Hypothesis testing and decision rules

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

Correction to use of z_α

- On lecture slides earlier in the semester, I used a definition of z_α that differs from that in the book. Let me call it z'_α,

$$\alpha = \int_{-\infty}^{z'_\alpha} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- The actual definition of z_α used by the book is, in fact,

$$\alpha = \int_{z_\alpha}^{+\infty} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- Note that $z_\alpha = z'_{1-\alpha} = -z'_\alpha$ since

$$1-\alpha = \int_{-\infty}^{+\infty} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) - \int_{z_\alpha}^{+\infty} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = \int_{-\infty}^{z_\alpha} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

- No credit was lost for mixing this up on homework
- I am slowly correcting the slides. We will use the book's definition henceforth.

Asymptotic unbiasedness

- For a sequence $\vec{Y}$ of n random variables
- Use $\hat{\sigma}^2_n(\vec{Y})$ to denote the MLE estimator for the variance of the normal.
- Recall $\hat{\sigma}^2_n(\vec{Y})$ is *not unbiased* since

$$E\left(\hat{\sigma}^2_n\right) = E\left[\frac{1}{n}\sum_{j=1}^{n}\left(Y_j - \overline{Y}\right)^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

- but it is *asymptotically unbiased* since $\lim_{n\to\infty} E\left(\hat{\sigma}^2_n\right) = \sigma^2$

- More generally, if $\hat{x}_n$ is a sequence of estimators for parameter x
  - If $E(\hat{x}_n) = x$ for all n, then $\hat{x}_n$ is unbiased.
  - If $\lim_{n\to\infty} E(\hat{x}_n) = x$, then $\hat{x}_n$ is asymptotically unbiased.

Consistency

- Consistency is a stronger condition
- For consistency, it is not enough that $\lim_{n\to\infty} E(\hat{x}_n) = x$
- For consistency, we demand that $\lim_{n\to\infty} \hat{x}_n = x$
- This means that not only must the mean of the estimator approach x, but the variance of the estimator must approach zero.
- Consistency implies asymptotic unbiasedness, but the reverse is not true.

Consistency (continued)

- For $y \in \mathbb{R}$, you can show that this pdf has mean μ and unit variance

$$p_Y(y;\mu) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|y-\mu|}$$

- Now consider a sequence of random variables $\vec{Y}$ where Y_n is sampled from $p\left(y; x + \frac{1}{n}\right)$
- For random variables $\vec{Y}$ generated in this way, can we find an estimator for x?
- Consider the estimator for x,

$$\hat{x}_n(\vec{Y}) = Y_n$$

- Note that $E(\hat{x}_n) = E(Y_n) = x + \frac{1}{n}$, so this is not unbiased, but it is asymptotically unbiased since

$$\lim_{n\to\infty} E(\hat{x}_n) = \lim_{n\to\infty} \left(x + \frac{1}{n}\right) = x.$$

Consistency (continued)

- We are considering a sequence of random variables $\vec{Y}$ where Y_n is sampled from $p\left(y; x + \frac{1}{n}\right)$, where in turn $p_Y(y;\mu) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|y-\mu|}$.
- We have shown the estimator for x given by $\hat{x}_n(\vec{Y}) = Y_n$ is *asymptotically unbiased*. But is it *consistent*?
- With elementary integration and a bit of work, you can show

$$P(|\hat{x}_n - x| < \epsilon) = \int_{x-\epsilon}^{x+\epsilon} dy\ p_Y(y;\mu) = \begin{cases} 1 - e^{-\sqrt{2}\epsilon} \cosh\left(\sqrt{2}/n\right) & \text{if } n\epsilon > 1 \\ e^{-\sqrt{2}/n} \sinh(\sqrt{2}\,\epsilon) & \text{otherwise} \end{cases}$$

- Fixing $\epsilon > 0$, however small, and letting $n \to \infty$, we find

$$\lim_{n\to\infty} P(|\hat{x}_n - x| < \epsilon) = 1 - e^{-\sqrt{2}\epsilon} \neq 1,$$

so $\hat{x}_n$ is *not consistent*, according to Definition 5.7.1 in the L&M text.

Example of Bayesian estimation

Bruce M. Boghosian

Loose ends

One more example of Bayesian estimation

Hypothesis Testing

Summary

- Exponentially distributed random variable W

$$f_W(w \mid \lambda) = \lambda e^{-\lambda w}$$

- Recall the mean is $E(w) = 1/\lambda$
- Suppose also that your prior for λ is the uniform distribution on $[a, b]$,

$$f_\Lambda(\lambda) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq \lambda \leq b \\ 0 & \text{otherwise} \end{cases}$$

- Suppose that you sample W and find a value w_s.

