A Theorist's Toolkit **(CMU 18-859T, Fall 2013)**

Lecture 2: CENTRAL LIMIT THEOREM

September 11th, 2013

Lecturer: Ryan O'Donnell *Scribe: Yu Zhao*

1 SUM OF RANDOM VARIABLES

Let $X_1, X_2, X_3, \dots$ be i.i.d. random variables (Here "i.i.d." means "independent and identically distributed"), s.t. $\mathbf{Pr}[X_i = 1] = p$, $\mathbf{Pr}[X_i = 0] = 1 - p$. X_i is also called Bernoulli random variable.

Let $S_n = X_1 + \cdots + X_n$. We will be interested in the random variable S_n which is called Binomial random variable ($S_n \sim B(n, p)$). If you toss a coin for n times, and $X_i = 1$ represents the event that the result is head in the ith turn, then S_n is just the total number of appearance of head in n times.

Recall some basic facts on expectation and variance, where Y, Y_1, Y_2 are random variables.

- $\mathbf{E}[Y_1 + Y_2] = \mathbf{E}[Y_1] + \mathbf{E}[Y_2]$
- $\mathbf{E}[Y_1 Y_2] = \mathbf{E}[Y_1]\,\mathbf{E}[Y_2]$ if random variables Y_1 and Y_2 are independent ($Y_1 \perp Y_2$)
- $\mathbf{E}[cY] = cE[Y]$, $\mathbf{E}[c + Y] = c + \mathbf{E}[Y]$, where c is a constant
- If we denote $\mu = \mathbf{E}[Y]$, the variance of Y

$$\begin{aligned}\mathbf{Var}[Y] &= \mathbf{E}[(Y-\mu)^2] = \mathbf{E}[Y^2 - 2\mu Y + \mu^2] \\ &= \mathbf{E}[Y^2] - 2\mu\,\mathbf{E}[Y] + \mu^2 = \mathbf{E}[Y^2] - \mathbf{E}[Y]^2\end{aligned}$$

- If $Y_1 \perp Y_2$

$$\begin{aligned}\mathbf{Var}[Y_1 + Y_2] &= \mathbf{E}[(Y_1 + Y_2)^2] - \mathbf{E}[(Y_1 + Y_2)]^2 \\ &= (\mathbf{E}[Y_1^2] + 2\,\mathbf{E}[Y_1 Y_2] + \mathbf{E}[Y_2^2]) - (\mathbf{E}[Y_1]^2 + 2\,\mathbf{E}[Y_1]\,\mathbf{E}[Y_2] + \mathbf{E}[Y_2]^2) \\ &= (\mathbf{E}[Y_1^2] - \mathbf{E}[Y_1]^2) + (\mathbf{E}[Y_2^2] - \mathbf{E}[Y_2]^2) \\ &= \mathbf{Var}[Y_1] + \mathbf{Var}[Y_2]\end{aligned}$$

- $\mathbf{Var}[cY] = c^2\,\mathbf{Var}[Y]$, $\mathbf{Var}[c + Y] = \mathbf{Var}[Y]$
- The standard derivation

$$\sigma = \mathbf{stddev}[Y] = \sqrt{\mathbf{Var}[Y]}$$

For any X_i we have the expectation $\mathbf{E}[X_i] = 1 \cdot \mathbf{Pr}[X_i = 1] + 0 \cdot \mathbf{Pr}[X_i = 0] = p$, therefore

$$\mathbf{E}[S_n] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np$$

Since the variance of X_i $\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 = p - p^2 = p(1-p)$ and X_i's are independent, the variance of S_n is

$$\mathbf{Var}[S_n] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = np(1-p)$$

We would like to somehow normalize the random variable S_n s.t. its mean is 0 and its variance is 1. Let

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

where μ and σ is the mean and standard derivation of S_n. It is easy to see that

$$\mathbf{E}[Z_n] = \mathbf{E}[(S_n - \mu)/\sigma] = (\mathbf{E}[S_n] - \mu)/\sigma = 0$$

and

$$\mathbf{Var}[Z_n] = \mathbf{Var}[(S_n - \mu)/\sigma] = \mathbf{Var}[S_n]/\sigma^2 = 1$$

Since $S_n = \sigma Z_n + pn$, we have

$$\mathbf{Pr}[S_n \le u] = \mathbf{Pr}[\sigma Z_n + pn \le u] = \mathbf{Pr}\left[Z_n \le \frac{u - pn}{\sigma}\right]$$

So if we know the probability distribution of Z_n, we may also know the probability distribution of S_n, vice versa.

