SOA and CAS: Exam P, Probability[1]
Chapter 18 and 19: Binomial and Negative Binominal Distributions

Yi Li [2]

January 13, 2024

Chapter 18: Binomial Distribution: Single Variable
Chapter 19: Negative Binominal Distribution: Single Variable

(1) *Chapter 18: Binomial Distribution: Discrete Single Variable*

(1.1) Binomial Distribution

Definition: $P(N = k)$ stands for the probability of "number of successful trails = k"

Example (1.1.a): An urn contains red and blue balls. $P(red) = p$ and $P(blue) = 1 - p$. Now, draw a ball from this urn, observe the color, put it back (*draw with repalcement*). Then, draw another ball, observe the color, put it back. Repeat this process n times. Then, $P(N = k)$ stands for P(out of n draws, the "number of red" = k). Note that we actually have n Bernoulli trails. Each trail has two outcomes: red ball or blue ball with $P(red) = p$ and $P(blue) = 1 - p$.

Formula:

$$Pr(N = k) = C_n^k * p^k * (1 - p)^{n-k} \tag{1}$$

where n: number of Bernoulli trails. Each Bernoulli trail has two outcomes, such as: success or fail, red or blue. $P(success/red) = p$ and $P(fail/blue) = 1 - p$

Example (1.1.b): 1000 machines, 15 are defective. Draw 20 times repeatedly.
Question: What is P(k ≤ 2)?

Solve: $\text{P}(k \le 2) = (\underbrace{k=0}_{\#defective\ is\ 0}) + (\underbrace{k=1}_{\#defective\ is\ 1}) + (\underbrace{k=2}_{\#defective\ is\ 2})$

$$= C_{20}^0 * p^0 * (1-p)^{20-0} + C_{20}^1 * p^1 * (1-p)^{20-1} + C_{20}^2 * p^2 * (1-p)^{20-2}$$

$where\ p = \frac{15}{1000}$

(1.2) Hypergeometric Distribution

Definition: $P(N = k)$ stands for the probability of "number of successful trails = k". Recall that the process in the binomial distribution is "draw with replacement". Here, the process is "draw without replacement".

Example (1.2.a): A set contains N samples, D defective. Now, draw one, draw another one, draw the third one (draw without replacement)..., draw n times in total.
Question: What is P(d defective in n samples)?

Solve: $P(d\ \ defective) = \frac{A}{B} = \frac{C_D^d C_{N-D}^{n-d}}{C_N^n} = \frac{\text{“}draw\ d\ from\ D\ defective\text{”} * \text{“}rest\ (n-d)\ draw\ from\ the\ rest\ (N-D)\text{”}}{\text{“}draw\ n\ from\ N\text{”}}$

Example (1.2.b): 1000 machines, 15 are defective. Draw 20 times without repalcement.
Question: What is P(k ≤ 2)?

By definition: we know the process is $Hypergeometric(n, p)$.

---

[1]The purpose of the use is non-commercial research and/or private study. Please do not copy or distribute without permission of the author.

[2]Email: liyifinhub@outlook.com. This note was drafted when I was preparing for the exam. Please email me if you find any errors. My personal website http://www.yilifinhub.com

Solve: $P(k \leq 2) = (\underbrace{k=0}_{\#defective\ is\ 0}) + (\underbrace{k=1}_{\#defective\ is\ 1}) + (\underbrace{k=2}_{\#defective\ is\ 2})$

$= \frac{C_{15}^{0} * C_{1000-15}^{20-0}}{C_{1000}^{20}} + \frac{C_{15}^{1} * C_{1000-15}^{20-1}}{C_{1000}^{20}} + \frac{C_{15}^{2} * C_{1000-15}^{20-2}}{C_{1000}^{20}}$

(1.3) Trinominal Distribution

For example: one urn contains three different classes of balls. First class has a same balls. Second class has b same balls. Third class has c same balls. $P(draw\ one\ a)$=0.6, $P(draw\ one\ b)$=0.3, $P(draw\ one\ c)$=0.1. Then, P(draw 3 first class, 2 second class, 1 third class)=$\frac{6!}{3!2!1!}(0.6)^3(0.2)^2(0.1)^1$

