Lecture Notes 7
Convergence and Limit Theorems

- Motivation
- Convergence with Probability 1
- Convergence in Mean Square
- Convergence in Probability, WLLN
- Convergence in Distribution, CLT

Motivation

- One of the key questions in statistical signal processing is how to estimate the *statistics* of a r.v., e.g., its mean, variance, distribution, etc.

  To estimate such a statistic, we collect *samples* and use an *estimator* in the form of a *sample average*

  - How good is the *estimator*? Does it "converge" to the true statistic?
  - How many samples do we need to ensure with some *confidence* that we are within a certain range of the true value of the statistic?

- Another key question in statistical signal processing is how to estimate a signal from noisy observations, e.g., using MSE or linear MSE

  - Does the estimator converge to the true signal?
  - How many observations do we need to achieve a desired estimation accuracy?

- The subject of convergence and limit theorems for r.v.s addresses such questions

Example: Estimating the Mean of a R.V.

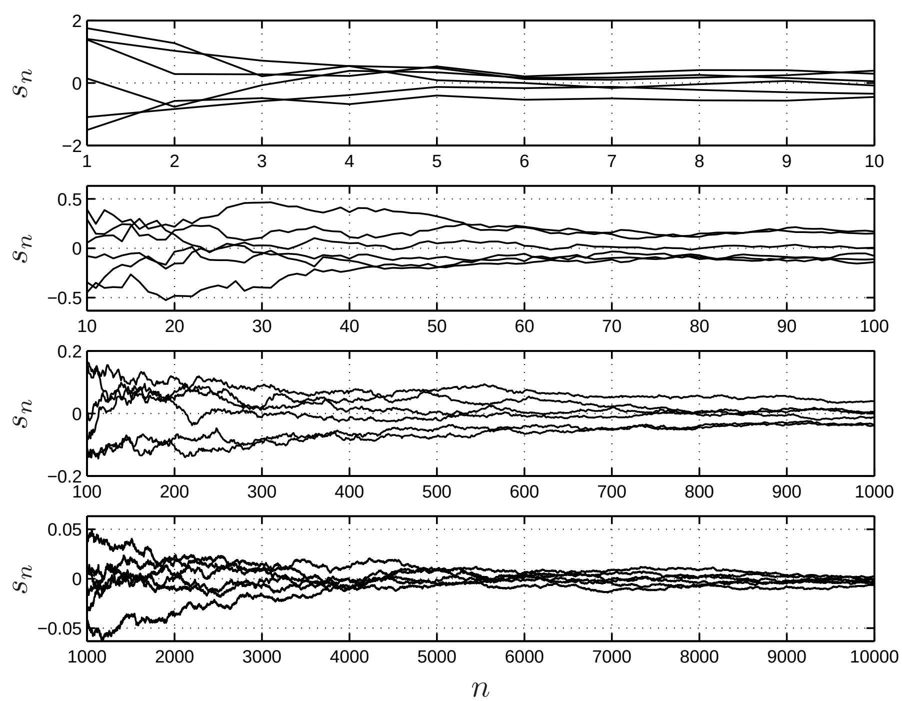
- Let X be a r.v. with finite but unknown mean $\mathrm{E}(X)$
- To estimate the mean we generate $X_1, X_2, \ldots, X_n$ i.i.d. samples drawn according to the same distribution as X and compute the *sample mean*

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

- Does S_n converge to $\mathrm{E}(X)$ as we increase n? If so, how fast?

  But what does it mean to say that a r.v. sequence S_n converges to $\mathrm{E}(X)$?
- First we give an example: Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\mathcal{N}(0,1)$
  - We use MATLAB to generate 6 sets of outcomes of $X_1, \ldots, X_n$
  - We then plot s_n for the 6 sets of outcomes as a function of n
  - Note that each s_n sequence appears to be converging to 0, the mean of the r.v., as n increases

