7.4 Stochastic Convergence and Limit Theorem

Equality vs. Approximate Equality of Two Real Numbers

- Equality of Two Real Numbers
  - Given two real numbers x and y, we have $x = y$ xor $x \neq y$.
  - Lemma. Two real numbers x and y, we have $x = y$ if and only if $x - y = 0$.
    * ... if and only if $x - y \leq 0$ and $x - y \geq 0$.
    * ... if and only if $|x - y| \leq 0$.
  - Proposition. Given two real numbers x and y, we have $x = y$ if and only if

$$|x - y| \leq \epsilon, \forall \epsilon > 0.$$

    * Proof by contraposition. $x \neq y$ iff $\exists \epsilon > 0, |x - y| > \epsilon$.
    * (o) ... if and only if $|x - y| < \epsilon, \forall \epsilon > 0$.
    * (x) ... if and only if $|x - y| < \epsilon, \forall \epsilon \geq 0$.
- Approximate Equality of Two Real Numbers
  - What about $x \approx y$ or $x - y \approx 0$?
  - Mathematical tools to handle ‘approximately equal’ or ’almost equal’ are
    * upper and lower bounding and

  * convergence of a sequence of real numbers

- Given a sequence $(x_n)_{n\geq 1}$ of real numbers and y, we may have the situation where
  - $x_n \neq y, \forall n$ but
  - we want to claim that $x_n \approx y$ for sufficiently large n.
    * Example. $x_n = e^{-n}$, $y = 0 \Longrightarrow x_n \neq y, \forall n$, but $x_n \approx y$ for sufficiently large n..
    * The notion of the convergence of a sequence of real numbers is to handle such an approximate equality mathematically precisely.
    * First, we need to review the convergence of a sequence of real numbers.

- Def. A sequence $(x_n)_{n\geq 1}$ of real numbers converges to a real number x if

$$\forall \varepsilon > 0, \exists N(\varepsilon) > 0, \text{ s.t. } n \geq N(\varepsilon) \Longrightarrow |x_n - x| \leq \varepsilon.$$

  - "For every epsilon positive,

    there exists N of epsilon positive such that

    the difference between x_n and x is less than epsilon

    for all n greater than or equal to N of epsilon."
  - x is called the limit of the sequence.
  - The definition can serve to check the convergence of a sequence only when the candidate limit is provided. $\rightarrow$ the Cauchy sequence in $\mathbb{R}$

- Notations
  * $\lim_{n\to\infty} x_n = x$
  * $x_n \to x$ as $n \to \infty$ (as n tends to infinity)
  * $x_n \to x$ (x_n converges to x)
  * $\forall \varepsilon > 0, |x_n - x| \leq \varepsilon$ for sufficiently large n
  * $x_n \approx x$ for sufficiently large n
- Properties
  * translation invariance: $x_n \to x$ iff $(x_n - x) \to 0$
  * linearity: $x_n \to x$ and $y_n \to y \Longrightarrow ax_n + by_n \to ax + by$
    · Consider $y_n = 1, \forall n$. Then, $x_n \to x \Longrightarrow ax_n + b \to ax + b$.

• Def. A sequence $(x_n)_n$ (not necessarily of real numbers) is called a Cauchy sequence if

$$\forall \varepsilon > 0, \exists N(\varepsilon) > 0, \text{ s.t. } m,n \geq N(\varepsilon) \Longrightarrow |x_n - x_m| \leq \varepsilon.$$

- Lemma. For a sequence $(x_n)_n$ of real numbers, it is Cauchy if and only if it converges.

  [Leon-Garcia] This lemma says that "the maximum variation in the sequence for points beyond $N(\varepsilon)$ is less than ε."

Equality vs. Approximate Equality of Two Real Functions

- Equality of Two Real Functions
  - Given two real functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$, we have $X = Y$ xor $X \neq Y$ on $\mathcal{D} \subset \mathbb{R}$.
  - Def. (everywhere equality) Given two real functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$, they are equal everywhere (e) on $\mathcal{D} \subset \mathbb{R}$ if

$$X(s) = Y(s), \forall s \in \mathcal{D}.$$

    * The equality of two functions involves the equality of two real numbers $X(s)$ and $Y(s)$ for each $s \in \mathcal{D}$.
    * Or it involves the equality of a real number $X(s) - Y(s)$ to zero for each $s \in \mathcal{D}$.
    * $\mathcal{D}$ could be $\mathbb{R}$ or an uncountable set. Then, the equality of two functions involves uncountably infinite number of equalities of two real numbers.
    * If there exists any single $s_o \in \mathcal{D}$ such that $X(s_0) \neq Y(s_0)$, then $X \neq Y$ everywhere on $\mathcal{D}$.
    * $X(s) = Y(s)$ can be replaced by $X(s) - Y(s) \leq 0, X(s) - Y(s) \geq 0$.
    * $X(s) = Y(s)$ can be replaced by $|X(s) - Y(s)| \leq \epsilon, \forall \epsilon > 0$.

* $X(s) = Y(s), \forall s \in \mathcal{D}$ can be replaced by $\{s \in \mathcal{D} : X(s) = Y(s)\} = \mathcal{D}$.
* Lemma. Two functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$ are equal everywhere iff

$$\{s \in \mathcal{D} : X(s) \neq Y(s)\} = \emptyset.$$

* iff $\{s \in \mathcal{D} : \forall \epsilon > 0, |X(s) - Y(s)| \leq \epsilon\} = \mathcal{D}$.
* This definition of 'equal everywhere' is too strong. Thus, people have introduced a weaker notion of the equality of two functions.

Introduction to real analysis

해석학

Measure Theory

Probability Theory I, II

Control

a decent Ph.D. in Sig. proc

Commun.

- Def. (almost everywhere equality) Given two functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$, they are equal almost everywhere (a.e.) if

$$m(\{s \in \mathcal{D} : X(s) \neq Y(s)\}) = 0,$$

where $m(\cdot)$ is the Borel measure.

