

Kajecik Fazbera na Metody Probabilistyczne

Spis treści

1 Pytania o wszystkim i o niczym (3)	7.4 Prognoza ataków dnia zerowego (16)
1.1 Probabilistyka - cele, umiejscowienie (3)	7.5 Przedziały ufności - wyznaczanie (17)
1.2 Mechanizmy losowe i pseudolosowe (3)	7.6 Robak sieciowy (17)
1.3 Mity a fakty (3)	8 Testowanie hipotez (17)
2 Rozkłady prawdopodobieństwa (4)	8.1 Policer SLA (17)
2.1 iTrust (4)	8.2 Test istotności (17)
2.2 Symulator w hiperkuli (4)	
2.3 Kod korekcyjny w przestrzeni Hamminga (4)	
2.4 Sieć sensorów (5)	
2.5 Wielodostęp rywalizacyjny (5)	
2.6 Filtr Blooma (5)	
2.7 Fuzja danych (5)	
2.8 Producenty i konsumenci (6)	
2.9 Rozproszony wybór lidera (6)	
2.10 Sieć społecznościowa (6)	
2.11 Szybkie sortowanie (6)	
2.12 Odtwarzanie tras pakietów IP (7)	
2.13 Optymalizacja polityki terminacji (7)	
2.14 Testowanie oprogramowania (7)	
2.15 Bizantyjscy generalowie (7)	
2.16 System reputacyjny Beta (8)	
2.17 Generacja liczb pseudolosowych (9)	
2.17 Metoda odwracania dystrybucyj (9)	
2.18 Naturalne zbiory danych (9)	
3 Sumy niezależnych wielkości losowych (9)	
3.1 Funkcje charakterystyczne i tworzące (9)	
3.1 Rozkłady sum (9)	
3.2 Równoważenie obciążenia: round robin (9)	
3.3 Zaległości w obsłudze (10)	
3.4 Utracone wywołania (10)	
3.5 Opróżnianie stosu (10)	
3.6 Słabe prawo wielkich liczb: eksperyment (11)	
3.7 Centralne twierdzenie graniczne (12)	
3.8 Serwer webowy (12)	
3.9 Eksperyment Wolfa (12)	
3.10 Czas trwania sesji (12)	
3.11 Atak na Wi-Fi (13)	
4 Ogony rozkładów prawdopodobieństwa (13)	
4.1 Częstość anomalii w sygnale (13)	
5 Łańcuchy Markowa (14)	
5.1 Jednorodny łańcuch Markowa I rzędu (14)	
5.2 Sztuczne źródła tekstowe (14)	
5.3 Ergodyczność łańcucha Markowa (14)	
5.4 PageRank (14)	
5.5 Rozpoznawanie mowy (15)	
6 Wektory losowe i korelacja (15)	
6.1 Korelacja: wariancja $X+Y$ (15)	
6.2 Korelacja, a zależność statystyczna (15)	
6.3 Regresja liniowa (15)	
6.4 Stratna kompresja obrazu (16)	
7 Estymatory (16)	
7.1 Próbkowanie losowe populacji (16)	
7.2 Estymatory maksimum wiarygodności (16)	
7.3 Monte Carlo: redukcja wariancji (16)	

Indeks pojęć

Atak na Wi-Fi (13)
Bizantyjscy generałowie (7)
Centralne twierdzenie graniczne (12)
Czas trwania sesji (12)
Częstość anomalii w sygnale (13)
Eksperyment Wolfa (12)
Ergodyczność łańcucha Markowa (14)
Estymatory maksimum wiarygodności (16)
Filtr Blooma (5)
Funkcje charakterystyczne i tworzące (9)
Fuzja danych (5)
Generacja liczb pseudolosowych (9)
iTrust (4)
Jednorodny łańcuch Markowa I rzędu (14)
Kod korekcyjny w przestrzeni Hamminga (4)
Korelacja, a zależność statystyczna (15)
Korelacja: wariancja $X+Y$ (15)
Mechanizmy losowe i pseudolosowe (3)
Metoda odwracania dystrybucyj (9)
Mity a fakty (3)
Monte Carlo: redukcja wariancji (16)
Naturalne zbiory danych (9)
Odtwarzanie tras pakietów IP (7)
Opróżnianie stosu (10)
Optymalizacja polityki terminacji (7)
PageRank (14)
Policer SLA (17)
Probabilistyka - cele, umiejscowienie (3)
Producenci i konsumenci (6)
Prognoza ataków dnia zerowego (16)
Przedziały ufności - wyznaczanie (17)
Próbka losowa populacji (16)
Regresja liniowa (15)
Robak sieciowy (17)
Rozkłady sum (9)
Rozpoznawanie mowy (15)
Rozproszony wybór lidera (6)
Równoważenie obciążenia: round robin (9)
Serwer webowy (12)
Sieć sensorów (5)
Sieć społecznościowa (6)
Stratna kompresja obrazu (16)
Symulator w hiperkuli (4)
System reputacyjny Beta (8)
Sztuczne źródła tekstowe (14)
Szybkie sortowanie (6)
Słabe prawo wielkich liczb: eksperyment (11)
Test istotności (17)
Testowanie oprogramowania (7)
Utracone wywołania (10)
Wielodostęp rywalizacyjny (5)
Zaległości w obsłudze (10)

1 Pytania o wszystkim i o niczym

1.1 Probabilistyka - cele, umiejscowienie

1.1.1 Wyjaśnij rolę probabilistyki jako “interfejsu”.

Probabilistyka jest interfejsem między procesami poznawczymi i decyzyjnymi w obecności mechanizmów losowych a matematyką. Pozwala dokonywać ocen ilościowych i jakościowych oraz ułatwia podejmowanie racjonalnych decyzji. Jako probabilistycy, jesteśmy klientami matematyki, a np. informatyka jest naszym klientem.

1.2 Mechanizmy losowe i pseudolosowe

1.2.1 Wyjaśnij obiektywny i subiektywny walor pojęcia prawdopodobieństwa.

Obiektywny - opisuje część danego zdarzenia przy wielokrotnym uruchamianiu mechanizmu losowego w doświadczeniu losowym. Subiektywny - opisuje stan wiedzy o danym mechanizmie losowym.

1.2.2 Jakie są elementy opisu modelu probabilistycznego doświadczenia losowego, czyli przestrzeni probabilistycznej?

Zbiór zdarzeń elementarnych oraz zdarzeń zdefiniowanych na podstawie zdarzeń elementarnych i ich prawdopodobieństwa.

1.2.3 Wymień podstawowe cechy dobrego generatora liczb pseudolosowych.

Każdą liczbę zwraca niemal tak samo często i bez uchwytynego związku z uprzednimi w bardzo długim cyklu.

1.2.4 Podaj znane Ci definicje prawdopodobieństwa.

- Klasyczna $P(A) = \frac{|A|}{|\omega|}$
- Geometryczna $P(A) = \frac{||A||}{||\omega||}$
- Aksjomatyczna
- Częstościowa

1.3 Mity a fakty

1.3.1 Czy i ewentualnie kiedy na wykładzie obaliliśmy mit, że jeśli w każdym pokoleniu średnio przypada jeden potomek na osobnika, to populacja ma szansę przetrwać?

Populacja ma szansę przetrwać w badanym skończonym przedziale czasowym, lecz nie w nieskończoności. Obaliliśmy ten mit przy okazji problemu opróżnienia stosu.

