The Geometric Distribution

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

Plan

Design a parametric model for our free-throw data

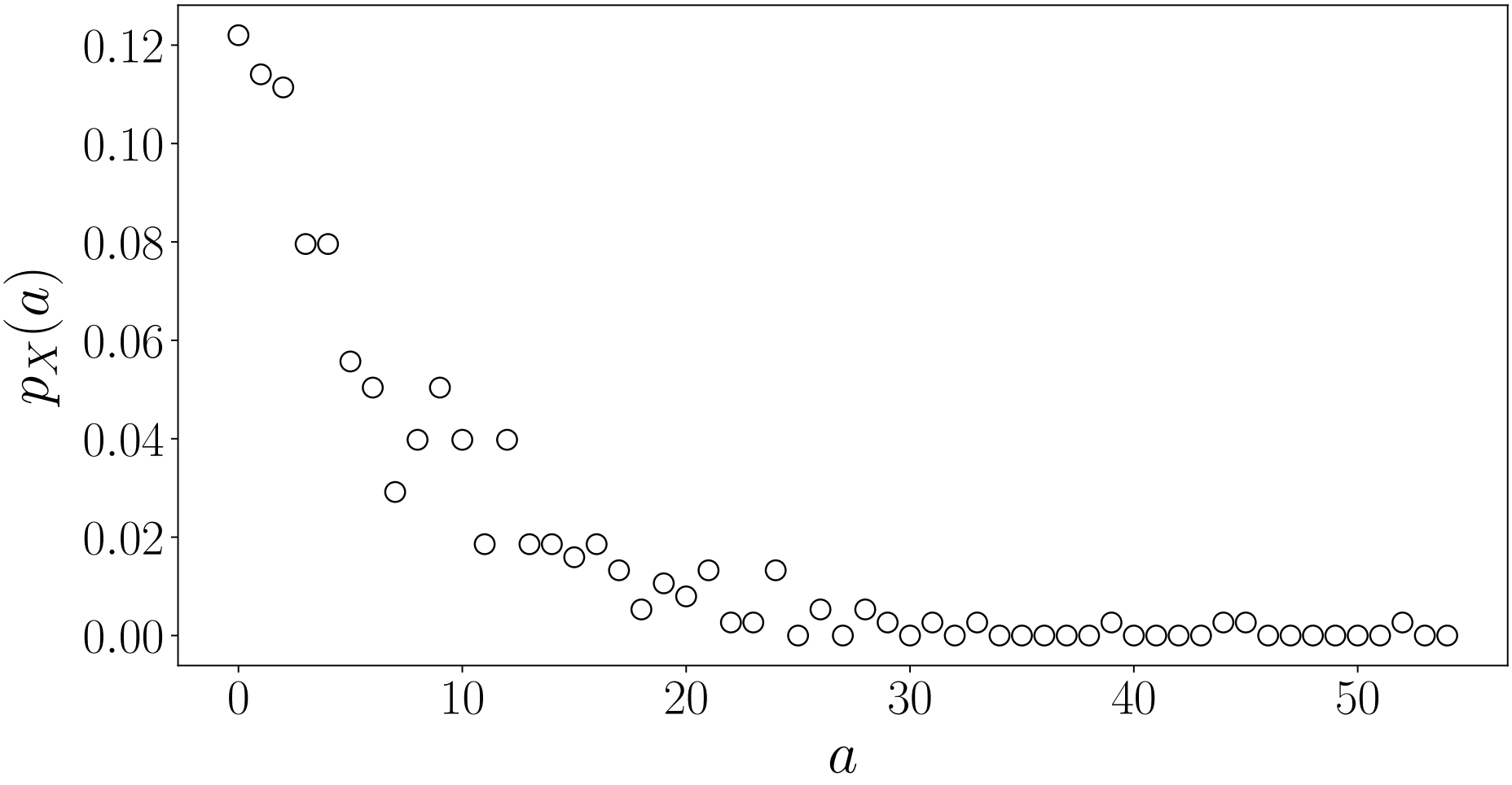
Free-throw data

Goal: Model streaks of consecutive free throws

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

$X := \{2, 4, 17, 3, 2, \ldots\}$

Nonparametric estimator



Alternative approach: Design parametric model

We need assumptions!

