Упражнение 8'. Нормално разпределение.

Ако X е сл. величина с ф.р. $F_X(x) = P(X < x)$, плътността и е $f_x(x) = F'_X(x)$. Когато тази производна съществува за всяко x, сл. величина се нарича абсолютно непрекъсната.

Ако $g(x)$ е реална измерима функция, то $Y = g(X)$ е друга сл. величина.

I) Ако $g(x)$ е строго растяща, тогава съществува обратната функция $g^{-1}(.)$ и за функцията на разпределение на Y имаме

$$F_Y(x) = P(Y < x) = P(g(X) < x) = P(X < g^{-1}(x)) = F(g^{-1}(x)).$$

Ако g и g^{-1} са диференцируеми, то плътността на Y се изразява чрез плътността на X като

$$f_Y(x) = f_X(g^{-1}(x))[(g^{-1}(x))'] = \frac{f_X(g^{-1}(x))}{g'(x)}. \tag{1}$$

II) Ако $g(x)$ е строго намаляваща, тогава отново съществува обратната функция $g^{-1}(.)$ и за функцията на разпределение на Y имаме

$$F_Y(x) = P(Y < x) = P(g(X) < x) = P(X > g^{-1}(x)) = 1 - F(g^{-1}(x)).$$

Ако g и g^{-1} са диференцируеми, то плътността на Y се изразява чрез плътността на X като

$$f_Y(x) = -f_X(g^{-1}(x))[(g^{-1}(x))'] = \frac{-f_X(g^{-1}(x))}{g'(x)}. \tag{2}$$

В общият случай, когато g е монотонна, и съществува g^{-1} и освен това те са диференцируеми, то уравнения (1) и (2) могат да се обобщят като

$$f_Y(x) = f_X(g^{-1}(x))|(g^{-1}(x))'| = \frac{f_X(g^{-1}(x))}{|g'(x)|}. \tag{3}$$

Нормално Разпределение

Случайната величина X има нормално разпределение с параметри a и σ^2, (означаваме също $X \sim N(a, \sigma^2)$), когато плътността и е

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Функцията на разпределение се записва по следния начин

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(u-a)^2}{2\sigma^2}} du, \quad -\infty < x < \infty.$$

Математическото очакване и дисперсията на сл. в. X са съответно: $E[X] = a$, $D[X] = \sigma^2$. Да пресметнем математическото очакване. Имаме

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx + a\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} d(x-a) + a$$

$$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{(x-a)}{\sqrt{2}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} d\frac{(x-a)}{\sqrt{2}\sigma} + a$$

[Полагаме $y = \frac{(x-a)}{\sqrt{2}\sigma}$ и получаваме]

$$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} y e^{-y^2} dy + a$$

$$= -\frac{\sqrt{2}\sigma}{2\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} d(-y^2) + a$$

$$= -\frac{\sqrt{2}\sigma}{\sqrt{2\pi}} \left(e^{-y^2} \Big|_{-\infty}^{\infty} \right) + a$$

$$= 0 + a = a.$$

Нормално разпределена сл. величина Z със средно $a = 0$ и дисперсия $\sigma^2 = 1$ се отбелязва $Z \sim N(0, 1)$ и се нарича стандартна нормална сл. величина. Нейната плътност се означава с

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$

и съответно функцията на разпределение

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du, \quad -\infty < x < \infty.$$

Тази функция на разпределение е табулирана и таблици има в почти всяка книга по вероятности и статистика, поради честото използуване на нормално разпределени сл. величини в различни приложения и поради следният факт:

Ако $X \sim N(a, \sigma^2)$, *то* $Z = \frac{X-a}{\sigma} \sim N(0,1)$. *С други думи ако центрираме с* a *и нормираме със* σ *новополучената сл. величина е стандартна нормална сл. величина.*
Нека да се убедим, че е така. Нека $X \sim N(a, \sigma^2)$. *Преобразуванието е* $g(x) = (x-a)/\sigma$ *и следователно* $g^{-1}(x) = \sigma x + a$ *и* $|g'(x)| = 1/\sigma$. *Тогава*

$$f_Z(x) = f_X(\sigma x + a)/(1/\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(\sigma x + a - a)^2}{2\sigma^2}}\right)/(1/\sigma)$$

$$= \frac{\sigma}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}} = \varphi(x).$$

Как да използуваме това преобразувание и таблиците на стандартната нормална функция на разпределение.

