# Lecture Summary

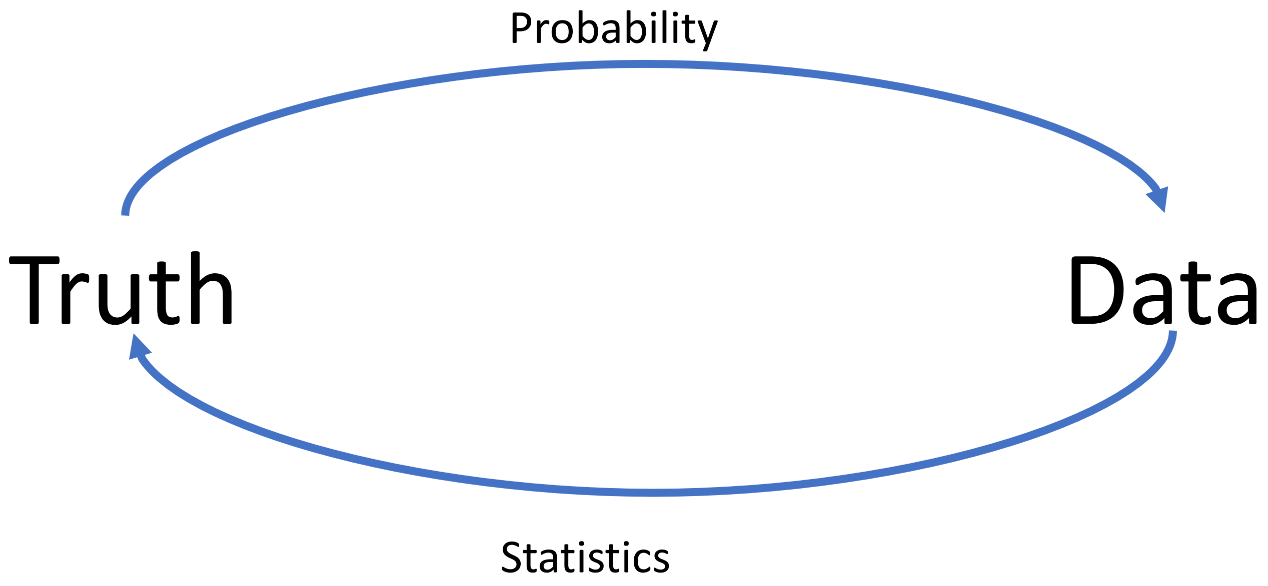
- Inference
- Bayesian vs Frequentist
- The Posterior
- Loss Function- Bayes Estimators
- Steps to Bayesian Inference
- Choices of priors
- Posterior predictive – Credible intervals

  - Material in DeGroot and Schervish Chapter 7.1-7.4

# What we did so far

- Probability
- Random Variables and their Distributions
- Expectation, Variance, Covariance
- Convergence of Random Variables

# Statistical Inference



# Types of Inference

- Estimation (prediction of a parameter)
- Hypothesis Testing
- Prediction

# Useful concepts

- Statistical Model:
  - identification of random variables of interest (observable or only hypothetically observable).
  - Specification of a family of possible distributions for the observable random variables.
  - Identification of the parameters of this distributions.
  - (Optional) Specification of a distribution for the unknown parameters.

# Useful Concepts

- Statistical Inference
  - A procedure that produces a probabilistic statement about some or all parts of a statistical model.

- Parameter
  - A characteristic that determine the joint distribution for the random variables of interest.

- Parameter space
  - The set of all possible values of a (vector of) parameter(s) $\theta$

# Example

- Suppose that 40 patients are going to be given a treatment for a condition and we will observe for each patient whether or not they recover from the condition. Then we have:
- Observed r.vs. $X_1, \dots X_{40}$
- Hypothetically observed r.vs $X_{41}, \dots$ the remaining patients that will receive the drug (not in the trial)
- Statistical Model: $X_i$ $i.i.d$, following a Bernoulli with parameter $\theta$
- Estimation: Find $\theta$

# Two approaches to statistical inference

- Classical (Frequentist) Inference
- $\theta$ is a number unknown to us
- You can compute $P(X_1, \dots X_{40}|\theta)$ for different possible $\theta's$

- Bayesian Inference
- Treat $\theta$ as a random variable (uncertainty)
- You can compute $P(\theta|X_1, \dots, X_{40})$ *(distribution)*

# Bayesian Inference

Likelihood

Prior

$$\underbrace{P(\theta|X_1,\ldots,X_n)}_{\text{Posterior}} = \frac{P(X_1,\ldots,X_n|\theta)P(\theta)}{P(X_1,\ldots,X_n)}$$