1.3.2 Jaki rozkład prawdopodobieństwa mają pierwsze cyfry znaczące zgodnie z prawem Benforda?

Rozkład logarytmiczny.

1.3.3 Jak z probabilistycznego punktu widzenia zaklasyfikujemy ciąg kolejnych przewag orłów nad reszkami w serii rzutów monetą?

Ciąg przewag orłów nad reszkami w serii rzutów monetą można analizować jako losową wędrówkę, gdzie każdy rzut to krok w górę dla orła (+1) lub w dół dla reszki (-1). Prawdopodobieństwo, że przewaga orłów utrzyma się przez n rzutów, wynosi około $\frac{1}{2^n}$. Średnia długość najdłuższego ciągu przewag orłów rośnie proporcjonalnie do pierwiastka kwadratowego liczby rzutów, czyli $\log_2(n)$.

1.3.4 Na podstawie rysunku zamieszczonego na slajdzie podaj probabilistyczną przyczynę, dla której “statystyczny pasażer” poczeka na przystanku dłużej niż 5 minut.

Jeśli przychodzą trzy autobusy po sobie, pierwszy, potem drugi z czasem X po pierwszym (mniejszym od średniej) a potem trzeci z czasem Y po drugim (większym od średniej) to częściej statystyczny pasażer będzie trafiał w przedział Y , czyli będzie czekał dłużej od wartości średniej.

1.3.5 W kontekście “prześladowania przez pecha” co dokładnie rozumiemy przez “symetryczną sytuację losową”?

Prześladowanie przez pecha to złudzenie. Spośród symetrycznych doświadczeń losowych, mocniej zapamiętujemy te, które są na naszą niekorzyść. Przykładowo, jeżeli codziennie na przystanek przyjeżdża autobus, to zapamiętamy tylko te sytuacje, kiedy nie przyjedzie na czas. Czyli symetryczna sytuacja losowa to powtórzone doświadczenie pod tymi samymi warunkami.

1.3.6 W teście wyboru z 5 wariantami odpowiedzi za poprawną odpowiedź przyznawany jest 1 punkt, za niepoprawną $-\frac{1}{4}$ punktu. Wyjaśnij, w jakim sensie neutralizuje to efekt przypadkowego wyboru odpowiedzi.

Ponieważ przewidywany zwrot jest równy zero. Inaczej - średnio strzelając uzyskamy 0 punktów.

2 Rozkłady prawdopodobieństwa

2.1 iTrust

2.1.1 Gdy liczby agentów objętych dystrybucją zapytań (z) i metadanych (m) są równe $4\sqrt{n}$ prawdopodobieństwo braku trafień oszacujemy jako...

$$P(T = 0) < e^{-\frac{mz}{n}} = e^{-\frac{4\sqrt{n} \cdot 4\sqrt{n}}{n}} = e^{-16}$$

2.2 Symulator w hiperkuli

2.2.1 Wydajność generatora dopuszczalnych układów odchyłek dla symulatora systemu zależy w następujący sposób od dopuszczalnego promienia hiperkuli.

Nie zależy (na wzorze można pokazać). Zależy jedynie od ilości wymiarów, w których jest określona.

2.2.2 W jaki sposób zastosowana została geometryczna definicja prawdopodobieństwa?

Prawdopodobieństwo jest równe stosunkowi miary obszaru hiperkuli do miary obszaru hiperkostki.

2.3 Kod korekcyjny w przestrzeni Hamminga

2.3.1 Dlaczego w użytym przykładzie tabeli dekodowania żaden korygowany ciąg błędów nie może zawierać trzech jedynek?

Bo dla trzech jedynek mamy do czynienia z fatalnym błędem decyzyjnym. Moc korekcyjna naszego kodu sięga niektórych dwóch błędów

2.3.2 Ciągiem nadanym jest 10000, ciągiem błędów jest 01001. Jaką decyzję podejmie dekodery?

Odbierzemy 11001, czyli koder podejmie decyzję 01001, czyli będzie to błąd fatalny.

$10000 \rightarrow$ ciąg nadany S , $01001 \rightarrow$ ciąg błędów Z
 $11001 \rightarrow$ ciąg odebrany $Y = S \oplus Z$, wybieramy S_d , gdzie najmniej jedynek w $S_d \oplus Y$

S_d	10111	10000	01110	01001
$S_d \oplus Y$	01110 (3)	01101 (2)	10111 (4)	10000 (1)

2.3.3 Wyjaśnij rolę zasady maksimum wiarygodności w konstrukcji kodu.

Na podstawie zasady maksimum wiarygodności wyprowadzamy metrykę Hamminga i wybieramy najbardziej wiarygodny ciąg bitów.

2.3.4 Jaka jest zasada konstrukcji tabeli dla dekodowania ciągów odebranych i na czym się ona opiera?

Pierwszy wiersz zawiera możliwe ciągi kodowe s . Pierwsza kolumna zawiera korygowalne ciągi błędów z . Pozostałe komórki zawierają odebrane sygnały y . Opiera się ona na zasadzie maksimum wiarygodności (sygnał y jest interpretowany jako najbardziej prawdopodobny ciąg s). Prawdopodobieństwo określane na podstawie liczby przekłamań (metryka Hamminga - im mniejsza tym bardziej prawdopodobny).

2.4.1 Co oznacza prawdopodobieństwo $P(Z_{50+} \geq 2)$?

Prawdopodobieństwo, że po czasie 2 w sieci będzie żyło ponad 50 sensorów.

2.4.2 Ile według przyjętego modelu wynosi wartość średnia czasu życia jednego sensora?

Całka z rozkładu życia sensora, czyli mianownik w wykładniku potęgi, czyli u nas: 1 Rozkład życia $P(X_i \geq x) = e^{-x}$ z tego komplementarna $F(X) = 1 - e^{-x}$ z tego $p(x)$ i całka $x p(x)$ przez części ale tego nie trzeba pewnie liczyć w ogóle.

$$\int_0^{\infty} x e^{-x} dx = 1$$

2.4.3 Jaki według przyjętego modelu jest rozkład czasu życia sensora, który przetrwał najdłużej spośród 100 sensorów?

$$P(Z_0 \geq x) = 1 - (1 - e^{-x})^{100}$$

2.5 Wielodostęp rywalizacyjny

2.5.1 Jakie byłoby prawdopodobieństwo $P(X = k)$ gdyby A były niezależne?

Takie samo, bo moglibyśmy je policzyć również z twierdzenia o kombinacji zdarzeń.

2.5.2 Jak wyglądałby rozkład $P(X = k), k = 0, 1, \dots, n$ gdyby zdarzenia A_i były rozłączne?

Gdy są nierozłączne może być kilka szczelin prostych naraz. Tylko jedno puste może zajść $k = 0$ lub $k = 1$.

2.5.3 Co jest większe: $|X = k|$ czy $\binom{n}{k}(n - k)^m$?

$\binom{n}{k}(n - k)^m$, bo doliczamy jeszcze, gdy $k + 1, k + 2, \dots, n$ szczelin pustych. Jest to uogólnienie wzoru Bernoulliego.

2.5.4 Uzasadnij, że zdarzenia A_1, \dots, A_n , gdzie A_i - szczelina i jest pusta, nie są rozłączne i nie są niezależne.

