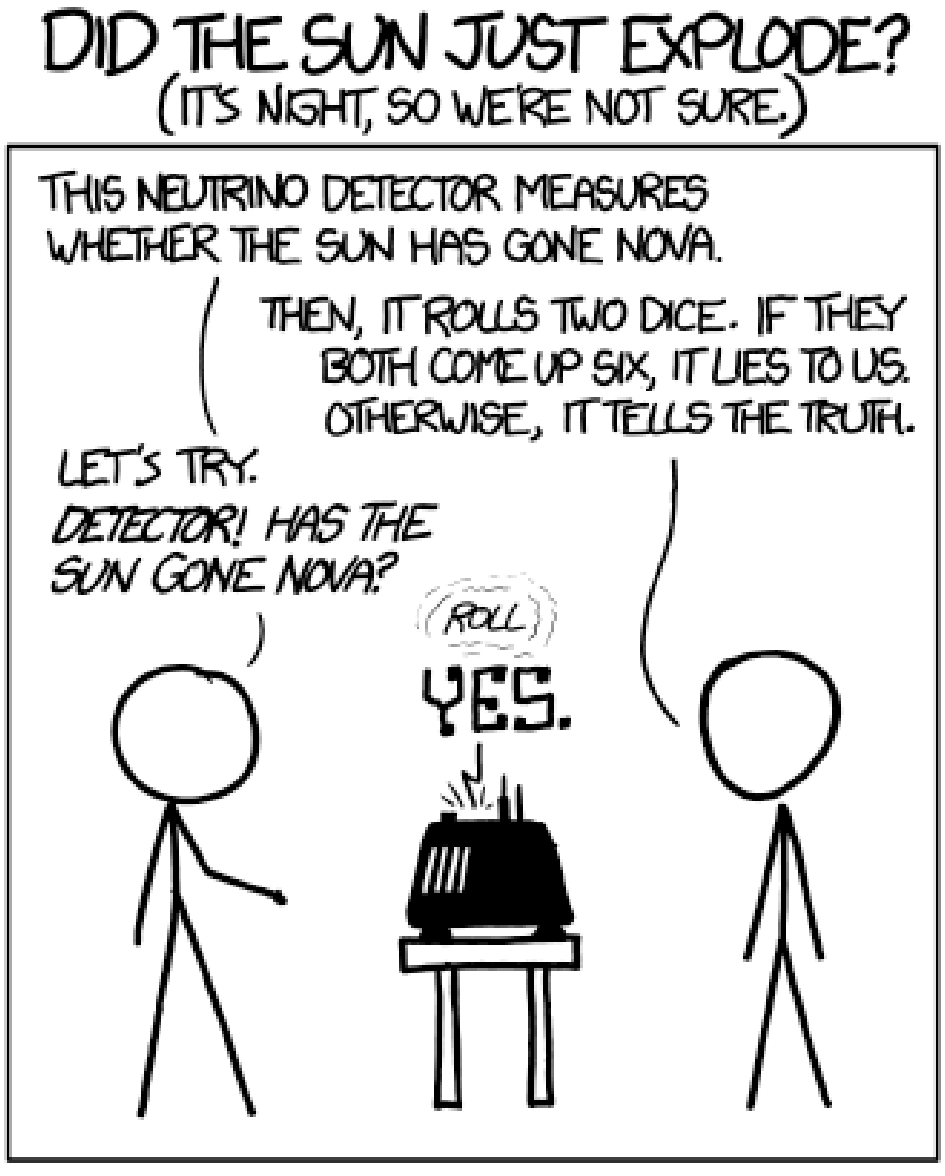
Data Generating Processes and Statistical Modeling

