Mathematics 340: Lecture 31

Central Limit Theorem: Examples

Dan Sloughter

Furman University

December 4, 2015

DeMoivre-Laplace Theorem

- Suppose S_n is a binomial random variable with n trials and probability of success p.
- For $i = 1, 2, \ldots, n$, let X_i be 1 if the ith trial is a success, and 0 otherwise.
- Note:
    - $X_1, X_2, \ldots, X_n$ are independent, identically distributed Bernoulli random variables.
    - $S_n = X_1 + X_2 + \cdots + X_n$.
- Hence, by the central limit theorem,

$$\lim_{n\to\infty} P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx.$$

Notes

- This is the DeMoivre-Laplace theorem.
- The theorem says S_n is approximately $\mathcal{N}(np, np(1-p))$.
- How large n has to be for a good approximation depends on the size of p.
- One rule of thumb: the approximation is reasonable provided both $np \ge 5$ and $n(1-p) \ge 5$.

Correction for continuity

- If X is binomial, then for any integer k,

$$P(X = k) = P\left(k - \frac{1}{2} \le X \le k + \frac{1}{2}\right).$$

- Hence when we approximate binomial probabilities using the normal distribution it is helpful to make a *correction for continuity*.
- Namely, if X is binomial with n trials and probability of success p, then for any integers a and b we have

$$\begin{aligned} P(a \le X \le b) &= P\left(a - \frac{1}{2} \le X \le b + \frac{1}{2}\right) \\ &\approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

Example

- Suppose X is the number of sixes observed in 100 rolls of a fair die.
- Then, for example,

$$\begin{aligned} P(10 < X < 20) &= P(11 \le X \le 19) \\ &= P(10.5 \le X \le 19.5) \\ &\approx \Phi\left(\frac{19.5 - \frac{100}{6}}{\sqrt{100\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}}\right) - \Phi\left(\frac{10.5 - \frac{100}{6}}{\sqrt{100\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}}\right) \\ &= \Phi(0.76) - \Phi(-1.65) \\ &= 0.7764 - 0.0495 \\ &= 0.7269. \end{aligned}$$

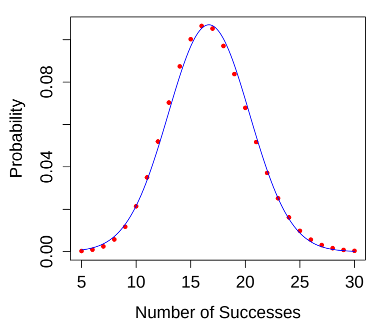
Example (cont'd)

- Note: the exact probability is

$$P(10 < X < 20) = P(11 \le X \le 19) = \sum_{k=11}^{19} \binom{100}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{100-k} = 0.7376.$$

Example (cont'd)

- Comparison of binomial mass function and approximating normal density:



Example

- An American roulette wheel has 38 pockets: 18 red pockets, 18 black pockets, and 2 green pockets (0 and 00).
- If a player bets one dollar on red, she wins a dollar if the ball lands in a red pocket and loses a dollar if the ball lands in any other pocket.
- Suppose repeated bets are placed and let X_i, $i = 1, 2, 3, \ldots, n$, be the amount won on the ith bet.
- Then the total amount won after n bets is given by

$$S_n = X_1 + X_2 + \cdots + X_n$$

- Note: for any i,

$$E(X_i) = -1 \times \frac{10}{19} + 1 \times \frac{9}{19} = -\frac{1}{19}$$

and

$$E(X_i^2) = 1 \times \frac{10}{19} + 1 \times \frac{9}{19} = 1.$$

Example (cont'd)

- Thus $\mathrm{Var}(X_i) = 1 - \frac{1}{361} = \frac{360}{361}$.
- So if σ_{X_i} is the standard deviation of X_i, $\sigma_{X_i} = \sqrt{\frac{360}{361}} = \frac{6\sqrt{10}}{19}$.
- Let σ_{S_n} be the standard deviation of S_n.
- Then, for example,

$$\mathrm{E}(S_{100}) = -\frac{100}{19} \text{ and } \sigma_{S_{100}} = 10 \times \frac{6\sqrt{10}}{19} = \frac{60\sqrt{10}}{19}.$$

- So

$$P(S_{100} < 0) = P\left(\frac{S_{100} + \frac{100}{19}}{\frac{60\sqrt{10}}{19}} < \frac{\frac{100}{19}}{\frac{60\sqrt{10}}{19}}\right) \approx \Phi\left(\frac{5}{3\sqrt{10}}\right) = \Phi(0.53) = 0.7019.$$

Example (cont'd)

- And

$$\mathrm{E}(S_{1000}) = -\frac{1000}{19} \text{ and } \sigma_{S_{1000}} = \sqrt{1000} \times \frac{6\sqrt{10}}{19} = \frac{600}{19}.$$

- So

$$P(S_{1000} < 0) = P\left(\frac{S_{1000} + \frac{1000}{19}}{\frac{600}{19}} < \frac{\frac{1000}{19}}{\frac{600}{19}}\right) \approx \Phi\left(\frac{5}{3}\right) = \Phi(1.67) = 0.9525$$

Example (cont'd)

- Also,

$$\mathrm{E}(S_{10000}) = -\frac{10000}{19} \text{ and } \sigma_{S_{10000}} = 100 \times \frac{6\sqrt{10}}{19} = \frac{600\sqrt{10}}{19}.$$

