

Metody Probabilistyczne

Spis treści

1. Pytania o wszystkim i o niczym	2
1.1. Probabilistyka - cele, umiejscowienie	2
1.2. Mechanizmy losowe i pseudolosowe	2
1.3. Mity a fakty	2
2. Rozkłady prawdopodobieństwa	3
2.1. iTrust	3
2.2. Symulator w hiperkuli	3
2.3. Kod korekcyjny w przestrzeni Hamminga	3
2.4. Sieć sensorów	3

Indeks pojęć

Kod korekcyjny w przestrzeni Hamminga	3
Mechanizmy losowe i pseudolosowe	2
Mity a fakty	2
Probabilistyka - cele, umiejscowienie	2
Sieć sensorów	3
Symulator w hiperkuli	3
iTrust	3

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).

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.

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. Sieć sensorów

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?

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