Нека да видим най-напред, че плътността $\varphi(x)$ *е четна функция, следователно графиката и е симетрична спрямо оста* Oy. *За дадено* $x \in (-\infty, \infty)$ *стойността на функцията на разпределение* $\Phi(x)$ *е равна на лицето под графиката на* $\varphi(x)$ *над интервала* $(-\infty, x)$. *Разбира се цялото лице е равно на единица.*

Като разгледаме и картинката можем да установим лесно следните връзки, които са полезни в случай, че таблицата, с която разполагаме съдържа само стойностите на $\Phi(x)$ *за* $x \in [0, \infty)$. *(както е например в учебника на Б.Димитров и Н.Янев).*

За $x < 0$ *е в сила* $P(Z < x) = P(Z > -x) = 1 - \Phi(-x)$.

За всяко x *в в сила* $P(Z \geq x) = 1 - P(Z < x) = 1 - \Phi(x)$.

Ако $X \sim N(a, \sigma^2)$ *то вероятността* $P(x_1 \leq X < x_2)$ *ще пресмятаме по следния начин:*

$$P(x_1 \leq X < x_2) = P(\frac{x_1 - a}{\sigma} \leq \frac{X - a}{\sigma} < \frac{x_2 - a}{\sigma})$$

$$= P(\frac{x_1 - a}{\sigma} \leq Z < \frac{x_2 - a}{\sigma}) = \Phi(\frac{x_2 - a}{\sigma}) - \Phi(\frac{x_1 - a}{\sigma}).$$

Задачи.

Зад. 1'. *Предполагаме, че височината на двадесет годишните младежи е нормално разпределена сл.в. $a = 170, = 5$. Да се определи вероятността от пет случайно избрани младежи поне един (точно двама) да има ръст от 165 до 175 см. Каква е вероятността младеж да бъде по-висок от 170 ако се знае, че той е по-висок от 160.*

Решение:
Нека да означим с X сл. в. равна на височината на младеж на 20 години. Според условието $X \sim N(170, 5^2)$. Нека да пресметнем вероятността

$$p = P(165 \leq X < 175) = P(165 - 170 \leq X - 170 < 175 - 170)$$

$$= P(\frac{165 - 170}{5} \leq \frac{X - 170}{5} < \frac{175 - 170}{5})$$

$$= P(-1 \leq \frac{X - 170}{5} < 1) = P(-1 \leq Z < 1) = P(Z < 1) - P(Z < -1)$$

$$= \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 =$$

$$2 \times 0.8413 - 1 = 1.6826 - 1 = 0.6826.$$

Сега опитът "Избрани са независимо един от друг 5 младежи на 20 години" е Бренулиева серия от независими опити и ако считаме за успех младежът да има ръст между 165 и 175 см, то вероятността успех е $p = 0.6826$.

Сега вероятността за поне един успех е

$$P_{\geq 1} = 1 - b(5, 0, 0.6826) = 1 - (0.6826)^5 = 1 - 0.148 = 0.852$$

Вероятността за точно два успеха е

$$b(5, 2, 0.6826) = \binom{5}{2} 0.6826^2 (1 - 0.6826)^3 = 10 \times (0.6826)^2 \times (0.3174)^3 = 0.469$$

Да намерим вероятността

$$P(X \geq 170 | X \geq 160) = P(X \geq 170, X \geq 160)/P(X \geq 160) = \frac{P(X \geq 170)}{P(X \geq 160)}.$$

Намираме

$$P(X \geq 170) = P((X - 170)/5 \geq (170 - 170)/5) = P(Z \geq 0) = 1 - \Phi(0) = 0.5$$

$$P(X \geq 160) = P((X - 170)/5 \geq (160 - 170)/5) = P(Z \geq -2)$$

$$= 1 - \Phi(-2) = 1 - (1 - \Phi(2)) = \Phi(2) = 0.9772$$

Така

$$P(X \geq 170 | X \geq 160) = P(X \geq 170, X \geq 160)/P(X \geq 160)$$

$$= P(X \geq 170)/P(X \geq 160) = 0.5/0.9772 = 0.512.$$

Зад. 2'. *Каква е вероятността нормално разпределена сл.в.* $X \sim N(a, \sigma^2)$ *да е в интервал от едно стандартно отклонение около средното си.*

Решение:

$$P(a- \leq X < a+\sigma) = P(\frac{(a-\sigma)-a}{\sigma} \leq \frac{X-a}{\sigma} < \frac{(a+\sigma)-a}{\sigma}) = P(-1 \leq Z < 1)$$

Тук $Z \sim N(0,1)$. Следователно

$$P(a- \leq X < a+\sigma) = P(-1 \leq Z < 1) = \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 2 \times 0.8413 - 1 = 0.6826$$

Следователно вероятността нормално разпределена сл.в. да попадне в интервал от едно стандартно отклонение около средното си е приблизително 68%.

