S&DS 241 Lecture 9

Poisson distribution

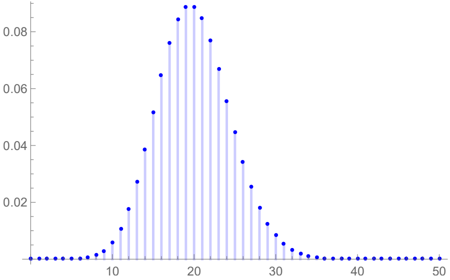
B-H: 4.7,4.8, math appendix A.8,A.9

Poisson distribution

A random variable X is said to have a Poisson distribution with parameter $\lambda \geq 0$, denoted by $X \sim \text{Pois}(\lambda)$, if[1]

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

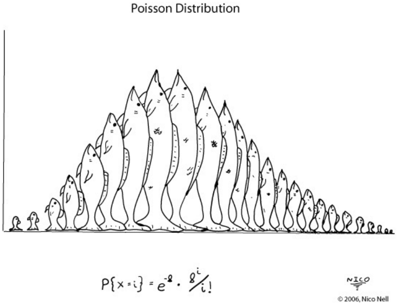
i.e., $P(X = 0) = e^{-\lambda}$, $P(X = 1) = \lambda e^{-\lambda}$, etc.

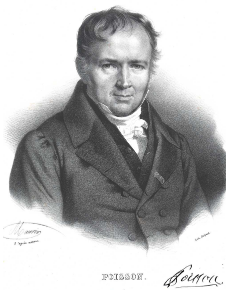


Pois(20)

[1] As a convention $0! = 1$.

Poisson distribution





Siméon Denis Poisson

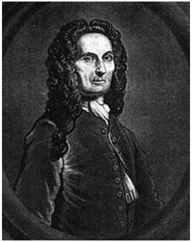
Application of Poisson distribution

Poisson distribution arises in studying many rare events

- Number of photons arriving at a detector in a fixed interval of time.
- Telephone traffic to base station / arrival of spam emails
- Modeling insurance claims
- Typos in a large document

Historical examples

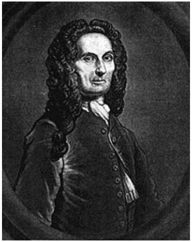
- Abraham de Moivre (1711) and Simon Denis Poisson (1837): Theorized about the number of wrongful convictions in a given country during a time interval of given length



[2] For dataset see Table 5.5 of Grinstead-Snell.

Historical examples

- Abraham de Moivre (1711) and Simon Denis Poisson (1837): Theorized about the number of wrongful convictions in a given country during a time interval of given length



- Ladislaus von Bortkiewicz (1898): Investigated the number of soldiers in the Prussian army killed accidentally by horse kicks in 1875-1894.[2]

[2] For dataset see Table 5.5 of Grinstead-Snell.

Relations between Poisson and Binomial distributions

- Poisson distribution is a good approximation of Bin(n, p) when n is large and p is small.
- Specifically,
  - when n is large (many independent trials)
  - success probability p is inversely proportional to n

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
    - in each subinterval there is either one photon or none

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
  - in each subinterval there is either one photon or none
  - arrivals are independent

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
  - in each subinterval there is either one photon or none
  - arrivals are independent
  - P(photon arrival) $\propto$ duration of interval, say,

$$p = \frac{\alpha T}{n}, \quad \alpha = \text{“arrival rate”}$$

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
  - in each subinterval there is either one photon or none
  - arrivals are independent
  - P(photon arrival) $\propto$ duration of interval, say,

$$p = \frac{\alpha T}{n}, \quad \alpha = \text{"arrival rate"}$$

- Then $X \sim \text{Bin}(n, \frac{\alpha T}{n})$

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
  - in each subinterval there is either one photon or none
  - arrivals are independent
  - P(photon arrival) $\propto$ duration of interval, say,

$$p = \frac{\alpha T}{n}, \quad \alpha = \text{“arrival rate”}$$

- Then $X \sim \text{Bin}(n, \frac{\alpha T}{n})$

Specific example

Let X be the number of photons arriving at a detector in a time window of length T.

