldots, Y_{10}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie Pareto o gęstości:

$$p_\theta(x) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}} & \text{gdy } x > 0 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Osobno, na podstawie prób losowych $X_1, X_2, \ldots, X_6$ i $Y_1, Y_2, \ldots, Y_{10}$, wyznaczono estymary największej wiarogodności T_X i T_Y parametru θ.
Prawdopodobieństwo $P(T_X < T_Y)$ jest równe: ?.

Rozwiązanie 7.4 *Policzymy estymatory największej wiarygodności. Najpierw* T_X.

$$L = \prod_{i=1}^{6} \frac{\theta}{(1 + x_i)^{\theta+1}} = \frac{\theta^6}{\prod_{i=1}^{6}(1 + x_i)^{\theta+1}}$$

$$\ln L = 6\ln(\theta) - \sum_{i=1}^{6}(\theta + 1)\ln(1 + x_i)$$

Maksymalizujemy L

$$\frac{\partial L}{\partial \theta} = \frac{6}{\theta} - \sum_{i=1}^{6} \ln(1+x_i) = 0$$

$$\hat{\theta} = \frac{6}{\sum_{i=1}^{6} \ln(1+x_i)} \rightarrow T_X = \frac{6}{\sum_{i=1}^{6} \ln(1+x_i)}$$

natomiast T_Y

$$T_Y = \frac{10}{\sum_{i=1}^{10} \ln(1+y_i)}$$

Czyli mamy obliczyć p-p

$$P\left(\frac{6}{\sum_{i=1}^{6} \ln(1+x_i)} < \frac{10}{\sum_{i=1}^{10} \ln(1+y_i)}\right)$$

jaki rozkład ma $\ln(1+x_i)$*? Zastosujemy wzory na trasnformację zmiennych*

$$Y = \ln(1+X) \rightarrow X = e^Y - 1$$

$$g(y) = f(h(y))|h'(y)|$$

$$g(y) = \frac{\theta}{(1+e^y-1)^{\theta+1}} \cdot e^y = \theta e^{-\theta y}$$

czyli to ma rozkład wykładniczy. Suma zmiennych będzie miała rozklad Gamma

$$\sum_{i=1}^{6} \ln(1+x_i) \sim G(6, \theta)$$

$$\sum_{i=1}^{10} \ln(1+y_i) \sim G(10, \theta)$$

żeby pozbyć się theta zamienimy na rozkłady chi-kwadrat (rozkład $G\left(\frac{n}{2}, \frac{1}{2}\right)$ *to* $\chi^2(n)$*):*

$$Z_1 = 2\theta \cdot \sum_{i=1}^{6} \ln(1+x_i) \sim G\left(6, \frac{\theta}{2/\theta}\right) \sim \chi^2(12)$$

$$Z_2 = 2\theta \cdot \sum_{i=1}^{10} \ln(1+y_i) \sim G\left(10, \frac{\theta}{2/\theta}\right) \sim \chi^2(20)$$

widać, że mamy

$$P\left(\frac{6}{Z_1} < \frac{10}{Z_2}\right) = (*)$$

tutaj widziemy, że trzeba będzie skorzystać z rozkładu F-Snedecora

Rozkład F Snedecora

Jeżeli X i Y są niezależne oraz $X \sim \chi^2(n_1)$ i $Y \sim \chi^2(n_2)$, to:

$$\frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \sim F(n_1, n_2)$$

$$(*) = P\left(\frac{12}{Z_1} < \frac{20}{Z_2}\right) = P\left(\frac{\frac{Z_2}{20}}{\frac{Z_1}{12}} < 1\right)$$

, mamy rozkład F Snedecora $F(20,12)$ z tablic (albo Excela =F.DIST(1,20,12,1)):

$$F(20,12) = 0.482684448$$

Zadanie 7.5 Niech (X, Y) będzie dwuwymiarową zmienną losowa o funkcji gęstości:

$$f(x,y) = \begin{cases} \frac{2}{\pi} & \text{gdy } x > 0 \text{ i } x^2 + y^2 = 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Niech $Z = \frac{Y}{X}$ i $V = X^2 + Y^2$. Wtedy łączny rozkład zmiennych Z, V jest taki, że:

(a) $EZ = 0$

(b) funkcja gęstości rozkładu brzegowego zmiennej Z wyraża się wzorem $g(z) = \frac{2}{\pi(1+z^2)}$ dla $z \in (0, +\infty)$

(c) mediana rozkładu brzegowego zmiennej Z jest równa $\frac{\sqrt{2}}{2}$

(d) zmienne Z i V są zależne

(e) kwantyl rzędu 0.25 rozkładu brzegowego zmiennej Z jest równy -1

Rozwiązanie 7.5 *Trzeba tu zauważyć, ze mamy do czynienia z rozkładem na okręgu czyli tak naprawdę mamy:*

$$x = r\cos\alpha$$

$$y = r\sin\alpha$$

wtedy rozkład $g(r,\alpha)$ wg wzoru

$$g(y) = f(h(y))|h'(y)|$$

$$g(r,\alpha) = \frac{2}{\pi} \cdot \left|\left| \begin{matrix} \frac{dx}{dr} & \frac{dx}{d\alpha} \\ \frac{dy}{dr} & \frac{dy}{d\alpha} \end{matrix} \right|\right| = \frac{2}{\pi} \cdot r$$

Wobec powyższego nasze zmienne to:

$$Z = \operatorname{tg}\alpha$$

$$V = r^2$$

Wiemy, że r i α są niezalezne stąd odpowiedź (d) odpada.

W odpowiedziach widać, że potrzebujemy rozkładu brzegowego zmiennej Z. Więc najpierw policzymy rozkład brzegowy od α z $g(r,\alpha)$, czyli całkę z rozkładu łącznego po r

$$f(\alpha) = \int_0^1 g(r,\alpha)dr = \int_0^1 \frac{2}{\pi} \cdot rdr = \frac{1}{\pi}$$

teraz określimy gęstość $Z = \operatorname{tg}\alpha$ ponownie używając wzoru na transformację zmiennych

$$Z = \operatorname{tg}\alpha \to \alpha = arctg(Z)$$

$$g(z) = \frac{1}{\pi} \cdot arctg'(z) = \frac{1}{\pi(1+z^2)}$$

Stąd widać, że odpada odpowiedź (b).

Gdybyśmy chcieli teraz policzyć:

$$EZ = \int_{-\infty}^{+\infty} \frac{1}{\pi(1+z^2)}dz = ?$$

*Odpowiedź jest taka, że ta wartość oczekiwana **NIE ISTNIEJE**. Podobnie jest w rozkładach Couchy'ego. Dalej*

$$\int_{-\infty}^{mediana} g(z) = \frac{1}{2}$$

gdy $mediana = 0$ to:

$$\int_{-\infty}^{0} \frac{1}{\pi(1+z^2)}dz = \frac{1}{\pi} arctg(z)|_{-\infty}^{0} = \frac{1}{\pi}\left(0 - \left(-\frac{\pi}{2}\right)\right) = \frac{1}{2}$$

wobec tego mediana równa zero a nie $\frac{\sqrt{2}}{2}$. Odpowiedź (c) odpada. Zostaje odpowiedź (e). Sprawdźmy:

$$\int_{-\infty}^{-1} g(z) = 0.25$$

$$\int_{-\infty}^{-1} \frac{1}{\pi(1+z^2)}dz = \frac{1}{\pi} arctg(z)|_{-\infty}^{-1} = \frac{1}{\pi}\left(-\frac{\pi}{4} - \left(-\frac{\pi}{2}\right)\right) = 0.25$$

Co należało udowodnić

Zadanie 7.6 Załóżmy, ze $X_1, \ldots, X_n, \ldots$ jest ciągiem niezależnych, dodatnich zmiennych losowych o jednakowym rozkładzie o gęstości

$$f(x) = x\exp(-x) \qquad \text{dla } x > 0$$

Niech $S_0 = 0$ i $S_n = X_1 + \ldots + X_n$ dla $n > 0$. Określmy zmienną losową N w następujący sposób:

$$N = max\{n \geqslant 0 : S_n \leqslant 4\}$$

Oblicz $P(N = 2)$.

Rozwiązanie 7.6

$$S_0 = 0$$

$$S_1 = X_1$$

$$S_2 = X_1 + X_2$$

$$S_3 = X_1 + X_2 + X_3$$

czyli prawdopodobieństwo, że

$$P(N = 2) = P(X_1 + X_2 \leqslant 4 \wedge X_1 + X_2 + X_3 > 4) =$$

Niech $X_1 + X_2 = Z$ *wtedy*

$$P(Z \leqslant 4 \wedge Z + X_3 > 4)$$

$$X_i$$

ma gęstość $Gamma(2, 1)$ *bo dla Gamma mamy*

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0$$

$$f(x) = \frac{1^2}{\Gamma(2)} x^{2-1} e^{-1\cdot x} = xe^{-x}$$

Czyli

$$Z \sim Gamma(4, 1)$$

$$f(z) = \frac{1^4}{\Gamma(4)} z^{4-1} e^{-1\cdot z} = \frac{1}{6} z^3 e^{-z}$$

Z *i* X_3 *są niezależne. Rysujemy rysunek pomocnicy i obliczamy prawdopodobieństwo:*

$$P(Z \leqslant 4 \wedge Z + X_3 > 4)$$

$$\int_0^4 \int_{4-z}^{\infty} xe^{-x} \cdot \frac{1}{6} z^3 e^{-z} dx dz =$$

$$= \int_0^4 \frac{1}{6} z^3 e^{-z} \left((4 - z)e^{-(4-z)} + e^{-(4-z)} \right) dz = 0.35166 = \frac{96}{5} e^{-4}$$

Zadanie 7.7 Obserwujemy niezależne zmienne losowe $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$. Zmienne losowe X_1, X_2, X_3, X_4 mają ten sam rozkład o dystrybuancie F_{μ_1}, a zmienne losowe $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$ mają ten sam rozkład o dystrybuancie F_{μ_2}. Dystrybuanta F_μ spełnia warunek:

$$F_\mu(x) = F(x - \mu)$$

dla pewnej ustalonej, nieznanej, ciągłej, ściśle rosnącej dystrybuanty F. Weryfikujemy hipotezę $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 > \mu_2$ stosując test o obszarze krytycznym

$$K = \{S : S \geqslant 19\}$$

gdzie S jest sumą rang tych spośród zmiennych X_1, X_2, X_3, X_4, w próce złożonej ze wszystkich obserwacji ustawnionych w ciąg rosnący, które są większe od $\max\{Y_1, Y_2, Y_3, Y_4, Y_5, Y_6\}$. Wyznaczyć rozmiar testu.