(1.4) Summary

	Bernoulli	Binomial	Hypergeometric
Mean	p	np	$\frac{nD}{N}$
Variance	1-p	np * (1-p)	$\frac{nD}{N} * \frac{N-D}{N} * \frac{N-n}{N-1}$

(2) *Chapter 19: Negative Binominal Distribution: Single Variable*

(2.1) Negative Binominal Distribution

Definition: (i) a total of (n+k) trails, the very last trail is "a successful trail". In the previous (n+k-1) trails, there are n unsuccessful trails; (ii) Each trail can only have 2 outcomes: successful trail or unsuccessful trail. P(successful)=p and P(unsuccessful)=$1-p$

$$P(N = k) = \underbrace{C_{n+k-1}^{n}(1-p)^n(p)^{k-1}}_{n+(k-1)\ has\ n\ unsuccessful\ trails,\ k-1\ successful\ trails} * \underbrace{p}_{last\ one\ is\ a\ successful\ trail}$$

$$= C_{n+k-1}^{n}(1-p)^n(p)^k$$

Example (2.1.a): Give $P(win)$=0.6
Question: P(4^{th} game win, occurs before, 3^{rd} lose)?

Solve: we know that the last trail (in the total 7 trails) must be a lose trail. Then, in order to have the 4^{th} win before the 3^{rd} lose, the previous (4+2) trails should have at least 4 wins.

$$Probability = P(N = 4) + P(N = 5) + P(N = 6)$$
$$where\ P(N = 4) = C_6^4(p)^4(1-p)^2$$
$$P(N = 5) = C_6^5(p)^5(1-p)^1$$
$$P(N = 6) = C_6^6(p)^6(1-p)^0$$
$$p = 0.6$$

Example (2.1.b): Give $P(win)$=0.6
Question: P(6^{th} game is win, which is also the 4^{th} win)?

Solve: we know that the last trail (in the total 6 trails) must be a win trail. Then, in order to make the 4^{th} win happens at the 6th trail, the previous 5 trails should have 3 wins and 2 loses.

$$Probability = P(N = 2) = P(2\ loses)$$
$$= \underbrace{C_5^2(1-p)^2(p)^3}_{total\ 5\ trails,\ 2\ unsuccessful\ and\ 3\ successful} * \underbrace{p}_{last\ one\ is\ a\ successful\ trail}$$

$$where\ p = 0.6$$

(2.2) Geometric Distribution

Definition: (i) a total of (n+1) trails, the very last trail is "a successful trail". The previous n trails are all unsuccessful trails; (ii) Each trail can only have 2 outcomes: successful and unsuccessful. P(successful)=p and P(unsuccessful)=$1-p$

$$\mathrm{P(N = k)} = \underbrace{(1-p)^n}_{n\ trails\ are\ all\ unsuccessful\ trails} * \underbrace{p}_{last\ one\ is\ the\ only\ successful\ trail}$$

$$= (1-p)^n p$$

where n=0, 1, 2,...

	Negative Binominal	Geometric Distribution Special case: (k=1: last trail is the only success trail)
Mean	$\frac{k}{p}(1-p)$	$\frac{1}{p}(1-p)$
Variance	$\frac{k}{p^2}(1-p)$	$\frac{1}{p^2}(1-p)$
Formula	$P(X=n) = C_{n+k-1}^{n}(1-p)^n p^k$	$P(X=n) = (1-p)^n p^1$

Example (2.2): With replacement: A shipment contains 20 packages, 7 are damaged. Randomly draw without replacement until the 4^{th} damaged package is discovered.

Question: What is the P(exactly 12 packages are inspected)?

Solve: we know that the 12^{th} package must be a damaged one. Thus, the previous 11 packages must contains 3 damage packages. (i) the previous 11 inspected packages: 3 damaged packages should be drawn from the total 7 damaged packages, while 8 packages should be drawn from the total 13 undamaged packages. That is $\frac{A}{B} = \frac{C_7^3 C_{13}^8}{C_{20}^{11}} = 0.26$; (ii) the last 12^{th} damaged package should be drawn from the total 4 remaining damaged packages. That is $\frac{C}{D} = \frac{C_4^1}{\underbrace{C_9^1}_{9\ left\ after\ 11\ draws}} = \frac{4}{9}$. Overall, P(exactly 12 packages are inspected)=$\frac{A}{B} * \frac{C}{D}$=0.26$*\frac{4}{9}$=0.12