Geometric, Negative Binomial, and HyperGeometric Distributions

Anastasiia Kim

February 17, 2020

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.

Negative Binomial Distribution

In a series of Bernoulli trials, the random variable X that equals the number of trials until r successes occur is a negative binomial random variable with parameters p and $r = 1, 2, 3, \ldots$ and

$$P(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \qquad x = r, r+1, r+2, \ldots$$

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

- a negative binomial (NB) random variable is a count of the number of trials required to obtain r successes
- the number of successes r is predetermined and the number of trials is random

Example

The probability that a camera passes the test is 0.8, and the cameras perform independently. What is the probability that the third failure is obtained in five or fewer tests?

Let X denote the number of cameras tested until three failures have been obtained.

$X \sim NB(0.2, 3)$

$$P(X \leq 5) = \sum_{x=3}^{5} = \binom{x-1}{3-1} 0.2^3 (0.8)^{x-3} = 0.056$$

HyperGeometric Distribution

Consider an urn with w white balls and b black balls. We draw n balls out of the urn at random without replacement. Let X be the number of white balls in the sample. Then X is said to have the Hypergeometric distribution with parameters w, b, and n
$X \sim HyperGeometric(w, b, n)$

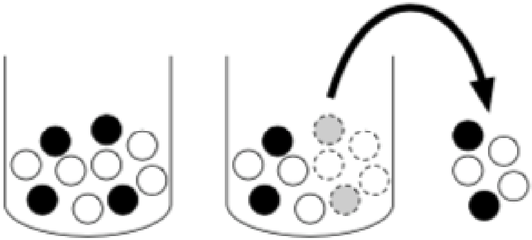


Figure 1: Hypergeometric story. An urn contains w = 6 white balls and b = 4 black balls. We sample n = 5 without replacement. The number X of white balls in the sample is Hypergeometric; here we observe X = 3.

HyperGeometric Distribution

If $X \sim HyperGeometric(w, b, n)$, then pmf of X is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}, \qquad 0 \leq k \leq w, 0 \leq n-k \leq b...$$

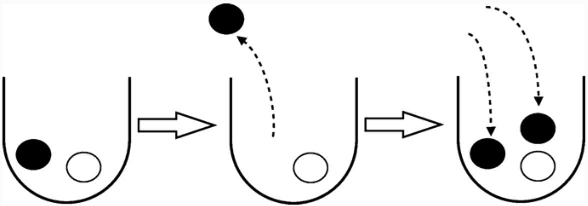
Example: in a five-card hand drawn at random from a well-shuffled standard deck, the number of aces in the hand X has the HyperGeometric(4, 48, 5) distribution, which can be seen by thinking of the aces as white balls and the non-aces as black balls. Using the Hypergeometric pmf, the probability that the hand has exactly three aces (X=3) is

$$P(X = 3) = \frac{\binom{4}{3}\binom{48}{5-3}}{\binom{4+48}{5}} = 0.0017$$

Important: the sampling should be done without replacement! Here, we are drawing cards from a deck of well-shulffed cards WITHOUT replacement.

Beta-binomial Distribution

In an urn containing w white balls and b black balls, random draws are made. If a white ball is observed, then two white balls are returned to the urn. Likewise, if a black ball is drawn, then two black balls are returned to the urn. If this is repeated n times, then the probability of observing k white balls follows a beta-binomial distribution with parameters n, w and b.



In Pólya's urn model, a ball is drawn at random from an urn and replaced and an additional ball of the same colour is added

Figure 2: The image is from https://link.springer.com/article/10.1007/s12064-019-00290-6