Probability vs Statistics

Probability If $X_1, X_2, ..., X_n \sim f(x_1, x_2, ..., x_n; \theta)$, what can we say about

$$\mathbb{P}[g(X_1, ..., X_n) \in A] = \int_{g(x_1,...,x_n)\in A} f(x_1, ..., x_n)\ dx_1...dx_n$$

or

$$\mathbb{E}[h(X_1, ..., X_n)] = \int h(x_1, ..., x_n) f(x_1, ..., x_n)\ dx_1...dx_n?$$

Statistics If $X_1, X_2, ..., X_n \sim f(x_1, x_2, ..., x_n; \theta)$, what can we say about f — "often", the unknown parameter θ — based on observed values of $X_1, X_2, ..., X_n$?

Remark The parameter $\theta \in \mathbb{R}^d$ may be multi-dimensional, i.e., $\theta = (\theta_1, \theta_2, ..., \theta_d)^\top$.

I. Frequentist Approach

- treat θ as a fixed (i.e., non-random) constant, albeit unknown
- want to estimate the value of θ
- in particular, by

  $$\widehat{\theta} = g(X_1, X_2, ..., X_n),$$

  referred to as an estimator
- a key question: how can we say $\widehat{\theta}_1 = g_1(X_1, ..., X_n)$ is better than $\widehat{\theta}_2 = g_2(X_1, ..., X_n)$, without knowing what θ actually is

"Hard" Probability Problem #1: Functions of RVs

- the estimator

$$\widehat{\theta} = g(X_1, X_2, ..., X_n)$$

is a function of RVs, so an RV itself

- interested in its distribution

- no way to assess numeric estimate, e.g.,

$$g(x_1, x_2, ..., x_n) = 8.45,$$

because don't know true value θ to start with

Sampling Distribution

$$
f(x_1, ..., x_n; \theta) \rightarrow \left\{ \begin{array}{ccc} \{x_1^{(1)}, \; ... \; , x_n^{(1)}\} & \rightarrow & \widehat{\theta}^{(1)} \\ \{x_1^{(2)}, \; ... \; , x_n^{(2)}\} & \rightarrow & \widehat{\theta}^{(2)} \\ \{x_1^{(3)}, \; ... \; , x_n^{(3)}\} & \rightarrow & \widehat{\theta}^{(3)} \\ \vdots & & \vdots \\ \{x_1^{(B)}, \; ... \; , x_n^{(B)}\} & \rightarrow & \widehat{\theta}^{(B)} \end{array} \right\} \text{distribution of } \widehat{\theta}
$$

imagine many data sets from the same data-generating process

Computer Simulation

```
for b = 1 to B
  data = generate(model)
  theta[b] = estimate(data)
end for
histogram(theta[1:B])
```

Exercise From $X_1, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$, estimate the unknown parameter θ with ...

II. Bayesian Approach

- $X_1, ..., X_n \sim f(x_1, ..., x_n; \theta)$ [really, $f(x_1, ..., x_n|\theta)$ here]
- parameter θ is unknown, so treat it as a random variable
- a random variable must have a distribution, so

$$\theta \sim \pi(\theta),$$

called the prior distribution

- "only" objective is to find

$$\pi(\theta|x_1, ..., x_n) = \frac{f(x_1, ..., x_n|\theta)\pi(\theta)}{\int f(x_1, ..., x_n|\theta)\pi(\theta)d\theta},$$

the posterior distribution, by the Bayes theorem

"Hard" Probability Problem #2: Compound Distributions

- quantity in denominator,

$$\underbrace{m(x_1,...,x_n)}_{\text{marginal of } x_1,...,x_n} \equiv \int \underbrace{f(x_1,...,x_n|\theta)\pi(\theta)}_{\text{joint of } x_1,...x_n,\theta} d\theta,$$

  is a compound distribution — the distribution of `something` (here, $X_1, ..., X_n$) depends on some parameter (here, θ), which is itself an RV; then, the (marginal, unconditional) distribution of `something` is a compound distribution

- main technical bottleneck, as RHS integral often intractable

- main philosophical bottleneck is choice of prior, $\pi(\theta)$