Stats 130
Day 7 Notes

Elijah Hantman

Binomial

Suppose we repeat n Bernoulli experiments, indpendently and with constant probability of success p. Define X as the random variable that counts the total number of successes. Then X takes possible values in the set $D = \{0, 1, ...n\}$.

To Calculate the probability function we observe that, in order to obtain exactly x successes, we need exactly $n - x$ failures. There are $\binom{n}{x}$ equally likely ways for this to happen. Thus,

$$f_x(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} | x > 0 \\ 0 \quad | \quad else \end{cases}$$

Notice that

$$\sum_{x=0}^{n} f_x(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1$$

Notated: $X \sim Bin(n, p)$

Example

The Giants need to win at least four of the remaining six games to make the playoffs. Given a record of 27-27 what is the probability they succeed?

We assume $p = 0.5$ and that X is the number of wins in the next six games. Then X $Bin(6, 0.5)$. We want

$$Pr(X \geq 4) = Pr(X = 4) + Pr(X = 5) + Pr(X = 6) \tag{1}$$

$$= (\frac{1}{2})^6 (\binom{6}{4} + \binom{6}{5} + \binom{6}{6}) \tag{2}$$

$$= \frac{11}{32} \tag{3}$$

Hypergeometric

A hypergeometric experiment is like multiple Bernoulli trials except there is dependence between the trials.

Suppose a box contains A red balls and B blue balls. n balls are drawn without replacement from the box. A random variable X is equal to the number of red balls drawn. $X \leq min\{n, A\}$. Also the sample has $n - X$ blue balls, thus, $n - X \leq B$, then

$$max\{0, n - B\} \leq X \leq min\{n, A\}$$

Then

$$Pr(X = x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$$

Notated: $X \sim Hyper(n, A, B)$

Example

The probability function is given as

$$f_x(x) = Pr(X = x), \quad max\{0, n - B\} \leq x \leq min\{n, A\}$$

We have patients suffering from depression. 34 recieved a placebo. Of those, 10 had no relapse, and 24 had a relapse.

Suppose we take a sample of patients with $n = 11$. Let $X = \#$ of successes. Then $A = 10$ and $B = 24$.

$$0 = max\{0, 11 - 24\} \leq X \leq min\{11, 10\} = 10$$

and

$$f_x(x) = \frac{\binom{10}{x}\binom{24}{11-x}}{\binom{34}{11}}, x = 0, ..., 10$$

Negative Binomial

A sequence of Bernoulli trials with probability of success p, that stops after a given number of successes is reached. A normal Binomial stops after n trials, a negative Binomial has n as a random variable.

$$\begin{aligned}
Pr(X = x) &= Pr(\text{x failures before r successes occur}) && (4)\\
&= Pr(\text{x failures and r-1 successes in x+r-1 trials})\times && (5)\\
Pr(\text{one success in the (x+r)-1 th trial}) &= \binom{x+r-1}{x} q^x p^{x+r-1-x} \times p && (6)\\
&= \binom{x+r-1}{x} q^x p^r, x = 0, 1, ... && (7)
\end{aligned}$$

Notated: $X \sim NegBin(p, r)$

Example

To pass a quiz you need five correct answers. The system provides a new question, until you reach the fifth correct answer. Your grade is (5/attempts)*100. What is the probabilty that you will score 50% assuming you are guessing the correct answers?

To score 50% you need 5 correct answers in 10 attempts. Assume you have a 50% chance of guessing the correct answer.

$$x = 5 \tag{8}$$
$$r = 5 \tag{9}$$
$$\binom{9}{5} \frac{1}{2}^5 \frac{1}{2}^5 \tag{10}$$
$$= 0.1230469 \tag{11}$$

Geometric

A Geometric experiment is a negative binomial with $r = 1$.

Therefore, the formula is

$$f_x(x) = pq^x, \quad x = 0, 1, ...$$

This works because

$$\sum_{x=0}^{\infty} f_x(x) = \sum_{x=0}^{\infty} pq^x = p \sum_{x=0}^{\infty} q^x = p(\frac{1}{1-q}) = \frac{p}{p} = 1$$

It is notated $X \sim Gep(p)$

Memory Loss Property

Consider $X \sim Geo(p)$

$$Pr(X \geq x) = 1 - Pr(X < x) = 1 - Pr(X \leq x-1) \tag{12}$$

$$= 1 - \sum_{i=0}^{x-1} q^i p = 1 - p\frac{1-q^x}{1-q} = q^x \tag{13}$$

Suppose $X \geq t$, what is the probabiilty that $X = t+h, h > 0$?

$$Pr(X = t+h | X \geq t) = \frac{Pr(X = t+h)}{Pr(X \geq t)} \tag{14}$$

$$= \frac{pq^{t+h}}{q^t} = pq^h = Pr(X = h) \tag{15}$$

The geometric "forgets" that there have been t attempts or more. Past attempts do not influence the probability of future attempts.

Another example is radioactive decay, each individual particle has a geometric probability function.

Poisson

A Poisson random variable counts the number of occurances of an event. It can therefore take any possible positive integer.

Some examples: # of new Covid positives, # of tweets, # of requests, etc.

The probability function of a Poisson can be obtained as the limit of the probability function of a binomial where the number of trials go to infinity. This results in

$$f_x(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, ...$$

Here λ is a parameter that determines the intensity of the process of occurrences per unit time.

Is $f_x(x)$ a probability function? $f_x(x) > 0, \forall x$. Then

$$\sum_{x=0}^{\infty} f_x(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$