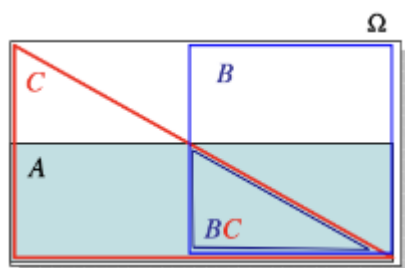
Nie są rozłączne, ponieważ może być kilka pustych szczelin naraz. Nie są niezależne, bo pustość jednej szczeliny zmniejsza prawdopodobieństwo pustości innej.

2.6 Filtr Blooma

2.6.1 Dokładne wyliczenie prawdopodobieństwa fałszywego alarmu jest trudne z uwagi na...

Fakt, że "zapalanie" kolejnych bitów nie jest niezależne statystycznie. Można liczyć z twierdzenia o kombinacji zdarzeń, ale jest to uciążliwe ze względu na naprzemienną sumę.

2.7 Fuzja danych

2.7.1 Wyjaśnij naiwność klasyfikatora

$$P(A | B) = P(A) = 1/2 \quad \checkmark$$

$$P(A | BC) = 1 \neq P(A | C) = 3/4 \quad \times$$

Niezależność zdarzeń nie implikuje niezależności tych zdarzeń pod wystąpieniem jakiegoś warunku jak ilustruje rysunek.

2.7.2 Podaj przykład relacji między zdarzeniami B i C zapewniającej warunkową niezależność statystyczną zdarzeń A i B na przykładowym wykresie Venna.

Powyższy rysunek: Wystąpienie/niewystąpienie Z może wprowadzać statystyczną zależność pomiędzy sygnałami detektorów. Niemniej warunkowa niezależność sygnałów często jest zakładana.

2.7.3 Dlaczego bezpośrednio zastosowanie zasady maksimum wiarygodności nazwaliśmy naiwnym?

Niezależne zdarzenia A i B nie muszą pozostać takimi przy warunku C - niezależność statystyczna na ogół nie implikuje warunkowej niezależności.

2.8 Producenty i konsumenci

2.8.1 Niech $\psi_{j,k}$ reprezentuje prawdopodobieństwo ostatecznego opróżnienia RK. Ułóż odpowiednią zależność rekurencyjną.

$$\begin{cases} \psi_{j,k} = p * \psi_{j+1,k} + q * \psi_{j,k-1} \\ \psi_{j,0} = 1 \\ \psi_{100,k} = 0 \end{cases} \quad \begin{cases} \Pi_{j,k} = p * \Pi_{j+1,k} + q * \Pi_{j,k-1} \\ \Pi_{100,k} = 1 \\ \Pi_{j,0} = 0 \end{cases}$$

ψ - prawdopodobieństwo opróżnienia RK

Π - prawdopodobieństwo przepełnienia RP

2.8.2 Co jest większe: $\Pi_{j,k}$ czy $\Pi_{j+1,k-1}$?

Większe jest $\Pi_{j+1,k-1}$, ponieważ jest większe prawdopodobieństwo kroku w stronę zwiększenia j a bufor konsumenta jest bliżej od opróżnienia.

Większe j - bliżej przepełnienia buforu producenta, mniejsze k - bliżej opróżnienia bufora konsumenta.

Prawdopodobieństwo rośnie.

$$\Pi_{j,k} < \Pi_{j+1,k-1} \quad \Pi_{j,k} > \Pi_{j-1,k+1}$$

2.8.3 Czym od przedstawionego różniłoby się rozumowanie prowadzące do prawdopodobieństwa opróżnienia RK?

Metoda ta różniłaby się warunkami granicznymi (zamiast $\psi_{j,0} = 1$ i $\psi_{100,k} = 0$ mamy $\Pi_{100,k} = 1$ i $\Pi_{j,0} = 0$).

2.8.4 W omawianym dwuwymiarowym błędzeniu przypadkowym występują dwa ekrany. Jakiego rodzaju?

Pierwszy ekran - przepełnienie RP $101k + 100$, drugi ekran opróżnienie RK $101 * 100 + j$.

2.9 Rozproszony wybór lidera

2.9.1 Jaki sens ma wartość l (małe L) w wyrażeniu na $P(T \leq k)$?

Jest to liczba wylosowana przez lidera, czyli ciąg orłów i reszek zamieniony na liczbę. Może ona być z przedziału $[0, 2^k - 1]$ i służy nam w twierdzeniu o prawdopodobieństwie całkowitym.

2.10 Sieć społecznościowa

2.10.1 Obliczając dynamikę $W(t)$ trzy razy korzystaliśmy z twierdzenia o prawdopodobieństwie całkowitym. Z jakich układów hipotez korzystaliśmy?

1. Zajście lub niezajście błędu implementacji lub interpretacji.
2. Spotkanie lub brak spotkania w ciągu dt .
3. Reputacje uczestnika i partnera, korzystając z $W(t)$ do obliczenia prawdopodobieństw hipotez.

2.10.2 Obliczając h_{xy} stosujemy twierdzenie o prawdopodobieństwie całkowitym. Co jest tutaj układem zupełnym zdarzeń?

Wystąpienie/nie wystąpienie błędu implementacji/interpretacji

2.10.3 Opisz układy zupełne zdarzeń wykorzystywane w rozumowaniu prowadzącym do określenia dynamiki procesu $W(t)$.

Na odcinku czasu wystąpi lub nie wystąpi interakcja. Jeśli wystąpi interakcja, to układ zupełny z 4 hipotezami \rightarrow losowy uczestnik i losowy partner mogą mieć niską lub wysoką reputację

2.11 Szybkie sortowanie

2.11.1 Wyjaśnij sens równania $ek|k = n - 1 + ek + en - 1 - k$.

Równanie rekurencyjne liczby potrzebnych porównań dla przypadku w którym wylosowana liczba jest większa od k liczb w sortowanym zbiorze. W tym wywołaniu wykonujemy $n - 1$ porównań (pivoting), oraz ilość porównań potrzebną do posortowania zbiorów o rozmiarze k oraz o rozmiarze $n - 1 - k$.

2.12 Odtwarzanie tras pakietów IP

2.12.1 Dlaczego słupki rozkładu odciętego są wyższe niż nieodciętego?

Ze względu na to, że masa prawdopodobieństwa przeniesiona została z ogona do tego przedziału, do którego obcięliśmy model. Jest to reprezentowane przez warunek normalizacyjny.

2.12.2 Jaki sens mają wielkości losowe X_1, \dots, X_d w problemie zbieracza kuponów?

Są to czasy czekania natrafienie na kolejny kupon (kolor), którego jeszcze nie mamy.

2.13 Optymalizacja polityki terminacji

2.13.1 Dlaczego dla rozważanej klasy polityk terminacji wystarczająca jest skończona tablica STOP/KONT?

Od pewnego momentu w wierszach i kolumnach tablicy są same wartości STOP.

2.13.2 Dlaczego nie jest poprawna polityka terminacji “znajdź pozycję w ciągu prób, dla której a jest maksymalne i podejmij wówczas decyzję STOP”?

Bo tę decyzję musimy podjąć na bieżąco, a nie po poznaniu całego ciągu. Jeżeli moglibyśmy podjąć ją po poznaniu całego ciągu, to zawsze bylibyśmy w stanie podjąć decyzję optymalną.

