The Poisson Distribution

Probability and Statistics for Data Science

Carlos Fernandez-Granda

NYU | COURANT INSTITUTE OF MATHEMATICAL SCIENCES

NYU DATA SCIENCE

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Earthquake

Goal: Model number of earthquakes in San Francisco over one year

Assumptions:

1. Earthquakes are independent
2. Probability of an earthquake in period of small length t is λt
3. Probability of more earthquakes is negligible when $t \to 0$

What is the probability of a earthquakes?

Discretizing into intervals

Strategy: Discretize year into n slots in the limit $n \to \infty$

$$
\begin{aligned}
&\mathrm{P}\left(a \text{ earthquakes }\right) \\
&= \lim_{n\to\infty} \mathrm{P}\left(a \text{ earthquakes in } n \text{ slots}\right) \\
&= \lim_{n\to\infty} \binom{n}{a} \left(\frac{\lambda}{n}\right)^{a} \left(1-\frac{\lambda}{n}\right)^{(n-a)} \\
&= \lim_{n\to\infty} \frac{n!\,\lambda^{a}}{a!\,(n-a)!\,(n-\lambda)^{a}} \left(1-\frac{\lambda}{n}\right)^{n} = \frac{\lambda^{a}\,e^{-\lambda}}{a!}
\end{aligned}
$$

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

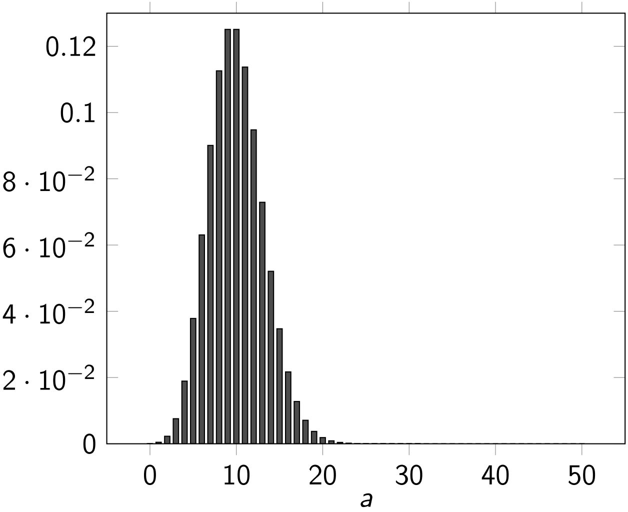
$$
\lim_{n\to\infty} \frac{n!}{(n-a)!\,(n-\lambda)^{a}} = \frac{n}{n-\lambda}\cdot\frac{n-1}{n-\lambda}\cdots\frac{n-a+1}{n-\lambda} = 1
$$

Poisson distribution

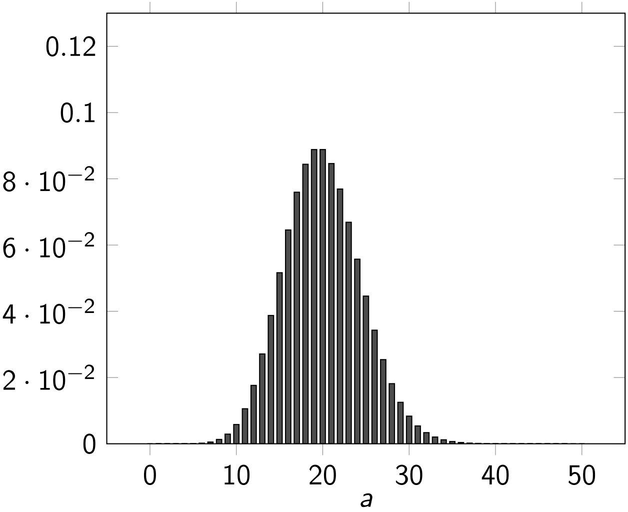
The pmf of a Poisson random variable with parameter λ is

$$p_{\tilde{a}}(a) = \frac{\lambda^{a} e^{-\lambda}}{a!} \qquad a = 0, 1, 2, \ldots$$

