Spis treści

1 Rozkłady

1.1 Rozkład dwumianowy

Rozkład dwumianowy to rozkład sumy n zmiennych losowych o rozkładzie Bernoulliego. Zmienna losowa X ma rozkład dwumianowy z parametrami n i p, zatem:

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

gdzie $\binom{n}{k}$ to liczba kombinacji k sukcesów w n próbach.

1.2 Rozkład normalny

Rozkład normalny (Gaussa) jest jednym z najważniejszych rozkładów statystycznych. Jest on określony przez dwa parametry: wartość oczekiwaną μ i wariancję σ^2. Gęstość rozkładu normalnego jest dana wzorem:

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

1.3 Rozkład chi-kwadrat

Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o rozkładzie normalnym $N(0,1)$. Wtedy zmienna losowa $X = \sum_{i=1}^{n} X_i^2$ ma rozkład chi-kwadrat z n stopniami swobody.

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

1.4 Rozkład wykładniczy

Jest to rozkład zmiennej, która opisuje czas między zdarzeniami w procesie Poissona. Zmienna losowa X ma rozkład wykładniczy z parametrem λ, zatem:

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0 \end{cases}$$

2 Statystyka Opisowa

Niech $X' = \{x_1, x_2, \ldots, x_n\}$ będzie zbiorem n obserwacji zmiennej losowej X. Zadaniem statystyki opisowej jest prezentacja rozkładu zmiennej losowej X w próbce X'.

2.1 Rodzaje statystyk opisowe

- Klasyczne - uśredniające wartość próbki. Na przykład momenty zwykłe r-tego rzędu:

$$m_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r$$

- Pozycyjne - oparte na pozycjach obserwacji w próbce. Na przykład mediana, kwartyle, percentyle.

2.2 Tendencja centralnej rozkładu empirycznego

- Średnia arytmetyczna:

$$\overline{X'} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Mediana

2.3 Charakterystyki rozrzutu rozkładu empirycznego

- Odchylenie standardowe:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{X'})^2}$$

- Współczynnik zmienności:

$$v = \frac{s}{\overline{X'}} \cdot 100\%$$

3 Model statystyczny

Jeżeli próba X' jest reprezentatywna, to można na jej podstawie wnioskować na temat populacji z której pochodzi. Aby określić zachowanie zmiennej losowej X w populacji, stosuje się model statystyczny. Zatem traktujemy wektor X' jako realizację zmiennej losowej X.

4 Estymacja Punktowa

Niech X' będzie próba populacji o rozkładzie P_θ gdzie $\theta \in \Theta$ jest parametrem. Estymatorem parametru θ nazywamy statystykę $\hat{\theta}$: $X' \to \Theta$ która pozwala na oszacowanie wartości parametru θ.

4.1 Metoda momentów

Metoda momentów polega na przyrównaniu kolejnych d momentów $m_1, \ldots, m_d$ do odpowiednich momentów rozkładu populacji $E(X^i) : i \in [1, d]$

4.2 Metoda największej wiarygodności

Funkcję $L(\theta, x) = p_\theta(x)$ nazywamy funkcją wiarygodności. Estymatorem największej wiarygodności parametru θ nazywamy nazywamy statystykę $\overset{\wedge}{\theta}$ która maksymalizuje funkcję wiarygodności.

$$\forall_{x \in X} L(\overset{\wedge}{\theta}, x) = \sup_{\theta \in \Theta} L(\theta, x)$$

4.3 Przykład

Estymatorem największej wiarygodności oraz metody momentów dla rozkładu wykładniczego z parametrem λ jest:

$$\overset{\wedge}{\lambda} = \frac{1}{\overline{\overline{X}}}$$

4.4 Estymatory nieobciążone

Estymator $\overset{\wedge}{\theta}$ nazywamy nieobciążonym, jeżeli $E(\overset{\wedge}{\theta}) = \theta$

4.5 Estymator modelu wykładniczego

Dla modelu wykładniczego, parametryzowanego przez λ, estymatorem nieobciążonym jest:

$$\overset{\wedge}{\lambda} = \frac{n-1}{n} \frac{1}{\overline{\overline{X}}}$$

4.6 Estymator modelu normalnego

Dla modelu normalnego, parametryzowanego przez μ i σ^2, estymatorem nieobciążonym jest:

$$\overset{\wedge}{\mu} = \overline{X}$$

$$\overset{\wedge}{\sigma}^2 = S^2$$

5 Metoda monte carlo

Niech X' będzie próbą populacji o rozkładzie P_θ, oraz niech $\overset{\wedge}{\theta}$ będzie estymatorem parametru θ. Załóżmy też, że mamy k niezależnych realizacji próby $x_1, \ldots x_k$. Wtedy histogram wartości $\hat{x}_n$: $n \in [1, k]$ jest przybliżeniem rozkładu $\overset{\wedge}{\theta}$.

6 Metoda bootstrap

Dystrybuanta empiryczna to statystyka o następującej postaci:

$$\hat{F}(x) = \frac{\#\{k : X_k \le x\}}{n}$$

Dla takiej dystrybuanty i próby X' zachodzi:

$$\sup_{x \in \mathbb{R}} | \hat{F}(x) - F(x) | \xrightarrow{1} 0$$

Próba bootstrapową X^* to próba losowa z rozkładu empirycznego. Ta próba musi powstać w wyniku n-krotnego losowania z zwracaniem. Rozkład statystyki $T(X^*) - \overset{\wedge}{\theta}$ jest bliski rozkładowi statystyki $T(X) - \theta$.

Mając k realizacji prób bootstrapowych $X_1^*, \ldots, X_k^*$, możemy przybliżyć rozkład statystyki $\overset{\wedge}{\theta} - \theta$, poprzez stworzenie histogramu $\overset{\wedge}{\theta} *_n : n \in [0, k]$

7 Przedziały ufności