LECTURE 24

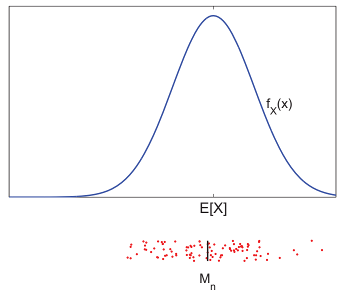
- **Readings:** Section 9.3
- Course VI Underground Guide Evaluations
  https://sixweb.mit.edu/student/evaluate/6.041-s2010
  until 11:59pm on May 16

Lecture outline

- Variance estimators revisited
- Linear regression
- Binary hypothesis testing
  - Types of errors
  - Likelihood ratio
  - Neyman–Pearson lemma

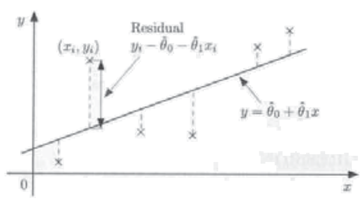
Variance estimation with unknown mean

- $X_1, X_2, \ldots, X_n$: i.i.d., mean μ, variance σ^2 (both unknown)
- $M_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ unbiased and consistent for μ



- $\overline{S}_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2$
  slightly biased:
  $$\mathbf{E}[\overline{S}_n^2] = \frac{n-1}{n}\sigma^2$$
- M_n deviates from μ in the direction that makes
  $$\sum_{i=1}^{n}(X_i - M_n)^2 < \sum_{i=1}^{n}(X_i - \mu)^2$$

Linear regression



- Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Model: $y \approx \theta_0 + \theta_1 x$
- Minimize sum of squared residuals:

$$(\hat\theta_0, \hat\theta_1) = \arg\min_{(\theta_0,\theta_1)} \sum_{i=1}^{n}(y_i - (\theta_0 + \theta_1 x_i))^2$$

Linear regression

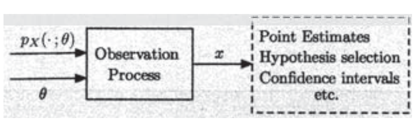
- Solution (set derivatives to zero):

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad \overline{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

$$\hat\theta_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad \hat\theta_0 = \overline{y} - \hat\theta_1 \overline{x}$$

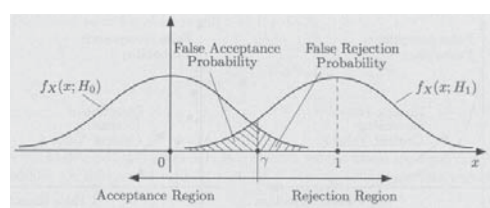
- Several problems, same solution
  - Maximum likelihood (linear model, normal noise)
  - Approximate Bayesian LMS estimation (linear model, i.i.d. zero-mean noise)
  - Approximate Bayesian linear LMS estimation

Binary hypothesis testing



- When θ takes two values, convention changes:
  - **null hypothesis** H_0:
    $X \sim p_X(\cdot\,; H_0)$ [or $f_X(\cdot\,; H_0)$]
  - **alternative hypothesis** H_1:
    $X \sim p_X(\cdot\,; H_1)$ [or $f_X(\cdot\,; H_1)$]
- Types of errors:
  - **Type I** (**false rejection**): rejecting H_0 when H_0 true
  - **Type II** (**false acceptance**): accepting H_0 when H_0 false

Binary hypothesis testing



- **Rejection region** R: H_0 is rejected when $X \in R$
- False rejection probability:
  $$\alpha(R) = \mathbf{P}(X \in R\,; H_0)$$
- False acceptance probability:
  $$\beta(R) = \mathbf{P}(X \notin R\,; H_1)$$

Review: (Bayesian) Binary hypothesis testing example

- Prior given: $\mathbf{P}(\Theta = 1) = p, \qquad \mathbf{P}(\Theta = 2) = 1 - p$
- Likelihoods given: $f_{X|\Theta}(x\,|\,1), \qquad f_{X|\Theta}(x\,|\,2)$
- Costs given: c_{12} (mistake 1 for 2), c_{21} (mistake 2 for 1)
- Minimize expected cost

