Communications & Networking Laboratory

ECE 528 – Introduction to Random Processes in ECE
Lecture 16: Sums of Random Variables
Laws of Large Numbers

Bijan Jabbari, PhD
Dept. of Electrical and Computer Eng.
George Mason University
Fairfax, VA 22030-4444, USA
bjabbari@gmu.edu
http://cnl.gmu.edu/bjabbari

December 2, 2019

Note

- These slides cover material partially presented in class. They are provided to help students to follow the textbook. The material here are partly taken from the book by A. Leon-Garcia, Probability and Random Processes for Electrical Engineering, 3rd edition, whom I am thankful.
- There are many other topics which have been covered in class using the blackboard as step-by-step derivation and detailed discussions were needed.

Sums of Random Variables

- Let X_1, X_2,…, X_n be a sequence of random variables, and let S_n be their sum:

$$S_n = X_1 + X_2 + \ldots + X_n$$

- S_n is a sequence of random variables
- What happens to CDF of S_n as n grows?
- How does sequence of S_n behave with n?

Mean of Sum of Random Variables

Regardless of statistical dependence

$E[X_1 + X_2 + \ldots + X_n] =$

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

Variance of a Sum of RVs

$\mathrm{VAR}(X_1 + X_2 + \ldots + X_n) =$

$$\begin{aligned}
\mathrm{VAR}(X_1 + X_2 + \cdots + X_n) &= E\left\{\sum_{j=1}^{n}(X_j - E[X_j])\sum_{k=1}^{n}(X_k - E[X_k])\right\} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}E[(X_j - E[X_j])(X_k - E[X_k])] \\
&= \sum_{k=1}^{n}\mathrm{VAR}(X_k) + \sum_{\substack{j=1 \\ j \neq k}}^{n}\sum_{k=1}^{n}\mathrm{COV}(X_j, X_k).
\end{aligned}$$

- If $X_1 + X_2 + \ldots + X_n$ are independent random variables then $\mathrm{COV}(X_j, X_k) = 0$ for $j \neq k$ and

$\mathrm{VAR}(X_1 + X_2 + \ldots + X_n) = \mathrm{VAR}(X_1) + \ldots + \mathrm{VAR}(X_n)$

Sum of IID RVs

- The mean and variance of the sum of *n* **independent, identically distributed** (iid) random variables are:

$$E[S_n] = E[X_1] + \ldots + E[X_n] = n\mu$$

$$VAR[S_n] = n\ VAR[X_j] = n\sigma^2$$

Sum of *n* Independent RVs

- Let $X_1, X_2, \ldots, X_n$ be n independent RVs.
- Then $S_n = X_1 + X_2 + \ldots + X_n$ has characteristic function:

$$
\begin{aligned}
\Phi_Z(\omega) &= E[e^{j\omega Z}] \\
&= E[e^{j\omega(X+Y)}] \\
&= E[e^{j\omega X} e^{j\omega Y}] \\
&= E[e^{j\omega X}] E[e^{j\omega Y}] \\
&= \Phi_X(\omega)\Phi_Y(\omega),
\end{aligned}
$$

Sample Mean

- Suppose X is a RV for which the mean $E[X] = \mu$ is unknown.
- $X_1, \ldots, X_n$ denote n independent, repeated measurements of X.
- The **sample mean** is used to estimate $E[X]$:

$$M_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

- How good is M_n as estimator for E[X]?
- What happens to M_n as n becomes large?

Mean & Variance of Sample Mean

$$E[M_n] = E\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{1}{n}\sum_{j=1}^{n} E[X_j] = \mu,$$

- Sample mean is an **unbiased estimator** for μ

$$\mathrm{VAR}[M_n] = \frac{1}{n^2}\mathrm{VAR}[S_n] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

- Variance of sample mean decreases with n.

Weak Law of Large Numbers

$$P[|M_n - E[M_n]| \geq \varepsilon] \leq \frac{VAR[M_n]}{\varepsilon^2}$$

- For any choice of error ε and probability 1 – δ, can select the number of samples n so that M_n is within ε of the true mean with probability 1 – δ or greater.

