CSCI-B609: A Theorist's Toolkit, Fall 2016 Aug 23

Lecture 01: the Central Limit Theorem

Lecturer: Yuan Zhou *Scribe: Yuan Xie & Yuan Zhou*

1 Central Limit Theorem for i.i.d. random variables

Let us say that we want to analyze the total sum of a certain kind of result in a series of repeated independent random experiments each of which has a well-defined expected value and finite variance. In other words, a certain kind of result (e.g. whether the experiment is a "success") has some probability to be produced in each experiment. We would like to repeat the experiment many times independently and understand the total sum of the results.

1.1 Bernoulli variables

We first consider the sum of a bunch of Bernoulli variables.

Specifically, let X_1, X_2, ..., X_n be i.i.d. random variables with

$$\Pr[X_i = 1] = p, \qquad \Pr[X_i = 0] = 1 - p.$$

Let $S = S_n = X_1 + X_2 + ... + X_n$ and we want to understand S.

According to the linearity of expectation, we have

$$\mathrm{E}[S] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + ... + \mathrm{E}[X_n] = np.$$

Since $X_1, X_2, ...X_n$ are independent, we have $\mathrm{Var}[S] = np(1-p)$.

Now let us use a linear transformation to make S mean 0 and variance 1. I.e. let us introduce Z_n, a linear function of S_n, to be

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Using $\mu = np$ and $\sigma = \sqrt{np(1-p)}$, we have

$$Z_n = \frac{S_n - \mu}{\sigma}.$$

Via this transformation, we do not lose any information about $S = S_n$. Specifically, for any u, we have

$$\Pr[S_n \leq u] = \Pr[\sigma Z_n + \mu \leq u] = \Pr\left[Z_n \leq \frac{u-\mu}{\sigma}\right].$$

Therefore, we proceed to study the distribution of Z_n.

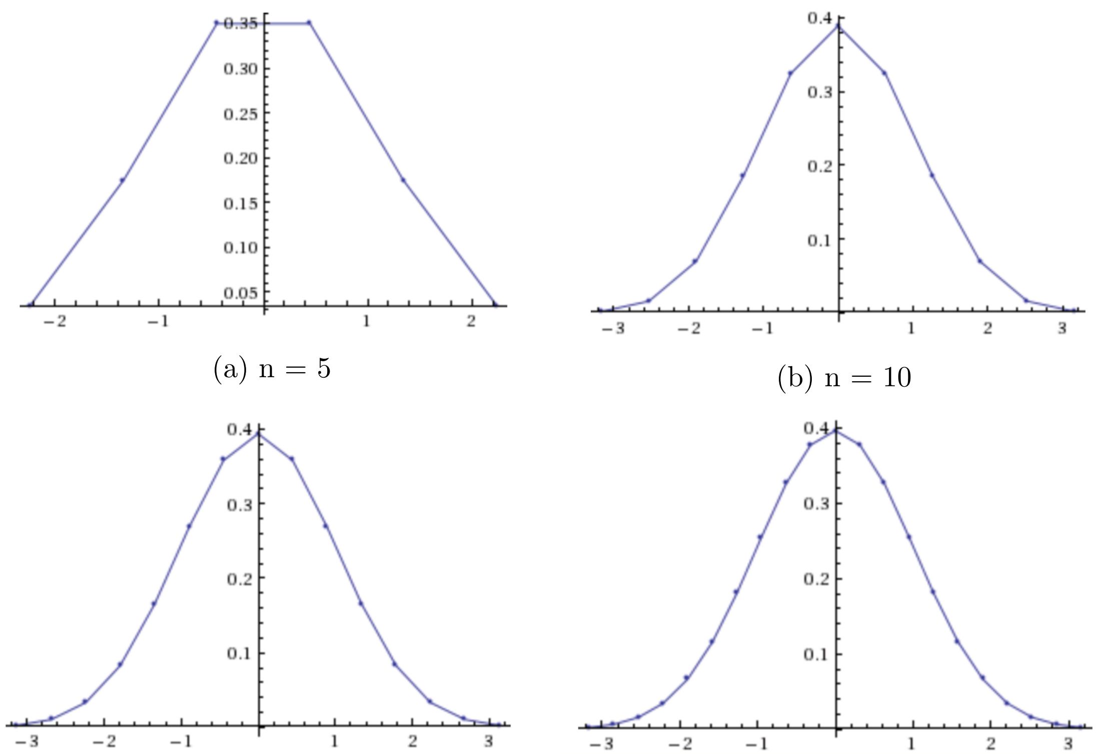
As a special instance, let us temporarily set $p = \frac{1}{2}$ so that X_i's become unbiased coin flips. In such case, we have

$$Z_n = \frac{X_1 + X_2 + ... + X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{1}{\sqrt{n}}((2X_1 - 1) + (2X_2 - 1) + ... + (2X_n - 1)).$$

For each integer $a \in [0, n]$, we have

$$\Pr\left[Z_n = \frac{2a-n}{\sqrt{n}}\right] = \frac{\binom{n}{a}}{2^n}.$$

Therefore, we can easily plot the probability density curve of Z_n. In Figure 1, we plot the density curve for a few values of n.



