Week 8 Stats 25-26

Srikanth Pai, Asst. Prof, MSE.

September 14, 2025

1 Homework for the week

1. Larsen and Marx 3.10.1, 3.10.12, 3.10.13,3.10.14,3.10.15.
2. Central limit theorem, Larsen and Marx: Example 4.3.2,4.3.4.

   Exercise 4.3.15,4.3.17,4.3.24,4.3.34.
3. This exercise compares the estimate of probabilities given by Chebyshev with the estimate from central limit theorem.

   Let $X_1, \ldots, X_N$ be i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$. Define the sample mean

   $$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

   (a) Use Chebyshev's inequality to show that

   $$\mathbb{P}\big(|\bar{X}_N - \mu| \geq \varepsilon\big) \;\leq\; \frac{\sigma^2}{N\varepsilon^2}.$$

   [Note: Done in class]

   (b) Now assume X_i have a distribution with finite third moment so that the Central Limit Theorem applies. Deduce that for large N,

   $$\mathbb{P}\big(|\bar{X}_N - \mu| \leq \varepsilon\big) \;\approx\; 2\Phi\left(\frac{\sqrt{N}\,\varepsilon}{\sigma}\right) - 1,$$

   where Φ is the standard normal cdf.

   (c) Compare the bounds in (a) and (b). Which estimate is sharper? Illustrate numerically for the case $X_i \sim$ Bernoulli(1/2), $N = 100$, $\varepsilon = 0.1$. [You can use a table, calculator, or a computer to compare.]

   Which estimate is better?

4. (Priyanka's estimator) Let $p \in (0,1)$ be the (unknown) probability of heads for a coin. Toss the coin N times and let

$$\hat{p}_N = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

where $X_1, \ldots, X_N \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. Let $z_{0.995}$ denote the 0.995-quantile of the standard normal distribution.

Prove, using the Central Limit Theorem, that asymptotically

$$\mathbb{P}\left(\hat{p}_N - z_{0.995}\sqrt{\tfrac{p(1-p)}{N}} \;\le\; p \;\le\; \hat{p}_N + z_{0.995}\sqrt{\tfrac{p(1-p)}{N}}\right) \;\approx\; 0.99.$$

In other words, show that with probability close to 99%, the true value p lies in the interval

$$\left[\hat{p}_N - z_{0.995}\sqrt{\tfrac{p(1-p)}{N}},\; \hat{p}_N + z_{0.995}\sqrt{\tfrac{p(1-p)}{N}}\right].$$

2 Bonus questions

In the last question in the previous section we saw that the Central Limit Theorem implies an *approximate* 99% bound

$$\mathbb{P}\left(\hat{p}_N - z_{0.995}\sqrt{\tfrac{p(1-p)}{N}} \;\le\; p \;\le\; \hat{p}_N + z_{0.995}\sqrt{\tfrac{p(1-p)}{N}}\right) \;\approx\; 0.99.$$

But the "approximate" value is vague and unless we have an estimate of the error as a function of N, we cannot be sure of our estimates. This problem is solved by **Berry-Esseen theorem:** For i.i.d. $X_1, \ldots, X_N$ with mean μ, variance σ^2, and third absolute moment $\rho = \mathbb{E}[|X_1 - \mu|^3]$, one has

$$\max_{x\in\mathbb{R}} \left|\mathbb{P}\left(\tfrac{S_N - N\mu}{\sigma\sqrt{N}} \le x\right) - \Phi(x)\right| \;\le\; C\frac{\rho}{\sigma^3\sqrt{N}},$$

for an absolute constant $C < 0.5$, where Φ is the standard normal cdf. We will not prove this result but specialize this bound to the Bernoulli(p) case. Show the following step by step.

1. Show that $\mu = p$ and $\sigma^2 = p(1-p)$.
2. Compute $\rho = \mathbb{E}[|X_1 - p|^3] = p(1-p)\big(p^2 + (1-p)^2\big)$.
3. Deduce that
$$\frac{\rho}{\sigma^3} = \frac{p^2 + (1-p)^2}{\sqrt{p(1-p)}}.$$
4. Conclude that the CLT-approximation error in the 99% interval is at most
$$2C\,\frac{p^2 + (1-p)^2}{\sqrt{p(1-p)}}\;\frac{1}{\sqrt{N}}.$$

For our case, if $p = 0.5$, then the error is less than $1\sqrt{N}$.For 10k samples it may add or subtract a maximum probability of 1%.