Assumption 1: Probability of making each attempt is fixed and equals θ

Assumption 2: All attempts are independent

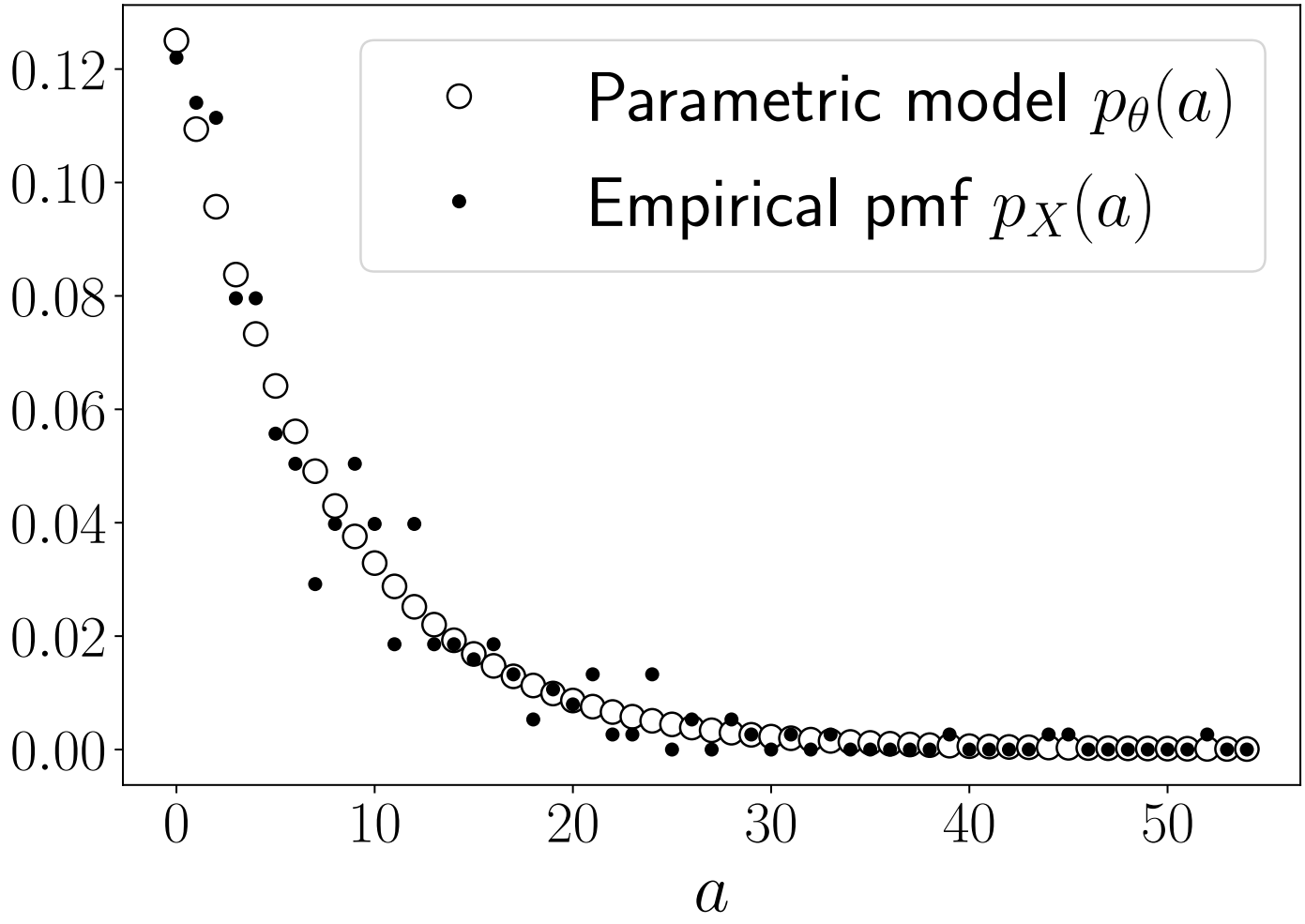
Is this true? No!