  - The Borel measure $m(\cdot)$ gives the size of a real set.
    * $m([0, 1]) = 1$
    * $m([0, 1] \backslash \mathbb{Q}) = 1$

    Thus, $X = Y$ a.e., iff $X(s) \neq Y(s)$ only on a set of measure zero.
  - When $m(\mathcal{D}) < \infty$ a.e. equal iff $m(\mathcal{D}) = m(\{s \in \mathcal{D} : X(s) = Y(s)\})$.
  - Lemma. Two functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$ are equal almost everywhere (a.e.) iff

$$\forall \varepsilon > 0, m(\{s \in \mathcal{D} : |X(s) - Y(s)| > \varepsilon\}) = 0.$$

    * When $m(\mathcal{D}) < \infty$ a.e. equal iff $\forall \varepsilon > 0, m(\{s \in \mathcal{D} : |X(s) - Y(s)| \leq \varepsilon\}) = m(\mathcal{D})$.
  - Lemma. Two functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$ are equal almost everywhere (a.e.) iff

$$\int_{-\infty}^{\infty} |X(s) - Y(s)|^p ds = 0$$

    for some $p \geq 1$, where the integral is the Lebesgue integral.

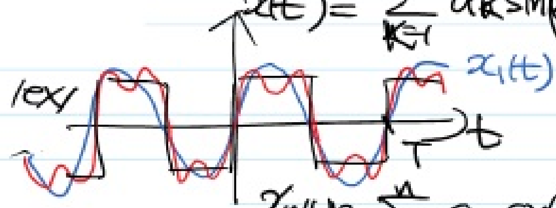
- Caution! The Def., Lemma, and Proposition have three equivalent expressions. However, the variations of them become no longer equivalent when the convergence of a sequence of functions is considered.
- Approximate Equality of Two Real Functions
  - What about $X(s) \approx Y(s)$?
  - Mathematical tools to handle 'approximately equal' or 'almost equal' are
    * upper and lower bounding and
    * convergence of a sequence of real functions
- Given a sequence $(X_n(s))_n$ of real functions and a real function $X(s)$, we may have the situation where
  - $X_n(s) \neq X(s), \forall n$ but
  - we want to claim that $X_n(s) \approx X(s)$ f.s.l. n.
  - Example. Given a square wave $X(s)$, the nth order Fourier series approximation $X_n(s)$ satisfies $X_n(s) \approx X(s)$ f.s.l. n.

- Later in you research, you may need the following notions of distance metric and equivalence class, which are related to the almost everywhere equality.
  - Distance Metric
    * We may introduce a metric $d(X(s), Y(s))$ to measure how far away two functions are located.
    * Def. Given a set A, a function $d(x, y)$ of $x, y \in A$ is a distance metric if
      · (non-negativity) $d(x, y) \geq 0$ with equality iff $x = y$ (in some sense),
      · (symmetry) $d(x, y) = d(y, x)$, and
      · (subadditivity=triangle inequality) $d(x, y) + d(y, z) \geq d(x, z)$ hold.
    * Due to the subadditivity requirement, $\int_{-\infty}^{\infty} |X(s) - Y(s)|^p ds$ is not a metric for $0 < p < 1$.
    * "in some sense" may be equal almost everywhere sense.
  - Equivalence Class
    * Given a set A of interest and a distance metric $d : A \times A \to \mathbb{R}_+$,
      · we may partition the set into equivalence classes.
      · A subset that contains all the elements of zero distance to each other is an equivalence class.
      · The equivalence class of an element $x \in A$ is $\{y \in A : d(x, y) = 0\}$.

Convergence of a Sequence of Functions

sequence of functions.
$\{x_n(t)\}_{n\ge 1}$

$x(t) = \sum_{k=1}^{\infty} a_k \sin\left(\frac{2\pi k t}{T}\right)$



$x_n(t) = \sum_{k=1}^{n} a_k \sin\left(\frac{2\pi k t}{T}\right)$

Gibbs' phenomenon.

— uniform convergence, For s.l.n, $|x_n(t) - x(t)| < \varepsilon, \forall t$

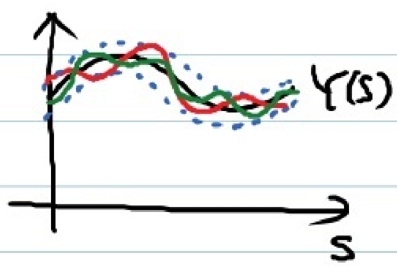
— pointwise convergence, "surely converge"

— almost everywhere " , "converges almost surely"

- Now, we introduce five DIFFERENT modes of convergence all making it mathematically precise that $X_n(s) \approx X(s)$ for sufficiently large n.
  - Though, the mode of convergence is only discussed here, the rate of convergence is also very important.
  - You need to study big O notations and little o notations to describe the rate of convergence.

1. Def. (uniform convergence) A sequence $(X_n(s))_{n\geq 1}$ of real functions converges uniformly to a real function $X(s)$ on $s \in \mathcal{D}$ if

$$\forall \varepsilon > 0, \exists N(\varepsilon) > 0, \text{ s.t. } n \geq N(\varepsilon) \Longrightarrow |X_n(s) - X(s)| \leq \varepsilon, \forall s \in \mathcal{D}.$$



$\varepsilon = 10^{-5}$

$N(\varepsilon) = 10^{6}$

$$X_n(s) \to \boxed{Y(s) = \lim_{n\to\infty} X_n(s)}$$

$$\int_{\Omega} Y(s)\, ds = \int_{\Omega} \lim_{n\to\infty} X_n(s)\, ds \neq \lim_{n\to\infty} \int_{\Omega} X_n(s)\, ds$$

/ex/

$$\lim_{m\to\infty} \lim_{n\to\infty} \frac{m}{m+n} \neq \lim_{n\to\infty} \lim_{m\to\infty} \frac{m}{m+n}$$

2. Def. (everywhere convergence) A sequence $(X_n(s))_{n\geq 1}$ of real functions converges everywhere (e) to a real function $X(s)$ if

$$X_n(s) \to X(s), \forall s \in \mathcal{D}.$$

- $\{s \in \mathcal{D} : X_n(s) \to X(s)\} = \mathcal{D}$.
- $\{s \in \mathcal{D} : X_n(s) \nrightarrow X(s)\} = \emptyset$.
- $\lim_{n\to\infty} X_n(s) = X(s), \forall s \in \mathcal{D}$
- $X_n(s) \to X(s)$ everywhere on $\mathcal{D}$ as $n \to \infty$
- $X_n(s) \to X(s)$ everywhere on $\mathcal{D}$
- pointwise convergence
- Lemma. If X_n converges uniformly to X on $\mathcal{D}$ then X_n converges everywhere to X on $\mathcal{D}$.