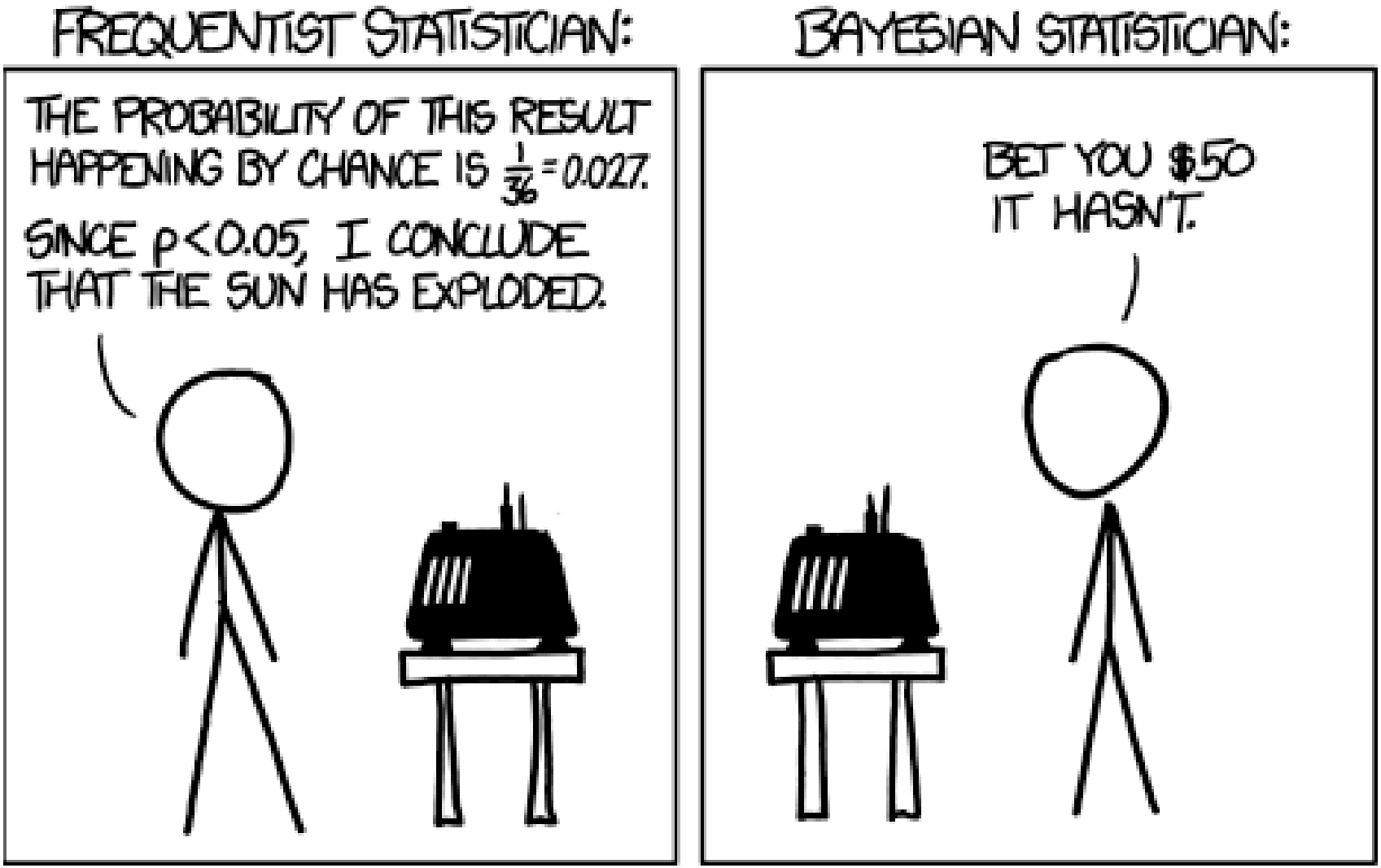
Jeffrey B. Arnold

January 28, 2016

Ontological Interpretations of Probability



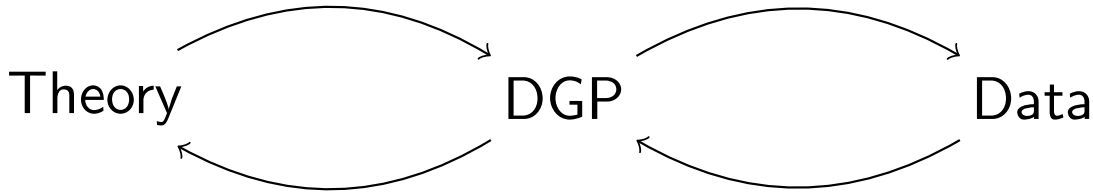
Ontological Interpretations of Probability



Ontological Interpretations of Probability

- Types of Probability
  - Frequentist (Classical)
  - Bayesian (Subjective)
- Probability of one-off events
- Beliefs of actors
- Mathematically equivalent rules

Data Generating Processes (DGP)



[A] data generating process (DGP) is rule or set of rules governing the social or political events that an analysst wishes to study and the rules by which observations of its results come tob be respresented in a dataset. A DGP . . . governs how a factors in a political process are related to each other. SMISS (p. 70)

DGP: Stochastic vs. Deterministic

- **Deterministic**: sufficient and necessary conditions to observe data
- **Stochastic**: probability of observing the data
- Why treat DGP as stochastic?
  - Sampling uncertainty
  - Theoretical uncertainty
  - Fundamental uncertainty

DGP and Statistical Models

- Stochastic DGP treats data as random variables
- Theoretically interested in specific parameters of the random variable: expectation
- *Model-based sampling*: assume a specific parametric probability distribution
- Problems with *model-based sampling*: DGP adds additional structure not implied by theory

Statistical Model with a Normal Distribution

$$
\begin{aligned}
Y &= \mu + \epsilon \\
\epsilon &\sim N(0, \sigma) \\
\mu &= \alpha + \beta X
\end{aligned}
$$

- Setting up a model is "statistical modeling" and defining our DGP
- How we estimate the parameters in the model α, β, σ to make inferences about the DGP is a separate problem.
- ϵ is the *stochastic component*
- μ is the *systematic component*