Maximum Likelihood Estimation for Continuous Models

Probability and Statistics for Data Science

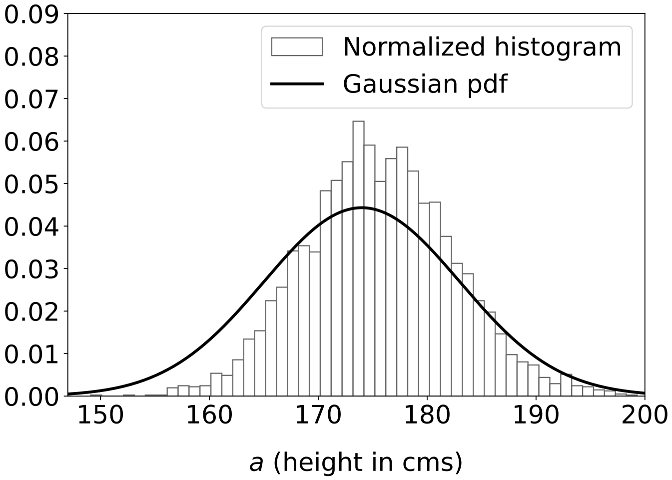
Carlos Fernandez-Granda

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

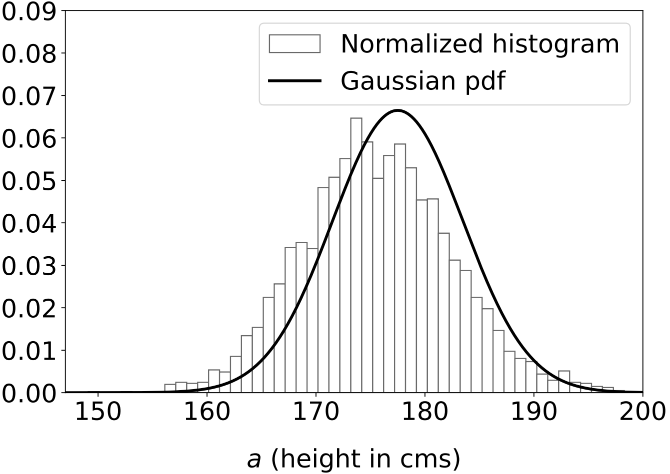
Goal

Learn how to fit continuous parametric models to data

$\mu_1 := 174,\ \sigma_1 := 9$



$\mu_2 := 177,\ \sigma_2 := 6$



One data point

Given a data point a and a parametric pdf f_θ, how should we choose θ?

Change of perspective: Interpret $f_\theta(a)$ as a function of θ

Assign the *highest possible density* to a

What if we have more data?

Assumptions

Let $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ be continuous random variables defined on the same probability space

They are identically distributed if they have the same pdf

They are independent, if events involving $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ are mutually independent

We often model data as i.i.d.

Likelihood

Data: x_1, x_2, ..., x_n

Under i.i.d. assumptions the *joint density* of the data under the parametric model is $\prod_{i=1}^{n} f_\theta(x_i)$

The likelihood of a model f_θ given a dataset $X := \{x_1, x_2, \dots, x_n\}$ is

$$\mathcal{L}_X(\theta) := \prod_{i=1}^{n} f_\theta(x_i)$$

The log-likelihood is

$$\log \mathcal{L}_X(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i)$$

Maximum likelihood

Given $f_\theta : A \to \mathbb{R}^+$ and a dataset $X := \{x_1, x_2, \ldots, x_n\}$, the ML estimate of θ is

$$\begin{aligned}\theta_{\mathrm{ML}} &:= \arg\max_\theta \mathcal{L}_X(\theta)\\ &= \arg\max_\theta \log \mathcal{L}_X(\theta)\end{aligned}$$

Exponential distribution

The pdf of an exponential random variable $\tilde{t}$ with parameter λ is

$$f_{\tilde{t}}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Log likelihood