Under the assumptions

What is the probability of a streak of length s?

$$
\begin{aligned}
p_{\tilde{s}}(s) &= \mathrm{P}(s \text{ free throws are made, followed by a miss}) \\
&= \mathrm{P}(\text{1st made} \cap \text{2nd made} \cap \cdots \cap s\text{th made} \cap s+1\text{th missed}) \\
&= \mathrm{P}(\text{1st made})\mathrm{P}(\text{2nd made}) \cdots \mathrm{P}(s\text{th made})\mathrm{P}(s+1\text{th missed}) \\
&= \theta^s(1-\theta)
\end{aligned}
$$

Parametric model



Geometric distribution

Flip a coin until we obtain heads (probability of heads = α)

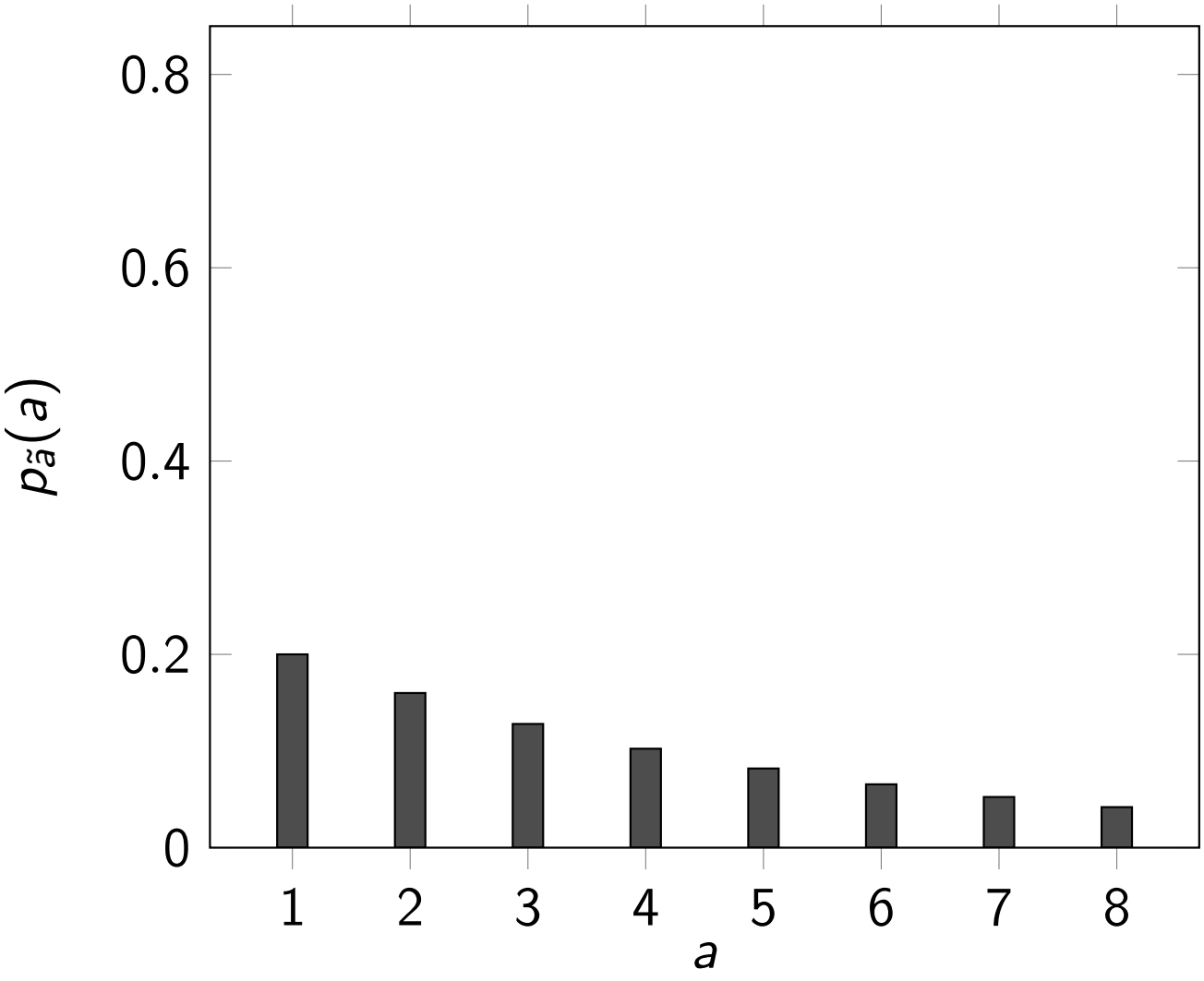
Probability of a flips?

