Quick Review of the Central Limit Theorem

Moment-Generating Functions and Derivation of Central Limit Theorem

Bruce M. Boghosian

Tufts University | School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

Motivation

- Point estimates (MLE or MM) yield a single result.
- There is no indication of how accurate that result is.
- Need way to quantify the level of uncertainty in the result.
- This is done by constructing a *confidence interval*.
- A confidence interval is an interval in which the parameter has a high probability of being found.
- For example, *95% confidence interval* for parameter p is an interval surrounding estimate, constructed such that actual p is in the interval with 95% probability (confidence).

Review of moment-generating functions I

- *Moment-generating function* for continuous r.v. X,

$$M_X(t) := E\left(e^{tX}\right) = \int dx\ f_X(x) e^{tx} = \int dx\ f_X(x) \sum_{j=0}^{\infty} \frac{t^j x^j}{j!} = \sum_{j=0}^{\infty} \frac{t^j}{j!} E\left(X^j\right)$$

$$= 1 + \frac{t}{1!}E(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \frac{t^4}{4!}E(X^4) + \cdots$$

- The above makes clear that

$$\begin{aligned} M_X(0) &= 1, \\ M'_X(0) &= E(X), \\ M''_X(0) &= E(X^2), \\ &\vdots \end{aligned}$$

$$\boxed{M_X^{(k)}(0) = E(X^k).}$$

Review of moment-generating functions II

Bruce M. Boghosian

- Suppose $Y = aX + b$ is a new r.v. linearly related to X,

$$M_Y(t) := E\left(e^{tY}\right) = E\left(e^{t(aX+b)}\right)$$
$$= E\left(e^{atX}e^{tb}\right) = e^{tb}E\left(e^{atX}\right) = e^{tb}M_X(at).$$

- Suppose X_1 and X_2 are uncorrelated, and $Y = X_1 + X_2$,

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(X_1+X_2)}\right)$$
$$= E\left(e^{tX_1}e^{tX_2}\right) = E\left(e^{tX_1}\right)E\left(e^{tX_2}\right) = M_{X_1}(t)M_{X_2}(t)$$

- The generalization if $Y = X_1 + X_2 + \cdots + X_n$ is then

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).$$

Bruce M. Boghosian

- Moment generating function of a standard normal Z

$$\begin{aligned}
M_Z(t) =& E\left(e^{tZ}\right) \\
=& \int dz \; \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) e^{tz} \\
=& \int dz \; \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2 - 2zt + t^2}{2} + \frac{t^2}{2}\right) \\
=& \exp\left(\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int dz \; \exp\left[-\frac{(z-t)^2}{2}\right] \\
=& e^{t^2/2}.
\end{aligned}$$

Outline of proof of Central Limit Theorem I

- Suppose $X_1, X_2, \ldots, X_n$ are *independent, identically distributed (iid)* r.v.s with mean μ and variance σ^2.
- Define the *standardized random variable* $S_j := \frac{X_j - \mu}{\sigma}$
- Note the S_j have zero mean and unit variance:

$$E(S_j) = E\left(\frac{X_j - \mu}{\sigma}\right) = \frac{1}{\sigma}\left[E(X_j) - \mu\right] = 0 \qquad \text{(zero mean)}$$

$$\mathrm{Var}(S_j) = \mathrm{Var}\left(\frac{X_j - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}\left(X_j\right) = \frac{\sigma^2}{\sigma^2} = 1 \qquad \text{(unit variance)}$$

Outline of proof of Central Limit Theorem II

- Define standardized r.v. for average $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$,

$$Z := \frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}\,\sigma} = \sum_{j=1}^{n} \frac{X_j - \mu}{\sqrt{n}\,\sigma} = \sum_{j=1}^{n} \frac{S_j}{\sqrt{n}}$$

Note *Z statistic* also has mean zero and unit variance,

$$E(Z) = \sum_{j=1}^{n} \frac{E(S_j)}{\sqrt{n}} = 0 \qquad \text{(zero mean)}$$

$$\text{Var}(Z) = \sum_{j=1}^{n} \text{Var}\left(\frac{S_j}{\sqrt{n}}\right) = \sum_{j=1}^{n} \frac{1}{n}\text{Var}\left(S_j\right) = \sum_{j=1}^{n} \frac{1}{n} = 1 \qquad \text{(unit variance)}$$

Outline of proof of Central Limit Theorem III

- We have $Z = \sum_{j=1}^{n} \frac{S_j}{\sqrt{n}}$, where $S_j := \frac{X_j - \mu}{\sigma}$
- The S_j are iid, with zero mean and unit variance.
- Let M_S be moment generating function of (all) S_j, so

$$\lim_{n\to\infty} M_Z = \left[M_S\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{t}{\sqrt{n}}0 + \frac{t^2}{2n}1 + \cdots\right]^n$$
$$= \lim_{n\to\infty}\left(1 + \frac{t^2}{2n} + \cdots\right)^n = e^{t^2/2}$$

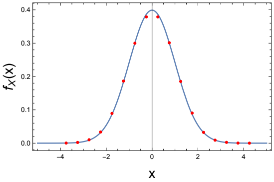
- Hence the Z statistic

$$Z := \frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{\sigma/\sqrt{n}} = \sum_{j=1}^{n} \frac{S_j}{\sqrt{n}}$$

must be distributed as a standard normal.

Example of the Central Limit Theorem in action

- Computer languages have random number generators
- Give uniformly distributed random numbers $X_j \in [0, 1]$.
- Each X_j has mean $1/2$ and standard deviation $\sigma = 1/(2\sqrt{3})$.
- Define $S_j := \frac{X_j - 1/2}{1/(2\sqrt{3})} = \sqrt{3}\,(2X_j - 1)$
- Define $Z := \frac{1}{\sqrt{n}}\,(S_1 + S_2 + \cdots + S_n)$ for some large n.
- Do this many times and histogram the results.
- For $n = 20$ and 10,000 histogrammed results:



- Red dots histogram peaks, blue curve standard normal.

Summary

- We have reviewed moment-generating functions.
- We have outlined the proof of the Central Limit Theorem.
- With these, we are prepared to study confidence intervals.