Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Nov. 3 2020

Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Chong Han

Outline

- 3. Multiple Random Variables
  - Two Random Variables
  - Marginal PDF
  - Conditional PDF
  - Functions of Two Random Variables
  - Joint Moments
  - Mean Square Error Estimation
  - Probability bound
  - Random Vectors
  - Sample Mean
  - Convergence of Random Sequences
  - Central Limit Theorem

Sample Mean

Sample Mean

Now, let us consider a different estimation problem. Earlier, we have looked at the problem of estimating a RV, Y, from a given RV, X. Here, we try to get an estimate for the mean of the RV, X, given the repeated independent trials, or outcomes of X, X_1, …X_n. So the X_i's are iid with the same PDF as X.

One way to estimate $\mu=E(X)$ is by using the sample mean:

Average

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The difference between this and the previous estimators is that here we are estimating a non-random quantity.

Performance

Three desirable qualities of an estimator of a non-random quantity are:

1. Unbiased: $E\{\hat{\mu}\} = \mu$

A biased estimator would have: $E\{\hat{\mu}\} = \mu + b$

2. Low MSE $E\left\{(\hat{\mu} - \mu)^2\right\}$

Notice that if $\hat{\mu}$ is unbiased, this is the variance of $\hat{\mu}$.

3. Consistency: $\lim_{n \to \infty} P\left[\left|\hat{\mu}_n - \mu\right| > \varepsilon\right] = 0$ (Convergent in Prob.)

Bias of the Sample Mean

Determine the bias of $\hat{\mu} = \overline{X}$

$$E\{\overline{X}\} = \frac{1}{n}\sum_i E\{X_i\} = \mu$$

Therefore, $\overline{X}$ is an unbiased estimator of μ.

MSE of the Sample Mean

MSE

$$E\left\{\left(\bar{X}-\mu\right)^2\right\} = E\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right]^2\right\}$$

$$= E\left\{\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} \mu\right)\right]^2\right\}$$

$$= E\left\{\left[\frac{1}{n}\sum_{i=1}^{n} \left(X_i - \mu\right)\right]^2\right\}$$

MSE of the Sample Mean

Use "the variance of the sum is the sum of the variances" for independent RVs

$$E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right]^2\right\}=\frac{1}{n^2}E\left\{\left[\sum_{i=1}^{n}(X_i-\mu)\right]^2\right\}=\frac{1}{n^2}\mathrm{var}(S_n)$$

where $S_n=\sum_{i=1}^{n}X_i$

$$\therefore MSE=\frac{1}{n^2}n\sigma_X^2=\frac{\sigma_X^2}{n}=\sigma_{\overline{X}}^2$$

MSE decreases as n increases.

Consistency

We can apply Chebyshev's inequality to $\overline{X}$:

We will use $\hat{\mu} = \overline{X}$ in place of X:

$\frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$

$$P\left\{\left|\overline{X}_n - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma_{\overline{X}}^2}{\varepsilon^2}$$

Negate both sides and add one:

$$1 - P\left\{\left|\overline{X}_n - \mu\right| \geq \varepsilon\right\} \geq 1 - \frac{\sigma_{\overline{X}}^2}{\varepsilon^2}$$

$$P\left\{\left|\overline{X}_n - \mu\right| < \varepsilon\right\} \geq 1 - \frac{\sigma_X^2}{n\varepsilon^2}$$

Interpretation

For any choice of ε and probability $1-\delta$, we can select n high enough so that the probability ~~$1-\delta$~~ is at least $1-\delta$ that $\bar{X}$ is within ε of the true value.

increase n. #.

Ex: Noisy measurements of a constant but unknown voltage, v, are taken.

$$X_i = V + N_i, \text{ where } N_i\text{'s are iid, Gaussian,}$$

$$E(N_i) = 0, \quad \sigma_{N_i} = 1\mu V$$

How many measurements are required so that $\bar{X}$ is within $\varepsilon = 0.5\mu V$ of the true mean with probability at least 0.99?

$$P\{|\bar{X}_n - \mu| < 0.5\mu V\} \geq 1 - \frac{\sigma_x^2}{n \times (0.5\mu V)^2} = 0.99.$$

$$1 - \frac{\sigma_x^2}{n \times 0.25\mu V^2} = 0.99$$

$$\frac{1}{0.25n} = 0.01$$

$$n = \frac{1}{0.0025} = 400$$

Short Summary

- The sample mean is a way to estimate a non-random quantity μ
- It is an unbiased estimator
- Its variance decreases by $1/n$, where n is the number of samples $\sigma_{\bar{x}} = \frac{\sigma_x^2}{N}$
- It is a consistent estimator
- The Chebyshev Inequality can be applied to get a sometimes loose lower bound on the n required to achieve a certain quality in the estimate