Plots of Sample Sequences of S_n



Convergence With Probability 1

- Recall that a sequence of numbers $x_1, x_2, \ldots, x_n, \ldots$ converges to x if for every $\epsilon > 0$, there exists an $n(\epsilon)$ such that $|x_n - x| < \epsilon$ for every $n \ge n(\epsilon)$
- Now consider a sequence of r.v.s $X_1, X_2, \ldots, X_n, \ldots$ all defined on the same probability space Ω. For every $\omega \in \Omega$ we obtain a *sample sequence* (sequence of numbers) $X_1(\omega), X_2(\omega), \ldots, X_n(\omega), \ldots$
- A sequence $X_1, X_2, X_3, \ldots$ of r.v.s is said to converge to random variable X *with probability 1* (w.p.1) if

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

- This means that the set of sample paths that converge to $X(\omega)$, in the sense of a sequence converging to a limit, has probability 1
- Equivalently, $X_1, X_2, \ldots, X_n, \ldots$ converges w.p.1 if for every $\epsilon > 0$,

$$\lim_{m\to\infty} \mathrm{P}\{|X_n - X| < \epsilon \text{ for every } n \ge m\} = 1$$

- Example 1: Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\mathrm{Bern}(1/2)$, and define $Y_n = 2^n \prod_{i=1}^n X_i$. Show that the sequence Y_n converges to 0 w.p.1

  Solution: To show this, let $\epsilon > 0$ (and $\epsilon < 2^m$), and consider

$$\begin{aligned}\mathrm{P}\{|Y_n - 0| < \epsilon \text{ for all } n \ge m\} &= \mathrm{P}\{X_n = 0 \text{ for some } n \le m\}\\ &= 1 - \mathrm{P}\{X_n = 1 \text{ for all } n \le m\}\\ &= 1 - (\tfrac{1}{2})^m \ \to 1 \text{ as } m \to \infty\end{aligned}$$

- An important example of convergence w.p.1: the *Strong Law of Large Numbers* (SLLN), which says that if $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. with finite mean $\mathrm{E}(X)$, then the sequence of sample means $S_n \to \mathrm{E}(X)$ w.p.1
  - The previous MATLAB example is a good demonstration of the SLLN — each of the 6 sample paths appears to be converging to 0, which is $\mathrm{E}(X)$
  - The proof of the SLLN and other convergence w.p.1 results are beyond the scope of this course. Take Stats 310 if you want to learn a *lot* more about this

Convergence in Mean Square

- A sequence of r.v.s $X_1, X_2, \ldots, X_n, \ldots$ converges to a random variable X *in mean square* (m.s.) if

$$\lim_{n\to\infty} \mathrm{E}\left[(X_n - X)^2\right] = 0$$

- Example: *Estimating the mean.*

  Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. with finite mean $\mathrm{E}(X)$ and variance $\mathrm{Var}(X)$. Then $S_n \to \mathrm{E}(X)$ in m.s

- Proof: Here we need to show that

$$\lim_{n\to\infty} \mathrm{E}\left[(S_n - \mathrm{E}(X))^2\right] = 0$$

  First note that

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

  So, S_n is an *unbiased* estimate of $\mathrm{E}(X)$

Now to prove convergence in m.s., consider

$$\begin{aligned}
\mathrm{E}\left[(S_n - \mathrm{E}(X))^2\right] &= \mathrm{E}\left[(S_n - \mathrm{E}(S_n))^2\right] \\
&= \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X)\right)^2\right] \\
&= \frac{1}{n^2}\mathrm{E}\left[\left(\sum_{i=1}^{n}(X_i - \mathrm{E}(X))\right)^2\right] \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i)\right) \quad \text{since } \{X_i\} \text{ are independent} \\
&= \frac{1}{n^2}\left(n\mathrm{Var}(X)\right) \\
&= \frac{1}{n}\mathrm{Var}(X) \to 0 \text{ as } n \to \infty
\end{aligned}$$

- Note that the proof works even if the r.v.s are only pairwise independent or even only uncorrelated

- Example: Consider the best linear MSE estimates found in the last example of Lecture Notes 6 as a sequence of r.v.s $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n, \ldots$, where $\hat{X}_n$ is the best linear estimate of X given the first n observations. This sequence converges in m.s. to X since $\mathsf{MSE}_n \to 0$
- Convergence in m.s. does not necessarily imply convergence w.p.1
- Example 2: Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent r.v.s such that

$$X_n = \begin{cases} 0 & \text{with probability } 1 - \frac{1}{n} \\ 1 & \text{with probability } \frac{1}{n} \end{cases}$$

Clearly this sequence converges to 0 in m.s., but does it converge w.p.1?