- Partition the time interval into n small subintervals of length T/n

- Assume that
  - in each subinterval there is either one photon or none
  - arrivals are independent
  - P(photon arrival) $\propto$ duration of interval, say,

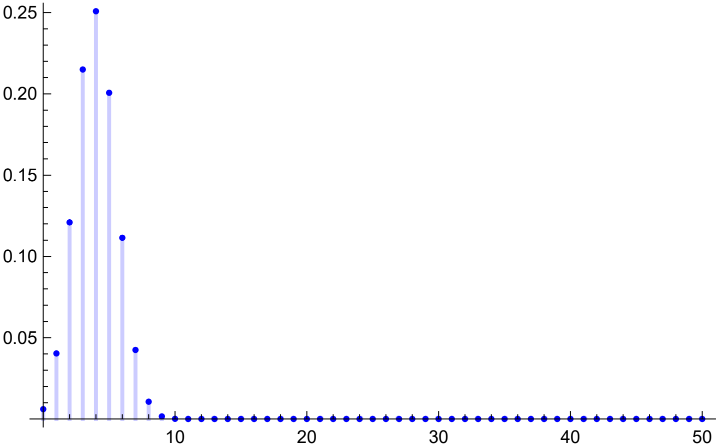
$$p = \frac{\alpha T}{n}, \quad \alpha = \text{“arrival rate”}$$

- Then $X \sim \mathrm{Bin}(n, \frac{\alpha T}{n})$

Question

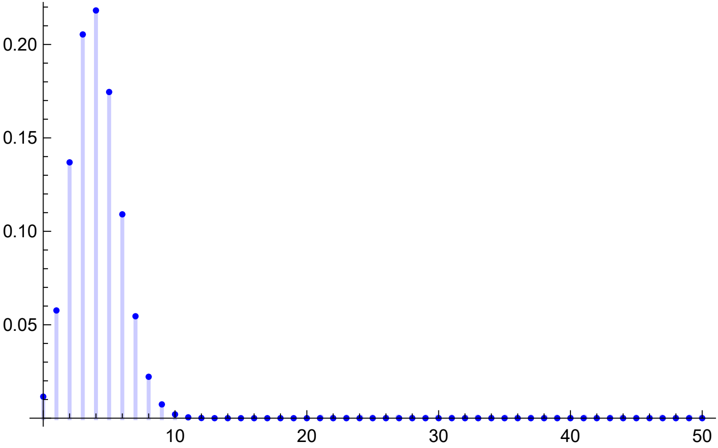
When n is large (partition by milliseconds, nanoseconds, etc), how does X behave?