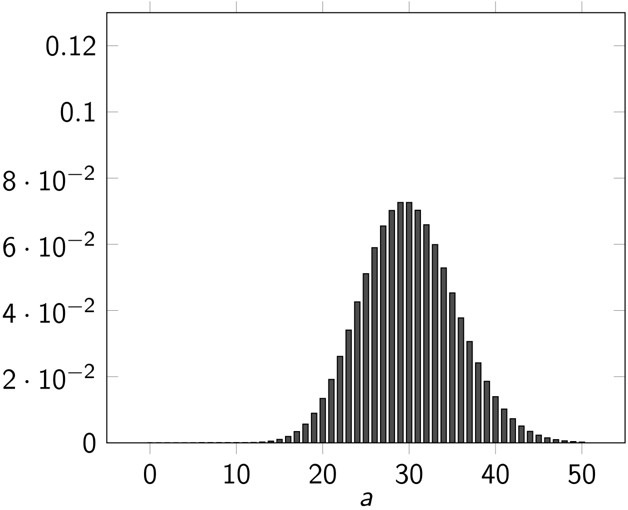
Poisson distribution $\lambda = 10$



Poisson distribution $\lambda = 20$



Poisson distribution $\lambda = 30$

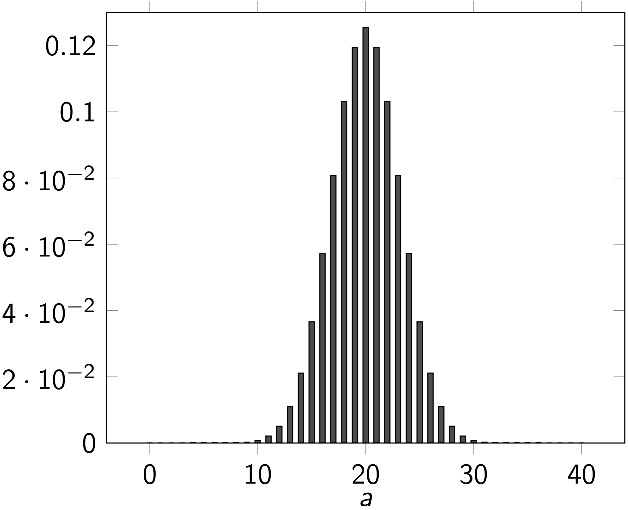


Convergence

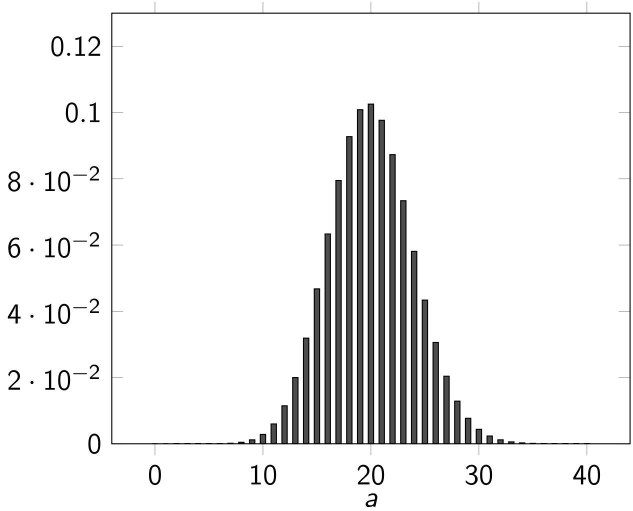
Pmf of binomial with parameters n and $\theta = \frac{\lambda}{n}$ converges to pmf of Poisson with parameter λ

