Interval Estimation II and Properties of Estimators I

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

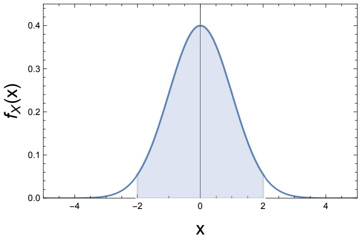
Notation for definite integrals of the standard normal

Bruce M. Boghosian

- Define z_α to be the number such that

$$\text{Prob}(Z < z_\alpha) = \int_{-\infty}^{z_\alpha} dz\, \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \alpha.$$

- Since the standard normal is symmetric about $x = 0$, we have $z_{1-\alpha} = -z_\alpha$.
  - For example, $z_{0.025} = -1.9599$ and $z_{0.975} = +1.9599$.



Interval estimation from last time

- Suppose that you have n measurements of experimental data.
- You have a priori knowledge that each datum is distributed normally.
- You know the variance $v_0 = \sigma_0^2$, but you do not know the mean μ.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left[-\frac{(x-\mu)^2}{2\sigma_0^2}\right]$$

- The MLE for the mean is the sample mean, $\mu_e = \frac{1}{n}\sum_{j=1}^n x_j$.
- By CLT, $Z = \frac{\mu_e - \mu}{\sigma_0/\sqrt{n}}$ is distributed like a standard normal for large n

Symmetric confidence intervals

- We know $Z = \frac{\mu_e - \mu}{\sigma_0/\sqrt{n}}$ is distributed like a standard normal
- Confidence $100(1-\alpha)\%$ that μ is in symmetric interval about μ_e
- Demand that $z_{\alpha/2} < Z < z_{1-\alpha/2}$, so we have

$$\begin{aligned}
\therefore \quad z_{\alpha/2} &< \frac{\mu_e - \mu}{\sigma_0/\sqrt{n}} < z_{1-\alpha/2} \\
\therefore \quad z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} &< \mu_e - \mu < z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} \\
\therefore \quad -z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} &< \mu - \mu_e < -z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \\
\therefore \quad -z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} &< \mu - \mu_e < z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}
\end{aligned}$$

$$\therefore \boxed{\mu \in \left[\mu_e - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}},\ \mu_e + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]}$$

Asymmetric confidence intervals

Bruce M. Boghosian

- Find the level of confidence that

$$\begin{aligned}
\mu_e - \zeta_- \frac{\sigma_0}{\sqrt{n}} < \mu < \mu_e + \zeta_+ \frac{\sigma_0}{\sqrt{n}} \\
-\zeta_- \frac{\sigma_0}{\sqrt{n}} < &\mu - \mu_e < +\zeta_+ \frac{\sigma_0}{\sqrt{n}} \\
-\zeta_+ \frac{\sigma_0}{\sqrt{n}} < &\mu_e - \mu < +\zeta_- \frac{\sigma_0}{\sqrt{n}} \\
z_{\alpha^-} := -\zeta_+ < &\frac{\mu_e - \mu}{\sigma_0/\sqrt{n}} < +\zeta_- =: z_{\alpha^+}
\end{aligned}$$

- Confidence is then $100(\alpha^+ - \alpha^-)\%$.
- Example: If $\zeta_- = \zeta_+ = 1.9599$, $\alpha_- = 0.025$ and $\alpha_+ = 0.975$, so confidence is $100(0.975 - 0.025)\% = 95\%$.

Confidence intervals for the binomial parameter p

- Suppose we conduct n Bernoulli trials with heads probability p.
- For one trial, the mean is p and the standard deviation is $\sqrt{p(1-p)}$
- For n trials, we have a binomial probability distribution with mean p and standard deviation $\sqrt{p(1-p)/n}$
- Using MLE or MM, we have $p_e = \frac{1}{n}\sum_{j=1}^{n} k_j$, so for large n

$$Z = \frac{p_e - p}{\sqrt{p_e(1-p_e)/n}}$$

will be distributed like a standard normal, by the Central Limit Theorem.

