* Announcement :

① Exam-prep section

Weekly Friday 2-3 pm

② Study group assignment

③ HW4 due next Mon (9/21)

After today's lecture

⇝ HW4 Q4

STAT 88: Lecture 10

Contents

Last time

<u>Sec 4.2</u> Waiting Times

Waiting time for first success: We have independent and identically distributed (i.i.d.) trials with probability p for success and q for failure.

Let $T_1 = \#$ trials until the first success. $q = 1-p$

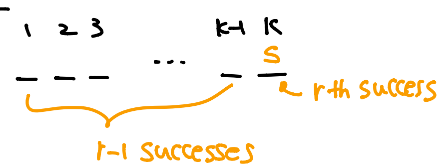
$$T_1 \sim \text{Geom}(p).$$

Then

$$P(T_1 = k) = q^{k-1}p.$$

$$P(T_1 > k) = q^k.$$

Picture



Waiting time for rth success:

Let $T_r = \#$ trials until the rth success.

$$\begin{aligned} P(T_r = k) &= P(r-1 \text{ of the first } k-1 \text{ trials are successes and trial } k \text{ is a success}) \\ &= \binom{k-1}{r-1} p^{r-1} q^{k-1-(r-1)} \cdot p = \binom{k-1}{r-1} p^{r-1} q^{k-r} \cdot p. \end{aligned}$$

$$P(T_r > k) = P(\text{at most } r-1 \text{ successes in the first } k \text{ trials}) = \sum_{j=0}^{r-1} \binom{k}{j} p^j q^{k-j}.$$

Warm up: (Exercise 4.5.10) Suppose you are running independent success/failure trials with probability 0.7 of success on each trial.

(a) What is the chance your first success is on the 3rd trial?

(b) What is the chance that you get 10 failures before the 15th success?

(a) $\underline{F}\ \underline{F}\ \underline{S}$ $\quad q^2 p = 0.3^2 \cdot 0.7$

(b)

$$\underbrace{\_\ \_\ \_\ \cdots\ \_}_{\substack{14 \text{ successes} \\ + 10 \text{ failures}}}\ \underline{S}$$

15th success

$$\binom{24}{14} p^{14} q^{10} \cdot p = \binom{24}{14} (0.7)^{15} (0.3)^{10}$$

4.3. Exponential Approximations

A useful approximation from Calculus:

$$\log(1+\delta) \approx \delta \text{ for small } \delta.$$

base e

Suppose we have n i.i.d. success/failure trials where n is large and p is small and average number of successes $\mu = np$.

$q = 1 - p$ Chance of failure

$(1-p)^n$ Chance no success in n trials

$(1-p)^n \approx e^{-np} = \boxed{e^{-\mu}}$ Approximation of chance of no success in n trials

Only depends on $\mu = np$

$x = (1-p)^n$

$\log x = n \log(1-p)$ $\delta = -p$

$\approx n(-p)$

$\rightsquigarrow x \approx e^{-np} = e^{-\mu}$

$\rightsquigarrow 1 - e^{-\mu} \simeq$ Chance that at least one success in n trials

Even though p small, if # trials n is large,

then at least one success w/ prob $\approx 1 - e^{-\mu}$

$\mu = 5$

$1 - e^{-\mu} > 0.99$

Example: A Chapter of a book has $n = 100,000$ words and the chance a word of the chapter has a misprint is very small $p = 1/1,000,000$. Give an approximation of the chance the chapter doesn't have a misprint.

A word is a trial, a misprint is a success

P(no misprint out of n words)

$= (1-p)^n$

$\approx e^{-np}$

$= e^{-0.1}$

4.4. The Poisson Distribution

count

Let X be the number of times a rare event occurs. Then

$$X \sim \text{Pois}(\mu).$$

parameter

Poisson formula:

$$P(X = k) = \frac{e^{-\mu}\mu^k}{k!} \text{ for } k = 0, 1, 2, \ldots.$$

Here μ is the average of counts

This sums to 1:

$$\sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} = e^{-\mu} \underbrace{\sum_{k=0}^{\infty} \frac{\mu^k}{k!}}_{= e^{\mu}} = e^{-\mu}e^{\mu} = 1.$$

In python, stats.poisson.pmf(k, μ) ← $P(X=k)$

stats.poisson.cdf(k, μ) ← $P(X \leq k) = F(k)$

Pois(μ) as an approximation of Binomial(n,p) when n is large and p is small

Let $X \sim \text{Pois}(\mu)$ and $Y \sim \text{Binomial}(n,p)$ where $\mu = np$ and n is large and p is small. We show that $P(X=0) \approx P(Y=0)$:

