Maximum-Likelihood Estimation

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

Fit parametric models to data

Free throws

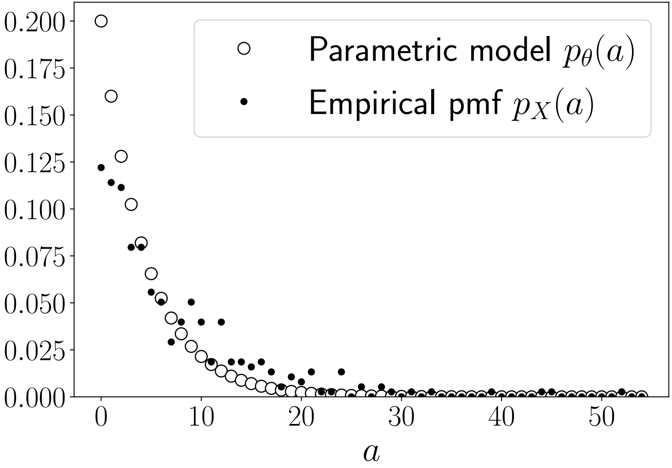
Goal: Model streaks of consecutive free throws

Data: 377 streaks from 3,015 free throws shot by Kevin Durant in the NBA

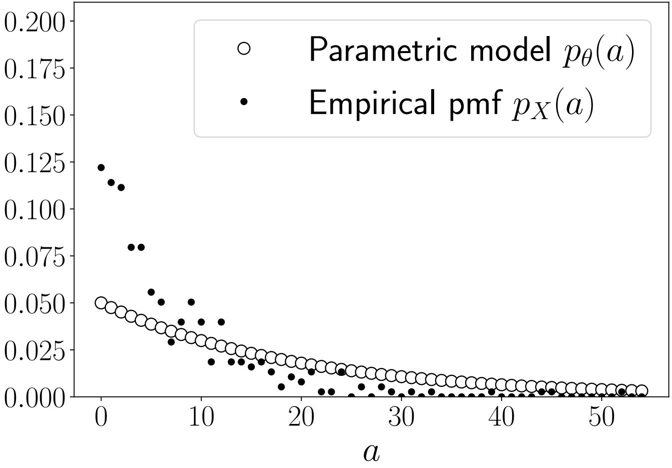
Parametric model

$$p_\theta(s) = \theta^s (1 - \theta)$$

Fit for $\theta := 0.8$



Fit for $\theta := 0.95$



One data point

Given a data point a and a parametric pmf p_θ, how should we choose θ?

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

Assign the *highest possible probability* to a

What if we have more data?

Assumptions

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

They are identically distributed if they have the same pmf

They are independent, if the events $\tilde{a}_1 = a_1, \tilde{a}_2 = a_2, \ldots, \tilde{a}_n = a_n$ are mutually independent

We often model data as i.i.d.

I.i.d. data

Data: x_1, x_2, ..., x_n

Under i.i.d. assumptions

$$\begin{aligned}\mathrm{P}(\tilde{a}_1 = x_1, \tilde{a}_2 = x_2, \ldots, \tilde{a}_n = x_n) &= \mathrm{P}(\tilde{a}_1 = x_1)\mathrm{P}(\tilde{a}_2 = x_2)\cdots\mathrm{P}(\tilde{a}_n = x_n) \\ &= \prod_{i=1}^{n} p_\theta(x_i)\end{aligned}$$

