Zestaw 1
Statystyczna Analiza Danych

Emilia Wiśnios

18 marca 2021

1. **Generowanie zmiennych losowych**

   Niech X będzie zmienną losową o ciągłej i ściśle rosnącej dystrybuancie F oraz niech U będzie zmienną losową o rozkładzie jednostajnym na odcinku $[0,1]$. Pokaż, że

   - zmienna losowa $F^{-1}(U)$ ma dystrybuantę F.
   - zmienna losowa $F(X)$ ma rozkład jednostajny na odcinku $[0,1]$.

   Rozwiązanie:

   > X - zmienna losowa
   > **Dystrybuanta** jest określona wzorem
   > $$F(t) = P(X \leq t)$$

   - $P(F^{-1}(U) \leq t) = P(F \circ F^{-1}(U) \leq F(t)) = P(U \leq F(t))$
     Korzystając ze znanej dystrybuanty dla rozkładu jednostajnego na odcinku $[0,1]$ (identyczność) mamy, że
     $$P(U \leq F(t)) = F(t)$$
   - $X \sim F$
     $$X \stackrel{d}{=} F^{-1}(U) \quad \Rightarrow \quad F(X) \stackrel{d}{=} F \circ F^{-1}(U) \stackrel{d}{=} U$$

2. **Rozkład dwumianowy. Choroba Taya-Sachsa.**

   Jeśli obydwoje rodziców jest dotkniętych chorobą, ich dziecko ma 1\4 szans odziedziczenia jej. Para chorych ma czwórkę dzieci. Jakie jest prawdopodobieństwo zachorowania dla danej liczby $l = 0, 1, ..., 4$ dzieci?

   Rozwiązanie:

   Chcemy obliczyć prawdopodobieństwo
   $$P(l \text{ chorych dzieci}) = \binom{4}{l}\left(\frac{1}{4}\right)^{l}\left(\frac{3}{4}\right)^{4-l} \quad l = 0, 1, 2, 3, 4$$
   .

3. Niech X będzie zmienną losową o rozkładzie Poissona, z parametrem λ. Wyprowadź

$$E(X) = Var(X) = \lambda$$

Rozwiązanie:

$X \sim$ Poiss (λ)

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, ...$$

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda e^{\lambda} e^{-\lambda} = \lambda$$

$$Var(X) = E(X - EX)^2 = EX^2 - (EX)^2$$

$$EX^2 = \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{k\lambda^k e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=1}^{\infty} \frac{k\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda E(X+1) = \lambda(\lambda+1)$$

$$Var(X) = \lambda(\lambda+1) - \lambda^2 = \lambda$$

4. **Rozkład dwumianowy vs Poissona**

Załóżmy, że prawdopodobieństwo uzyskania sukcesu w schemacie Bernoulliego maleje wraz ze wzrostem liczby doświadczeń w ten sposób, że $np_n = \lambda$ dla dowolnego n gdzie λ jest ustaloną liczbą dodatnią. Pokaż, że

$$\lim_{n\to\infty} P(K_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, ...$$

gdzie K_n to zmienna losowa o rozkładzie Binomial(n, p_n).

Rozwiązanie:

$$P(K_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n!}{(n-k)!k!} \frac{(np_n)^k}{n^k} \left(1 - \frac{np_n}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)...(n-k+1)}{n \cdot n \cdot ... \cdot n} \frac{\lambda^k}{k!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \overset{n\to\infty}{\to} \frac{\lambda^k}{k!} e^{-\lambda}$$

5. **Brak pamięci rozkładu wykładniczego.**

- Rozważmy komponent konstrukcyjny jakiejś maszyny i wybierzmy rozkład wykładniczy z parametrem λ jako model jego wytrzymałości (ile czasu T działa).
- Załóżmy, że komponent działa już czas s.
- Jakie jest prawdopodobieństwo, że będzie działał jeszcze dodatkowo t jednostek czasu?

Rozwiązanie:

Gęstość $X \sim f$

$$P(X \in B) = \int_B f(x)dx$$

$$X \sim \exp(\lambda)$$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$T \sim \exp(\lambda)$$

Prawdopodobieństwo warunkowe

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

$$P(T > t+s \,|\, T > s) = \frac{P(T > t+s \ \cap \ T > s)}{P(T > s)} = \frac{P(T > t+s)}{P(T > s)}$$

$$P(T > s) = \int_s^\infty \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_s^\infty = 0 - (-e^{-\lambda s}) = e^{-\lambda s}$$

$$P(T > t+s \,|\, T > s) = \frac{P(T > t+s \ \cap \ T > s)}{P(T > s)} = \frac{P(T > t+s)}{P(T > s)} = e^{-\lambda t}$$

Tylko ten rozkład ma własność braku pamięci.