Poisson approximation of Binomial



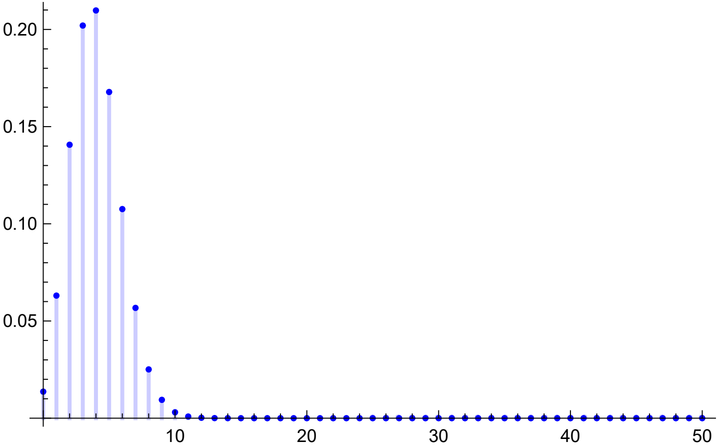
Bin$(10, 4/10)$

Poisson approximation of Binomial



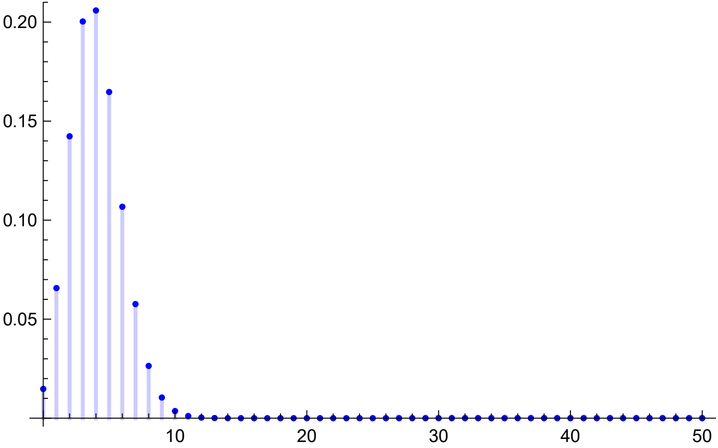
$\text{Bin}(20, 4/20)$

Poisson approximation of Binomial



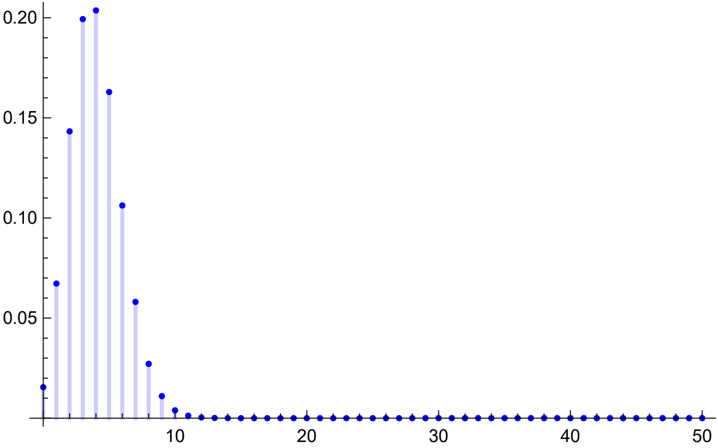
$\text{Bin}(30, 4/30)$

Poisson approximation of Binomial



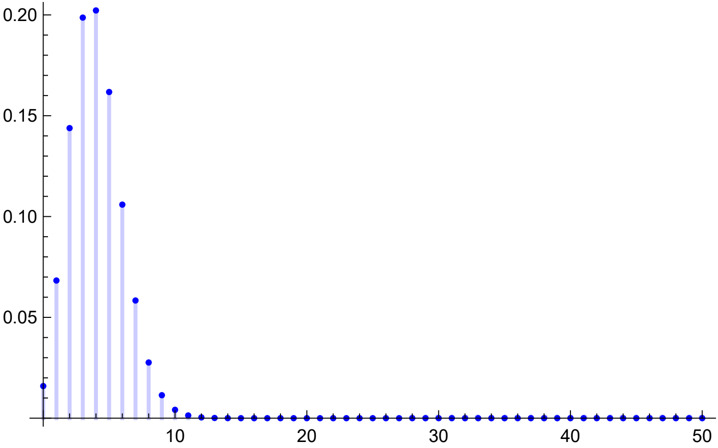
$\text{Bin}(40, 4/40)$

Poisson approximation of Binomial



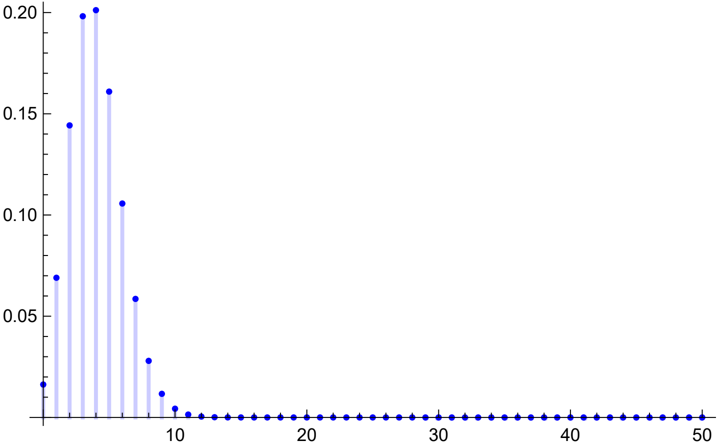
$\mathrm{Bin}(50, 4/50)$

Poisson approximation of Binomial



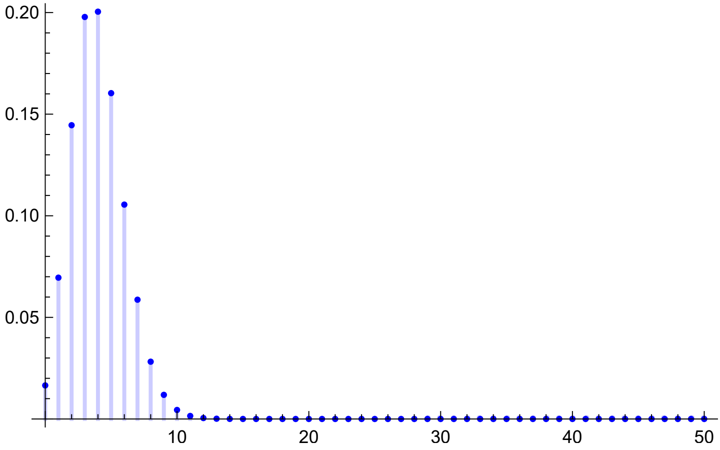
$\text{Bin}(60, 4/60)$

Poisson approximation of Binomial



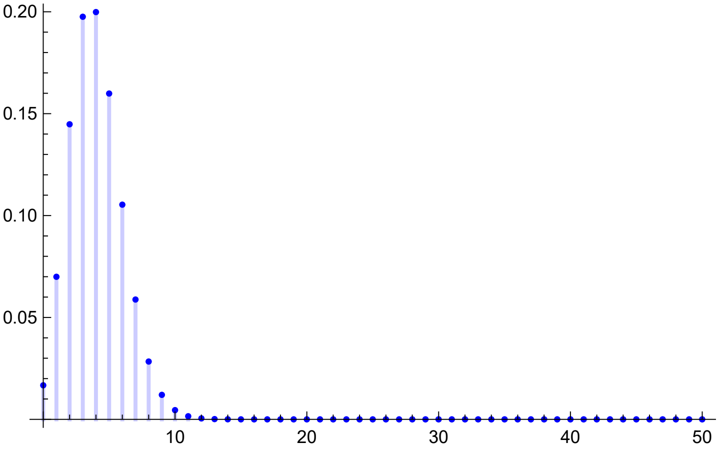
$\text{Bin}(70, 4/70)$

Poisson approximation of Binomial



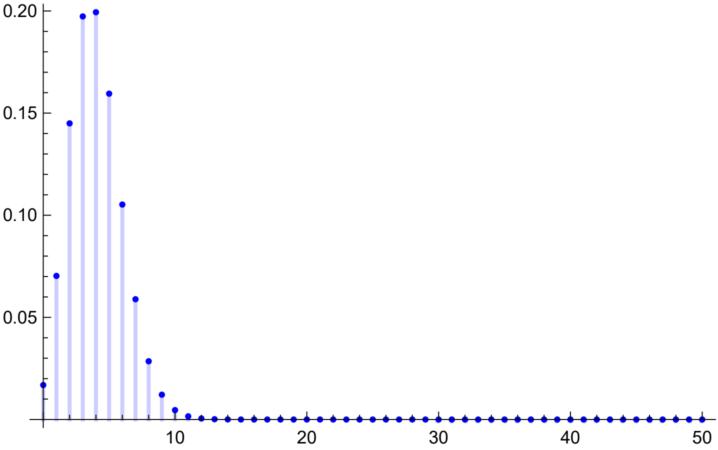
Bin$(80, 4/80)$

Poisson approximation of Binomial



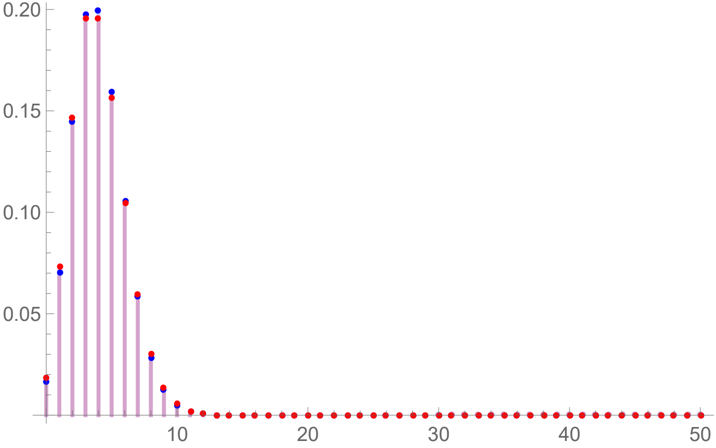
$\text{Bin}(90, 4/90)$

Poisson approximation of Binomial



Bin$(100, 4/100)$

Poisson approximation of Binomial



Bin$(100, 4/100)$ vs Pois(4)

Poisson approximation of binomial distribution

- Precise statement:

$$\mathsf{Bin}\,(n, \lambda/n) \to \mathsf{Pois}(\lambda), \quad \text{as } n \to \infty$$

