STAT 230 - Probability

Cameron Roopnarine

Last updated: November 9, 2019

Contents

1 Lecture 13

Recall

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

1.1 Negative Binomial Distribution (5.5)

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

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

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. $X \sim \mathrm{NB}(10, \frac{1}{2})$

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 $\mathrm{NB}(40, p)$.

1.4 Range and Probability Function of the Negative Binomial Distribution

range $x \in \{0, 1, \dots\}$ (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:

$$\overbrace{\underbrace{\_\,\_\,\_ \cdots \_}_{(k-1)\text{ S's, } x\text{ F's}} \mid \underbrace{\text{S}}_{k\text{th S}}}^{x+(k-1)\text{ Trials}}$$

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. $Y \sim \mathrm{NB}(5, 0.2)$, 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

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. $Y \sim \mathrm{Bin}(50, 0.01)$.

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

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

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

2.2 Geometric Distribution (5.6)

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. $X \sim \mathrm{Geo}(p)$

2.3 Range and Probability Function of the Geometric Distribution

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

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

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

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

Proof.

$$\sum_{x=0}^{\infty}(1-p)^x p = \underbrace{p + p(1-p) + \dots}_{\text{(geo. series: } a = p,\ r = 1-p)}$$
$$= \frac{p}{1-(1-p)}$$
$$= 1$$

□

Find the cumulative distribution function.

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

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

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

	Discrete Uniform	Hypergeometric	Binomial	Negative Binomial	Geometric	Poisson
function range parameters	$\mathrm{DU}[a,b]$ $a, a+1, \dots, b$	$\mathrm{Hyp}(N,r,n)$ bad	$\mathrm{Bin}(n,p)$ $0, 1, \dots, n$	$\mathrm{NB}(k,p)$ $0, 1, \dots$	$\mathrm{Geo}(p)$ $0, 1, \dots$	$\mathrm{Poi}(\mu)$ $0, 1, \dots$ $\mu = np$, $\mu = \lambda t$
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

3 Lecture 15

3.1 Example

Naomi invites 12 people to her party. If each independently comes with probability p. Let $X = \#$ of guests.

Binomial: $X \sim \mathrm{Bin}(12, p)$

3.2 Example

20 toys in a machine. Each time you grab one with a claw. Let $X = \#$ of tries to get one toy you want.

None.

3.3 Example

Trying to catch a pokemon, each time has a probability p of succeeding. Let $X = \#$ of failed attempts.

Geometric: $X \sim \mathrm{Geo}(p)$

3.4 Example

You have 5 classes randomly scheduled in a row. Let $X = \#$ of classes before your favourite.

range: $0, 1, 2, 3, 4$, and the probability is $1/5$ for each of the range.

Discrete Uniform: $X \sim \mathrm{DU}[1, 4]$

3.5 Poisson Distribution from Binomial (5.7)

Suppose we have a $X \sim \mathrm{Bin}(n, p)$ where n is very large and p is very small. Then, as $n \to \infty$ and $p \to 0$ such that np remains constant, the probability function of X approaches a limit.

Let $np = \mu$, so $p = \frac{\mu}{n}$. Then

$$\begin{aligned}
\lim_{n\to\infty} f(x) &= \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\mu^x}{n^x} (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n\to\infty} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-x+1}{n} (1-\frac{\mu}{n})^n (1-\frac{\mu}{n})^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n\to\infty} (1-\frac{\mu}{n})^n \\
&= \frac{e^{-\mu}\mu^x}{x!}
\end{aligned}$$

We write: $X \sim \mathrm{Poi}(\mu)$, range: $0, 1, \cdots$

We can use the Poisson random variable as an approximation to the Binomial when n is large, and p is small. The only thing we need to do is $\mu = np$.

3.6 Example

Tim Hortons roll up the rim says 1 in 6 cups win a prize. Suppose you have 80 cups. Find the probability that you get 10 or fewer winners.

Let $X = \#$ of winning cups. $X \sim \text{Bin}(80, 1/6)$ We want

$$\begin{aligned}
F(10) &= P(X \leq 10) \\
&= \sum_{x=0}^{10} f(x) \\
&= \binom{80}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{80} + \cdots + \binom{80}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^{70} \\
&= 0.2002 \text{ (tedious)}
\end{aligned}$$

Try a Poisson approximation. $Y \sim \text{Poi}(\mu = np = \frac{80}{6} \approx 13.33)$. Then,

$P(Y \leq 10) = e^{-13.33}\left[1 + \frac{13.33}{1!} + \cdots + \frac{13.33^10}{10!}\right] = 0.224$

Not a good approximation since p was too large.