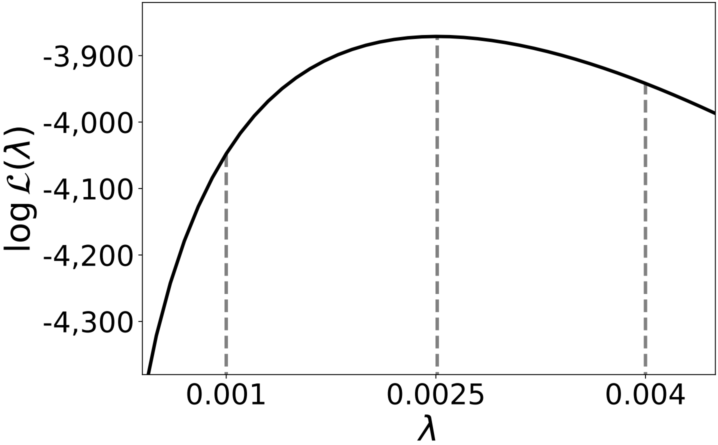
Dataset $X := \{x_1, x_2, \ldots, x_n\}$

$$\begin{aligned}
\log \mathcal{L}_X(\lambda) &= \sum_{i=1}^{n} \log f_\lambda(x_i) \\
&= \sum_{i=1}^{n} \log \lambda \exp(-\lambda x_i) \\
&= n \log \lambda - \lambda \sum_{i=1}^{n} x_i
\end{aligned}$$

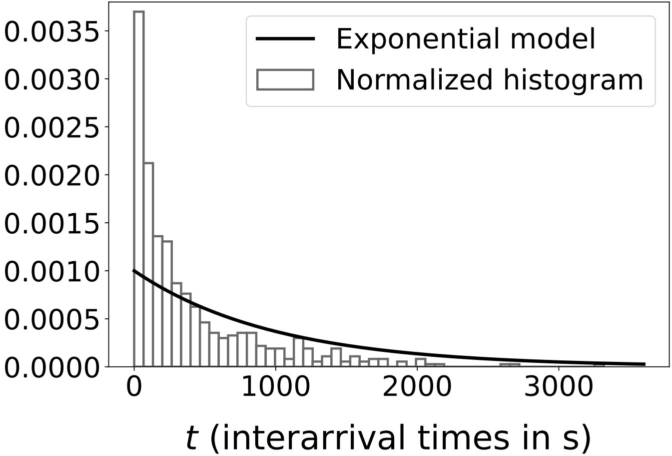
Call center in bank

Goal: Model time between calls (6 am-7 am on weekdays)

Log likelihood (call center data)

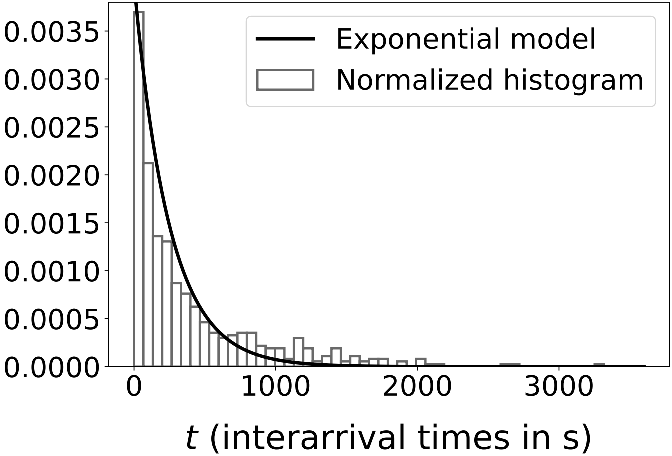


$\lambda := 10^{-3}$



Exponential model
Normalized histogram
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
1000
2000
3000
t (interarrival times in s)

$\lambda := 4 \cdot 10^{-3}$



ML estimate of exponential

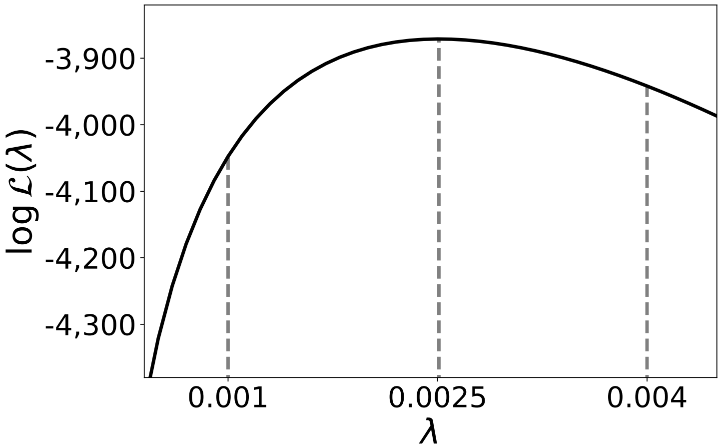
$$\log \mathcal{L}_X(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\mathrm{d} \log \mathcal{L}_{x_1,\ldots,x_n}(\lambda)}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\frac{\mathrm{d}^2 \log \mathcal{L}_{x_1,\ldots,x_n}(\lambda)}{\mathrm{d}\lambda^2} = -\frac{n}{\lambda^2} < 0 \quad \text{for all } \lambda > 0$$

