Laws of large numbers and the Central Limit Theorem (CLT) with examples

Marianna Bolla, Prof, DSc.
Institute of Mathematics, BME

October 14, 2024

- **Useful Inequalities**

  1. *Markov's Inequality:* Let X be a r.v. of finite first moment and taking on nonnegative values. Then
  $$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(X)}{c}, \qquad \forall c > 0.$$
  2. *Chebyshev's Inequality:* Let X be a r.v. with finite second moment. Then
  $$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}\,(X)}{\varepsilon^2}, \qquad \forall \varepsilon > 0.$$

- **Laws of Large Numbers, Central Limit Theorem**

  1. *Weak Law*: If $X_1, \dots, X_n$ are i.i.d. with finite $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}\,(X_i) = \sigma^2$, then $\overline{X}_n \to \mu$ with high probability (w.h.p.):
  $$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0, \text{ where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
  Equivalently,
  $$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \leq \varepsilon) = 1, \quad \forall \varepsilon > 0.$$
  Indeed, by the Chebyshev's Inequality
  $$\mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$
  2. *Strong Law*: If $X_1, \dots, X_n$ are i.i.d. with finite $\mathbb{E}(X_i) = \mu$, then $\overline{X}_n \to \mu$ almost surely:
  $$\mathbb{P}(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$
  3. **CLT**: If $X_1, \dots, X_n$ are i.i.d. with finite $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}\,(X_i) = \sigma^2$, then
  $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \to \mathcal{N}(0,1) \quad \text{in distribution (convergence of c.d.f.'s)}, \quad n \to \infty.$$

Equivalently,

$$\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \to \mathcal{N}(0,1) \quad \text{in distribution (convergence of c.d.f.'s)}, \quad n \to \infty,$$

i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} < x) = \Phi(x) \quad \forall x \in \mathbb{R}.$$

Applying it to $X_1, X_2, \dots$ i.i.d. Bernoulli with parameter p: for large n, $\sum_{i=1}^{n} X_i$ is approximately $\mathcal{N}(np, \sqrt{np(1-p)})$ (also called De Moivre–Laplace theorem) and $\overline{X}_n$ is approximately $\mathcal{N}(p, \sqrt{\frac{p(1-p)}{n}})$. Also, in this case, the weak low of large numbers boils down to

$$\mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}. \tag{1}$$

Exercises:

1. A fair die is rolled 100 times. Find the exact and the approximate probability that the number of '6' outcomes is between 10 and 20.

   Solution: let X be the number of the outcomes '6'. Since $X \sim \mathcal{B}_{100}\left(\frac{1}{6}\right)$, the exact probability is

   $$\mathbb{P}(10 \leq X \leq 20) = \sum_{k=10}^{20} \binom{100}{k} \left(\frac{1}{6}\right)^k \left(1 - \frac{5}{6}\right)^{100-k}.$$

   This formula is correct to be written in the exam. The approximate probability can be calculated by the De Moivre–Laplace theorem, where $n = 100$ is "large".

   $$X \sim \mathcal{N}\left(100\frac{1}{6}, \sqrt{100\frac{1}{6}\frac{5}{6}}\right)$$

   which is,

   $$\approx \mathcal{N}\left(\frac{50}{3}, \frac{\sqrt{5}10}{6}\right).$$

   Therefore, the probability is

   $$\mathbb{P}(10 \leq X \leq 20) = \mathbb{P}(\frac{10 - \frac{50}{3}}{\sqrt{5}\frac{10}{6}} \leq \frac{X - \frac{50}{3}}{\sqrt{5}\frac{10}{6}} \leq \frac{20 - \frac{50}{3}}{\sqrt{5}\frac{10}{6}})$$

   By using the standard normal distribution function, this equals to

   $$\Phi(\frac{2}{\sqrt{5}}) - \Phi(-\frac{4}{\sqrt{5}}) = \Phi(\frac{2}{\sqrt{5}}) + \Phi(\frac{4}{\sqrt{5}}) - 1 = 0.777$$

2. There are 300 parking permits and each permit holder comes to the university with probability 0.7, independently of the others. How many parking lots (l) to establish so that

   $$\mathbb{P}(\text{someone with permit cannot find a place to park}) = 0.01.$$

Solution: Let X be the number of permit holders coming in. Where $n = 300$ and $p = 0.7$, therefore, $X \sim \mathcal{B}_{300}(0.7)$. Then by the Central limit Theorem, $X \sim \mathcal{N}(210, \sqrt{63})$, and so,

$$\mathbb{P}(X > l) = 0.01$$

and

$$\mathbb{P}(X \leq l) = 0.99.$$

But

$$\mathbb{P}\left(\frac{X - 210}{\sqrt{63}} \leq \frac{l - 210}{\sqrt{63}}\right) = 0.99$$

$$\Phi\left(\frac{l - 210}{\sqrt{62}}\right) = 0.99$$

From the standard normal table, the value that corresponds to the probability 0.99 is 2.33. Hence,

$$\frac{l - 210}{\sqrt{63}} = 2.33$$

and $l = 2.33 \cdot \sqrt{63} + 210 \approx 229$.

3. 1000 persons arrive to the left or right entrance of a theater independently (they choose between the entrances with 0.5-0.5 probability). How many hangers (h)to place into the left and right cloak rooms, if they want to give only a 1 percent chance to the event that somebody cannot place his/her coat in the nearest cloakroom. (Each person has a coat.)

   Solution: let X be the number of people who arrive to the left. Since $X \sim \mathcal{B}_{1000}(0.5)$, by using the Central Limit Theorem, $X \sim \mathcal{N}(500, \sqrt{250} = 5 \cdot \sqrt{10})$.

$$\mathbb{P}(X > h \text{ or } 1000 - X > h) = .05$$

   Using the complementary event,

$$\mathbb{P}(1000 - h \leq X \leq h) = \mathbb{P}(\frac{-h - 500}{5\sqrt{10}} \leq \frac{X - 500}{5\sqrt{10}} \leq \frac{h - 500}{5\sqrt{10}}) = 0.95$$

   and

$$\Phi\left(\frac{h - 500}{5\sqrt{10}}\right) - \left(1 - \Phi\left(\frac{h - 500}{5\sqrt{10}}\right)\right) = .95$$

   By solving this equation, $h = 531$.

4. *Elections 2000.* In a state (call it Florida) the voters vote for two candidates randomly (with 0.5-0.5 probability), independently of each other. If there are 5 million voters, what is the probability that the difference of the votes given for the two candidates is less than 300 in absolute value?

5. *Overbooking.* There are 300 seats on an airplane, but the airline company gives out more than 300 reservations as (because of connected flights and other issues) passengers may not show up in time with probability 0.2 (say, independently). How many reservations can be made if the company wants to make the probability that a passenger with valid reservation cannot get a boarding pass (overbooking) be 0.01?

6. *Opinion poll.* Find the number n of persons to be interviewed (independently) so that the true (but unknown) population support p of a candidate and its relative frequency based on this poll differ at most 0.01 with probability at least 90 percent? (Hint: use formula (1).)