(a) n = 5

(b) n = 10

(c) n = 20

(d) n = 40

Figure 1: Probability density curves of Z_n for a few values of n

We can see that as $n \to \infty$, the probability density curve converges to a fixed continuous curve as illustrated in Figure 2.

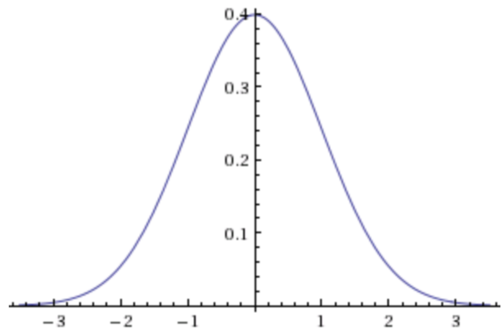


Figure 2: The famous "Bell curve" – the probability density function of a standard Gaussian variable

Indeed, even when $p = \Pr[X_i = 1]$ is a constant in $(0, 1)$ other than $\frac{1}{2}$, the probability density curve of Z_n still converges to the same curve as $n \to \infty$. We call the probability distribution using such curve as pdf the *Gaussian distribution* (or *Normal distribution*).

1.2 The Central Limit Theorem

The Central Limit Theorem (CLT) for i.i.d. random variables can be stated as follows.

Theorem 1 (the Central Limit Theorem)**.** *Let Z be a standard Gaussian. For any i.i.d $X_1, X_2, ..., X_n$ (not necessarily binary valued), as $n \to \infty$, we have $Z_n \to Z$ in the sense that $\forall u \in \mathbb{R}, \Pr[Z_n \le u] \to \Pr[Z \le u]$.*

More specifically, for each $\epsilon > 0$, there exists $N \in \mathbb{N}$ so that for every $n > N$ and every $u \in \mathbb{R}$, we have

$$|\Pr[Z_n \le u] - \Pr[Z \le u]| < \epsilon.$$

Definition 2. *We use $Z \sim \mathcal{N}(0, 1)$ to denote that Z is a standard Gaussian variable. More specifically, Z is a continuous random variable with probability density function*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

We also use $Y \sim \mathcal{N}(\mu, \sigma)$ to denote that Y is a Gaussian variable with mean μ and variance σ^2, i.e. $Y = \sigma Z + \mu$ where Z is a standard Gaussian.

Now we introduce a few facts about Gaussian variables.

Theorem 3. *Let $\vec{Z} = (Z_1, Z_2, ..., Z_d) \in \mathbb{R}^d$, where $Z_1, Z_2, ..., Z_d$ are i.i.d. standard Gaussians. Then the distribution of $\vec{Z}$ is rotationally symmetric. I.e., the probability density will be the same for $\vec{z_1}$ and $\vec{z_2}$ when $\|\vec{z_1}\| = \|\vec{z_2}\|$.*

Proof. The probability density function of $\vec{Z}$ at $\vec{z} = (z_1, z_2, ..., z_d)$ is

$$\phi(z_1)\phi(z_2)...\phi(z_d) = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-(z_1^2+z_2^2+...+z_d^2)/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\|\vec{z}\|^2},$$

which only depends on $\|\vec{z}\|$. □

The following corollary says that the function $\phi(\cdot)$ is indeed a probability density function.

Corollary 4. $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \mathrm{d}z = 1$

Corollary 5. *Linear combination of independent gaussians is still gaussian.*

2 The Berry-Esseen Theorem (CLT with error bounds)

When designing and analyzing algorithms, we usually need to know the convergence rate in order to derive a guarantee on the performance (e.g. time/space complexity) of the algorithm. In this sense, the Central Limit Theorem (Theorem 1) may not be practically useful. The following Berry-Esseen theorem strengthens the CLT with concrete error bounds.

