# Lecture 14: Central Limit Theorem

*4 April, 2018*

*Sunil Kumar Gauttam*

*Department of Mathematics, LNMIIT*

**Theorem 14.1 (Central Limit Theorem)** *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean $\mu$ and non-zero variance $\sigma^2$. Define*

$$S_n := X_1 + X_2 + \cdots + X_n, \quad Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

*Then*

$$\lim_{n\to\infty} P(Z_n \leq x) = N(x), \ \forall x \in \mathbb{R},$$

*where* $N(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$

Another frequently used notation for $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ is $\Phi(x)$.

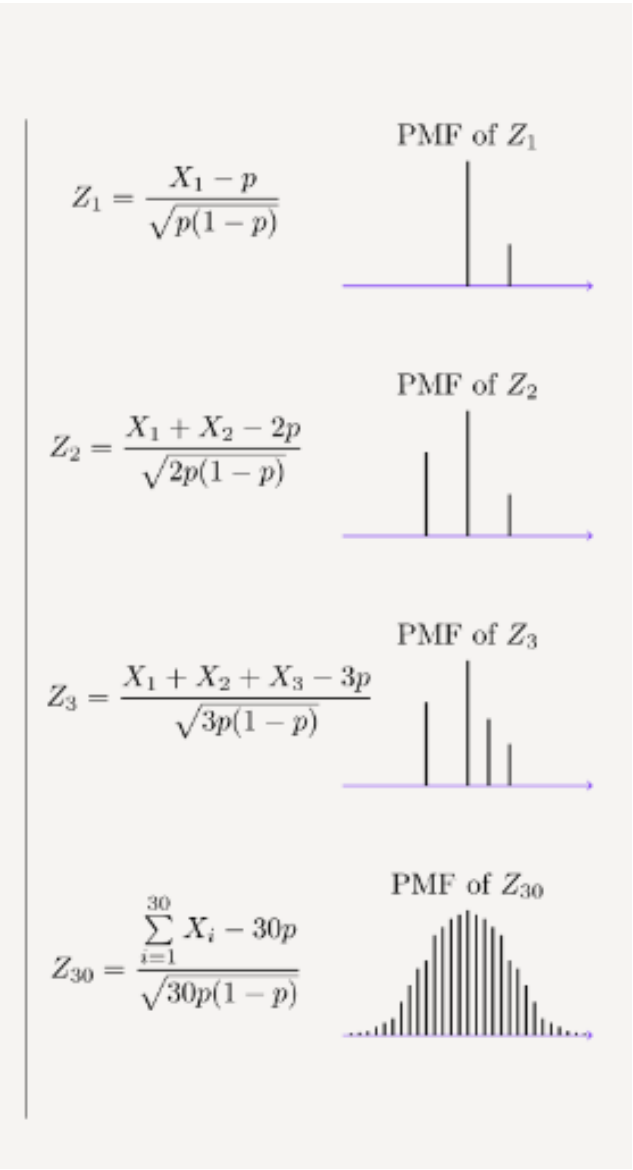
The central limit theorem is surprisingly general. Besides independence, and the implicit assumption that the mean and variance are finite, it places no other requirement on the distribution of the $X_i$, which could be discrete, continuous, or mixed.

To get a feeling for the CLT, let us look at some examples.

**Example 14.2** *Let $X_i$'s be independent Bernoulli($p$). Then $EX_i = p, Var(X_i) = p(1-p)$. Also, $S_n = X_1 + X_2 + \cdots + X_n$ has Binomial($n, p$) distribution. Thus,*

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

*We plot the PMF of $Z_n$ for different values of $n$ by choosing $p = \dfrac{1}{3}$.*




*As you see, the shape of the PMF gets closer to a normal PDF curve as n increases. Hence, the CDF of $Z_n$ will converges to the standard normal CDF.*