in the sense of convergence of PMF: for any fixed k,

$$P\left(\mathsf{Bin}\,(n, \lambda/n) = k\right) \xrightarrow{n\to\infty} P(\mathsf{Pois}(\lambda) = k)$$

Poisson approximation of binomial distribution

- Precise statement:

$$\mathsf{Bin}\,(n, \lambda/n) \to \mathsf{Pois}(\lambda), \quad \text{as } n \to \infty$$

in the sense of convergence of PMF: for any fixed k,

$$P\left(\mathsf{Bin}\,(n, \lambda/n) = k\right) \xrightarrow{n\to\infty} P(\mathsf{Pois}(\lambda) = k)$$

- Example ($k = 0$):

$$P\left(\mathsf{Bin}\,(n, \lambda/n) = 0\right) = \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda} = P(\mathsf{Pois}(\lambda) = 0)$$

Fact from calculus [B-H A.2.5]: $\lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x$

Poisson approximation of binomial distribution

In general: for fixed $k \geq 0$,

$$
\begin{aligned}
& P\left(\mathsf{Bin}\left(n, \lambda/n\right) = k\right) \\
= \quad & \binom{n}{k} \left(\frac{\lambda}{n}\right)^{k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
= \quad & \frac{n\left(n-1\right) \cdot \cdots \cdot \left(n-k+1\right)}{k!} \left(\frac{\lambda}{n}\right)^{k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
