Distributions July, 27, 2020

---

1) Geometric Distribution

Application: The Copoun Collector's Problem

2) Poisson Distribution

Question:
What is the expected number of times we have to roll a fair 6-sided die until we roll a 6?

$Pr[\omega \neq 6] = \left(\frac{5}{6}\right)^{n}$

$\lim_{n \to \infty} \left(\frac{5}{6}\right)^{n} = 0$

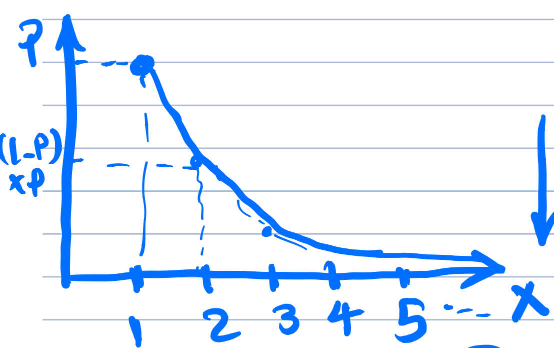
$< 1$

## 1) Geometric Distribution

Flip a coin with $Pr[H]=p$ until we get H.

r.v. $X$ = the number of flips until the first Heads appears.

Pr For instance




$$\left.\begin{array}{l} \omega_1 = H \\ \omega_2 = TH \\ \omega_3 = TTH \\ \vdots \\ \omega_i = T\cdots TH \end{array}\right\} \quad \begin{array}{l} Pr \\ p \\ (1-p)p \\ (1-p)^2 p \\ \vdots \\ (1-p)^{i-1} p \end{array}$$

$\Omega = \{\omega_i,\ i=1,2,\dots\}$

$X(\omega_i) = i$

Then $Pr[X=i] = (1-p)^{i-1}p$.

Definition: A random variable $X$ for which

$$Pr[X=i] = (1-p)^{i-1}p$$

is said to have the geometric distribution with parameter $p$

$$X \sim Geometric(p)$$

Sanity check: $\sum_{i=1}^{\infty} Pr[X=i] = 1$ ?

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1-a} \qquad |a|<1$$

$$\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1} = p\sum_{i=0}^{\infty}(1-p)^i = p \times \frac{1}{1-(1-p)}$$

$$= p \times \frac{1}{p} = 1$$

$$\boxed{Pr[X \geq i] = (1-p)^{i-1}}$$

If $X \sim \text{Geometric}(p)$,

$$E[X] = ? \qquad Var(X) = ?$$

$$E[X] = \sum_{a} a \Pr[X=a] = \sum_{i=1}^{\infty} i \Pr[X=i] = \sum_{i=1}^{\infty} i \underline{(1-p)^{i-1} p}$$

Theorem: For $X \sim \text{Geometric}(p)$

$$E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$$

$$(1-p)^{i-1} - (1-p)^{i} = (1-p)^{i-1} p$$

$$E[X] = \sum_{i=1}^{\infty} i \Pr[X=i] = \sum_{i=1}^{\infty} i \left( \Pr[X \geq i] - \Pr[X \geq i+1] \right)$$

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

$$= \sum_{i=1}^{\infty} i \Pr[X \geq i] - \sum_{i=1}^{\infty} i \Pr[X \geq i+1]$$

$$= \sum_{i=1}^{\infty} i \Pr[X \geq i] - \sum_{i=1}^{\infty} (i-1) \Pr[X \geq i]$$

$$= \sum_{i=1}^{\infty} \Pr[X \geq i] \left( i - (i-1) \right) = \sum_{i=1}^{\infty} \Pr[X \geq i]$$

Alternatively: $$= \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

$$E[X] = p + 2p(1-p) + 3p(1-p)^2 + 4p(1-p)^3 + \cdots$$

$$-\ (1-p)E[X] = 0 + p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 + \cdots$$

$$pE[X] = p + p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots$$

$$= \sum_{i=1}^{\infty} (1-p)^{i-1} p = 1 \Rightarrow pE[X] = 1 \Rightarrow \boxed{E[X] = \frac{1}{p}}$$

Remember $E[X] = \sum_{a \in A} a \Pr[X=a]$ for r.v. $X$.