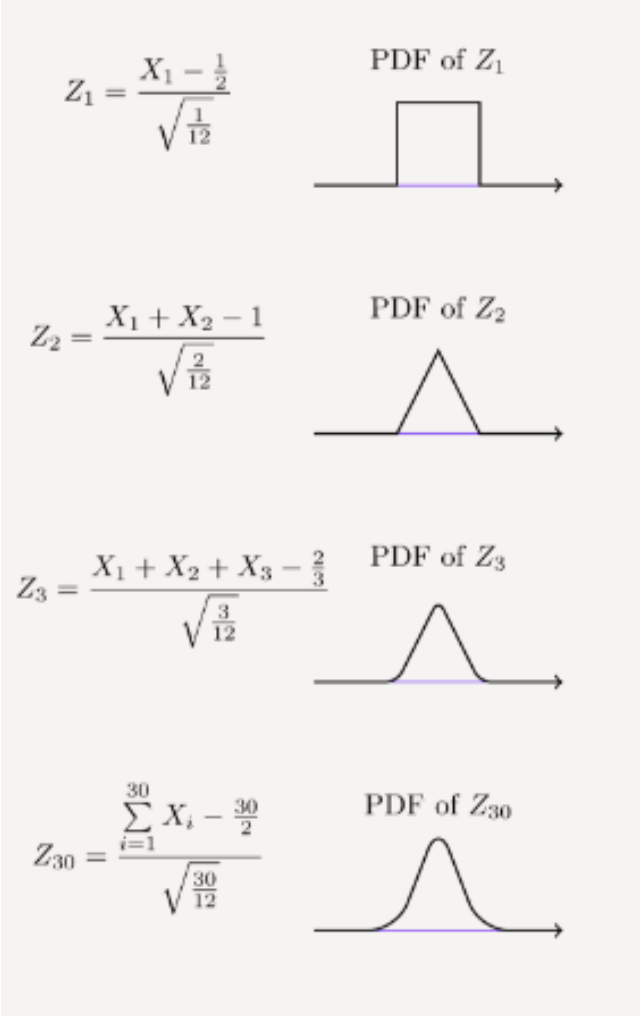
$$F_{Z_n}(x) = \sum_{z \in R_{Z_n}: z \le x} f_{Z_n}(z) \to N(x) \text{ or } \Phi(x)$$

*That what the CLT states.*

**Example 14.3** *Let $X_i$'s be independent Uniform$(0,1)$. Then $EX_i = \frac{1}{2}, Var(X_i) = \frac{1}{12}$. Let $S_n = X_1 + X_2 + \cdots + X_n$. In this case,*

$$Z_n = \frac{S_n - \frac{n}{2}}{\sqrt{n/12}}.$$

*We have derived a general formula for the pdf of sum of two independent random variables with pdf's. Hence random variable $Z_n$ has pdf. We plot the PDF of $Z_n$ for different values of $n$*




*As you see, the shape of the PDF gets closer to a normal PDF curve as n increases. Hence, the CDF of $Z_n$ will converges to the standard normal CDF.*

$$F_{Z_n}(x) = \int_{-\infty}^{x} f_{Z_n}(z)dz \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

*That what the CLT states.*

**Normal Random variable** A normal random variable $X$ with zero mean and unit variance is said to be a standard normal. It's distribution function is denoted by $N(\cdot)$:

$$N(x) = P(X \leq x) = P(X < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

We can not get closed form for $N(x)$. One use numerical technique to find the approximate values of $N(x)$. It is recorded in a table and is a very useful tool for calculating various probabilities involving normal random variables. Note that the table only provides the values of $N(x)$ for $x > 0$, because the omitted values can be found using the symmetry of the pdf. For example, if $X$ is a standard normal random variable, for $x > 0$ we have

$$N(-x) = P(X \leq -x) = P(X \geq x) = 1 - P(X < x) = 1 - N(x).$$

**TABLE 1 Values of the standard normal distribution function**

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| .0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| .1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5363 | .5675 | .5714 | .5753 |
| .2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| .3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| .4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| .5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| .6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| .7 | .7580 | .7611 | .7642 | .7673 | .7703 | .7734 | .7764 | .7794 | .7823 | .7852 |
| .8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| .9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9278 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9430 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9648 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9700 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9762 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9874 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3. | .9987 | .9990 | .9993 | .9995 | .9997 | .9998 | .9998 | .9999 | .9999 | 1.0000 |

