

MATH 180A: INTRO TO PROBABILITY (FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: § 3.3

Next: § 3.4

HW3 due **TONIGHT!**

Lab 3.2 due Monday.

* Regrade requests for HW2: window

↳ separate request for each problem

↳ detailed, polite responses please.

Tuesday, 10/22

8am - 11pm

* **Midterm 1**: Wednesday, 10/23, 8-10pm, PCYNH 109

Assigned seats (see Triton Ed)

1 double sided sheet of (hand-written) notes allowed

Practice Midterms posted on webpage

Poisson Distribution

A random variable X has the Poisson(λ) distribution if

$$\lim_{n \rightarrow \infty} P(S_{n, \frac{\lambda}{n}} = k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{= e^{\lambda}} = 1 \quad \checkmark$$

E.g. A **100 year storm** is a storm magnitude expected to occur in any given year with probability $1/100$.

Over the course of a century, how likely is it to see at least 4 100 year storms?

$$P(S_{100, \frac{1}{100}} \geq 4) = \sum_{k=4}^{100} P(S_{100, \frac{1}{100}} = k) \approx \sum_{k=4}^{100} e^{-1} \frac{(1)^k}{k!}$$
$$\sum_{k=4}^{100} \binom{100}{k} \left(\frac{1}{100}\right)^k \left(1 - \frac{1}{100}\right)^{100-k} \approx 1.8374\% = 0.018374$$
$$\approx \sum_{k=4}^{\infty} e^{-1} \frac{1}{k!} = 1 - \sum_{k=0}^3 e^{-1} \frac{1}{k!} = 1.8288\%$$

Summary

Sampling independent trials, the most important (discrete) probability distributions are:

- $\text{Ber}(p)$: $P(X=1)=p, P(X=0)=1-p$ $0 \leq p \leq 1$
(single trial with success probability p)
- $\text{Bin}(n,p)$: $P(S_n=k) = \binom{n}{k} p^k (1-p)^{n-k}$ $0 \leq k \leq n$
(number of successes in n independent trials with rate p)
- $\text{Geom}(p)$ $P(N=k) = (1-p)^{k-1} p$ $k=0,1,2,\dots$
(first successful trial in repeated independent trials with rate p)
- $\text{Poisson}(\lambda)$ $P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $k=0,1,2,\dots$ $\lambda > 0$.
(Approximates $\text{Bin}(n, \lambda/n)$; number of rare events in many trials)

Expectation

3.3

Toss a fair coin 1000 times, and record the sequence of outcomes.

110100011011001011...

Average them. $\frac{1}{1000}(1+1+0+1+0+0+0+1+1+0+1+1 \dots)$

What size do you **expect** this number to have?

[About ~~$\frac{1}{2}$~~ of the outcomes are 1, about ~~$\frac{1}{2}$~~ are 0. \therefore Average $\approx \frac{1}{2}(1) + \frac{1}{2}(0) = p(1) + (1-p) \cdot 0$

What if the coin is biased $P(X_j=1)=p, P(X_j=0)=1-p$.

Definition: Let X be a discrete random variable with possible values t_1, t_2, t_3, \dots . The **expectation** or **expected value** of X is

$$E(X) := \sum_j t_j P(X=t_j)$$

weighted average
"balance point"

Question: Is the expectation $\mathbb{E}(X)$ the value X is equal to most often?

(a) Yes, always.

(b) No, not generally.

Eg. Let X be the number rolled on a fair die. $X \in \{1, 2, 3, 4, 5, 6\}$

$$\mathbb{E}(X) = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1}{6} (1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}.$$

Eg. Let Y be $\text{Ber}(p)$. $\mathbb{E}(Y) = p \cdot 1 + (1-p) \cdot 0 = p$

Eg. You toss a biased coin (Y) repeatedly until the first heads. How long do you expect it to take?

$N =$ the time the 1st heads comes up. $N \sim \text{Geom}(p)$

$$\mathbb{E}(N) = \sum_{k=1}^{\infty} k \cdot \underbrace{P(N=k)}_{(1-p)^{k-1} p} = p \sum_{k=1}^{\infty} k (1-p)^{k-1}$$

$= p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p} \frac{d}{dx} \sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$

$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$

E.g. $S_n \sim \text{Bin}(n, p)$ ($S_n = X_1 + X_2 + \dots + X_n$ for X_j independent $\text{Ber}(p)$)

$$\mathbb{E}(S_n) = \sum_{k=0}^n k \cdot \mathbb{P}(S_n = k) = \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np$$

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

$$\begin{aligned} & \mathbb{E}(X_1 + \dots + X_n) \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n) \end{aligned}$$

$$\mathbb{E}(X_j) = p$$

E.g. $X \sim \text{Poisson}(\lambda)$

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

↳ E.g. A factory has, on average, 3 accidents per month. Estimate the probability that there will be exactly 2 accidents this month.

$$\left. \begin{aligned} X &= \# \text{accidents/month} & \mathbb{E}(X) &= 3 = \lambda \\ X &\sim \text{Poisson}(3) \end{aligned} \right\} \mathbb{P}(X=2) = e^{-3} \frac{3^2}{2!} = 22.4\%$$

E.g. Toss a fair coin until tails comes up. If this is on the first toss, you win \$2 and stop. If heads comes up, the pot doubles, and you continue. That is, if the first tails is on the k^{th} toss, you win 2^k dollars.

What is your expected winnings?

$W = \{ 2^k \text{ if the first tails is on the } k^{\text{th}} \text{ toss} \}$

$$E(W) = \sum_{k=1}^{\infty} 2^k \cdot \underbrace{P(W=2^k)}_{\left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} = \frac{1}{2^k}} = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \infty.$$