Lecture 5 Jan 30, 2025

Poissonization

Binomial / Multinomial $\qquad\qquad$ Poisson

$$X \sim \text{Bin}(n,p) \longrightarrow Y \sim \text{Poi}(np)$$

$$\text{answer} \longleftarrow \text{answer}$$

solve

nicer world:
independence

$$E[Y] = \text{Var}[Y] = np$$

$$\left.\begin{array}{l} Y_1 \sim \text{Poi}(\lambda_1) \\ Y_2 \sim \text{Poi}(\lambda_2) \end{array}\right\} \quad Y_1 + Y_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$$

$$\text{Binomial}(n, p) \underset{}{\overset{np = \lambda}{\approx}} \text{Poisson}(\lambda)$$

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{e^{-\lambda} \lambda^k}{k!}$$

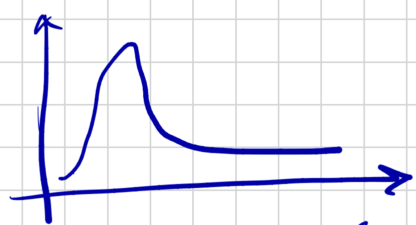
## why does this even make sense?

Binomials are usually pretty normal,

unless you push them to the limit

— then they turn into a Poisson

(and make quite the splash ...  )



small $np$

## Example  balls and bins

$m$ balls are thrown into $n$ bins uniformly at random.

$X_i :=$ # balls in bin $i$

$X_i \sim \mathrm{Bin}(m, 1/n)$

$X_i$'s are not independent.

($X_1 = m$ implies all other $B_i$'s are zero.)

In Poisson world:

$Y_i \sim \mathrm{Poi}(m/n)$

↳ all independent

Go and solve your favorite problem in this new world ....

Binomial $(n, p) \approx$ Poisson $(n, p)$

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad \frac{e^{-np} (np)^k}{k!}$$

$X \sim \mathrm{Bin}(n, p)$

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

$$= \frac{np \cdot (n-\cancel{1})p \cdot (n-\cancel{2})p \cdots (n-\cancel{k+1})p}{k!} \left(1 - \frac{np}{n}\right)^{n-\cancel{k}}$$

$$\simeq \frac{(np)^k e^{-np}}{k!}$$

$n$ is very large and $p$ is small

**Theorem:** Let $X_n \sim \mathrm{Bin}(n, p)$ where $p$ is a function of $n$ and $\lim_{n \to \infty} np = \lambda$ is a constant that is independent of $n$. Then, for any fixed $k$,

$$\lim_{n \to \infty} \Pr[X_n = k] = \frac{e^{-\lambda} \lambda^k}{k!}$$

# Poissons with fixed sum

Suppose we have two independent r.v.

$Y_1 \sim Poi(\lambda_1)$

$Y_2 \sim Poi(\lambda_2)$

claim 1

first obsereve that $Y := Y_1 + Y_2$ is a random variable from $Poi(\lambda_1 + \lambda_2)$.

proof

for any integer $t \geq 0$:

$$Pr[Y = t] = \sum_{r=0}^{t} Pr[Y = t \mid Y_1 = r] . Pr[Y_1 = r]$$

$$= \sum Pr[Y_2 = t - r \mid Y_1 = r] . Pr[Y_1 = r]$$

using independence of $Y_1$ and $Y_2$

$$= \sum_{r=0}^{t} Pr[Y_2 = t - r] . Pr[Y_1 = r]$$

$$= \sum_{r=0}^{t} \frac{e^{-\lambda_2}\lambda_2^{t-r}}{(t-r)!} \cdot \frac{e^{-\lambda_1}\lambda_1^{r}}{r!}$$

$$= \sum_{r=0}^{t} \frac{1}{(t-r)!\,r!} \cdot \overset{1 \,=}{\frac{t!}{t!}} \cdot e^{-(\lambda_1+\lambda_2)} \cdot \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{t-r} \cdot \left(\frac{\lambda_1}{(\lambda_1+\lambda_2)}\right)^{r} \overset{1 \,=}{\cdot \,(\lambda_1+\lambda_2)^{t}}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1+\lambda_2)^{t}}{t!} \underbrace{\sum_{r=0}^{t} \binom{t}{r} \cdot \left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{t-r} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{r}}$$

$$= \left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}+\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{t} = 1^{t} = 1$$

via binomial expansion $(x+y)^t = \sum_{r=0}^{t} \binom{t}{r} x^r y^{t-r}$

$$= \frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1+\lambda_2)^{t}}{t!} \implies Y \sim \mathrm{Poi}(\lambda_1+\lambda_2) \quad \square$$