= \quad & \frac{\lambda^k}{k!} \cdot 1 \cdot \underbrace{\left(1 - \frac{1}{n}\right)}_{\to 1} \cdot \cdots \cdot \underbrace{\left(1 - \frac{k-1}{n}\right)}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n}}_{\to e^{-\lambda}} \\
\to \quad & \frac{e^{-\lambda}\lambda^k}{k!}
\end{aligned}
$$

Properties of Poisson

Verify normalization of Poisson PMF

- Recall: Taylor expansion

$$
\begin{aligned}
e^{\lambda} &= 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots + \frac{\lambda^k}{k!} + \cdots \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}
\end{aligned}
$$

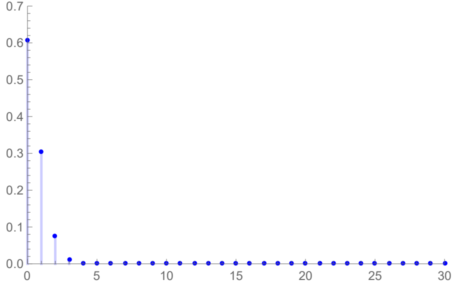
- Hence

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

Expectation

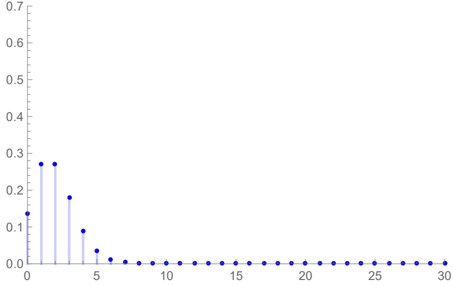
- $X \sim \text{Pois}(\lambda)$. Then
$$E(X) = \lambda,$$
<u>Intuition</u>: the expectation of of $\text{Bin}(n, \frac{\lambda}{n})$ is λ, for any n.
- Direct verification:
$$\begin{aligned}
E(X) =& \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\
=& \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} \\
=& \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^{j+1}}{j!} \\
=& \lambda \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} = \lambda
\end{aligned}$$