Example 1.1. *Suppose* $p = \frac{1}{2}$, $\mathbf{E}[S_n] = \frac{n}{2}$, $\mathbf{Var}[S_n] = \frac{\sqrt{n}}{2}$,

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

It can be seen as

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

Recall from Lecture 1, the probability that Z_n *is 0 is* $\mathbf{Pr}[Z_n = 0] = \Theta(\frac{1}{\sqrt{n}})$, *when* n *is even, as in Figure 1.*

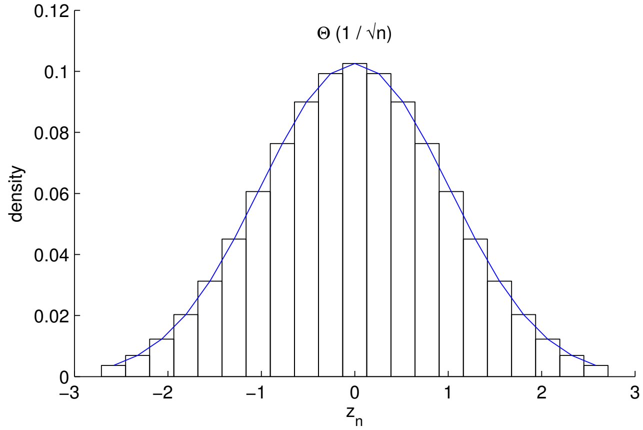

Figure 1: Histogram of Z_n for $n = 60$ and $p = 1/2$

2 GAUSSIAN DISTRIBUTION

Definition 2.1 (Gaussian Distribution)**.** A random variable Z is Gaussian distributed with parameters μ and σ^2, (abbreviated $N(\mu, \sigma^2)$), if it is continuous with p.d.f. (probability density function)

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

where μ refers to the mean and σ^2 refers to the variance of Gaussian. Particularly we call $Z \sim N(0, 1)$ a standard Gaussian.

Fact 2.2. *Let $Z_1, \ldots, Z_d$ be i.i.d. standard Gaussians, $\overrightarrow{Z} = (Z_1, \ldots, Z_d)$. Then $\overrightarrow{Z}$'s distribution is rotationally symmetric, which means for all $\|\overrightarrow{Z}\| = r$, the probability density of $\overrightarrow{Z}$ is the same . Figure 2 shows the probability density of 2-dimension standard Gaussian.*

Proof. p.d.f. of $|\overrightarrow{Z}|$ at $(Z_1, \ldots, Z_d)$ with $\|\overrightarrow{Z}\| = r$ is

$$\begin{aligned}\phi(\overrightarrow{Z}) &= \phi(Z_1)\phi(Z_2)\ldots\phi(Z_d) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \prod_{i=1}^{d} e^{-\frac{Z_i^2}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\frac{1}{2}(Z_1^2+\cdots+Z_d^2)} = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\frac{1}{2}\|(Z_1,\ldots,Z_d)\|^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-r^2/2}\end{aligned}$$

Therefore the probability density of $\overrightarrow{Z}$ only depends on r, which means $\overrightarrow{Z}$'s distribution is rotationally symmetric.

□

Corollary 2.3.

$$\int_{-\infty}^{\infty} \phi(x)dx = 1$$

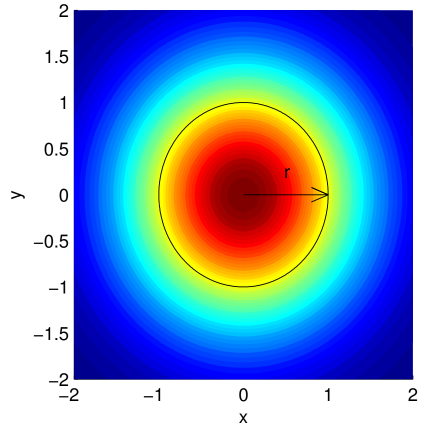

Figure 2: Distribution of 2D-standard Gaussian is rotationally symmetric

Proof. Consider about the 2-dimension standard Gaussian $\overrightarrow{Z} = (Z_1, Z_2)$, which has p.d.f. $\frac{1}{2\pi}e^{-\frac{1}{2}(Z_1^2+Z_2^2)}$. We can first integrate the p.d.f. in a natural way. The integral becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}dz_1dz_2 = \left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz\right)^2 = \left(\int_{-\infty}^{\infty}\phi(z)dz\right)^2$$