**Claim 2** condition on the sum of $Y_1 + Y_2$

These r.v. are coming from a Binomial distribution.

Assume $Y_1 + Y_2 = m$ for a fixed $m$,

then
$$\begin{cases} Y_1 \sim Bin\left(m, \dfrac{\lambda_1}{\lambda_1+\lambda_2}\right) \\[2ex] Y_2 \sim Bin\left(m, \dfrac{\lambda_1}{\lambda_1+\lambda_2}\right) \end{cases}$$

**Proof** for any integer $0 \le t \le m$

$$Pr[\,Y_1 = t \mid Y_1 + Y_2 = m\,]$$

$$= \frac{Pr[\,Y_1 = t \text{ and } Y_1 + Y_2 = m\,]}{Pr[\,Y_1 + Y_2 = m\,]}$$

$$= \frac{Pr[\,Y_1 = t \text{ and } Y_2 = m-t\,]}{Pr[\,Y = m\,]}$$

$Y \sim Poi(\lambda_1+\lambda_2)$

via claim 1

$$= \frac{\Pr[Y_1 = t] \cdot \Pr[Y_2 = m - t]}{\Pr[Y = m]}$$

via independence of $Y_1$ and $Y_2$

$Y \sim poi(\lambda_1 + \lambda_2)$

$$= \frac{\cancel{e^{-\lambda_1}} \cdot \lambda_1^{t}}{t!} \cdot \frac{\cancel{e^{-\lambda_2}} \lambda_2^{m-t}}{(m-t)!} \cdot \frac{m!}{\cancel{e^{-(\lambda_1+\lambda_2)}} (\lambda_1+\lambda_2)^m}$$

$$= \binom{m}{t} \cdot \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{t} \cdot \left(1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{m-t}$$

$$= \Pr[X = t]$$

$X \sim Bin\left(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

$\Rightarrow$ $Y_1$ condition of $Y_1 + Y_2$ acts like a binomial. $\square$

How about the joint distribution?

binomial world

$\forall i \in [n] \quad X_i^{(k)} \sim Bin(k, 1/m)$

Poisson world

$Y_i^{(m)} \sim Poi(m/n)$

**Theorem** The distribution $(Y_1^{(m)}, \ldots, Y_n^{(m)})$ conditioned on $\sum_i Y_i^{(m)} = k$ is the same as $(X_1^{(k)}, \ldots, X_n^{(k)})$.

**Proof** Suppose we have a tuple $(k_1, \ldots, k_n)$ such that $\sum_{i=1}^{n} k_i = k$

Binomial world:

$$\Pr[X_1^{(k)} = k_1, \ldots, X_n^{(k)} = k_n] = \frac{\binom{k}{k_1; k_2; \ldots k_n}}{n^k} = \frac{k!}{\prod k_i! \; n^k}$$

Poisson world:

$$\Pr[Y_1^{(m)} = k_1, \ldots, Y_n^{(m)} = k_n \mid \sum_i Y_i^{(m)} = k]$$

$$= \frac{P\left[Y_1^{(m)} = k_1, \ldots, Y_n^{(m)} = k_n\right]}{\Pr\left[\sum Y_i^{(m)} = k\right]}$$

$$= \frac{\prod_{i=1}^{n} \Pr\left[Y_i^{(m)} = k_i\right]}{\Pr\left[\sum Y_i^{(m)} = k\right]} = \left(\prod_{i=1}^{n} \frac{\cancel{e^{-m/n}} \left(m/n\right)^{k_i}}{k_i!}\right) \cdot \frac{k!}{\cancel{e^{-m}}\, m^k}$$

$$= \frac{k!}{\left(\prod_i k_i!\right) \cdot n^k} = \Pr\left[X_1^{(k)} = k_1, \ldots, X_n^{(k)} = k_n\right]$$

$\square$

How to go from binomial world to poisson world ?

