LECTURE 23

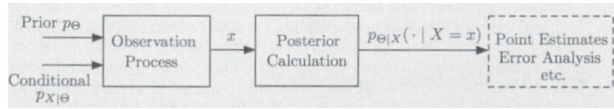
- **Readings:** Section 9.1 through p. 470
  Section 9.2 through p. 482
- Course VI Underground Guide Evaluations
  https://sixweb.mit.edu/student/evaluate/6.041-s2010
  until 11:59pm on May 16

Lecture outline

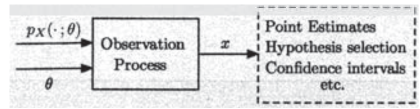
- Classical statistical inference
- Classical parameter estimation
  - bias
  - consistency
- Maximum likelihood estimation
- Confidence intervals

Bayesian vs. classical inference

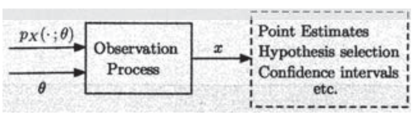
- Want to make inferences about *parameter(s)* θ
- Bayesian: θ is a realization of random variable Θ



- Classical: θ is unknown but not random



Classical parameter estimation



- θ is not a random variable
- Estimator $\widehat{\Theta}$ is a (different) random variable for each θ
- Must care about performance for all possible θ
  - Cannot average over θ because θ is not random!
- Robustness?
  - No sensitivity to prior because there is no prior
  - Still depends entirely on model for observation generation

Terminology/properties of estimators

- $\widehat{\Theta}_n$: estimator of θ from $X_1, X_2, \ldots, X_n$
- **Estimation error**: $\widetilde{\Theta}_n = \widehat{\Theta}_n - \theta$
- **Bias**: $b_\theta(\widehat{\Theta}_n) = \mathbf{E}_\theta[\widetilde{\Theta}_n] = \mathbf{E}_\theta[\widehat{\Theta}_n] - \theta$
- **Unbiased:** $b_\theta(\widehat{\Theta}_n) = 0$ or $\mathbf{E}_\theta[\widehat{\Theta}_n] = \theta$ (for **all** θ)
- **Asymptotically unbiased:** $\lim_{n\to\infty} b_\theta(\widehat{\Theta}_n) = 0$ for **all** θ
- **Consistent:** $\widehat{\Theta}_n$ converges in probability to θ, for **all** θ
- **Mean squared error (MSE):** (function of θ)

$$\begin{aligned}\mathbf{E}_\theta[(\widehat{\Theta}_n - \theta)^2] &= \mathrm{var}_\theta(\widehat{\Theta}_n - \theta) + (\mathbf{E}_\theta[\widehat{\Theta}_n - \theta])^2 \\ &= \mathrm{var}_\theta(\widehat{\Theta}_n) + \left(b_\theta(\widehat{\Theta}_n)\right)^2\end{aligned}$$

Examples

- Parameter θ is known to be positive
- X is exponentially distributed with parameter θ
- $\mathbf{E}_\theta[X] = 1/\theta$, so $\widehat{\Theta} = 1/X$ seems reasonable
  - $b_\theta(\widehat{\Theta}) = \int_0^\infty \frac{1}{x}\theta e^{-\theta x}\,dx = \infty$
- $\mathbf{P}_\theta\left(X < \frac{1}{\theta \ln 2}\right) = \frac{1}{2}$, so $\widehat{\Theta} = \frac{1}{X \ln 2}$ seems reasonable
- n obs: $\widehat{\Theta}_n = n/(X_1 + X_2 + \cdots + X_n)$ seems reasonable
  - $(X_1 + X_2 + \cdots + X_n)/n$ converges in probability to $1/\theta$
  - $n/(X_1 + X_2 + \cdots + X_n)$ converges in probability to θ

Maximum likelihood (ML) estimation

- Pick θ that "makes data most likely":

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_\theta p_X(x\,;\theta)$$

- Compare to (Bayesian) MAP estimation:

$$\hat{\theta}_{\mathrm{MAP}} = \arg\max_\theta \frac{p_{X|\Theta}(x\,|\,\theta)\,p_\Theta(\theta)}{p_X(x)}$$

- Advanced properties:
  - **Invariance** to invertible change of parameterization $\zeta = h(\theta)$, i.e., ML estimate will be $\hat{\zeta}_{\mathrm{ML}} = h(\hat{\theta}_{\mathrm{ML}})$
  - **Consistency:** ML estimates $\widehat{\Theta}_n$ from $(X_1, X_2, \ldots, X_n)$ form a consistent sequence
  - **Asymptotic normality:** $(\widehat{\Theta}_n - \theta)/\sigma(\widehat{\Theta}_n)$ approaches standard normal

Example: Exponential distribution parameter

- $X_1, X_2, \ldots, X_n$: indep. with exponential(θ) distribution
- ML estimate is

$$\begin{aligned}\hat{\theta}_{\text{ML}} &= \arg\max_{\theta} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n\,;\,\theta) \\ &= \arg\max_{\theta} \prod_{i=1}^{n} \theta e^{-\theta x_i} \\ &= \arg\max_{\theta} \left(n \log\theta - \theta \sum_{i=1}^{n} x_i \right)\end{aligned}$$

