ГОДИШНИК НА МИННО-ГЕОЛОЖКИЯ УНИВЕРСИТЕТ “СВ. ИВАН РИЛСКИ”, Том 55, Св. II, Добив и преработка на минерални суровини, 2012
ANNUAL OF THE UNIVERSITY OF MINING AND GEOLOGY “ST. IVAN RILSKI”, Vol. 55, Part II, Mining and Mineral processing, 2012

ИЗСЛЕДВАНЕ НА ВЪЗМОЖНИТЕ ГРЕШКИ ПРИ ЗАМЯНА НА ЕДНО ВЕРОЯТНОСТНО РАЗПРЕДЕЛЕНИЕ С ДРУГО ПРИ РЕШАВАНЕ НА ЗАДАЧИ ОТ ДОБИВА НА ПОЛЕЗНИ ИЗКОПАЕМИ

Георги Трапов

Минно-геоложки университет „Св. Иван Рилски”, 1700 София; Trapov@abv.bg

РЕЗЮМЕ. Известно е, че при изследване на процеси и явления с вероятностен характер, се стремим да достигнем до използване на някои от известните разпределения. Това е голямо предимство тъй като техните свойства са добре изучени. При това много често се опитваме да използваме нормалното разпределение. Но това крие доста голям риск от допускане на грешка ако е направено без извършване на подходящи предварителни изследвания. В настоящата разработка се прави опит да се оцени тази грешка при положение, че е налице равномерно разпределение, но по някакви причини е възприет модела на нормалното разпределение. Целта не е да се омаловажи значението на нормалното разпределение, а да се посочат аргументи против преувеличаване на ролята му и да се покаже необходимостта от подходяща обосновка, преди да се използва който и да е от известните вероятностни закони. За съжаление често недостига на натрупани данни от наблюдения и липсата на теоретични съображения не ни дава възможност да използваме известните строги статистически критерии за избор на разпределение. Прилагането им в такива случаи е по-скоро опит да се прикрие липсата на информация.

STUDY OF POSSIBLE ERRORS WHEN WE REPLACE DEFINITELY PROBABILITY DISTRIBUTION WITH OTHER WHEN SOLVING PROBLEMS OF MINING

Georgi Trapov
University of Mining and Geology “St. Ivan Rilski”, 1700 Sofia; Trapov@abv.bg

ABSTRACT. It is known that in the study of processes and phenomena with probabilistic nature, strive to reach using some of the known distributions. This is a great advantage because their properties have been well studied. In this very often try to use the normal distribution. But on of this lies much high risk of errors if it is done without making the proper preliminary research. In this paper an attempt is made to assess this error if there is a uniform distribution, but for some reason has adopted the model of the normal distribution. The goals is not to minimize the importance of the normal distribution, but stated arguments against exaggerating its role and show the need for appropriate justification before using any of the known probability laws. Unfortunately, often the lack of accumulated observational data and the lack of theoretical considerations allow us not to use certain from statistical criteria for selection of distribution. Their application in such cases is rather an attempt to conceal the lack of information.

Въведение

При изследване на процеси и явления с вероятностен характер, се стремим да достигнем до използване на някои от известните теоретични разпределения. Това е голямо предимство тъй като техните свойства са добре изучени. При това много често се опитваме да използваме нормалното разпределение. Когато имаме достатъчен брой наблюдения, можем успешно да изберем теоретично вероятностно разпределение, което достатъчно добре да обяснява свойствата на изследваната величина. Но ако наблюденията са малко, при избор на което и да е вероятностно разпределение рискуваме да допуснем значителна грешка. Тук се прави анализ и оценка на грешката тогава, когато са налице съображения, че определена величина има вероятностно разпределение различно от нормалното, но ние я заменяме със съответна на нея нормално разпределена случайна величина. Тази грешка в много случай довежда да значително отклоняване на изчислените стойности на изследваните величини - производителност, коефициент на устойчивост и др.

Големите предимства да се използва нормалното разпределение са добре известни. Ето обаче съображения, които ни карат да сме с повишено внимание:
- Много често за изследваните величини нямаме теоретични предпоставки или натрупани предишни наблюдения, които убедително да сочат, че е налице нормално разпределение. Нещо повече, редица автори като В.Т. Глушков и др. [1], З.Д. Низгуредкий [2], Н.Г. Петров [3], W. Foster, E. Weber [4], O.G. Oswald [5] и др. показват емпирично, че законът на разпределение на вероятностите например за съдържанието на полезен компонент в руда или за определени свойства на скалите е твърде далеч от нормалния закон и може да се апроксимира много по-добре с други разпределения;
- На практика изследваните величини много често (например физико-механични свойства на скалите) , за разлика от нормално разпределените, имат само положителни стойности;
- Дори и тогава, когато от теоретични съображения знаем, че дадена величина има нормално разпределение, за да осигурим представителна грешка на наблюденията