This is an example of convergence in distribution

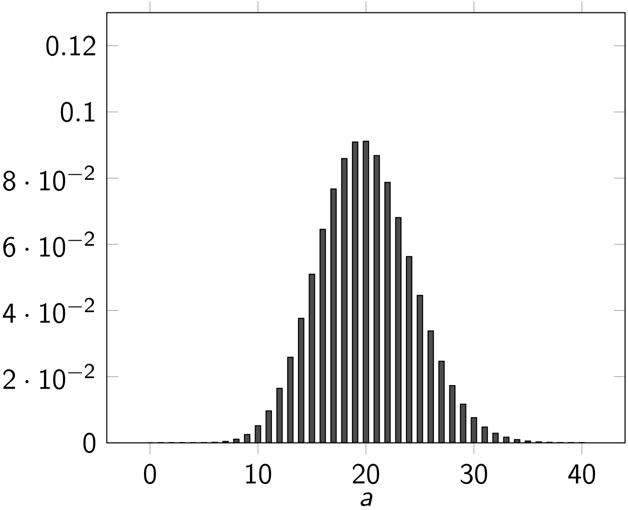
Binomial $n = 40$, $\theta = \frac{20}{40}$



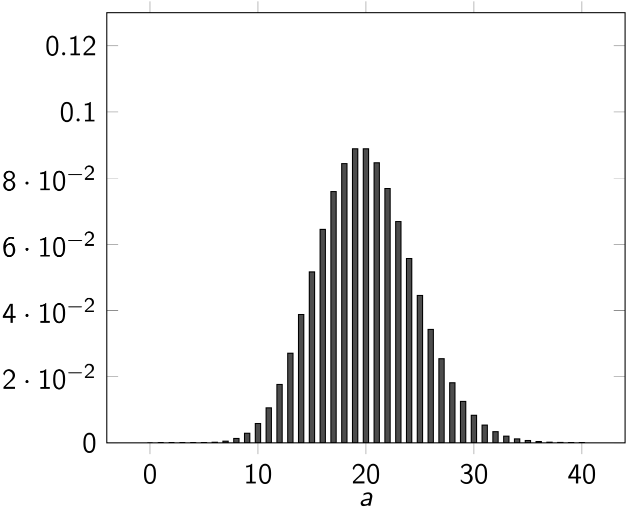
Binomial $n = 80$, $\theta = \frac{20}{80}$



Binomial $n = 400$, $\theta = \frac{20}{400}$



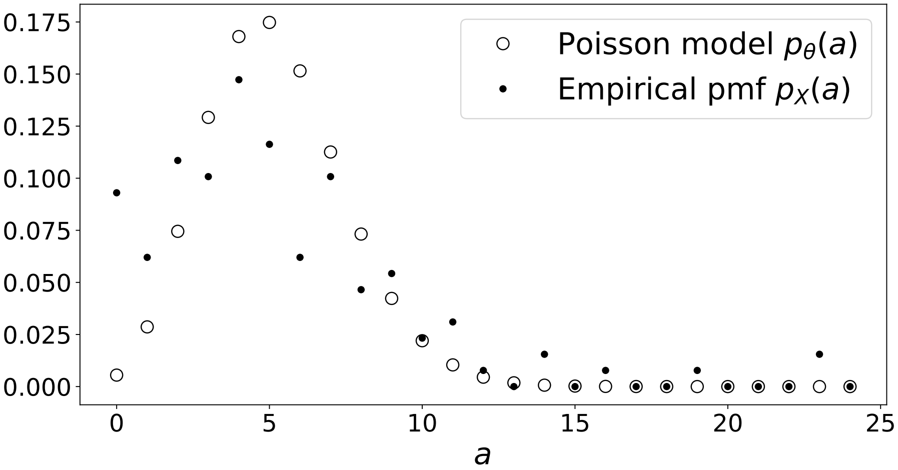
Poisson distribution $\lambda = 20$



Call center in bank

Goal: Model number of calls between 6 am and 7 am on weekdays

Poisson parametric model



What have we learned?

Derivation of the Poisson distribution