It actually does not, since for $0 < \epsilon < 1$ and any m

$$\begin{aligned}
\mathrm{P}\{|X_n - 0| < \epsilon \text{ for all } n \geq m\} &= \lim_{n\to\infty} \prod_{i=m}^{n} \left(1 - \frac{1}{i}\right) \\
&= \lim_{n\to\infty} \prod_{i=m}^{n} \left(\frac{i-1}{i}\right) \\
&= \lim_{n\to\infty} \frac{(m-1)}{m}\frac{m}{(m+1)} \cdots \frac{(n-1)}{n} \\
&= \lim_{n\to\infty} \frac{m-1}{n} \to 0 \neq 1
\end{aligned}$$

- Also convergence w.p.1 does not imply convergence in m.s.; for example, consider the sequence in Example 1. Since

$$\mathrm{E}\left[(Y_n - 0)^2\right] = \left(\tfrac{1}{2}\right)^n 2^{2n} = 2^n ,$$

the sequence does not converge in m.s. even though it converges w.p.1

Convergence in Probability

- A sequence of r.v.s $X_1, X_2, \ldots, X_n, \ldots$ converges to a r.v. X *in probability* if for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\{|X_n - X| < \epsilon\} = 1$$

- Convergence w.p.1 implies convergence in probability. The converse is not necessarily true, so convergence w.p.1 is stronger than in probability
- Example 3: Let $X_1, X_2, \ldots, X_n, \ldots$ be independent such that

$$X_n = \begin{cases} 0 & \text{with probability } 1 - \frac{1}{n} \\ n & \text{with probability } \frac{1}{n} \end{cases}$$

Clearly, this sequence converges in probability to 0, since

$$\mathrm{P}\{|X_n - 0| > \epsilon\} = \mathrm{P}\{X_n > \epsilon\} = \frac{1}{n} \to 0 \text{ as } n \to \infty$$

But does it converge w.p.1? The answer is no (see Example 2)

- Convergence in m.s. implies convergence in probability. To show this we use the Markov inequality. For any $\epsilon > 0$,

$$P\{|X_n - X| > \epsilon\} = P\{(X_n - X)^2 > \epsilon^2\} \le \frac{E(X_n - X)^2}{\epsilon^2}$$

  If $X_n \to X$ in m.s., then

$$\lim_{n\to\infty} E\left[(X_n - X)^2\right] = 0 \;\Rightarrow\; \lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0\,,$$

  i.e., $X_n \to X$ in probability

- The converse is not necessarily true. In Example 3, X_n converges in probability. Now consider

$$E\left[(X_n - 0)^2\right] \;=\; 0 \cdot \left(1 - \frac{1}{n}\right) + n^2 \cdot \frac{1}{n} = n \to \infty \text{ as } n \to \infty$$

  Thus X_n does not converge in m.s.

- So convergence in probability is weaker than both convergence w.p.1 and in m.s.

The Weak Law of Large Numbers

- The WLLN states that if $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of i.i.d. r.v.s with finite mean $E(X)$ and variance $\mathrm{Var}(X)$, then

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to E(X) \text{ in probability}$$

- We already proved that $S_n \to E(X)$ in m.s., and since convergence in m.s. implies convergence in probability, $S_n \to E(X)$ in probability

  So, WLLN requires only uncorrelation of the r.v.s (SLLN requires independence)

- Confidence intervals: Given $\epsilon, \delta > 0$, how large should n, the number of samples, be so that

$$P\{|S_n - E(X)| \le \epsilon\} \ge 1 - \delta\,,$$

  i.e., S_n is within $\pm\,\epsilon$ of $E(X)$ with probability $\ge 1 - \delta$?

