Introduction to Probability Theory
13th week: The Law of Large Numbers and the Central Limit Theorem.

Theorem (**Law of Large Numbers**). *Let $X_1, X_2, \dots,$ denote the sequence of i.i.d. random variables, such that $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2 < \infty$. Moreover, let $S_n = X_1 + \cdots + X_n$. Then, for every constant $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0.$$

Definition. *We say that a continuous random variable X has a* **normal distribution** $N(\mu, \sigma^2)$ *if the density of X is given by*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}),$$

for every $x \in \mathbb{R}$. Then $\mathbb{E}X = \mu$ and $\operatorname{Var} X = \sigma^2$.

If $X \sim N(0,1)$ we say that X has a **standard normal distribution**. *The CDF of such X is usually denoted by $\Phi(\cdot)$.*

Remark. Unfortunately, Φ cannot be expressed by elementary functions, so its values should be found in either in the tables, or using an applet such as
`http://onlinestatbook.com/2/normal_distribution/standard_normal.html`
The most commonly used values of Φ are the following: $\Phi(1.2816) \sim 0.9$, $\Phi(1.6449) \sim 0.95$, $\Phi(1.9600) \sim 0.975$, $\Phi(2.3263) \sim 0.99$, $\Phi(2.5758) \sim 0.995$.

Theorem (**Central Limit Theorem**). *Let $X_1, X_2, \dots,$ denote the sequence of i.i.d. random variables, such that $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2 < \infty$ and $\mathbb{E}|X_1|^3 < \infty$. Moreover, let $S_n = X_1 + \cdots + X_n$. Then for every constant $x \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx = \Phi(x).$$

The CLT approximation can be used when both n and $\mathbb{E}S_n$ are large; typically it gives a reasonable estimates when $n \ge 30$ and $\mathbb{E}S_n \ge 10$. Moreover, one should remember then when we used CLT to approximate an integer-valued random variable then we should apply $1/2$ correction term. For instance, if we want to approximate $\mathbb{P}(S_n = k)$, then we should replace this probability by $\mathbb{P}(k - 1/2 < S_n \le k + 1/2)$, which, in turn, we estimate by

$$\begin{aligned}\mathbb{P}(k - 1/2 < S_n \le k + 1/2) &= \mathbb{P}\left(\frac{k - n\mu - 1/2}{\sigma\sqrt{n}} < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{k - n\mu + 1/2}{\sigma\sqrt{n}}\right) \\ &= \Phi\Big(\frac{k - n\mu + 1/2}{\sigma\sqrt{n}}\Big) - \Phi\Big(\frac{k - n\mu + 1/2}{\sigma\sqrt{n}}\Big).\end{aligned}$$

Homework assignment for Quiz 10, May 30th.

Important! For the quiz you might need a simple calculator, as the one you can find in mobile phones.

Problem 1. We toss a symmetric die 100 times. Let S_{100} denote the number of times we get 6. Use a calculator and a table to:
(i) find $\mathbb{P}(S_{100} = 15)$;
(ii) estimate $\mathbb{P}(S_{100} = 15)$ using the CLT.

Problem 2. We flip a coin 400 times. Let S_{400} denote the number of times we get head.
(i) Use the CLT (and the statistical tables) to estimate $\mathbb{P}(205 < S_{400} \le 222)$.
(ii) Using the CLT to find k such that $\mathbb{P}(S_{400} \le k) \sim 0.95$.

Problem 3. How many times do we need to toss a symmetric die so that the probability that we get 6 more than 20% of time is smaller than 0.1? Estimate it using: (i) Chebyshev's inequality, (ii) the CLT.

Problem 4. The time T Mr. January waits for the bus (measured in minutes) in a given day is exponentially distributed with parameter 0.1, i.e. T has the density

$$f(x) = \begin{cases} 0.1e^{-x/10} & \text{for } x \geq 0 \\ 0 & \text{for } x \leq 0. \end{cases}$$

Find the probability that in the last 100 days Mr. January waited for his bus less than 1200 minutes altogether (i.e. 12 minutes in average).

Remark You can find $\mathbb{E}T$ and $\operatorname{Var} T$ in the notes for the 10th week.

Problems we shall do in class on May 30th

Problem 1. The diameter of a sequoia tree grows each year, and each growth from 1 to 2 cm is equally likely. Estimate the probability that in the next 300 years the diameter of the tree will increase by at least 4,6 meters using: (i) Markov's inequality (ii) Chebyshev's inequality (iii) the Central Limit Theorem.

Problem 2. The head of a post office declares that the average time a customer waits for a service is shorter than 15 minutes. A TV crew wants to challenge her claim and measure this time for 50 customers. What is the chance that the average time they will measure is shorter than 15 minutes if the real time T (measured in minutes) a customer is waiting for her/his turn in this post office is a continuous non-negative random variable such that $\mathbb{E}T = 14$ and $\operatorname{Var} T = 23$? Suppose that the crew asks the head of the office how many customers they should examined. What should she answer them to be 99% sure that the average time they will measure is indeed shorter than 15 minutes?

Problem 3. We toss a die 100 times. Estimate the probability that we get 6 fewer than 10 times using: (i) Chebyshev's inequality (ii) the Central Limit Theorem.