Therefore, in order to prove $\int_{-\infty}^{\infty}\phi(x)dx = 1$, it suffices to prove

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(z_1^2+z_2^2)}dz_1dz_2 = 2\pi$$

since $\int_{-\infty}^{\infty}\phi(x)dx > 0$.

On the otherhand, as in Figure 3, we can intergrate the function $e^{-\frac{1}{2}(z_1^2+z_2^2)}$ by height. The height of each cylinder is dh, while the radium of it is r which satisfies $e^{-r^2/2} = h$. Therefore we have

$$r^2 = 2\ln\frac{1}{h}$$

Since $0 < h = e^{-r^2/2} \leq 1$, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(z_1^2+z_2^2)}dz_1dz_2 = \int_0^1 \pi r^2 dh = \int_0^1 2\pi\ln\frac{1}{h}dh = 2\pi\left(h\ln\frac{1}{h}+h\right)\Big|_0^1 = 2\pi$$

□

Corollary 2.4 (Sum of Independent Gaussian)**.** *Suopose $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ are independent. We have*

$$aX + bY \sim (a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

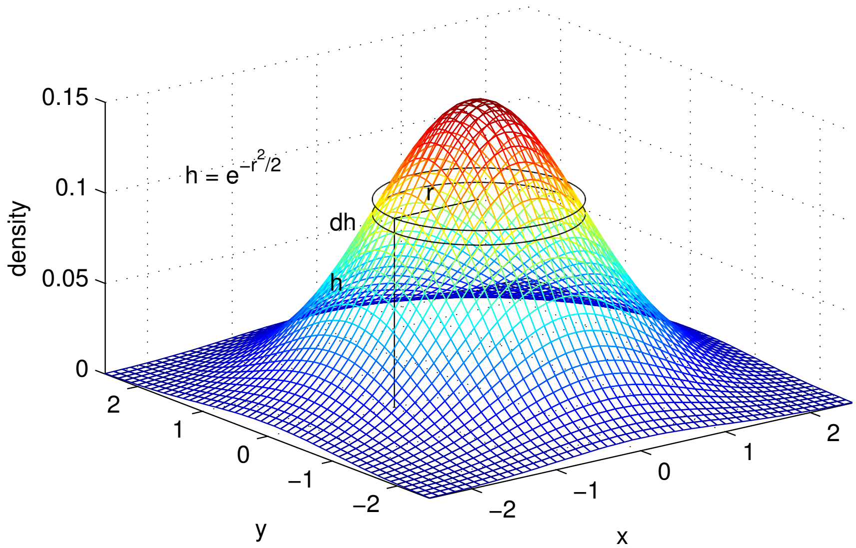

Figure 3: Integration of 2D Gaussian by height

Proof. If $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, then $X, Y \sim N(0, 1)$. We want to prove $aX + bY \sim N(0, a^2 + b^2)$, and we have

$$Z = aX + bY = (a, b) \cdot (X, Y)$$

Because 2D standard Gaussian (X, Y) is rotationally symmetric, we can rotate (a, b) to $(\sqrt{a^2 + b^2}, 0)$ as in Figure 4. Now $Z = (\sqrt{a^2 + b^2}, 0) \cdot (X, Y) = \sqrt{a^2 + b^2}X = N(0, a^2 + b^2)$.

Suppose $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$. Since $(X - \mu_1)/\sigma_1, (Y - \mu_2)/\sigma_2 \sim N(0, 1)$, we have

$$Z - a\mu_1 - b\mu_2 = a(X - \mu_1) + b(Y - \mu_2) = a\sigma_1 \frac{X - \mu_1}{\sigma_1} + b\sigma_2 \frac{Y - \mu_2}{\sigma_2} \sim N(0, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Therefore $Z \sim (a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.

