

## Metoda Monte-Carlo

Są przypadki kiedy zamiast wykonać jakiś eksperyment chcielibyśmy symulować jego wyniki używając komputera i generatora liczb (pseudo)losowych. Większość bibliotek programów komputerowych zawiera procedury numeryczne generujące liczby losowe o rozkładzie równomiernym w przedziale 0 do 1  $r \in \langle 0,1 \rangle$ . Stosując odpowiednie algorytmy możemy takie liczby przekształcić w liczby losowe o dowolnym rozkładzie.

Założmy, że chcemy generować liczby losowe o pewnym rozkładzie (gęstości) prawdopodobieństwa

$$f(x)$$

Dystrybuantą  $F(x)$  zmiennej losowej nazywamy funkcję

$$F(x) \equiv P(t \leq x) = \int_{-\infty}^x f(t) dt,$$

której wartości są równe prawdopodobieństwu, że wartość  $t$  zmiennej losowej nie przekracza  $x$ . Jeżeli

$$y = F(x),$$

to oczywiście  $y \in \langle 0,1 \rangle$  i prawdopodobieństwo, że, na przykład

$$y > 0,6 \text{ i } y \leq 0,7$$

wynosi 0,1 (10%) i jest takie samo dla każdego innego przedziału o szerokości 0,1 zawartego w  $\langle 0,1 \rangle$ . Oznacza to, że transformacja

$$y = F(x)$$

przekształca zmienną losową  $x$  w zmienną  $y$  o rozkładzie równomiernym w przedziale  $\langle 0,1 \rangle$ . Wynika z tego dalej, że transformacja odwrotna

$$x = F^{-1}(r)$$

przekształca zmienną losową o rozkładzie równomiernym  $r \in \langle 0,1 \rangle$  w zmienną  $x$  o żądanym rozkładzie  $f(x)$ .

*Przykład.*

Zmienną o rozkładzie wykładniczym

$$f(t) = \frac{e^{-t/\tau}}{\tau}$$

możemy otrzymać stosując transformację

$$t = -\tau \ln(r).$$

Wynika to z faktu, że dystrybuanta zmiennej wykładniczej ma postać

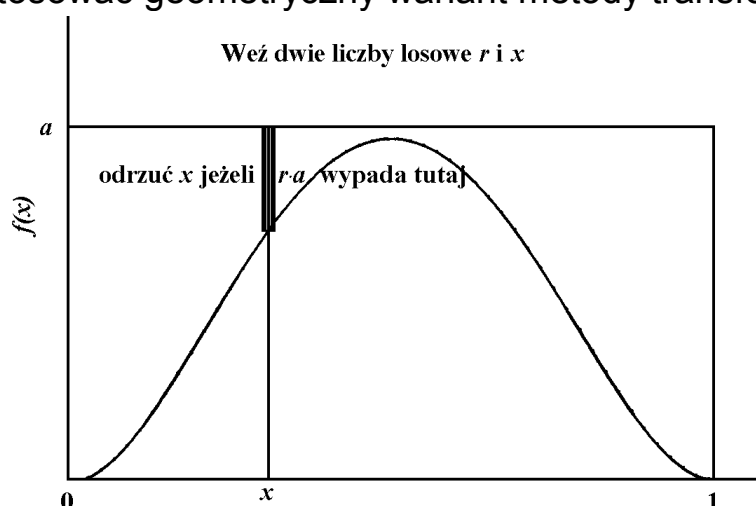
$$F(t) = \int_0^t f dt = 1 - e^{-t/\tau},$$

co dalej oznacza, że zarówno  $r = 1 - e^{-t/\tau}$  jak i  $r = e^{-t/\tau}$  mają rozkład równomierny w przedziale  $\langle 0,1 \rangle$ . Zatem rzeczywiście jeżeli  $r = e^{-t/\tau}$  ma rozkład równomierny, to  $t = -\tau \ln(r)$  ma rozkład wykładniczy.

Metoda transformacji wykorzystująca funkcję odwrotną do dystrybuanty pożądanego rozkładu jest zupełnie ogólna i możemy ją praktycznie stosować bez kłopotów wszędzie tam, gdzie znana jest analityczna postać funkcji odwrotnej  $F^{-1}$ . Jeżeli postać analityczna nie jest znana, to rozkład  $f(x)$  możemy całkować numerycznie i znaleźć w ten sposób rozwiązanie równania

$$r = F(x).$$

Jeżeli całkowanie numeryczne jest niemożliwe lub niepraktyczne, to możemy zastosować geometryczny wariant metody transformacji.