Convergence of Random Sequences

Convergence of Random Sequences

Consider a sequence of RVs, each defined on the same sample space ***S***

$$X_1(s), X_2(s), \ldots, X_n(s), \ldots$$

Each outcome in ***S*** corresponds to a particular sequence of real numbers

5 types of convergence:

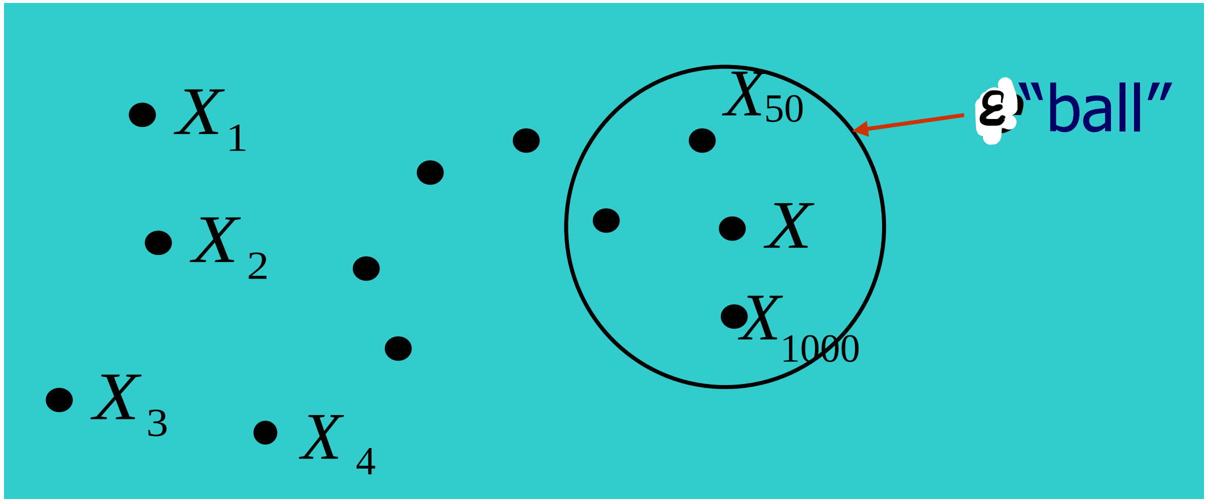
- Sure
- Almost-sure
- In probability
- Mean square
- In distribution

Convergence of a Non-random Sequence

"Sure" convergence

$X_n \to X$ if for any $\varepsilon > 0$, there exists an integer $N(\varepsilon)$ such that

$$|X_n - X| < \varepsilon \text{ for } n > N(\varepsilon)$$



Sure (s) Convergence

Also known as "Convergence Everywhere"

element in sample space

$$X(n, s) \to X(s) \text{ for all } s \in \mathbf{S}.$$

index

We may write $X(n) \to X$ surely

Every outcome converges to its respective limit

Almost Sure (a.s.) Convergence

Also known as

- "Convergence Almost Everywhere" (a.e.) or
- "with probability 1 (wp1)":

except for several points, non-continuous

not s but A

$A \subset \Omega$, $A \neq \Omega$.

$$X(n,s) \to X(s) \text{ for all } s \in \mathbf{A}, \text{ such that } P(\mathbf{A}) = 1$$

$P(A)=1$

or
$$P\left[\lim_{n\to\infty} X(n) = X\right] = 1$$

$P(\{w : \lim_{n\to\infty} X_n(w) = X(w)\}) = 1$

We may write $X(n) \to X$ wp1, a.s. or a.e.