We choose θ to maximize this probability

Likelihood

The likelihood of a model p_θ given data $X := \{x_1, x_2, \ldots, x_n\}$ is

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

The log-likelihood function is

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

Maximum likelihood

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

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

Bernoulli distribution

Bernoulli pmf with parameter θ

$$p_\theta(1) = \theta$$

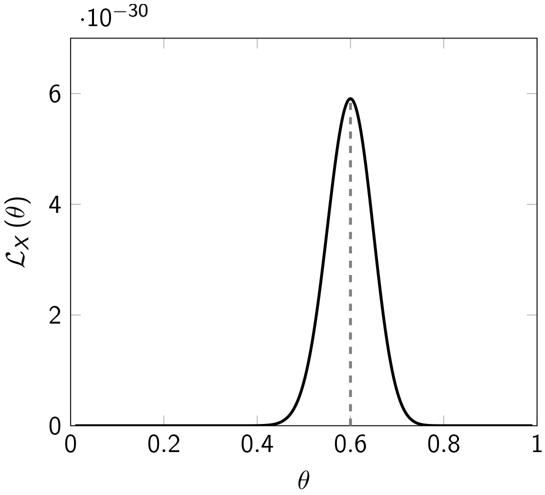
$$p_\theta(0) = 1 - \theta$$

Likelihood

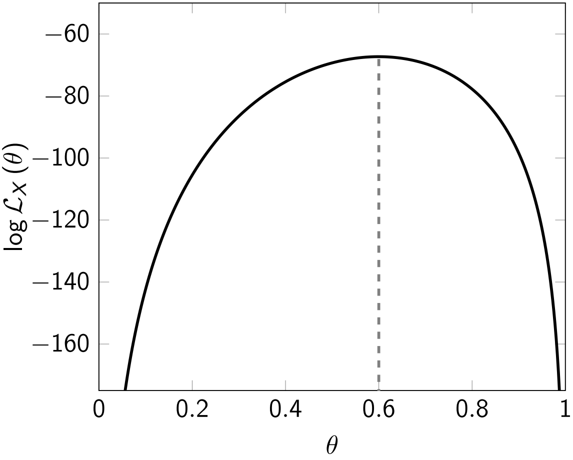
$$\mathcal{L}_{\{x_1,\ldots,x_n\}}(\theta) = \prod_{i=1}^{n} p_\theta(x_i)$$
$$= \theta^{n_1}(1-\theta)^{n_0}$$

$$\log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\theta) = n_1 \log \theta + n_0 \log(1-\theta)$$

Likelihood (60 ones, 40 zeros)



Log likelihood (60 ones, 40 zeros)



ML estimate

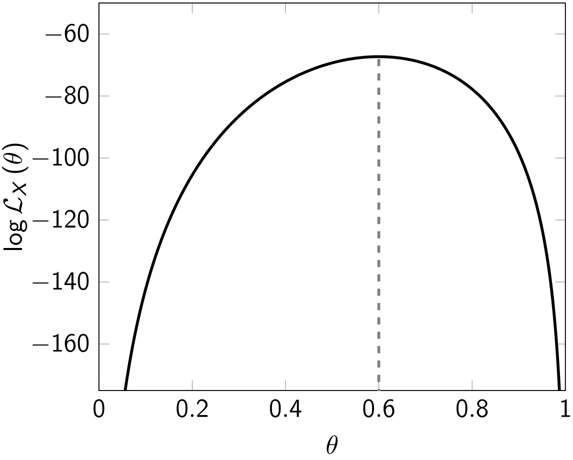
$$\log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\theta) = n_1 \log \theta + n_0 \log (1-\theta)$$

$$\frac{\mathrm{d} \log \mathcal{L}_{x_1,\ldots,x_n}(\theta)}{\mathrm{d}\theta} = \frac{n_1}{\theta} - \frac{n_0}{1-\theta}$$

$$\frac{\mathrm{d}^2 \log \mathcal{L}_{x_1,\ldots,x_n}(\theta)}{\mathrm{d}\theta^2} = -\frac{n_1}{\theta^2} - \frac{n_0}{(1-\theta)^2} < 0 \qquad \text{for all } \theta \in [0,1]$$

$$\theta_{\mathrm{ML}} = \frac{n_1}{n_0 + n_1}$$

Log likelihood (60 ones, 40 zeros)