Rozwiązanie 7.7 *Warunek podany na początku zadania oznacza, że rozkłady mają ten sam kształt (są przesunięte), przy takiej samej wartości oczekiwanej (H_0) te rozkłady są identyczne. Ranga w wylosowanej próbce to pozycja zmiennej w próbce. Np.gdy*

$$X_1, Y_1, X_2, Y_3, ...$$

to Y_3 ma rangę 4. My mamy wyznaczyć rozmiar testu czyli:

$$P(K) = P(S \geqslant 19) =?$$

żeby suma rang była większa od 19 to mamy następujące przypadki (Y_m to maksymalne Y_i):

$$[], [], [], [], [], [], [], [Y_m], [X], [X] \to S = 19$$

$$[], [], [], [], [], [], [Y_m], [X], [X], [X] \to S = 27$$

$$[], [], [], [], [], [Y_m], [X], [X], [X], [X] \to S = 34$$

w powyższym należy zauważyć, że Y_m może być co najmniej na 6 pozycji! Wiemy, że skoro rozkłady są takie, to wszystkie losowania są jednakowo prawdopodobne. Zdefinijmy wobec tego wszystkie możliwe ułożenia (tu w ogole nie musimy patrzyć na numery zmiennych, mamy ciągi $X, X, Y, X, Y, \ldots$). Wszystkie możliwe układy (permutacje z powtórzeniami):

$$\frac{10!}{6!4!} = 210$$

Dla przypadku pierwszego zostają nam układy gdzie zostaje 7 miejsc do ułożenia i 5 zmiennych Y oraz 2 zmienne X

$$\frac{7!}{5!2!} = 21$$

Przypadek drugi:

$$\frac{6!}{5!1!} = 6$$

Przypadek trzeci (tylko jedna permutacja same Yki)

$$1$$

Szukane prawdopodobieństwo:

$$P(S \geqslant 19) = \frac{21+6+1}{210} = \frac{14}{105}$$

Do zadania można podejść też inaczej. Prawdopodobieństwo przypadku 1 to (losujemy najpierw X, potem X, potem Y; losujemy 'od końca'):

$$\frac{4}{10} \cdot \frac{3}{9} \cdot \frac{6}{8} = \frac{1}{10}$$

Przypadek dwa (losujemy X,X,X,Y)

$$\frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{6}{7} = \frac{1}{35}$$

Przypadek 3:

$$\frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{1}{7} \cdot \frac{6}{6} = \frac{1}{210}$$

Sumując prawdopodobieństwa dostajemy $\frac{14}{105}$.

Zadanie 7.8 Łańcuch Markowa ma trzy stany E_1, E_2, E_3 i macierz przejścia: Niech X_n oznacza stan, w którym znajduje się łańcuch po dokonaniu n kroków

1/3	2/3	0
1/4	1/4	1/2
1/2	0	1/2

$n = 0, 1, \ldots$. Funkcję f na zbiorze stanów określamy wzorem $f(E_i) = i - 1$ dla $i = 1, 2, 3$.
Niech $c = \lim_{n\to\infty} Cov(f(X_n), f(X_{n+1}))$. Granica c jest równa: ?

Rozwiązanie 7.8 *Z definicji kowariancji:*

$$Cov(f(X_n), f(X_{n+1}) = E(f(X_n) \cdot f(X_{n+1})) - E(f(X_n)) \cdot E(f(X_{n+1}))$$

Obliczymy prawdopodobieństwa bycia w stanie po n krokach:

$$p_1 = \frac{1}{3}p_1 + \frac{1}{4}p_2 + \frac{1}{2}p_3$$

$$p_2 = \frac{2}{3}p_1 + \frac{1}{4}p_2$$

$$p3 = \frac{1}{2} + \frac{1}{2}p_3$$

$$p_1 + p_2 + p_3 = 1$$

Po rozwiązaniu mamy: $p_1 = \frac{9}{25}$, $p_2 = \frac{8}{25}$, $p_3 = \frac{8}{25}$. *Dalej:*

$$f(E_1) = 0$$

$$f(E_2) = 1$$

$$f(E_3) = 2$$

mamy:

$$E(f(X_n)) = E(f(X_{n+1})) = 1 \cdot p_2 + 2 \cdot p_3 = \frac{24}{25}$$

Mamy macierz wartości dla $f(X_n) \cdot f(X_{n+1})$ *Biorąc pod uwagę powyższą tabelę*

0	0	0
0	1	2
0	2	4

mamy:

$$E(f(X_n) \cdot f(X_{n+1})) = p_2 \cdot 1 \cdot \frac{1}{4} + p_2 \cdot 2 \cdot \frac{1}{2} + p_3 \cdot 2 \cdot 0 + p_3 \cdot 4 \cdot \frac{1}{2} = \frac{26}{25}$$

$$Cov(f(X_n), f(X_{n+1}) = \frac{26}{25} - \left(\frac{24}{25}\right)^2 = \frac{74}{625}$$

8 Egzamin z 23 marca 2015

Zadanie 8.1 Rozważamy model regresji liniowej postaci $Y_i = a + bx_i + \epsilon_i$, $i = 1, 2, 3, 4, 5$, gdzie b jest nieznanym parametrem rzeczywistym, $x_1 = x_2 = 1$, $x_3 = 2$, $x_4 = x_5 = 3$, 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 $\{|\frac{\hat{b}}{\hat{\sigma}}| > c\}$, gdzie $\hat{b}$, $\widehat{sigma}$, są estymatorami największej wiarygodności parametrów b i σ, a stała c dobrana jest tak, aby test miał rozmiar 0.05. Stała c jest równa:?

Rozwiązanie 8.1 *PODOBNE zadanie 2, 26.05.2014, prawie identyczne jak zadanie 3, 28.05.2012*

Dla przypomnienia w rozkładzie normalnym:

$$aX + b \sim N(a\mu + b, (a\sigma)^2)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

W rozkładzie normalnym:

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

W regresji liniowej funkcja wiarygodności jest postaci:

$$L = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-a-bx_i)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5 e^{\sum_{i=1}^{5} -\frac{(y_i-a-bx_i)^2}{2\sigma^2}}$$

Bierzemy logarytm z funkcji wiarygodności:

$$\ln L = -5\ln\sqrt{2\pi} - 5\ln\sigma +$$

$$-\frac{(y_1-a-b)^2 + (y_2-a-b)^2 + (y_3-a-2b)^2 + (y_4-a-3b)^2 + (y_5-a-3b)^2}{2\sigma^2}$$

Pochodne:

$$\frac{\partial \ln L}{\partial b} = 0$$

$$\frac{\partial \ln L}{\partial b} =$$

$$= -\frac{-2(y_1-a-b) - 2(y_2-a-b) - 2\cdot 2(y_3-a-2b) - 2\cdot 3(y_4-a-3b) - 2\cdot 3(y_5-a-3b)}{2\sigma^2}$$

stąd

$$24b = y_1 + y_2 + 2y_3 + 3y_4 + 3y_5 - 10a$$

$$\frac{\partial \ln L}{\partial a} = 0$$

liczymy analogicznie jak wyżej i uzyskujemy:

$$5a = y_1 + y_2 + y_3 + y_4 + y_5 - 10b$$

wstawiamy do równania z b i uzyskujemy:

$$4b = -y_1 - y_2 + y_4 + y_5 \to b = \frac{-y_1 - y_2 + y_4 + y_5}{4}$$

Stąd uwzględniając rozkład ϵ_i

$$b \sim N\left(0, 4 \cdot \frac{\sigma^2}{4^2}\right) = N\left(0, \frac{\sigma^2}{4}\right)$$

Jaki jest estymator odchylenia standardowego?

$$\frac{\partial \ln L}{\partial \sigma} = 0$$

Stąd

$$\sigma^2 = \frac{1}{5}\sum_{i=1}^{5}(y_i - a - bx_i)^2 =$$

$$= \frac{1}{5}\left((y_1 - a - b)^2 + (y_2 - a - b)^2 + (y_3 - a - 2b)^2 + (y_4 - a - 3b)^2 + (y_1 - a - 3b)^2\right) =$$

mamy $(y_1 - a - b)(y_1 - a - b) = y_1^2 - 2y_1a - 2y - 1b + 2ab + b^2$ *TU SIE STRASZNIE KOMPLIKUJE*

$y_i \sim N(0, \sigma^2)$. *Dowodzimy, że* $\frac{\hat{b}}{\hat{\sigma}}$ *ma rozkład t-Studenta.*

DOKOŃCZYĆ

Zadanie 8.2 Niech X i Y będa niezaleznymi zmienymi losowymi każda z rozkładu wykładniczego o wartości oczekiwanej 1.
Niech $U = 2X + Y$ i $V = X - 2Y$.
Wtedy prawdopodobieństwo $P(U \in (0, 5) \wedge V \in (0, 5))$ jest równe:?

Rozwiązanie 8.2 *Rozpiszemy:*

$$P(2X + Y \in (0, 5) \wedge X - 2Y \in (0, 5)) =$$

$$P(0 < 2X + Y < 5 \wedge 0 < X - 2Y < 5) =$$

$$P\left(-2X < Y < 5 + 2X \wedge \frac{X}{2} > Y > \frac{X - 5}{2}\right)$$

Rysujemy rysunek pomocniczy z obszarami i dostajemy całkę podwojną:

$$\int_0^2 \int_0^{\frac{x}{2}} e^{-x}e^{-y}dydx + \int_2^{2.5} \int_0^{5-2x} e^{-x}e^{-y}dydx =$$

$$= \frac{1}{3} - 2e^{-2.5} + \frac{5}{3}e^{-3}$$

Zadanie 8.3 Rozważmy nastepujące zagadnienie testowania hipotez statystycznych. Dysponujemy próbką $X_1, \ldots, X_n$ z rozkładu normalnego o nieznanej średniej μ i znanej wariancji równej 9. Przeprowadzamy najmocniejszy test hipotezy $H_0 : \mu = 0$ przeciwko alternatywie $H_1 : \mu = 2$ na poziomie istotności $\alpha = 1/2$. Niech β_n oznacza prawdopodobieństwo błędu drugiego rodzaju, dla rozmiaru próbki n.
Wybierz poprawne stwierdzenie: 1.

$$\lim_{n\to\infty} \beta_n \exp(\frac{2n}{9})\sqrt{2\pi} = 1$$

2.

$$\lim_{n\to\infty} \beta_n \exp(\frac{2n}{9})\frac{2\sqrt{2\pi n}}{3} = 1$$

itd.