We can use the Chebyshev inequality:

$$\begin{aligned}\mathrm{P}\{|S_n - \mathrm{E}(X)| \le \epsilon\} &= \mathrm{P}\{|S_n - \mathrm{E}(S_n)| \le \epsilon\} \\ &\ge 1 - \frac{\mathrm{Var}(S_n)}{\epsilon^2} = 1 - \frac{\mathrm{Var}(X)}{n\epsilon^2}\end{aligned}$$

So n should be large enough that

$$\frac{\mathrm{Var}(X)}{n\epsilon^2} \le \delta \Rightarrow n \ge \frac{\mathrm{Var}(X)}{\delta\epsilon^2}$$

- Example: Let $\epsilon = 0.1\sigma_X$ and $\delta = 0.001$. The number of samples should satisfy

$$n \ge \frac{\sigma_X^2}{0.001 \times 0.01\sigma_X^2} = 10^5 ,$$

i.e., 10^5 samples ensure that S_n is within $\pm 0.1\sigma_X$ of $\mathrm{E}(X)$ with probability ≥ 0.999, *independent* of the distribution of X

Example of Convergence to a Random Variable

- Except for the example of a sequence of linear MSE estimates, the sequences of r.v.s in our examples converged to constants
- Here is another example where a sequence of r.v.s converges in m.s. to a r.v:

Flip a coin with random bias P conditionally independently to obtain the sequence $X_1, X_2, \ldots, X_n, \ldots$, where as usual $X_i = 1$ if the ith coin flip is heads and $X_i = 0$ otherwise

As we already know, the r.v.s $X_1, X_2, \ldots, X_n$ are not independent, but given $P = p$ they are i.i.d. $\mathrm{Bern}(p)$

It is easy to show using conditional expectation that $\mathrm{E}(S_n) = \mathrm{E}(X_1) = \mathrm{E}(P)$

In a homework exercise, you will show that $S_n \to P$ (not to $\mathrm{E}(P)$) in m.s.

Convergence in Distribution

- A sequence of r.v.s $X_1, X_2, \ldots, X_n, \ldots$ converges *in distribution* to a r.v. X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for every x at which $F_X(x)$ is continuous
- Convergence in probability implies convergence in distribution — so convergence in distribution is the weakest form of convergence we discuss
- The most important example of convergence in distribution is the Central Limit Theorem (CLT). Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. r.v.s with finite mean $\mathrm{E}(X)$ and variance σ_X^2. Consider the *normalized* sum

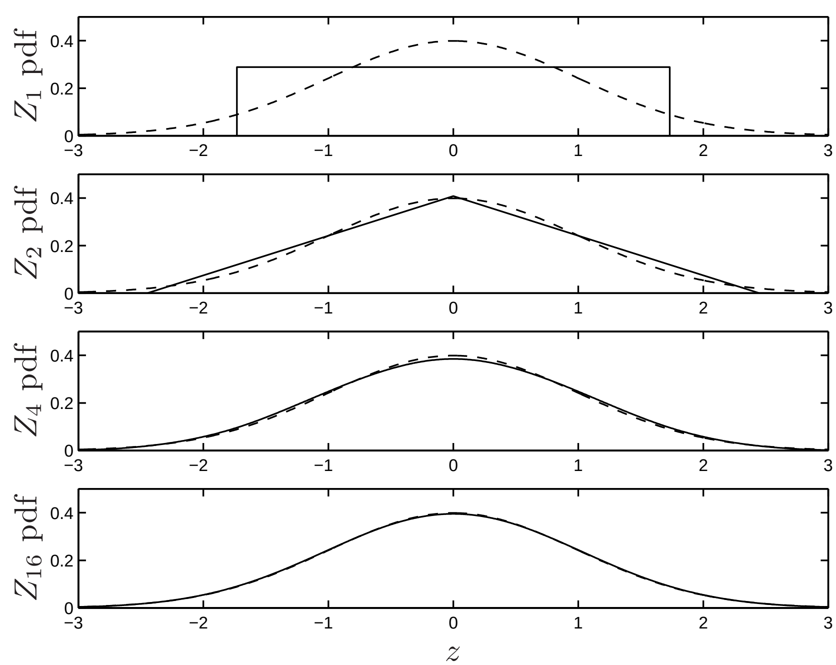
  $$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mathrm{E}(X)}{\sigma_X}$$