Free throws

Goal: Model streaks of consecutive free throws

Data: 377 streaks from 3,015 free throws shot by Kevin Durant in the NBA

Parametric model

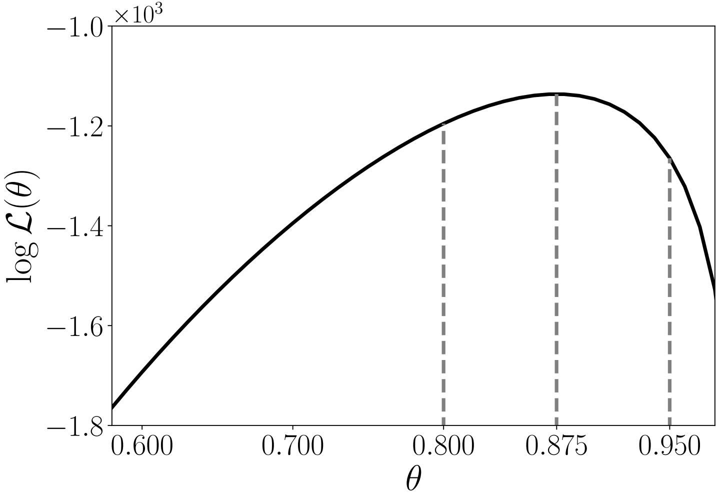
$$p_\theta(s) = \theta^s(1-\theta)$$

Log-likelihood

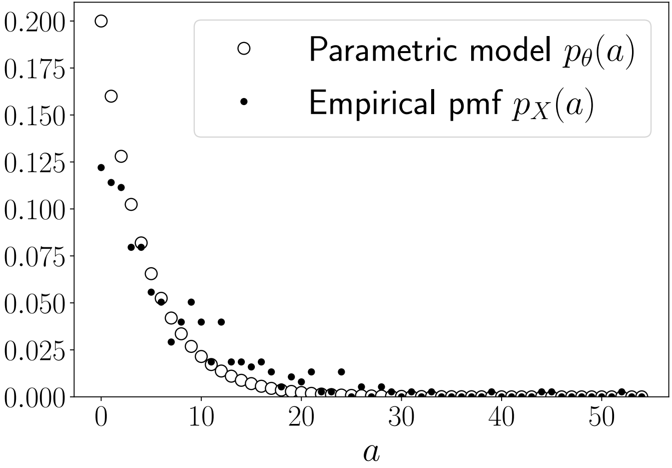
$$\begin{aligned}
\log \mathcal{L}_{\{x_1,\ldots,x_n\}}(\theta) &= \sum_{i=1}^{n} \log p_\theta(x_i) \\
&= \sum_{i=1}^{n} \log\left(\theta^{x_i}(1-\theta)\right) \\
&= \sum_{i=1}^{n} \left(x_i \log\theta + \log(1-\theta)\right) \\
&= \left(\sum_{i=1}^{n} x_i\right) \log\theta + n\log(1-\theta) \\
&= n_{\text{made}} \log\theta + n_{\text{missed}} \log(1-\theta)
\end{aligned}$$