Rozwiązanie 8.3 *Dla przypomnienia:*

Błąd I rodzaju: ***odrzucenie*** H_0***, gdy jest prawdziwa***
Błąd II rodzaju: ***nie odrzucenie*** H_0***, gdy jest fałszywa***

Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia H_1 ***przy jej prawdziwości (odrzucenie*** H_0 ***przy jej fałszywości)****. Jest to 1 - p-p błędu II rodzaju.*

Test najmocniejszy (o najmniejszym błędzie drugiego rodzaju) budujemy poprzez badanie:

$$\frac{L_1}{L_0} > c$$

wtedy odrzucamy H_0 *przy jej prawdziwości.* L_1 *to funkcja wiarygodności przy* H_1*,* L_0 *to funkcja wiarygodności przy* H_0 *Mamy:*

$$\frac{L_1}{L_0} = \frac{(\frac{1}{\sqrt{2\pi\sigma}})^n \exp(-\sum_{i=1}^n \frac{(x_i-2)^2}{2\sigma^2})}{(\frac{1}{\sqrt{2\pi\sigma}})^n \exp(-\sum_{i=1}^n \frac{x_i^2}{2\sigma^2})} > c$$

$$\exp\left(-\sum_{i=1}^n \frac{(x_i-2)^2}{2\sigma^2} + \sum_{i=1}^n \frac{x_i^2}{2\sigma^2} +\right) > c$$

$$-\sum_{i=1}^n \frac{(x_i-2)^2}{2\sigma^2} + \sum_{i=1}^n \frac{x_i^2}{2\sigma^2} > c$$

$$\sum_{i=1}^n x_i > c$$

Jakie jest p-p, że

$$P\left(\sum_{i=1}^n x_i > c\right) = 0.5$$

Suma ma rozkład $N(0, 9n)$ można znormalizować i mamy:

$$P\left(\frac{\sum x_i}{\sqrt{9n}} > \frac{c}{\sqrt{9n}}\right) = 0.5$$

co oznacza, że $\frac{c}{\sqrt{9n}} = 0$ czyli mamy:

$$P(\frac{\sum x_i}{\sqrt{9n}} > 0) = 0.5$$

co dalej daje np.

$$P(\sum x_i > 0) = 0.5$$

Stąd też można dojśc do wniosku, że

$$P(\bar{X} > 0) = 0.5$$

czyli, że średnia też jest testem najmocniejszym.

Teraz chcemy policzyć błąd drugiego rodzaju: 'przyjęcie' H_0 gdy jest fałszywa:

$$P(\bar{X} < 0|H_1)$$

Średnia przy H_1 ma rozkład $N(2, n \cdot \frac{9}{2}) = N(2, \frac{9}{n})$. Normalizujemy:

$$P\left(\frac{\bar{X} - 2}{\sqrt{9/n}} < \frac{-2}{\sqrt{9/n}}\right) = \beta_n$$

$$\beta_n = \int_{-\infty}^{-\frac{2}{\sqrt{9/n}}} \frac{1}{\sqrt{2\pi} \cdot 1} \exp\left(-\frac{x^2}{2}\right) dx =?$$

tej całki analitycznie nie policzymy. Widzimy, że przy dążeniu do nieskończoności β_n dąży do zera. W odpowiedziach mamy np

$$\lim_{n\to\infty} \beta_n \cdot n$$

więc mamy iloczyny czegoś co dąży do zera i drugiego czynnika dążącego do nieskończoności. Będziemy stosowali twierdzenie de l'Hospitala. Policzymy pochodną z B_n

$$B_n' = -\frac{1}{3\sqrt{2\pi}\sqrt{n}} e^{-\frac{2n}{9}}$$

i podstawiamy do odpowiedzi (mianownik przekształcamy do postaci 1 / wyrażenie. Wychodzi dobra odpowiedź:

$$\lim_{n\to\infty} \beta_n \exp(\frac{2n}{9}) \frac{2\sqrt{2\pi n}}{3} = 1$$

Zadanie 8.4 Na podstawie próby losowej $X_1, \ldots, X_n$ gdzie X_i, $i = 1, \ldots, n$ są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na przedziale $(0, \theta)$ i $\theta > 0$ jest nieznanym parametrem, zbudowono przedział ufności dla parametru θ na poziomie ufności $1-\alpha$ postaci: $[T_X, aT_X]$ gdzie T_X jest estymatorem największej wiarogodnosci parametru θ na podstawie próby $X_1, \ldots, X_n$.
Następnie uzyskano niezależnie drugą próbę losową $Y_1, \ldots, Y_m$ z tego samego rozkładu i zbudowano przedział ufności dla parametru θ na poziomie ufności $1 - \alpha$ postaci $[T_{XY}, bT_{XY}]$,
gdzie T_{XY} jest estymatorem największej wiarogodności parametru θ na podstawie próby $X_1, \ldots, X_n, Y_1, \ldots, Y_m$.

Oblicz prawdopodobieństwo, że tak utworzone przedziały będą rozłączne.

Rozwiązanie 8.4 *W rozkładzie jednostajnym estymator największej wiarygodności parametru θ to maksimum. Stąd:*

$$T_X = \max(X_1, \ldots, X_n) = X_m$$

$$T_{XY} = \max(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$$

Z pierwszego przedziału ufności wiemy, że

$$P(X_m < \hat{\theta} < aX_m)$$

możemy narysować mały rysunek pomocniczy. Teraz mamy określić prawdopodobieństwo, że przedziały będą rozłączne, oznacza to, że

$$P(T_{XY} > aX_m) = P(Y_m > aX_m)$$

bo, któryś Y_i musiał zwiększyć nasz estymator największej wiarygodności . Wyznaczymy najpierw a

$$P(X_m < \hat{\theta} < aX_m) = 1 - \alpha$$

$$P(X_m < \hat{\theta} < aX_m) = P(\hat{\theta} < aX_m) - \underbrace{P(\hat{\theta} < X_m)}_{=0} = 1 - \alpha$$

Drugie p-p równe zero bo rozkład ciągły.

$$Pr\left(\frac{\hat{\theta}}{a} < X_m\right) = 1 - \alpha$$

Rozkład maksimum $P(X_m < t) = (\frac{t}{\theta})^n$, $P(X_m > t) = 1 - (\frac{t}{\theta})^n$. Stąd

$$1 - \left(\frac{\theta/a}{\theta}\right)^n = 1 - \alpha \to \alpha = \left(\frac{1}{a}\right)^n \to a = \alpha^{-1/n}$$

Dalej już prosto:

$$P(Y_m > \alpha^{-1/n} X_m) =?$$

Rysunek pomocniczy. Generalnie $P(X \leqslant Y) = \int_0^\infty \int_0^y f(x)f(y)dxdy$. *Mamy:*

$$f_{x_m}(x) = n \cdot \frac{x^{n-1}}{\theta^n}$$

$$f_{y_m}(x) = m \cdot \frac{y^{m-1}}{\theta^m}$$

$$\int_0^\theta \int_0^{y/a} n \cdot m \cdot \frac{x^{n-1}y^{m-1}}{\theta^{n+m}} dxdy = \frac{\alpha m}{m+n}$$

Zadanie 8.5 Niech $Z_1, Z_2, \ldots, Z_n, \ldots$ będą niezależnymi zmiennymi losowymi o rozkładzie Pareto o gęstości

$$p_{\lambda,\theta}(x) = \begin{cases} \frac{\lambda^\theta \theta}{(x+\lambda)^{\theta+1}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}$$

gdzie $\theta > 1$, $\lambda > 0$ są ustalonymi liczbami. Niech N będzie zmienną losową niezależną od $Z_1, \ldots, Z_n, \ldots$, o rozkładzie gemetrycznym

$$P(N = n) = (1-q)q^{n-1}$$

gdy $n = 1, 2, 3, \ldots,$, gdzie $q \in (0,1)$ jest ustaloną liczbą.
Wyznaczyć $E(Z_1 + Z_2 + \ldots + Z_N | \min(Z_1, Z_2, \ldots, Z_N) = t)$, gdzie t jest ustaloną liczbą większą od 0.

Rozwiązanie 8.5 *Wiemy, że*

$$E(Z_1 + \ldots + Z_N | \min(Z_1, \ldots, Z_N) = t) = t + (N-1)EZ_1$$

pod warunkiem, że N.

Należy zwrócić uwagę, że $\min(Z_1, \ldots, Z_N) = t$, *czyli rozkład jest ucięty! Przeskalujemy gęstość:*

$$\int_t^\infty c \cdot \frac{\lambda^\theta}{(x+\lambda)^{\theta+1}} dx = 1$$

stąd

$$c = \frac{(t+\lambda)^\theta}{\lambda^\theta}$$

czyli przeskalowana gęstość:

$$f(x) = \frac{(t+\lambda)^\theta \lambda}{(x+\lambda)^{\theta+1}}$$

dla $x > t$.

