3.4 The Binomial Distribution

In many experiments, we perform independent trials, like flipping a fair coin 5 times.

Say we perform n independent trials. Each trial is either a Success or Failure, and each trial has the same probability of success, p. Let X be the number of successes obtained in the n trials. Then X is a Binomial RV with parameters n and p. $X \sim Bin(n, p)$.

Ex: How many Heads do we get if we flip a fair coin 10 times? $X \sim Bin(10, 0.5)$

How many 6's do we get if we roll a fair die 15 times? $X \sim Bin(15, \frac{1}{6})$

Sometimes the trials are not independent, like if we sample without replacement from a population. If the number of trials is at most 5% of the population, we can still approximate the number of successes with the Binomial distribution.

Ex: Survey students at a university, if we sample (without replacement) less than 5% of the population. Count the number of math majors in our sample.

Let $X \sim Bin(3, \frac{1}{4})$. What is $P(X=2)$?

F	S	S
S	F	S
S	S	F

$$P(X=2) = \boxed{3 \cdot \left(\tfrac{1}{4}\right) \cdot \left(\tfrac{1}{4}\right) \cdot \left(1-\tfrac{1}{4}\right)}$$

$$P(X \geq 1) = 1 - P(X=0)$$

$$= P(X=1) + P(X=2) + P(X=3)$$

Let $X \sim Bin(n,p)$. Then for $k = 0, 1, 2, \dots, n$,

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

combination

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

Example 1: A laptop manufacturer knows from experience that 10% of their laptops will need service during the warranty period. A company buys 5 of their laptops, chosen independently from the manufacturer's supply.

a) What is the probability exactly 3 laptops will need service during the warranty period?

b) What is the probability no more than 2 laptops will need service during the warranty period?

Let X be the number of laptops needing service.

a) $X \sim Bin(5, 0.1)$ $\quad P(X=3) = \binom{5}{3} \cdot 0.1^3 \cdot (1-0.1)^{5-3} = \boxed{0.0081}$

b) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$$= \binom{5}{0} \cdot 0.1^0 \cdot 0.9^5 + \binom{5}{1} 0.1^1 \cdot 0.9^4 + \binom{5}{2} 0.1^2 \cdot 0.9^3 \approx \boxed{0.991}$$

Mean and Variance of the Binomial distribution:

If $X \sim Bin(n,p)$, then

$E[X] = np$

$Var(X) = np(1-p)$

$SD(X) = \sqrt{np(1-p)}$

Example 2: At a certain store, 0.5% of all payments will be charged back by the credit card company. If the store processes 1000 payments each month, what is

a) the probability exactly 3 chargebacks will occur this month?

b) the mean and SD of the number of chargebacks this month?

Let X be the number of chargebacks.

$X \sim Bin(1000, 0.005)$

a) $P(X=3) = \binom{1000}{3} \cdot 0.005^3 \cdot (1-0.005)^{1000-3} \approx \boxed{0.1403}$

b) $E[X] = n \cdot p = 1000 \cdot 0.005 = \boxed{5}$ $\quad SD(X) = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0.005 \cdot 0.995} \approx \boxed{2.23}$

3.5 Hypergeometric and Negative Binomial Distributions :

Hypergeometric Distribution : Say we have a population of size N. In the population, there are M successes and $N-M$ failures. We pick a random sample of size n. Let X be the number of successes in the sample.

$X \sim \text{Hyper Geom}(N, M, n)$

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad \text{for } x \text{ that make sense}$$

x successes

$n-x$ failures

combinations

Example 1: At an animal shelter, there are 10 cats and 8 dogs. If we pick 5 of the animals at random, what is the probability that we pick 3 dogs?

Hypergeometric, $N=18$, $M=8$, $N-M=10$, $n=5$, $x=3$

$$P(X=3) = \frac{\binom{8}{3}\binom{10}{2}}{\binom{18}{5}} = \frac{5}{17} \approx \boxed{0.294}$$

Online/Computer

Wolfram Alpha

Math Way

R

Mean and Variance for the Hypergeometric Distribution

Let X be a Hypergeometric random variable with

- population size N
- M successes in the population
- sample size n
- X is the number of successes in the sample.

$$E[X] = n \cdot \frac{M}{N}, \quad Var(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N}\left(1-\frac{M}{N}\right)$$

Example 2: Find $E[X]$ and $V(X)$ for example 1.

