THE Geometric Distribution

If trials are independent and probability of success is constant

X = # of trials until 1st success is observed

$x = 1, 2, 3 \ldots$ (no upper bound)

$X \sim Geo(p)$

Defining S = success $\quad F$ = failure

$P(S) = p$
$P(F) = 1-p$

$$P(X=s) = P(F_1 \cap F_2 \cap \ldots \cap F_{x-1} \cap S_x)$$
$$= P(F_1)\,P(F_2)\,P(F_3) \ldots P(F_{x-1})\,P(S_x)$$
$$= p(1-p)^{x-1}$$

THIS DISTRIBUTION **CANNOT** BE USED ON YOUR CALCULATOR.

$P(X \leq x)$ (success in less than or equal to x trials)

$$= \sum_{k=1}^{x} p(1-p)^{k-1} = p\sum_{k=1}^{x}(1-p)^{k-1} = p\big(\underbrace{1 + (1-p)^1 + (1-p)^2 + \ldots + (1-p)^{x-1}}_{\text{geometric series}}\big)$$

$a = 1, \; r = 1-p \qquad S_x = \dfrac{1(1-(1-p)^x)}{1-(1-p)} = \dfrac{1-(1-p)^x}{p}$

$$P(X \leq x) = p\left(\frac{1-(1-p)^x}{p}\right) = 1-(1-p)^x$$
$$P(X > x) = 1 - P(X \leq x) = 1 - (1-(1-p)^x) = (1-p)^x$$
$$P(X \geq x) = P(X > x-1) = (1-p)^{x-1}$$
$$P(X < x) = P(X \leq x-1) = 1-(1-p)^{x-1}$$

All the geometric distribution formulas you will ever need

Mean of Geometric Distribution.

reminder: $\frac{d}{dx}(x^n) = nx^{n-1}$

$\frac{d}{dp}((1-p)^x) = -x(1-p)^{x-1}$

$$E[X] = \sum_{x=1}^{\infty} xP(X=x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = p\sum_{x=1}^{\infty} x(1-p)^{x-1}$$

$\frac{d}{dx}(y_1) + \frac{d}{dx}(y_2) = \frac{d}{dx}(y_1+y_2)$

$$= -p\sum_{x=1}^{\infty} \frac{d}{dp}(1-p)^x = -p\frac{d}{dp}\left\{\underbrace{\sum_{x=1}^{\infty}(1-p)^x}_{\text{geometric series}}\right\}$$

$a = 1-p \quad r = 1-p$

$S_\infty = \frac{a}{1-r} = \frac{1-p}{1-(1-p)} = \frac{1-p}{p} = \frac{1}{p} - 1$

$$= -p\frac{d}{dp}\left(\frac{1}{p} - 1\right)$$

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

Variance of Geometric Distribution

$$Var(X) = E[X(X-1)] + E[X] - (E[X])^2$$

$\frac{d}{dp}\{(x-1)(1-p)^x\} = -x(x-1)(1-p)^{x-1}$

$\frac{d}{dp}\{(1-p)^{x-1}\} = -(x-1)(1-p)^{x-2}$

$$E[X(X-1)] = p\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-1}$$

$$= -p\sum_{x=2}^{\infty} \frac{d}{dp}\left((x-1)(1-p)^x\right)$$

$$= -p\frac{d}{dp}\left\{\sum_{x=2}^{\infty}(x-1)(1-p)^x\right\}$$

take out $(1-p)^2$

$$= -p\frac{d}{dp}\left\{(1-p)^2\sum_{x=2}^{\infty}(x-1)(1-p)^{x-2}\right\} = p\frac{d}{dp}\left\{(1-p)^2\sum_{x=2}^{\infty}\frac{d}{dp}\left((1-p)^{x-1}\right)\right\}$$

$$= p\frac{d}{dp}\left\{(1-p)^2\underbrace{\frac{d}{dp}\left(\sum_{x=2}^{\infty}(1-p)^{x-1}\right)}\right\}$$

$\frac{d}{dp}\left\{\sum_{x=2}^{\infty}(1-p)^{x-1}\right\} = \frac{d}{dp}\left\{\underbrace{(1-p) + (1-p)^2 + \ldots}_{\text{geometric series}}\right\}$

$= \frac{d}{dp}\left(\frac{1}{p} - 1\right) = -\frac{1}{p^2}$

$a = 1-p \quad r = 1-p$

$S_\infty = \frac{a}{1-r} = \frac{1-p}{p} = \frac{1}{p} - 1$

$$= p\frac{d}{dp}\left\{(1-p)^2\left(-\frac{1}{p^2}\right)\right\}$$

$$= -p\frac{d}{dp}\left(\frac{1-2p+p^2}{p^2}\right) = -p\frac{d}{dp}\left(p^{-2} - 2p^{-1} + 1\right) = -p\left(-2p^{-3} + 2p^{-2}\right) = \frac{2}{p^2} - \frac{2}{p}$$

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

Ex (3A)

(4) Pass an exam to get into law school, probability of passing = 0.3

a) i) $X = \#$ of attempts $X \sim Geo(0.3)$

$P(X=3) = (0.3)(0.7)^2 = 0.147$

ii) $P(X \geq 4) = P(X \leq 3) = 1-(0.7)^3 = 0.657$

b) Independent trials, constant probability.