Let $X_1, \ldots, X_n$ be a sequence of iid RVs with finite mean E[X] = μ, then for ε > 0,

$$\lim_{n \to \infty} P[|M_n - \mu| < \varepsilon] = 1$$

Example: What Weak Law Says

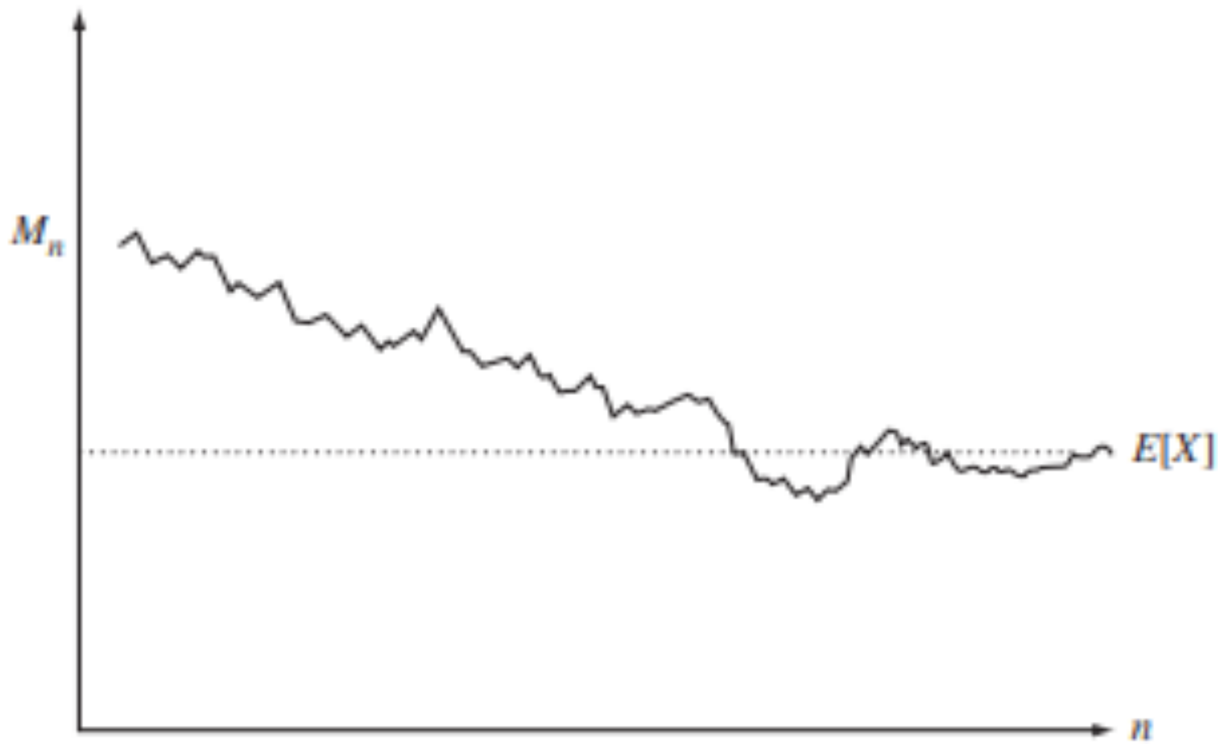


FIGURE 7.1
Convergence of sequence of sample means to $E[X]$.

Strong Law of Large Numbers

Let $X_1, \ldots, X_n$ be a sequence of iid RVs with finite mean $E[X] = \mu$, then for $\varepsilon > 0$,

$$P[\lim_{n\to\infty} \left| M_n - \mu \right| < \varepsilon] = 1$$

Gaussian RVs: They' re Everywhere!

- In nature, many macroscopic phenomena result from summation of numerous independent, microscopic processes.
- In many man-made problems, the averages often consist of the sum of independent RVs.
  - Let $X_1, X_2, \ldots, X_n$ be a sequence of iid RVs with finite mean μ and finite variance σ^2, and let S_n be their sum:

    $S_n = X_1 + X_2 + \ldots + X_n$

  - As n becomes large the cdf of a properly normalized S_n approaches that of a Gaussian RV.

Central Limit Theorem

- Let Z_n be the zero-mean, unit-variance random variable defined by

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Central Limit Theorem

$$\lim_{n\to\infty} P[Z_n \le z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

- The summands X_j can have any distribution as long as they have a finite mean and finite variance.