2.13.3 Wyjaśnij probabilistyczny sens równania $e(o, r) = \frac{1}{2}e(o+1, r) = \frac{1}{2}e(o, r+1)$.

Równanie to można interpretować probabilistycznie jako przewidywanie prawdopodobieństwa wystąpienia zdarzenia e w punkcie (o, r) na podstawie prawdopodobieństw wystąpienia tego samego zdarzenia w sąsiednich punktach: $(o+1, r)$ i $(o, r+1)$.

2.14 Testowanie oprogramowania

2.14.1 Co reprezentują wielkości losowe B i G ?

B to liczba niepoprawnych modułów jeśliby zakończyć testowanie na etapie i . G to liczba przetestowanych modułów do następnego etapu, czyli do natrafienia na niepoprawny. Jest to koszt przejścia z danego etapu do następnego jeśli wybrano kontynuację.

2.14.2 Wyjaśnij słownie w terminach wartości oczekiwanych optymalną regułę decyzji KONT/STOP w kolejnych etapach.

Wybieramy tą decyzję, która ma mniejszy koszt. Kontynuacja to koszt wykonania testów. Stop to ryzyko awarii krytycznej spowodowanej błędnym modulem

2.14.3 W którym momencie zastosowaliśmy uogólnione pojęcie wartości średniej?

Liczenie kosztu kontynuacji: $Eh\left(\frac{B_{n-y,p}}{h}\right)$

2.15 Bizantyjscy generałowie

2.15.1 Jakiej informacji nie wykorzystuje zasada maksimum wiarygodności w porównaniu do twierdzenia Bayesa?

Nie korzysta ona z informacji o prawdopodobieństwie zostania bizantyjskim generałem przez agenta. Nie zakłada, że dzieje się to w sposób niezależny.

2.15.2 Jakie różnice między wynikami dla twierdzenia Bayesa i zasady maksimum wiarygodności zaobserwowaliśmy przyjmując, że liczba agentów typu BG może być dowolna od 0 do 100 z jednakowym prawdopodobieństwem?

Wówczas wyniki byłyby do siebie proporcjonalne, nie obserwowalibyśmy różnic w kształtach tzn. $P\left(\frac{H_i}{R} = k\right) \sim P(R = k|H_i)$ (przeskalowane o współczynnik normalizacji).

2.15.3 Jak mierzyć niepewność uzyskanych wyników? Dlaczego w przypadku zasady maksimum wiarygodności jest ona większa w porównaniu z twierdzeniem Bayesa?

Niepewność uzyskanych wyników możemy mierzyć przez rozproszenie słupków histogramu od maksimum. Dla zasady maksimum wiarygodności nie wykorzystujemy znajomości o prawdopodobieństwie zostania przez agenta bizantyjskim generałem, dlatego nie jest faworyzowana mniejsza ich ilość dla mniejszej liczby niepoprawnych odpowiedzi, co mogłoby stanowić czynnik przyczyniający się do zredukowania niepewności.

2.15.4 Zastosowanie zasady maksimum wiarygodności zamiast twierdzenia Bayesa może być motywowane tym, że...

Dla zasady maksimum wiarygodności nie wykorzystujemy znajomości o prawdopodobieństwie zostania przez agenta bizantyjskim generałem, dlatego nie jest faworyzowana mniejsza ich ilość dla mniejszej liczby niepoprawnych odpowiedzi, co mogłoby stanowić czynnik przyczyniający się do zredukowania niepewności.

2.16 System reputacyjny Beta

2.16.1 Jak na jakość eliminacji nieuczciwych klientów wpłynęłoby zastąpienie parametrów kwantyl 0.95 i kwantyl 0.05 przez kwantyl 0.9 i kwantyl 0.1?

Więcej klientów zostałoby uznanych za oszczerców i współników.

2.16.2 Czy zastosowanie kwantyli rzędu 0.01 i 0.99 zamiast 0.05 i 0.95 w skróciłoby, czy wydłużyło opisaną procedurę eliminacji nieuczciwych klientów?

Wydłużyłoby, ponieważ w każdej iteracji system byłby bardziej uważny przy odrzucaniu nieuczciwej liczby klientów i odrzucał by ich mniej.

2.16.3 Do wypracowania oceny z^* wykorzystuje się rozkład a posteriori $p(z|w, n)$. Z punktu widzenia twierdzenia Bayesa, co reprezentuje z , a co w i n ?

z należy do zbioru ciągłego układu zupełnego hipotez. Jego wartość należy do $[0, 1]$ i jest hipotezą, że poziom zaufania agenta = z .

w - liczba raportów o wysokiej satysfakcji

n - liczba raportów o niskiej satysfakcji

w, n - to co już zaszło i o tym wiemy

2.16.4 Dla ograniczenia efektów nieuczciwych raportów ocenę z^* porównuje się z jaką charakterystyką rozkładu $p(z|w, n)$?

Ocenę większościową $z^*(M)$ porównujemy z kwantylami rozkładu otrzymanego dla w i m tylko danego klienta. Jeśli ocena jest mniejsza lub większa od progowego kwantyla to klienta uznajemy za nieuczciwego.

2.16.5 Rozkład beta jest pewnym warunkowym rozkładem gęstości prawdopodobieństwa - jakiej wielkości losowej i przy jakim warunku?

Rozkład beta jest określony dla dwóch parametrów, α i β , które są parametrami kształtu. Wartości tych parametrów wpływają na kształt rozkładu gęstości prawdopodobieństwa. Zmienna losowa X z rozkładem beta ma gęstość prawdopodobieństwa określoną wzorem:

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

x jest wartością zmiennej losowej,

$\alpha > 0$ jest parametrem kształtu,

$\beta > 0$ jest drugim parametrem kształtu,

$B(\alpha, \beta)$ jest funkcją beta, która normalizuje rozkład (całka gęstości prawdopodobieństwa od 0 do 1 równa 1).

2.17 Generacja liczb pseudolosowych

2.17 Metoda odwracania dystrybucyj

2.17.1 Jakie przekształcenie generuje liczby losowe o rozkładzie $f(x)$?

Liczby losowe o rozkładzie $f(x)$ generuje funkcja odwrotna. Czasami trzeba zrobić całkę $f(x)$ jak $p(x)$ podane.

$$p(x) = \frac{1}{a} e^{-\frac{x}{a}} \Rightarrow X = -a \ln(1 - U)$$

$$F(x) = 0.5 + \frac{\arctan(x)}{\pi} \Rightarrow X = \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$

$$F(x) = 1 - e^{-\frac{x^b}{a}} \Rightarrow X = \sqrt[b]{-a \ln(U)}$$

$$p(x) = 2x \Rightarrow X = \sqrt{U}$$

2.18 Naturalne zbiory danych

2.18.1 Jaki jest fizyczny sens stałej k w wyrażeniu $P(X < x) = P(kX < x)$?

Naturalna zmienna losowa nie może zależeć od skali (jednostki), czyli prawdopodobieństwo X musi być takie samo jak kX , gdzie $k > 0$ jest dowolną stałą.

3 Sumy niezależnych wielkości losowych

3.1 Funkcje charakterystyczne i tworzące

3.1 Rozkłady sum

3.1.1 Jakiej operacji na niezależnych wielkościach losowych odpowiada mnożenie ich funkcji charakterystycznych?

Dodawaniu tych wielkości.

