Chapter 8.1

Geometric Random Variable

Geometric experiment

The **geometric** experiment is very similar to a binomial experiment.

The difference is that instead of deciding in advance how many trials to perform and counting the number of successes, now we will repeat the trials until we observe the first success.

Geometric random variable

Definition 8.1 Suppose that independent trials, each of which results in a success with probability π and in a failure with probability $1-\pi$, are to be performed until we observe the first success. If X represents the number of failures that occur, then X is said to be a geometric random variable with parameter π. Its PMF is given by

$$f(x) = (1-\pi)^x \pi, \quad x = 0, 1, 2, \ldots$$

We write $X \sim Geom(\pi)$.

Possible outcomes

ex: FFS $X = 2$

Example 8.1

Suppose you roll a fair six-sided dice until you get a "6". What is probability that it will take you 20 rolls?

X = # failures before the 1st "6"

Model $X \sim Geom(\pi = 1/6)$ [we are assuming that the die rolls are ind]

$P(X=19) = \left(1-\frac{1}{6}\right)^{19} \cdot \frac{1}{6}$

Example 8.1 introduces the PMF of the geometric random variable:

$$\begin{aligned} P(X = x) &= P(\text{x failures followed by a success}), \\ &= (1-\pi)^x \pi, \ x = 0, 1, 2, \dots \qquad (8.1) \end{aligned}$$

which can be recognized as the terms in a geometric series:

$$a + ar + ar^2 + ar^3 + \dots$$

with ratio $r = (1-\pi)$ and coefficient $a = \pi$. Hence the name "geometric" for this random variable.

$r = (1 - \pi)$

If $|r| < 1$

$$a + ar + ar^2 + ar^3 + \dots = \sum_{x=0}^{\infty} a \cdot r^x = \frac{a}{1-r}.$$

We can use this result to show that the PMF of a geometric random variable in fact does sum to 1.

$$\sum_{x=0}^{\infty} \pi(1-\pi)^x = \frac{\pi}{1-(1-\pi)} = \frac{\pi}{\pi} = 1.$$

Example 8.2

$f(x) = (1-\pi)^x \cdot \pi, \ x = 0, 1, \dots$

Suppose you roll a fair six-sided dice until you get a "6". Calculate the probability that it will take you at least 20 rolls.

$X \sim \text{Geom}(\pi = 1/6)$

$P(X \geq 19) = P(X = 19) + P(X = 20) + P(X = 21) + \dots$

$X \sim Geom(\pi = 1/6)$

$P(X \geq 19) = P(X=19) + P(X=20) + P(X=21) + \cdots$

$= \left(\frac{5}{6}\right)^{19} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^{20} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^{21} \cdot \frac{1}{6} + \cdots$

$= \left(\frac{5}{6}\right)^{19} \left(\frac{1}{6} + \left(\frac{5}{6}\right) \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^{2} \cdot \frac{1}{6} + \cdots \right)$

$= \left(\frac{5}{6}\right)^{19} \frac{1/6}{1-5/6} = \left(\frac{5}{6}\right)^{19} \cdot \frac{\cancel{1/6}}{\cancel{1/6}} = \left(\frac{5}{6}\right)^{19}$

First "s" happens after 19 trials $\iff$ you must have failed in the first 19 trials

Redo: using summation notation

$$P(X \geq 19) = \sum_{x=19}^{\infty} P(X=x) = \sum_{x=19}^{\infty} \left(\frac{5}{6}\right)^{x} \cdot \frac{1}{6}$$

$$= \frac{1}{6} \sum_{x=19}^{\infty} \left(\frac{5}{6}\right)^{x}$$

$$= \frac{1}{6} \sum_{u=0}^{\infty} \left(\frac{5}{6}\right)^{u+19}$$

$\boxed{\text{change of variable } u = x - 19}$

$$\frac{1}{6} \sum_{u=0}^{\infty} \left(\frac{5}{6}\right)^{u} \left(\frac{5}{6}\right)^{19}$$

$$= \left(\frac{5}{6}\right)^{19} \cdot \frac{1}{6} \sum_{u=0}^{\infty} \left(\frac{5}{6}\right)^{u}$$

$$= \left(\frac{5}{6}\right)^{19} \frac{1}{6} \frac{1}{(1-5/6)}$$

[since $\sum_{u=0}^{\infty} r^u = \frac{1}{1-r}$] by geometric series.

$$= \left(\frac{5}{6}\right)^{19}$$

Example 8.2 show that for any non-negative integer k,

$$P(X \geq k) = (1-\pi)^k$$

(relevant for HW5 problem 3)

where $X \sim Geom(\pi)$.