например не е по-голяма от $\frac{\sigma}{5}$ (което при $\sigma = 1$ е 20%), която да бъде гарантирана с вероятност $\gamma = 0{,}9$ ще трябва да имаме извадка с обем поне

$$n_{\min} = \frac{Z^2_{\frac{1+\gamma}{2}}.\sigma^2}{\delta^2} \geq \frac{Z^2_{\frac{1+0{,}9}{2}}.\sigma^2}{\left(\frac{\sigma}{5}\right)^2} = 1{,}64^2.25 = 67{,}24,$$

където Z_α е квантил от ниво α на стандартното нормално разпределение.

И така броят на наблюденията трябва да е поне 68! Но това е твърде голямо число. Лабораторните и особено някои натурни наблюдения са или невъзможни, или много скъпи.

Например:

- Ние не можем да си позволим да предизвикаме многократно свлачище, за да направим необходимите ни наблюдения;
- Провеждането на необходимите сондажи за изследване на скалната ядка е свързано с разход на много ресурси, включително и финансови;
- Изпитанието в лаборатория на един скален образец струва доста пари и обем на извадка от порядъка на 5 до 10 ни се струва твърде много, а както видяхме от теоретични съображения броят на наблюдаваните образци е желателно да бъде поне 25 - 30, а най добре би било да е два или три пъти повече!

Оценка на грешката при замяна на равномерно разпределение с нормално

В настоящата разработка ще проследим какво се получава тогава, когато са налице съображения, че определена величина η има равномерно разпределение с възможни стойности в интервал $[\alpha, \beta]$, но ние я заменяме със съответна на нея нормално разпределена случайна величина ξ.

Тъй като разпределението на величината η е равномерно, то вероятността нейните стойности да са извън интервала $[\alpha, \beta]$ е равна на нула. Но когато например от извадка намерим оценки за математическото очакване и дисперсията (съответно a и σ^2) и минаваме не към равномерно а към нормално разпределение, ние „улавяме" трисигмувия интервал, извън който стойностите на нормално разпределената случайна величина попадат с вероятност почти нула. С други думи връзката между параметрите на двете съответни едно на друго разпределения се дава с равенствата

$$\left|\begin{array}{l}\alpha = a - 3\sigma \\ \beta = a + 3\sigma\end{array}\right. .$$

Следователно по /2.7/ законът на разпределение на величината η можем да запишем във вида

$$p_\eta(x) = \begin{cases} \frac{1}{6\sigma}, & x \in [a-3\sigma, a+3\sigma] \\ 0, & x \notin [a-3\sigma, a+3\sigma] \end{cases},$$

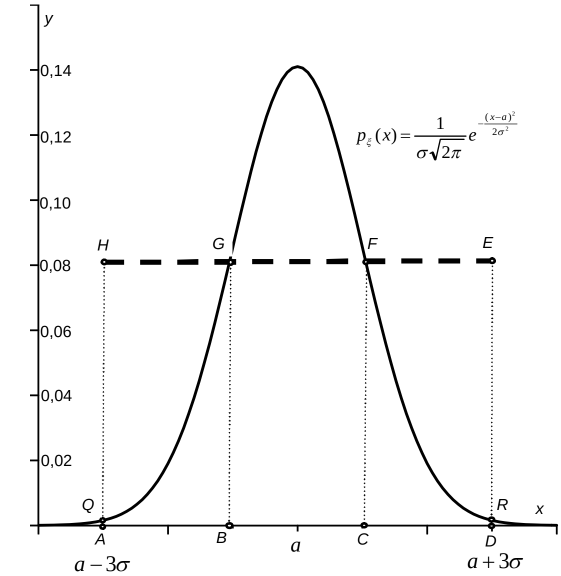
от където намираме:

$$E\eta = \frac{\alpha+\beta}{2} = \frac{a-3\sigma+a+3\sigma}{2} = a;$$

$$D\eta = \frac{(\beta-\alpha)^2}{12} = 3\sigma^2;\ \sigma_\eta = \sigma\sqrt{3}.$$

И така $E\eta = a = E\xi$, $D\eta = 3\sigma^2 = 3D\xi$, $\sigma_\eta = \sigma_\xi\sqrt{3}$.