Let $X$ be a normal random variable with mean $\mu$ and variance $\sigma^2$. We define a new random variable $Z := \dfrac{X-\mu}{\sigma}$. Then $Z$ is a linear function of $X$ hence normal. Also $Z$ mean zero and variance is one. To see this,

$$E[Z] = \frac{E[X]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0.$$
$$var(Z) = \frac{var(X-\mu)}{\sigma^2} = \frac{var(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

Thus $Z$ is a standard normal random variable. This fact allows us to calculate the probability of any event defined in term of $X$: we redefine the event in terms of $Z$, then use the standard normal normal table. This is how it is done:

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = N\left(\frac{x-\mu}{\sigma}\right)$$

**Normal Approximation Based on the Central Limit Theorem:** The central limit theorem allows us to calculate probabilities related to $Z_n$ as if $Z_n$ were normal, CLT says $P(Z_n \le x) \approx N(x)$ for large values of $n$. Note that $S_n = \sigma\sqrt{n}Z_n + n\mu$. Since normality is preserved under linear transformations, Sicne $Z_n \sim N(0,1)$, this is equivalent to treating $S_n$ as a normal random variable with mean $n\mu$ and variance $n\sigma^2$.

Let $S_n = X_1 + \cdots + X_n$, where the $X_i$ are independent identically distributed random variables with mean $\mu$ and variance $\sigma^2$. If $n$ is large, the probability $P(S_n \le c)$ can be approximated by treating $S_n$ as if it were normal, according to the following procedure.

Step 1 Calculate the mean $n\mu$ and the variance $n\sigma^2$ of $S_n$.

Step 2 Use the approximation

$$P(S_n \le c) \approx N\left(\frac{c-n\mu}{\sigma\sqrt{n}}\right)$$

**Example 14.4** *We load on a plane* 100 *packages whose weights are independent random variables that are uniformly distributed between* 5 *and* 50 *kg. What is the probability that the total weight will exceed 3000 kg?*

**Solution:** Let us translate the problem in probabilistic model. Let $X_i$ denotes the weight of $i$th packages. $X_1, X_2, \cdots, X_{100}$ are iid uniform random variables with density

$$f(x) = \begin{cases} \dfrac{1}{45}, & \text{if } 5 \le x \le 50 \\ 0, & \text{otherwsie} \end{cases}$$

Let $S = X_1 + X_2 + \cdots + X_{100}$ denote the total weight. Then question is to calculate the $P(S > 3000)$. Let $f$, $g$, and $h$ be functions on the reals, and suppose the convolutions $(f * g) * h$ and $f * (g * h)$ exist. Then we have $(f * g) * h = f * (g * h)$. Using this result and $S$ is sum of independent random variable we can see that pdf of $S$ is 100-fold convolution of $f$. It is not easy to calculate this one. Hence it is very difficult to find the desired probability, but an approximate answer can be quickly obtained using the central limit theorem. Treat $S$ as normal random variable. So now we find it's mean and variance, which is $100\mu$ and $100\sigma^2$ where $\mu = E[X_i], \sigma^2 = var(X_i)$.

$$\begin{aligned}
E(X_i) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \frac{1}{45} \int_5^{50} x dx \\
&= \frac{5 + 50}{2} = 27.5 \\
E(X_i^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\
&= \frac{1}{45} \int_5^{50} x^2 dx \\
&= \frac{(50)^3 - 5^3}{3 \times 45} \\
&= \frac{(50)^2 + 50 \times 5 + 5^2}{3} \\
&= 925 \\
var(X_i) &= 925 - (27.5)^2 = 925 - 756.25 = 168.75
\end{aligned}$$

Now

$$P(S > 3000) = 1 - P(S \leq 3000) = 1 - N\left(\frac{3000 - 2750}{10\sqrt{168.75}}\right) = 1 - N\,(1.92)$$

■