3.1.2 Dodawanie niezależnych wielkości losowych odpowiada następującym operacjom na ich rozkładach prawdopodobieństwa i funkcjach charakterystycznych:

na rozkładach: spłot, na funkcjach charakterystycznych: mnożenie

3.1.3 Jak z funkcji charakterystycznej otrzymać rozkład prawdopodobieństwa dyskretnej wielkości losowej?

Korzystając z twierdzenia o jednoznaczności

3.1.4 Jak z funkcji tworzącej otrzymać rozkład prawdopodobieństwa dyskretnej wielkości losowej?

Korzystając z rozwinięcia w szereg potęgowy

3.1.5 Czy słuszne jest stwierdzenie, że funkcja tworząca jest szczególnym przypadkiem funkcji charakterystycznej, gdzie $ejx = z$? Uzasadnij.

$$g(X) = \begin{cases} e^{jvX} X \text{ ciągła o zbiorze realizacji } x \in [a, b] \\ z^X X \text{ dyskretna o zbiorze realizacji } x \in \{0, 1, 2, \dots\} \end{cases}$$

Jest, charakterystyczna to całkowanie, tworząca to dodawanie. Różnica jest w dziedzinie.

3.2 Równoważenie obciążenia: round robin

3.2.1 Jaki rozkład ma interwał między dwoma kolejnymi wywołaniami do danego serwera w obecności 5 równoległych serwerów?

Rozkład Erlanga-5.

3.2.2 Jaka własność funkcji charakterystycznej jest wykorzystywana przy dodawaniu niezależnych wielkości losowych?

Funkcja charakterystyczna sumy niezależnych wielkości losowych jest równa iloczynowi funkcji charakterystycznych tych zmiennych.

3.2.3 W jaki sposób liczba równoległych serwerów wpływa na rozkład interwału widzianego przez ustalony serwer?

Czym więcej serwerów, tym jest to rozkład Erlanga- n o większym n . Wówczas, masa prawdopodobieństwa przesuwana się w kierunku wyższych wartości.

3.2.4 Rozkład Erlanga powstaje w wyniku dodawania niezależnych wielkości losowych o rozkładach... wykładniczych

3.3 Zaległości w obsłudze

3.3.1 W wyniku czego powstaje rozkład Erlanga rzędu k ?

W wyniku k -krotnego splotu rozkładu wykładniczego.

3.3.2 Skąd bierze się jakościowa różnica w wykresach prawdopodobieństwa naruszenia reżimu czasu rzeczywistego?

Z różnicy w czasach wykonania pojedynczego przerwania. Jeśli przychodzące przerwy zajmują nam dużo czasu to prawdopodobieństwo zaległości jest wysokie.

3.3.3 W wyniku dodawania niezależnych wielkości losowych o rozkładach równomiernych powstaje rozkład...

Irwin-Halla

3.4 Utracone wywołania

3.4.1 Jaki jest związek między rozkładem Poissona a rozkładem geometrycznym?

Jeżeli dla ilości wywołań opisywanych rozkładem Poissona przyjmiemy losowy przedział czasu o rozkładzie wykładniczym (czyli T zamiast t) to ilość wywołań w tym losowym przedziale będzie miała rozkład geometryczny (można to wyprowadzić z twierdzenia o prawdopodobieństwie całkowitym).

3.4.2 Jeżeli liczba wywołań w danym przedziale czasowym ma rozkład Poissona, to jaki jest rozkład liczby wywołań utraconych w czasie awarii o losowym czasie trwania o rozkładzie wykładniczym?

Jest to rozkład geometryczny.

3.4.3 Jeżeli liczba wywołań w danym przedziale czasu ma rozkład Poissona, to jaki jest rozkład interwału między dwoma kolejnymi wywołaniami?

Jest to rozkład wykładniczy.

3.4.4 Jaki rozkład prawdopodobieństwa odpowiada funkcji charakterystycznej będącej pojedynczą eksponentą?

Atom w jednym punkcie, a reszta jest rozłożona jak krzywa wykładnicza.

3.5 Opróżnianie stosu

3.5.1 W przykładzie dla $P(N = 0) = 0.3$, $P(N = 1) = 0.2$, $P(N = 2) = 0.2$, $P(N = 3) = 0.1$ prawdopodobieństwo, że populacja w drugim pokoleniu wymrze wynosi...

$$P(Y^2 = 0) = 0.4407$$

3.5.2 Jakie jest prawdopodobieństwo ostatecznego wymarcia populacji przy $P(N = 0) = 0.3, P(N = 1) = 0.4, P(N = 2) = 0.2, P(N = 3) = 0.05$.

Korzystamy z metody Newtona (iteracyjnym przybliżaniu pierwiastka równania)

$q = 0.3 + 0.4q + 0.2q^2 + 0.05q^3$ (przyrównujemy to do 0, czyli przenosimy q na drugą)

$f(q) = 0.05q^3 + 0.2q^2 - 0.6q + 0.3$ (obliczamy z tego pochodną)

$f'(q) = 0.15q^2 + 0.4q - 0.6$

$q_{n+1} = q_n - \frac{f(q_n)}{f'(q_n)}$

Jako q_0 wybieramy 0 i podstawiając do $f(q)$ i $f'(q)$ obliczamy wartości. Podstawiamy je do wzoru na kolejne q i mamy q_1 . Następnie obliczamy q_2 za pomocą q_1 itd.

$q_0 = 0 \rightarrow q_1 = 0.5 \rightarrow q_2 = 0.65517 \rightarrow q_3 = 0.67964$

$\Pi = 0.68078$

3.5.3 Prawdopodobieństwa liczby bezpośrednich potomków są następujące: 0 - 30%, 1 - 40%, 2 - 20% i 3 - 10%. Jaka jest szansa, że populacja nie wymrze?

Szansa że wymrze, to 0.791, czyli że przeżyje to $1 - 0.791 = 0.209$

3.5.4 Jeżeli równanie $\Pi = G_N(\Pi)$ ma dwa rozwiązania, to co można powiedzieć o prawdopodobieństwie, że stos ponownie się opróżni?

Jest ono równe mniejszemu z tych dwóch prawdopodobieństw, czyli $\Pi = \min\{\Pi_1, \Pi_2\}$. Nie ma pewności czy się opróżni, czy nie.

3.5.5 Jaka jest probabilistyczna interpretacja nachylenia krzywej $G_n(z)$ w punkcie 1?

Jest to wartość średnia liczby bezpośrednich potomków $G'_n(1) = EN$. Jeśli $EN \leq 1$ to populacja na pewno wymrze.

W przeciwnym wypadku, istnieje niezerowe prawdopodobieństwo, że stos się już nigdy nie opróżni.

Dla tych parametrów wymrze: 0.791, przetrwa: 0.209

3.6 Słabe prawo wielkich liczb: eksperyment

3.6.1 Gdyby rozkład X był wykładniczy, to do którego z pokazanych przebiegów podobny byłby analogicznie otrzymany przebieg w funkcji n ?

Błby analogiczny do rozkładu równomiernego, bo ma skończoną wartość średnią.

3.6.2 Co możemy powiedzieć o wartości średniej rozkładu Cauchy'ego?

Nie istnieje, ponieważ nie działa prawo wielkich liczb. Wielokrotnie powtórzone doświadczenie daje statystycznie tak samo dobry wynik jak pojedyncza iteracja.

3.6.3 Dlaczego przebieg $\frac{S_n}{n}$ w funkcji n nie ustala się dla wielkości losowej X o rozkładzie Cauchy'ego?

Rozkład Cauchy'ego oscyluje wokół 0.5 i się nie zbiega. W przypadku tego rozkładu wartość średnia nie istnieje więc nie działa prawo wielkich liczb, dowolna ilość obserwacji gwarantuje statystycznie tak samo dobry wynik jak jedna obserwacja.