Margin of error

Bruce M. Boghosian

- The *margin of error* is half the maximum width of a confidence interval.
- Let k be the number of successes in n Bernoulli trials. Estimate is $p_e = k/n$.
- Confidence interval is $\left[p_e - z_{1-\alpha/2}\frac{\sigma_e}{\sqrt{n}},\ p_e + z_{1-\alpha/2}\frac{\sigma_e}{\sqrt{n}}\right]$
- Width of confidence interval is $2z_{1-\alpha/2}\frac{\sigma_e}{\sqrt{n}} = \frac{z_{1-\alpha/2}(4\sigma_e)}{2\sqrt{n}}$
- Estimate of standard deviation is $\sigma_e = p_e(1-p_e)$
- This is a problem, because we don't know in advance what p_e will be.
- The largest that $4\sigma_e = 4p_e(1-p_e)$ could be, however, is one.
- The margin of error is $100d\%$ where $d = \frac{z_{1-\alpha/2}}{2\sqrt{n}}$
- Usually $\alpha = 0.05$, but definition can be generalized to other values of α

Choosing sample sizes

- We have seen that the largest interval half width possible is $d = \frac{z_{1-\alpha/2}}{2\sqrt{n}}$
- We have in general

$$\text{Prob}\left(-d \leq \frac{1}{n}\sum_{j=1}^{n} x_j - p \leq +d\right) = 1 - \alpha.$$

- This can be regarded as an equation for the minimum value of n needed to attain the confidence α, and margin of error $100d\%$
- For fixed n, you can have more confidence in estimates with larger margins of error
- Likewise, you can have smaller margins of error, but you may have less confidence in those.

Comparing interval estimation for normal and binomial dists.

- In all examples with normal distributions, we *specified* the variance σ_0
- We found confidence intervals for the estimate of the mean

$$\mu_e = \frac{1}{n}\sum_{j=1}^{n} x_j$$

- Instead of insisting on prior knowledge of σ_0, why didn't we use the estimate,

$$\sigma_e = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(x_j - \mu_e\right)^2} \quad ?$$

- After all, for the binomial distribution, we had no hesitation about using both

$$p_e = \frac{1}{n}\sum_{j=1}^{n} k_j \quad \text{and} \quad \sigma_e = \sqrt{p_e(1-p_e)}.$$

An important distinction

- For the binomial distribution, the mean is p and the standard deviation is $\sqrt{p(1-p)}$. The latter is completely determined by the former.
- For the normal distribution, mean μ and standard deviation σ are two separately specifiable parameters, each with its own estimator.
- When we use an estimator to find μ_e from our n data points, we effectively "use up" a data point.
- When we use μ_e in the calculation of an average to obtain σ_e, our average is effectively over only $n-1$ points.
- For this reason, the *sample standard deviation* used for interval estimation for normally distributed data is not that given by the MLE (or MM) estimator.
- We will see what the correct expression is in the future. In the meantime, you can use σ_e for estimation in Homework 2, with this understanding.
- If n is very large, this makes very little difference.

Ambiguity associated with estimators

- In the example of the uniform distribution

$$f_X(x) = \left\{ \begin{array}{ll} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{array} \right\}$$

  - MLE estimator $\hat{a}(\vec{x}) = \min_j x_j$
  - MM estimator $\hat{a}(\vec{x}) = M_1 - \sqrt{3}\sqrt{M_2 - M_1^2}$, for sample moments M_1 and M_2
- Which one is "right"?
- There is no single answer to that question. We must instead identify desirable properties of estimators, and see which estimators have which of those properties.

Estimators are themselves random variables!

Bruce M. Boghosian

- Note that estimators such as $\hat{a}(\vec{x})$ are functions of random variables.
- Hence estimators are themselves random variables.
- They presumably have probability density functions, though it is not always easy to figure out what they are
- For estimating the sample mean $\hat{\mu}(\vec{x})$, we were able to use the CLT to study its distribution
- For other estimators, such an approach may not be possible
- Because they have density functions, however, we know certain things about them.
  - They have means. It is possible to speak of $E(\hat{a}(\vec{x}))$ and $E(\hat{\mu}(\vec{x}))$.
  - They have standard deviations. It is possible to speak of $\sigma_{\hat{a}}$ and $\sigma_{\hat{\mu}}$

What are some desirable properties of estimators?

