Dr. Z.'s Probability Lecture 23 Handout:
The Central Limit Theorem; The Strong Law of Large Numbers

By Doron Zeilberger

Very Important Theorem (The Central Limit Theorem):

Let $X_1, \dots, X_n$ be a sequence of **independent** and **identically distributed** random variables, each with mean μ and variance σ^2, then the distribution

$$\frac{X_1 + \ldots + X_n \, - \, n\mu}{\sigma \sqrt{n}} \quad ,$$

tends to the **standard normal distribution** as n goes to infinity.

Note: Another way of saying it is that the **sample average** $(X_1 \, + \, \ldots \, + \, X_n)/n$, for large n, may be approximated by a Normal distribution with parameters $(\mu, \frac{\sigma^2}{n})$ (i.e. mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. It is an **amazing** theorem, since it is so general. The individual X_i can be **anything** (with finite variance). It also shows that the *larger* n is, the *smaller* is the standard deviation of the sample mean, i.e. the probability that the sample average would be far from the mean gets smaller and smaller.

Frequently Used Abbreviation: iid means "independent, identically distributed".

Problem 23.1: A seamstress wants to measure the length of a dress. Each measurement is independent of the others, and have a common mean d inches and common standard-deviation of 0.1 inches. She estimates the length of the dress by taking the average of all the measurements. How many measurements does she have to make in order to make %97 sure that her estimate is accurate to within ± 0.01inches?

(i) If she trusts the Central Limit Theorem?

(ii): If she uses the more conservative Chebyshev inequality?

Sol. to 23.1: If she makes n measurements, the **sample mean**, $\bar{X}$, has mean d, and standard deviation $\frac{\sigma}{\sqrt{n}}$ The seamstress wants

$$P\{|\bar{X} - d| \leq 0.01\} \geq 0.97 \quad .$$

Dividing by $\frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{n}}$, this is equivalent to

$$P\{|\frac{\bar{X} - d}{\frac{0.1}{\sqrt{n}}}| \leq \frac{0.01}{\frac{0.1}{\sqrt{n}}}\} \geq 0.97 \quad .$$

Since $\frac{\bar{X}-d}{\frac{0.1}{\sqrt{n}}}$is roughly the standard normal distribution, Z, she needs

$$P\{\,|Z| \leq 0.1\sqrt{n}\,\} \,\geq\, 0.97 \quad .$$

But $P\{|Z| \leq a\} = 2\,\Phi(a) - 1$, so she needs

$$2\,\Phi(0.1\sqrt{n}) - 1 \geq 0.97 \quad ,$$

which is the same as

$$\Phi(0.1\sqrt{n}) \geq 0.985 \quad .$$

From the Φ table, $\Phi(2.17) = 0.985$, so she needs

$$0.1\sqrt{n} \geq 2.17 \quad ,$$

that implies

$$n \geq (21.7)^2 = 470.89 \quad .$$

Ans. to 23.1(i): If the seamstress trusts the Central Limit Theorem, she should take (at least) 471 measurements.

For part **(ii)**, we have thanks to Chebyshev,

$$P\{|\bar{X} - d| \geq 0.01\} \leq \frac{\sigma^2/n}{0.01^2} = \frac{(0.1)^2/n}{0.01^2} = \frac{100}{n} \quad .$$

The seamstress wants the left side to be ≤ 0.03 so she needs

$$\frac{100}{n} \leq 0.03 \quad ,$$

hence

$$n \geq \frac{100}{0.03} = 3333.3333 \quad .$$

Ans. to 23.1(ii): If the seamstress wants to be absolutely sure that the probability of the sample average deviating from the true value by ± 0.01 inches be at least 0.97, she should take (at least) 3334 measurements.

Problem 23.2: A loaded tetrahedral die has four faces marked with one, three, five, and seven dots. The probability that it lands on the "one-dot face" is 0.4, the probability that it lands on the "three-dots face" is 0.3, the probability that it lands on the "five-dot" face is 0.2 and the probability that it lands on the "seven-dots" face is 0.1.

If this loaded tetrahedral die is rolled 64 times, find the approximate probability that the average number of dots is between 2.7 and 3.4.

