Modelowanie stochastyczne I*

PPT IM IIIr.
Lista II†

28.10.2004

1 O pewnych rozkładach zmiennych losowych

1.1 Transformacje

W wielu przypadkach próbuje się przekształcać obserwowane dane w nadziei, że transformowane dane będzie można łatwiej ziterpretować. Typowymi przekształceniami stosowanymi w praktyce są $g(x) = a + bx$, x^a, $\exp(x)$ i $\log(x)$ $(a, b \in \Re)$. Zauważmy, że statystyki porządkowe są przekształcane w odpowiednie statystyki porządkowe, gdy funkcja g jest rosnąca, tj. $Y_1^* = g(X_1^*), Y_2^* = g(X_2^*), \ldots, Y_n^* = g(X_n^*)$. W przypadku malejącej funkcji g mamy: $Y_n^* = g(X_1^*), \ldots, Y_1^* = g(X_n^*)$.

1. Zmienną losową o rozkładzie Pareto $Pa(\alpha)$ otrzymujemy przekształcając zmienną losową $X \sim Exp(\alpha)$ przez $g(x) = \exp(x)$. Niech $Y = \exp(X)$.

   $$1 - G_Y(t) = t^{-\alpha} \text{ gdy } t > 1,\ \alpha > 0$$

   Parametr α nazywamy indeksem Pareto.

2. Niech $X \sim Exp(\lambda)$ i $Y = g(X) = X^{\frac{1}{\tau}}$, gdzie $\tau > 0$, $\lambda > 0$. Y ma rozkład Weibulla:

   $$1 - F(y) = \exp(-\lambda x^{\tau}).$$

   τ nazywamy indeksem Weibulla.

3. Niech $X \sim Exp(1)$ i $Y = g(X) = \frac{X^{-\gamma}-1}{\gamma}$, gdzie $\gamma \in \Re$. Rozkład zmiennej losowej Y ma postać

   $$G_\gamma(y) = 1 - \exp(-(1 + \gamma y)^{-\frac{1}{\gamma}}) \text{ gdy } 1 + \gamma y > 0.$$

   Rozkład o dystrybuancie $G_\gamma(y)$ nazywamy rozkładem wartości ekstremalnej.

*Zadania do 21 listopada '04.

†Krzysztof Szajowski, Instytut Matematyki, Politechnika Wrocławska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, E-mail: *szajow@im.pwr.wroc.pl*

4. Szczególny przypadek rozkładu wartości ekstremalnej $G_\gamma(y)$ omówionego w punkcie 3 dla $\gamma = 0$ nazywamy rozkładem Gumbela albo podwójnie wykładniczym. Transformacja $g(x)$ z punktu 3 przy $\gamma \to 0$ ma postać $g(x) = -\log(x)$. Wówczas

$$G_0(y) = 1 - \exp(-e^{-y}) \text{ dla } y \in \Re.$$

5. Przekształcenie liniowe $g(x) = a + bx$ zmiennej losowej Y zdefiniowanej w punkcie 3 daje zmienną losową o uogólnionym rozkładzie wartości ekstremalnej:

$$F(t) = 1 - \exp(-[1 + \gamma\frac{t-a}{b}]^{-\frac{1}{\gamma}}) \text{ dla } 1 + \gamma\tfrac{t-a}{b} > 0.$$

   Szczególny przypadek uogólnionego rozkładu wartości ekstremalnej przy $\gamma = b$ i $a = 1$ nazywamy rozkładem Frécheta:

$$F(t) = 1 - \exp(-t^{-\frac{1}{b}}) \text{ dla } t > 0.$$

6. Niech zmienna losowa X ma rozkład Pareto $Pa(\lambda)$ i niech $Y = g(X) = (\beta(X-1))^{\frac{1}{\tau}}$ z $\beta, \tau > 0$. Mamy

$$1 - F_Y(t) = (\frac{\beta}{\beta + t^\tau})^\lambda. \tag{1}$$

7. Przekształcenie liniowe $g(x) = \frac{x-b}{a}$, $a > 0$, zmiennej losowej X o rozkładzie Pareto daje zmienną losową o alternatywnym rozkładzie Pareto

$$1 - F(t) = (b + at)^{-\alpha}.$$

8. Zmienną Y o rozkładzie logarytmiczno-normalny otrzymujemy ze zmiennej X o rozkładzie normalnym przekształceniem $g(x) = \exp(x)$.

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t}\exp(-\frac{1}{2\sigma^2}(\log(t) - \mu)^2) \text{ dla } t > 0.$$

9. Niech X ma rozkład gamma z parametrami $b > 0$ i $p > 0$. Zmienna losowa $Y = g(X) = \exp(X)$ ma rozkład logarytmiczno gamma o gęstości

$$f(t) = \frac{b^p}{\Gamma(p)}(\log t)^{p-1}t^{-b-1} \text{ dla } t \geqslant 1.$$

Dodatkowe wiadomości na temat wprowadzonych w tym rozdziale rozkładów można znaleźć w książkach [2] i [1].