На фигура 1 са дадени графиките на законите на разпределение на ξ - камбановидната линия, и на η - отсечката *HE* е частта от графиката на закона на разпределение на величината η в интервала $[a-3\sigma; a+3\sigma]$. Следователно тя е успоредна на оста *Ox* и е на височина $\frac{1}{6\sigma}$. Извън този интервал графиката на закона на η е по абсцисната ос.



Фиг. 1.

Решаваме системата уравнения

$$\left|\begin{array}{l} p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} \\ p_\eta(x) = \begin{cases} \frac{1}{6\sigma}, & x \in [a-3\sigma, a+3\sigma] \\ 0, & x \notin [a-3\sigma, a+3\sigma] \end{cases} \end{array}\right. \quad (1)$$

спрямо x, т. е. от равенството

$$\frac{1}{6\sigma} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$$

намираме абсцисите на точките *G* и *F*:

$$x_G = a - \sigma\sqrt{2}\ln\frac{3\sqrt{2\pi}}{\pi} = a - 1{,}234355205\sigma;$$
$$x_F = a + \sigma\sqrt{2}\ln\frac{3\sqrt{2\pi}}{\pi} = a + 1{,}234355205\sigma$$
.

Сега можем да посочим кои са координатите на точките $A(a-3\sigma;0)$, $B(x_G;0)$, $C(x_F,0)$, $D(a+3\sigma;0)$, $E\left(a+3\sigma;\frac{1}{6\sigma}\right)$, $F\left(x_F;\frac{1}{6\sigma}\right)$, $G\left(x_G;\frac{1}{6\sigma}\right)$, $H\left(a-3\sigma;\frac{1}{6\sigma}\right)$.

Да проследим каква грешка правим, ако заменим равномерно разпределение със съответното му нормално вероятностно разпределение. За целта например нека търсим вероятността разглежданата случайна величина да попадне в интервал $[\alpha;\beta]$, на който и двата края са наляво от точката $a-3\sigma$, или надясно от $a+3\sigma$. Тогава

$$P(\alpha \le \xi \le \beta) - P(\alpha \le \eta \le \beta) = P(\alpha \le \xi \le \beta) < \frac{1-0{,}9973}{2} = 0{,}00135,$$

т. е. в този случай грешката е нищожно малка. Но на практика търсенето на вероятност в такъв интервал е по-скоро изключение.

За по-обстоен анализ да разгледаме двете функции на разпределение, съответно на нормалното разпределение

$$F_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt;$$

и на равномерното разпределение

$$F_\eta(x) = \begin{cases} 0, & \text{при } x < a-3\sigma \\ \dfrac{x-(a-3\sigma)}{6\sigma}, & \text{при } x \in [a-3\sigma; a+3\sigma] \\ 1, & \text{при } x > a+3\sigma \end{cases}$$

Да образуваме разликата

$$f(x) = F_\xi(x) - F_\eta(x). \tag{2}$$

За да ползваме по-лесно таблиците за функцията на Лаплас $[\Phi(x)]$, или функцията на разпределение $[F_Z(x)]$ на стандартната нормално разпределена случайна величина правим замяната $x = a + t\sigma$.

Тогава

$$F_\xi(a+t\sigma) = F_Z\left(\frac{(a+t\sigma)-a}{\sigma}\right) = F_Z(t).$$

Като положим $f(a+t\sigma) = \varphi(t)$, разликата (2) добива вида

$$\varphi(t) = \begin{cases} F_Z(t), & \text{при } t < -3 \\ F_Z(t) - \dfrac{t+3}{6}, & \text{при } -3 \le t \le -1{,}234355 \\ F_Z(t) - \dfrac{t+3}{6}, & \text{при } -1{,}234355 < t \le +1{,}234355 \\ F_Z(t) - \dfrac{t+3}{6}, & \text{при } -1{,}234355 < t \le +3 \\ F_Z(t), & \text{при } t > +3 \end{cases}. \tag{3}$$

Не е трудно да се провери, че тази функция достига минимум в точката

$$x_G = x_1 = a - \sigma\sqrt{2}\ln\frac{3\sqrt{2\pi}}{\pi} = a - 1{,}234355205\sigma,$$