3. Def. (almost everywhere convergence) A sequence $(X_n(s))_{n\geq 1}$ of real functions converges almost everywhere (a.e.) to a real function $X(s)$ on $\mathcal{D}$ if

$$m(\{s \in \mathcal{D} : X_n(s) \nrightarrow X(s)\}) = 0$$

- $m(\cdot)$ is the Borel measure.
- When $m(\mathcal{D}) < \infty$, a.e. iff $m(\{s \in \mathcal{D} : X_n(s) \to X(s)\}) = m(\mathcal{D})$.
- $X(s)$ does not converge to $X(s)$ only on a set of measure zero.
- $\lim_{n\to\infty} X_n(s) = X(s)$ for almost every $s \in \mathcal{D}$
- $X_n(s) \to X(s)$ a.e. on $\mathcal{D}$ as $n \to \infty$
- $X_n(s) \to X(s)$ a.e. on $\mathcal{D}$
- Lemma. If X_n converges everywhere to X then X_n converges almost everywhere to X.

4. Def. (convergence in the pth mean) A sequence $(X_n(s))_{n\geq 1}$ of real functions converges in the pth mean to a real function $X(s)$ on $\mathcal{D}$ if

$$\int_{s\in\mathcal{D}} |X_n(s) - X(s)|^p ds \to 0$$

for $p \geq 1$.

- Note that if we view $\int |X_n(s) - X(s)|^p ds$ as x_n then the convergence in the pth mean is nothing but the convergence of a real sequence $(x_n)_{n\geq 1}$ to 0.
- Def. (mean square convergence) When $p = 2$, the convergence is in the mean square (MS) sense.
  - The MS convergence is denoted by

$$\underset{n\to\infty}{\text{l.i.m.}}\ X_n(s) = X(s)$$

  on $\mathcal{D}$. and reads, "the limit in the mean of $X_n(s)$ equals $X(s)$ on $\mathcal{D}$."
- $X_n(s) \to X(s)$ in the pth mean on $\mathcal{D}$ as $n \to \infty$
- $X_n(s) \to X(s)$ in the pth mean on $\mathcal{D}$.
- Lemma. X_n converges almost everywhere to X does not imply X_n converges in the pth mean to X. Conversely, X_n converges in the pth mean to X does not imply X_n converges almost everywhere to X.
- Lemma. For a sequence $(X_n(s))_{n\geq 1}$ of real functions, it converges in the MS

sense iff the sequence is Cauchy, i.e.,

$$\forall \varepsilon > 0, \exists N(\varepsilon) > 0, \text{ s.t. } m,n \geq N(\varepsilon) \Longrightarrow \int_{s\in\mathcal{D}} |X_n(s) - X_m(s)|^2 ds \leq \varepsilon.$$

or simply

$$\int_{s\in\mathcal{D}} |X_n(s) - X_m(s)|^2 ds \to 0$$

as $n \to \infty$ and $m \to \infty$, or for sufficiently large m and n.

5. Def. (convergence in measure) A sequence $(X_n(s))_{n\geq 1}$ of real functions converges in measure to a real function $X(s)$ on $\mathcal{D}$ if

$$\forall \varepsilon > 0, m(\{s \in \mathcal{D} : |X_n(s) - X(s)| > \varepsilon\}) \to 0$$

- Note that if we view $m(\{s \in \mathbb{R} : |X_n(s) - X(s)| > \varepsilon\})$ as $x_n(\varepsilon)$ then the convergence in measure is nothing but the convergence of a sequence $(x_n(\varepsilon))_n$ to zero.
- When $m(\mathcal{D}) < \infty$, converges in measure iff $\forall \varepsilon > 0, m(\{s \in \mathcal{D} : |X_n(s) - X(s)| \leq \varepsilon\}) \to m(\mathcal{D})$.
- $X_n(s) \to X(s)$ in measure on $\mathcal{D}$ as $n \to \infty$
- $X_n(s) \to X(s)$ in measure on $\mathcal{D}$
- Lemma. If X_n converges almost everywhere to X then X_n converges in measure to X.
- Lemma. If X_n converges in the pth mean to X then X_n converges in measure to X.

uniform convergence

$\Downarrow$

convergence everywhere

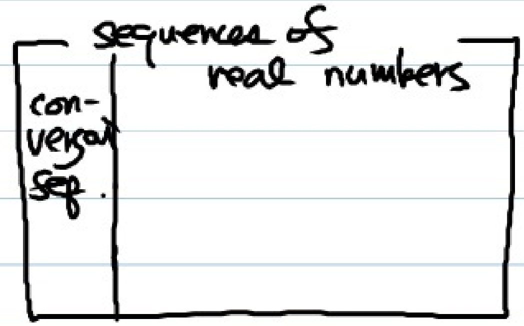
$\Downarrow$

convergence a.e. $\qquad$ convergence in the p-th mean $(p \geq 1)$

$\searrow \qquad \swarrow$

convergence in measure

a sequence of real numbers $(x_n)_{n\geq 1}$ $x_n \in \mathbb{R}$



limit

$\limsup_{n\to\infty} x_n$

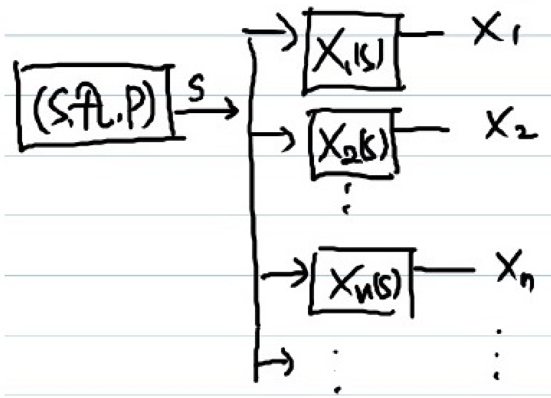
$\liminf_{n\to\infty} x_n$

a sequence of real functions

a sequence of real r. variables

→ Convergence of a sequence of random variables

Ch.7 random vector
random sequence



$(X_n)_{n\geq 1}$

$(X_n)_n$

(X_n)

X_n

$\{X_n\}$

$\{X_n\}_n$

$\{X_n\}_{n\geq 1}$

Q. a sequence of real r. variables converges to ?

Equality vs. Approximate Equality of Two Random Variables

- Apparently,

  (i) the equality of two random variables is nothing but the equality of two functions because a random variable is a function and

  (ii) the convergence of a sequence of random variables is nothing but the convergence of a sequence of functions.