Poisson formula
$P(X=k) = \frac{e^{-\mu} \cdot \mu^k}{k!}$

$$P(Y=0) = \underbrace{\binom{n}{0}}_{=1}\underbrace{p^0}_{=1}(1-p)^n = (1-p)^n \underset{\text{Exp. Approx.}}{\approx} e^{-np} = e^{-\mu} = P(X=0).$$

Similarly,

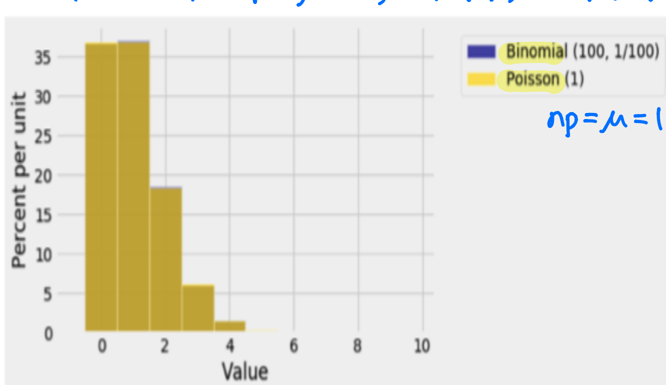
$$P(Y=1) = \underbrace{\binom{n}{1}}_{=n}p^1\underbrace{(1-p)^{n-1}}_{=\frac{(1-p)^n}{1-p}} = \underbrace{np}_{=\mu}\frac{(1-p)^n}{\underbrace{1-p}_{\approx 1}} \approx \mu e^{-\mu} = P(X=1).$$

Similarly,

$$P(Y=2) = \binom{n}{2} p^2(1-p)^{n-2} = \frac{n\overbrace{(n-1)}^{\approx n}}{2} p^2 \frac{(1-p)^n}{\underbrace{(1-p)^2}_{\approx 1}} \approx \frac{n^2p^2}{2}(1-p)^n$$
$$\approx \frac{n^2p^2}{2} e^{-np} = \frac{\mu^2}{2}e^{-\mu}$$
$$= P(X=2)$$

Poisson distribution is approximately Binominal distribution for small p and large n.

Formally, as $n \to \infty$, $p \to 0$ w/ $np \to \mu > 0$, $P(Y=k) \to P(X=k)$



In other words, if $X \sim \text{Pois}(\mu)$, then it represents the number of an event with small chance of success out of many independent trials.

It is rare that a word in a chapter of a book has a misprint. There are many words in a chapter and each is an independent trial for a misprint. The number of words in a Chapter that are misprinted can be modeled by a Poisson distribution.

A word is a trial, a misprint is a success

n words / p success prob.

$n = 100{,}000$, $p = \frac{1}{1000{,}000}$

$X = \#$ misprinted words $\sim \text{Binomial}(n, p)$

$\sim \text{pois}(\mu)$

$\mu = np = 0.1$

Sums of independent Poisson random variables

A useful property of the Poisson distribution is that if X and Y are random variables such that

- X and Y are independent;
- $X \sim \text{Pois}(\mu)$;
- $Y \sim \text{Pois}(\lambda)$;

then $S = X + Y$ has the Poisson distribution with parameter $\mu + \lambda$.

Example: (Exercise 4.5.8) In the first hour that a bank opens, the customers who enter are of three kinds: those who only require teller service, those who only want to use the ATM, and those who only require special services (neither the tellers nor the ATM). Assume that the numbers of customers of the three kinds are independent of each other and also that:

- $X =$ the number that only require teller service has the Poisson (6) distribution,
- $Y =$ the number that only want to use the ATM has the Poisson (2) distribution, and
- $Z =$ the number that only require special services has the Poisson (1) distribution.

Suppose you observe the bank in the first hour that it opens. In each part below, find the chance of the event described.

(a) 12 customers enter the bank

(b) more than 12 customers enter the bank

(c) customers do enter but none requires special services

(a) W = # customers enter the bank

$= X+Y+Z \sim \text{Pois}(9)$, $\quad 9 = 6+2+1$

$$P(W=12) = \frac{e^{-9}\cdot 9^{12}}{12!}$$

(b) $P(W>12) = 1 - P(W \le 12)$

$= 1 - F(12)$

(c) $P(Z=0 \text{ and } W>0) = P(Z=0,\ X+Y>0)$, $\quad W = X+Y+Z$, $Z = 0$

* Z and W are not independent, but Z and $X+Y$ are independent!

$= P(Z=0)\,P(X+Y>0)$ $\quad$ ∵ Indep.

$= e^{-1} * (1 - P(X+Y=0))$ $\quad$ $X+Y \sim \text{pois}(8)$

$= e^{-1} * (1 - e^{-8})$

$P(X+Y=0) = e^{-8}$ $\quad$ ($\mu = 8, k = 0$)

$X \sim \text{Pois}(\mu)$

$$P(X=k) = \frac{e^{-\mu}\cdot \mu^k}{k!}$$