Rozważmy w tym celu rozkład  $f(x)$  o wartościach niezerowych w przedziale  $\langle 0,1 \rangle$  i taki, że  $f(x) < a$  w tym przedziale. Generujemy parę liczb losowych  $(x, r)$  o rozkładzie równomiernym w przedziale  $\langle 0,1 \rangle$ . Jeżeli  $r \cdot a \leq f(x)$ , to wartość  $x$  pozostawiamy, w przeciwnym przypadku odrzucamy. Pozostawione wartości  $x$  tworzą zbiór o pożądanym rozkładzie  $f(x)$ .

Generowanie normalnych liczb losowych.

1. Jeżeli weźmiemy parę liczb losowych  $(x_1, x_2)$  o rozkładzie równomiernym w przedziale  $(0,1)$ , to przekształcenie

$$y_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2)$$

$$y_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2)$$

daje parę liczb losowych  $(y_1, y_2)$  niezależnych o rozkładzie normalnym  $N(0,1)$ .

2. Wartość średnia i dyspersja zmiennej o rozkładzie równomiernym w przedziale  $(0,1)$  wynoszą

$$\mu = 0,5$$

$$\sigma = \frac{1}{\sqrt{12}} \approx 0,288675\dots$$

Zatem zmienna  $z$  równa sumie 12 takich liczb losowych minus 6

$$z = \sum_{i=1}^{12} x_i - 6$$

ma średnią

$$\mu_z = 0$$

i dyspersję

$$\sigma_z = 1$$

i na mocy twierdzenia granicznego ma rozkład zbliżony do normalnego.

## Testowanie zgodności średnich i wariancji

Mamy dwa niezależne zbiory wyników pomiarów tej samej wielkości:

$$\{x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}\} \text{ oraz } \{x_{21}, x_{22}, x_{23}, \dots, x_{2n_2}\}$$

o liczebnościach odpowiednio  $n_1$  i  $n_2$ .

Każdy z nich charakteryzuje się swoją średnią i wariancją:

$$\mu_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i} \quad s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \mu_1)^2$$

$$\mu_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i} \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \mu_2)^2$$

Chcemy wiedzieć czy różnice między średnimi i wariancjami są statystycznie istotne.

### 1. Test zgodności wariancji

Zmienna  $F$  ma wartość:

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}.$$

Przy założeniu, że prawdziwa jest hipoteza zerowa

$$\sigma_1^2 = \sigma_2^2$$

$F$  staje się równe

$$F = \frac{s_1^2}{s_2^2}$$

i ma rozkład  $F$  (Fishera-Snydecora) o liczbach stopni swobody  $n_1 - 1$  i  $n_2 - 1$ .

Jeżeli poziom istotności ma wynosić  $\alpha$  (zwykle 0,05), to

$$P(F(n_1 - 1, n_2 - 1, \alpha/2) \leq F \leq F(n_1 - 1, n_2 - 1, 1 - \alpha/2)) = 1 - \alpha$$

prawdopodobieństwo, że przypadkowo wartość  $F$  wyjdzie poza przedział ufności wynosi  $\alpha$ .

Jeżeli wyliczona dla naszych zbiorów wartość  $F$  nie wypada poza przedziałem ufności, to hipotezę o równości wariancji przyjmujemy. W przeciwnym przypadku odrzucamy ją i przyjmujemy alternatywną o istotnej różnicy wariancji.

## 2. Test zgodności średnich dla równych wariancji

Jeżeli wariancje obu ciągów wyników są jednakowe, to obliczamy średnią ważoną wariancję

$$s_D^2 = \frac{\sum_{i=1}^{n_1} (x_{1i} - \mu_1)^2 + \sum_{i=1}^{n_2} (x_{2i} - \mu_2)^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right).$$

Zmienna

$$t = \frac{\mu_1 - \mu_2}{s_D}$$

ma rozkład Studenta o  $n_1 + n_2 - 2$  stopniach swobody w przypadku kiedy oba ciągi pochodzą z tej samej populacji (mają te same rozkłady).

Jeżeli hipotezę zerową o równości średnich  $\mu_1 = \mu_2$  testujemy na poziomie istotności  $\alpha$ , to szukamy przedziału ufności

$$P(t(n_1 + n_2 - 2, \alpha/2) \leq t \leq t(n_1 + n_2 - 2, 1 - \alpha/2)) = 1 - \alpha.$$

Jeżeli wartość statystyki  $t$  wypada w tym przedziale, to przyjmujemy hipotezę zerową  $\mu_1 = \mu_2$ . W przeciwnym przypadku odrzucamy ją i uznajemy średnie za istotnie różne.