In other words, the probability that there will be at least k failures before the first success is equal to the probability of k failures.

Calculations in R

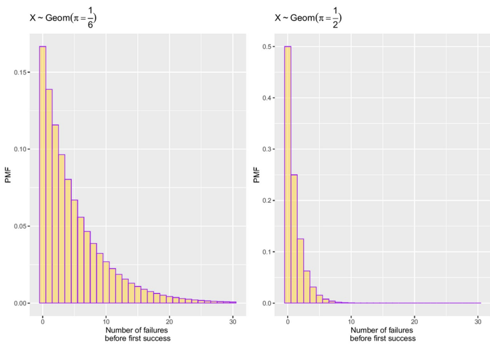
The R functions related to geometric random variable follow the same syntax as the binomial distribution and are shown below.

```
dgeom(x = 20, prob = 1/6)  #P(X = x)
## [1] 0.00435

pgeom(q = 20, prob = 1/6)  #P(X <= q)
## [1] 0.978

pgeom(q = 20, prob = 1/6, lower.tail = F)  #P(X > q)
## [1] 0.0217
```

Calculations in R



Calculations in R

```r
#attach packages in setup code chunk with include=FALSE
#library(patchwork)   #places graphs side by side

geom_df <- tibble(
  x = 0: 30,
  geomPMF1 = dgeom(x = x, prob = 1/6),
  geomPMF2 = dgeom(x = x, prob = 1/2)
)

p1 <- ggplot(data = geom_df,
             mapping = aes(x = x, y = geomPMF1) ) +
        geom_col(width = 1, alpha=0.5, fill = "gold", color="purple")+
  labs(x = "Number of failures \nbefore first success",
       y = "PMF",
      title=expression(X %~% Geom(pi == frac(1,6) )))

p2 <- ggplot(data = geom_df,
             mapping = aes(x = x, y = geomPMF2) ) +
        geom_col(width = 1, alpha=0.5, fill = "gold", color="purple")+
  labs(x = "Number of failures \nbefore first success",
      y = "PMF",
      title=expression(X %~% Geom(pi == frac(1,2) )))

p1 + p2
```

Expected value of $X \sim Geom(\pi)$

Theorem 8.2 Let $X \sim Geom(\pi)$. Then

$$E[X] = \frac{1-\pi}{\pi}.$$

Before we dive into the derivation, let's recall an earlier stated fact about the geometric series with ratio r and coefficient a:

$$a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1-r}$$

so long as $|r| < 1$. Differentiating both sides with respect to r, we have:

$$a + 2ar + 3ar^2 + \ldots = \frac{a}{(1-r)^2}. \tag{8.2}$$

Expected value of $X \sim Geom(\pi)$

$$\begin{aligned}
E[X] &= \sum_x x \cdot f(x) = \sum_{x=0}^{\infty} x(1-\pi)^x \pi, \\
&= \pi \sum_{x=1}^{\infty} x q^x, \quad q = 1-\pi \\
&= \pi \left[q + 2q^2 + 3q^3 + 4q^4 + \ldots \right], \\
&= \pi \frac{q}{(1-q)^2}
\end{aligned}$$

where we have used the result in (8.2) with $a = q$ and $r = q$.

Replacing q with $(1-\pi)$ yields:

$$E[X] = \pi \cdot \frac{1-\pi}{(1-(1-\pi))^2} = \frac{1-\pi}{\pi}.$$

The expected value calculation is intuitive. It emphasizes that our waiting time for the first success depends on the *odds* of a failure.

In the case of the fair six sided die, we should expect

$$\frac{\frac{5}{6}}{\frac{1}{6}} = 5$$

failures before we roll a "6".