Same situation with $\theta := 1 - \alpha$ and $s := a - 1$

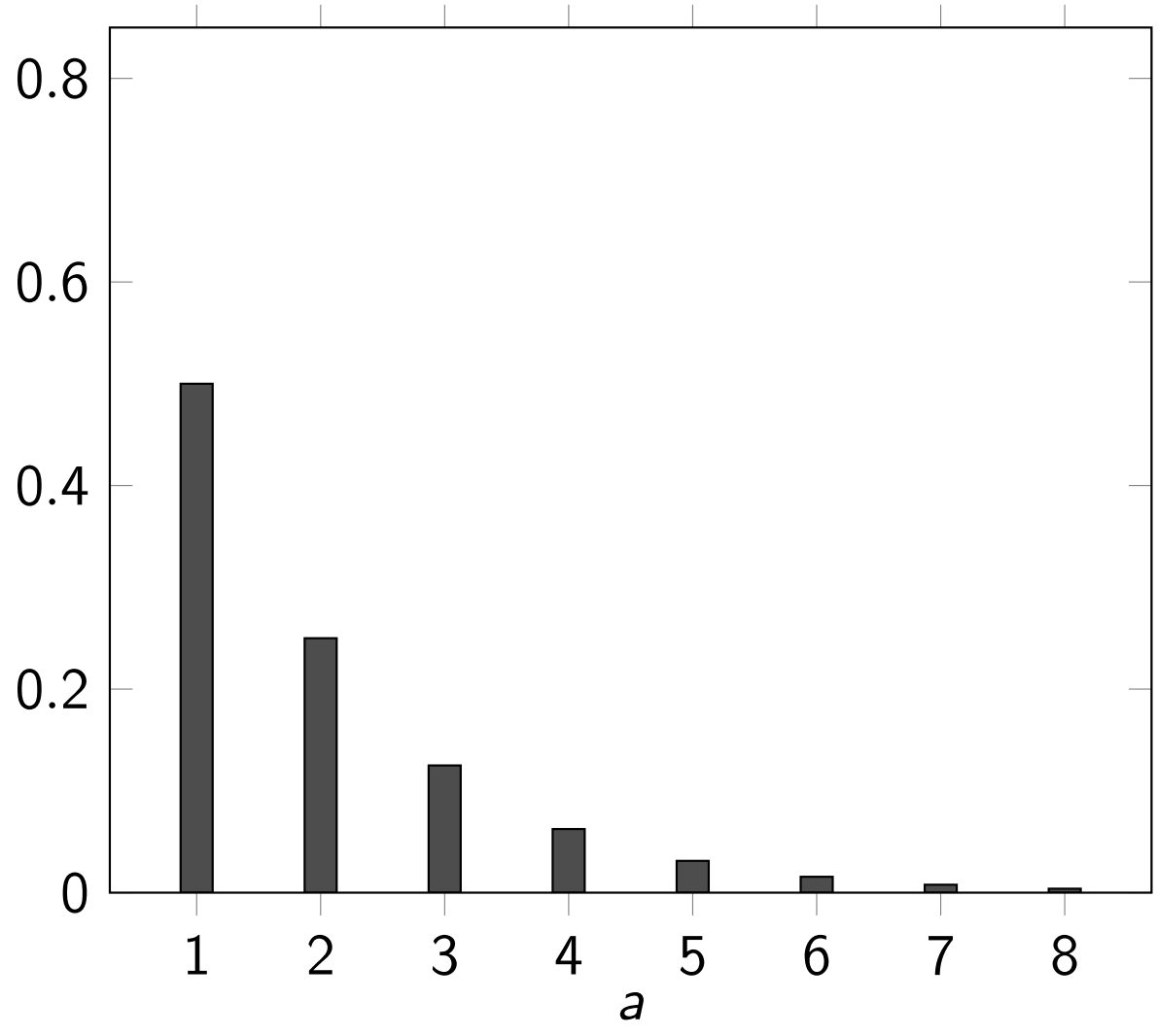
The pmf of a geometric random variable $\tilde{a}$ with parameter α is