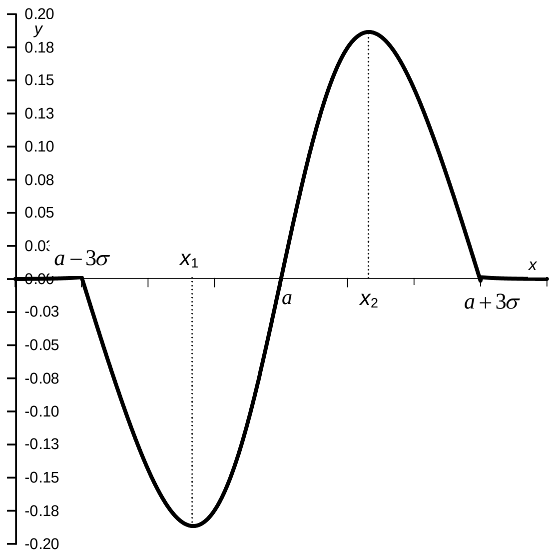
и максимум в точката

$$x_F = x_2 = a + \sigma\sqrt{2}\ln\frac{3\sqrt{2\pi}}{\pi} = a + 1{,}234355205\sigma,$$

като самите екстремуми са

$$\varphi_{\min} = \varphi(-1{,}234355205) = F_Z(-1{,}23) - \frac{1{,}765644769}{6} = 0{,}10935 - 0{,}29427 = -0{,}18492$$

$$\varphi_{\max} = \varphi(1{,}234355205) = F_Z(1{,}23) - \frac{4{,}23}{6} = 0{,}89065 - 0{,}70573 = 0{,}18492.$$



Фиг. 2.

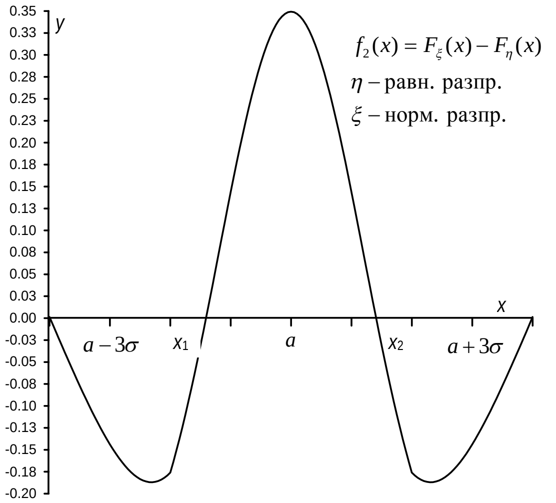
Този резултат показва, че абсолютната стойност на грешката, която се получава от замяната, достига около 18,5%, а осцилацията е приблизително 37%.

Графиката на функцията $f(x)$ е дадена на фигура 2.

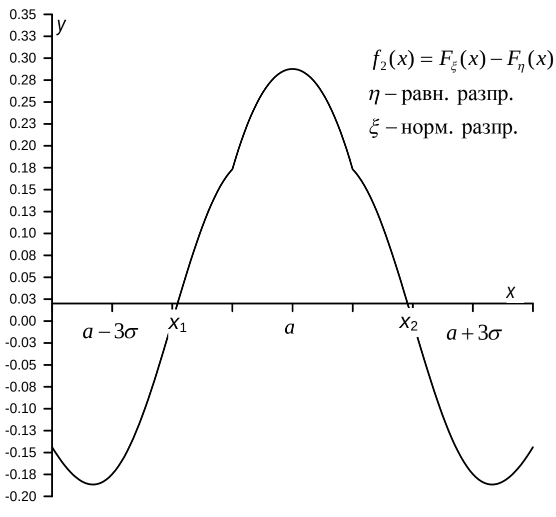
Ние на практика обикновено искаме да намерим вероятността случайна величина да попада в интервал с големина от σ до 6σ. Да означим средата на този интервал с x, а краищата му с $\alpha = x - t\sigma$, $\beta = x + t\sigma$. С фиксиране стойността на *t*, задаваме ширината на интервала, след което изменяме стойностите на *x*, които са средата на интервала, примерно отляво надясно. Така ние придвижваме и самия интервал. Интересно е при това как се изменя разликата от вероятностите

$$[P(\alpha \le \xi < \beta) - P(\alpha \le \eta < \beta)], \qquad (4)$$

тъй като тя показва каква е стойността на грешката, която правим когато осъзнато или не, заменяме едното разпределение с другото.



Фиг. 3.



Фиг. 4.

За целта използваме формулите

$$P(\alpha \le \xi < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) = \Phi\left(\frac{x + t\sigma - a}{\sigma}\right) - \Phi\left(\frac{x - t\sigma - a}{\sigma}\right),$$

$$P(\alpha \le \eta < \beta) = F_\eta(\beta) - F_\eta(\alpha) = F_\eta(x + t\sigma) - F_\eta(x - t\sigma).$$

Ще се спрем на три характерни случая за дължина на интервала.

