LECTURE 21

- Readings: Sections 8.1–8.2

Lecture outline

- Statistical inference
  - Contrast with probability theory
  - Bayesian vs. classical
- Bayesian inference
  - Four primary forms of Bayes’ rule
  - Types of problems/outputs
  - MAP estimation

Statistics

- Drawing inferences from limited and imperfect data
- Design and interpretation of experiments
  - polling, census, medical/pharmaceutical trials
- Netflix competition
- Finance
- Signal processing
  - Tracking, detection, speaker identification, forensics

Bayesian vs. classical inference

- Want to make inferences about parameter(s) $\theta$
- Bayesian: $\theta$ is a realization of random variable $\Theta$
- Classical: $\theta$ is unknown but not random

Four versions of Bayes’ rule

- $\Theta$ discrete, $X$ discrete: $p_{\Theta|X}(\theta|x) = \frac{p_{\Theta}(\theta)p_{X|\Theta}(x|\theta)}{\sum_k p_{\Theta}(k)p_{X|\Theta}(x|k)}$
- $\Theta$ discrete, $X$ cont.: $p_{\Theta|X}(\theta|x) = \frac{p_{\Theta}(\theta)f_{X|\Theta}(x|\theta)}{\int f_{\Theta}(t)f_{X|\Theta}(x|t)dt}$
- $\Theta$ cont., $X$ discrete: $f_{\Theta|X}(\theta|x) = \frac{f_{\Theta}(\theta)p_{X|\Theta}(x|\theta)}{\int f_{\Theta}(t)p_{X|\Theta}(x|t)dt}$
- $\Theta$ cont., $X$ cont.: $f_{\Theta|X}(\theta|x) = \frac{f_{\Theta}(\theta)f_{X|\Theta}(x|\theta)}{\int f_{\Theta}(t)f_{X|\Theta}(x|t)dt}$

Parameter of a coin

- Suppose a coin has probability of heads $\theta$. What do we believe about $\theta$ after observing $X$ heads in $n$ tosses?
- What is the Bayesian approach?
  - $f_{\Theta|X}(\theta|x)$ = Beta($\alpha, \beta$) prior:
  - $f_{\Theta}(\theta) = \begin{cases} \frac{1}{B(\alpha, \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, & \text{if } 0 < \theta < 1; \\ 0, & \text{otherwise} \end{cases}$
  - $B(\alpha, \beta) = \frac{(\alpha-1)(\beta-1)!}{(\alpha+\beta-1)!}$

Common mean of normal random variables

- Suppose normal $X_1, X_2, \ldots, X_n$ have an unknown common mean $\theta$ and known variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$
- If $\Theta$ is normal and $X_i$s are conditionally independent given $\Theta$, then $f_{\Theta|X}$ is normal
  - [Algebraic details in Example 8.3]
- Every observed $X_i$ ⇒ posterior update within normal class
  - Only mean update and variance update
- Important in engineering applications
Questions asked about $\theta$

- **Binary hypothesis testing:**
  Choose between two possibilities for $\theta$

- **$m$-ary hypothesis testing:**
  Choose between $m$ possibilities for $\theta$

- **Estimation:**
  Pick a number $\hat{\theta}$ that approximates $\theta$
    - **Estimator:** random variable $\hat{\Theta} = g(X)$ for some $g$
    - **Estimate:** value $\hat{\theta}$ of an estimator (determined by realization $x$ of $X$)

Maximum a posteriori probability (MAP) rule

- Posterior distribution: PMF $p_{\Theta|X}(\cdot|x)$ or PDF $f_{\Theta|X}(\cdot|x)$

- Pick $\hat{\theta}$ such that
  \[
  p_{\Theta|X}(\hat{\theta}|x) = \max_{\theta} p_{\Theta|X}(\theta|x) \\
  f_{\Theta|X}(\hat{\theta}|x) = \max_{\theta} f_{\Theta|X}(\theta|x)
  \]