Bruce M. Boghosian

- Because estimators are random variables with density functions, we know certain things about them:
  - They have means. It is possible to speak of $E(\hat{a}(\vec{x}))$ and $E(\hat{\mu}(\vec{x}))$.
  - They have standard deviations. It is possible to speak of $\sigma_{\hat{a}}$ and $\sigma_{\hat{\mu}}$
- And these lead to some ideas for desirable properties of estimators:
  - If data $\vec{x}$ is generated from density function $f_X(x;\theta)$ with parameter θ, it would be really nice if $E(\hat{\theta}(\vec{x}))=\theta$. This property is called *unbiasedness*.
  - It would be nice if $\sigma_{\hat{\theta}}$ were as small as possible. This is related to the property of *efficiency*. We will look at both relative and absolute notions of efficiency. The latter will lead to a bound on just how efficient an unbiased operator can be, called the *Cramér-Rao bound*.
- These topics will keep us busy for the remainder of this lecture and next lecture, at the very least.

Unbiasedness: Example PDF

- Consider the one-parameter probability density function

$$f_Y(y;\theta) = \begin{cases} \frac{2y}{\theta^2} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

- Normalization: $\int_0^\theta dy\, \frac{2y}{\theta^2} = 1$
- Mean: $\mu = \int_0^\theta dy\, \frac{2y}{\theta^2} y = \frac{2}{3}\theta$
- Mean square: $E(Y^2) = \int_0^\theta dy\, \frac{2y}{\theta^2} y^2 = \frac{1}{2}\theta^2$
- Variance: $\mathrm{Var}(Y) = \int_0^\theta dy\, \frac{2y}{\theta^2} (y-\mu)^2 = \frac{1}{18}\theta^2$
- Standard deviation: $\sigma_Y = \sqrt{\mathrm{Var}(Y)} = \frac{1}{3\sqrt{2}}\theta$

Unbiasedness: Method of moments

- Let $M_1 = \frac{1}{n}\sum_{j=1}^{n} y_j$ be the sample mean.
- Set the theoretical mean equal to the sample mean: $\frac{2}{3}\theta_e = M_1$
- Hence $\theta_e = \frac{3}{2}M_1$
- MM estimator is then

$$\hat{\theta}_{\text{mm}}(\vec{y}) = \frac{3}{2n}\sum_{j=1}^{n} y_j$$

- MM estimator is unbiased

$$E(\hat{\theta}_{\text{mm}}(\vec{y})) = \frac{3}{2n}\sum_{j=1}^{n} E(y_j) = \frac{3}{2n}\sum_{j=1}^{n}\frac{2}{3}\theta = \frac{3}{2n}n\left(\frac{2}{3}\theta\right) = \theta.$$

Unbiasedness: Maximum likelihood estimation

Bruce M. Boghosian

- If $\max_j y_j > \theta$, the likelihood will be zero
- So suppose that $\theta > \max_j y_j$
- Likelihood is $L(\vec{y};\theta) = \prod_{j=1}^{n} \left(\frac{2y_i}{\theta^2}\right)$
- This clearly increases as θ decreases, so the MLE estimator is

$$\hat{\theta}_{\mathrm{mle}}(\vec{y}) = \max_j y_j$$

- It may be shown that the MLE estimator is biased, since

$$E(\hat{\theta}_{\mathrm{mle}}(\vec{y})) = E(\max_j y_j) = \frac{2n}{2n+1}\theta$$

(proof to come later).

Unbiasedness: Maximum likelihood estimation

- It may be shown that the MLE estimator is biased, since

$$E(\hat{\theta}_{\text{mle}}(\vec{y})) = E(\max_j y_j) = \frac{2n}{2n+1}\theta$$

- It is *asymptotically unbiased*, since it approaches θ as $n \to \infty$.
- We can construct an unbiased version of the MLE estimator by defining

$$\hat{\theta}_3(\vec{y}) := \frac{2n+1}{2n} \max_j y_j.$$

- It is then clear that $E(\hat{\theta}_3(\vec{y})) = \theta$, as desired.

Summary

Bruce M. Boghosian

- We quickly reviewed interval estimation from last time.
- We reviewed z_α notation for indefinite integrals of the standard normal PDF.
- We learned both symmetric and asymmetric interval estimation
- We reviewed interval estimation for the binomial distribution.
- We learned about margin of error, and the tradeoff between it and confidence
- We learned about the distinction between estimated standard deviation and sample standard deviation
- We learned about the properties of *unbiasedness* and *efficiency* of estimators.