- However,
  - the Borel measure needs to be changed to the probability measure of each probability space, and $P(S) = 1 < \infty$,

  (0) equal everywhere is not used, the uniform convergence is not used,

  (1) almost everywhere equality/convergence is called almost sure equality/convergence,

  (2) the almost everywhere equality/convergence is called the equality/convergence with probability one

  (3) the convergence in measure is called the convergence in probability, and

  (4) there is a new mode of convergence called convergence in distribution.

- In summary, almost surely $=$ with probability 1 is used for equality, and there are 4 modes of convergence.

- Def. (almost sure equality) Given a probability space $(S, \mathcal{A}, P)$, two random variables X and Y are equal almost surely (a.s.) if

$$P(\{s \in S : X(s) = Y(s)\}) = 1,$$

  - ... iff $P(\{s \in S : X(s) \neq Y(s)\}) = 0$, i.e., $X \neq Y$ only on a set of probability zero.
  - $X = Y$ a.s. is alternatively called $X = Y$ with probability one (w.p. 1).

- Lemma. Two random variables X and Y are equal almost surely (a.s.) iff

$$\mathbf{E}\{|X - Y|^p\} = 0$$

  for some $p \geq 1$.

- Lemma. Given a probability space $(S, \mathcal{A}, P)$, two random variables X and Y are equal almost surely (a.s.) iff

$$\forall \varepsilon > 0, P(\{s \in S : |X(s) - Y(s)| > \varepsilon\}) = 0.$$

  - $\forall \varepsilon > 0, P(|X - Y| > \varepsilon) = 0$
  - $\forall \varepsilon > 0, P(|X - Y| \leq \varepsilon) = 1$

- Lemma. $X = Y$ a.s. $\Longrightarrow F_X(x) = F_Y(x)$. However, the converse is not true in general.

$\square$ $\underset{\sim}{X=Y}$ vs $\{X=Y\}$

(i) $X(\omega)=Y(\omega),\ \forall\omega\in\Omega$ equal surely

(ii) $P(\{\omega\in\Omega:\ X(\omega)\neq Y(\omega)\})=0$, " almost surely

(iii) $E[|X-Y|^p]=0$, $p\geq 1$. $\left(\Leftrightarrow \int_\Omega |X(\omega)-Y(\omega)|^p dP(\omega)=0\right)$

(iv) $\forall\varepsilon>0$ $P(|X-Y|>\varepsilon)=0$

Convergence of a Random Sequence

- Four modes of the convergence of a sequence of random variables
  - Given a probability space $(S, \mathcal{A}, P)$, a sequence $(X_n)_{n\geq 1}$ of random variables $X_1(s), X_2(s), \ldots$, and the limit random variable $X(s)$, we have the following four modes of convergence.
    1. Def. $X_n \to X$ almost surely (a.s.)
$$P(\{s : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1$$
    2. Def. $X_n \to X$ in the pth mean ($p \geq 1$)
$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = 0$$
    3. Def. $X_n \to X$ in probability
$$\forall \varepsilon > 0, \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$
    4. Def. $X_n \to X$ in distribution/law (weak convergence)
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \text{ for all continuous points}$$
  - (convergence a.s.) $X_n \to X$ w.p. 1
  - (convergence in the pth mean) $X_n \to X$ in L^p, $X_n \xrightarrow{L^p} X$
  - (convergence in probability) $X_n \xrightarrow{\mathcal{P}} X$; convergence in measure, convergence in probability measure

– (convergence in distribution) $X_n \xrightarrow{\mathcal{D}} X$

- Properties
  - a.s. convergence is often proved by using the Borel-Cantelli lemma.
  - a.s. $\Rightarrow$ in probability
  - $L^p \Rightarrow$ in probability
  - in probability $\Rightarrow$ in distribution
  - in distribution $\Leftrightarrow \lim_{n\to\infty} \Phi_{X_n}(\omega) = \Phi_X(\omega)$

$X_n \xrightarrow{d} X \Leftrightarrow E[f(X_n)] \to E[f(X)]$ as $n \to \infty$

for every bounded continuous ft.

- Example. X_n's are i.i.d. uniform on $[0,1]$ and $Y_n = \min\{X_1, X_2, ...X_n\}$. Then, $Y_n \to 0$ with probability 1.
  - See Example 5.14 in p. 281 of Bertsekas and Tsitsiklis, *Introduction to Probability*, 2nd ed.
- Example. Convergence in probability does not necessarily mean convergence with probability 1.

$$\begin{array}{ccc} \text{a.s.} & & \text{in } L^p \text{ sense} \\ & & \text{in the } p\text{-th mean} \quad (p \geq 1) \\ & \searrow \quad \swarrow & \\ & \text{in prob.} & \\ & \Downarrow & \\ & \text{in distribution} & \end{array}$$

Asymptotic 점근적 analysis.

4 modes of convergence

$\boxed{\text{rate of convergence.}}$

big O, little O notations

– See Example 5.15 in p. 282 of Bertsekas and Tsitsiklis, *Introduction to Probability*, 2nd ed.

- Proofs
  - a.s. $\Rightarrow$ in probability

"$\forall \varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) \to 0$"

$$0 \leq P(|X_n - X| \geq \varepsilon) \leq P\left(\bigcup_{m \geq n} \{|X_m - X| \geq \varepsilon\}\right)$$

Recall

$$\lim_{n\to\infty} P\left(\bigcup_{m\geq n}\{\ \}\right) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{m\geq n}\{\ \}\right)$$

$$= P(\{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}) = 0 \quad \because X_n \to X \text{ a.s.}$$

$$\limsup_{n\to\infty} C_n = \bigcap_{n=1}^{\infty}\bigcup_{m\geq n} C_m$$

$$\liminf_{n\to\infty} C_n = \bigcup_{n=1}^{\infty}\bigcap_{m\geq n} C_m$$

where we used the Borel-Cantelli lemma.