Liczymy wartość oczekiwaną z Z_1:

$$EZ_1 = \int_t^\infty x \frac{(t+\lambda)^\theta \lambda}{(x+\lambda)^{\theta+1}} dx = (t+\lambda)^\theta \lambda \int_t^\infty \frac{x}{(x+\lambda)^{\theta+1}} dx =$$

Liczymy przez części i dostajemy:

$$= \frac{\theta t + \lambda}{\theta - 1}$$

Dla przypomnienia:

Rozkład gemetryczny

$$P(X = k) = (1 - p)^{k-1} \cdot k$$

dla $k = 1, 2, 3, \ldots$. *Intepretowany jako p-p pierwszego sukcesu w k-tej próbie*

$$E(X) = \frac{1}{p}$$

$$Var(X) = \frac{1 - p}{p^2}$$

Liczona wartość oczekiwana z początku zadania była pod warunkiem N, wobec tego liczymy wartość oczekiwaną:

$$E(t + (N - 1)EZ_1) = t + E(N)E(Z_1) - EZ_1 = \frac{\lambda q + \theta t - (1 - q)t}{(1 - q)(\theta - 1)}$$

Zadanie 8.6 Niech X będzie zmienną losową o funkcji gęstości

$$f_\theta(x) = \begin{cases} \frac{1}{\theta^2}(\theta - |x|) & \text{gdy } |x| < \theta \\ 0 & \text{w przeciwnym wypadku} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Na podstawie pojedynczej obserwacji weryfikujemy hipotezę $H_0 : \theta = 1$ przy alternatywie $H_1 : \theta \neq 1$ testem opartym na ilorazie wiarogodności na poziomie istotności 0.1. Moc tego testu przy alternatywie $\theta = 2$ jest równa ?.

Rozwiązanie 8.6 *Dla przypomnienia:*

Błąd I rodzaju: ***odrzucenie*** H_0***, gdy jest prawdziwa***

Błąd II rodzaju: ***nie odrzucenie*** H_0***, gdy jest fałszywa***

Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia H_1 ***przy jej prawdziwości (odrzucenie*** H_0 ***przy jej fałszywości).***

Test jest oparty na ilorazie wiarygodności więc potrzebujemy dwóch funkcji wiarygodności (i ich pochodnych do policzenia supremum):

$$L_0(x, \theta = 1) = \frac{1}{1}(1 - |x|) = 1 - |x|$$

$$L_1(x, \theta \neq 1) = \frac{1}{\theta^2}(\theta - |x|)$$

bo test oparty na ilorazie wiarygodności:

$$\frac{\sup_{\theta \in \Theta_1} L(\theta, x)}{\sup_{\theta \in \theta_0} L(\theta, x)}$$

Test oparty na ilorazie wiarygodności: gdy wynik L_1/L_0 jest większy od pewnej liczby to ***odrzucamy*** *hipotezę zerową. Tym samym, gdy L_0/L_1 mniejsze od tej liczy to również odrzucamy.*

$$\frac{L_1}{L_0} = \frac{\frac{1}{\theta^2}(\theta - |x|)}{1 - |x|}$$

Do testu potrzebujemy supremum

$$\sup_{\theta \in \Theta_0} L_0 = 1 - |x|$$

$$\sup_{\theta \in \Theta_1} L_1 = ?$$

$$\frac{\partial}{\partial \theta}\left(\frac{1}{\theta^2}(\theta - |x|)\right) = -\frac{1}{\theta^2} + 2\frac{|x|}{\theta^3} = 0 \rightarrow \theta = 2|x|$$

wstawiamy to do ilorazu wiarygodności i mamy:

$$\frac{L_1}{L_0} = \frac{1}{4(|x| - |x|^2)}$$

H_0 odrzucimy gdy

$$\frac{1}{4(|x| - |x|^2)} > \gamma$$

czyli

$$4(|x| - |x|^2) < \gamma$$

Przy H_0 mamy:

$$P(4|x| - |x|^2 < \gamma | \theta = 1) = 0.1$$

sprawdźmy dla jakich γ powyższe zachodzi

$$|x| - |x|^2 < \gamma/4 \rightarrow |x| - |x|^2 - \gamma/4 < 0$$

$$|x_1| = 1/2 - 1/2\sqrt{1 - \gamma}$$

$$|x_2| = 1/2 + 1/2\sqrt{1 - \gamma}$$

to oznacza, że pomiędzy tymi dwoma punktami funkcja kwadratowa jest wieksza od zera co oznacza, że w tym obszarze przyjmujemy H_0. Test jest dwustronny (mamy wartości bezwzględne), więc o ile dla całego obszaru $P(x \in D) = 1 - 0.1 = 0.9$ to dla powyższego jest to połowa $P(x \in D) = 0.45$. Stąd gęstość:

$$\frac{1}{\theta^2}(\theta - |x|)$$

przy $H_0 : \theta = 1$ jest $1 - |x|$, w połowie obszaru $1 - x$

$$\int_{1/2-1/2\sqrt{1-\gamma}}^{1/2+1/2\sqrt{1-\gamma}} (1-x)dx = 0.44 \to \sqrt{1-\gamma} = 0.9$$

Moc testu to odrzucenie H_0 przy jej fałszywości czyli, 1 - przyjęcie H_0 przy $\theta = 2$ czyli

$$\int_{1/2-1/2\sqrt{1-\gamma}}^{1/2+1/2\sqrt{1-\gamma}} \frac{1}{2^2}(2-|x|)dx = 0.675$$

$$ODP = 1 - 0.675 = 0.325$$

Zadanie 8.7 Rozwazmy ciąg niezależnych dwuwymiarowych zmiennych losowych $(X_n, Y_n)_{n=1}^{+\infty}$, gdzie (X_n, Y_n) mają rozkłady jednostajne na zbiorze $[-2, 2]x[-2, 2]$. Niech

$$S_n = (S_{n,1}, S_{n,2}) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i)$$

oraz

$$|S_n| = \sqrt{S_{n,1}^2 + S_{n,2}^2}$$

Stałą c dobrano tak, aby:

$$\lim_{n\to\infty} P(|S_n| < c\sqrt{n}) = 0.95$$

Stała c jest równa ?.

Rozwiązanie 8.7 ***Twierdzenia graniczne***

SUMA DUŻEJ LICZBY ZMIENNYCH LOSOWYCH Z JEDNAKOWEGO ROZKŁADU MA ROZKŁAD NORMALNY

$S_{n,1}$ ma rozkład jednostajny o wartości oczekiwanej $\frac{-2+2}{2} = 0$ oraz wariancji $Var(S_{n,1}) = \frac{(b-a)^2}{12} = \frac{16}{12} = 4/3$, $S_{n,1}$ przy $n \to \infty$ będzie miała rozkład $N(0, 4/3 \cdot n)$ analogicznie $S_{n,2}$ będzie miała rozkład $N(0, 4/3 \cdot n)$. Mamy:

$$P(|S_n| < c\sqrt{n}) = P\left(\sqrt{S_{n,1}^2 + S_{n,2}^2} < c\sqrt{n}\right) =$$

obie strony sa dodanie więc można podnieśc do kwadratu:

$$= P(S_{n,1}^2 + S_{n,2}^2 < c^2 \cdot n)$$

Znormalizujmy $S_{n,1}$ i $S_{n,2}$:

$$\frac{S_{n,1}}{\sqrt{4/3 \cdot n}} \sim N(0, 1)$$

$$\frac{S_{n,2}}{\sqrt{4/3 \cdot n}} \sim N(0, 1)$$

Jeżeli podniesiemy te zmienne do kwadratu to otrzymamy zmienne o rozkładzie χ^2*:*

$$\left(\frac{S_{n,1}}{\sqrt{4/3 \cdot n}}\right)^2 \sim \chi^2(1)$$

$$\left(\frac{S_{n,2}}{\sqrt{4/3 \cdot n}}\right)^2 \sim \chi^2(1)$$

Rozkład χ^2 *to szczególny przypadek rozkładu Gamma zachodzi:*

$$\chi^2(n) + \chi^2(k) = \chi^2(n+k)$$

czyli mamy:

$$P\left(\underbrace{\frac{S_{n,1}^2}{4/3 \cdot n} + \frac{S_{n,2}^2}{4/3 \cdot n}}_{\sim\chi^2(2)} < \frac{c^2 \cdot n}{4/3 \cdot n}\right) = 0.95$$

Z tablic dla $\alpha = 0.05$ *wartości krytyczne to* 5.99146

$$\frac{c^2}{4/3} = 5.99146 \to c = 2.826$$

Zadanie 8.8 Niech $X_1, \ldots, X_n, X_{n+1}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie jednostajnym na przedziale $(0, \theta)$. Parametr $\theta > 0$ jest nieznany i jest realizacją zmiennej losowej o rozkładzie o gęstości

$$p(\theta) = \begin{cases} \frac{4}{\theta^5} & \text{gdy } \theta > 1, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Wyznaczamy predyktor bayesowski $\hat{X}_{n+1}$ zmiennej X_{n+1} w oparciu o próbę $X_1, X_2, \ldots, X_n$ przy kwadratowej funkcji straty. Wartość oczekiwana tego predyktora czyli wielkość $E(\hat{X}_{n+1}|\theta)$ jest równa: ?.

Rozwiązanie 8.8 *Kwadratowa funkcja straty, czyli będziemy minimalizowali błąd kwadratowy* $E((\hat{X}_{n+1} - X_{n+1})^2)$*. Bezpośrednio jednakże w tym zadaniu nie skorzystamy z tego. Zmienna* X_{n+1} *pochodzi z rozkładu jednostajnego, więc moglibyśmy powiedzieć, że pod warunkiem, że* θ *jest:*

$$E(X_{n+1}|\theta) = \frac{\theta}{2}$$

Przy kwadratowej funkcji straty ***predyktor Bayesowski*** *to (korzystamy z iteracyjności):*

$$\hat{X}_{n+1} = E(X_{n+1}|X_1, \ldots, X_n) = E(E(X_{n+1}|\theta)|X_1, \ldots, X_n) = E\left(\frac{\theta}{2}|X_1, \ldots, X_n\right)$$

Iteracyjność: Gdy

$$F_1 \subset F_2 \subset M \to E(Y|F_1) = E(E(Y|F_2)|F_1) = E(E(Y|F_1)|F_2)$$

Rozkład θ jest zdefiniowany przez próbę $X_1, \dots, X_n$, tzn. jeżeli okazuje się, że wszystkie X_i są bardzo małe, można spodziewać się, że θ też jest mała. Potrzebne nam będzie:

$$f(\theta|X_1, \dots, X_n)\frac{f(\theta \wedge X_1, \dots, X_n)}{f(X_1, \dots, X_n)} = \frac{f(\theta \wedge X_1, \dots, X_n)}{\int f(X_1, \dots, X_n, \theta)d\theta}$$

mamy:

$$f(X_1, \dots, X_n|\theta) = \frac{1}{\theta^n}$$

$$f(X_1, \dots, X_n \wedge \theta) = \frac{1}{\theta^n} \cdot \frac{4}{\theta^5}$$