  The sum is called normalized because $\mathrm{E}(Z_n) = 0$ and $\mathrm{Var}(Z_n) = 1$

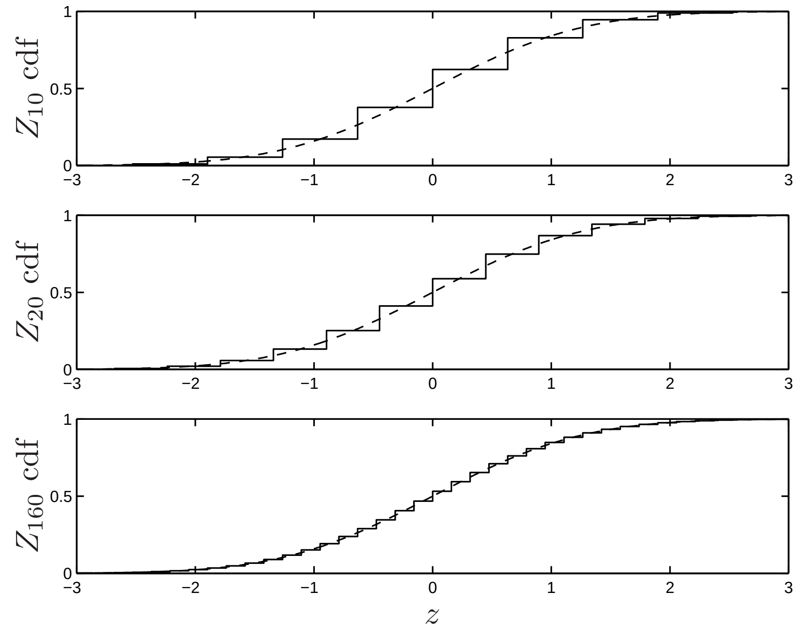
  The *Central Limit Theorem* states that $Z_n \to Z \sim \mathcal{N}(0,1)$ in distribution, i.e.,

  $$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z) = \begin{cases} 1 - Q(z) & z \ge 0 \\ Q(-z) & z < 0 \end{cases}$$

- Example: Let $X_1, X_2, \ldots$ be i.i.d. $\mathrm{U}[-1,1]$ r.v.s. The normalized sum is $Z_n = \sum_{i=1}^{n} X_i / \sqrt{n/3}$. The following plots show the pdf of Z_n for $n = 1, 2, 4, 16$. Note how quickly the pdf of Z_n approaches the Gaussian.



- Example: Let $X_1, X_2, \ldots$ be i.i.d. $\mathrm{Bern}(\frac{1}{2})$. The normalized sum is $Z_n = \sum_{i=1}^{n}(X_i - 0.5)/\sqrt{n/4}$. The following plots show the cdf of Z_n for $n = 10, 20, 160$. Z_n is discrete and thus has no pdf, but its cdf converges to the Gaussian cdf.



Application: Confidence Intervals

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. with finite mean $\mathrm{E}(X)$ and variance $\mathrm{Var}(X)$ and let S_n be the sample mean
- Given ϵ, $\delta > 0$, how large should n, the number of samples, be so that
$$\mathrm{P}\{|S_n - \mathrm{E}(X)| \le \epsilon\} \ge 1 - \delta\,?$$
- We can use the CLT to find an estimate of n as follows:
$$\begin{aligned}
\mathrm{P}\{|S_n - \mathrm{E}(S_n)| \le \epsilon\} &= \mathrm{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathrm{E}(X))\right| \le \epsilon\right\} \\
&= \mathrm{P}\left\{\left|\frac{1}{\sigma_X\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathrm{E}(X))\right| \le \frac{\epsilon\sqrt{n}}{\sigma_X}\right\} \\
&\approx 1 - 2Q\left(\frac{\epsilon\sqrt{n}}{\sigma_X}\right)
\end{aligned}$$
- Example: For $\epsilon = 0.1\sigma_X$, $\delta = 0.001$, set $2Q(0.1\sqrt{n}) = 0.001$, so $0.1\sqrt{n} = 3.3$ or $n = 1089$ — much smaller than $n \ge 10^5$ obtained by the Chebyshev inequality