  - in the 2nd mean $\Rightarrow$ in probability

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X|^2 \geq \varepsilon^2) \leq \frac{E\{|X_n - X|^2\}}{\varepsilon^2} \to 0$$

– in probability $\Rightarrow$ in distribution

$$F_{X_n}(x) = P(X_n \le x)$$

$$X > x+\varepsilon$$
$$+ \underline{\;) \; -X_n \ge -x\;}$$
$$X - X_n > \varepsilon$$

(i) upper bound

$$P(X_n \le x) = P(\cancel{X_n \le x}, X \le x+\varepsilon) + \Pr(X_n \le x, X > x+\varepsilon)$$
$$\le P(X \le x+\varepsilon) + \Pr(|X_n - X| > \varepsilon)$$
$$= F_X(x+\varepsilon) + P(|X_n - X| > \varepsilon)$$

(ii) lower bound

$$F_X(x-\varepsilon) = \Pr(X \le x-\varepsilon)$$
$$= \Pr(X \le x-\varepsilon, X_n \le x) + \Pr(X \le x-\varepsilon, X_n > x)$$
$$\le \Pr(X_n \le x) + \Pr(|X_n - X| > \varepsilon)$$

$$F_X(x-\varepsilon) \le F_{X_n}(x) + P(|X_n - X| > \varepsilon)$$

$$-P(|X_n - X| > \varepsilon) + F_X(x-\varepsilon) \le F_{X_n}(x)$$

$n \to \infty$

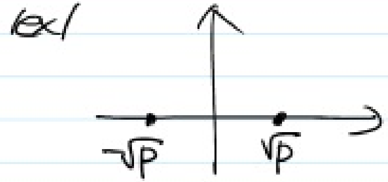
$\varepsilon \downarrow 0$

$$F_X(x) \le \lim_{n\to\infty} F_{X_n}(x) \le F_X(x).$$

• Which mode of convergence is simplest to handle?

• Convergence in distribution and Probability of error

ex)



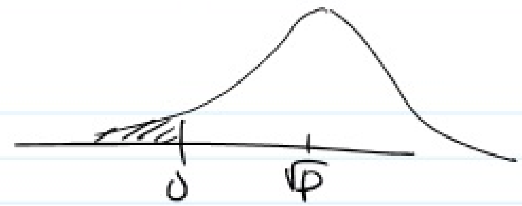
$Y_n = \sqrt{P}\, b + N + \cancel{\sqrt{P_I}\, N_I}$

$b \in \{1, -1\}$

$N \sim N(0, \sigma_N^2)$

$N_I^{(n)} = \dfrac{(2X_1-1)+(2X_2-1)+\cdots+(2X_n-1)}{n} = \dfrac{2\sum X_i - n}{n}$

X_i i.i.d. $B(1, \frac{1}{2})$



$P_b = \Pr(Y_n < 0 \mid b=1)$

$= E\left[1_{(-\infty,0]}(Y_n) \mid b=1\right]$

$\cong \Pr(Y < 0 \mid b=1)$ (X)

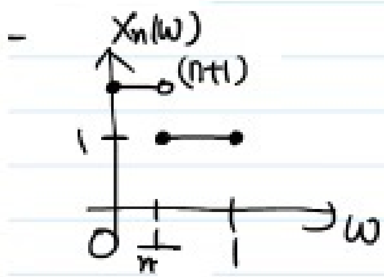
$\simeq \Pr(\sqrt{P}\, b + N + \sqrt{P_I}\, N_I^{(n)} \mid b=1)$

$= Q\left(\sqrt{\dfrac{P}{\sigma_N^2 + \frac{P_I}{n}}}\right)$

$\dfrac{\sqrt{n}\, N_I^{(n)}}{\sqrt{n}} = \dfrac{1}{\sqrt{n}} \dfrac{(2X_1-1)+(2X_2-1)+\cdots(2X_n-1)}{\sqrt{n}}$

$\dfrac{(2X_1-1)+(2X_2-1)+\cdots(2X_n-1)}{\sqrt{n}} \to N(0,1)$

- Examples
  - The strong law of large numbers (SLLN) is an almost sure convergence.
  - The weak law of large numbers (WLLN) is a convergence in measure.
  - The central limit theorem (CLT) is a convergence in distribution.
- Counter-Examples
  - a.s. $\not\Rightarrow$ in the 2nd mean



(i) $\{\omega\in\Omega : X_n(\omega)\to X(\omega)\} = (0,1]$

$P((0,1]) = 1$ $\quad\therefore\ X_n \to X$ w.p.1

(ii) $E\{|X_n-X|^2\} = \int_0^1 |X_n(\omega)-X(\omega)|^2 d\omega$

$= \int_0^{\frac{1}{n}} (n)^2 \, d\omega = n \not\to 0$ as $n\to\infty$.

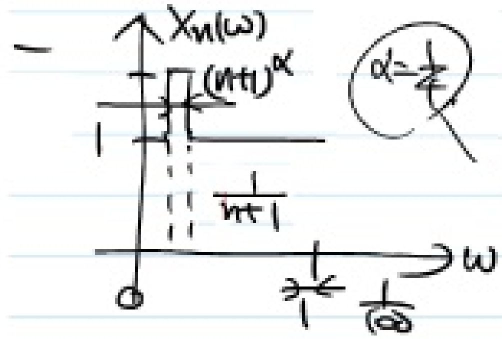
(iii) $P(\{s\in S : |X_n(s)-X(s)|>\varepsilon\})$

$= P(\{s\in S : 0\le s\le \frac{1}{n}\})$

$= \frac{1}{n} \to 0, \quad \forall \varepsilon>0.$

– a.s. $\not\Rightarrow$ in the pth mean $(p \geq 4)$

– in the pth mean $(1 \leq p < 4)$ $\not\Rightarrow$ a.s.



$E\{|X_n - X|^p\} = \frac{1}{n+1}\left((n+1)^{\alpha} - 1\right)^p$

$\to 0$ as $n \to \infty$ $\quad 1 \leq p < 4.$

$X_n \not\to X$ a.s.

- Cauchy Criterion

  – Recall that

Def. $(x_n)_{n\geq 1}$ is a Cauchy sequence if

$\forall \varepsilon > 0, \ \exists N(\varepsilon) \in \mathbb{N}, \ \forall m \in \mathbb{N}, \ n \geq N(\varepsilon) \Rightarrow |x_{m+n} - x_n| < \varepsilon$

Theorem. $(x_n)_{n\geq 1}$ $\quad x_n \in \mathbb{R}.$ $\quad x \in \mathbb{R}.$ $\qquad$ Existence Theorem.

$(x_n)_{n\geq 1}$ is Cauchy $\Leftrightarrow$ x_n converges to $\boxed{\text{some}}$ $x \in \mathbb{R}.$

$\exists x \in \mathbb{R}.$

  – For a sequence of random variables,

(i) $(X_n)_{n\ge1}$ converges a.s. $\Leftrightarrow P(\{\omega : \boxed{X_n(\omega) \text{ is a Cauchy seq}}\})=1$.