Musimy wycałkować ostatnie po $d\theta$, tutaj należy ostrożnie podejść do granic całkowania bo θ generalnie może być nie mniejsza niż x_{max}, ale należy zwrócić uwagę, że x_m może być mniejsze od 1, wtedy θ i tak zgodnie z rozkładem musi być większa od 1. Wobec tego:

$$\int_{x_m}^{\infty} \frac{4}{\theta^{n+5}} d\theta = \frac{4}{n+4} \cdot \frac{1}{x_n^{n+4}}$$

$$\int_{1}^{\infty} \frac{4}{\theta^{n+5}} d\theta = \frac{4}{n+4}$$

$$f(\theta|X_1, \dots, X_n) = \begin{cases} \frac{n+4}{\theta^{n+5}} & gdy\ x_m < 1, \theta \in (1, \infty) \\ \frac{(n+4)x_m^{n+4}}{\theta^{n+5}} & gdy\ x_m \geqslant 1, \theta \in (x_m, \infty) \end{cases}$$

Teraz jak wrócimy i policzymy ile równa się $\hat{X}_{n+1}$ to mamy $E(\theta/2)$:

$$\int_{1}^{\infty} \frac{\theta}{2} \frac{n+4}{\theta^{n+5}} d\theta = \frac{1}{2} \frac{n+4}{n+3}$$

$$\int_{x_m}^{\infty} \frac{\theta}{2} \frac{(n+4)x_m^{n+4}}{\theta^{n+5}} = \frac{1}{2} \frac{n+4}{n+3} \cdot x_m$$

Widać, że wartość oczekiwana zależy od X_m, mamy policzyć

$$E(\hat{X}_{n+1}|\theta) = ?$$

Z Law of total expectations:

$$E(\hat{X}_{n+1}) = E(\hat{X}_{n+1}|x_m < 1) \cdot P(x_m < 1) + E(\hat{X}_{n+1}|x_m \geqslant 1)P(x_m \geqslant 1)$$

Jaki rozkład ma x_m?

$$P(\max(X_1, \dots, X_n) < t) = \left(\frac{x}{\theta}\right)^n$$

$$f(x|\theta) = n\frac{x^{n-1}}{\theta^n}$$

stąd

$$\hat{X}_{n+1} = \begin{cases} \frac{1}{2}\frac{n+4}{n+3} & gdy\ x_m < 1,\ \ z\ p\text{-}p\ \frac{1}{\theta^n} \\ \frac{1}{2}\frac{n+4}{n+3}x_m & gdy\ x_m \geqslant 1,\ \ z\ p\text{-}p\ 1-\frac{1}{\theta^n} \end{cases}$$

$$E(\hat{X}_{n+1}|\theta) = \frac{1}{2}\frac{n+4}{n+3}\cdot\frac{1}{\theta^n} + E(\hat{X}_{n+1}|x_m \geqslant 1)P(x_m \geqslant 1) =$$

po prawej stronie zostanie tylko całka z licznika. Mamy:

$$= \frac{1}{2}\frac{n+4}{n+3}\cdot\frac{1}{\theta^n} + \int_1^\theta \frac{1}{2}\frac{n+4}{n+3}x\cdot n\cdot\frac{x^{n-1}}{\theta^n}dx = \frac{1}{2}\frac{n+4}{(n+3)(n+1)}\left(n\theta + \frac{1}{\theta^n}\right)$$

Zadanie 8.9 Rzucamy czterema symetrycznymi monetami. Następnie rzucamy ponownie tymi monetami, na których nie wypadły 'orły'. W trzeciej rundzie rzucamy tymi monetami, na których do tej pory nie wypadly 'orły'. Oblicz prawdopodbieństwo, że po trzech rundach na wszystkich monetach będą 'orły' (wybierz najbliższą wartość).

Rozwiązanie 8.9 *Najpierw rozpatrzymy pojedynczą monetę. Jakie jest p-p uzyskania orła w 3 rzutach?:*

$$1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$

Wobec tego jakie jest prawdopobieństwo sukcesu dla 4 monet?

$$\left(\frac{7}{8}\right)^4 = 0.586$$

Zadanie 8.10 Niech A i C będą zdarzeniami niezależnymi oraz $P(A) = \frac{1}{3}$ i $P(A \cup C) = 7/9$. Niech $P(B|A) = P(B|C) = P(B|A \cap C) = 1/2$ i $P(B'|A' \cap C') = 3/4$. Wtedy $P(A|B)$ jest równe.

Rozwiązanie 8.10 *Do takich zadań podchodzi licząc się wszystkie możliwe prawdopodbieństwa po drodze. Ważne wzory:*

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

$$P(X \cup Y \cup Z) = P(X) + P(Y) + P(Z) - P(X \cap Y) - P(X \cap Z) - P(Y \cap Z) + P(X \cap Y \cap Z)$$

$$P(X' \cap Y') = 1 - P(X \cup Y)$$

ostatni wzór działa również dla 3.

mamy:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{1/3} = 1/2 \rightarrow P(A \cap B) = 1/6$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{P(B)}$$

Z niezależności:

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = P(A) + P(C) - P(A) \cdot P(C) = 7/9$$

Stąd

$$P(C) = 2/3$$

i stąd

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = 1/2 \rightarrow P(B \cap C) = 1/3$$

Dalej

$$P(B|A \cap C) = \frac{P(A \cap B \cap C)}{P(A) \cdot P(C)} = 1/2 \rightarrow P(A \cap B \cap C) = 1/2 \cdot 1/3 \cdot 2/3 = 1/9$$

Dalej

$$P(B'|A' \cap C') = 3/4 = \frac{P(A' \cap B' \cap C')}{P(A' \cap C')} \rightarrow P(A' \cap B' \cap C') = 1/6$$

Z wzoru:

$$P(A' \cap C') = 1 - P(A \cup C) = 1 - 7/9 = 2/9$$

$$1/6 = P(A' \cap B' \cap C') = 1 - P(A \cup B \cup C)$$

$$5/6 = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

stąd

$$P(B) = 4/9$$

czyli finalnie

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 3/8$$

9 Egzamin z 15 czerwca2015

Zadanie 9.1 Niech $X_1, \ldots, X_9$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa:

$$P(X_i = 1) = 3/5 \quad \text{i} \quad P(X_i = 1) = 2/5$$

Niech $S_k = \sum_{i=1}^{k} X_i$ dla $k = 1, 2, \ldots, 9$ Prawdopodobieństwo:

$$P(S_9 = 3 \wedge S_1 < 5, S_2 < 5, \ldots, S_8 < 5)$$

jest równe?

Rozwiązanie 9.1 *Jakie jest prawdopodobieństwo, że $S_9 = 3$, ale bez dodakowych warunków na pozostałe sumy? Tzn w 9 próbach mamy wylosować 6 jedynek i 3 minus jedynki (jedyna opcja żeby uzyskać 3).*

$$P(S_9 = 3) = \binom{9}{6}\left(\frac{3}{5}\right)^6\left(\frac{2}{5}\right)^3$$

gdzie $\binom{9}{6}$ odpowiada ze liczbę kombinacji, a pozostała cześć za prawdopodobieństwo. Wobec tego musimy policzyć kombinacje, które nie spełniają: niech 1 oznacza 1 a - oznacza -1:

1) 1 1 1 1 1 1 - - -
2) 1 1 1 1 1 - 1 - -
3) 1 1 1 1 1 - - 1 -
4) 1 1 1 1 1 - - - 1
5) 1 1 1 1 - 1 1 - -
6) 1 1 1 - 1 1 1 - -
7) 1 1 - 1 1 1 1 - -
8) 1 - 1 1 1 1 1 - -
9) - 1 1 1 1 1 1 - -

Pozostałe kombinacje nie odpadają. Kombinacja np. nr 5 odpada bo mamy $S_1, S_2, \ldots, S_9$ równe odpowiednio $1, 2, 3, 4, 3, 4, 5, 4, 3$ w ciągu pojawia się 5tka dlatego kombinacja musi odpaść. Mamy:

$$P(S_9 = 3 \wedge S_1 < 5, S_2 < 5, \ldots, S_8 < 5) =$$

$$= \left(\binom{9}{6} - 9\right)\left(\frac{3}{5}\right)^6\left(\frac{2}{5}\right)^3 = 0.2239$$

Zadanie 9.2 W urnie znajduje się 10 kul Zielonych, 10 kul Białych i 10 kul Czarnych. Losujemy bez zwracania 9 kul. Niech

- Z oznacza liczbę wylosowanych kul Zielonych,
- B oznacza liczbę wylosowanych kul Białych

- C oznacza liczbę wylosowanych kul Czarnych

Wtedy współczynnik kowriancji $Cov(Z, B)$ jest równy?

Rozwiązanie 9.2 *Na cale założmy, że kule są ponumerowane i ułozone w kolejności:*

$$X_1, \ldots, X_{10}, X_{11}, \ldots, X_{20}, X_{21}, \ldots, X_{30}$$

$X_1, \ldots, X_{10}$ *to kule zielone,* $X_{11}, \ldots, X_{20}$ *to kule białe,* $X_{21}, \ldots, X_{30}$ *to kule czarne. Jakie jest prawdopodobieństwo dla pojedynczej kulki, że zostanie wylosowana (bez patrzenia na kolor)? Analogiczny przypadek to gdybym miał grupę 30 osób i na super wycieczke losowane było 9 osób. Jakie mam prawdopodbieństwo, że zostanę wylosowany?*

$$P_1 = \frac{\binom{N-1}{k-1}}{\binom{N}{k}} = \frac{k}{N}$$

Gdzie $k = 9, N = 30$. *W liczniku są wszystkie 9 osobowe drużyny w których ja jestem, w mianowniku są wszystkie możliwe drużyny. Niech teraz w klasie będzie 10 aktuariuszy. Każdy aktuariusz ma prawdopodobieństwo pojechania na wycieczkę* $9/30$ *jaka jest wartość oczekiwana liczby wylosowanych aktuariuszy? Niech teraz w naszym przykładzie* X_i *oznacza czy i-ty aktuariusz został wylosowany (przyjmuje wartości* 0 *- nie wylosowany,* 1 *- wylosowany; czyli rozkład dwupuntkowy). Wtedy* Z *to liczba wylosowanych aktuariuszy i* Z *jest sumą* X_i *po* $i = 1, \ldots, 10$. *Wtedy:*