CLT for Random Vectors

- The CLT applies to i.i.d. sequences of random vectors
- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ be a sequence of i.i.d. k-dimensional random vectors with finite mean $\boldsymbol{\mu}$ and nonsingular covariance matrix Σ. Define the sequence of random vectors $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n, \ldots$ by

$$\mathbf{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\mu})$$

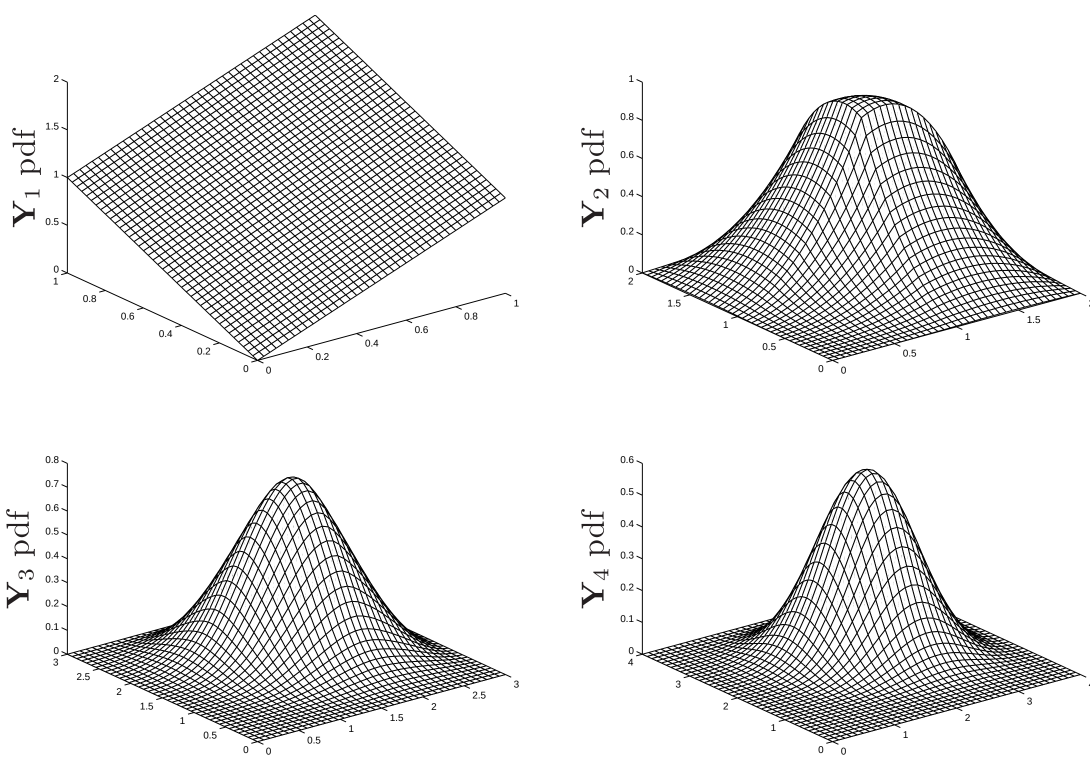
- The Central Limit Theorem for random vectors states that as $n \to \infty$

$$\mathbf{Z}_n \to \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma) \text{ in distribution}$$

- Example: Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ be a sequence of i.i.d. 2-dimensional random vectors with

$$f_{\mathbf{X}_1}(x_{11}, x_{12}) = \begin{cases} x_{11} + x_{12} & 0 < x_{11} < 1,\, 0 < x_{12} < 1 \\ 0 & \text{otherwise} \end{cases}$$

  The following plots show the joint pdf of $\mathbf{Y}_n = \sum_{i=1}^{n} \mathbf{X}_i$ for $n = 1, 2, 3, 4$. Note how quickly it looks Gaussian.



Relationships Between Types of Convergence

- The following figure summarizes the relationships between the different types of convergence presented in Lecture Notes 7