Choose $\hat{\theta} = 2$ when

$$c_{21}\frac{\mathbf{P}(\Theta = 2) f_{X|\Theta}(x\,|\,2)}{f_X(x)} > c_{12}\frac{\mathbf{P}(\Theta = 1) f_{X|\Theta}(x\,|\,1)}{f_X(x)}$$

Equivalently, choose $\hat{\theta} = 2$ when

$$\underbrace{\frac{f_{X|\Theta}(x\,|\,2)}{f_{X|\Theta}(x\,|\,1)}}_{\text{likelihood ratio } L(x)} > \underbrace{\underbrace{\frac{c_{12}}{c_{21}}}_{\text{cost ratio}}\frac{\mathbf{P}(\Theta = 1)}{\mathbf{P}(\Theta = 2)}}_{\text{critical value}}$$

(Textbook does not have costs)

Bayesian binary hypothesis testing—new terminology

- Prior given: $\mathbf{P}(H_0 \text{ true}) = p, \qquad \mathbf{P}(H_1 \text{ true}) = 1 - p$
- Likelihoods given: $f_X(x\,;\,H_0), \qquad f_X(x\,;\,H_1)$
- Costs given: c_{I} (mistake H_0 for H_1), c_{II} (mistake H_1 for H_0)
- Minimize expected cost

Reject H_0 when

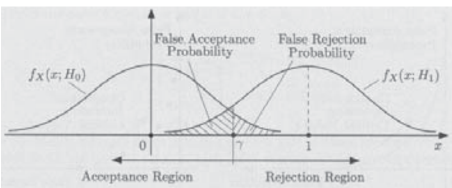
$$c_{\mathrm{II}}\frac{\mathbf{P}(H_1 \text{ true}) f_X(x\,;\,H_1)}{f_X(x)} > c_{\mathrm{I}}\frac{\mathbf{P}(H_0 \text{ true}) f_X(x\,;\,H_0)}{f_X(x)}$$

Equivalently, reject H_0 when

$$\underbrace{\frac{f_X(x\,;\,H_1)}{f_X(x\,;\,H_0)}}_{\text{likelihood ratio } L(x)} > \underbrace{\underbrace{\frac{c_{\mathrm{I}}}{c_{\mathrm{II}}}}_{\text{cost ratio}}\frac{\mathbf{P}(H_0 \text{ true})}{\mathbf{P}(H_1 \text{ true})}}_{\text{critical value}}$$

Rule: reject H_0 when likelihood ratio exceeds some value

Classical binary hypothesis testing



- A **likelihood ratio test** rejects H_0 when $L(x) > \xi$
  - $\xi = 1$ corresponds to ML rule
  - Relative costs of Type I and Type II errors could drive selection of ξ
  - Often minimize probability of one error type while meeting requirement on other

Neyman–Pearson lemma

- In principle, rejection region R could be any n-dim. set
- **Neyman–Pearson lemma**: likelihood ratio tests give all the best rejection regions
- Formally: Suppose for some ξ,

  $$\mathbf{P}(L(X) > \xi\,;\,H_0) = \alpha, \qquad \mathbf{P}(L(X) \le \xi\,;\,H_1) = \beta.$$

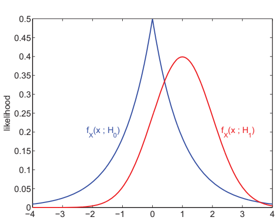
  Any test with rejection region R such that

  $$\mathbf{P}(X \in R\,;\,H_0) \le \alpha \quad \text{will have} \quad \mathbf{P}(X \notin R\,;\,H_1) \ge \beta.$$

- Huge simplification: always just one scalar ξ to vary

Example (one observation)

- Null hypothesis: $f_X(x\,;\,H_0) = \frac{1}{2}e^{-|x|}$
- Alternative hypothesis: $f_X(x\,;\,H_1) = \frac{1}{\sqrt{2\pi}}e^{-(x-1)^2/2}$



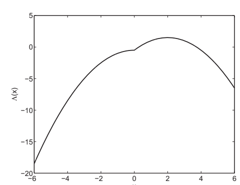
Since $f_X(x\,;\,H_1)$ decays faster than $f_X(x\,;\,H_0)$ when $|x| \to \infty$, not good to reject H_0 for very large $|x|$.