Proof of Central Limit Theorem

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu)$$

$$\begin{aligned}
\Phi_{Z_n}(\omega) &= E[e^{j\omega Z_n}] \\
&= E\left[\exp\left\{\frac{j\omega}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu)\right\}\right] \\
&= E\left[\prod_{k=1}^{n} e^{j\omega(X_k-\mu)/\sigma\sqrt{n}}\right] \\
&= \prod_{k=1}^{n} E\left[e^{j\omega(X_k-\mu)/\sigma\sqrt{n}}\right] \\
&= \left\{E\left[e^{j\omega(X-\mu)/\sigma\sqrt{n}}\right]\right\}^n
\end{aligned}$$

Central Limit Theorem (cont' d)

$$E\left[e^{j\omega(X-\mu)/\sigma\sqrt{n}}\right]$$

$$= E\left[1+\frac{j\omega}{\sigma\sqrt{n}}(X-\mu)+\frac{(j\omega)^2}{2!n\sigma^2}(X-\mu)^2+R(\omega)\right]$$

$$= 1+\frac{j\omega}{\sigma\sqrt{n}}E\left[(X-\mu)\right]+\frac{(j\omega)^2}{2!n\sigma^2}E\left[(X-\mu)^2\right]+E\left[R(\omega)\right]$$

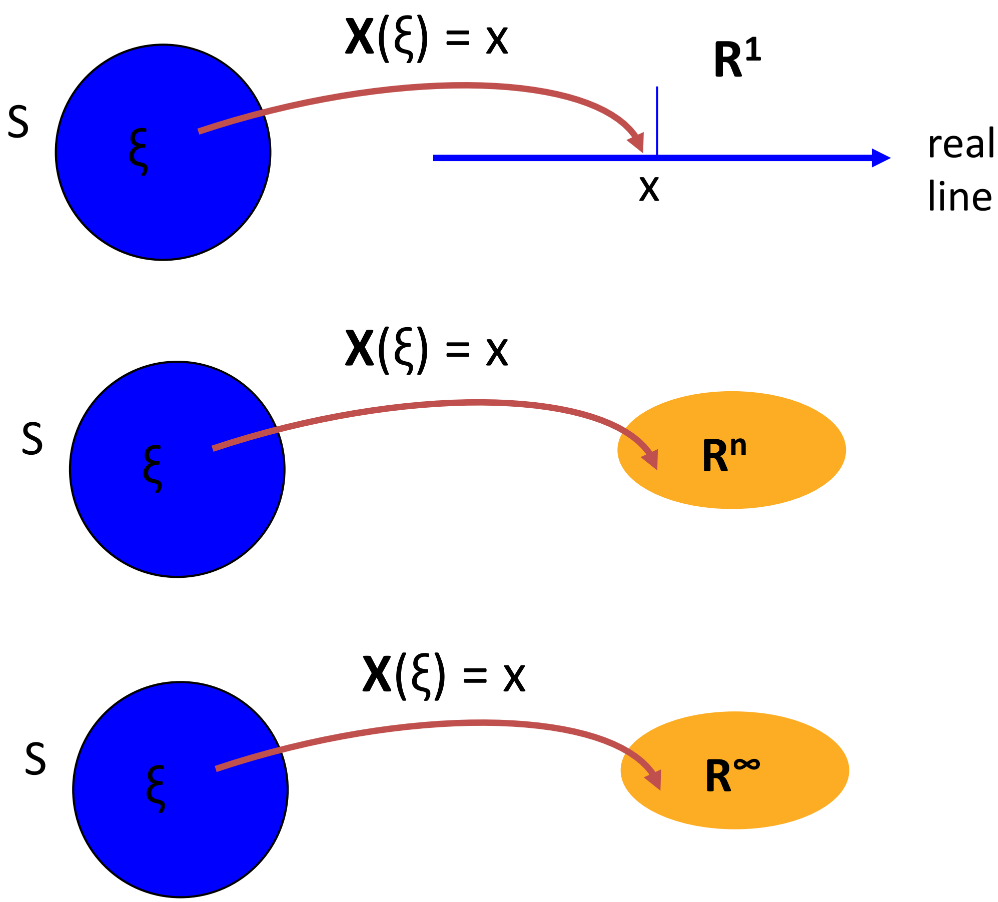
$$E\left[e^{j\omega(X-\mu)/\sigma\sqrt{n}}\right]=1-\frac{\omega^2}{2n}+E\left[R(\omega)\right]$$

