Lecture 4

Hypergeometric Distribution

Lets return to a situation we studied earlier: a randomly shuffled deck of cards.

- 4 aces, 48 other cards.

- Suppose we draw (a sample) four cards, without replacement (i.e. we don't put the first card back before we draw the second, third, fourth card).

- Let X = the number of aces in the sample.

$$P(X=0) = P(\text{none of the 4 aces}) = \frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50}\cdot\frac{45}{49}$$

$$P(X=1) = \underbrace{\frac{4}{52}\,\frac{48}{51}\cdot\frac{47}{50}\cdot\frac{46}{49}}_{\text{first one was ace}} + \underbrace{\frac{48}{52}\cdot\frac{4}{51}\cdot\frac{47}{50}\cdot\frac{46}{49}}_{\text{second card was ace.}}$$

$$+\underbrace{\frac{48}{52}\cdot\frac{47}{51}\cdot\frac{4}{50}\cdot\frac{46}{49}}_{\text{third one is an ace}} + \underbrace{\frac{48\cdot 47\cdot 46}{52\cdot 51\cdot 50}\cdot\frac{4}{49}}_{\text{4th one is an ace.}}$$

$$P(X=2) = \underbrace{\frac{4}{52}\cdot\frac{3}{51}\cdot\frac{48}{50}\cdot\frac{47}{49}}_{\text{first two are aces}} + \underbrace{\frac{4}{52}\cdot\frac{48}{51}\cdot\frac{3}{50}\cdot\frac{47}{49}}_{\text{1st and 3rd.}} +$$

$$+\frac{4}{52}\cdot\frac{48}{51}\cdot\frac{47}{50}\cdot\frac{3}{49} + \frac{48}{52}\cdot\frac{4}{51}\cdot\frac{3}{50}\cdot\frac{47}{49}$$

$$+\frac{48}{52}\cdot\frac{4}{51}\cdot\frac{47}{50}\cdot\frac{3}{49} + \frac{48}{52}\cdot\frac{47}{51}\cdot\frac{4}{50}\cdot\frac{3}{49}.$$

$$P(X=3) = \ldots$$

$$P(X=4) = \frac{4}{52}\cdot\frac{3}{51}\cdot\frac{2}{50}\cdot\frac{1}{49}$$

This gives an example of a <u>hypergeometric</u> distribution.

<u>Defn</u>: (Hypergeometric Distribution)

Given a set of N objects, with K objects considered successful and $N-K$ failures, our experiment consists of choosing n objects (out of N) <u>without replacement.</u> ($n \le N$, $K \le N$).

Then, if X = "the number of successes in the sample",

we say X is hypergeometric.

We have a p.m.f given by

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n+K-N\} \text{ to } \min\{K, n\}$$

If X is hypergeometric with parameters N, K, and n, then

$$E(X) = np \qquad V(X) = np(1-p)\left(\frac{N-n}{N-1}\right).$$

where $p = K/N$ (the probability of success).

Note that if n is small relative to N, $\left(\frac{N-n}{N-1}\right)$ is close to 1 and X is similar to the Binomial random variable:

Poisson Distribution

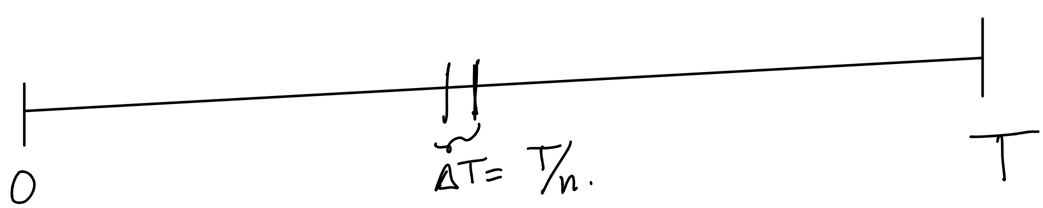
- The Poisson distribution describes events which occur randomly in a given interval.

Ex:

- # flaws in a length of wire.
- # of earthquakes in 5 years.
- # of buses at a stop in a week, etc.

P.m.f: $\rightsquigarrow$ need to know average # of events per unit length.

Fix an interval of length T



- divide T into n pieces, $\Delta T = T/n$.
- by choosing n sufficiently large, we may assume that the probability of > 1 event in ΔT is very small (tends to 0 as $n \to \infty$).

- Let λ = average # of flaws per unit length.

- $p = \lambda \Delta T = \frac{\lambda T}{n}$ is the probability of an event in a lenght ΔT.

- Since there are n intervals, and each interval either has an event or does not
  - ↳ i.e. each ΔT is a Bernoulli trial with probability T.
  - ↳ So $X \sim Bin(n, p)$ for large enough n.

- If we refine our interval by taking $n \to \infty$ we get

$$f(x) = \lim_{n \to \infty} \underbrace{\binom{n}{x} p^x (1-p)^{n-x}}_{\text{Binomial pmf.}}$$

$$= \lim_{n \to \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda T}{n}\right)^x \left(1 - \frac{\lambda T}{n}\right)^{n-x}$$

$$= \frac{(\lambda T)^x}{x!} e^{-\lambda T}.$$

So pmf for the Poisson distribution is

$$f(x) = \frac{(\lambda T)^x}{x!} e^{-\lambda T}$$

Since the Poisson random variable is approximated by the Binomial random variable we have:

$$E(X) = np \xrightarrow[n\to\infty]{} \cancel{n}\cdot\left(\frac{\lambda T}{\cancel{n}}\right) = \lambda T$$

$$V(X) = np(1-p) \xrightarrow[n\to\infty]{} \cancel{n}\left(\frac{\lambda T}{\cancel{n}}\right)\left(1-\cancelto{0}{\frac{\lambda T}{n}}\right)$$

$$= \lambda T.$$

So, to specify a Poisson random variable, we specify the average/unit, λ,

---

<u>Ex:</u> A real estate agency sells on average $\underbrace{\text{2 houses/day}}_{\lambda}$. What is the probability that they sell 10 houses $\underbrace{\text{next week}}_{T.}$?

$T = 7$ days. $\lambda = 2$ houses/day.

Let $X = \#$ of houses sold per week.

Then $E(X) = \lambda T = 2\cdot 7 = 14$ houses/week.

We want $P(X) = 10$.

i.e. $f(10)$ where $f(x) = \dfrac{e^{-14}(14)^x}{x!}$

$$f(10) = \frac{e^{-14}(14)^{10}}{10!} \approx 0.066$$

i.e. 6.6%.