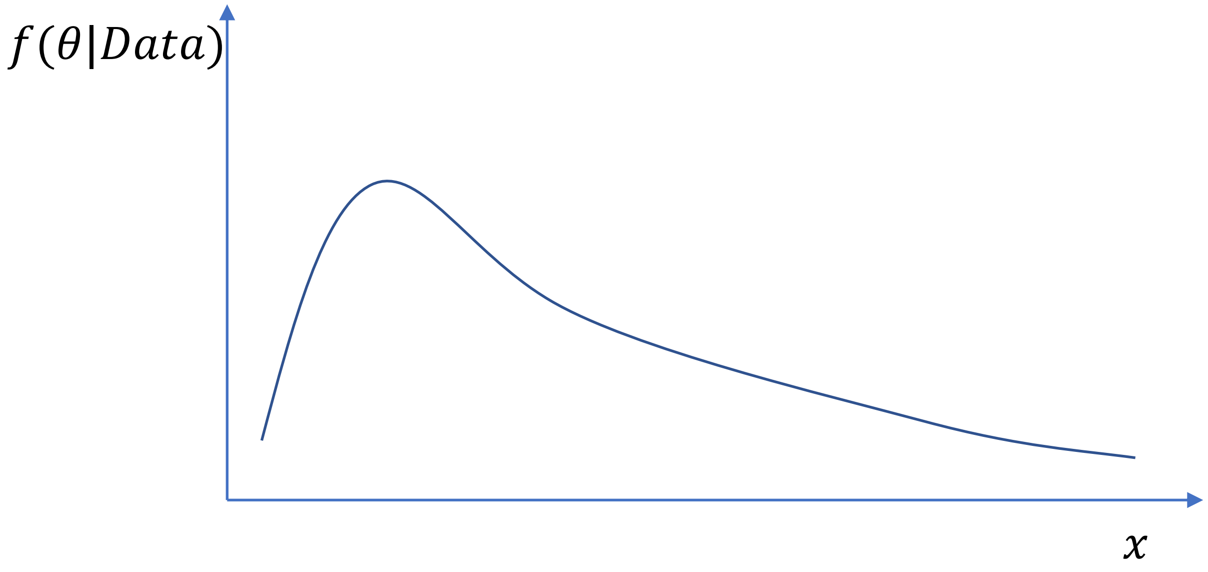
Posterior

Normalizing Constant

Bayes Rule

# Posterior

- The posterior is a distribution




How to pick a single $\hat{\theta}$?

- Maximum a posteriori (MAP) estimator: Find the value that maximizes the posterior

- Other?

# Loss Function

- $L(\theta, a)$: Quantifies how far  your estimate $a$ is from the true value $\theta$.

- Examples of loss functions:
- Mean Squared Error: $(a-\theta)^2$
- Mean Absolute Error: $|a-\theta|$
- Zero-one loss: $0, if\ a=\theta, 1\ otherwise$.

- The loss is a random variable
- We are looking for the estimate $\alpha$ that minimizes $E(L(\theta, a)|Data)$
- Bayes estimator: Minimizes $E(L(\theta, a)|Data)$ for all possible $Data$.

# How do we perform Bayesian Inference

- Pick a prior $P(\theta)$
- Compute the likelihood $P(X_1, \ldots, X_n|\theta)$
- Compute the normalization constant

  $P(X_1, \ldots, X_n) = \int P(X_1, \ldots, X_n|\theta) f(\theta) d\theta$

  (Also known as the marginal likelihood)

- Difficult, not always necessary

# Example: Bernoulli Distribution with Uniform prior

- $X_1, \dots, X_n$ follow a Bernoulli distribution
- We want to estimate $P(\theta | X_1, \dots, X_n)$

$$P(\theta | X_1, \dots, X_n) = \frac{P(X_1, \dots, X_n | \theta) P(\theta)}{P(X_1, \dots, X_n)}$$

- No prior information: All $p$ in [0,1] are equally likely: $P(\theta) \sim Beta(1, 1)$
- Posterior follows $Beta(1 + \sum X_i, 1 + n - \sum X_i)$

# Choice of Prior

- When the prior and the posterior follow the same family of distributions, the prior is called a conjugate prior.

- Sometimes it is convenient to pick a prior that does not have a proper distribution (e.g., “uniform” for the mean of a Normal). This is called an improper prior.

- In the large sample limit, the effect of the prior vanishes.

# Using the posterior

- Predicting the next observation (posterior predictive):

$$P(X_{n+1}\,|X_1, \dots, X_n) = \int P(X_n|\theta)f(\theta|X_1, \dots, X_n)d\theta$$

- Credible Interval -
- For α ∈ (0, 1), a Bayesian confidence region with level $\alpha$, or an $100(1-\alpha)$-credible interval is a random subset $R$ of the parameter space, which depends on the sample $X_1, \dots, X_n$, such that:

$$P(\theta \in R|X_1, \dots, X_n) = 1 - \alpha$$