Example sampling from a discrete distribution

$s_1, s_2, \dots, s_m \sim P$ over $[n]$

$\forall\ i \subseteq [n]$ $X_i :=$ Frequency of $i$ among samples

$= \sum_{j=1}^{m} \mathbb{1}\{s_j = i\}$

not independent

$E[X_i] = mp_i$ $Y_i = Poi(mp_i)$

Multinomial world:

Fix $m$

Draw $s_1, \dots, s_m$

$X_i \sim \mathrm{Bin}(m, p_i)$

Poisson world

Fix $m$

$\hat{m} \leftarrow \mathrm{Poi}(m)$

Draw $s_1, \dots, s_{\hat{m}}$

$Y_i =$ frequency of $i$

$\boxed{= \mathrm{Poi}(m p_i)}$

Theorem $Y_i \sim \mathrm{Poi}(m p_i)$

$$\Pr[Y_i = k] = \sum_{t=k}^{\infty} \Pr[\hat{m} = t] \Pr[Y_i = k \mid \hat{m} = t]$$

$$= \sum_t \frac{e^{-m} m^t}{t!} \cdot \binom{t}{k} (1-p_i)^{t-k} p_i^{\,t}$$

$$= \frac{e^{-m} (p_i m)^k}{k!} \underbrace{\sum_t \frac{(p_i m)^{t-k} (1-p_i)^{t-k}}{(t-k)!}}$$

$$\sum_{t'} \frac{(m\,p_i(1-p_i))^{t'}}{t'!} = e^{m(1-p_i)p_i}$$

$$= \frac{e^{-mp_i}(mp_i)^k}{k!} = \Pr[Y = k] \qquad \square$$

$Y \sim Poi(mp_i)$

* Since the probability distribution of $Y_i$ does not involve any $Y_j$ $(j \neq i)$, one can imply that $Y_i$'s are independent.

How to translate back?

Example: we focus on balls and bins setting and mention few theorems without proofs.

setup:

Suppose we have thrown $m$ balls into $n$ bins uniformly at random.

$X_i$ = # balls in bin $i$ is a random variable drawn from $Bin(m, 1/n)$

we approximate $X_i$ with $Y_i$

where $Y_i$ is drawn from $Poi(m/n)$

notice the means are identical.

Theorem 1

Let $f(x_1, \ldots, x_n)$ be a non-negative function. Then for the balls and bins setup stated above:

$$E\left[f(X_1, \ldots, X_n)\right] \leq e\sqrt{m}\; E\left[f(Y_1, \ldots, Y_n)\right]$$

Corollary if an event $A$ happens with probability $p$ in the poisson setup, then $A$ happens with probability at most $p\,e\sqrt{m}$ in the binomial setup.

Proof. set $f(x_1, \ldots, x_n) = 1$

if $A$ considered as "occured" when we have $x_i$ balls in bin $i$. and set $f(x_1, \ldots, x_n) = 0$ otherwise

Clearly, we have

$$E[f(X_1, \dots, X_n)] = \Pr[A \text{ in binomial world}]$$

$$E[f(Y_1, \dots Y_n)] = \Pr[A \text{ in Piosson world}]$$

Applying theorem 1 implies the statement □

**Theorem 2** Let $f(x_1, \dots, x_m)$ be a non-negative function s.t. $E[f(X_1, \dots, X_m)]$ in either monotonically increasing or monotonically decreasing in $m$.

Then

$$E[f(X_1, \dots, X_n)] \le 2 \cdot E[f(Y_1, \dots, Y_n)]$$

**Corollary:** Let $A$ be an event whose probability is either monotonically increasing or monotonically decreasing. in the number of balls. If $A$ has probability $\le p$ is in the poissonized world. then $A$ has probability $\le 2p$ in the Binomial world.