Последователни експерименти

Както вече коментирахме в урока за събития и вероятности, един експеримент може да бъде провеждан многократно и честотата на появяване на определен резултат да се тълкува като събитие. Например, ако при хвърляне на монета всеки път вероятността да се падне ”ези“ е 1/2, не е трудно да съобразим, че теоретичната вероятност след n последователни хвърляния да имаме k пъти ”ези“ е пропорционална на броя възможни комбинации C_n^k и се дава с формулата

$$p\,(n,k) = \frac{1}{2^n}\binom{n}{k}.$$

Така например, за вероятността да видим ”ези“ в 3 от 5 поредни хвъляния имаме

$$p\,(5,3) = \binom{5}{3}\left(\frac{1}{2}\right)^5 = \frac{10}{32} \approx 31\%.$$

По-общо, ако имаме събитие с два възможни изхода X и $\bar{X}$, с индивидуални вероятности при всеки опит съответно p и $\bar{p} = 1 - p$, биномната формула на *Newton* помага да определим вероятността X да се наблюдава точно в k от общо n последователни експеримента, която се дава от (биномния) закон на *Bernoulli*

$$p\,(n,k) = \binom{n}{k} p^k (1-p)^{n-k}. \qquad (0.1)$$

Нека илюстрираме това с един малко по-нестандартен вариант на играта ”ези-тура“, при който монетата има форма на полукълбо и ако се пренебрегнат силите на триене, можем да приемем, че вероятността p да падне на плоската си част е на половина по-малка от тази да падне на облата $1 - p$, тъй като такова е съотношението на двете площи, следователно имаме съответно $p = \frac{1}{3}$ и $\bar{p} = \frac{2}{3}$. Така например, за да определим вероятността в 3 от 5 последователни хвърляния монетата да падне върху плоската си част, ползвайки формула (0.1) получаваме

$$p\,(5,3) = \binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2 = \frac{40}{243} \approx 16.5\%.$$

Тази идея се обобщава с така нареченото *полиномно разпределение*, описващо експеримент с общо k на брой различни и независими помежду си изхода X_i. Ако означим съответните индивидуални (еднократни) вероятности за случване с p_i, очевидно имаме $\sum\limits_{i=1}^{k} p_i = 1$. Тогава вероятността след n проведени опита да отчетем всяко събитие X_i точно m_i пъти, където $\sum\limits_{i=1}^{k} m_i = n$, се дава с формулата

$$p\,(m_1, m_2 \ldots m_k) = \frac{n!}{m_1!\,m_2! \cdots m_k!}\, p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \qquad (0.2)$$

като при $k = 2$ съвпада с (0.1). Тук един класически и много опростен пример е зарче за игра, в което вероятността да се падне всяко едно от числата, обозначени на страните му, е $p_i = \frac{1}{6}$, което дава за вероятността при 6 последователни хвърляния всяко от тези числа да се падне съответно $0, 1, 2, 2, 1, 0$ пъти се дава с

$$p\,(0, 1, 2, 2, 1, 0) = \frac{6!}{0!\,1!\,2!\,2!\,1!\,0!}\left(\frac{1}{6}\right)^6 = \frac{5}{1296} \approx 0.4\%.$$

Нека обърнем внимание, че както биномният, така и полиномният закон предполагат една и съща индивидуална вероятност p (съответно p_i) при провеждането на всеки експеримент. В практиката обаче често се случва това да не е изпълнено. Ще разгледаме един актуален пример: нека 5 от всеки 1000 души в популацията на някакъв град са носители на вирус. Ако знаем, че средното време за удвояване броя на заразените е два дни, а разпределението им смятаме за равномерно, каква е вероятността жител на този град да се сблъска с вируса ако контактува с по един случайно избран свой съгражданин дневсно в продължение на седмица? Най-лесно ще отговорим на въпроса каква е вероятността $\bar{p}_7$ човекът да не влезе в контакт с вируса в продължение на седмица, след което ще изразим $p = 1 - \bar{p}_7$. Тъй като всеки един е заразен с вероятност $p_0 = 0.5\%$, още първият ден имаме $\bar{p}_1 = \bar{p}_0^2 = (1 - p_0)^2$, и поради средното време за удвояване, всеки следващ ден вероятността за единично заразяване ще нараства с коефициент $\sqrt{2}$, т.е. имаме

$$\bar{p}_{k+1} = \bar{p}_k(1 - 2^{\frac{k}{2}}p_0), \qquad \bar{p}_0 = 1 - p_0$$

което дава вероятността за контакт с вируса до седмия ден включително като

$$p = 1 - \bar{p}_7 \approx 12.3\%$$

или приблизително колкото в игра на руска рулетка с барабан за осем патрона.

Многократните експерименти дават връзката между вероятности и статистика при съпоставянето на теоретичната вероятност към емпирична, дефинирана като честота на случване. Формула (0.1) и (0.2) дават и първи примери за (дискретни) статистически разпределения, които ще бъдат във фокуса на следващия ни урок.

Задачи

1. Решете горната задача без отчитане на времето за удвояване и 1% заразени.

2. Пресметнете примера по-горе за кух зар, в който броят точки на всяка стена показва масата ѝ. Намерете и вероятността ако зарът е с форма на тетраедър, да се отчат за m_i съответно $0, 1, 2$ и 3 случвания от общо шест.

3. Ако в кутия има седем бели и шест черни топки, каква е вероятността след пет случайни тегления без връщане да са останали по равно от двата вида?