Example of Bayesian estimation (continued)

- Suppose that you sample W and find a value w_s.
- The posterior distribution of λ is

$$g_\Lambda(\lambda \mid W = w_s) = \frac{f_W(w_s \mid \lambda) f_\Lambda(\lambda)}{\int_{-\infty}^{+\infty} d\xi \; f_W(w_s \mid \xi) f_\Lambda(\xi)}$$

$$= \begin{cases} \dfrac{\lambda e^{-\lambda w_s} \frac{1}{b-a}}{\int_a^b d\xi \; \xi e^{-\xi w_s} \frac{1}{b-a}} & \text{if } a \le \lambda \le b \\ 0 & \text{otherwise} \end{cases}$$

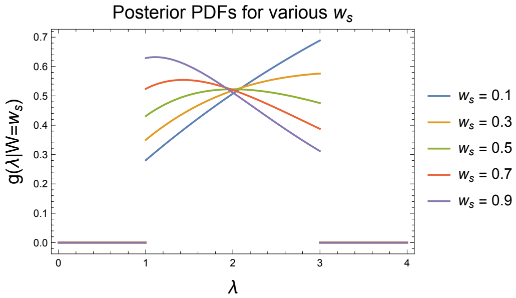
$$= \begin{cases} \dfrac{\lambda w^2 e^{-\lambda w_s}}{(1+aw_s)e^{-aw_s} - (1+bw_s)e^{-bw_s}} & \text{if } a \le \lambda \le b \\ 0 & \text{otherwise} \end{cases}$$

- What does this posterior distribution look like for various samples w_s?

Example of Bayesian estimation (continued)

- Suppose your prior for $f_\Lambda(\lambda)$ has $a = 1$ and $b = 3$
- The posterior distribution of λ is

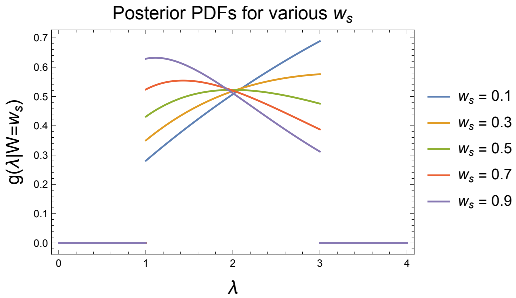
$$g_\Lambda(\lambda \mid W = w_s) = \begin{cases} \dfrac{\lambda w^2 e^{-\lambda w_s}}{(1+aw_s)e^{-aw_s} - (1+bw_s)e^{-bw_s}} & \text{if } a \le \lambda \le b \\ 0 & \text{otherwise} \end{cases}$$



A potential drawback of Bayesian estimation

- There is no way that the posterior $g_\Lambda(\lambda \mid W = w_s)$ can be nonzero anywhere outside of the region $[a, b]$, where the prior $f_\Lambda(\lambda)$ was nonzero.
- It is clear that this is generally true from the equation

$$g_\Lambda(\lambda \mid W = w_s) = \frac{f_W(w_s \mid \lambda) f_\Lambda(\lambda)}{\int_{-\infty}^{+\infty} d\xi \; f_W(w_s \mid \xi) f_\Lambda(\xi)}$$



The Decision Rule

Bruce M. Boghosian

- Imagine a *yes/no question* about random data.
  - Deciding whether or not a person on trial *is or is not* guilty
  - Testing mileage on cars to decide if a fuel additive *is or is not* effective
- In both examples, there is a *null hypothesis* H_0: This is the thing you would suppose to be true by default – in the absence of evidence to the contrary.
  - In a trial, the null hypothesis in the U.S. is "innocent until proven guilty"
  - In a fuel additive test, the null hypothesis is that the additive is ineffective.
- There is also an *alternative hypothesis* H_1: This might be the thing you are trying to use evidence to prove is true.
  - In a trial, the alternative hypothesis might be that the defendant is guilty.
  - In fuel additive test, alternative hypothesis is that additive is proven effective.

The fuel additive problem

- We have a fleet of n cars whose mileage is $\overline{y} = 25$ mpg.
- In test with the additive, we found mileages $y_1, \ldots, y_n$, with *sample mean*

$$\overline{y} = \frac{1}{n} \sum_{j=1}^{n} y_j = 26.3 \text{ mpg}$$

- Sample mean is *normally distributed* with *known variance* of $\sigma = 2.4$ mpg.
- The competing hypotheses are
  - H_0: $\mu = 25.0$ so the additive is ineffective
  - H_1: $\mu > 25.0$ so the additive is effective
- Note $\overline{y} < 25.0$ is certainly not grounds to reject H_0.
- Even values of $\overline{y}$ slightly greater than 25.0 might not contradict H_0.
- The question is "How large must $\overline{y}$ be before we stop believing H_0?"