$$E(Z) = 10 \cdot E(X_i) = 10 \cdot \frac{9}{30}$$

Powyższe rozumowanie stosujemy do kul i mamy

$$E(Z) = 10 \cdot \frac{9}{30}$$

$$E(B) = 10 \cdot \frac{9}{30}$$

Teraz wiemy z definicji kowariancji, że

$$Cov(Z, B) = E(Z \cdot B) - E(Z) \cdot E(B)$$

ale nie jesteśmy w stanie policzyć $E(Z \cdot B)$. *W związku z tym podejdziemy do tematu inaczej:*

$$Var(B + Z) = Var(B) + Var(Z) + 2Cov(Z, B)$$

Potrzebujemy policzyć $Var(B)$, *wiemy, że pojedyczna kulka ma rozkład dwupunktowy:*

$$P(X_i = 1) = \frac{k}{N}$$

$$P(X_i = 0) = 1 - \frac{k}{N}$$

Dla pojedycznej kulki

$$E(X_i^2) = \frac{k}{N}$$

a wariancja:

$$Var(X_i) = \frac{k}{N} - \frac{k^2}{N^2} = \frac{k(N-k)}{N^2}$$

Jaka jest wobec tego wariancja

$$Var(X_1 + \ldots + X_n) =$$

$$= Var(X_1) + Var(X_2) + \ldots + Var X_n + n \cdot (n-1) Cov(X_i, X_j) =$$
$$= nVar(X_i) + n(n-1)Cov(X_i, X_j)$$

przy założeniu, że wszystkie wariancje i kowariancje pojedynczych zmiennych są takie same (bo są). Liczba kowariancji wynika z faktu: $\sum 2Cov(X_i, X_j) = 2 \cdot \binom{n}{2} = n(n-1)$ *Wariancje* X_i *mamy to teraz jaka jest kowariancja*

$$Cov(X_i, X_j) =?$$

Teraz:

$$P(X_i \cdot X_j = 1) = \frac{\binom{2}{2}\binom{N-2}{k-2}}{\binom{N}{k}} = \frac{k}{N} \cdot \frac{k-1}{N-1}$$

czyli wracając do przykładu z klasą, jeżeli w klasie jest Jaś i Małgosia i chcemy żeby oni zostali wylosowani, to interesują nas w liczniku wszystkie 9 osobowe drużyny, w których są Jaś i Małgosia a w mianowniku wszystkie możliwe 9 osobowe drużyny.

$$P(X_i \cdot X_j = 0) = 1 - \frac{k}{N} \cdot \frac{k-1}{N-1}$$

Stąd

$$E(X_i \cdot X_j) = 1 \cdot \frac{k}{N} \cdot \frac{k-1}{N-1}$$

$$Cov(X_i, X_j) = \frac{k}{N} \cdot \frac{k-1}{N-1} - \frac{k}{N} \cdot \frac{k}{N}$$

Mamy:

$$Var(B) = Var(X_{11} + \ldots + X_{20}) =$$

$$= 10 \cdot \frac{9(30-9)}{30^2} + 10(10-1) \cdot (9/30 \cdot 8/29 - 9^2/30^2) = \frac{42}{29} = Var(Z)$$

$$Var(B+Z) = Var(X_1 + \ldots + X_{20}) = 20 \cdot \frac{9(30-9)}{30^2} + 20(20-1)(9/30 \cdot 8/29 - 9^2/30^2) = \frac{42}{29}$$

$$Cov(Z, B) = \frac{Var(B_Z) - Var(B) - Var(Z)}{2} = -\frac{21}{29}$$

Można to zadanie również rozwiązać bardzo sprytnie, zauważając, że

$$Var(B + Z + C) = 0$$

$$Var(B + Z + C) = 3Var(B) + 3 \cdot 2 \cdot Cov(Z, B) = 0$$

stąd

$$Cov(Z, B) = -\frac{1}{2}Var(B)$$

Zadanie 9.3 Zakładając, że obserwacje $x_1, x_2, \ldots, x_{12}$ stanowią próbkę losową z rozkładu Pareto o gęstości

$$f_\theta(x) = \begin{cases} \frac{4^\theta \theta}{(4+x)^{\theta+1}} & \text{gdy } x > 0, \\ 0 & \text{gdy } x \leqslant 0 \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem, wyznaczono wartość estymatora największej wiarygodności parametru θ i otrzymano $\hat{\theta} = 1.5$. W próbce były dwie obserwacje o wartości 12, a pozostałe dziesięć obserwacji miało wartości mniejsze od 12. Okazało się, że w rzeczywistości zaobserwowane wartości stanowiły próbkę z uciętego rozkładu Pareto, czyli były realizacjami zmiennych losowych $X_i = \min\{Y_i, 12\}$ gdzie Y_i, $i = 1, 2, \ldots, 12$, są niezależnymi zmiennymi losowymi z rozkładu o gęstości f_θ. Wyznaczyć wartość estymatora największej wiarogodności parametru θ po uwzględnieniu modyfikacji założeń.

Rozwiązanie 9.3 *Funkcja wiarygodności:*

$$L = \prod_{i=1}^{12} \frac{4^\theta \theta}{(4 + x_i)^{\theta+1}} = \frac{4^{12\theta}\theta^{12}}{\prod_{i=1}^{12}(4 + x_i)^{\theta+1}}$$

logarytm z funkcji wiarygodności

$$\ln L = 12\theta \ln 4 + 12 \ln \theta - \sum_{i=1}^{n}(\theta + 1) \ln(4 + x_i)$$

Pochodna (maksymalizujemy funkcję wiarygodności):

$$\frac{\partial L}{\partial \theta} = 12 \ln 4 + \frac{12}{\theta} - \sum_{i=1}^{12} \ln(4 + x_i) = 0$$

stąd

$$\hat{\theta} = \frac{12}{\sum_{i=1}^{n} \ln(4 + x_i) - 12 \ln 4}$$

wiemy, z treści, że $\hat{\theta} = \frac{3}{2}$ *stąd:*

$$\frac{12}{\sum_{i=1}^{n} \ln(4 + x_i) - 12 \ln 4} = \frac{3}{2}$$

stąd

$$\frac{24}{3} + 12 \ln 4 = \sum_{i=1}^{12} \ln(4 + x_i)$$

Teraz korzystając z treści zadania wiemy, że w rzeczywistości mamy rozkład ucięty Pareto, to znaczy, że na przedziale $(0, 12)$ *istnieje gęstość, a w punkcie* $x = 12$ *mamy prawdopodobieństwo punktowe. Obliczmy wobec tego*

$$P(X = 12) = P(Y \geqslant 12) = \int_{12}^{\infty} \frac{4^\theta \theta}{(4+\theta)^{\theta+1}} dx = \frac{1}{4^\theta}$$

Obliczymy wobec tego funkcję wiarygodności:

$$L = \prod_{i=1}^{10} \frac{4^\theta \theta}{(4+x_i)^{\theta+1}} \cdot \prod_{i=1}^{2} P(X_i = 12) =$$

$$= \frac{4^{10\theta}\theta^{10}}{\prod_{i=1}^{10}(4+x_i)^{\theta+1}} \cdot \frac{1}{4^{2\theta}}$$

$$\ln L = 10\theta \ln 4 + 10 \ln \theta - (\theta+1) \sum_{i=1}^{12} \ln(4+x_i) - 2\theta \ln 4 = (*)$$

Podstawimy teraz

$$\frac{24}{3} + 12 \ln 4 = \sum_{i=1}^{12} \ln(4+x_i)$$

czyli

$$\frac{24}{3} + 12 \ln 4 = \sum_{i=1}^{10} \ln(4+x_i) + 2\ln(4+12)$$

$$\sum_{i=1}^{10} \ln(4+x_i) = \frac{24}{3} + 12 \ln 4 - 2 \ln 16$$

$$(*) = 10\theta \ln 4 + 10 \ln \theta - (\theta+1)\left(\frac{24}{3} + 12 \ln 4 - 2 \ln 16\right) - 2\theta \ln 4$$

Pochodna:

$$\frac{\partial L}{\partial \theta} = 10 \ln 4 + 10\frac{1}{\theta} - \left(\frac{24}{3} + 12 \ln 4 - 2 \ln 16\right) - 2 \ln 4 = 0$$

$$\theta = \frac{10}{\left(\frac{24}{3} + 12 \ln 4 - 2 \ln 16\right) - 10 \ln 4 + 2 \ln 4} = 1.25$$

Zadanie 9.4 Niech $X_1, X_2, \ldots, X_n, \ldots,\ I_1, I_2, \ldots, I_n, \ldots, N$ będą niezależnymi zmiennymi losowymi. Zmienne $X_1, X_2, \ldots, X_n, \ldots$ mają rozkład o wartości oczekiwanej 4 i wariancji 1. Zmienne $I_1, I_2, \ldots, I_n, \ldots$ mają rozkład jednostajny na przedziale $(0, 1)$. Zmienna N ma rozkład ujemny dwumianowy:

$$P(N = n) = \frac{\Gamma(2+n)}{n!} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^n$$

dla $n = 0, 1, 2, \ldots$.Niech

$$S_N = \begin{cases} 0 & \text{gdy } N = 0, \\ \sum_{i=1}^{n} I_i X_i & \text{gdy } N > 0 \end{cases}$$