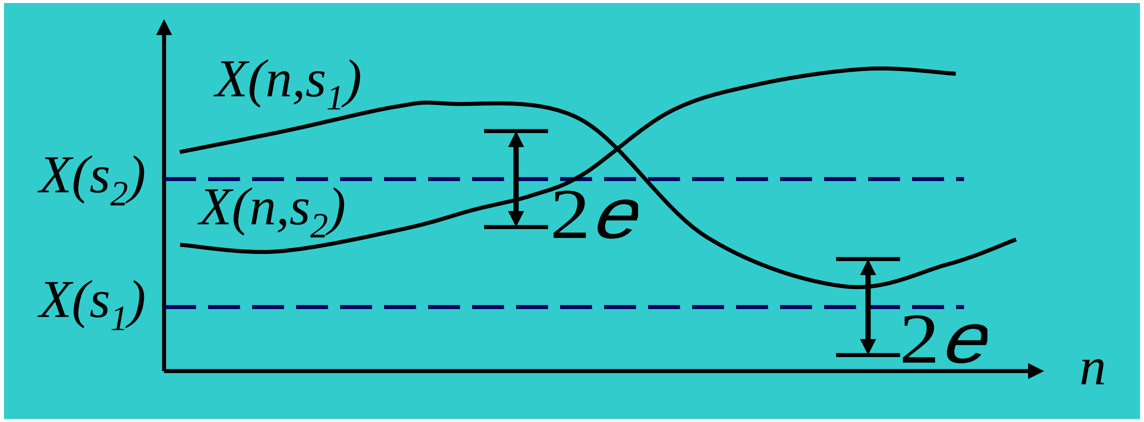
Convergence in Probability (p)

For any $e > 0$

$$\lim_{n\to\infty} P\left(\left|X(n) - X\right| > \varepsilon\right) = 0$$

The limit is "on the outside"

This does not guarantee that any particular sample function converges.



Example: p Does Not Imply a.s.

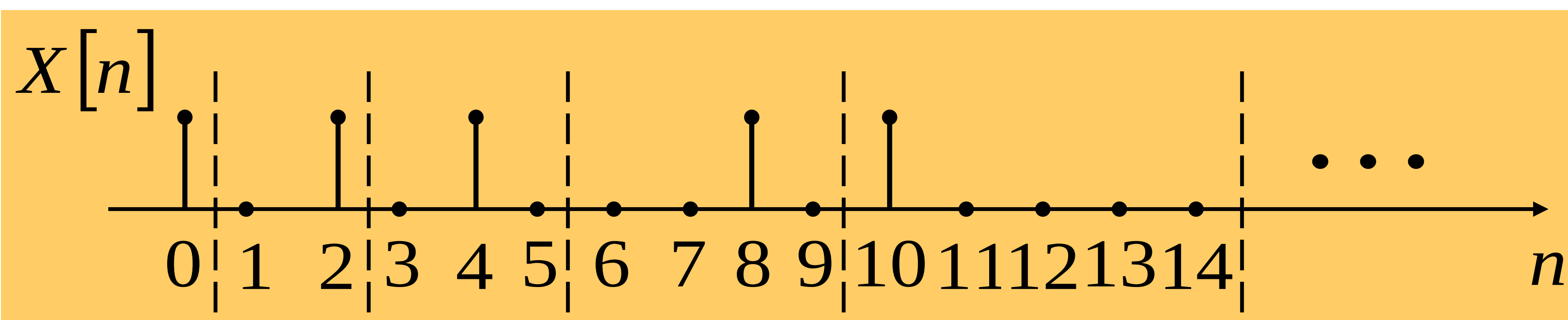
almost sure

Define a random binary pulse sequence $X[n] \in \{0,1\}$, on $n \geq 0$ as follows:

Set $X[0]=1$

Then for next two points, set exactly one of the $X[n]$'s to one, equally likely.

For the next three points, set exactly one to "one" and the others to "zero," and so forth.



Example, Concluded

$$P\left(\left|X[n]-0\right|>\varepsilon\right)=\frac{1}{m} \qquad \text{for } \frac{(m-1)m}{2}\le n<\frac{(m+1)m}{2}$$

For $m=2$, endpoints are 1 and 3.
For $m=3$, endpoints are 3 and 6.

$$P\left(\left|X[n]-0\right|>\varepsilon\right)\to 0 \quad \text{as n} \to \infty$$

This sequence converges in probability $\quad \therefore X[n]\to 0 \quad \text{(p)}$

Since no sequence outcome converges to zero,

$$\therefore X[n] \not\to 0 \text{ a.s.}$$

Mean Square Convergence (MS)

$$E\left[\left(X(n,s)-X(s)\right)^2\right]\to 0 \text{ as } n\to\infty$$

Denoted: $\text{l.i.m.}\, X_n = X \quad \text{as} \quad n\to+\infty$

"Average error power goes to zero"

Relation Between MS and "In Probability"