Sol. to 23.2: The **first** step is to find μ and σ^2 for a single throw.

$$\mu = 0.4 \cdot 1 + 0.3 \cdot 3 + 0.2 \cdot 5 + 0.1 \cdot 7 = 3 \quad .$$

$$\sigma^2 = 0.4 \cdot (1-3)^2 + 0.3 \cdot (3-3)^2 + 0.2 \cdot (5-3)^2 + 0.1 \cdot (7-3)^2 = 4 \quad .$$

let $\bar{X}$ be the sample mean, then it is approximately normal with mean 3 and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{64}} = \frac{2}{8} = \frac{1}{4} = 0.25$.

We need

$$Pr\{\, 2.7 \le \bar{X} \le 3.4 \,\} \;=\; Pr\{\, -0.3 \le \bar{X} - 3 \le 0.4 \,\} \;=\; Pr\{\, -\frac{0.3}{0.25} \le \frac{\bar{X}-3}{0.25} \le \frac{0.4}{0.25} \,\} \;=$$

$$=\; Pr\{\, -1.2 \le \frac{\bar{X}-3}{0.25} \le 1.6 \,\} \;=\; \Phi(1.6) - \Phi(-1.2) \;=\; 0.9452 - 0.1151 \;=\; 0.8301 \quad .$$

Ans. to 23.2: The approximate probability that the average number of dots is between 2.7 and 3.4 is %83.01.

Note: According to Maple, the exact probability is 0.8222906620, not bad, but not great.

Problem 23.3: Let X_i , $i = 1, \ldots, 81$, be independent, identically distributed, continuous random variables each with probability density function

$$f(x) \;=\; \begin{cases} 2x, & if \quad 0 < x < 1 \quad ; \\ 0, & otherwise. \end{cases}$$

Calculate an approximation to

$$P\left\{ 50 \;\le \sum_{i=1}^{81} X_i \le 56 \right\} \quad .$$

Sol. to 23.3: We first find μ and σ^2 (and hence σ) for the individual X_i.

$$\mu \;=\; E[X_i] \;=\; \int_0^1 x \cdot 2x \, dx \;=\; \int_0^1 2\,x^2 \, dx = \frac{2}{3} \quad ,$$

$$E[X_i^2] \;=\; \int_0^1 x^2 \cdot 2x \, dx \;=\; \int_0^1 2\,x^3 \, dx \;=\; \frac{x^4}{2}\Big|_0^1 = \frac{1}{2} \quad .$$

Hence

$$\sigma^2 = E[X_i^2] \;-\; E[X_i]^2 = \frac{1}{2} - (\frac{2}{3})^2 \;=\; \frac{1}{18} \quad .$$

So $\sigma = \frac{\sqrt{2}}{6}$.

By the Central Limit Theorem $\sum_{i=1}^{81} X_i$ is approximately a normal distribution with mean $n\mu = 81 \cdot \frac{2}{3} = 54$ and standard deviation $\sigma\sqrt{n} = \frac{\sqrt{2}}{6} \cdot \sqrt{81} = \frac{3\sqrt{2}}{2}$.

We have, as usual

$$P\left\{ 50 \;\le \sum_{i=1}^{81} X_i \le 56 \right\} \;=\; P\left\{ 50 - 54 \;\le (\sum_{i=1}^{81} X_i) - 54 \le 56 - 54 \right\} \;=\; P\left\{ -4 \;\le (\sum_{i=1}^{81} X_i) - 54 \le 2 \right\} \;=$$

$$= P\left\{-\frac{4}{\frac{3\sqrt{2}}{2}} \le \frac{(\sum_{i=1}^{81} X_i) - 54}{\frac{3\sqrt{2}}{2}} \le \frac{2}{\frac{3\sqrt{2}}{2}}\right\}$$

This is approximately (by the Central Limit Theorem)

$$= P\{-1.885618082 \le Z \le 0.9428090414\} = \Phi(0.94) - \Phi(-1.88) = 0.8264 - 0.0301 = 0.7954 \quad .$$

Ans. to 23.3: $P\left\{50 \le \sum_{i=1}^{81} X_i \le 56\right\}$ is approximately 0.7954.

Famous Theorem (The Strong Law of Large Numbers):

Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed (aka iid) random variables, each having finite mean μ. Then with probability 1,

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \quad as \quad n \to \infty \quad .$$

Comment: This theorem, while famous, is only of theoretical interest. It is only *qualitative* not *quantitative*.