§10. Geometric and Negative Binomial Random Variables

Geometric random variables capture the waiting time for the occurrence of a binomial event.

Example 1

Toss an unbalanced coin; probability for H is p.

Let X be the number of tosses when H shows up for a first time. Then

$$P(X = x) = (1-p)^{x-1}p$$

Definition 1

A **geometric random variable**, $X \sim \mathrm{Geom}(p)$ has pmf

$$P(x) = (1-p)^{x-1}p \qquad ; \qquad x = 1, 2, 3, 4, \ldots$$

Proposition 1

The expected value and the variance are

$$E(X) = \frac{1}{p} \qquad\qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Example 2

Ice-fishing in the Lake of Two Mountains: the probability of catching sturgeon is $p = 0.02$ on any given day.

a. What is the expected value and standard deviation for the number of days to catch a sturgeon?

b. What is the probability Rejean will catch sturgeon for the first time on day 50 of his ice - fishing?

Solution

Let X = the number of days until the first catch; with $p = 0.02$

a. $E(X) = \dfrac{1}{0.02} = 50$

$$\text{Var}(X) = \frac{1-0.02}{(0.02)^2} = 2450 \quad \Rightarrow \quad \sigma_X = \sqrt{2450} = 49.5$$

b. $P(X = 50) = (1-0.02)^{49}(0.02) = 0.007432$

Example 3

Toss an unbalanced coin; probability for H is p. Let X be the number of tosses when the r^{th} head appears. Then

$$P(X = x) = {}_{x-1}C_{r-1}(1-p)^{x-r}p^r \qquad x = r, r+1, \ldots$$

Definition 2

A **negative binomial random variable** $X \sim \text{NegBinom}(r, p)$ has pmf

$$P(x) = {}_{x-1}C_{r-1}(1-p)^{x-r}p^r \qquad\qquad x = r, r+1$$

Note: When $r = 1$ this is geometric random variable.

Let's compute the expected value

$$\begin{aligned} E(X) = \sum_{x=r}^{\infty} x \cdot {}_{x-1}C_{r-1}(1-p)^{x-r}p^r &= p^r \sum_{x=r}^{\infty} x \cdot \frac{(x-1)!}{(x-r)!(r-1)!}(1-p)^{x-r} \\ &= rp^r \sum_{x=r}^{\infty} {}_xC_r(1-p)^{x-r} \\ &= rp^r\{1 + {}_{r+1}C_1(1-p) + {}_{r+2}C_2(1-p)^2 + \cdots\} \\ &= rp^r(1-(1-p))^{-(r+1)} \\ &= rp^r p^{-(r+1)} \\ &= \frac{r}{p} \end{aligned}$$

Proposition 2

The expected value and the variance of a negative binomial RV are

$$E(X) = \frac{r}{p} \qquad\qquad \text{Var}(X) = \frac{r(1-p)}{p^2}$$

Note: If $X \sim \text{NegBinom}(r, p)$ then $X = X_1 + \cdots + X_r$; $X_i \sim \text{Geom}(p)$

$$E(X) = E(X_1 + X_2 + \cdots + X_r) = E(X_1) + E(X_2) + \cdots + E(X_r) = r \cdot \frac{1}{p}$$

$$\text{Var}(X) = \text{Var}(X_1 + X_2 + \cdots + X_r) \underset{\uparrow \atop independence}{=} \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_r)$$

$$= r \cdot \frac{1-p}{p^2}$$

Example 4

Catching sturgeon on the Lake of Two mountains; $p = 0.02$

a. What is the expected value for the number of days to catch 2 sturgeon?

b. What is the expected number of failures before 2 sturgeons are caught?

Solution

a. $E(X) = \dfrac{2}{0.02} = 100$

b. $E(Y) = \dfrac{r}{p} - r = 100 - 2 = 98.$

Example 5

Bob has two fair coins. Bob tosses them and sets a side any coins that come up heads. Bob continues tossing (on each toss removing the heads) until both coins are showing heads. On average, how many rounds of tosses does Bob have to make?

Solution

$Y = max\{X_1, X_2\}$ $\qquad X_1, X_2 \sim \text{Geom}(p)$

Two cases:

a) Both coins come up tails for $y - 1$ tosses and then both coins come up heads on the y^{th} toss.

$$P_1 = \left(\frac{1}{4}\right)^{y-1} \frac{1}{4}$$

b) There are $j-1$ tosses both coins show tails; then one coin comes up heads on the j^{th} toss. The remaining coin shows heads on the $y > j$ toss for the first time.

$$
\begin{aligned}
P_2 &= \sum_{j=1}^{y-1} \left(\frac{1}{4}\right)^{j-1} \left\{ {}_2C_1 \cdot \frac{1}{2} \cdot \frac{1}{2} \right\} \cdot \left(\frac{1}{2}\right)^{y-j-1} \cdot \frac{1}{2} \\
&= \sum_{j=1}^{y-1} {}_2C_1 \cdot \left(\frac{1}{4}\right)^{j} \cdot \left(\frac{1}{2}\right)^{y-j}
\end{aligned}
$$

Then

$$
\begin{aligned}
P = P_1 + P_2 &= \left(\frac{1}{4}\right)^{y} + \sum_{j=1}^{y-1} {}_2C_1 \left(\frac{1}{4}\right)^{j} \left(\frac{1}{2}\right)^{y-j} \\
&= \left(\frac{1}{4}\right)^{y} + 2\sum_{j=1}^{y-1} \left(\frac{1}{2}\right)^{y+j} \\
&= \left(\frac{1}{4}\right)^{y} + 2\left(\frac{1}{2}\right)^{y}\sum_{j=1}^{y-1} \left(\frac{1}{2}\right)^{j} \\
&= \left(\frac{1}{4}\right)^{y} + 2\left(\frac{1}{2}\right)^{y}\left[1 - \left(\frac{1}{2}\right)^{y-1}\right] \\
&= \frac{2^{y+1}-3}{4^y}
\end{aligned}
$$

Thus,

Y	1	2	3	$\cdots$
$P(y)$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{13}{64}$	$\cdots$

Example 6

A fair coin is tossed repeatedly. What is the probability that the H count reaches 5 before the T count reaches 3?

Solution

The 5^{th} head must be preceded by $0, 1$, or 2 tails.

$$
P = \underbrace{\left(\frac{1}{2}\right)^{5}}_{5H,0T} + {}_5C_4 \underbrace{\left(\frac{1}{2}\right)^{5}}_{5H} \cdot \underbrace{\left(\frac{1}{2}\right)}_{1T} + {}_6C_4 \underbrace{\left(\frac{1}{2}\right)^{5}}_{5H} \cdot \underbrace{\left(\frac{1}{2}\right)^{2}}_{2T} = \frac{29}{198}
$$