$\boxed{\forall m\in\mathbb{N},\ |X_{m+n}(\omega)-X_n(\omega)|\to 0 \text{ as } n\to\infty}$

(ii) $(X_n)_{n\ge1}$ " in the mean-square sense.

$\Leftrightarrow \forall m\in\mathbb{N}, E\{(X_{m+n}-X_n)^2\}\to 0$ as $n\to\infty$

(iii) $(X_n)_{n\ge1}$ converges in p. $\Leftrightarrow \forall\varepsilon>0,\ P(|X_{m+n}-X_n|>\varepsilon)\to 0,\ \forall m$

Law of Large Numbers (LLN)

- Simply put, a law of large numbers is about
  - the convergence of the arithmetic average of random variables
  - to the statistical average
  - in some sense.
    - If the mode is the convergence a.s., then it is called a strong law of large numbers (SLLN).
    - If the mode is the convergence in probability, then it is called a weak law of large numbers (WLLN).
- To define the arithmetic average, we first need a sequence of random variables $(X_i)_{i=1}^{\infty}$.
  - There are many different versions of LLNs.

- Each LLN has a different set of restrictions on the random variables $X_1, X_2, \ldots$.
- Thus, you must be very careful in invoking a law of large numbers.
- However, they mostly have the common condition of $\mathbb{E}[X_i] = \mu_X$, a constant for all i.

• Thus, an LLN is to show

$$M_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu_X$$

in some sense under certain conditions on $X_1, X_2, \ldots$.

I SLLN, WLLN

(ex) X_i's are i.i.d. r.v.'s with $E\{X_i\}=\mu,\ \forall i$

(ex) X_i's are pairwise uncorrelated r.v.'s w/ $E\{X_i\}=\mu$, $Var\{X_i\}=\sigma^2$.

sample mean

arithmetic average

statistical (ensemble) average

$$\overline{X}_n \triangleq \frac{X_1+X_2+\cdots+X_n}{n} \longrightarrow \mu$$

a.s. SLLN

p. WLLN

II CLT

X_i's i.i.d. r.v.'s w/ $E\{X_i\}=\mu$, $Var\{X_i\}=\sigma^2$

$$\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma}}_{\text{mean } 0,\ \text{variance } 1} \xrightarrow{d.} X \sim N(0,1)$$

$$\Leftrightarrow \frac{1}{\sqrt{n}}\Sigma(X_i-\mu) \to X \sim N(0,\sigma^2)$$

$$\not\Leftrightarrow \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i}_{\text{mean } \sqrt{n}\mu,\ \text{variance } \sigma^2} \longrightarrow \cancel{X \sim N(\mu,\sigma^2)}$$

$$= \sqrt{n}\,\overline{X}_n \qquad \frac{N(\sqrt{n}\mu,\sigma^2)}{}$$

$$\overline{X}_n \qquad N(\mu,\tfrac{\sigma^2}{n})$$

- Theorem. (an SLLN) Given a sequence $(X_i)_{i=1}^{\infty}$ of independent and identically distributed random variables with mean μ_X and variance $\sigma_X^2 < \infty$, the arithmetic average converges almost surely to the statistical average μ_X, i.e.,

$$M_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu_X \text{ w.p. 1.}$$

  – Proof. Omitted.

- Theorem. (a WLLN) Given a sequence $(X_i)_{i=1}^{\infty}$ of pairwise uncorrelated random variables with same mean μ_X and same variance $\sigma_X^2 < \infty$, the arithmetic average converges in probability to the statistical average μ_X, i.e.,

$$M_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu_X \text{ in prob.}$$

  – Proof. Use the Chebychev inequality.

- Theorem (a WLLN)

Given a seq. of pairwise uncorrelated r.v.'s X_i w/ $E[X_i] = \mu_X$ & $Var[X_i] = \sigma_X^2 \quad \forall i$,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu_X \text{ in } P.$$

Proof)

To show $\forall \varepsilon > 0$, $P(\{s \in S : |Y_n(s) - \mu_X| > \varepsilon\}) \to 0$

Note that

$$Pr(|Y_n - \mu_X| > \varepsilon) = Pr(|Y_n - \mu_X|^2 > \varepsilon^2) \leq \frac{E\{|Y_n - \mu_X|^2\}}{\varepsilon^2} = \frac{\sigma_X^2}{n\varepsilon^2} \to 0, \forall \varepsilon > 0$$

where

$$E\{|Y_n - \mu_X|^2\} = E\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu_X\right|^2\right\}$$

$$= E\left\{\left|\frac{(X_1 - \mu_X) + (X_2 - \mu_X) + \cdots + (X_n - \mu_X)}{n}\right|^2\right\} = E\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\right|^2\right\}$$

$$= \frac{1}{n^2} E\left\{\left(\sum_{i=1}^{n}(X_i - \mu_X)\right)\left(\sum_{j=1}^{n}(X_j - \mu_X)\right)\right\} = \frac{1}{n^2} \cdot n \cdot \sigma_X^2 = \frac{\sigma_X^2}{n}$$

Central Limit Theorem (CLT)

- Simply put, a central limit theorem is about
  - the convergence of $\sqrt{n}$ times the arithmetic average of zero-mean random variables
  - to a Gaussian random variable
  - in distribution.
- Again, to define the arithmetic average, we first need a sequence of random variables $(X_i)_{i=1}^{\infty}$.
  - There are many different versions of CLTs.
  - Each CLT has a different set of restrictions on the random variables $X_1, X_2, ...$
  - Thus, you must be very careful in invoking a CLT.
- Although
  - the exact CLT is to justify a Gaussian approximation to $\sqrt{n}$ times the arithmetic average of zero-mean random variables,
  - it is often used to justify a Gaussian approximation to the arithmetic average of nonzero-mean random variables.

- Theorem. (a CLT) Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with mean μ_X and variance σ_2, and define

$$Z_n \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_X}{\sigma}.$$

Then, the characteristic function $\Phi_{Z_n}(\omega)$ of Z_n converges to that of a Gaussian random variable with mean 0 and variance 1, i.e.,

$$\lim_{n\to\infty} \Phi_{Z_n}(\omega) = \exp\left(-\frac{\omega^2}{2}\right).$$

  – Proof. Omitted.