Wtedy $Var(S_N)$

Rozwiązanie 9.4 ***Rozkład ujemny dwumianowy***

$$p_k = \frac{\Gamma(r+k)}{\Gamma(r)k!} p^r (1-p)^k$$

$$E(X) = \frac{r(1-p)}{p}$$

$$VAR(X) = \frac{r(1-p)}{p^2}$$

Mamy:

$$Var(S_N) = E(S_N^2) - (E(S_N))^2$$

$$E(S_N) = E(E(S_N|N))$$

$$E(S_N|N) = E\left(\sum_{i=1}^{N} I_i X_i\right) = N \cdot E(I_i)E(X_i) = N \cdot \frac{1}{2} \cdot 4 = 2N$$

$$E(2N) = ?$$

Z wzoru na rozkład dwumianowy mamy:

$$E(N) = \frac{2(1-3/4)}{3/4}$$

$$Var(N) = \frac{2(1-3/4)}{(3/4)^2}$$

Czyli

$$E(2N) = 2E(N) = 2\frac{2(1-3/4)}{3/4} = \frac{4}{3}$$

Dalej

$$E(S_N^2) = E(E(S_N^2|N))$$

$$E(S_N^2|N) = E\left(\left(\sum_{i=1}^{N} I_i X_i\right)^2\right) = NE(I_i^2 X_i^2) + N(N-1)E(I_i X_i I_j X_j) =$$

powyższe łatwo zauważyć z rozpisania sumy, mamy dalej:

$$= NE(I_i^2)E(X_i^2) + N(N-1)E(I_i)^2 E(X_i)^2 = (*)$$

Teraz: wariancja w rozkładzie jednostajnym to $(b-a)^2/12$, *w naszym przypadku to* $Var(I_i) = 1/12$, *skoro* $E(I_i) = 1/2$ *to* $E(I_i^2) = 1/12 + (1/2)^2 = 1/3$, *dla* X_i *mamy* $E(X^2) = 1 + 4^2 = 17$

$$(*) = \frac{17}{3}N + N(N-1)\frac{1}{4} \cdot 16$$

$$E\left(\frac{17}{3}N + N(N-1)4\right) =$$

$$= \frac{17}{3} \cdot E(N) + E(N^2) \cdot 4 - E(N) \cdot 4 = (*)$$

$$E(N^2) = Var(N) + (E(N))^2 = \frac{2(1-3/4)}{(3/4)^2} + \left(\frac{2(1-3/4)}{3/4}\right)^2 = \frac{4}{3}$$

$$E(N) = \frac{2(1-3/4)}{3/4} = \frac{2}{3}$$

$$(*) = \frac{58}{9}$$

$$Var(S_N) = \frac{58}{9} - (4/3)^2 = \frac{14}{3}$$

Zadanie 9.5 Niech $X_1, X_2, \ldots, X_n, \ldots$ będą niezależnymi zmiennymi losowymi o identycznym rozkładzie o gęstości

$$f(x) = \begin{cases} \frac{3}{x^4} & \text{gdy } x > 1, \\ 0 & \text{gdy } x \leqslant 1 \end{cases}$$

Niech $U_n = (X_1 \cdot X_2 \cdot \ldots \cdot X_N)^{\frac{1}{n}}$. Wtedy:

(a) $\lim_{n\to\infty} P\left((U_n - 1.5)\sqrt{n} > \frac{9}{4}\right) = 0.1587$

(b) $\lim_{n\to\infty} P\left((U_n - 1.5)\sqrt{n} > 3e^3\right) = 0.1587$

(c) $\lim_{n\to\infty} P\left((U_n - e^{1/3})\sqrt{n} > 3e^{-1/3}\right) = 0.1587$

(d) $\lim_{n\to\infty} P\left((U_n - e^{-1/3})\sqrt{n} > 3e^{1/3}\right) = 0.1587$

(e) $\lim_{n\to\infty} P\left(3(U_n - e^{1/3})\sqrt{n} > e^{1/3}\right) = 0.1587$

Rozwiązanie 9.5 *Widać, że mamy* U_n *gdzie jest iloczyn, żeby to zmienić na sumę weźmiemy logarytm z tej zmiennej*

$$\ln(U_n) = \frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

jaki rozkład ma $\ln X_i$*? Skorzystajmy ze wzorów na transformację:*

$$Y = lnX$$

$$X = e^Y \to h(y) = e^y$$

Wtedy gęstość ma postać:

$$g(y) = f(h(y)) \cdot |h'(y)|$$

$$g(y) = \frac{3}{e^{4y}} \cdot e^y = \frac{3}{e^{3y}} = 3 \cdot e^{-3y}$$

popatrzmy teraz czy gęstość się zgadza, $x > 1$ *to oznacza, że* y *zaczyna się od zera. Jest ok. Widać, że mamy rozkład Beta*

$$EX = \frac{1}{3}$$

$$VarX = \frac{1}{9}$$

Gdy n dąży do nieskończoności suma zmiennych losowych o dowolnym rozkładzie ma rozkład normalny.

Czyli $\sum lnX_i \sim N(1/3n, 1/9n)$*, a przed sumą mamy jeszcze* $1/n$

$$\ln U_n \sim N\left(\frac{1}{3}, \frac{1}{9n}\right)$$

Patrząc na odpowiedzi mamy:

$$P\left((U_n - a)\sqrt{n} > b\right) =$$

$$= P\left(U_n\sqrt{n} > b + a\sqrt{n}\right) =$$

$$= P\left(\ln U_n > \ln\left(\frac{b + a\sqrt{n}}{\sqrt{n}}\right)\right)$$

Znormalizujmy $\ln U_n$

$$\frac{\ln U_n - 1/3}{\frac{1}{3\sqrt{n}}} \sim N(0,1)$$

$$P\left(\frac{\ln U_n - 1/3}{\frac{1}{3\sqrt{n}}} > \frac{\ln\left(\frac{b+a\sqrt{n}}{\sqrt{n}}\right) - 1/3}{\frac{1}{3\sqrt{n}}}\right)$$

$$P(X > C) = 0.1587$$

z tablic rozkładu normalnego $P(X < C) = 0.8413$ *wynika, że C=1. Wszystko przy założeniu, że* $n \to \infty$

$$\frac{\ln\left(\frac{b+a\sqrt{n}}{\sqrt{n}}\right) - 1/3}{\frac{1}{3\sqrt{n}}} = 1$$

$$3\sqrt{n}\left(\ln\left(\frac{b + a\sqrt{n}}{\sqrt{n}}\right) - 1/3\right) = 1$$

$$\left(3\sqrt{n}\ln\left(\frac{b+a\sqrt{n}}{\sqrt{n}}\right)-\sqrt{n}\right)=1$$

$$\left(3\ln\left(\frac{b+a\sqrt{n}}{\sqrt{n}}\right)^{\sqrt{n}}-\sqrt{n}\right)=1$$

rozważmy

$$3\ln\left(\frac{b+a\sqrt{n}}{\sqrt{n}}\right)^{\sqrt{n}}=3\ln\left(a+\frac{b}{\sqrt{n}}\right)^{\sqrt{n}}=$$

przypominamy sobie, że $\lim_{n\to\infty}\left(1+\frac{k}{n}\right)^n=e^k$

$$=3\cdot\ln\left(a\left(1+\frac{b}{a\sqrt{n}}\right)\right)^{\sqrt{n}}=3\sqrt{n}\cdot\ln a+3\cdot\ln(e^{b/a})$$

mamy w granicy $n\to\infty$*:*

$$3\sqrt{n}\cdot\ln a+3\cdot\frac{b}{a}-\sqrt{n}=1$$

stąd dochodzimy do wniosku

$$a=e^{1/3}$$

stąd

$$3\cdot\frac{b}{e^{1/3}}=1\to b=\frac{1}{3}\cdot e^{\frac{1}{3}}$$

Odpowiedź E.

Zadanie 9.6 Niech $X_1,\ldots,X_{10}$ będą niezależnymi zmiennymi losowymi o identycznym rozkładzie normalnym $N(\mu,\sigma^2)$ z nienanymi parametrami $\mu\in R$ i $\sigma>0$. Budujemy przedział ufności dla parametru μ postaci:

$$[X_{3:10},X_{7:10}],$$

gdzie $X_{k:10}$ oznacza k-tą statystykę pozycyjną z próby $X_1,X_2,\ldots,X_{10}$. Wtedy prawdopodobieństwo

$$P_{\mu,\sigma}(\mu\in[X_{3:10},X_{7:10}])$$

jest równe: ?.

Rozwiązanie 9.6 *Tutaj mamy następującą sytuację, szukamy:*

$$P(X_{3:10}\leqslant\mu\leqslant_{7:10})$$

Z warunków zadania mamy:

$$X_{3:10}\leqslant\mu$$

Wartości sa z rozkładu normalnego więc nie ma prawdopodobieństwa, że liczba jest dokładnie równa μ. Dodatkowo rozkład jest symetryczny czyli

$$P(X_i<\mu)=1/2$$

dla każdego i. Warunek $X_{3:10} \leqslant \mu$ oznacza, że trzy najmniejsze zmienne musza być mniejsze od μ, warunek $X_{7:10}$ oznacza, że 4 największe zmienne muszą być większe od μ. Jeżeli ułożymy (jedno z możliwych losowań) zmienne:

$$\underbrace{X_1 < X_2 < X_3}_{X_3<\mu} < X_4 < X_5 < X_6 < \underbrace{X_7 < X_8 < X_9 < X_{10}}_{X_7>\mu}$$

Rozpatrzmy wobec tego przypadki kiedy próbka nie spełnia warunków, żeby być w przedziale ufności. Nie może być:

1. *10 zmiennych większych od μ, czyli p-p $(1/2)^{10}$, co oznacza, że wszystkie są mniejsze*
2. *9 zmiennych większych od μ, czyli p-p $\binom{10}{9} \cdot (1/2)^9 \cdot (1/2)^1$, co oznacza, że tylko jedna jest mniejsza*
3. *8 zmiennych większych od μ, czyli p-p $\binom{10}{8} \cdot (1/2)^8 \cdot (1/2)^2$, co oznacza, że tylko dwie są mniejsze*
4. *10 zmiennych mniejszych od μ, czyli p-p $(1/2)^{10}$*
5. *9 zmiennych mniejszych od μ, czyli p-p $\binom{10}{9} \cdot (1/2)^9 \cdot (1/2)^1$*
6. *8 zmiennych mniejszych od μ, czyli p-p $\binom{10}{8} \cdot (1/2)^8 \cdot (1/2)^2$*
7. *7 zmiennych mniejszych od μ, czyli p-p $\binom{10}{7} \cdot (1/2)^7 \cdot (1/2)^3$, co oznacza, że tylko 3 są większe od μ*

Suma powyższych prawodpodobieństw da nam prawdopodbieństwo $P(\notin [X_{3:10}, X_{7:10}])$. Mamy:

$$P(\notin [X_{3:10}, X_{7:10}]) = (1/2)^{10}(1 + 10 + 45 + 1 + 10 + 45 + 120) = \frac{29}{128}$$

$$ODP = 1 - \frac{29}{128} = \frac{99}{128}$$

Zadanie 9.7 Niech (X, Y) będzie dwuwymiarową zmienną losową z rozkładu o gęstości

$$p(x, y) = \begin{cases} 2x & \text{gdy } x \in (0,1) \wedge y \in (0,1), \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Niech $U = X + Y$ i $V = X - Y$. Wtedy $E\left(U|V = \frac{1}{3}\right)$ jest równa ?