□

3 CENTRAL LIMIT THEOREM

Theorem 3.1 (Central Limit Theorem)**.** *For any i.i.d.* $X_1, \ldots, X_n$, $Z_n \xrightarrow[n\to\infty]{} Z$, *where* Z *is a standard Gaussian* $N(0, 1)$,

i.e. $\forall u \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{Pr}[Z_n \le u] = \mathbf{Pr}[Z \le u]$$

Remember the Central Limit Theorem is kind of useless since we don't know how quickly Z_n will converge to a standard Gaussian.

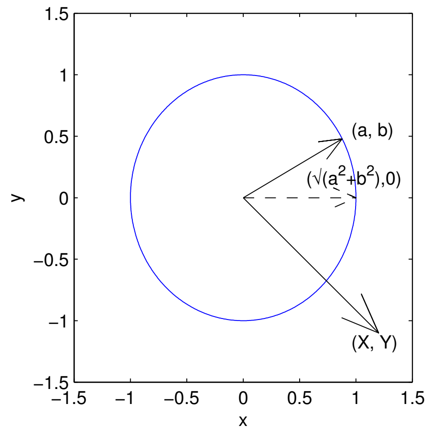

Figure 4: Rotating (a, b) to $(\sqrt{a^2+b^2}, 0)$ since (X, Y) is rotationally symmetric

The name of the useful version is Berry-Esseen Theorem.

Theorem 3.2 (Berry-Esseen Theorem). *Let $X_1, \ldots, X_n$ be independent r.v.s, Assume (W.O.L.O.G.) $\mathbf{E}[X_i] = 0$. we write $\sigma_i^2 = \mathbf{E}[X_i^2] = \mathbf{Var}[X_i]$ and assume $\sum_{i=1}^n \sigma_i^2 = 1$.*

Let

$$S = X_1 + \ldots X_n$$

so $\mathbf{E}[S] = 0, \mathbf{Var}[S] = 1$. Then $\forall u \in \mathbb{R}$,

$$|\,\mathbf{Pr}[S \le u] - \mathbf{Pr}[Z \le u]| \le O(1) \cdot \beta$$

where $Z \sim N(0, 1)$, $\beta = \sum_{i=1}^n \mathbf{E}[|X_i|^3]$

Remember β is not always small. Here is an example.

Example 3.3. *Let $X_1 = \begin{cases} +1 & w.p. \frac{1}{2} \\ -1 & w.p. \frac{1}{2} \end{cases}$ and $X_2, \ldots, X_n \equiv 0$. In this scenario, S just has the same distribution as X_1. Then $\beta = 1$ is a big number.*

On the other hand, this theorem does work in some cases.

Example 3.4. $X_i = \begin{cases} +\frac{1}{\sqrt{n}} & w.p. \frac{1}{2} \\ -\frac{1}{\sqrt{n}} & w.p. \frac{1}{2} \end{cases}$,

$$\mathbf{E}[X_i^2] = \frac{1}{n}, \mathbf{E}[|X_i|^3] = \frac{1}{n^{-\frac{3}{2}}}, \forall i$$

Therefore $\beta = 1/\sqrt{n}$. In this case β is small, and

$$|\,\mathbf{Pr}[S \le u] - \mathbf{Pr}[Z \le u]| = O\left(\frac{1}{\sqrt{n}}\right)$$

The most recent upper bound of $O(1)$ is $O(1) \approx .5514$ [She13]. In this scenario, we have

$$|\mathbf{Pr}[S \le u] - \mathbf{Pr}[Z \le u]| \le \frac{.56}{\sqrt{n}}$$

We can find that $\mathbf{Pr}[Z \le u] \approx 0.001$ when $u = -3$ by computer or in standard normal table. Suppose H is the number of appearance of head when tossing a coin for n times. As in Example 1.1, we have

$$\frac{2H - n}{\sqrt{n}} = S \le u$$

which means

$$H \le \frac{n}{2} + u\frac{\sqrt{n}}{2}$$

Assign $u = -3$, we have $\mathbf{Pr}[H \le n/2 - 1.5\sqrt{n}] \approx 0.001$, which is quite small.

Sometimes β might be extremely large.

