STAT 230 - Probability

Cameron Roopnarine

Last updated: October 28, 2019

Contents

1 Lecture 13

Recall

Binomial approximation to Hypergeometric distribution: If $X \sim \text{Hyp}(N, r, n)$, we can approximate it with $\text{Bin}(n, \frac{r}{N})$ if n is a small proportion of N.

1.1 Negative Binomial Distribution

Setup: Bernoulli Trials

- independent
- each trial is a success or fail (S or F)
- $P(\text{success}) = p = \text{constant}$

Suppose we want to get k S's. We do trials until we get k S's and let $X = \#$ of F's. We get

$$X \sim \text{NB}(k, p)$$

in a total of $k + X$ trials.

Binomial	Negative Binomial
know # of trials	unknown # of trials
unknown # of S's	known # of S's
$\binom{n}{x} p^x (1-p)^{n-x}$	$\binom{x+k-1}{k-1} p^k (1-p)^x$

1.2 Example

How many tails until we get the 10th head on a fair coin. $X \sim \text{NB}(10, \frac{1}{2})$

1.3 Example

If courses were independent with probability p of passing and you need 40 courses, then the number of failed courses would be $\text{NB}(40, p)$.

1.4 Range and Probability Function of the Negative Binomial Distribution

range $x \in \{0, 1, ...\}$ (countably infinite)

$$\begin{aligned} f(x) = P(X = x) &= p(x \text{ F's before } k\text{th S}) \\ &= \binom{x+k-1}{x} p^k (1-p)^x \\ &= \binom{x+k-1}{k-1} p^k (1-p)^x \end{aligned}$$

In a picture:

$$\underbrace{\_\ \_\ \_\ \cdots\ \_\,|}_{k-1 \text{ S's, } x \text{ F's}}\ \underbrace{\text{S}}_{k\text{th S}}$$

with $x + k - 1$ trials.

1.5 Example

Suppose a startup is looking for 5 investors. They ask investors repeatedly where each independently has a 20% chance of saying yes. Let X = total # of investors that they ask and note that X does not follow a negative binomial distribution. Find $f(x)$ and $f(10)$.

Let Y = # who say no before 5 say yes. $Y \sim \text{NB}(5, 0.2)$, and $X = Y + 5$. So,

$$\begin{aligned} f(x) &= P(X = x) \\ &= P(Y + 5 = x) \\ &= P(Y = x - 5) \\ &= \binom{(x-5)+5-1}{5-1}(0.2)^5(0.8)^{x-5} \\ &= \binom{x-1}{4}(0.2)^5(0.8)^{x-5} \qquad \text{for } x = 5, \ldots \end{aligned}$$

$$f(10) = \binom{9}{4}(0.2)^5(0.8)^5$$

note that it's $\binom{9}{4}$ and not $\binom{10}{5}$ because the 10th investor must have said yes.

2 Lecture 14

Recall

$$X \sim \text{NB}(k, p)$$

$$f(x) = \binom{x+k-1}{k-1} p^k (1-p)^x$$

with $k - 1$ S's, x F's and the last one is a S.

2.1 Example

Suppose you send a bit string over a noisy connection with each bit independently having a probability 0.01 of being flipped. What is the probability that

(a) it takes 50 bits to get 5 errors?

(b) a 50 bit message has 5 errors?

(b) Let Y = # of errors in 50 bits. $Y \sim$ Bin$(50, 0.01)$.

Then, $P(Y = 5) = \binom{50}{5}(0.01)^5(0.99)^{45}$

(a) Let X = # of correct bits until 5 errors. $X \sim$ NB$(5, 0.01)$.

Then, $P(X = 45) = \binom{49}{4}(0.01)^5(0.99)^{45}$

2.2 Geometric Distribution

The Geometric Distribution is just a special case of the Negative Binomial Distribution with $k = 1$. Let X = # of F's in Bernoulli trials before the first S. $X \sim$ Geo(p)

2.3 Range and Probability Function of the Geometric Distribution

range: $x \in \{0, 1, ...\}$

$$\begin{aligned} f(x) &= P(X = x) \\ &= P(\underbrace{\text{F, F, ...}}_{\text{all F's}}, \text{S}) \\ &= (1-p)^x p \end{aligned}$$

or sub $k = 1$ into the NB probability function.

Prove $\sum_{\text{all } x} f(x) = 1$

Proof.

$$\begin{aligned} \sum_{x=0}^{\infty}(1-p)^x p &= \underbrace{p + p(1-p) + ...}_{\text{(a geometric series with } a = p,\ r = 1-p)} \\ &= \frac{p}{1-(1-p)} \\ &= 1 \end{aligned}$$

□

Find the cumulative distribution function.

$$\begin{aligned} F(x) &= P(X \le x) \\ &= 1 - P(X > x) \\ &= 1 - [f(x+1) + ...] \\ &= \underbrace{1 - [p(1-p)^{x+1} + p(1-p)^{x+2} + ...]}_{\text{(a geometric series with } a = p(1-p)^{x-1},\ r = 1-p)} \\ &= 1 - \frac{p(1-p)^{x+1}}{1-(1-p)} \\ &= 1 - (1-p)^{x+1} \text{ for } x = 0, 1, ... \end{aligned}$$

if $x \in \mathbb{R}$, then

$$F(x) = \begin{cases} 1-(1-p)^{\lfloor x \rfloor + 1}, \text{ if } x \geq 0 \\ 0, \text{ if } x < 0 \end{cases}$$

	Discrete Uniform	Hypergeometric	Binomial	Negative Binomial	Geometric	Poisson
function range parameters	DU$[a, b]$ $a, a+1, \dots, b$	Hyp(N, r, n) bad	Bin(n, p) $0, 1, \dots, n$	NB(k, p) $0, 1, \dots$	Geo(p) $0, 1, \dots$	Poi(μ) $0, 1, \dots$ $\mu = np$, $\mu = \lambda t$
pf, $f(x)$	$\frac{1}{b-a+1}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\binom{x+k-1}{k-1}p^k(1-p)^x$	$p(1-p)^x$	$\frac{e^{-\mu}\mu^x}{x!}$
cdf, $F(x)$	$\frac{x-a+1}{b-a+1}$				$1-(1-p)^{x+1}$	$e^{\mu}[1+\frac{\mu^1}{1!}+\cdots\frac{\mu^x}{x!}]$
how to tell	"equally likely" know min. & max.	know total # objects know # S's know # trials without replacement selecting a subset	Bernoulli trials know # trials count # S's	Bernoulli trials "until" "it take... to get" "before" know how many S's we want	"until we get" "before the first"	Bin. with large amount of trials, small prob rate specified (#events/time) no pre-specified max. events occurring at any time (randomly) Poisson process & know time & count events doesn't make sense to ask how often an event did **not** occur

Bernoulli trials:

- independent
- each outcome is a S or F
- P(success)=p=constant