- The above CLT says

$$\frac{X_1 + X_2 + ...X_n}{n} \approx Y \sim \mathcal{N}\left(\mu_X, \frac{\sigma^2}{n}\right).$$

However, an LLN says

$$\frac{X_1 + X_2 + ...X_n}{n} \approx \mu_X.$$

Which approximation do you prefer?

- Theorem. (a CLT) If X_i's are i.i.d. random variables with $\Pr(X_i = -1) = \Pr(X_i = 1) = 1/2$, then
$$\frac{X_1 + X_2 + ...X_n}{\sqrt{n}}$$
converges in distribution to a Gaussian random variable with mean zero and variance 1.
  - Proof. Since the characteristic function of $X_i/\sqrt{n}$ is
  $$\Phi(\omega) = \cos\left(\frac{\omega}{\sqrt{n}}\right),$$
  the characteristic function of $(X_1 + X_2 + ...X_n)/\sqrt{n}$ is given by
  $$\Phi_n(\omega) = \left\{\cos\left(\frac{\omega}{\sqrt{n}}\right)\right\}^n.$$
  By using l'Hospital's rule, we can show that
  $$\ln \Phi_n(\omega) = \frac{\ln\cos\left(\frac{\omega}{\sqrt{n}}\right)}{\frac{1}{n}}$$
  converges to $-\omega^2/2$. Thus, by the continuity of the logarithmic function, we have
  $$\Phi_n(\omega) \longrightarrow \exp\left(-\frac{\omega^2}{2}\right).$$
  Therefore, the conclusion follows.

· Trick.

"$g(x)$ is continuous."

$$\equiv \boxed{\lim_{x\to x_0} g(x) = g\left(\lim_{x\to x_0} x\right)} = g(x_0)$$

So, instead of showing

$$\left\{\cos\left(\frac{\omega}{\sqrt{n}}\right)\right\}^n \to \exp\left(-\frac{\omega^2}{2}\right). \quad e^{\ln\left\{\cos\left(\frac{\omega}{\sqrt{n}}\right)\right\}^n} \to e^{-\frac{\omega^2}{2}}$$

$$\ln\left\{\cos\left(\frac{\omega}{\sqrt{n}}\right)\right\}^n \to -\frac{\omega^2}{2}. \quad n\ln\cos\left(\frac{\omega}{\sqrt{n}}\right) = \frac{\ln\cos\left(\frac{\omega}{\sqrt{n}}\right)}{\frac{1}{n}}$$

l'Hospital's rule

$$= \frac{\frac{-\sin\left(\frac{\omega}{\sqrt{n}}\right)}{\cos\left(\frac{\omega}{\sqrt{n}}\right)}\left(-\frac{1}{2}\right)n^{-\frac{3}{2}}\omega}{-\frac{1}{n^2}} \qquad \frac{\frac{1}{n^{\frac{3}{2}}}}{n^2}$$

$$= \frac{\left(-\frac{1}{2}\right)\sin\left(\frac{\omega}{\sqrt{n}}\right)\omega}{\cos\left(\frac{\omega}{\sqrt{n}}\right)\frac{1}{\sqrt{n}}} \to -\frac{\omega^2}{2}.$$

uniform convergence $\Leftarrow$ (X)

$\Downarrow$

pointwise " (e.) $\Leftarrow$ (X)

$\Downarrow$

almost everywhere " (a.e.)

$p=\infty$

$\Downarrow$

$p \geq 1$

SLLN a.s. $\qquad E\{|X_n - X|^p\} \to 0.$ $p \geq 1.$

X_n converges in the p-th mean to X

WLLN in probability (p.)

$\forall \varepsilon > 0.$

$P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty.$

CLT convergence in distribution (d.) $\qquad X_n \xrightarrow{d.} X$

SLLN a.s. $\qquad L^p \qquad$ convergence in the L^p-norm

in the p-th mean

WLLN p

$\Downarrow$

CLT $d \qquad$ convergence in law

weak convergence

- Be careful. The arithmetic mean converges to 0 by an LLN. $\sqrt{n}$ times the arithmetic average converges to a Gaussian.
- This theorem is often used to justify the approximation of a binomial random variable $S_n \sim \mathcal{B}(n, 1/2)$ to the Gaussian random variable $N_n \sim \mathcal{N}(n/2, n/4)$.
  * Be careful. S_n does not converge because its mean diverges to infinity. N_n does not converge because its mean and variance diverge to infinity.
  * (Gaussian approximation to binomial) Similarly, we can invoke the approximation of a binomial random variable $S_n \sim \mathcal{B}(n, p)$ to the Gaussian random variable $N_n \sim \mathcal{N}(np, npq)$.

• (Poisson approximation to binomial) We've already seen a Poisson approximation to a binomial random variable, where the difference is on that np is fixed as λ when n tends to infinity.

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k e^{-np}}{k!} \qquad n \text{ large and } p \text{ small}$$

Gaussian approximation to binomial

$Y_n \sim B(n, \frac{1}{2})$

$$Y_n = X_1 + X_2 + \cdots + X_n = \sqrt{n} \cdot \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

X_i's are i.i.d. w/

$\Pr(X_i = 0) = \Pr(X_i = 1) = 1/2$

$$Y_n - n/2 = \sqrt{n} \cdot \frac{(X_1 - \frac{1}{2}) + (X_2 - \frac{1}{2}) + \cdots + (X_n - \frac{1}{2})}{\sqrt{n}}$$

$$\sim N(0, \tfrac{n}{4})$$

$$Y_n \approx N\left(\frac{n}{2}, \frac{n}{4}\right).$$

- $B(n, p) \approx N(np, np(1-p))$.