$$p_{\tilde{a}}(a) = (1 - \alpha)^{a-1} \alpha, \quad a = 1, 2, \ldots$$

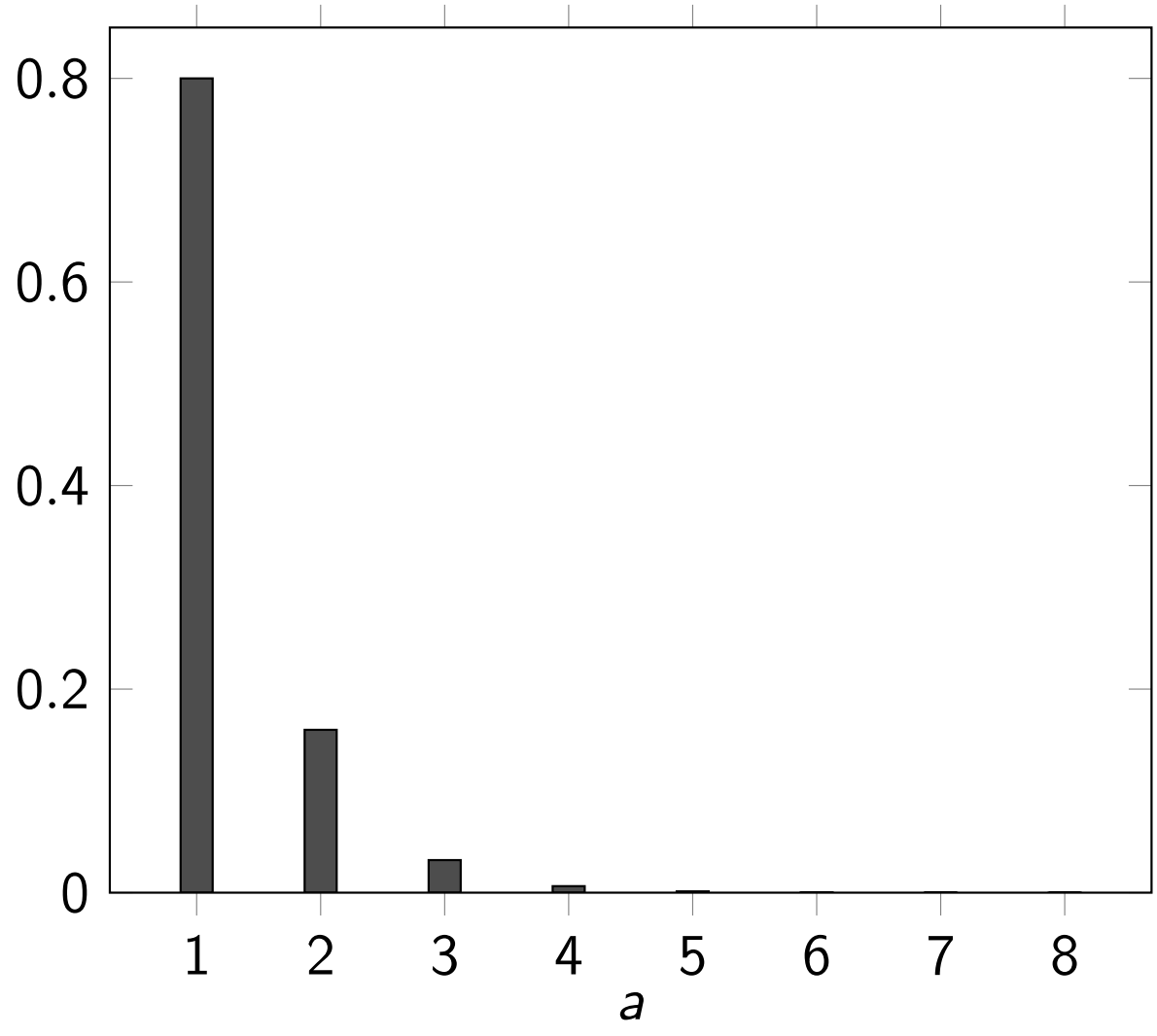
Geometric distribution $\alpha = 0.2$



Geometric distribution $\alpha = 0.5$



Geometric distribution $\alpha = 0.8$



What have we learned?

How to design a parametric model

Definition of the geometric distribution