SGH, Zwiad, Symulacje w arkuszu kalkulacyjnym – sygn. 4622-01

Własności wybranych rozkładów prawdopodobieństwa

Rozkłady dyskretne

1. Bernoulliego – dwupunktowy: opis sytuacji, w których wynikiem może być sukces lub porażka (np. zabieg chirurgiczny)

   p – prawdopodobieństwo sukcesu

   x – sukces: 1, porażka: 0

$$f(x) = p^x(1-p)^{1-x} \qquad E(x) = p \qquad D^2(x) = p(1-p)$$

2. Binomial – dwumianowy: liczba sukcesów w n –elementowej próbie (np. liczba wadliwych elementów w próbie 20 sztuk pobranych w procesie produkcyjnym)

   p – prawdopodobieństwo pojedynczego sukcesu

   x – liczba sukcesów

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad E(x) = np \qquad D^2(x) = np(1-p)$$

3. Discrete uniform – jednostajny dyskretny: opis sytuacji, w których jest kilka możliwych wyników, a każdy z nich jest równie prawdopodobny (nie ma przesłanek, aby pewne zdarzenia traktować jako bardziej prawdopodobne niż inne), np. rzut kostką do gry

$$f(x) = \frac{1}{n} \qquad E(x) = \frac{a+b}{2} \qquad D^2(x) = \frac{n^2-1}{12}$$

4. Geometric – geometryczny: jakie jest prawdopodobieństwo, że w próbie numer x+1 zostanie osiągnięty sukces,

   p – prawdopodobieństwo pojedynczego sukcesu

$$f(x) = p(1-p)^x \qquad E(x) = \frac{1}{p} \qquad D^2(x) = \frac{1-p}{p^2}$$

5. Negative binomial – ujemny dwumianowy: liczba prób do osiągnięcia sukcesu x razy

   p – prawdopodobieństwo pojedynczego sukcesu

   x – liczba sukcesów

   n – liczba prób

$$f(x) = \binom{x+n-1}{n} p^x (1-p)^n \qquad E(x) = x\frac{1-p}{p} \qquad D^2(x) = x\frac{1-p}{p^2}$$

6. Poissona: rozkład zdarzeń rzadkich: liczba zdarzeń w czasie, np. liczba wypadków drogowych, pożarów, wygranych w Lotto;

   Rozkład graniczny dla r. Bernoulliego:

   $n \to \infty$

   $p \to 0$

   $np \to \lambda$

$$f(x) = \frac{\lambda^k}{x!} e^{-\lambda} \qquad E(x) = \lambda \qquad D^2(x) = \lambda$$

Rozkłady ciągłe

7. Beta:
   - modelowanie rozkładów unimodalnych oraz procesów z naturalnymi ograniczeniami z dołu i z góry
   - np. rozkład statystyk pozycyjnych (min,max,E(x)), czas przebywania w sieci PERT

8. Chi kwadrat
   - testy statystyczne χ^2 (zgodności rozkładów, analiza wariancji)
   - rozkład zmiennej $Y=X_1^2 +\ldots+ X_n^2$, gdzie X_i ~IND, i=1,…,n

9. Erlang:
   historycznie – liczba rozmów telefonicznych, które mogą być wykonywane w danym momencie czasowym,
   obecnie – w inżynierii ruchu, jako czas przebywania w systemie kolejkowym

10. Exponential – wykładniczy:
    - czas pomiędzy kolejnymi wystąpieniami rzadkiego zdarzenia, które zachodzi średnio λ razy na jednostkę czasu
    - np. odstęp czasu pomiędzy przejazdami samochodów przez skrzyżowanie, czas bezawaryjnej pracy urządzeń elektronicznych, czas pojawienia się klientów w sklepie spożywczym

11. F Snedecora
    - iloraz dwóch zmiennych o rozkładzie chi-kwadrat
    - jest wykorzystywany najpowszechniej do oceny wariancji (np. test na równośc wariancji, ANOVA)

12. Gamma
    - suma k zmiennych losowych o rozkładzie wykładniczym
    - w r. wykładniczym moda =0, natomiast w r. gamma moda $\neq 0$
    - np. rozkład czasu bezawaryjnej pracy żarówki elektrycznej, czas obsługi w kasie biletowej na meczu piłkarskim

13. Lognormal
    - jeżeli x ~N(μ,σ), to e^x ma rozkładzie logarytmiczno-normalnym
    - modelowanie zmiennych, które są iloczynem wielu niezależnych czynników
    - np. iloczyn dziennych stóp zwrotu, błędy pomiaru, rozmiar pól ropy naftowej, dochody osobiste, wiek w momencie zawierania pierwszego małżeństwa

14. Normal
    - dobre przybliżenie rozkładu z próby, jeśli
      a) występuje silna tendencja do przyjmowania wartości blisko środka rozkładu
      b) dodatnie i ujemne odchylenia względem środka są równie prawdopodobne
      c) liczność odchyleń gwałtownie spada wraz ze wzrostem ich wielkości
    - z uwagi na CTW r.n. można wyobrazić sobie jako nieskończoną sumę niezależnych zdarzeń losowych (np. dwumianowych)
    - np. wzrost, waga człowieka

15. t Studenta (Gosset w browarach Guinessa, 1908)
    - symetryczny kształt podobny do r. normalnego
    - testy dotyczących wartości średniej w populacji (gdy nieznane jest odch. std.)

16. Trójkątny
    - subiektywny opis populacji, gdy znane są tylko wyrywkowe dane
    - oparty na znajomości min, max oraz przewidywanej wartości modalnej

17. Jednostajny
    - opis zm. los. o stałej gęstości prawdopodobieństwa w obrębie określonego przedziału (a;b)

18. Weibulla
    - wykorzystywany w analizie danych behawioralnych, ze względu na dużą elastyczność, może naśladować r. normalny, wykładniczy
    - najczęściej w ocenie niezawodności urządzeń (np. przekaźników elektronicznych, łożysk kulkowych), czas produkcji, moment ukończenia, prędkość wiatru, prawdopodobieństwo utraty sygnału w telekomunikacji
    - jest wykorzystywany, gdy prawdopodobieństwo defektu zmienia się (maleje) w czasie

Testy zgodności

Weryfikacja, czy elementy próby mają rozkład zgodny z zakładanym rozkładem teoretycznym

1. Chi kwadrat – różnica liczności klas

   Statystyka testująca: $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{n_i}$ ~ χ^2_{k-m-1}, gdzie

   i=1,…, k – numer klasy

   m – liczba parametrów rozkładu

   n_i - empiryczna liczebność klasy i w próbie

   $\hat{n}_i$ - teoretyczna liczebność klasy i w próbie

2. Kołmogorowa-Smirnowa – różnica dystrybuant
   - F_0 – dystrybuanta teoretyczna
   - oblicz dystrybuantę empiryczną

     $F_n(x) = {}^j/_n$, dla $x_j \leq x < x_{j+1}$
   - statystyka testująca $D_n = \max\{ D_n^-, D_n^+\}$,gdzie

     $D_n^- = \max_{1\leq j\leq n} \{ {}^j/_n - F_0(x_j) \}$

     $D_n^+ = \max_{1\leq j\leq n} \{ F_0(x_j) - {}^{j-1}/_n \}$
   - test dostosowano również dla dużych prób – test λ-Kołmogorowa

3. Anderson-Darling: test oparty na kwadracie różnicy dystrybuant

$$Q = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 \psi(x) d^2 F_0(x)$$

$$A^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 [F_0(x)(1 - F_0(x))]^{-1} d^2 F_0(x)$$