3.7 Poisson Distribution from Poisson Process (5.8)

The Poisson Process: Suppose events occur randomly in time or space according to three conditions:

(1) Independence: the number of events in one period cannot affect another non-overlapping period

(2) Individuality: events occur one at a time (cannot have two at the exact same time)

(3) Homogeneity or Uniformity: events occur at a constant rate

4 Lecture 16

Consider a Poisson Process with rate λ, (i.e. λ events occur on average per unit time). Observe the process for t units of time. Let $X = \#$ of events that occur. Then, $X \sim \text{Poi}(\mu)$, where $\mu = \lambda t$. That is,

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}$$

4.1 Example

Request coming in from a web server at a rate of 100 requests per minute. $\lambda = 100, t = \frac{1}{60}$ The # of requests per second would be

$$\text{Poi}\left(\mu = \frac{100}{60} = \frac{5}{3}\right)$$

4.2 Combining Other Models with the Poisson Process (5.9)

Problems may involve many different random variables!

4.3 Example (Continued)

We say that a second is quiet if it has no requests.

(a) Find probability that a second is quiet

(b) In a minute (60 non-overlapping seconds), find the probability of 10 quiet seconds

(c) Find the probability of having to wait 30 non-overlapping seconds to get 2 quiet seconds

(d) Given (c), find the probability of 1 quiet second in the first 15 seconds

(a) Let $X = \#$ requests in a second. $X \sim \text{Poi}(5/3)$.

We want $P(X=0) = \frac{e^{-\frac{5}{3}}(\frac{5}{3})^0}{0!} = 0.189$

(b) Let $Y = \#$ quiet seconds out of 60. $Y \sim \text{Bin}(60, 0.189)$.

We want $P(Y=10) = \binom{60}{10}(0.189)^{10}(0.811)^{50} = 0.124$

(c) Let $Z = \#$ non-quiet seconds before getting 2 quiet seconds. $Z \sim \text{NB}(2, 0.189)$.

We want $P(Z=28) = \binom{29}{1}(0.189)^2(0.811)^{28} = 0.003$

(d) Let $D = 1$ quiet second in the first 15 seconds, $D_x = \#$ of quiet seconds out of 15. $D_x \sim \text{Bin}(15, 0.189)$.

$$P(D_x = 1) = \binom{15}{1}(0.189)^1(0.811)^{14}$$

We get,

$$\begin{aligned}
P(D \mid \text{wait 30 to get 2 quiet}) &= \frac{P(D \text{ AND wait 30 to get 2 quiet})}{P(\text{wait 30 to get 2 quiet})} && (1)\\
&= \frac{P(D \text{ AND wait an additional 15 to get 1 additional quiet})}{P(C)} && (2)\\
&= \frac{P(D)P(\text{wait an additional 15 to get 1 additional quiet})}{P(C)} && (3)\\
&= \frac{\binom{15}{1}(0.189)^1(0.811)^{14} \times (0.811)^{14}(0.189)}{\binom{29}{28}(0.189)^2(0.811)^{28}} && (4)\\
&= \frac{\binom{15}{1}}{\binom{29}{28}} && (5)\\
&= \frac{15}{29} && (6)
\end{aligned}$$

In (3) we used the independence of non-overlapping time intervals and constant probability of events.

4.4 Summarizing Data on Random Variables (7.1)

Let $X = \#$ of kids in a family.

Value	Frequency
1	3
2	10
3	1
4	1

4.4.1 Definition (Median)

The *median* of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

4.4.2 Definition (Mode)

The *mode* of the sample is the value which occurs most often. There is no guarantee there will be only a single mode.

Mean: average $\rightarrow \frac{1\times3+2\times10+3\times1+4\times1}{15}$

Median: 2

Mode: 2

5 Lecture 17

5.1 Expectation of a Random Variable (7.2)

Imagine we know the theoretical probability of each # of kids in a family.

x	1	2	3	4	5
$f(x)$	0.43	0.4	0.12	0.04	0.01

Now we replace the observed proportion in the sample mean with $f(x)$

$$\sum_{\text{all } x} xf(x) = (1)(0.43) + (2)(0.4) + (3)(0.12) + (4)(0.04) + (5)(0.01) = 1.8$$

which is the theoretical mean.

Why do we have sample mean > theoretical mean?

- urban vs rural population
- income level
- sampled max family size but theoretical includes growing families
- selection bias: more likely to choose one with more sister/brothers from population

5.1.1 Definition (Expected Value)

Let X be a discrete random variable and probability function $f(x)$. The *expected value* (also called the mean or the expectation) of X is given by

$$\mu = E[x] = \sum_{\text{all } x} xf(x)$$

Remark 1. μ will be within the range but not necessarily equal to a possible value of x.

We might be interested in the expected value of some function of X, $g(X)$.

5.2 Example

Tax credit of $\$1000$ plus $\$250$ per kid. Find the average cost.

x	1	2	3	4	5
$g(x)$	1250	1500	1750	2000	2250

Average cost = weighted average of $g(x)$ values $= (1250)(0.43) + \cdots + (2250)(0.01) = 1450$

5.2.1 Theorem

Let X be a discrete random variable and probability function $f(x)$. The expected value of a some function $g(X)$ of X is given by

$$E[g(X)] = \sum_{\text{all } x} g(x)f(x)$$

Note that $g(x) = 1000 + 250x$ from last example.

$$E[g(X)] = 1000 + 250E[x] = 1450$$