RPiS

Generator liczb losowych z rozkładu $\mathbf{N}(0,1)$

Przykład:

Załóżmy, że mamy do dyspozycji generator liczb losowych z rozkładu $U[0,1]$ i wylosowaliśmy dwie wartości u_1, u_2. Dwuwymiarowa zmienna losowa (U_1, U_2) ma zatem rozkład o gęstości $f_{U_1,U_2}(u_1,u_2) = 1$ dla $((u_1,u_2) \in [0,1]\times[0,1]$. Rozważmy nowe zmienne $Y_1 = -2\ln U_1,\ Y_2 = 2\pi U_2$. Oczywiście $Y_1 \in [0,\infty)$ oraz $Y_2 \in [0, 2\,\pi)$. Interpretując Y_1, Y_2 jako współrzędne biegunowe punktu na płaszczyźnie można powiedzieć iż losujemy kwadrat promienia i argument punktu. Wyznaczmy gęstość $f_{Y_1,Y_2}(y_1,y_2)$ zmiennej (Y_1,Y_2).

$$\begin{cases} U_1 = \exp\left(-\dfrac{Y_1}{2}\right) \\ U_2 = \dfrac{Y_2}{2\,\pi} \end{cases}, \text{ skąd } \operatorname{abs}(J) = \operatorname{abs}\begin{vmatrix} -\dfrac{1}{2}\exp\left(-\dfrac{Y_1}{2}\right) & 0 \\ 0 & \dfrac{1}{2\,\pi} \end{vmatrix} = \frac{1}{4\cdot\pi}\exp\left(-\frac{Y_1}{2}\right). \tag{1}$$

W powyższym wzorze chcemy policzyć wartość bezwzględną z wyznacznika Jacobianu. Niestety, obydwie operacje (wartość bezwzględna i wyznacznik) oznaczane są często tym samym znakiem $|\ |$. Wskutek tego: $\operatorname{abs}(\det(A)) \equiv ||A||$ – co z kolei mogłoby sugerować, że mówimy o **normie** macierzy A.
Dla gęstości $f_{Y_1,Y_2}(y_1,y_2)$ mamy zatem wzór

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{4\cdot\pi}\exp\left(-\frac{Y_1}{2}\right). \tag{2}$$

Od współrzędnych biegunowych (Y_1,Y_2) przejdźmy teraz do współrzędnych kartezjańskich (X_1,X_2), to znaczy

$$\begin{cases} X_1 = \sqrt{Y_1}\,\cos Y_2 \\ X_2 = \sqrt{Y_1}\,\sin Y_2 \end{cases}, \text{ i stąd } J = \begin{vmatrix} \dfrac{\cos Y_2}{2\sqrt{Y_1}} & -\sqrt{Y_1}\sin Y_2 \\ \dfrac{\sin Y_2}{2\sqrt{Y_1}} & \sqrt{Y_1}\cos Y_2 \end{vmatrix} = \frac{1}{2}. \tag{3}$$

Kończymy przekształcenia dwiema uwagami:

1. Wyznaczony powyżej Jacobian należy ODWRÓCIĆ. Zwyczajowo liczymy Jacobian "starych" zmiennych względem "nowych". Tutaj: wygodniej jest wyznaczyć odwrotność Jacobianu "nowych" zmiennych względem "starych" zmiennych.
2. Korzystamy również z zależności: $Y_1 = X_1^2 + X_2^2$ [a].

Wynik końcowy ≡ gęstość $f_{X_1,X_2}(x_1,x_2)$ ma postać:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_1^2}{2}\right)\times\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_2^2}{2}\right), \tag{4}$$

co oznacza, że zmienne X_1, X_2 są niezależne i mają rozkład $\mathrm{N}(0,1)$ każda.

[a] (wzór (3))

Różne przybliżenia wariancji

Zakładamy, że dane są niezależne obserwacje pochodzące z tego samego rozkładu $\mathrm{N}(\mu, \sigma^2)$. Wiadomo,[1] że zmienna $\dfrac{nS_\mu^2}{\sigma^2}$ ma rozkład $\chi^2(n)$ a także[2] $\dfrac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$.
Wspomnijmy również, że $\chi^2(k) \equiv \mathrm{Gamma}(1/2, k/2)$. Stąd, dla rozkładu $\chi^2(k)$: $M_{\chi^2(k)}(t) = (1-2t)^{-k/2}$.
To umożliwia sformułowanie – w miarę prostego do dowodu – twierdzenia

Twierdzenie 1. *Jeżeli zmienna losowa $X \sim \chi^2(k)$, to $\mathrm{E}(X) = k$ oraz $\mathrm{V}(X) = 2k$.*

Rozpatrzmy trzy estymatory wariancji σ^2, mianowicie $S_{n-1}^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$, $S_n^2 = \dfrac{1}{n}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$ oraz $S_{n+1}^2 = \dfrac{1}{n+1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$. Z początkowej uwagi wynika, że zmienne $\dfrac{(n-1)S_{n-1}^2}{\sigma^2}$, $\dfrac{nS_n^2}{\sigma^2}$, $\dfrac{(n+1)S_{n+1}^2}{\sigma^2}$ mają rozkład $\chi^2(n-1)$ każda.