(7) $X \sim Geo(0.032)$

a) $P(X=x) = p(1-p)^{x-1} = 0.032(0.968)^{x-1} = 0.0203$

$\ln 0.032 + (x-1)\ln 0.968 = \ln 0.0203$

$x = 1 + \dfrac{\ln 0.0203 - \ln 0.032}{\ln 0.968} = 14.99 \approx 15$

b) largest x s.t. $P(X \leq x) < 0.1$

$P(X \leq x) = 1-(1-p)^x = 1-0.968^x < 0.1$

$0.968^x > 0.9$

$x \ln 0.968 < \ln 0.9$

$x < \dfrac{\ln 0.9}{\ln 0.968} < 3.23955...$

$x_{max} = 3$

c) smallest x s.t. $P(X \geq x) < 0.05$

$P(X \geq x) = 1 - P(X \leq x-1) = 1-(1-(1-p)^{x-1}) = (1-p)^{x-1} = 0.968^{x-1} < 0.05$

$(x-1)\ln 0.968 > \ln 0.05$

$x > 1 + \dfrac{\ln 0.05}{\ln 0.968} = 93.11$

$\therefore x_{min} = 94$

(10) Roll 3 dice: O=odd, E= even

$\{OOO \text{ or } EEE\}$: fail $\{OOE, OEO, OEE, EOO, EOE, EEO\}$: success

$P(\text{success}) = 3/4$

$X \sim Geo(3/4)$

a) $P(X=3) = \frac{3}{4}\left(\frac{1}{4}\right)^2 = \frac{3}{64} = 0.046875$

b) $P(X \geq 4) = \left(\frac{1}{4}\right)^3 = \frac{1}{64} = 0.015625$

(11) a) X= # of customers per hour $X \sim Po(4)$

$P(X \geq 5) = 1 - P(X \leq 4) = 0.3712$

b) Y = # of hours until ≥5 customers enter $Y \sim Geo(0.3712)$

$P(Y=5) = (0.3712)(1-0.3712)^4 = 0.05803$

$P(Y>8) = P(Y \geq 9) = (1-0.3712)^8 = 0.02444$

Ex (3B)

(7) a) X= # of attempts to parallel park the car

$X \sim Geo(P)$

b) P is constant, parallel parking is independent

c) $P(X=2) = 0.16 = p(1-p)$ $p = 0.8,\ 0.2 = 0.2$ (rej. 0.8 since $p < 0.5$)

$p^2 - p + 0.16 = 0$

d) $E[X] = \frac{1}{p} = 5$

e) $Var(X) = \frac{1-p}{p^2} = \frac{1-0.2}{0.2^2} = 20$

THE NEGATIVE BINOMIAL DISTRIBUTION

let X = # of trials until the r^{th} success is observed (Same assumptions as geometric)

if $r=1$, this is geometric

- p constant
- independent trials

If $X=K$, then there are precisely $r-1$ successes observed in the first $K-1$ trials

$Y \sim B(K-1, p)$

$$P(Y=r-1) = \binom{K-1}{r-1} p^{r-1} (1-p)^{(K-1)-(r-1)}$$

model Y using normal binomial

$$P(X=K) = P(Y=r-1) \times \underline{P(K^{th} \text{ trial is a success})}$$

$$= \binom{K-1}{r-1} p^{r-1} (1-p)^{K-r} \times p$$

$$= \binom{K-1}{r-1} p^{r} (1-p)^{K-r}$$

this is just p.

for r successes to be observed in K trials, there must be r-1 successes in K-1 trials **AND** the K^{th} trial is a success.

$X \sim NB(r, p)$ special case ($NB(1, p) = Geo(p)$)

$$P(X=K) = \binom{K-1}{r-1} p^{r} (1-p)^{K-r}$$

"Negative" binomial because instead of fixed # of trials $\longrightarrow$ # of successes,

we use this for fixed # of successes $\longrightarrow$ # of trials

Expected Value of $X \sim NB(r, p)$

If $G_i \sim Geo(p)$ where $i = 1, 2, \ldots, r$ and all are independent of each other,

$$X = G_1 + G_2 + G_3 + \ldots + G_r = \sum_{i=1}^{r} G_i \Rightarrow X \sim NB(r, p)$$

# of trials for 1st success + # of trials for 2nd success + ... + # of trials for rth success

= # of trials for r successes = NB(r, p)

$$E[X] = E\left[\sum_{i=1}^{r} G_i\right] = \sum_{i=1}^{r} E[G_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

Variance of $X \sim NB(r, p)$

using the geometric distribution sum from above

$$Var(X) = Var\left(\sum_{i=1}^{r} G_i\right) = \sum_{i=1}^{r} Var(G_i) = \sum_{i=1}^{r} \left(\frac{1-p}{p^2}\right) = \frac{r(1-p)}{p^2}$$

Exercise 3D

5. $X \sim NB(r, 0.4)$

a) $E(X) = \frac{r}{p} = \frac{r}{0.4} = 15 \qquad r = 15 \times 0.4 = 6$

b) i) $P(X=10) = \binom{k-1}{r-1} p^r (1-p)^{k-r} = \binom{9}{5} 0.4^6 \times 0.6^4 = 0.0669$

ii) $P(X \leq 8) = \sum_{k=6}^{8} \left\{ \binom{k-1}{5} \times 0.4^6 \times 0.6^{k-6} \right\} = 0.0498$

8. a) $2 \times \left(\frac{1}{2}\right)^4 = \frac{1}{2^3} = \frac{1}{8} = 0.125$

b) let X = # of trials to observe 3 successes (success = 3 coins on same side)

$X \sim NB(3, 0.125) \quad P(X=6) = \binom{5}{2} \times 0.125^3 \times 0.875^3 = 0.0131$

c) let Y = # of trials to observe 12 successes $\quad Y \sim NB(12, 0.125)$

$E[Y] = \frac{r}{p} = \frac{12}{0.125} = 96$