## 3. Test zgodności średnich dla różnych wariancji

W takim przypadku wielkość

$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

ma w przybliżeniu rozkład Studenta o liczbie stopni swobody danej wzorem

$$\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{[s_1^2/n_1]^2}{n_1 - 1} + \frac{[s_2^2/n_2]^2}{n_2 - 1}},$$

którą zaokrąglamy w dół do całkowitej, żeby móc skorzystać z odpowiednich tablic.

## Monte-Carlo 2

The model of generating data points.

Using "artificial" data points generated by the computer programme it is possible to test the proposed techniques and to compare their results. It is also possible to prove the adequacy of formulae giving the estimations of regression coefficients and their dispersions.

In this work data were generated according to the simple linear model of the TL growth curve. It was assumed that

$$I = I_0 + b \cdot D,$$

where:  $I$  denotes the measured TL intensity growing linearly with the additional dose  $D$ ,  $I_0$  is the initial (natural) level of TL intensity and  $b$  is the proportionality coefficient (TL sensitivity).

In the following text the above equation will be used in the rewritten form:

$$y = a + b \cdot x$$

and the values of  $a$  and  $b$  were chosen to be equal 1

$$a = 1, \quad b = 1$$

It was assumed that  $y$  values were measured for five chosen  $x$  values, namely:

$$X_i = i - 1, \quad i = 1, 2, 3, 4, 5$$

and that these measurements were repeated five times for each  $X_i$  giving 25 values total.

The resulting values were obtained as follows:

$$Y_{ij} = 1 + X_i + \varepsilon_{1j} + \varepsilon_{2j}, \quad i, j = 1, 2, 3, 4, 5$$

or

$$Y_{ij} = i + \varepsilon_{1j} + \varepsilon_{2j}, \quad i, j = 1, 2, 3, 4, 5$$

where  $\varepsilon$ 's are the pseudo random numbers taken from the normal distribution  $N(0, \sigma)$ .

That means that the model values resulting from the eqn. :

$$y_i = 1 + X_i = i$$

are considered as the expected values of  $Y_{ij}$  and the added pseudo random numbers  $\varepsilon$  simulate the measurement error. It was assumed no systematic error in  $Y_{ij}$  and that the overall error is constituted from two components. These are the absolute error which doesn't depend on the  $y$  value and the relative error proportional to  $y$ . The values  $X_i$  were assumed not to be subject to an error.

The programme was written in BASIC for the IBM PC/XT compatible computer and the internal random generator was used. It gave the

pseudo random numbers  $r$  uniformly distributed in the interval  $(0,1)$ . The simple linear transformation

$$\eta = \sum_{i=1}^{12} r_i - 6$$

yielded pseudo random numbers  $\eta$  having the distribution which approximates the normal  $N(0,1)$  distribution with the accuracy good enough for the purpose of the present work (Muller, 1959). Numbers  $\eta$  obtained in the described way were then used to simulate measurement errors  $\varepsilon_1$  and  $\varepsilon_2$ . The fixed absolute error with the standard deviation  $\sigma$  is equal to

$$\varepsilon_1 = \sigma \cdot \eta$$

and the fixed relative error ( $\delta$  %) with the standard deviation proportional to the measured value  $y$  is given by

$$\varepsilon_2 = \frac{\delta}{100} (1 + X_i) \cdot \eta .$$

Therefore the final results

$$Y_{ij} = i + \sigma \cdot \eta' + \frac{\delta}{100} \cdot i \cdot \eta'' = i \cdot \left( 1 + \frac{\delta}{100} \cdot \eta'' \right) + \sigma \cdot \eta'$$

where  $\eta'$  and  $\eta''$  are independent numbers generated for each needed  $Y_{ij}$  value.

In the authors opinion the set of 25 "measurement results" generated in this way corresponds well to the situation of actual TL measurements. In practice the random errors of doses can be neglected as compared to those of TL intensities and regarding the measurement errors as resulting from two sources corresponding to eps1 and eps2 components seems reasonable.

---

Furthermore it can be noted that doses are not individually measured thus any random errors in their assessments propagate finally to the apparent TL intensity errors and such a situation is also included in the proposed model.