Department of Mathematics
College of William & Mary

MATH 451/551

Chapter 4. Common Discrete Distributions

4.3 Geometric Distribution

GuanNan Wang
gwang01@wm.edu

Mean

$X \sim Geo(p) \Rightarrow \mathcal{X} = \{0, 1, 2, \dots\} \Rightarrow X$ is discrete.

$$f(x) = p(1-p)^x$$

$$E(X) = \sum_{\mathcal{X}} x f(x) = \sum_{x=0}^{\infty} x\, p(1-p)^x = p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 + \cdots$$

$$\begin{array}{rlll}
= & p(1-p) + p(1-p)^2 + p(1-p)^3 + p(1-p)^4 + \cdots & + & p\,\dfrac{(1-p)}{1-(1-p)} = 1-p \\
 & \phantom{p(1-p) + {}} p(1-p)^2 + p(1-p)^3 + p(1-p)^4 + \cdots & + & p(1-p)\,\dfrac{(1-p)}{1-(1-p)} = (1-p)^2 \\
 & \phantom{p(1-p) + p(1-p)^2 + {}} p(1-p)^3 + p(1-p)^4 + \cdots & + & p(1-p)^2\,\dfrac{1-p}{1-(1-p)} = (1-p)^3 \\
 & \phantom{p(1-p) + p(1-p)^2 + p(1-p)^3 + {}} p(1-p)^4 + \cdots & & (1-p)^4 + \cdots
\end{array}$$

$$= (1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{(1-p)}{1-(1-p)} = \frac{1-p}{p}$$

Variance

$$V(X) = E(X^2) - \underbrace{\{E(X)\}^2}$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 p(1-p)^x = p(1-p) + \underbrace{4p(1-p)^2} + 9p(1-p)^3 + \cdots$$

$$= p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots + \qquad \frac{p(1-p)\, 1}{1-(1-p)} = 1-p$$

$$3p(1-p)^2 + 3p(1-p)^3 + \cdots + \qquad \frac{3p(1-p)^2 \cdot 1}{1-(1-p)} = 3(1-p)^2$$

$$5p(1-p)^3 + \cdots + \qquad \frac{5p(1-p)^3 \cdot 1}{1-(1-p)} = 5(1-p)^3 \cdots$$

$$=(1-p) + (1-p)^2 + (1-p)^3 + \cdots + \qquad \frac{1-p}{1-(1-p)} = \frac{1-p}{p}$$

$$2(1-p)^2 + 2(1-p)^3 + \cdots + \qquad \frac{2(1-p)^2}{1-(1-p)} = \frac{2(1-p)^2}{p}$$

$$+\ 2(1-p)^3 + \cdots + \qquad \frac{2(1-p)^3}{1-(1-p)} = \frac{2(1-p)^3}{p} + \cdots$$

$$= \frac{2\{(1-p) + (1-p)^2 + (1-p)^3 + \cdots\} - (1-p)}{p}$$

$$= \frac{2\,\frac{1-p}{1-(1-p)} - (1-p)}{p} = \frac{2(1-p) - p(1-p)}{p^2} = \frac{(1-p)(2-p)}{p^2}$$

$$V(X) = E(X^2) - \{E(X)\}^2$$

$$= \frac{(1-p)(2-p)}{p^2} - \frac{(1-p)^2}{p^2} = \frac{(1-p)(2-\cancel{p}-1+\cancel{p})}{p^2}$$

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

$$p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 + \cdots \qquad S \qquad S - (1-p)S = 1-p$$

$$S = \frac{1-p}{p}$$

$$p(1-p)^2 + 2p(1-p)^3 + \cdots \qquad (1-p)S$$

$$\underbrace{p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots}_{} = \frac{p(1-p)}{p} = (1-p)$$