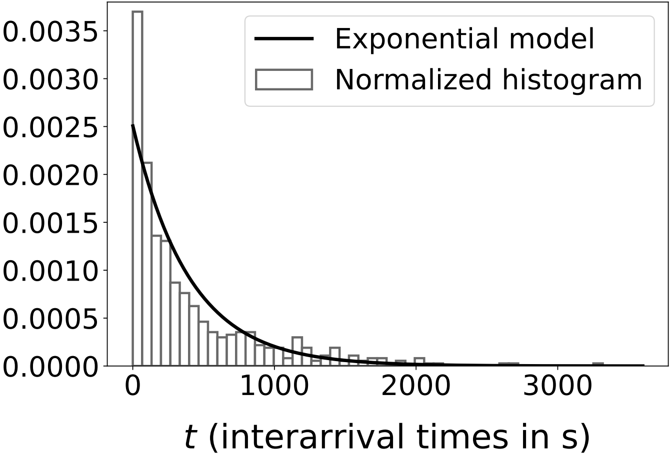
$$\lambda_{\mathrm{ML}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

Log likelihood (call center data)



Maximum-likelihood estimate

$\lambda_{\mathsf{ML}} := 2.5\,10^{-3}$



Gaussian distribution

The Gaussian or normal parametric pdf with mean μ and standard deviation σ is

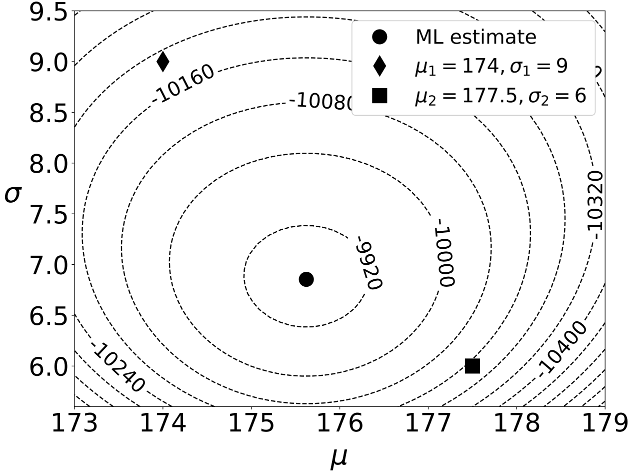
$$f_{\tilde{a}}(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(a-\mu)^2}{2\sigma^2}}$$

Log likelihood

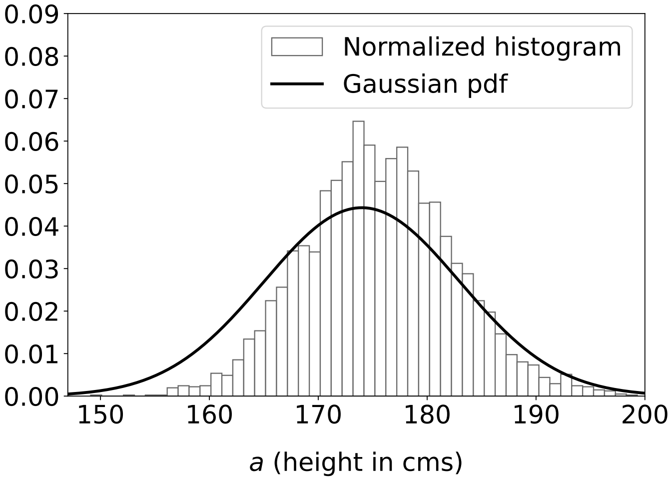
Dataset $X := \{x_1, x_2, \ldots, x_n\}$

$$\begin{aligned}
\log \mathcal{L}_X(\mu, \sigma) &= \sum_{i=1}^{n} \log f_{\mu,\sigma}(x_i) \\
&= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(a-\mu)^2}{2\sigma^2}} \\
&= -\frac{n \log(2\pi)}{2} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}
\end{aligned}$$