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

Convergence of Sequences of Random Variables

Random Processes for
Engineering Applications

Random Variables, Vectors, and Sequences as Mappings



$\mathbf{X}(\xi) = x$

$\mathbf{X}(\xi) = (x_1, x_{2,\ldots}, x_n)$

$\mathbf{X}(\xi) = (x_1, x_2, \ldots, x_{n,}\ldots)$

- Mapping is deterministic
- Randomness is induced by ξ

Sequences of Random Variables

- Vector RV $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a function that assigns a vector of real values to each outcome ξ from some sample space S:

$$\mathbf{X}(\xi) = \left(X_1(\xi), X_2(\xi), \ldots, X_n(\xi)\right)$$

  - Randomness in **X** induced by randomness in the probability law governing selection of ξ.
  - A **sequence of random variables X** is a function that assigns a countably infinite number of real values to each outcome ξ from sample space S:

$$\mathbf{X}(\xi) = \left(X_1(\xi), X_2(\xi), \ldots, X_n(\xi), \ldots\right)$$

Convergence of Sequences of RVs

- We are interested in a sequence of random variables (usually not iid) X_1, X_2,… that converges to some random variable X:

$$X_n \to X \qquad \text{as } n \to \infty$$

- What does convergence mean?

Example: Explicit Mapping of ξ

- Outcome ξ selected at random from the interval S = [0,1].

$$V_n(\xi) = \xi\left(1 - \frac{1}{n}\right)$$

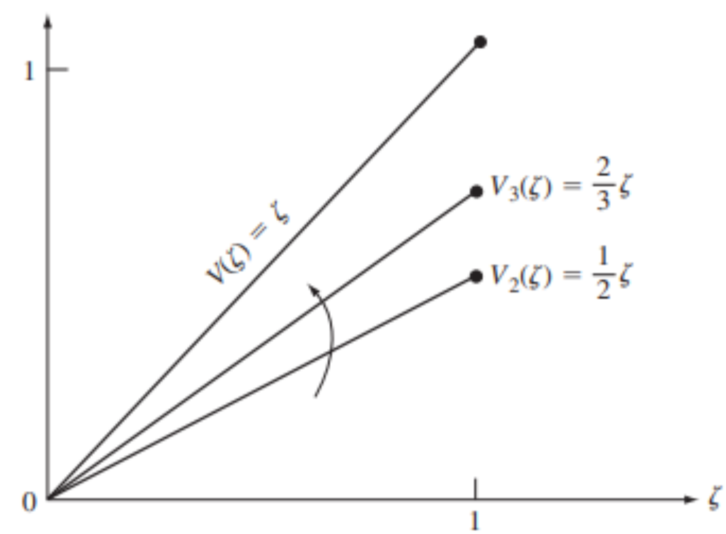
- Sequences of functions of ξ

Example: Explicit Mapping of ξ

- Outcome ξ selected at random from the interval S = [0,1].

$$V_n(\xi) = \xi\left(1 - \frac{1}{n}\right)$$

- Sequences of random variables



Sequence of random variables as a sequence of functions of ζ

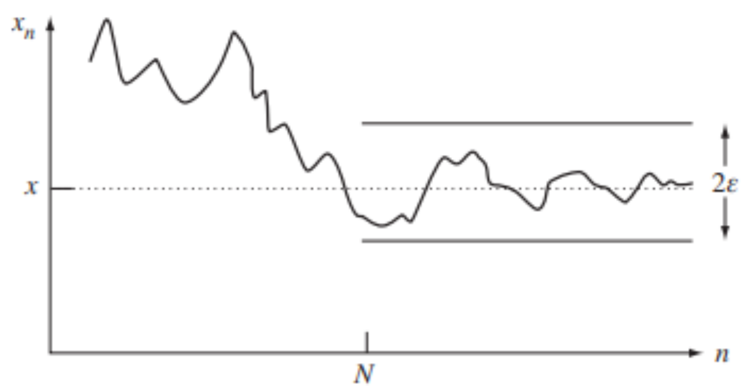
Example: Urn Experiment

- Urn has 2 black balls and 2 white balls.
- At time n a ball is selected at random and color is noted.
- If # of balls of this color > # of balls of other color, the ball is put back; otherwise ball is left out.
- Let $X_n(\xi)$ be # of black balls after nth draw.