Зад. 3'. *Каква е вероятността нормално разпределена сл.в.* $X \sim N(a, \sigma^2)$ *да е в интервал от* k *стандартни отклонения около средното си, където:*
а) $k = 2$,
б) $k = 3$.

Решение:
а) $P(a - 2 \leq X < a + 2\sigma) = 2\Phi(2) - 1 = 2 \times 0.9772 - 1 = 0.9544 \sim 95\%$
б) $P(a - 3 \leq X < a + 3\sigma) = 2\Phi(3) - 1 = 2 \times 0.9987 - 1 = 0.9974 \sim 100\%$

Зад. 4'. *Нека е дадена нормално разпределена сл.в.* $X \sim N(70, 12^2)$. *Какъв е интервалът около средото, който в който случайната величина попада поне с* 50% *вероятност. Или с други думи поне* 50% *от реализациите и са в този интервал.*

Решение:
Търсим такова $x > 0$, че да е вярно $P(70 - x \leq X < 70 + x) \geq 0.5 \Rightarrow$

$$P(70-x \leq X < 70+x) \geq 0.5 \Leftrightarrow P(\frac{(70-x)-70}{12} \leq (X-70)/12 < \frac{(70+x)-70}{12}) \geq 0.5 \Leftrightarrow$$

$$P(-x/12 \leq Z < x/12) \geq 0.5 \Leftrightarrow 2\Phi(x/12) - 1 \geq 0.5 \Leftrightarrow \Phi(x/12) \geq 0.75$$

Сега в таблицата намираме при кой квантил на станрартното номрално разпределение вероятността е поне 0.5. Това се оказва че е $q_0.5 = 0.675$. Тогава получаваме $x = 0.675 \times 12 = 8.1$ и от тук $P(X \in [61.9; 78.1)) \geq 0.5$.

Зад. 5'. *Резултатът от един тест е нормално разределен със средна 80т. и стандартно отклонение от 13т.*
a) Какъв резултат трябва да е изкарал студент за да е сред най-добрите 15% *на този тест.*
b) Какъв резултат трябва да е изкарал студент за да е сред най-слабите 22% на този тест.
c) Каква е вероятността измежду случайно изтеглени 5 теста да има поне един който има повече от 90т.

Решение:
a) Търсим такова $x > 0$, което да удовлетворява $P(X \geq x) \leq 0.15$. Следователно имаме

$$P((X-80)/13 \geq (x-80)/13) \leq 0.15 \Leftrightarrow P(X < (x-80)/13) \geq 0.85.$$

Сега от таблицата за стандартно нормално разпределение получаваме:

$$\frac{x-80}{13} \geq 1.035 \Leftrightarrow x \geq 93.455.$$

Следователно трябва да изкара поне 94 точки за да е в първите 15%.

b) Търсим такова $x > 0$, което да удовлетворява $P(X < x) \leq 0.22$. Следователно имаме

$$P((X-80)/13 < (x-80)/13) < 0.22 \Leftrightarrow P(X < (x-80)/13) < 0.22. \Leftrightarrow \Phi((x-80)/13) < 0.22$$

$$1 - \Phi((x-80)/13) = \Phi((80-x)/13) \geq 0.78$$

Тогава имаме $(80-x)/13 \geq 0.775$ и следователно $x \leq 69.925$. Следователно трябва да има по-малко от 69 точки за да е измежду най-слабите 22%.

c) Първо да сметнем каква е вероятносста един тест да има поне 90т. За целта трябва да намерим вероятността $P(X \geq 90)$. Последователно намираме

$$P(X \geq 90) = P(Z \geq (90-80)/13) = P(Z \geq 0.769) = 1-P(Z < 0.769) = 1-0.7794 = 0.23.$$

Сега вероятността измежду случайно изтеглени 5 теста да има поне един който има повече от 90т. се дава чрез $1 - b(5, 0, 0.23) = 1 - 0.2707 = 0.7293$.