Binomial and Geometric distributions

Anastasiia Kim

February 12, 2020

Example

There are n eggs, each of which hatches a chick with probability p (independently). Each of these chicks survives with probability r, independently.

- What is the distribution of the number (H) of chicks that hatch?
- What is the expected number of chicks that hatch?
- What is the distribution of the number (S) of chicks that survive?

R: Binomial distribuion

Useful functions in R: *dbinom*, *pbinom*, *rbinom*:

- *dbinom* is the Binomial pmf p(x). It takes three inputs: the first is the value of x at which to evaluate the pmf, and the second and third are the parameters n and p. For example, $dbinom(3, 5, 0.2)$ returns the probability $P(X = 3)$ where $X \sim Bin(5, 0.2)$

$$dbinom(3, 5, 0.2) = \binom{5}{3}(0.2)^3(0.8)^2 = 0.0512$$

R: Binomial distribuion

- *pbinom* is the Binomial cdf F(x). It takes three inputs: the first is the value of x at which to evaluate the CDF, and the second and third are the parameters. For example, $pbinom(3, 5, 0.2)$ is the probability $P(X \leq 3)$ where $X \sim Bin(5, 0.2)$

$$pbinom(3, 5, 0.2) = \sum_{k=0}^{3} \binom{5}{k} (0.2)^k (0.8)^{5-k} = 0.9933$$

R: Binomial distribuion

- *rbinom* is a function for generating Binomial random variables. For *rbinom*, the first input is how many r.v.s we want to generate, and the second and third inputs are still the parameters. For example, *rbinom*$(3, 5, 0.2)$ produces realizations of three i.i.d. Bin(5, 0.2) r.v.s

$$2 \quad 1 \quad 1$$

Example

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Let X = the number of samples that contain the pollutant in the next 18 samples analyzed.

- Find the probability that in the next 18 samples, exactly 2 contain the pollutant
- Determine the probability that at least four samples contain the pollutant
- Find $P(3 \leq X < 7)$

Example

A new treatment for a disease is being tested, to see whether it is better than the standard treatment. The existing treatment is effective on 50% of patients. It is believed initially that there is a $2/3$ chance that the new treatment is effective on 60% of patients, and a $1/3$ chance that the new treatment is effective on 50% of patients. In a pilot study, the new treatment is given to 20 random patients, and is effective for 15 of them.

Geometric distribution

Suppose that independent trials, each having a probability p of being a success, are performed until a success occurs. If we let X equal the number of trials required, then

$$P(X = x) = (1 - p)^{x-1}p \qquad x = 1, 2, 3, ...$$

with $E(X) = 1/p$ and $Var(X) = (1 - p)/p^2$

- pmf follows because, in order for X to equal x, it is necessary and sufficient that the first x - 1 trials are failures and the xth trial is a success. The outcomes of the successive trials are assumed to be independent.

Prove that $\sum_{n=1}^{\infty} P(X = x) = 1$ meaning that with probability 1, a success will eventually occur. Derive cdf.

Example

The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the first error. Then P (X = 5) is the probability that the first four bits are transmitted correctly and the fifth bit is in error.

$$P(X = 5) = 0.9^4 0.1 = 0.066$$

The mean number of transmissions until the first error is $E(X) = 1/p = 1/0.1 = 10$. The standard deviation of the number of transmissions before the first error is $SD(X) = \sqrt{Var(X)} = \sqrt{(1-p)/p^2} = \sqrt{(1-0.1)/0.01} = \sqrt{90} = 9.49$.

Lack of Memory Property

A geometric random variable has been defined as the number of trials until the first success.

- Because the trials are independent, the count of the number of trials until the next success can be started at any trial without changing the probability distribution of the random variable.

For example, if 100 bits are transmitted, the probability that the first error, after bit 100, occurs on bit 105 is the probability that the next six outcomes are OOOOE. This probability is $(0.9)^4(0.1) = 0.066$, which is identical to the probability that the initial error occurs on bit 5.