Funkcja przeżycia

$$\overline{F}(t) = 1 - F(t)$$

$$\overline{F}(t+s) = \overline{F}(t) \cdot \overline{F}(s)$$

Ile jest funkcji ciągłych spełniających te warunki? Jedna co do podstawy.

6. **Kwantyl rozkładu Laplace'a**

Podać wzór na kwantyl rzędu p w rozkładzie o gęstości

$$f(x) = 1\backslash 4 e^{-|x|/2}$$

Rozwiązanie:

mediana = kwantyl rzędu 1\2
kwantyl rzędu p

$$P(X \leq q(p)) \geq p$$

$$P(X \leq q(p)) \geq 1 - p$$

$$F^{-1}(p) = q(p)$$

Kwantyle dla rozkładu Laplace'a

$$F(t) = \int_{-\infty}^t \frac{1}{4} e^{-|x|/2} dx$$

Z symetrii:

$$F(0) = \frac{1}{2}$$

$$F(t) = \frac{1}{2} + sgn(t) \int_0^{|t|} \frac{1}{4} e^{-x/2} dx = \frac{1}{2} + sgn(t) \left[-\frac{1}{2} e^{-x/2} \Big|_0^{|t|} \right] = \frac{1}{2} + sgn(t) \left[-\frac{1}{2} e^{-|t|/2} + \frac{1}{2} \right] =$$

$$= \begin{cases} \frac{1}{2} e^{t/2} & t < 0 \\ 1 - \frac{1}{2} e^{-t/2} & t \geq 0 \end{cases}$$

Otrzymana dystrybuanta jest ciągła i jest ściśle rosnąca. Możemy zatem policzyć funkcję odwrotną, czyli

$$F(t) = p$$

Z warunku $F(0) = \frac{1}{2}$ mamy dla $p \leq \frac{1}{2}$

$$\frac{1}{2}e^{t/2} = p \;\Rightarrow\; t/2 = \log(2p) \;\Rightarrow\; 2\log(2p)$$

Dla $p > 1/2$ mamy analogicznie $-2\log(2(1-p))$.

7. **Standaryzacja**

Pokazać, korzystając z pojęcia dystrybuanty, że standardyzowana zmienna losowa $Y = (X-\mu)/\sigma$, powstająca ze zmiennej X o rozkładzie $N(\mu,\sigma)$ ma rozkład $N(0,1)$.

Rozwiązanie:

$$X \sim N(\mu, \sigma^2)$$
$$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$EX = \mu, \quad VarX = \sigma^2$$

X, Y- niezależne zmienne losowe o rozkładzie normalnym

$$X + Y \text{ mają rozkład normalny}$$

$$E(X+Y) = EX + EY$$
$$\mu_{X+Y} = \mu_X + \mu_Y$$
$$Var(X+Y) = VarX + VarY$$
$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

X - zmienna losowa o rozkładzie normalnym, $ax + b$ też ma rozkład normalny

$$Y = aX + b$$

$$\mu_Y = EY = E(aX+b) = aEX + b = a\mu_X + b$$
$$\sigma^2_Y = VarY = Var(aX+b) = Var(aX) = a^2 VarX = a^2\sigma^2_X$$

$X \sim N(\mu, \sigma^2), Y = \frac{X-\mu}{\sigma}$

$$EY = 0$$
$$VarY = ?$$

Uwaga - wystarczy sprawdzić dla sytuacji gdy $X \sim N(0,1)$

$$Y = aX + b$$
$$P(Y \leq t) = P\left(X \leq \frac{t-b}{a}\right) = F_X\left(\frac{t-b}{a}\right)$$
$$f_Y(t) = \frac{d}{dt}P(Y \leq t) = f_X\left(\frac{t-b}{a}\right)\frac{1}{a}$$
$$f_X(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$
$$f_y(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{t-b}{a})^2}{2}}\frac{1}{a} = \frac{1}{\sqrt{2\pi}a}e^{-\frac{(t-b)^2}{2a^2}} \quad \Rightarrow \quad N(b, a^2)$$

8.

9.

10. W urnie są trzy czarne kule i jedna biała. Losujemy kolejno i bez zwracania kule aż do momentu wyciągnięcia białej. Niech X - liczba wyciągniętych kul. Oblicz $E(X)$.

Rozwiązanie:

$$P(X=1)=\frac{1}{4}$$
$$P(X=2)=\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{4}$$
$$P(X=3)=\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{4}$$
$$P(X=4)=\frac{1}{4}$$
$$EX=2,5$$

11. Co jest bardziej prawdopodobne: wygrać z równorzędnym przeciwnikiem dwie partie z trzech, czy cztery partie z sześciu?

Rozwiązanie:

$$P(2 \text{ z } 3)=\binom{3}{2}\left(\frac{1}{2}\right)^3=\frac{3}{8}$$
$$P(4 \text{ z } 6)=\binom{6}{4}\left(\frac{1}{2}\right)^6=\frac{15}{64}$$

12.

13.