Example 3.5. *Let*

$$X_1 = \begin{cases} +n & w.p. \frac{1}{2n^2} \\ -n & w.p. \frac{1}{2n^2} \\ 0 & otherwise \end{cases}$$

and $X_2, \ldots, X_n \equiv 0$. $\sum_{i=1}^n \mathbf{E}[X_i^2] = \mathbf{E}[X_1^2] = 1$. *In this scenario,* $\mathbf{Pr}[S = 0] \to 1$, *which means* S *does not converge to a normal distribution. In this case* $\beta \to +\infty$.

From this example, we can see that the constraint on $\beta = \sum_{i=1}^n \mathbf{E}[|X_i|^3]$ capture two things: The r.v.s will not become extremely huge with small probability; The sum does not only depend on finite number of random variables.

Notice that Berry-Esseen Theorem is good because it does not care about the value of u. We have

$$\mathbf{Pr}\left[H \le \frac{n}{2} + u\frac{\sqrt{n}}{2}\right] \overset{O(\frac{1}{\sqrt{n}})}{\approx} \int_{-\infty}^{u} \phi(z)dz$$

even though $u = -0.2\sqrt{n}$ which is supertiny.

4 CUMULATIVE DISTRIBUTION

Definition 4.1 (Cumulative Distribution Function of Standard Gaussian)**.** We denote $\Phi(u)$ as the c.d.f. (cumulative distribution function) of standard Gaussian $N(0, 1)$

$$\Phi(u) = \mathbf{Pr}[Z \le u] = \int_{-\infty}^{u} \phi(z)dz$$

and define

$$\bar{\Phi}(u) = \sum_{u}^{+\infty} = \mathbf{Pr}[Z \ge u] = \Phi(-u)$$

It is trivial that $\Phi(u) = \bar{\Phi}(u) = \frac{1}{2}$.

Fact 4.2.

$$\bar{\Phi}(u) = O\left(\frac{\phi(u)}{u}\right)$$

when $u > 0$.

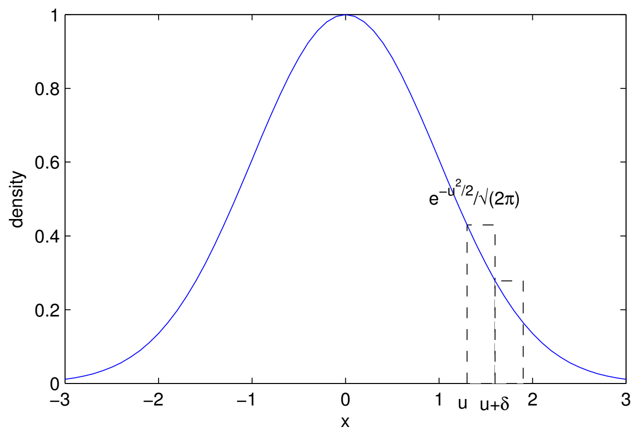

Figure 5: Find a δ s.t. $\phi(u+\delta) \approx c \cdot \phi(u)$, where c is a constant and less than 1

Proof. We want to find a δ s.t. $\phi(u+\delta) \approx c \cdot \phi(u)$, where c is a constant and less than 1 (Figure 5). $\delta = 1/u$ satisfies our conditions.

$$\phi(u+\frac{1}{u}) = \frac{1}{\sqrt{2\pi}} e^{-(u+\frac{1}{u})^2/2} = e^{-1}\phi(u) \cdot e^{-\frac{1}{2u^2}} \le e^{-1}\phi(u)$$

In general, we have $\phi(u+\frac{k}{u}) \le e^{-k}\phi(u)$ where $u > 0$ and $k \le \mathbb{N}$. Since $\phi(u)$ is descreasing when $u > 0$, using the method in Lecture 1, we have

$$\bar{\Phi}(u) = \int_u^{+\infty} \phi(u)du \le \sum_{k=0}^{\infty} \frac{1}{u}\phi(u+\frac{k}{u}) \le \frac{\phi(u)}{u}\sum_{k=0}^{\infty} e^{-k} = \frac{e}{e-1} \cdot \frac{\phi(u)}{u} = O\left(\frac{\phi(u)}{u}\right)$$

□

Proposition 4.3.

$$\bar{\Phi}(u) \sim \frac{\phi(u)}{u}$$

when $u \to +\infty$.

In fact,

$$(\frac{1}{u} - \frac{1}{u^3})\phi(u) \le \bar{\Phi}(u) \le \frac{1}{u}\phi(u)$$

when $u \to +\infty$.

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.