The fuel additive problem (continued)

Bruce M. Boghosian

- How large must $\overline{y}$ be before we stop believing H_0?
- There is no right answer to this question.
- It depends how much confidence you want in your answer.
- Suppose we want 95% confidence. Then we seek y^* such that we can reject H_0 whenever $\overline{y} > y^*$, and have 95% confidence that our judgement is correct.

$$
\begin{aligned}
P(\text{We reject } H_0 \mid H_0 \text{ is true}) &= 0.05 \\
\therefore \quad P(\overline{y} > y^* \mid \mu = 25.0) &= 0.05 \\
\therefore \quad P(\overline{y} > y^* \mid \mu = 25.0) &= 0.05 \\
\therefore \quad P\left(\frac{\overline{y} - \mu}{2.4/\sqrt{30}} > \frac{y^* - \mu}{2.4/\sqrt{30}} \mid \mu = 25.0\right) &= 0.05 \\
\therefore \quad P\left(Z > \frac{y^* - \mu}{2.4/\sqrt{30}}\right) &= 0.05
\end{aligned}
$$

The fuel additive problem (continued)

- We can have the desired confidence if

$$P\left(Z > \frac{y^* - \mu}{2.4/\sqrt{30}}\right) = 0.05$$

$$\therefore \quad \frac{y^* - 25.0}{2.4/\sqrt{30}} = z_{0.05} = 1.64$$

$$\therefore \quad y^* = 25.0 + 1.64\frac{2.4}{\sqrt{30}} = 25.7186$$

- So we should reject the null hypothesis if $\overline{y} > 25.7186$.

Testing for equality

- Suppose we wish to test $H_0 : \ \mu = \mu_0$.
- This might have been $\mu_0 = 25$ in the previous example.
- We wish to have 95% confidence that we are right, so $\alpha = 0.05$.
- Suppose we accept H_0 if $\mu - \Delta y \leq \overline{y} \leq \mu + \Delta y$, and reject otherwise.
- Demand

$$\begin{aligned}
P(\text{We reject } H_0 \mid H_0 \text{ is true}) &= \alpha \\
\therefore \quad P(\overline{y} < \mu - \Delta y \mid \mu = \mu_0) + P(\overline{y} > \mu + \Delta y \mid \mu = \mu_0) &= \alpha \\
\therefore \quad P(\overline{y} > \mu + \Delta y \mid \mu = \mu_0) &= \alpha/2 \\
\therefore \quad P\left(\frac{\overline{y} - \mu}{\sigma/\sqrt{n}} > \frac{\Delta y}{\sigma/\sqrt{n}} \mid \mu = \mu_0\right) &= \alpha/2
\end{aligned}$$

- We know $Z = \frac{\overline{y}-\mu}{\sigma/\sqrt{n}}$ is distributed as a standard normal, so $\frac{\Delta y}{\sigma/\sqrt{n}} + z_{\alpha/2}$, or

$$\Delta y = +z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Summary of the above

- Easier to state in terms of $z := \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$
- Let $y_1, \ldots, y_n$ be a random sample from a normal distribution for which σ is known.
- To test $H_0: \ \mu = \mu_0$ versus $H_1: \ \mu > \mu_0$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.
- To test $H_0: \ \mu = \mu_0$ versus $H_1: \ \mu < \mu_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
- To test $H_0: \ \mu = \mu_0$ versus $H_1: \ \mu \neq \mu_0$ at the α level of significance, reject H_0 if either $z \leq -z_{\alpha/2}$ or $z \geq +z_{\alpha/2}$.

Caveats and pitfalls

- All of the above tests are designed to accept or reject the null hypothesis.
- The above analysis worked because assuming the null hypothesis H_0 means assuming that $\mu = \mu_0$. We could then subtract μ_0 across the inequality and divide by the variance to obtain a random variable distributed like a standard normal.
- We do not have a way to accept or reject the alternative hypothesis.
- If H_0 is not rejected, it is better to characterize that as "failure to reject the null hypothesis," rather than "accept the null hypothesis".

Summary

- We straightened out the definition of the z_n.
- We discussed the difference between asymptotic unbiasedness and consistency.
- We gave a concrete example of Bayesian parameter estimation.
- We discussed null and alternate hypotheses.
- We discussed hypothesis testing and decision rules.