Wartość oczekiwana $\mathrm{E}\left(\dfrac{nS_n^2}{\sigma^2}\right) = n-1$. Stąd, $\mathrm{E}(S_n^2) = \dfrac{n-1}{n}\sigma^2$. Wartość oczekiwana przybliżenia S_n^2 jest zatem różna od wartości przybliżanego parametru σ^2. O takim estymatorze mówimy, że jest **estymatorem obciążonym**. Jeżeli $n \to \infty$, to $\mathrm{E}(S_n^2) \to \sigma^2$. Mówimy wówczas, że S_n^2 jest estymatorem **asymptotycznie nieobciążonym** parametru σ^2.

Ponieważ $S_{n-1}^2 = \dfrac{n}{n-1}S_n^2$, więc $\mathrm{E}(S_{n-1}^2) = \sigma^2$. Mówimy, że S_{n-1}^2 jest **nieobciążonym estymatorem** parametru σ^2. Podobnie $S_{n+1}^2 = \dfrac{n}{n+1}S^2$ czyli $\mathrm{E}(S_{n+1}^2) = \dfrac{n-1}{n+1}\sigma^2$ (S_{n+1}^2 jest estymatorem obciążonym dla σ^2). Najlepszym estymaterem (uwzględniając wartość oczekiwaną) jest zatem S_{n-1}^2, najgorszym S_{n+1}^2.

Porównajmy teraz wariancje rozpatrywanych estymatorów. Wiemy, że $\dfrac{n\,S_n^2}{\sigma^2} \sim \chi^2(n-1)$. Stąd wynika, że $\mathrm{V}(S_n^2) = \dfrac{2(n-1)}{n^2}\sigma^4$, $\mathrm{V}(S_{n-1}^2) = \dfrac{2(n-1)}{(n-1)^2}\sigma^4$ oraz $\mathrm{V}(S_{n+1}^2) = \dfrac{2(n-1)}{(n+1)^2}\sigma^4$. Im mniejsza wariancja, tym bardziej zmienna losowa jest "stabilna". Najlepszym estymatorem (kierując się wariancją) jest S_{n+1}^2, najgorszym S_{n-1}^2.

[Popularne|Ulubione] wzory i rozkłady

1. Załóżmy, że zmienne losowe X, Y są niezależne i podlegają rozkładom $X \sim \chi^2(n)$, $Y \sim \chi^2(k)$. Wówczas zmienna losowa $Z = X + Y$ podlega rozkładowi $Z \sim \chi^2(n+k)$.
2. Załóżmy, że zmienna X podlega rozkładowi $\mathrm{N}(\mu, \sigma^2)$. Niech dodatkowo $Y = \dfrac{X-\mu}{\sigma}$. Zachodzi Fakt: $X \sim \mathrm{N}(\mu, \sigma^2) \iff Y \sim \mathrm{N}(0,1)$.
3. $\mathrm{Gamma}\,(1/2, n/2) \equiv \chi^2(n)$.
4. Załóżmy, że zmienne $X_1, \ldots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu, \sigma^2)$ każda. Wówczas zmienna $Z = \sum_{k=1}^{n}\left(\dfrac{X_k-\mu}{\sigma}\right)^2$ ma rozkład $\chi^2(n)$.
5. Niezależne zmienne X, Y mają rozkłady $X \sim \chi^2(k)$, $Y \sim \chi^2(l)$ odpowiednio. Mówimy, że zmienna $F(k,l) = \dfrac{X}{Y}\cdot\dfrac{l}{k}$ ma rozkład F-Fishera z (k,l) stopniami swobody.
6. Niezależne zmienne X, Y mają rozkłady $X \sim \mathrm{N}(0,1)$, $Y \sim \chi^2(k)$ odpowiednio. Mówimy, że zmienna $t(k) = \dfrac{X}{\sqrt{Y/k}}$ ma rozkład t-Studenta z k stopniami swobody.

---

[1] Notatka 6, wzór (3).
[2] Notatka 6, tw. 4.

7. Intuicja: iloraz niezależnych i normalizowanych rozkładów χ^2 to rozkład F-Fishera zaś iloraz standardowego rozkładu normalnego i pierwiastka normalizowanego rozkładu χ^2 to rozkład t-Studenta.
8. Załóżmy, że zmienne $X_1, \dots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu, \sigma^2)$ każda. Niech dodatkowo $S_\mu^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \mu)^2$. Wówczas $\frac{nS_\mu^2}{\sigma^2} \sim \chi^2(n)$.
9. Załóżmy, że zmienne $X_1, \dots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu, \sigma^2)$ każda. Niech dodatkowo $S^2 = \frac{1}{n} \sum_{k=1}^{n} \left(X_k - \bar{X}\right)^2$. Wówczas $\frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$.
10. ...tysiąc i jeden wzór (jak w orientalnych baśniach).

Witold Karczewski