PMF vs expectation



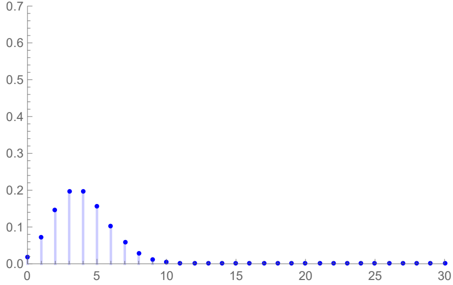
Pois(0.5)

PMF vs expectation



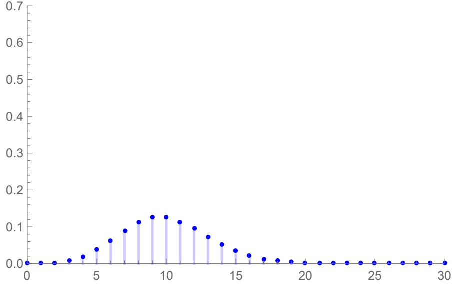
Pois(2)

PMF vs expectation



Pois(4)

PMF vs expectation



Pois(10)

Sum of independent Poissons

Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. Then

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

Sum of independent Poissons

Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. Then

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

- This property is also inherited from Binomials: when $p \to 0$,

$$\text{Bin}\Big(\frac{\lambda}{p}, p\Big) \to \text{Pois}(\lambda)$$
$$\text{Bin}\Big(\frac{\mu}{p}, p\Big) \to \text{Pois}(\mu)$$

and from last lecture we know that

$$\underbrace{\text{Bin}\Big(\frac{\lambda}{p}, p\Big) + \text{Bin}\Big(\frac{\mu}{p}, p\Big)}_{\text{independent}} = \text{Bin}\Big(\frac{\lambda + \mu}{p}, p\Big) \to \text{Pois}(\lambda + \mu)$$

Direct verification

Let $X \sim$ Pois(λ) and $Y \sim$ Pois(μ) be independent.

$$
\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X=j)\, P(Y=k-j) \\
&= \sum_{j=0}^{k} \frac{e^{-\lambda}\lambda^j}{j!} \frac{e^{-\mu}\mu^{k-j}}{(k-j)!} \\
&= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} k! \frac{\lambda^j}{j!} \frac{\mu^{k-j}}{(k-j)!} \\
&= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda^j \mu^{k-j} \\
&= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda+\mu)^k \qquad \text{binomial expansion (Lec 8)}
\end{aligned}
$$

Conditioned on the sum

Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What's the distribution of X conditioned on $X + Y$?

Conditioned on the sum

Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What's the distribution of X conditioned on $X+Y$?

$$\begin{aligned} & P(X=j|X+Y=k) \\ = & \frac{P(X=j, X+Y=k)}{P(X+Y=k)} = \frac{P(X=j)P(Y=k-j)}{P(X+Y=k)} \\ = & \frac{\frac{e^{-\lambda}\lambda^j}{j!}\frac{e^{-\mu}\mu^{k-j}}{(k-j)!}}{\frac{e^{-(\lambda+\mu)}}{k!}(\lambda+\mu)^k} = \binom{k}{j}\left(\frac{\lambda}{\lambda+\mu}\right)^j\left(\frac{\mu}{\lambda+\mu}\right)^{k-j} \end{aligned}$$

Conditioned on the sum

Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What's the distribution of X conditioned on $X+Y$?