- What is behavior of outcomes as n becomes large?

Convergence of Sequences of Numbers

- Suppose that each point in S, say ξ, produces a particular sequence of real numbers, $x_1, x_2, \ldots x_n$
- The sequence x_n converges to x if, given any $\varepsilon > 0$, we can specify an integer N such that for all values of n beyond N we can guarantee that $|x_n - x| < \varepsilon$.



Convergence of a sequence of numbers

Convergence of Sequences of Numbers II

- Cauchy criterion:

  The sequence x_n converges if and only if, given $\varepsilon > 0$, we can specify an integer N′ such that for m and n greater than N′, $|x_n - x_m| < \varepsilon$.

Types of Convergence

- **Sure Convergence:** The sequence of random variables $\{X_n(\xi)\}$ converges surely to the random variable $X(\xi)$ if the sequence of functions $X_n(\xi)$ converges to the function $X(\xi)$ as $n \to \infty$ ***for all ξ in S***:

$$X_n(\xi) \to X(\xi) \quad \text{as } n \to \infty \qquad \text{for all } \xi \in S$$

- **Almost-Sure Convergence:** The sequence of random variables $\{X_n(\xi)\}$ converges almost surely to the random variable $X(\xi)$ if the sequence of functions $X_n(\xi)$ converges to the function $X(\xi)$ as $n \to \infty$ for all ξ in S, except possibly on a set of probability zero; that is,

$$P[\xi : X_n(\xi) \to X(\xi) \quad \text{as } n \to \infty] = 1$$

Example: Explicit Mapping of ξ

$$V_n(\xi) = \xi\left(1 - \frac{1}{n}\right) \qquad\qquad V(\xi) = \xi$$

Example 7.17

Let $V_n(\zeta)$ be the sequence of random variables from Example 7.16. Does the sequence of real numbers corresponding to a fixed ζ converge?

From Fig. 7.8(a), we expect that for a fixed value ζ, $V_n(\zeta)$ will converge to the limit ζ. Therefore, we consider the difference between the nth number in the sequence and the limit:

$$|V_n(\zeta) - \zeta| = \left|\zeta\left(1 - \frac{1}{n}\right) - \zeta\right| = \left|\frac{\zeta}{n}\right| < \frac{1}{n},$$

where the last inequality follows from the fact that ζ is always less than one. In order to keep the above difference less than ε, we choose n so that

$$|V_n(\zeta) - \zeta| < \frac{1}{n} < \varepsilon;$$

that is, we select $n > N = 1/\varepsilon$. Thus the sequence of real numbers $V_n(\zeta)$ converges to ζ.

Example: Urn Experiment

- Urn has 2 black balls and 2 white balls.
- At time n a ball is selected at random.
- If # of balls of this color > # of balls of other color, ball put back; otherwise ball left out.
- Let $X_n(\xi)$ be # of black balls after nth draw.
- Does sequence of random variables converge?

Sequences of Random Variables

- A **sequence of random variables X** is a function that assigns a countably infinite number of real values to each outcome ξ from sample space S:

$$\mathbf{X}(\xi) = \left(X_1(\xi), X_2(\xi), \ldots, X_n(\xi), \ldots\right)$$

- **Almost-Sure Convergence:** The sequence of random variables $\{X_n(\xi)\}$ converges almost surely to the random variable $X(\xi)$ if the sequence of functions $X_n(\xi)$ converges to the function $X(\xi)$ as $n \to \infty$ for all ξ in S, except possibly on a set of probability zero; that is,

$$P[\xi : X_n(\xi) \to X(\xi) \quad \text{as } n \to \infty] = 1$$

Mean Square Convergence

- The sequence of random variables $\{X_n(\xi)\}$ converges in the mean square sense to the random variable $X(\xi)$ if

$$E[(X_n(\xi) \to X(\xi))^2] \to 0 \quad \text{as } n \to \infty$$

  - Denoted by (**l**imit **i**n the **m**ean):

$$\text{l.i.m. } X_n(\xi) \to X(\xi) \quad \text{as } n \to \infty$$

  - Mean square convergence occurs if the second moment of the error approaches zero as n → ∞.
  - Implies that as n increases, an increasing proportion of sample sequences are close to X.