1.2 Funkcja średniego nadmiaru

W wielu zastosowaniach wykorzystywana jest funkcja średniego nadmiaru (patrz [1] str.37) (średniego pozostałego czasu życia, średniej nadwyżki) $e(a)$ definiowana jako

$$e(a) = \mathbf{E}(X - a | x > a),$$

przy modelu analizowanej wielkości danej zmienną losową X i założeniu, że $\mathbf{E}(X) < \infty$.

Ważnym w praktyce zadaniem jest estymacja funkcji $e(a)$ na podstawie próby reprezentacyjnej $x_1, x_2, \ldots, x_n$. Estymator $\hat{e}_n$ określamy jako

$$\hat{e}_n(a) = \frac{\sum_{i=1}^n x_i \mathbb{I}_{(a,\infty)}(x_i)}{\#\{i : x_i > a\}} - a.$$

Często wykres empirycznej funkcji $\hat{e}_n(a)$ wykonywany jest w punktach $a = x^*_{n-k}$, $k = 1, 2, \ldots, n-1$. Wówczas

$$E_{k,n} := \hat{e}_n(x^*_{n-k}) = \frac{1}{k} \sum_{j=1}^{k} x^*_{n-j+1} - x^*_{n-k}.$$

$E_{k,n}$ można interpretować jako ocenę tangensa nachylenia wykresu kwantylowego na prawo od punktu o współrzędnych $(-\log(\frac{k+1}{n}), x^*_{n-k})$. Nachylenie to oceniamy jako

$$\tilde{E}_{k,n} = \frac{\frac{1}{k}\sum_{j=1}^k x^*_{n-j+1} - x^*_{n-k}}{-\frac{1}{k}\sum_{j=1}^k \log(\frac{j}{n+1}) + \log(\frac{k+1}{n+1})}. \tag{2}$$

Przy założeniu, że $\mathbf{E}(X) < \infty$ mamy, że $\lim_{x\to\infty} x(1 - F(x)) = 0$ i

$$e(a) = \frac{\int_a^{x_+} (1 - F(u)) du}{1 - F(a)}.$$

x_+ to prawy koniec nośnika rozkładu $F(u)$. Mamy ponadto

$$\int_0^a \frac{1}{e(u)} du = \log e(0) - \log((1 - F(a))e(a)),$$

czyli

$$1 - F(a) = \frac{e(0)}{e(a)} \exp(-\int_0^a \frac{1}{e(u)} du).$$

1.3 Zadania

1. Napisać procedury obliczające wartości dystrybuant oraz obliczające kwantyle dla funkcji z 1.1.

2. Napisać procedury rysujące gęstości rozkładów wprowadzonych w 1.1.

3. Skonstruować procedurę pozwalającą symulować stan konta instytucji finansowej na koniec dnia. Zakładamy, że zaczynamy ze stanem kasy 0, wysokość pojedynczych operacji jest zmienną losową X_i o rozkładzie zadanym dystrybuantą $F(x)$. Liczba operacji N w ciągu dnia jest zmienną losową niezależną od X_i, $i = 1, 2, \ldots$.

   Przeprowadzić symulacje, gdy

   (a) N ma rozkład Poissona ze znanym parametrem λ a X_i mają rozkład normalny o znanej średniej i wariancji.

   (b) N ma rozkład Poissona z rozkładem λ zadanym rozkładem gamma o znanych parametrach $b > 0$ i $p > 0$, a wielkości operacji X_i jak w pkt. 3a.

   (c) Przeanalizować stan konta na koniec dnia w funkcji parametrów rozkładów występujących w zastosowanych modelach w pkt. 3a i 3b.

4. Dla wprowadzonych w rozdziale 1.1 rozkładów zaproponować funkcje generujące zmienne losowe o tych rozkładach. Wykonać symulacje za pomocą skonstruowanych generatorów. Dla różnych wartości parametrów testem graficznym ocenić jakość symulacji.

5. Zilustrować centralne twierdzenie graniczne z wykorzystaniem symulacji zmiennych losowych. Wybrać jeden z rozkładów, którego generator zaprogramowano w ramach ćwiczeń. Dobrać przykłady tak, aby pokazać istotność założeń.

   Wskazówka: Jako przykład zmiennej nie mającej wartości oczekiwanej przyjąć zmienną o rozkładzie Cauchy'ego.

6. * Wykorzystując rezultaty zadania 4 przeprowadzić estymację funkcji średniego nadmiaru określonej w rozdziale 1.2 dla rozkładów z rozdziału 1.1.

Literatura

[1] Jan Beirlant, Jozef L. Teugels, and Petra Vynckier. *Practical Analysis of Extreme Values.* Leuven University Press, Leuven, 1996.

[2] Bolesław Kopociński. *Zarys teorii odnowy i niezawodności.* PWN, Warszawa, 1973.