Then

Define r.v. $Y = g(X) \Rightarrow E[Y] = E[g(X)] = \sum_{a \in A} \overset{Y}{g(a)} \Pr[X=a]$ → LOTUS

↳ function

---

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

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

$$E[X^2] = p + 4(1-p)p + 9(1-p)^2 p + 16(1-p)^3 p + \dots$$

$$-\quad (1-p)E[X^2] = 0 + (1-p)p + 4(1-p)^2 p + 9(1-p)^3 p + \dots$$

---

$$pE[X^2] = p + 3(1-p)p + 5(1-p)^2 p + 7(1-p)^3 p + \dots$$

$$= \sum_{i=1}^{\infty} (2i-1)(1-p)^{i-1} p$$

$$= 2\sum_{i=1}^{\infty} i(1-p)^{i-1} p - \sum_{i=1}^{\infty} (1-p)^{i-1} p$$

$$= 2 \quad E[X] - 1 = \frac{2}{p} - 1.$$

$$\Rightarrow pE[X^2] = \frac{2}{p} - 1 \Rightarrow \boxed{E[X^2] = \frac{2}{p^2} - \frac{1}{p}}$$

So

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

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

The coupon Collector's Problem:

- There are n different types of baseball cards.
- We get the cards by buying boxes of cereal
- Each box contains exactly one card
- This card is equally likely to be any of the n cards.

$S_n$ = The number of boxes we need to buy in order to collect all n cards.

What is $E[S_n]$ ?

Define: $X_i$ = # of cards we buy before we get the $i^{th}$ new card.

Then

$$S_n = X_1 + X_2 + \cdots + X_n$$

So $E[S_n] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$

What is the distribution for $X_i$?

$X_1$:

$Pr[X_1=1] = 1 \Rightarrow E[X_1] = 1 \times Pr[X_1=1] = 1$

$X_2$:

$Pr[\text{old card}] = \frac{1}{n}$

$Pr[\text{new card}] = \frac{n-1}{n}$

$X_2 \sim \text{Geometric}\left(\frac{n-1}{n}\right)$

$E[X_2] = \frac{n}{n-1}$.

$X_3$:

$Pr[\text{old card}] = \frac{2}{n}$

$Pr[\text{new card}] = \frac{n-2}{n}$

$X_3 \sim \text{Geometric}\left(\frac{n-2}{n}\right)$

$E[X_3] = \frac{n}{n-2}$

$\vdots$

$X_i$: $Pr[\text{old card}] = \frac{i-1}{n}$

$Pr[\text{new card}] = \frac{n-(i-1)}{n}$

$X_i \sim \text{Geometric}\left(\frac{n-(i-1)}{n}\right)$

$E[X_i] = \frac{n}{n-(i-1)}$

$$E[S_n] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{i}$$

exercise

## 2) Poisson Distribution:

Assum: The average number of passing cars passing through a tunnel per uint time is $\lambda$.

Define $X$ = The number of cars passing per unit time.

Question: what is the probability distribution of $X$?

Poisson distribution.

Definition: A random variable $X$ for which

$$\boxed{Pr[X=i] = \frac{\lambda^i}{i!} e^{-\lambda}} \quad i = 0, 1, \dots, n$$

is said to have Poisson distribution.

$$X \sim \text{Poisson}(\lambda)$$

Sanity check: $\sum_{i=0}^{\infty} Pr[X=i] = 1$

$$\sum_{i=0}^{\infty} Pr[X=i] = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \times e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{e^{\lambda}} = e^{-\lambda} \times e^{\lambda} = 1$$

what is $E[X]$? $\lambda$ the average

$$E[X] = \underbrace{\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!}}_{} e^{-\lambda} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!}$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{}$$

$$= \lambda \underbrace{e^{-\lambda} \times e^{\lambda}}_{} = \lambda.$$

What about $Var(X)$?