Recall Markov's Inequality. X is a non-negative RV with mean m_X. For any $a>0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Now let Y be a RV and Y_n a sequence.

$$P(|Y_n - Y| > \varepsilon) = P([Y_n - Y]^2 > \varepsilon^2) < \frac{E\{[Y_n - Y]^2\}}{\varepsilon^2}$$

$E\{[Y_n - Y]^2\} \to 0$

MS

implies

$P(|Y_n - Y| > \varepsilon) \to 0$

In probability

Example

Let S=[0,1], and let a point be selected at random. Consider the random sequence

$$Z(n,s) = e^{-n(ns-1)}$$

For $s = 0$, the sequence diverges → a single outcome diverge

For $s \neq 0$, the sequence converges to zero

Therefore, $Z(n) \to 0$ wp1

↓ with probability 1

$X(n,s) \to 0$

$A = (0,1]$

↘ almost sure

Example, Concluded

就算 converge surely
也未必 converge in the mean square sense

$$E\left\{\left[e^{-n(ns-1)}-0\right]^2\right\}=\int_0^1 e^{-2n(ns-1)}ds=e^{2n}\int_0^1 e^{-2n^2 s}ds$$

$$=e^{2n}\left.\frac{e^{-2n^2 s}}{-2n^2}\right|_0^1$$

$$=e^{2n}\left(\frac{e^{-2n^2}-1}{-2n^2}\right)\to\infty \text{ as } n\to\infty$$

Therefore, $Z(n,s)$ does not converge in the mean square sense

Convergence In Distribution (d)

The CDFs of the RVs in the sequence converge to the CDF of the limiting RV

$$F_{X_n}(x) \rightarrow F_X(x) \text{ as } n \rightarrow \infty$$

This is the weakest type of convergence

Example: The CLT

central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of iid RVs with $m < ¥ , s^2 < ¥$

Consider

$$S_n = \sum_{i=1}^{n} X_i, \quad E\{S_n\} = n\mu \quad \mathrm{var}\{S_n\} = n\sigma^2$$

Then $Z_n = \dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$

is a zero mean, unit variance version of S_n.

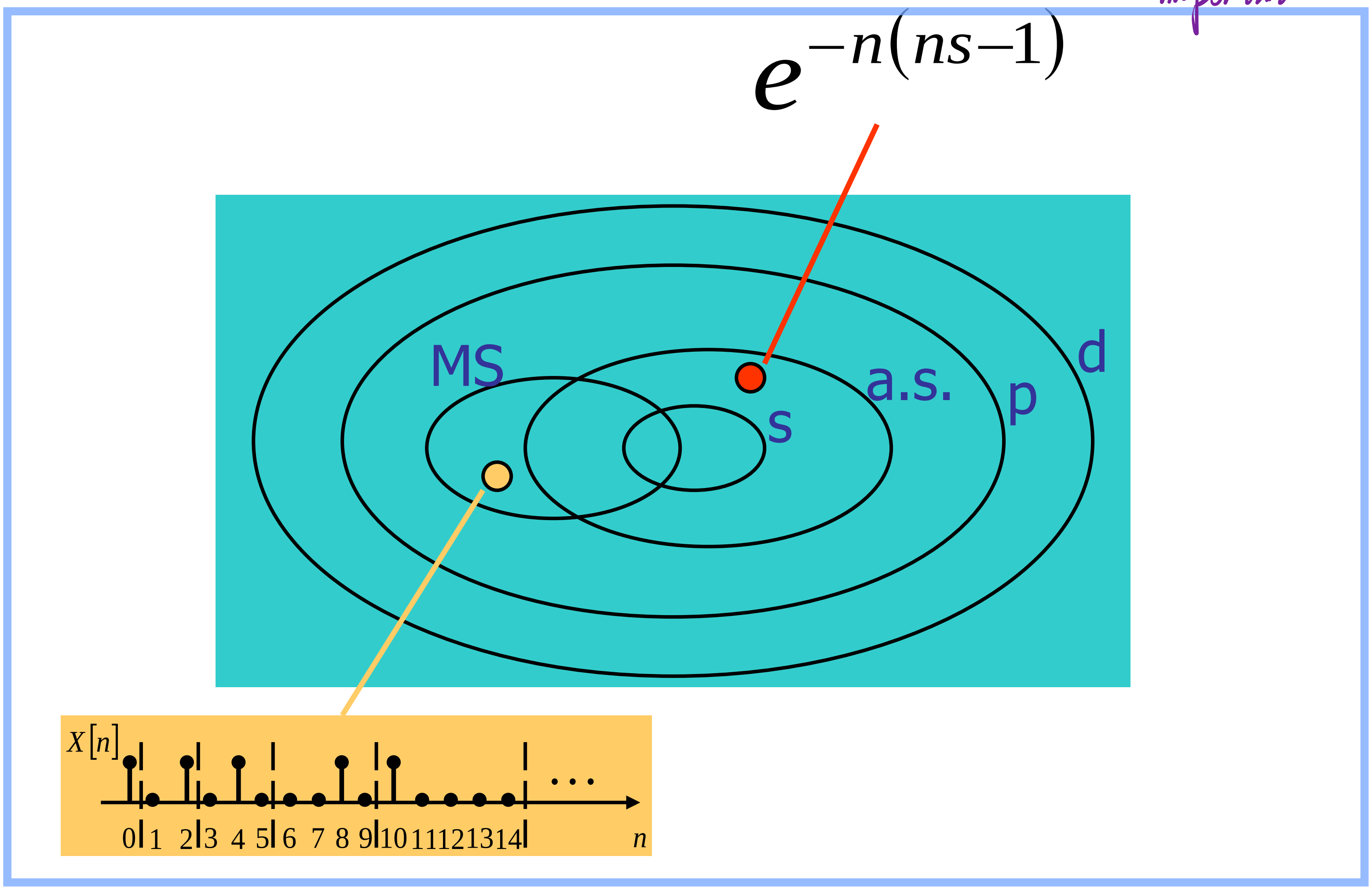
The Central Limit Theorem (CLT) says:

$$F_{Z_n}(z) \to \mathrm{F}(z) \text{ as } n \to \infty$$

Recall $\mathrm{F}(z)$ is the standard normal CDF.

Comparison

important



Weak Law of Large Numbers

Let X_i, $i=1,2,\ldots,n$ be iid RVs with finite mean μ .Then the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$$ in probability, or

$$\lim_{n\to\infty} P\left(\left|\overline{X} - \mu\right| < \varepsilon\right) = 1$$

This is the “Weak Law of Large Numbers”.

Strong Law of Large Numbers

If X_i's have finite variance in addition to finite mean, then we can strengthen the result:

$$P\left(\lim_{n\to\infty} \overline{X} = \mu\right) = 1$$

The limit is inside!

This is the "Strong Law of Large Numbers".

X converges to n "with probability 1" or "almost surely."

The strong and weak laws tell us that it pays to take a lot of measurements if you are trying to estimate the true mean.

Short Summary

Convergence of a random sequence is not as simple as convergence of a sequence of real numbers

There are a number of different senses of convergence, and they are not totally "nested" within one another

m.s. is not nested.

The Weak and Strong Laws of Large Numbers give us conditions for two different types of convergence of the sample mean to the true mean

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid RVs with $\mu < \infty$, $\sigma^2 < \infty$

Consider

$$S_n = \sum_{i=1}^{n} X_i, \quad E\{S_n\} = n\mu \quad \mathrm{var}\{S_n\} = n\sigma^2$$

Then $Z_n = \dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$ $\qquad \dfrac{\frac{1}{n}S_n - \mu}{\frac{\sigma}{\sqrt{n}}} = Z_n$

is a zero mean, unit variance version of S_n.

The Central Limit Theorem (CLT) says:

$$F_{Z_n}(z) \to \phi(z) \text{ as } n \to \infty$$

Recall $\phi(z)$ is the standard normal CDF.

Application of CLT

Can approximate sums of a large number of RVs as Gaussian RVs with appropriate μ and σ^2

Example:

Raindrops have an expected weight of μ_W and a weight std dev of σ_W. Approximate the probability that 1000 drops collected weight more than 1,200 μ_W.

$$S_{1000} = \sum_{i=1}^{1000} W_i$$

$P(S_{1000} > 1200\mu_w)$

$= P\left(Z_{1000} > \frac{200\mu_w}{\sqrt{1000\sigma_w^2}}\right) \approx 1-\phi(\quad)$