3.6.4 Co jest powodem narastania trendu przebiegu $\frac{S_n}{n}$ w funkcji n dla wielkości losowej X o rozkładzie Pareto?

Jest to spowodowane faktem, że rozkład Pareto z takim współczynnikiem kształtu ma wartość średnią nieskończoną, zatem dąży do nieskończoności.

3.6.5 Który parametr rozkładu Pareto należałoby zmienić, aby uzyskać przebieg $\frac{S_n}{n}$ w funkcji n podobny jak dla wielkości losowej X o rozkładzie równomiernym?

Należałoby zmienić współczynnik kształtu tak, żeby otrzymać wartość średnią skończoną. Należy go ustalić $k > 1$.

3.6.6 Czy w wariancie Chinczyzna słabe prawo wielkich liczb działa dla każdego typu rozkładu prawdopodobieństwa obserwacji?

Nie, nie będzie on działał dla rozkładu, który nie ma średniej, czyli np. dla Cauchy'ego.

3.7 Centralne twierdzenie graniczne

3.7.1 W odróżnieniu od słabego prawa wielkich liczb, w wariancie Chinczyna centralne twierdzenie graniczne wymaga, aby...

wielkość losowa miała skończoną wariancję.

3.7.2 Czy statystycznie niezależne wielkości losowe o jednakowych rozkładach spełniają słabe prawo wielkich liczb i twierdzenie graniczne i pod jakimi warunkami?

Słabe prawo wielkich liczb tak, pod warunkiem, że istnieje wartość średnia. Centralne twierdzenie graniczne (Ctg) tak, pod warunkiem, że istnieje skończona wariancja.

3.7.3 Co zaobserwujemy dodając do siebie wiele rozkładów wykładniczych?

Zaobserwujemy kolejne rozkłady Erlanga-n, które będą coraz bardziej przypominać krzywą dzwonową. Rozkład wykładniczy ma skończoną wariancję.

3.8 Serwer webowy

3.8.1 Aby obliczyć $P(99000 \leq S_{40000} \leq 101000)$ potrzebna nam jest wartość funkcji Laplace'a dla...

$$\Phi\left(\frac{101000 - 40000 * 2.5}{\sqrt{40000 * 1.369}}\right) - \Phi\left(\frac{99000 - 40000 * 2.5}{\sqrt{40000 * 1.369}}\right) = 2\Phi(3.652)$$

$$2\Phi\left(\frac{101000 - 99000}{2 * \sqrt{40000 * 1.369}}\right) = 2\Phi\left(\frac{101000 - 99000}{547.6}\right) - \text{szybkie liczenie gdy } \frac{a+b}{2} = 100000$$

$$\sigma = \sqrt{10 * 0.25 * (1 - 0.25)}$$

$$P(a \leq S_n < b) = \Phi\left(\frac{b - n * m}{\sqrt{n * \sigma}}\right) - \Phi\left(\frac{a - n * m}{\sqrt{n * \sigma}}\right)$$

Wynik na slajdzie: $x = 3.652$

3.9 Eksperyment Wolfa

3.9.1 Jakie odchylenie standardowe miałby rozkład wielkości losowej $\frac{S_n}{n}$ przy założeniu, że eksperyment obejmował nie 20000, lecz 200 rzutów kostką?

$$s = \frac{0.373}{\sqrt{200}}$$

3.9.2 Co gdyby rozkład obejmował 200 rzutów?

Hipoteza, że kostka jest uczciwa byłaby bardziej wiarygodna, bo mielibyśmy większe odchylenie standardowe.

3.10 Czas trwania sesji

3.10.1 Gęstość prawdopodobieństwa przekształconej wielkości losowej otrzymujemy zakładając, że przekształcenie g spełnia warunki...

gładkość, odwracalność

3.10.2 Dystrybuantę przekształconej wielkości losowej otrzymujemy zakładając, że przekształcenie g spełnia warunki...

odwracalność

3.10.3 Dla jakiej funkcji g rozkład wielkości losowej $Y = g(X)$ ma ten sam kształt co rozkład X ?

Dla funkcji liniowej. Pochodna musi być stała.

3.10.4 Jaka jest zasada konstrukcji wykresu Q-Q i co taki wykres nam daje?

Jest to wykres kwantyl funkcji dopasowywanej do kwantyla empirycznego. Dzięki temu możemy mierzyć jakość dopasowania - idealnie prosta. Zasada tworzenia - w celu dopasowania jak najlepszego rozkładu.

3.10.5 Jaki byłby kształt rozkładu Z , gdyby przyjąć $g(x) = ax + b$ (a, b - stałe)?

Taki sam jak kształt S , bo pochodna stała.

3.10.6 Przyjmując $g(s) = s^2$ otrzymaliśmy następujący rozkład Z :

$$p_{Z(z)} = F'_{Z(z)} = \frac{F'_{S(s)}}{|g'(s)|} \Big|_{s=g^{-1}(z)}$$

$$\text{Dla } g(s) = s^2 \text{ oraz } p_{S(s)} = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-(s-m)^2/2\sigma^2}$$

$$g^{-1}(s) = \sqrt{s}$$

$$p_{Z(z)} = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot 2\sqrt{z}} e^{-(\sqrt{z}-m)^2/2\sigma^2}$$

3.11 Atak na Wi-Fi

3.11.1 Jakie zagadnienie rozstrzyga nierówność Popoviciu?

Pozwala nam określić maksymalną wariancję dla wszystkich rozkładów $\{X_1\}, \dots, \{X_n\}$.

3.11.2 Czym należy zastąpić warunek $n \geq 663.6\sigma^2$, jeżeli wymagany procent pewności zmniejszymy do 98%?

Odczytujemy argument funkcji Laplace'a dla 0.49 i oznaczamy go przez k . Wtedy $n \geq k^2 \frac{\sigma^2}{0.01}$.

4 Ogony rozkładów prawdopodobieństwa

4.1 Częstość anomalii w sygnale

4.1.1 Podaj szkic obliczenia wartości 0.25% dla szumu w kanale.

$$P(\text{szum} > 50) = P(\text{szum} > 20\sigma_{\text{szum}}) \leq \frac{1}{20^2} = 0.0025$$

4.1.2 Gdyby odchylenie sygnału wejściowego było o 50% większe $P(|\text{out} - E| > 50)$ byłoby ograniczone nie przez 2.4%, tylko przez

$$P(|X - EX| \geq c\sigma) \leq \frac{1}{c^2}$$

$$\text{Wartość dla } P(|\text{sygnał} - E| > 50) \leq 2.4\%$$

$$50 = c * 7.76 \Rightarrow c = 6.44$$

$$P(|\text{sygnał} - E| > 50) \leq \frac{1}{6.44^2} = 2.4\%$$

W zadaniu σ ma się równać $1.5 * 7.76$ czyli 11.64

$$50 = \sigma * c, \text{ a więc } c = \frac{50}{11.64} = 4,296$$

$$P(|\text{sygnał} - E| > 50) \leq \frac{1}{4.296^2} = 5.4\%$$

5.1 Jednorodny łańcuch Markowa I rzędu

Macierz przejść jednorodnego łańcucha Markowa I rzędu ma następujące własności:

- Jest to macierz kwadratowa, stochastyczna, czyli $\sum_{j=0}^s m_{ij} = 1$
- Komórki reprezentują niezmiennie w czasie prawdopodobieństwa przejść między stanami: $P(X_t = j \mid X_{t-1} = i) = m_{ij}$
- Niech $p(0)$ oznacza rozkład stanów w chwili początkowej, wówczas korzystając z macierzy M możemy policzyć $p(t) = p(0)M^t$
- Natomiast M^t jest również macierzą stochastyczną. Wartości własne tej macierzy spełniają równanie $1 = \lambda_0 \geq \dots \geq \lambda_s$

5.2 Sztuczne źródła tekstowe

5.2.1 Jak wielką przestrzeń stanów musielibyśmy rozważać, gdyby generacji tekstu na podstawie częstości n -gramów dokonywać przy pomocy łańcuchów Markowa 1 rzędu?

$$|S|^{n-1}, S \text{ to liczba możliwych znaków}$$

5.2.2 Jaki byłby rozmiar macierzy przejść w sztucznym generatorze tekstu opartym na łańcuchach Markowa I rzędu operującym częstościami n -gramów?

$$|S|^{n-1} \times |S|^{n-1}$$

5.2.3 Wyjaśnij, dlaczego do analizy w zasadzie wystarczają nam łańcuchy Markowa I rzędu.

Bo każdy łańcuch wyższego rzędu można zredukować do łańcucha 1. rzędu kosztem wykładniczego wzrostu ilości stanów.

5.2.4 Liczba stanów łańcuchów Markowa I rzędu do analizy quadrogramów na tekstach angielskich wynosi

26^3 lub 27^3 (uwzględniając spacje, na slajdach uwzględniona)

$|S|^2$ dla trigramów w I rzędzie (pamiętamy 2 ostatnie)

5.3 Ergodyczność łańcucha Markowa

5.3.1 Co przesądza o szybkości zbieżności do rozkładu granicznego?

Druga wartość własna λ_1 macierzy M . Im mniejsza, tym szybsza zbieżność.

5.3.2 Jaki jest rozkład graniczny jednorodnego dwustanowego łańcucha Markowa o macierzy przejść

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

$$(\pi_1, \pi_2) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (\pi_1, \pi_2)$$

$$\begin{cases} \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2 = \pi_1 \\ \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2 = \pi_2 \end{cases}$$

$$\pi_1 = \frac{3}{4}, \pi_2 = \frac{1}{4} \Rightarrow \pi = \left(\frac{3}{4}, \frac{1}{4}\right)$$

5.3.3 Który parametr macierzy prawdopodobieństw przejść decyduje o szybkości zbieżności rozkładu prawdopodobieństwa stanów jednorodnego łańcucha Markowa I rzędu do rozkładu granicznego $p(\infty)$?

Szybkość zbieżności łańcucha Markowa odpowiada drugiej wartości własnej M czyli λ_1 . Czym λ_1 mniejsze, tym szybciej się zbiega.

5.4 PageRank

5.4.1 Jak z punktu widzenia markowskiego modelu probabilistycznego należy interpretować współczynnik tłumienia w podstawowej formule PageRank?

Jest to prawdopodobieństwo, że internauta nieskończenie błędzący po internecie nie zrestartuje przeglądania.

Prawdopodobieństwo to stanowi współczynnik proporcjonalności między PageRank, a częstością odwiedzin strony w błędzeniu losowym.

5.4.2 Współczynniki PageRank odpowiadają następującej charakterystyce łańcucha Markowa I rzędu...

$$\frac{p_j}{d} = \frac{1-d}{s} + \sum_i \frac{p_i}{a_i}$$

5.5 Rozpoznawanie mowy

5.5.1 Gdyby jako sygnaturę ciągu X ustalić rozkład długości 3-paczek to w porównaniu z otrzymanymi na wykładzie byłby on...

o wolniejszym spadku wraz ze wzrostem k , ponieważ prawdopodobieństwo 3-paczki długości k jest większe niż ciągu jedynek długości k . 3-paczki obejmują też wszystkie 1-paczki, a ciąg jedynek to właśnie 1-paczka.

5.5.2 Opisz zasadę tworzenia ciągu binarnego zgodnie z mechanizmem ukrytego łańcucha Markowa.

Mamy dwa stany, jeżeli znajdujemy się w danym kroku w stanie 1 to P wypisania jedynek jest równe b_1 , a zera $1 - b_1$, natomiast w stanie 2 wypisanie jedynek jest równe b_2 , a zera $1 - b_2$ (czyli generujemy z pewnym prawdopodobieństwem zależnym od stanu).

6 Wektory losowe i korelacja

6.1 Korelacja: wariancja $X+Y$

6.1.1 Dane są wielkości losowe X i Y o znanych odchyleniach standardowych σ_X i σ_Y . Jaka jest największa możliwa wartość $V(X+Y)$?

$$V(X+Y) = VX + VY + 2\gamma_{XY} \leq \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y = (\sigma_X + \sigma_Y)^2$$

6.2 Korelacja, a zależność statystyczna

6.2.1 Określ korelację i zależność statystyczną pomiędzy wielkościami losowymi X i Y , gdzie $X = \cos(U)$, $Y = \sin(U)$, $p_U = \frac{1}{2}\pi$, $u \in [-\pi, \pi]$

Są nieskorelowane, ale zależne (jedynka trygonometryczna). $X^2 + Y^2 = 1$ - zależne. Po scałkowaniu i obliczeniu $E(XY) - E(X)E(Y)$ wychodzi 0 - nieskorelowane.

6.2.2 Jak mają się do siebie własności korelacji i zależności statystycznej składowych wektora losowego?

Jeżeli składowe wektora są niezależne, to są nieskorelowane ($\text{cov} = 0$). Jeżeli są nieskorelowane, to z tego nie wynika niezależność - patrz zad. wyżej.

6.3 Regresja liniowa

6.3.1 Jaki wpływ ma unormowany współczynnik kowariancji na prostą regresji i błąd średniokwadratowy regresji?

Od tego współczynnika zależy współczynnik kierunkowy prostej. Przykładowo, ujemnie skorelowane skutkują w prostej malejącej. Ponadto, im bardziej skorelowane zmienne, tym mniejszy błąd średniokwadratowy. Błąd ten wynosi 0 dla korelacji liniowej. Natomiast jest maksymalny, gdy współczynnik korelacji wynosi 1. Czym bardziej skorelowane tym lepiej (nie ważne czy dodatnio czy ujemnie).

6.3.2 Jak wygląda prosta regresji w przypadku, gdy X i Y są nieskorelowane?

W przypadku, gdy zmienne X i Y są nieskorelowane, współczynnik kierunkowy β_1 w prostym modelu regresji liniowej będzie bliski lub równy zeru, co oznacza, że nie ma liniowego związku między X i Y . W takiej sytuacji, najlepszym estymatorem dla Y byłaby po prostu średnia Y , ponieważ X nie dostarcza żadnej dodatkowej informacji, która mogłaby pomóc w przewidywaniu Y .

Zatem równanie prostej regresji przy braku korelacji między X i Y sprowadza się do: $Y = \beta_0 + \varepsilon$, gdzie β_0 będzie równy średniej Y .