Same as Bernoulli! $\theta_{\text{ML}} = 0.875$ is fraction of made free throws

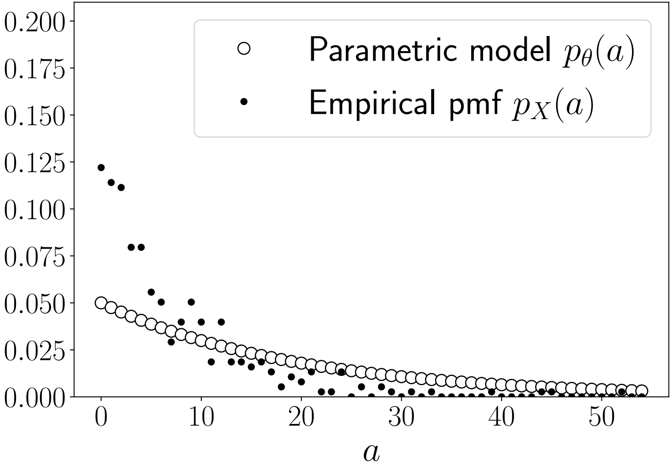
Log-likelihood



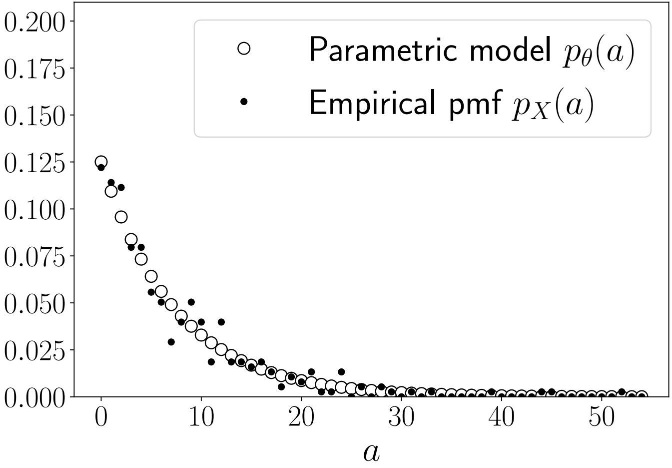
Fit for $\theta := 0.8$



Fit for $\theta := 0.95$



Fit for $\theta := 0.875$



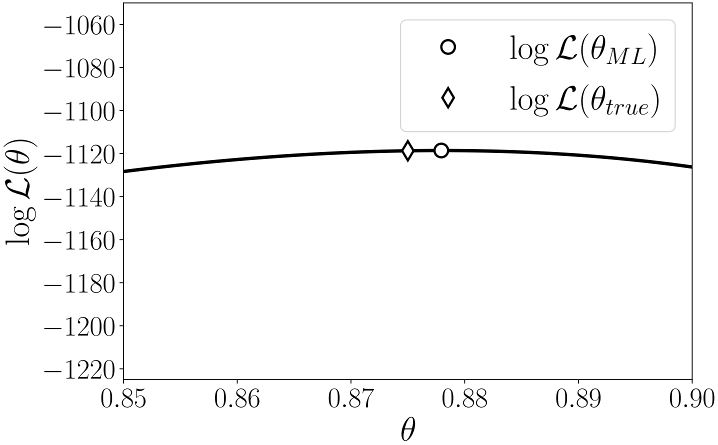
How stable is the ML estimate?

We simulate 3,015 i.i.d. free throws from

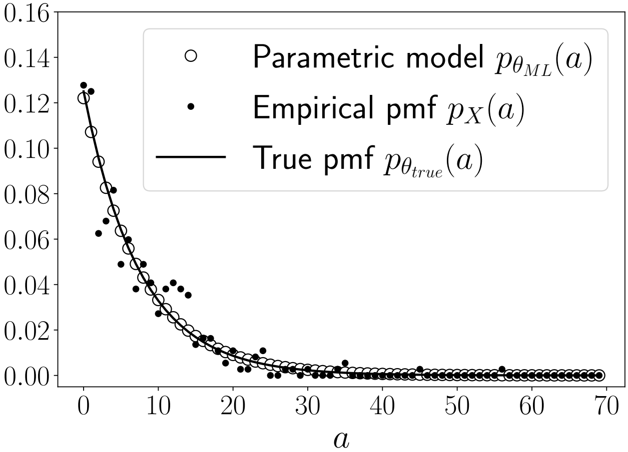
$$p_\theta(s) = \theta_{\text{true}}^{s}(1 - \theta_{\text{true}})$$