Convergence in Probability

- The sequence of random variables $\{X_n(\xi)\}$ converges in probability to the random variable X (ξ) if, for any $\varepsilon > 0$,

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

- Mean square convergence implies convergence in probability since

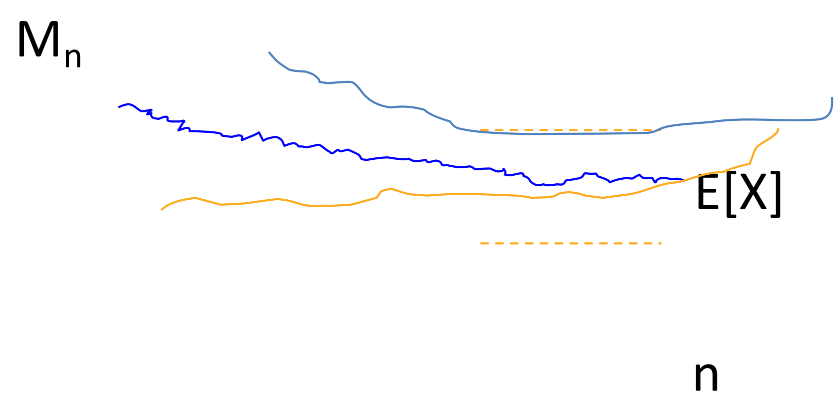
$$P[|X_n(\xi) - X(\xi)| > \varepsilon] = P[|(X_n(\xi) - X(\xi))^2| > \varepsilon^2] < \frac{E[(X_n - X)^2)}{\varepsilon^2}$$

Weak Law of Large Numbers

- Let $X_1,\ldots, X_n$ be a sequence of iid RVs with finite mean E[X] = μ and finite variance, then for large n, the sample mean is close to E[X] with high probability

$$P[\,|\,M_n = \mu\,| < \varepsilon] > 1 - \delta \quad \text{for } n > n_0$$

- States that most sample sequences are close to E[X], but not that they necessarily remain close.



Example: Mean-square and Almost-sure Convergence

- Let $R_n(\xi)$ be the error introduced by a communication channel in nth transmission.
- Errors introduced as follows:
  - 1st transmission channel introduces an error
  - In next 2 transmissions, channel randomly selects one transmission to introduce an error. Other transmission is error free.
  - In next 3 transmissions, only one has error…
- Does the sequence of transmission errors converge, and if so, in what sense?

Example: Mean-square and Almost-sure Convergence (cont'd)

Example 7.21

Does the sequence $V_n(\zeta)$ in Example 7.18 converge in the mean square sense?

In Example 7.18, we found that $V_n(\zeta)$ converges surely to ζ. We therefore consider

$$E[(V_n(\zeta) - \zeta)^2] = E\left[\left(\frac{\zeta}{n}\right)^2\right] = \int_0^1 \left(\frac{\zeta}{n}\right)^2 d\zeta = \frac{1}{3n^2},$$

where we have used the fact that ζ is uniformly distributed in the interval $[0, 1]$. As n approaches infinity, the mean square error approaches zero, and so we have convergence in the mean square sense.