CLT applied

Sum of Bernoulli RVs

Consider $Y = \sum_{i=1}^{n} X_i$

where the X_i's are iid with: $P(X_i = 1) = p$

$$\mathrm{P}(X_i = 0) = 1 - p$$

Therefore, Y is a binomial RV with probability of success p.
If $k \leq n$, then

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

How to apply CLT? $= P(k \leq Y < k-1)$

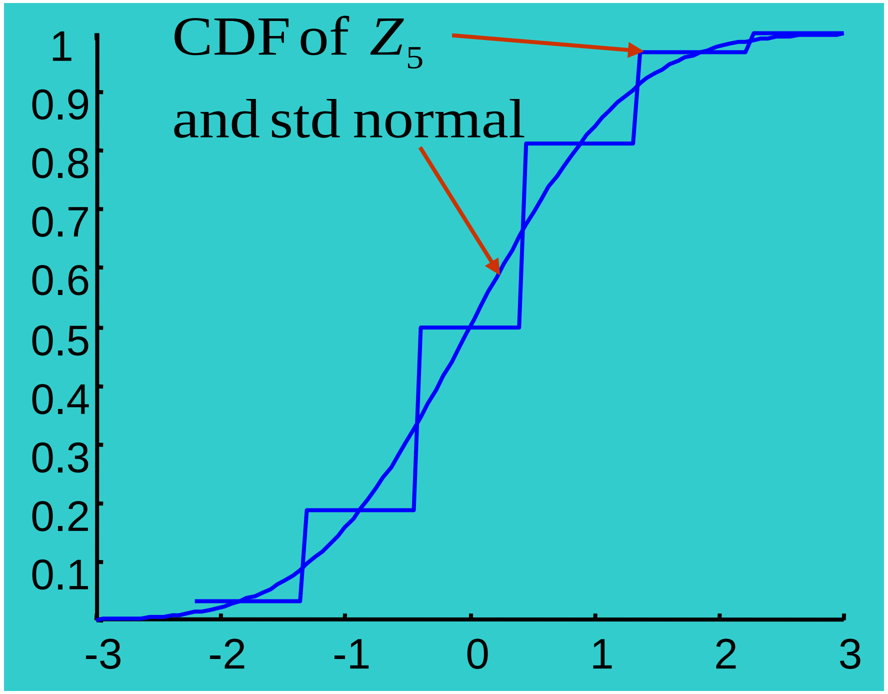
Graphical Illustration

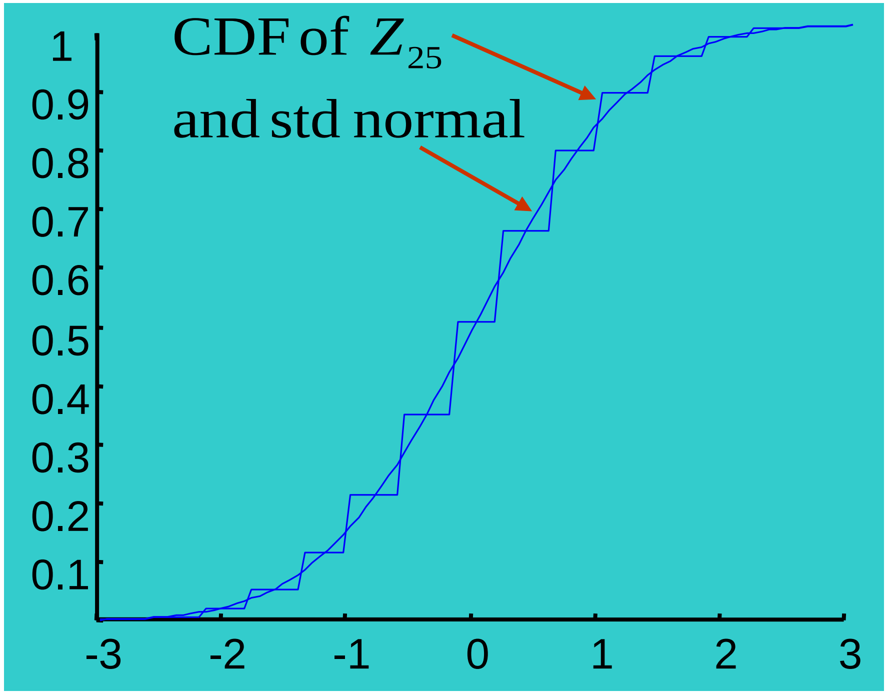
$$S_5 = \sum_{i=1}^{5} X_i$$

$$Z_5 = \frac{S_5 - 5\mu}{\sqrt{5\sigma^2}}$$

$$S_{25} = \sum_{i=1}^{25} X_i$$

$$Z_{25} = \frac{S_{25} - 25\mu}{\sqrt{25\sigma^2}}$$





Thank You!