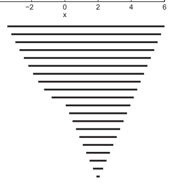
What rejection region $R = [a, b]$ is good?

$$L(x) = \frac{f_X(x\,;\,H_1)}{f_X(x\,;\,H_0)} = \sqrt{\tfrac{2}{\pi}}\exp(|x| - \tfrac{1}{2}(x-1)^2)$$

Example (one observation), continued

- Let $\Lambda(x) = |x| - \frac{1}{2}(x-1)^2$, so $L(x) = \sqrt{2/\pi}\,\exp(\Lambda(x))$

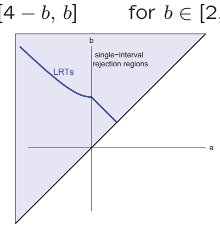




Both (!) endpoints of all good rejection regions are obtained by varying $\Lambda(x)$

$$[-\sqrt{b^2 - 4b},\, b] \quad \text{for } b \in [4, \infty)$$

$$[4 - b,\, b] \quad \text{for } b \in [2, 4]$$



Example (two observations)

- Null hypothesis: $f_X(x\,;\,H_0) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- Alternative hypothesis: $f_X(x\,;\,H_1) = \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2}$

What is the right way to use two samples to make a decision?

$$
\begin{aligned}
L(x) &= \frac{f_{X_1,X_2}(x_1,x_2\,;\,H_1)}{f_{X_1,X_2}(x_1,x_2\,;\,H_0)} = \frac{\exp(-\frac{1}{2}x_1^2)\,\exp(-\frac{1}{2}x_2^2)}{\exp(-\frac{1}{2}(x_1-1)^2)\,\exp(-\frac{1}{2}(x_2-1)^2)} \\
&= \frac{\exp(-\frac{1}{2}(x_1^2+x_2^2))}{\exp(-\frac{1}{2}((x_1^2-2x_1+1)+(x_2^2-2x_2+1)))} \\
&= \exp\left(-\tfrac{1}{2}(2(x_1+x_2)-2)\right)
\end{aligned}
$$

All (!) good tests depend on $x_1 + x_2$

(equivalently, depend on sample mean $\frac{1}{2}(x_1+x_2)$)

Final exam preparation

- Solutions to PS11 posted today
- Review session:
  Thursday, May 13, 7:30pm–9:30pm, 32-123
- Office hours

Su	8:00p–10:00p	Park	Student Center Reading Room
Mo	1:00p– 3:00p	Willsky	32-D582
	1:00p– 3:00p	Orji	24-308
	2:00p– 4:00p	Njoroge	32-D640
	4:00p– 6:00p	Megretski	32-D730
Tu	10:30a–11:30a	Goyal	36-428 (Haus)
	12:00p– 2:00p	Maymon	38-466 (Jackson)
	1:00p– 3:00p	Orji	24-312
	2:00p– 3:00p	Goyal	36-680
	4:30p– 6:30p	Dwivedi	24-312
	6:30p– 8:30p	Krishnaswamy	24-312

Parting words

Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

– Stephen Jay Gould

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.

– Pierre Simon Laplace

Probability does pervade the universe, and in this sense, the old chestnut about baseball imitating life really has validity. The statistics of streaks and slumps, properly understood, do teach an important lesson about epistemology, and life in general. The history of a species, or any natural phenomenon, that requires unbroken continuity in a world of trouble, works like a batting streak. All are games of a gambler playing with a limited stake against a house with infinite resources. The gambler must eventually go bust. His aim can only be to stick around as long as possible, to have some fun while he's at it, and, if he happens to be a moral agent as well, to worry about staying the course with honor!

– Stephen Jay Gould

Advice for the final exam

Chance favors the prepared mind.

– Louis Pasteur

Good luck!