Convergence in Distribution

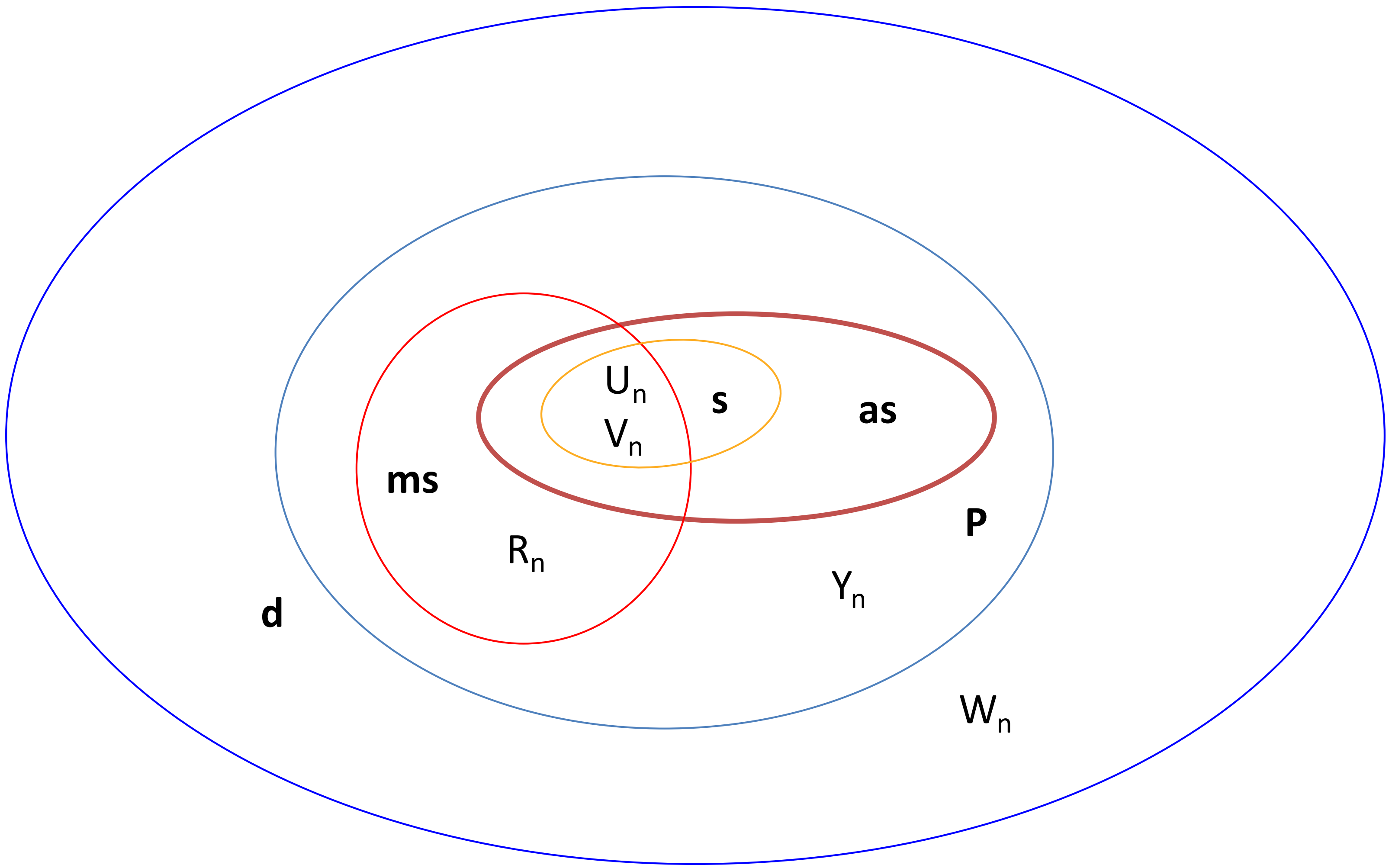
- The sequence of random variables X_n with cdfs $\{F_n(x)\}$ converges in distribution to the random variable X with cumulative distribution F(x) if

$$F_n(x) \to F(x) \quad \text{as } n \to \infty$$

for all x at which the cdf is continuous.

- Addresses convergence of cdf' s, not of RVs
- Convergence in probability implies convergence in distribution
- Central limit theorem is an example of convergence in distribution.

Convergence Types



Lecture Summary

- Convergence in mean square sense and in probability do not address the convergence behavior of entire sequences, but rather the behavior of the ensemble of sequences at a large values of *n*.
- Mean square sense convergence implies convergence in probability
- Mean square convergence does not imply convergence almost surely and vice versa.
- Convergence in distribution does not address the behavior of sequences of random variables, but rather of their distribution functions.