- So

$$P(S_{1000} < 0) = P\left(\frac{S_{1000} + \frac{10000}{19}}{\frac{600\sqrt{10}}{19}} < \frac{\frac{10000}{19}}{\frac{600\sqrt{10}}{19}}\right) \approx \Phi\left(\frac{5\sqrt{10}}{3}\right) = \Phi(5.27) \approx 1.$$

Example

- Let X_k be the outcome of the kth roll of a fair die, $k = 1, 2, \ldots, n$.
- Let

$$S_n = X_1 + X_2 + \cdots + X_n.$$

- Then $\mathrm{E}(S_n) = \frac{7n}{2}$ and $\mathrm{Var}(S_n) = \frac{35n}{12}$.
- For example, if $n = 100$, we have $\mathrm{E}(S_{100}) = 350$, $\mathrm{Var}(S_{100}) = \frac{875}{3}$.
- Then, for example,

$$\begin{aligned} P(320 \le S_{100} \le 380) &= P(319.5 \le S_n \le 380.5) \\ &= P\left(\frac{319.5 - 350}{\sqrt{\frac{875}{3}}} \le \frac{S_{100} - 350}{\sqrt{\frac{875}{3}}} \le \frac{380.5 - 350}{\sqrt{\frac{875}{3}}}\right) \\ &\approx \Phi(1.79) - \Phi(-1.79) \\ &= 0.9633 - 0.0367 = 0.9266. \end{aligned}$$

Example (cont'd)

- Equivalently, this says that

$$P(3.2 \leq \bar{X}_{100} \leq 3.8) \approx 0.9266,$$

where, as usual,

$$\bar{X}_n = \frac{S_n}{n}.$$

Example

- Suppose $X_1, X_2, \ldots, X_{50}$ are independent exponential random variables, each with mean 1000.
- That is, each X_i has parameter $\lambda = \frac{1}{1000}$.
- Let $S_{50} = X_1 + X_2 \cdots + X_{50}$ and

$$\bar{X}_{50} = \frac{S_{50}}{50}.$$

- Then

$$\begin{aligned} E(S_{50}) &= 50 \times 1000 = 50,000, \\ \text{Var}(S_{50}) &= 50 \times (1000)^2 = 50,000,000, \\ E(\bar{X}_{50}) &= 1000, \end{aligned}$$

and

$$\text{Var}(\bar{X}_{50}) = \frac{(50)(1000)^2}{50^2} = 20,000.$$

Example (cont'd)

- Hence, for example,

$$\begin{aligned} P(\bar{X}_{50} > 900) &= P\left(\frac{\bar{X}_{50} - 1000}{\sqrt{20,000}} > \frac{900 - 1000}{\sqrt{20,000}}\right) \\ &\approx 1 - \Phi(-0.71) \\ &= 1 - 0.2389 \\ &= 0.7611. \end{aligned}$$

- Note:
  - S_{50} is gamma with parameters $\alpha = 50$ and $\lambda = \frac{1}{1000}$.
  - The exact probability is

$$P(\bar{X}_{50} > 900) = P(S_{50} > 45000) = 0.7532.$$

Approximating a Poisson distribution

- Recall: The sum of independent Poisson random variables is again Poisson.
- In particular, if Y is Poisson with mean λ, then we may write

$$Y = X_1 + X_2 + \cdots + X_n,$$

where $X_1, X_2, \ldots, X_n$ are independent Poisson random variables each having mean $\frac{\lambda}{n}$.
- It follows that Y is approximately $N(\lambda, \lambda)$ if λ is sufficiently large.
- Rule of thumb: the approximation is reasonable when $\lambda \geq 5$.
- Equivalently,

$$\frac{Y - \lambda}{\sqrt{\lambda}}$$

is approximately $N(0, 1)$ for sufficiently large λ.

Example

- Suppose X is Poisson with mean 100.
- Then, for example,

$$P(X \le 90) = P(X \le 90.5) = P\left(\frac{X-100}{10} \le \frac{90.5-100}{10}\right) \approx \Phi(-0.95) = 0.1711.$$

- Note: exactly,

$$P(X \le 90) = \sum_{k=0}^{90} \frac{100^k e^{-100}}{k!} = 0.1714.$$

Example

- Suppose $X_1, X_2, \ldots, X_n$ represent the errors from n measurements.
- We will suppose that X_k, $k = 1, 2, \ldots, n$, are independent and uniformly distributed on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ (in appropriate units).
- Then $\mathrm{E}(X_k) = 0$ and $\mathrm{Var}(X_k) = \frac{1}{12}$.
- If

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

then

$$\mathrm{E}(\bar{X}_n) = 0 \text{ and } \mathrm{Var}(\bar{X}_n) = \frac{1}{12n}.$$

Example (cont'd)

- For example, if $n = 25$,

$$\mathrm{E}(\bar{X}_{25}) = 0 \text{ and } \mathrm{Var}(\bar{X}_{25}) = \frac{1}{300}.$$

- Then, for example,

$$P(-0.1 \le \bar{X}_{25} \le 0.1) = P(-0.1\sqrt{300} \le \sqrt{300}\bar{X}_{25} \le 0.1\sqrt{300}) \approx \Phi(1.73) - \Phi(-1.73) = 0.9582 - 0.0418 = 0.9164.$$

- Note: contrast this with the probability that an individual measurement errs by less than 0.1:

$$P(-0.1 \le X_k \le 0.1) = 0.2.$$