Find critical point by differentiating w.r.t. θ:

$$\begin{aligned}\hat{\theta}_{\text{ML}} &= \frac{n}{x_1 + x_2 + \cdots + x_n} \\ \widehat{\Theta}_n &= \frac{n}{X_1 + X_2 + \cdots + X_n}\end{aligned}$$

Example: Normal distribution mean

- $X_1, X_2, \ldots, X_n$: indep. with mean θ and variance 1
- ML estimate is

$$\begin{aligned}\hat{\theta}_{\text{ML}} &= \arg\max_{\theta} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n\,;\,\theta) \\ &= \arg\max_{\theta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x_i-\theta)^2/2} \\ &= \arg\min_{\theta} \sum_{i=1}^{n} (x_i - \theta)^2\end{aligned}$$

Find critical point by differentiating w.r.t. θ:

$$\begin{aligned}\hat{\theta}_{\text{ML}} &= \frac{x_1 + x_2 + \cdots + x_n}{n} \\ \widehat{\Theta}_n &= \frac{X_1 + X_2 + \cdots + X_n}{n}\end{aligned}$$

Confidence interval (CI)

- An estimate $\widehat{\Theta}_n$ is only a number with no reliability
- $[\widehat{\Theta}_n^-, \widehat{\Theta}_n^+]$ is a $1-\alpha$ **confidence interval** for θ when

$$\mathbf{P}_\theta\left(\widehat{\Theta}_n^- \le \theta \le \widehat{\Theta}_n^+\right) \ge 1 - \alpha \qquad \text{for } \textbf{all } \theta$$

  - often $\alpha = 0.01$ or 0.05
  - interpretation is subtle
  - sometimes want one-sided confidence interval

$$\left(-\infty, \widehat{\Theta}_n^+\right] \qquad \text{or} \qquad \left[\widehat{\Theta}_n^-, \infty\right)$$

Example: Exponential distribution parameter

- X: exponential with parameter θ
- Analyze $[a/X, b/X]$ as a confidence interval for θ:

$$\begin{aligned}\mathbf{P}_\theta\left(\frac{a}{X} \le \theta \le \frac{b}{X}\right) &= \mathbf{P}_\theta\left(\frac{a}{\theta} \le X \le \frac{b}{\theta}\right) \\ &= \int_{a/\theta}^{b/\theta} \theta e^{-\theta x}\, dx \\ &= -e^{-\theta x}\Big|_{x=a/\theta}^{x=b/\theta} = e^{-a} - e^{-b}\end{aligned}$$

No dependence on θ, so have a confidence interval

- Example: $\left[\frac{1}{4X}, \frac{4}{X}\right]$ is a 0.76 confidence interval ("76% CI")

Example: Normal distribution mean

- $X_1, X_2, \ldots, X_n$: indep. with mean θ and variance 1
- $\widehat{\Theta}_n = (X_1 + \cdots + X_n)/n$ has mean θ, variance $1/n$
- Analyze $[\widehat{\Theta}_n - \delta, \widehat{\Theta}_n + \delta]$ as a confidence interval for θ:

$$\begin{aligned}\mathbf{P}_\theta\left(\widehat{\Theta}_n - \delta \le \theta \le \widehat{\Theta}_n + \delta\right) &= \mathbf{P}_\theta\left(-\delta \le -\widehat{\Theta}_n + \theta \le \delta\right) \\ &= \mathbf{P}_\theta\left(-\delta \le \widehat{\Theta}_n - \theta \le \delta\right) \\ &= \mathbf{P}_\theta\left(-\delta\sqrt{n} \le \underbrace{\sqrt{n}(\widehat{\Theta}_n - \theta)}_{\text{standard normal}} \le \delta\sqrt{n}\right)\end{aligned}$$

  - For 0.95 CI, want $\Phi(-\delta\sqrt{n}) = 0.025$, or $\delta\sqrt{n} \approx 1.96$

$$\left[\widehat{\Theta}_n - \frac{1.96}{\sqrt{n}}, \widehat{\Theta}_n + \frac{1.96}{\sqrt{n}}\right]$$

Mean and variance estimation with unknown distribution

- $X_1, X_2, \ldots, X_n$: i.i.d., mean μ, variance σ^2 (both unknown)
- Sample mean $M_n = \dfrac{X_1 + \cdots + X_n}{n}$ well studied
  - $\mathbf{E}[M_n] = \mu$ (unbiased)
  - M_n converges in probability to θ (WLLN, consistency)
- Natural variance estimate $\overline{S}_n^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - M_n)^2$
  - $\mathbf{E}[\overline{S}_n^2] = \dfrac{n-1}{n}\sigma^2$ (biased, asymptotically unbiased)
  - $\hat{S}_n^2 = \dfrac{n}{n-1}\overline{S}_n^2$ unbiased
  - $\overline{S}_n^2$ and $\hat{S}_n^2$ both consistent