6.4.1 Optymalna stratna kompresja obrazu wymaga, by kolumny macierzy do obrotu wektora pikseli spełniały następujące warunki:

żeby były m największymi wartościami własnymi macierzy kowariancji Γ . Wtedy jako kolumny macierzy wstawiamy wektory własne macierzy kowariancji o tych wartościach.

6.4.2 Opisz operację macierzową prowadzącą do dekokorelacji ciągu pikseli.

Operacja ta polega na obrocie układu współrzędnych za pomocą macierzy Q , której kolumny są największymi wektorami własnymi macierzy kowariancji Γ . Obrót ten prowadzi do zapisania n wartości pikseli za pomocą m wartości, gdzie $m \leq n$.

6.4.3 Które charakterystyki macierzy kowariancji ciągu pikseli są istotne z punktu widzenia optymalizacji stratnej kompresji obrazu?

Istotne są odpowiednie wektory własne macierzy.

6.4.4 Jak należy dobrać kolumny macierzy przekształcenia Q , aby zdekokorelować piksele i zminimalizować stratność kompresji?

Błąd ten zminimalizujemy, gdy kolumnami macierzy Q uczynimy wektory własne macierzy.

7 Estymatory

7.1 Próbką losową populacji

7.1.1 Podaj warunki probabilistyczne, jakie powinna spełniać prawidłowo pobrana próba losowa z populacji.

Prosta - X_i są niezależne. Reprezentatywna - $X_i \sim X$ (rozkład cech osobnika i taki jak rozkład cech całej populacji).

7.2 Estymatory maksimum wiarygodności

7.2.1 Jaki jest wpływ błędów statystycznych poszczególnych obserwacji na ocenę wartości średniej w populacji o modelu $N(a, \sigma)$?

Czym większy błąd statystyczny danej próbki, z tym mniejszą wagą uwzględniamy ją przy estymacji wartości średniej. Ma to sens, gdyż obserwacje są dla nas mniej wiarygodne, skoro obarczone są większym błędem.

7.3 Monte Carlo: redukcja wariancji

7.3.1 Która z dwóch wielkości: $VX + VY$ oraz $V(X + Y)$ jest większa?

To zależy od ich korelacji.

Jeżeli jest ujemna to $V(X + Y) < VX + VY$

Jeżeli dodatnia to $V(X + Y) > VX + VY$

Jeżeli zerowa to $V(X + Y) = VX + VY$ czyli wtedy, kiedy X i Y są niezależne.

7.3.2 Jaki efekt przyniosłaby metoda zmiennych parami przeciwnych, gdyby dla $k = 1, 3, 5, \dots$ przyjąć $U_{(k+1)} = U_{(k)}$?

Zmniejszenia odchylenia standardowego.

7.4 Prognoza ataków dnia zerowego

7.4.1 Jak wzrost liczby typów zaobserwowanych ataków przekładałby się na jakość prognozy ataków dnia zerowego?

Otrzymałobyśmy wówczas mniej efektywny estymator.

7.4.2 Co reprezentuje otrzymana wartość $14376 - 4343 + 2292 - \dots = 11430$: estymator? obciążenie estymatora? ocenę? błąd statystyczny oceny? coś innego?

Estymatę, czyli konkretną ocenę dokonaną przez estymator. A konkretnie liczbę nowych typów. Natomiast 178 to błąd statystyczny oceny.

7.5 Przedziały ufności - wyznaczanie

7.5.1 Dla ustalonych n i z szerokość przedziału ufności rośnie wraz z...

Ze wzrostem δ , ponieważ wówczas pozwalamy na to, że a może przyjmować szersze spektrum wartości.

7.6 Robak sieciowy

7.6.1 Gdyby przy $n = 1000$ i $z = 0.2$ jako poziom ufności przyjąć 99%, to czy mogłaby być poprawna wartość $P(0.1697 < a < 0.2324) = 0.99$?

Nie, ponieważ nie wyznaczamy prawdopodobieństwa tego, że a należy do jakiegoś przedziału, bo a nie jest zmienną losową. Zamiast tego wyznaczamy prawdopodobieństwo tego, że kolejne wyznaczone przedziały ufności będą zawierały a .

7.6.2 Gdyby przy $n = 1000$ i $z = 20\%$ jako poziom ufności przyjąć $\delta = 90\%$, to czy spodziewany rezultatem jest $J_\delta(z) = [0.16, 0.24]$? Uzasadnij.

Nie, prawdziwy wynosi $[0, 1801, 0, 2216]$. Ponieważ wraz ze spadkiem poziomu ufności, "ufamy" mniejszej ilości rozwiązań, więc przedział się zwęża a nie rozszerza.

n	z	δ	$J_\delta(z)$
1000	0.2	0.95	[0.1764, 0.2260]
1000	0.2	0.90	[0.1801, 0.2216]
10000	0.2	0.95	[0.1923, 0.2080]
1000	0.25	0.95	[0.2242, 0.2778]

7.6.3 Gdyby przy $n = 1000$ i $z = 20\%$ jako poziom ufności przyjąć $\delta = 99\%$, to czy mogłaby być prawdziwa równość $P(a \in [0.1619, 0.232]) = 0.99$?

Nie, jeżeli np. $a = 0.17$ to $P = 1$, jeżeli $a = \text{np. } 0.1$ to $P = 0$, a nie $\frac{50}{50}$. Patrz wyżej.

8 Testowanie hipotez

8.1 Policer SLA

8.1.1 Czy hipotezę H_0 częściej będziemy odrzucać na poziomie istotności 5% niż na poziomie istotności 1%?

Częściej będziemy odrzucać dla 5%, ponieważ ma większy obszar odrzucenia.

8.1.2 Kiedy definiujemy prawostronny, a w jakiej sytuacji lewostronny obszar odrzucenia O_α ?

Prawostronny, jeśli duże wartości \hat{A} przemawiają przeciwko H_0 . Lewostronny, jeśli małe.

8.1.3 Dlaczego w teście serii przyjęliśmy dwustronny obszar odrzucenia?

Bo i małe i duże ilości serii przemawiają przeciwko temu, że próba jest prosta.

8.1.4 Opisz idealny i najdalszy od idealnego przebieg krzywej ROC. Uzasadnij

Idealna krzywa ROC zaczyna się w punkcie (0,0), przechodzi pionowo do punktu (0,1), a następnie poziomo do punktu (1,1), co oznacza doskonałą klasyfikację z AUC równym 1. Najgorsza krzywa ROC to linia prosta od (0,0) do (1,1), wskazująca na losowe przypisywanie klas z AUC równym 0.5, czyli brak zdolności do nauki i rozróżniania klas.

8.2 Test istotności

8.2.1 Gdy wyznaczona wartość p -value jest wyższa niż przyjęty poziom istotności, to czy powinniśmy odrzucić hipotezę zerową?

Gdy jest wyższa to bezpieczniej czujemy się nie odrzucając.

8.2.2 Gdy wyznaczona wartość p -value jest niższa niż przyjęty poziom istotności, powinniśmy odrzucić hipotezę zerową?

Tak, powinniśmy odrzucić hipotezę zerową, ponieważ dla mniejszych poziomów istotności byśmy to odrzucili. Odrzucając to czujemy się tym bezpieczniej, im mniejsza jest p -value.

8.2.3 Gdy wyznaczona wartość p -value jest wyraźnie niższa niż przyjęty poziom istotności, czujemy się bardziej, czy mniej komfortowo odrzucając hipotezę zerową?

Gdy jest niższa to odrzucając czujemy się bezpieczniej.