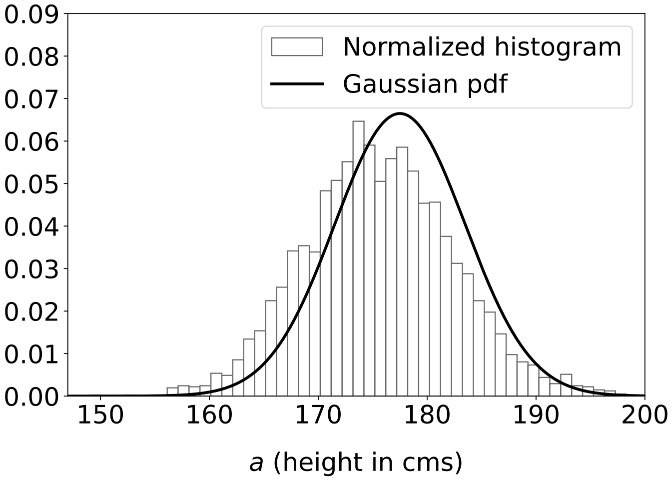
Log likelihood (height data)



$\mu_1 := 174,\ \sigma_1 := 9$



$\mu_2 := 177,\ \sigma_2 := 6$



ML estimate of Gaussian

Dataset $X := \{x_1, x_2, \ldots, x_n\}$

$$\log \mathcal{L}_X(\mu, \sigma) = -\frac{n \log(2\pi)}{2} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial \log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2}$$

$$\frac{\partial^2 \log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\mu, \sigma)}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\mu_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

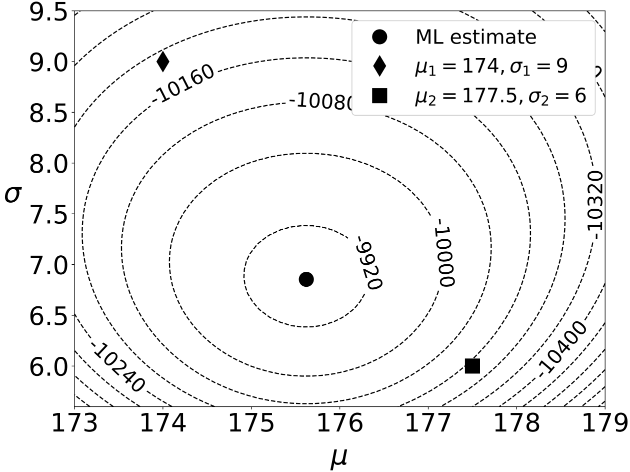
ML estimate of Gaussian

$$\log \mathcal{L}_X(\mu_{\text{ML}}, \sigma) = -\frac{n \log(2\pi)}{2} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu_{\text{ML}})^2}{2\sigma^2}$$

$$\frac{\partial \log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\mu_{\text{ML}}, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu_{\text{ML}})^2}{\sigma^3}$$

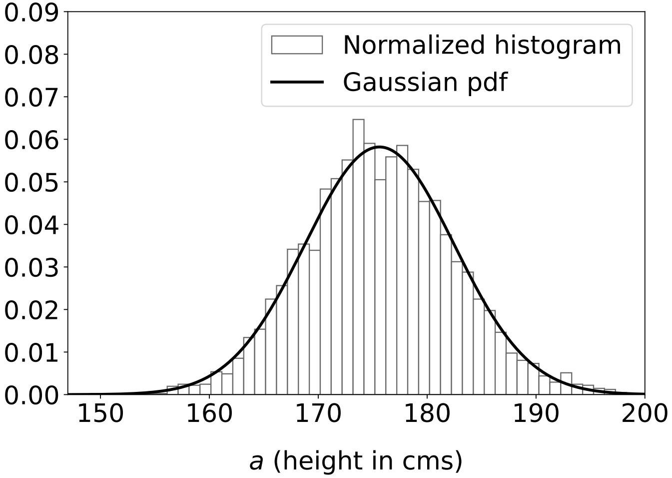
$$\sigma_{\text{ML}}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{\text{ML}})^2$$

Log likelihood (height data)



Maximum-likelihood estimate

$\mu_{\mathsf{ML}} := 177$, $\sigma_2 := 6$



What have we learned?

To fit continuous parametric models using maximum likelihood