§5.1, 5.3 Geometric and negative binomial distributions

Example Consider the random experiment of repeatedly tossing a coin with heads probability $p \in (0,1)$ until the first time you get heads. Let X count the number of tosses made (including the toss that gave the first heads). Find the ① range of X and ② the pmf of X.

① range is $\{1,2,3,\dots\}$

② $P(X=k) = (1-p)^{k-1}p$

Def A random variable X has the geometric distribution with parameter $p \in (0,1)$ if its range is $\{1,2,3,\dots\}$ and its probability mass function is $P(X=k) = (1-p)^{k-1}p$ for $k \geq 1$.

Shorthand: $X \sim Geom(p)$.

Remark Geometric distribution models number of trials up to and including first success in repeated independent trials of a success/failure experiment.

R commands: if $X \sim Geom(p)$

$P(X=k)$	dgeom(k-1, p)
$P(X \leq k)$	pgeom(k-1, p)

__Example__ Let $X \sim \text{Geom}(p)$. Show $E[X] = \frac{1}{p}$.

First, recall that $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ when $|x|<1$

Therefore, $\frac{1}{(1-x)^2} = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \sum_{k=1}^{\infty} kx^{k-1}$ when $|x|<1$.

Now, we see that

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k \cdot P(X=k) \\ &= \sum_{k=1}^{\infty} k(1-p)^{k-1} p \\ &= p \sum_{k=1}^{\infty} k(1-p)^{k-1} \\ &= p \frac{1}{(1-(1-p))^2} \\ &= \frac{1}{p} \end{aligned}$$

__Example__ Consider the random experiment of repeatedly tossing a coin with heads probability $p \in (0,1)$ until getting r heads, for a given integer $r \geq 1$. Let X count the number of tosses made (up to and including the toss that gave the rth head). Find the ① range of X and ② pmf of X.

① range is $\{r, r+1, r+2, \dots\}$

② $P(X=k) = \underbrace{\binom{k-1}{r-1}}_{\text{number of ways to distribute the first } r-1 \text{ heads in first } k-1 \text{ tosses.}} (1-p)^{k-r} p^r$ for $k \geq r$.

Def A random variable X has the <u>negative binomial distribution</u> with parameters $r \geq 1$ and $p \in (0,1)$ if its range is $\{r, r+1, r+2, \ldots\}$ and its probability mass function is given by $P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ for $k \geq r$.

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

Remark: ① counts independent trials until rth success

② if $X \sim NegBin(r,p)$, then $X = G_1 + \cdots + G_r$ where $G_1, \ldots, G_r \sim Geom(p)$ are i.i.d.

R commands: if $X \sim NegBin(r,p)$

$P(X=k)$	dnbinom(k-r, r, p)
$P(X \leq k)$	pnbinom(k-r, r, p)

Example Let $X \sim NegBin(r,p)$. Show $E[X] = \frac{r}{p}$.

Note that $X = G_1 + G_2 + \cdots + G_r$ where $G_1, G_2, \ldots, G_r \sim Geom(p)$ are i.i.d., with G_i counting the trials after the $(i-1)$st success to the ith success. Thus

$$E[X] = E[G_1 + \cdots + G_r]$$
$$= E[G_1] + \cdots + E[G_r]$$
$$= \frac{r}{p}.$$

<u>Example</u> Consider repeating independent trials of a success/failure experiment with success prob. $p \in (0,1)$.

① If there are r or fewer successes in the first n trials, what can you conclude about when the $(r+1)$st success occurs?

must be after nth trial

② If the $(r+1)$st success after the nth trial, what can you conclude about the number of successes in the first n trials?

must be r or fewer

③ Relate these ideas in terms of binomial and negative binomial random variables.

Let X count number of successes in first n trials.

Let Y count number of trials until $(r+1)$st success.

Then $X \sim Bin(n,p)$, $Y \sim NegBin(r+1, p)$

and ① says event $\{X \le r\}$ implies $\{Y > n\}$

② says event $\{Y > r\}$ implies $\{X \le r\}$

So they're the same event and $P(X \le r) = P(Y > r)$.

<u>Remark</u> This "dual" relationship between X and Y is where the name "negative binomial" comes from.

Problem 1. Suppose applicants for a job are hired with probability $p = 0.15$ independently from person to person, one at a time as they come in. Unless otherwise stated, use geometric or negative binomial random variables to do the following.

a. Find the probability of each of the following events.

1. It took exactly 3 applicant until the first person was hired.
2. It took at least 3 applicants until the first person was hired.
3. It took exactly 10 applicants until 3 people were hired.
4. It took at least 10 applicants until 3 people were hired.
5. It took exactly 15 applicants until 7 people were hired.
6. It took at least 15 applicants until 7 people were hired.
7. Repeat parts 4 and 6. using a binomially distributed random variable.

b. Find the expected value of the number of applicants seen until

1. a hire is made.
2. 3 hires are made.
3. 7 hires are made.

(a) Let $X \sim Geom(0.15)$, $Y \sim NegBin(3, 0.15)$,

$Z \sim NegBin(7, 0.15)$

(1) $P(X = 3)$

(2) $P(X \geq 3) = 1 - P(X \leq 2)$

(3) $P(Y = 10)$

(4) $P(Y \geq 10) = 1 - P(Y \leq 9)$

(5) $P(Z = 15)$

(6) $P(Z \geq 15) = 1 - P(Z \leq 14)$

(7) Let $B_1 \sim Bin(9, 0.15)$, $B_2 (14, 0.15)$

$P(Y \geq 10) = P(B_1 \leq 2)$

$P(Z \geq 15) = P(B_2 \leq 6)$

Problem 1

Part a:

```{r}
dgeom(3-1, 0.15) # part 1
1-pgeom(2-1,0.15) # part 2
dnbinom(10-3, 3, 0.15) # part 3
1-pnbinom(9-3, 3, 0.15) # part 4
dnbinom(15-7, 7, 0.15) # part 5
1-pnbinom(14-7,7,0.15) # part 6
pbinom(2,9,0.15) # part 7
pbinom(6,14,0.15) # part 7
```

```
[1] 0.108375
[1] 0.7225
[1] 0.03895012
[1] 0.8591466
[1] 0.001398124
[1] 0.9977925
[1] 0.8591466
[1] 0.9977925
```

(b) (1) $E[X] = \frac{1}{0.15}$

(2) $E[Y] = \frac{3}{0.15}$

(3) $E[Z] = \frac{7}{0.15}$