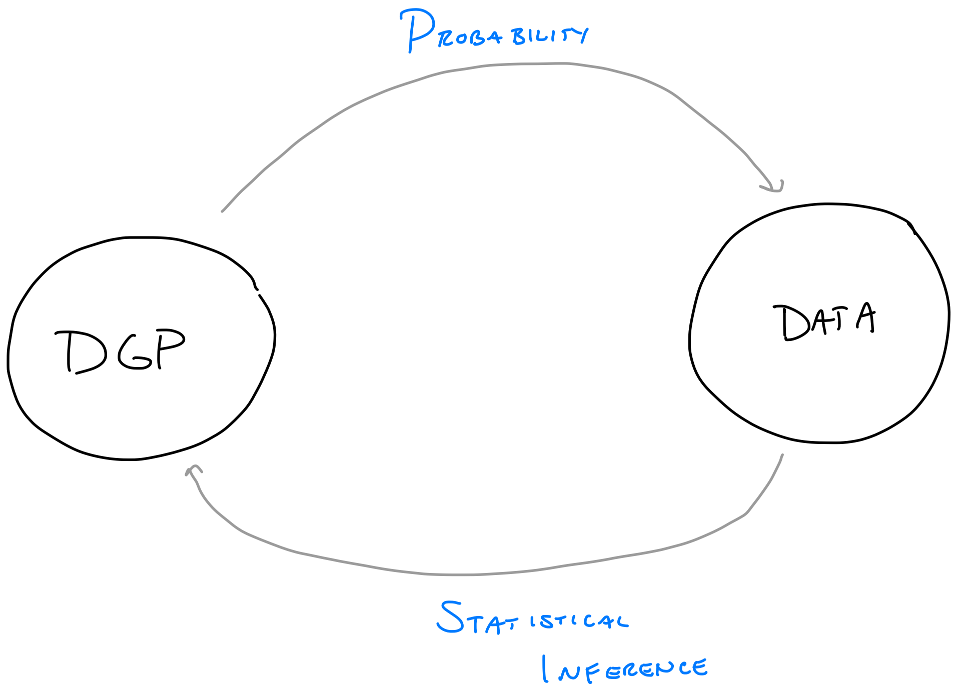
Bayesian Inference

Stat 510



Probability
DGP
DATA
Statistical
Inference

DEFINITION

A FUNCTION $\mathbb{P}$ IS A PROBABILITY DISTRIBUTION IF

1. $\mathbb{P}(A) \geq 0$ FOR ALL A

2. $\mathbb{P}(\Omega) = 1$

   SAMPLE SPACE

3. $A_1, A_2, A_3, \ldots$ DISJOINT

   THEN $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

PROBABILITY

THEORY → RULES → MATH

APPLICATION → INTERPRETATION → STATISTICS

So far...

Frequentist

Now...

Bayesian

Frequentist Statistics

OBJECTIVE

- Probabilities are limiting relative frequencies.
- Parameters are fixed, unknown constants.
- Want methods with well-defined long run frequency properties.

BAYESIAN STATISTICS

SUBJECTIVE

- PROBABILITIES ARE DEGREES OF BELIEF.
- WE CAN MAKE PROBABILITY STATEMENTS ABOUT PARAMETERS.
- INFERENCE IS DONE BASED ON DISTRIBUTION OF PARAMETERS.

THM BAYES' THEOREM

LET $A_1, A_2, \ldots, A_k$ BE A PARTITION OF Ω

$\mathbb{P}(A_i) > 0$ FOR EACH i

$\mathbb{P}(B) > 0$

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_j)\,\mathbb{P}(A_j)}{\mathbb{P}(B)}$$

$$= \frac{\mathbb{P}(B \mid A_j)\,\mathbb{P}(A_j)}{\sum_{i=1}^{k} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)}$$

MAIN IDEA

- EXPRESS BELIEF THROUGH PRIOR.
- MODEL DATA USING LIKELIHOOD
- USE BAYES THEOREM TO UPDATE BELIEF.
  THAT IS, OBTAIN THE POSTERIOR.