with $\theta_{\text{true}} := 0.875$ and compute θ_{ML}

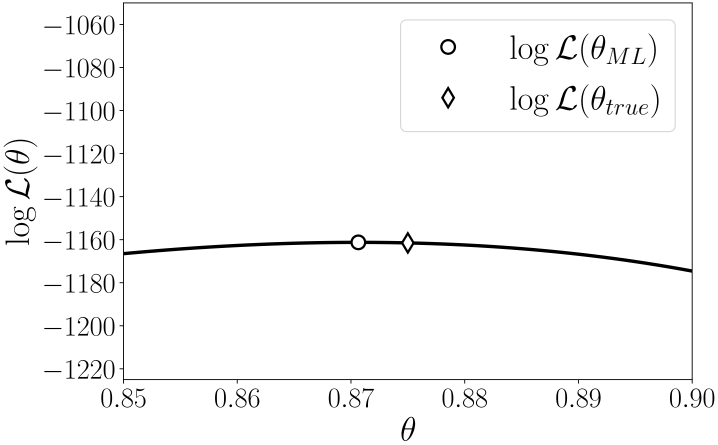
Log-likelihood



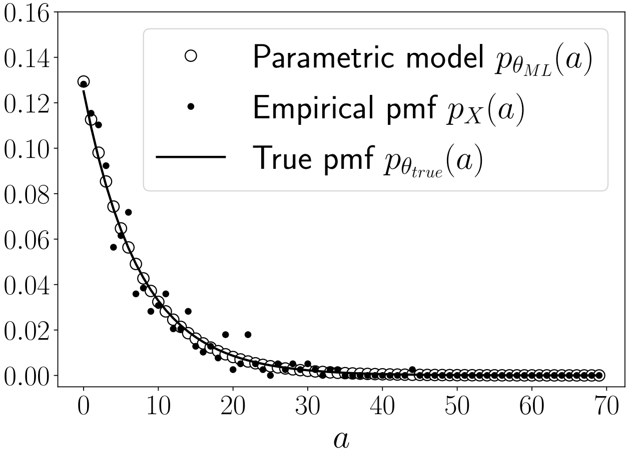
$\theta_{\text{ML}} = 0.873$



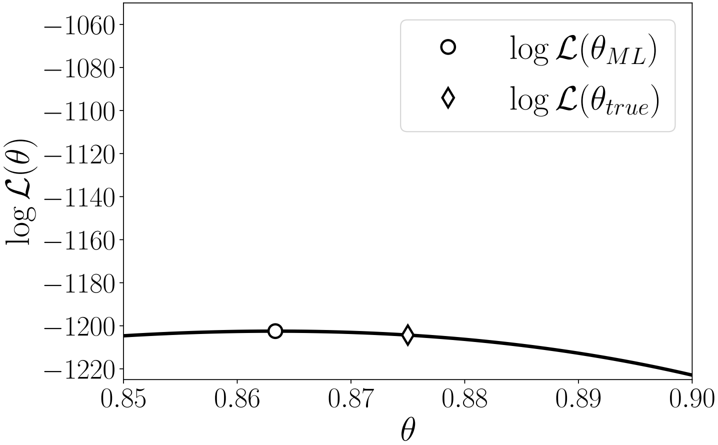
Log-likelihood



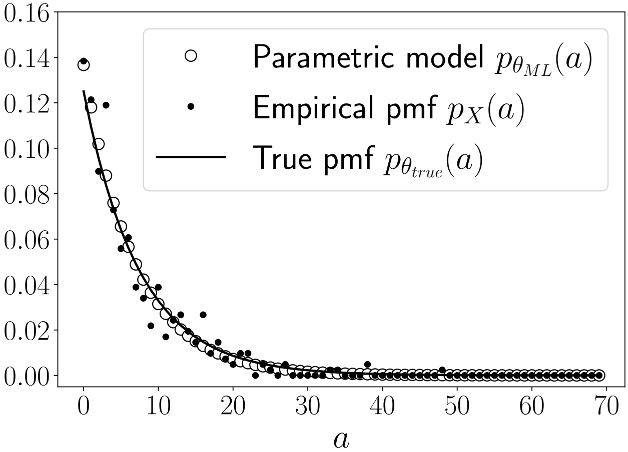
$\theta_{\mathsf{ML}} = 0.879$



Log-likelihood



$\theta_{\mathsf{ML}} = 0.867$



What have we learned?

To fit parametric models using maximum likelihood