$$
\begin{aligned}
&P(X=j|X+Y=k) \\
&= \frac{P(X=j, X+Y=k)}{P(X+Y=k)} = \frac{P(X=j)P(Y=k-j)}{P(X+Y=k)} \\
&= \frac{\frac{e^{-\lambda}\lambda^j}{j!}\frac{e^{-\mu}\mu^{k-j}}{(k-j)!}}{\frac{e^{-(\lambda+\mu)}}{k!}(\lambda+\mu)^k} = \binom{k}{j}\left(\frac{\lambda}{\lambda+\mu}\right)^j\left(\frac{\mu}{\lambda+\mu}\right)^{k-j}
\end{aligned}
$$

Conditioned on the sum $X+Y=k$, X is distributed as $\text{Bin}(k, \frac{\lambda}{\lambda+\mu})$.

Poisson approximation of binomial distribution

- Replace $\mathsf{Bin}(n, p)$ by $\mathsf{Pois}(np)$.
- We know this is accurate when $n \to \infty$ and np converges to a constant.

Poisson approximation

Poisson paradigm

Let $A_1, \ldots, A_n$ be a collection of events (e.g. indexed by time or space)

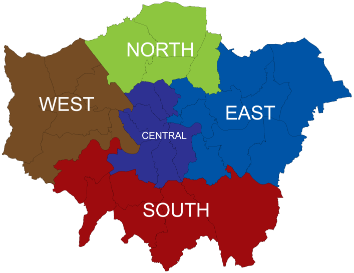
- n is large
- $p_j = P(A_j)$ small
- A_j are (approximately) independent

Let $X = \sum_{j=1}^{n} \mathbf{1}_{A_j}$ count how many of the events A_j's occur. Then X is approximately Poisson distributed as $\text{Pois}(\sum_{j=1}^{n} p_j)$

Example

Consider the statistics of flying bomb hits in the south of London during World War II.

- The entire area is divided into a grid of $N = 576$ small areas of size $\frac{1}{4}$km^2 each.
- The total number of hits is 537.
- The average number of hits per square is $537/576 = 0.93$ hits per square.



Grinstead-Snell Example 5.4 and Feller vol I Sec VI.7(b).

London bombing data fitted by Poisson

Number of hits k	0	1	2	3	4	5+
Number of areas with k hits	229	211	93	35	7	1
$576 \cdot P(X = k)$, $X \sim \text{Pois}(0.93)$	227	211	99	31	7	2

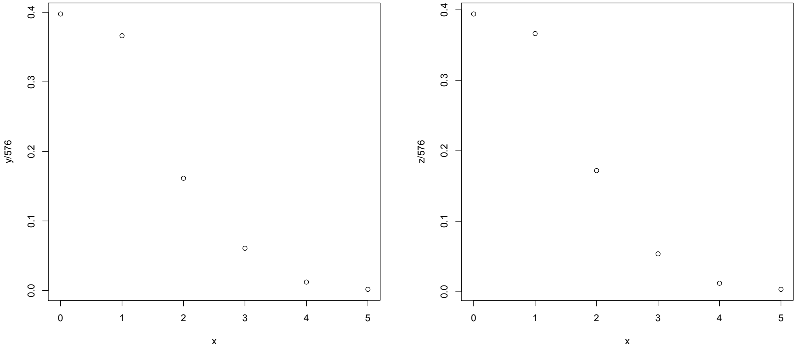


Figure: Empirical proportions vs. Pois(0.93)

Interpretation

Poisson distribution is surprisingly good; as judged by the χ^2-criterion, under ideal conditions some 88 per cent of comparable observations should show a worse agreement. It is interesting to note that most people believed in a tendency of the points of impact to cluster. If this were true, there would be a higher frequency of areas with either many hits or no hit and a deficiency in the intermediate classes. Table 4 indicates perfect randomness and homogeneity of the area; we have here an instructive illustration of the established fact that to the untrained eye randomness appears as regularity or tendency to cluster.

TABLE 4

EXAMPLE (*b*): FLYING-BOMB HITS ON LONDON

k	0	1	2	3	4	5 and over
N_k	229	211	93	35	7	1
$Np(k; 0.9323)$	226.74	211.39	98.54	30.62	7.14	1.57