For example 1, $N=18$, $M=8$, $n=5$

$$E[X] = 5 \cdot \frac{8}{18} = \frac{20}{9} \qquad Var(x) = \boxed{\left(\frac{18-5}{18-1}\right) \cdot 5 \cdot \frac{8}{18} \cdot \left(1-\frac{8}{18}\right)}$$

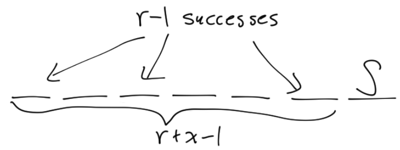
The Negative Binomial Distribution:

Say we conduct independent trials, each of which is a success with probability p and a failure with probability $1-p$. We keep doing trials until we get r successes. Let X be the number of failures before we get r successes. X is a Negative Binomial RV.

$$X \sim NB(r, p)$$

$$P(X=x) = \binom{r+x-1}{r-1} \cdot p^r \cdot (1-p)^x$$

$$x = 0, 1, 2, 3, \ldots$$



Example 3: A charity is calling for donations. Assume each call there is a 0.1 chance to receive a donation, independently of the other calls. The charity will make calls until it receives 5 donations. Let X be the number of calls without donations until 5 donations are reached. What is $P(X=20)$?

$r=5$, $p=0.1$, $x=20$. $X \sim NB(5, 0.1)$

$$P(X=20) = \binom{5+20-1}{5-1}(0.1)^5(1-0.1)^{20} \approx \boxed{0.0129}$$

Mean and Variance of the Negative Binomial Distribution

Let $X \sim NB(r, p)$. Then

$$E[X] = \frac{r(1-p)}{p}, \quad Var(X) = \frac{r(1-p)}{p^2}$$

Geometric Distribution: When $r=1$, the negative binomial distribution is often called Geometric instead.

$X \sim NB(1, p)$ is the same as $X \sim Geom(p)$.

3.6 Poisson Distribution:

A discrete RV X has the Poisson distribution with parameter μ if X has pmf

$$p(x) = P(X=x) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{for } x = 0, 1, 2, 3 \ldots$$

Why do these probabilities add up to 1?

$$\sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^x}{x!} = e^{-\mu} \underbrace{\sum_{x=0}^{\infty} \frac{\mu^x}{x!}}_{\substack{\text{Taylor Series} \\ e^{\mu}}} = e^{-\mu} \cdot e^{\mu} = \frac{e^{\mu}}{e^{\mu}} = 1.$$

Example 1: Let X be a Poisson RV with parameter $\mu = 3$. Find $P(X \leq 2)$.

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) . \qquad \frac{e^{-\mu}\mu^x}{x!}$$

$$= \boxed{\frac{e^{-3} \cdot 3^0}{0!} + \frac{e^{-3} \cdot 3^1}{1!} + \frac{e^{-3} \cdot 3^2}{2!}}$$

Poisson Distribution as a Limit:

Suppose that in the binomial distribution $Bin(n,p)$, we let $n \to \infty$ and $p \to 0$ in such a way that $np \to \mu > 0$. Then these Binomial distributions converge to a Poisson(μ) distribution.

In practice, we can approximate $Bin(n,p)$ with $Pois(\mu)$ where $\mu = np$ if $n > 50$ and $np < 5$.

Example 2: At a certain company, when they publish a book there is a 0.01 probability that there is an error on any page, independently of the other pages. If a published book is 200 pages long, what is the probability there are exactly 3 errors in the book?

Rare events, $200 \cdot 0.01 = 2$

so we can approximate with Poisson(2).

$$P(X=3) = \boxed{\frac{e^{-2} \cdot 2^3}{3!}}$$

Mean and Variance for the Poisson Distribution

If $X \sim \text{Poisson}(\mu)$ then

$E[X] = \mu$ and

$Var(X) = \mu$.

Poisson RV over Time (Poisson Process)

Say we are counting events over time (e.g. meteors, emails received).

Let's assume

1) For a short interval of time, the probability for a single event occurring is approximately proportional to the length of the time interval.

$$P(\text{one event occurs during time } \Delta t) \approx \alpha \cdot \Delta t$$

2) During short time Δt, the probability of 2 or more events occurring is negligible.

3) The number of events occurring during Δt is independent of the past.

Then the number of events occurring in a time interval of length t is $X \sim \text{Poisson}(\alpha t)$.

When $t=1$, $E[X] = \alpha$, so α is the expected number of events in 1 unit of time.

Example 3: For a certain email address, 4 emails on average are received per hour. What is the probability no emails are received over a 30 minute period?

$$\alpha t = 4 \cdot 1 \qquad \alpha = 4$$

$$\mu = \alpha \cdot 0.5 = 4 \cdot 0.5 = 2$$

Let X be the number of emails in the 30 minute window

$$X \sim \text{Poisson}(2)$$

$$P(X=0) = \frac{e^{-2}\, 2^0}{0!} = \frac{e^{-2} \cdot 1}{1} = \boxed{e^{-2}}$$