Parameter θ

Observation x

$$\underbrace{f(\theta \mid x)}_{\text{Posterior}} = \frac{\overbrace{f(x \mid \theta)}^{\text{Likelihood}} \; \overbrace{f(\theta)}^{\text{Prior}}}{\underbrace{\int f(x \mid \theta)\, f(\theta)\, d\theta}_{\text{Normalizing constant}}}$$

PARAMETER θ

OBSERVATIONS $X_1, X_2, \dots X_n$

X^n

$$f(x_1, \dots x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \mathcal{L}_n(\theta)$$

x^n

$$f(\theta \mid x^n) = \frac{f(x^n \mid \theta) f(\theta)}{\int f(x^n \mid \theta) f(\theta) \, d\theta} = \frac{\mathcal{L}_n(\theta) f(\theta)}{c_n}$$

IS PROPORTIONAL TO

$$f(\theta \mid x^n) \propto \mathcal{L}(\theta)\, f(\theta)$$

POSTERIOR

LIKELIHOOD

PRIOR

Point Estimates

Examples: mean / median / mode of posterior

$$\bar{\theta}_n = \int \theta f(\theta | x^n) \, d\theta$$

This math could get hard...

INTERVAL ESTIMATES

FIND a, b SUCH THAT

$$\int_{-\infty}^{a} f(\theta \mid x^n)\, d\theta = \int_{b}^{\infty} f(\theta \mid x^n)\, d\theta = \alpha/2$$

$C = (a, b)$ THEN

$$\mathbb{P}(\theta \in C \mid x^n) = \int_{a}^{b} f(\theta \mid x^n)\, d\theta = 1 - \alpha$$

C IS A $1-\alpha$ <u>POSTERIOR INTERVAL</u>

Bayesian Testing

$$H_0: \theta \geq \theta_0$$

$$H_1: \theta < \theta_0$$

Can calculate probabilities like

$$\mathbb{P}(H_0 \mid x^n)$$

$$\mathbb{P}(H_1 \mid x^n)$$

EXAMPLE

$$X_1, \ldots, X_n \sim \text{BERNOULLI}(p)$$

"FLAT PRIOR"

$$p \sim \text{UNIF}(0,1) \qquad f(p) = 1, \quad 0 \le p \le 1$$

$$f(p \mid x^n) \propto f(p)\,\mathcal{L}_n(p)$$

$$= p^s (1-p)^{n-s} = p^{(s+1)-1}(1-p)^{(n-s+1)-1}$$

$$S = \sum_{i=1}^{n} x_i$$

- VALID LIKELIHOOD
- NOT VALID DENSITY

Beta Distribution

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad \mathbb{E}[x] = \frac{\alpha}{\alpha+\beta}$$

$$f(p \mid x^n) \propto p^{(s+1)-1}(1-p)^{(n-s+1)-1}$$

$$f(p \mid x^n) = \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} p^{(s+1)-1}(1-p)^{(n-s+1)-1}$$

Posterior mean

$$p \mid x^n \sim \text{Beta}(s+1, n-s+1) \qquad \bar{p} = \frac{s+1}{n+2}$$

$$\boxed{\text{NO INTEGRATION!}}$$

$$X_1, \ldots, X_n \sim \text{Bernoulli}(p)$$ LIKELIHOOD

$$p \sim \text{Beta}(\alpha, \beta)$$ PRIOR

$$p \mid x^n \sim \text{Beta}(\alpha + s, \beta + n - s)$$ POSTERIOR

CONJUGATE PRIOR

POSTERIOR MEAN

$$\bar{p} = \frac{\alpha + s}{\alpha + \beta + n} = \left(\frac{n}{\alpha + \beta + n}\right) \hat{p} + \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right) p_0$$

MLE

PRIOR MEAN

$$p_0 = \frac{\alpha}{\alpha + \beta}$$