При $t = 1$ дължината на интервала е 2σ. Тогава разликата (4) ще се даде с функцията

$$f_1(x) = \left[\Phi\left(\frac{x + \sigma - a}{\sigma}\right) - \Phi\left(\frac{x - \sigma - a}{\sigma}\right)\right] - \left[F_\eta(x + \sigma) - F_\eta(x - \sigma)\right].$$

Графиката на тази функция е дадена на фигура 3.

Интервал с дължина 4σ имаме, когато изберем $t = 2$. В този случай (4) добива вида

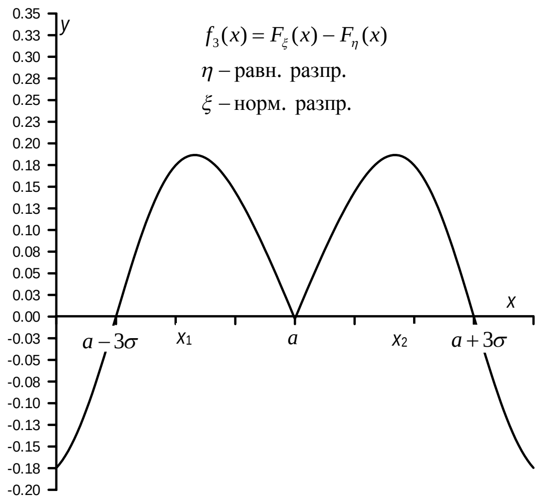
$$f_2(x) = \left[\Phi\left(\frac{x + 2\sigma - a}{\sigma}\right) - \Phi\left(\frac{x - 2\sigma - a}{\sigma}\right)\right] - \left[F_\eta(x + 2\sigma) - F_\eta(x - 2\sigma)\right]$$

Графиката на функцията е дадена на фигура 4.

При $t = 3$ дължината на интервала е 6σ. Тогава разликата от вероятностите (4) ще бъде

$$f_3(x) = \left[\Phi\left(\frac{x + 3\sigma - a}{\sigma}\right) - \Phi\left(\frac{x - 3\sigma - a}{\sigma}\right)\right] - \left[F_\eta(x + 3\sigma) - F_\eta(x - 3\sigma)\right].$$

Графиката на функцията е дадена на фигура 5.



Фиг. 5.

От разгледаните три случая се забелязва очакваната тенденция с увеличаване дължината на интервала, в който искаме да оценим вероятността, грешката да намалява. Това е така, защото когато разглеждаме по-големи интервали, в тях по-често се съдържат абсцисите на точките на пресичане на графиките на двете разпределения. В тези точки двете графики си разменят във вертикално отношение местата, което довежда до намаляване на абсолютната стойност на разликата

$[P(\alpha \leq \eta < \beta) - P(\alpha \leq \xi < \beta)]$. Въпреки това се забелязва, че грешката по абсолютна стойност пада от около 35% едва на 18%, т. е. остава твърде голямо число.

Заключение

По аналогичен начин могат да се оценят възможните грешки, при преминаване от което и да е вероятностно разпределение към нормално или към някое друго.

Като анализираме получените резултати, достигаме до извода, че при избор на теоретично вероятностно разпределение трябва да сме особено внимателни. Ако имаме достатъчен брой наблюдения, можем успешно да изберем теоретично вероятностно разпределение (може да се окаже и нормално), което достатъчно добре да обяснява свойствата на изследваната величина. Но ако наблюденията са малко на брой, то при избор на което и да е вероятностно разпределение рискуваме да допуснем значителна грешка, тъй като е твърде възможно да не попаднем на „истинското" разпределение. В такива случаи можем към натрупаните данни от наблюдения да добавим разумни съображения за поведението на изследвания процес и да изберем разпределение, което в много случаи може да се окаже различно от нормалното.

Литература

1. Глушков, В.Т. и др. Статистический метод обработки данных о прочностным свойств реалных горных пород. В сб. науч.тр.Н.-и.горнорудн.ин-т УССР, 1971, №16, 59-63 с.
2. Низгуредкий, З.Д. Учет неедноролностей при формировании статистических функций распределения геологических показателей. В сб. Маркшайд.дело, вып.1, Л., 1974 -135 с.
3. Петров, Н.Г. О параметрах распределения крепости угля и вмешающих пород. Уголь, 7, 1981, 7-9 с.
4. Forster, W., E. Weber. Influences of the probability of failures of slopes. "Soil Mech. and Found. Eng. Proc, 10 Int. Cons., Stockholm, 15-19 juni 1981, vdl.I" Rotterdam, 1981, 127-130 c.
5. Oswald. O.G. Wahrscheinlichkeitstheoretische Bet rachungen zur Ermittlung bodenmechanischer Kennwerte. Neue bergbautechnik, januar 1975, 17-21 c.