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

$$E[X^2] = \sum_{i=0}^{\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=1}^{\infty} i \frac{\lambda^i}{(i-1)!}$$

$$= e^{-\lambda} \sum_{i=1}^{\infty} (i-1+1) \frac{\lambda^i}{(i-1)!} = e^{-\lambda} \sum_{i=1}^{\infty} (i-1) \frac{\lambda^i}{(i-1)!} + e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!}$$

$$= e^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^i}{(i-2)!} + e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!}$$

$$= e^{-\lambda} \lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} + e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

$$= e^{-\lambda} \lambda^2 \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{\to e^{\lambda}} + e^{-\lambda} \lambda \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{e^{\lambda}}$$

$$= e^{-\lambda}\lambda^2 e^{\lambda} + e^{-\lambda}\lambda e^{\lambda} = \lambda^2+\lambda = E[X^2]$$

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

$$\boxed{Var(X) = \lambda}$$

Theorem: Let $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$ be independent Poisson random variables.

Then

$$Z = X+Y \sim Poisson(\lambda+\mu)$$

$$Pr[Z=k] = \frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)}$$

Proof:

$$Pr[Z=k] = Pr[X+Y=k]$$

total Probability

$$= \sum_{i=0}^{k} Pr[X=i, Y=k-i]$$

independence

$$= \sum_{i=0}^{k} Pr[X=i] \times Pr[Y=k-i]$$

$$= \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} \times \frac{\mu^{k-i}}{(k-i)!} e^{-\mu}$$

$$= e^{-(\lambda+\mu)} \sum_{i=0}^{k} \frac{1}{i!\,(k-i)!} \lambda^i \mu^{k-i}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda^i \mu^{k-i}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \underbrace{\sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i}}_{(\lambda+\mu)^k}$$

$$= \frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)}$$

## Poisson as a Limit of Binomial

**Theorem:** Let $X \sim \text{Binomial}(n, \frac{\lambda}{n})$ where $\lambda > 0$ is a fixed constant. Then

$$n \to \infty \Rightarrow P[X=i] \to \frac{\lambda^i}{i!} e^{-\lambda}$$

**Proof:**

$$Pr[X=i] = \binom{n}{i} p^i (1-p)^i \quad , \quad p = \frac{\lambda}{n}$$

The idea is that we do a testing on many sample points (n) and we get $\lambda$ success for the occurance of the n events

So $\lambda$ is defined

$$\underbrace{\lambda}_{\text{average numbers of success in } n.} = \underbrace{n}_{\text{sample Points}} \underbrace{p}_{\text{success rate}}$$

where we don't know $p$ but we can get an estimate of it by testing many $(n \to \infty)$

Points and Counting the successful event as $\lambda$ so $\quad p = \frac{\lambda}{n} \Rightarrow X \sim \text{Binomial}(n, \frac{\lambda}{n})$

now we take the limit of $n \to \infty$.

$$\Pr[X=i] = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}$$

$$= \frac{n!}{n^i (n-i)!} \frac{\lambda^i}{i!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-i}}_{1}$$

$$n \to \infty \begin{cases} \dfrac{n!}{n^i (n-i)!} = \dfrac{n(n-1)\cdots(n-i+1)}{n^i} \to 1 \\ \left(1-\dfrac{\lambda}{n}\right)^n \to e^{-\lambda} \\ \left(1-\dfrac{\lambda}{n}\right)^i \longrightarrow (1-0)^{-i} = 1 \end{cases}$$

$$\Rightarrow \Pr[X=i] = 1 \times \frac{\lambda^i}{i!} \times e^{-\lambda} \times 1$$

$$\Rightarrow \Pr[X=i] = \frac{\lambda^i}{i!} e^{-\lambda}$$