Theorem 6 (the Berry-Esseen Theorem)**.** *Let $X_1, X_2, ..., X_n$ be independent. Assume w.l.o.g. that $\mathrm{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = \sigma_i^2$ and $\sum_{i=1}^n \sigma_i^2 = 1$. Let $Z = X_1 + X_2 + ... + X_n$. (Note that $\mathrm{E}[Z] = 1, \mathrm{Var}[Z] = 1$.) Then $\forall u \in \mathbb{R}$, we have*

$$\left| \Pr[S \leq u] - \Pr_{Z \sim \mathcal{N}(0,1)}[Z \leq u] \right| \leq O(1) \cdot \beta,$$

where $\beta = \sum_{i=1}^{n} \mathrm{E}\,|X_i|^3$.

Remark 1. *The hidden constant in the upperbound of the theorem can be as good as .5514 by [She13].*

Remark 2. *The Berry-Esseen theorem does not need X_i's to be identical. Independence among variables is still essential.*

We still use the unbiased coin flips example to see how this bound works.

Let

$$X_i = \begin{cases} +\frac{1}{\sqrt{N}}, & w.p. \frac{1}{2} \\ -\frac{1}{\sqrt{N}}, & w.p. \frac{1}{2} \end{cases}$$

be independent random variables.

We can check that $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = \frac{1}{n}, \sum \sigma_i^2 = 1$ satisfy the requirement in the Berry-Esseen theroem. We can also compute that $\mathrm{E}\,|X_i|^3 = \frac{1}{n^{\frac{3}{2}}}$, and therefore $\beta = \frac{1}{\sqrt{n}}$.

According to the Berry-Esseen theorem, we have

$$\forall u \in \mathbb{R}, \left| \Pr[S \leq u] - \Pr_{Z \sim \mathcal{N}(0,1)}[Z \leq u] \right| \leq \frac{.56}{\sqrt{n}}. \tag{1}$$

The right-hand side ($\frac{.56}{\sqrt{n}}$) gives a concrete convergence rate.

Now let us investigate whether the $O\left(\frac{1}{\sqrt{n}}\right)$ upper bound can be improved. Say n is even, then $S = \frac{\#\text{Heads} - \#\text{Tails}}{\sqrt{n}}$. Then $S = 0 \Leftrightarrow \#H = \#T = \frac{n}{2}$. Now let us estimate this probability using (1). For $\epsilon > 0$, we have

$$\begin{aligned} \Pr[\#H = \#T] &= \Pr[S = 0] = \Pr[S \leq 0] - \Pr[S \leq -\epsilon] \\ &= (\Pr[S \leq 0] - \Pr[Z \leq 0]) - (\Pr[S \leq -\epsilon] - \Pr[Z \leq -\epsilon]) + (\Pr[Z \leq 0] - \Pr[Z \leq -\epsilon]) \\ &\leq |\Pr[S \leq 0] - \Pr[Z \leq 0]| - |\Pr[S \leq -\epsilon] - \Pr[Z \leq -\epsilon]| + \Pr[-\epsilon < Z \leq 0]. \end{aligned}$$

Taking $\epsilon \to 0^+$, we have

$$\begin{aligned} \Pr[\#H = \#T] &\leq |\Pr[S \leq 0] - \Pr[Z \leq 0]| - |\Pr[S \leq -\epsilon] - \Pr[Z \leq -\epsilon]| \\ &\leq \frac{.56}{\sqrt{n}} + \frac{.56}{\sqrt{n}} = \frac{1.12}{\sqrt{n}}, \end{aligned} \tag{2}$$

where the last inequality is because of (1).

On the other hand, it is easy to see that

$$Pr[\#H = \#T] = \frac{\binom{n}{\frac{n}{2}}}{2^n}.$$

Using Sterling's approximation, when $n \to \infty$, we have

$$\Pr[\#H = \#T] \to \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{2\pi \cdot \frac{n}{2} \cdot (\frac{n}{2e})^n \cdot 2^n} = \frac{\sqrt{2}}{\sqrt{\pi n}} \approx \frac{.798}{\sqrt{n}}. \tag{3}$$

If we had a essentially better upper bound (say $o\left(\frac{1}{\sqrt{n}}\right)$) in (1), we would get an upper bound of $o\left(\frac{1}{\sqrt{n}}\right)$ in (2). This would contradict (3). Therefore the upper bound in (1) given by the Berry-Esseen theorem is asymptotically tight.

References

[She13] I. G. Shevtsova. On the absolute constants in the Berry–Esseen inequality and its structural and nonuniform improvements. *Inform. Primen.*, **7**(1):124–125, 2013.