$$1 + z + z^2 + z^3 + \cdots = \frac{1}{\cancel{1-z}}$$

$$1 + 2z + 3z^2 + 4z^3 + \cdots = \frac{1}{(1-z)^2}$$

MGF

$$M(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x$$

$$= p \sum_{x=0}^{\infty} \{e^t(1-p)\}^x$$

$$= p\left[1 + e^t(1-p) + \{e^t(1-p)\}^2 + \cdots\right]$$

$$= p \frac{1}{1-e^t(1-p)} = \frac{p}{1-(1-p)e^t}, \quad (1-p)e^t < 1$$

$$E(X) = \left.\frac{dM(t)}{dt}\right|_{t=0} = \left.\frac{d \frac{p}{1-(1-p)e^t}}{dt}\right|_{t=0} = \left.\frac{+p(1-p)e^t}{\{1-(1-p)e^t\}^2}\right|_{t=0}$$

$$= \frac{p(1-p)}{(1-1+p)^2} = \frac{1-p}{p}$$

Skewness

$$E\left\{\left(\frac{X-\mu}{\delta}\right)^3\right\} = \frac{2-p}{\sqrt{1-p}}$$

Kurtosis

$$E\left\{\left(\frac{X-\mu}{\sigma}\right)^4\right\} = \frac{p^2-9p+9}{1-p}$$

Memoryless Property

Memoryless Property

For $X \sim Geo(p)$ and any two nonnegative integers x and y,

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Interpretation

Consider a sequence of repeated, mutually independent, and identically distributed Bernoulli trials and a random variable X that is the number of failures before the first success. If you know that X is greater than or equal to x, then the distribution of the **remaining** number of Bernoulli trials before the first success has the same distributioon as if the original x trials had never occurred. This interpretation is consistent with intuition: the previous history of the sequence of Bernoulli trials has no effect on the outcomes of future Bernoulli trials.

LHS: $P(\underbrace{X \geq x+y}_{A} \mid \underbrace{X \geq x}_{B}) = \dfrac{P(X \geq x+y \cap X \geq x)}{P(X \geq x)}$

$A \subset B$

$$= \frac{P(X \geq x+y)}{P(X \geq x)}$$

$$= \frac{1-P(X < x+y)}{1-P(X<x)}$$

$$= \frac{1-P(X \leq x+y-1)}{1-P(X \leq x-1)} = \frac{1-F_X(x+y-1)}{1-F_X(x-1)}$$

$$= \frac{1-\{1-(1-p)^{x+y\cancel{-1+1}}\}}{1-\{1-(1-p)^{x\cancel{-1+1}}\}} = \frac{(1-p)^{\cancel{x}+y}}{(1-p)^{\cancel{x}}}$$

$$= (1-p)^y$$

RHS: $P(x \geq y)$

$= 1-F_X(y-1)$

$= (1-p)^y$

Example 1

$P(\text{"double six"}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Roll a pair of fair dice repeatedly until a "double six" appears. Let X be the number of rolls prior to the appearance of the first double six. Find $f(x)$, $E(X)$, $V(X)$, and $P(X < 24)$.

$X \sim Geo\left(\frac{1}{36}\right)$

$f(x) = \frac{1}{36}\left(\frac{35}{36}\right)^x, \quad x = 0, 1, 2, \cdots$

$E(X) = \frac{1-p}{p} = \frac{35/36}{1/36} = 35$

$V(X) = \frac{1-p}{p^2} = \frac{35/36}{(1/36)^2} = 1260$

$P(X < 24) = P(X \le 23) = F(23) = 1 - \left(1 - \frac{1}{36}\right)^{24} \approx 0.4914$

R Functions

Function	Returned Value
dgeom(x, p)	calculates the probability mass function $f(x)$
pgeom(x, p)	calculates the cumulative mass function $F(x)$
qgeom(u, p)	calculates the percentile (quantile) $F^{-1}(u)$
rgeom(m, p)	generates m random variates

X = # failures before 1st success

X = # trails to get 1st success

Alternative Definition

The geometric distribution can also be parameterized as the trial number of the first success.

- **PMF:** A discrete random variable X with PMF

$$f(x) = p(1-p)^{(x-1)}, \qquad x = 1, 2, \ldots$$

  for $0 < p < 1$ is a *Geometric*(p) random variable.
- This shifts the distribution one unit to the right.

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1-p}{p^2}.$$

Example 2

How many tosses of a pair of fair dice are necessary to be 99% certain that a double six will appear?

Thank You

THANK YOU!