Rozwiązanie 9.7 *Mamy:*

$$E\left(X + Y|X - Y = \frac{1}{3}\right) =$$

$$= E\left(X + Y|Y = X - \frac{1}{3}\right) =$$

$$= E\left(2X - \frac{1}{3}|Y = X - \frac{1}{3}\right)$$

Rysujemy rysunek pomocniczy i uwzględniamy, że $x \in (0,1) \wedge y \in (0,1)$. Liczymy z definicji:

$$E(2X - 1/3|Y = X - 1/3) = \frac{\int_{1/3}^{1}(2x-1)\cdot 2x\cdot dx}{\int_{1/3}^{1} 2x\cdot dx} = \frac{10}{9}$$

Zadanie 9.8 Niech $X_1, X_2, \ldots, X_{10}$ bedą niezależnymi zmiennymi losowymi o identycznym rozkładzie Weibulla o gęstości

$$f_\theta(x) = \begin{cases} 2\theta x \exp(-\theta x^2) & \text{gdy } x \in (0, +\infty) \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Testem jednostajnie najmocniejszym na poziomie istotności 0.05 weryfikujemy hipotezę $H_0 : \theta \leqslant 2$ przy alternatywie $H_1 : \theta > 2$. Niech $F_k(x)$ oznacza dystrybuantę rozkładu chi-kwadrat z k stopniami swobody w punkcie x.
Moc tego testu przy alternatywie $H_1 : \theta = 6$ jest równa ?.

Rozwiązanie 9.8 ***Moc testu statystycznego rozumiana jest jako prawdopodobieństwo nieodrzucenia*** H_1 ***przy jej prawdziwości (odrzucenie*** H_0 ***przy jej fałszywości).***

Test jednostajnie najmocniejszy tworzymy poprzez zbudowanie ilorazu wiarygodności. Gdy:

$$\frac{L_1}{L_0} > c$$

to odrzucamy H_0. Mamy:

$$L_1 = \prod_{i=1}^{10} 2\theta_1 e^{-\theta_1 x_i^2} = 2^{10}\theta_1^{10} e^{-\sum_{i=1}^{10}\theta_1 x_i^2}$$

$$L_0 = \prod_{i=1}^{10} 2\theta_0 e^{-\theta_0 x_i^2} = 2^{10}\theta_0^{10} e^{-\sum_{i=1}^{10}\theta_0 x_i^2}$$

mamy:

$$\frac{L_1}{L_0} = \frac{\theta_1^{10}}{\theta_0^{10}} e^{(\theta_0-\theta_1)\sum_{i=1}^{10} x_i^2} =$$

$$\frac{\theta_1^{10}}{\theta_0^{10}} \exp\left((\theta_0 - \theta_1)\cdot \sum_{i=1}^{10} x_i^2\right) > c$$

stąd możemy podzielić przez $\theta_1^{10}/\theta_0^{10}$ a następnie wziąć logarytm i mamy:

$$(\theta_0 - \theta_1)\cdot \sum_{i=1}^{10} x_i^2 > c$$

wiemy z treści, że $\theta_0 < \theta_1$ czyli po podzieleniu przez $(\theta_0 - \theta_1)$ odwracamy znak nierówności:

$$\sum_{i=1}^{10} x_i^2 < c$$

Jaki rozkład ma suma kwadratów zmiennych o rozkładzie Weibulla? Wymyślmy najpierw jaki rozkład ma x_i^2 Skorzystamy ze wzorów na transformację:

$$Y = X^2 \rightarrow x = \sqrt{Y}$$

gęstość Y:

$$g(y) = f(h(y))|h'(y)|$$

$$h(y) = \sqrt{y}$$

mamy:

$$h'(y) = \frac{1}{2\sqrt{y}}$$

$$g(y) = 2\theta \cdot \sqrt{y} \cdot e^{-\theta x} \cdot \frac{1}{2\sqrt{y}} = \theta \cdot e^{-\theta x}$$

czyli

$$g(y) = \theta \cdot e^{-\theta x}$$

co oznacza, że X_i^2 ma rozkład wykładniczy z $\beta = \theta$, suma zmiennych o rozkładzie wykładniczym daje rozkład Gamma:

$$\sum_{i=1}^{10} X_i^2 \sim Gamma(10, \theta)$$

Więc chcemy wyznaczyć c gdy:

$$P\left(\sum_{i=1}^{10} X_i^2 > c\right) = 0.05$$

oczywiście nie mam tablic gamma, więc muszę przejść przez rozkład χ^2. Gdy $\alpha = \frac{n}{2}$ oraz $\beta = \frac{1}{2}$ to mamy rozkład ch-kwadrat z n stopniami swobody. Niech $\sum_{i=1}^{10} X_i^2 = Z$

$$2\theta \cdot Z \sim Gamma\left(10, \frac{\theta}{2\theta}\right) = \chi^2(20)$$

bo gdy $X \sim Gamma(a,b) \rightarrow kX \sim Gamma(a, \frac{b}{k})$

$$P(2 \cdot \theta \cdot Z < 2\theta c) = 0.05$$

z tablic

$$2 \cdot \theta \cdot c = 10.8508$$

gdy podstawimy $\theta = 2$ (z H_0) to mamy $2 \cdot 2 \cdot c = 10.8508$ Teraz gdy podstawimy $\theta = 6$ mamy:

$$2 \cdot 6 \cdot c = 10.8508 \cdot 3 = 32.5524$$

czyli prawdopodobieństwo odrzucenia H_0 przy jej fałszywości to:

$$F_{20}(32.5524)$$

Zadanie 9.9 Załóżmy, że $X_1, X_2, \ldots, X_n, \ldots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie jednostajnym na przedziale $[0;1]$, zaś N jest mzienną losową o rozkładzie Poissona z wartością oczekiwaną 2, niezależną od zmiennych losowych $X_1, X_2, \ldots, X_n, \ldots$. Niech:

$$Y_N = \begin{cases} \min\{X_1, X_2, \ldots, X_N\} & \text{gdy } N > 0, \\ 0 & \text{gdy} N = 0 \end{cases}$$

$$Z_N = \begin{cases} \max\{X_1, X_2, \ldots, X_N\} & \text{gdy } N > 0, \\ 0 & \text{gdy} N = 0 \end{cases}$$

Obliczyć $E(Z_N - Y_N)$

Rozwiązanie 9.9 *Czyli mamy obliczyć:*

$$E(Z_N - Y_N) = E(Z_N) - E(Y_N)$$

Najpierw policzymy

$$E(Z_N) = E(E(Z_N|N))$$

jaki rozkład ma maksimum? Własności maksimum i rozkładu jednostajnego:

$$P(\max\{X_1, \ldots, X_N\} < t) = P(X_i < t)^N = t^N$$

$$f_{max}(x) = N \cdot t^{N-1}$$

Przy okazji rozkład minimum:

$$P(\min\{X_1, \ldots, X_N\} < t) = 1 - P(\min\{X_1, \ldots, X_N\} > t) =$$

$$= 1 - P(X_i > t)^N = 1 - (1-t)^N$$

$$f_{min}(x) = N \cdot (1-t)^{N-1}$$

Wobec tego:

$$E(Z_N|N) = \int_0^1 t \cdot N \cdot t^{N-1} dt = \frac{N}{N+1}$$

tu należy być ostrożny z przypadkiem co się dzieje gdy $N = 0$ widzimy, że Z_N przyjjmuje wartość zero więc jest ok. Z rozkładu poissona $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$$E\left(\frac{N}{N+1}\right) = \sum_{n=0}^{\infty} \frac{n}{n+1} \frac{2^n e^{-2}}{n!} =$$

mały trick:

$$= \sum_{n=0}^{\infty} \frac{n+1-1}{n+1} \frac{2^n e^{-2}}{n!} = \sum_{n=0}^{\infty} \frac{n+1}{n+1} \frac{2^n e^{-2}}{n!} - \sum_{n=0}^{\infty} \frac{1}{n+1} \frac{2^n e^{-2}}{n!} =$$

$$= 1 - \sum_{n=0}^{\infty} \frac{1}{n+1} \frac{2^n e^{-2}}{n!} = (*)$$

mamy:

$$\sum_{n=0}^{\infty} \frac{1}{n+1} \frac{2^n e^{-2}}{n!} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{2^{n+1} e^{-2}}{(n+1)!} = \frac{1}{2} \sum_{k=1}^{\infty} \frac{2^k e^{-2}}{k!} = \frac{1}{2}(1 - e^{-2})$$

stąd

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

Teraz

$$E(Y_N|N) = \int_0^1 t \cdot N \cdot (1-t)^{N-1} dt = \frac{1}{N+1}$$

*i tu **mamy problem z zerem** bo gdy $N = 0$ powinno być $E(Y_0) = 0$. Powyższe rozwiązanie całki jest prawdłowe gdy $N > 0$. Stąd mamy:*

$$E(Y_N|N) = \underbrace{E(E(Y_N|N)|N=0)}_{=0} \cdot P(N=0) + E(E(Y_N|N)|N>0) \cdot P(N>0) =$$

$$= 0 + \sum_{n=1}^{\infty} \frac{1}{n+1} \cdot \frac{2^n e^{-2}}{n!} = (*)$$

Wiemy, że od zera:

$$\sum_{n=0}^{\infty} \frac{1}{n+1} \frac{2^n e^{-2}}{n!} = \frac{1}{2}(1 - e^{-2})$$

Wobec tego jak od wyniku odejmiemy pierwszy wyraz to mamy to co chcemy:

$$\sum_{n=1}^{\infty} \frac{1}{n+1} \cdot \frac{2^n e^{-2}}{n!} = \frac{1}{2}(1 - e^{-2}) - e^{-2}$$

$$(*) = \frac{1}{2}(1 - e^{-2}) - e^{-2}$$

Czyli mamy:

$$E(Z_N - Y_N) = E(Z_N) - E(Y_N) = \frac{1}{2} + \frac{1}{2} e^{-2} - \frac{1}{2}(1 - e^{-2}) + e^{-2} = 2e^{-2}$$