1. finite number of sets
   - $\bigcup_{i=1}^{n} \mathcal{A}_i \triangleq \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n$: a finite union
   - $\bigcap_{i=1}^{n} \mathcal{A}_i \triangleq \mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n$: a finite intersection
   - The sets $\mathcal{A}_i$ in a class $\mathcal{A}$ are pairwise disjoint if $\mathcal{A}_n \cap \mathcal{A}_m = \emptyset, \forall n \neq m$.← Draw a Venn diagram.
2. countably infinite number of sets $(\mathcal{A}_n)_{n\in\mathbb{N}}$
   - $\bigcup_{n=1}^{\infty} \mathcal{A}_n \triangleq \{x : x \in \mathcal{A}_n, \exists n \in \mathbb{N}\}$: a countable union
   - $\bigcap_{n=1}^{\infty} \mathcal{A}_n \triangleq \{x : x \in \mathcal{A}_n, \forall n \in \mathbb{N}\}$: a countable intersection
   - Def. $(\mathcal{A}_n)_{n\in\mathbb{N}}$ is an increasing sequence of subsets of $\mathcal{S}$ if $\mathcal{A}_n \subset \mathcal{A}_{n+1}, \forall n$.
     – If $\mathcal{A} \triangleq \cap_{n=1}^{\infty} \mathcal{A}_n$ then we denote $\mathcal{A}_n \downarrow \mathcal{A}$.
   - Def. $(\mathcal{A}_n)_{n\in\mathbb{N}}$ is a decreasing sequence of subsets of $\mathcal{S}$ if $\mathcal{A}_n \supset \mathcal{A}_{n+1}, \forall i$.
     – If $\mathcal{A} \triangleq \cup_{n=1}^{\infty} \mathcal{A}_n$ then we denote $\mathcal{A}_n \uparrow \mathcal{A}$.
   - Def. The limsup (limit supremum) and liminf (limit infimum) of a sequence $(\mathcal{A}_n)_{n\in\mathbb{N}}$ of subsets of $\mathcal{S}$:

– $\limsup_{n\to\infty} \mathcal{A}_n \triangleq \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} \mathcal{A}_m = \{s \in \mathcal{S} : s \in \mathcal{A}_n$ for infinitely many $n\}$

– $\liminf_{n\to\infty} \mathcal{A}_n \triangleq \bigcup_{n=1}^{\infty} \bigcap_{m\geq n} \mathcal{A}_m = \{s \in \mathcal{S} : s \in \mathcal{A}_n$ for all but finitely many $n\}$

– $\cap_{n=1}^{\infty}\mathcal{A}_n \subset \liminf_{n\to\infty} \mathcal{A}_n \subset \limsup_{n\to\infty} \mathcal{A}_n \subset \cup_{n=1}^{\infty}\mathcal{A}_n$

- Def. $\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A}$ if $\lim_{n\to\infty} 1_{\mathcal{A}_n}(s) = 1_{\mathcal{A}}(s), \forall s \in \mathcal{S}$.
  – $\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A}$ iff $\limsup_{n\to\infty} \mathcal{A}_n = \liminf_{n\to\infty} \mathcal{A}_n = \mathcal{A}$

3. uncountably infinite number of sets $(\mathcal{A}_\alpha)_{\alpha\in\mathcal{I}}$, where $\mathcal{I}$ is an uncountable index set.

- $\bigcup_{\alpha\in\mathcal{I}} \mathcal{A}_\alpha \triangleq \{x : x \in \mathcal{A}_\alpha, \exists\alpha \in \mathcal{I}\}$: an uncountable union
- $\bigcap_{\alpha\in\mathcal{I}} \mathcal{A}_\alpha \triangleq \{x : x \in \mathcal{A}_\alpha, \forall\alpha \in \mathcal{I}\}$: an uncountable intersection
- /ex/ When $\mathcal{A}_\alpha = [2, \alpha)$, find $\cap_{\alpha>3}\mathcal{A}_\alpha$

- Q. Given

$$A_n = \left\{\frac{0}{n}, \frac{1}{n}, \cdots, \frac{n^2}{n}\right\},$$

– show that $\limsup_n A_n = \mathbb{Q}_+$ and

– $\liminf_n A_n = \mathbb{Z}_+$

- Q. Given

$$B_n = \left[0, \frac{n}{n+1}\right],$$

  - show that $\limsup_n B_n = [0, 1)$, $\liminf_n B_n = [0, 1)$, and hence
  - $\lim_n B_n = [0, 1)$
  -

- Q. (5 points: 21-1 HW#2) When you repeatedly toss an unbiased coin independently, you stop at the first show of heads. Answer the following questions.

  (a) Classify the sample space.

  (b) Find the prob. that you stop at an even number of tosses.

  (c) Find the prob. that you stop at an odd number of tosses.

  (d) Find the prob. that you toss infinitely many times.

- Q. (17 points: 21-1 HW#2) We have a deck of cards, on the face of each card an integer is printed; thus the cards are $\{1, 2, 3...\}$. At the nth round of this game, the first n cards are taken, they are shuffled. You pick one of them. If your pick is 1, you win that round. Let $\mathcal{B}_n$ denote the event that you win the nth round. Each round is independent of all the other rounds. Let $\mathcal{B}_n = \{Y_n = 1\}$ and $\mathcal{B}_n^c = \{Y_n = 0\}$ and define $M = \sum_{n=1}^{\infty} Y_n$. Answer the following questions.

  (a) (2 points) Are $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n, ...$ independent events? Are $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n, ...$ pairwise disjoint events?

  (b) (5 points) What is the meaning of $\cap_{n=1}^{\infty} \mathcal{A}_n$ and find $P(\cap_{n=1}^{\infty} \mathcal{A}_n)$.

  (c) (5 points) Find $P(M = 1)$.

  (d) (5 points) What is the meaning of $\cup_{n=1}^{\infty} \mathcal{A}_n$ and find $P(\cup_{n=1}^{\infty} \mathcal{A}_n)$.

  (e) (0 point) Show $P(M = m) = 0$ for all $m > 1$ and show $P(M = \infty) = 1$ if you can. (Hint: Borel-Cantelli lemma)

- (Wiki)
  - In probability theory, the Borel-Cantelli lemma is a theorem about sequences of events. In general, it is a result in measure theory.
  - It is named after Emile Borel and Francesco Paolo Cantelli, who gave statement to the lemma in the first decades of the 20th century. A related result, sometimes called the second Borel-Cantelli lemma, is a partial converse of the first Borel-Cantelli lemma.
  - The lemma states that, under certain conditions, an event will have probability of either zero or one. Accordingly, it is the best-known of a class of similar theorems, known as zero-one laws. Other examples include Kolmogorov's zero-one law and the Hewitt-Savage zero-one law.
  - Let $E_1, E_2, \ldots$ be a sequence of events in some probability space. The Borel-Cantelli lemma states:

    If the sum of the probabilities of the event E_n is finite

    $$\sum_{n=1}^{\infty} \Pr(E_n) < \infty,$$

    then the probability that infinitely many of them occur is 0. That is,

    $$\Pr\left(\limsup_{n\to\infty} E_n\right) = 0,$$

or, equivalently, the probability that only finitely many of them occur is 1, that is,

$$\Pr\left(\liminf_{n\to\infty} E_n^c\right) = 1.$$

– A related result, sometimes called the second Borel-Cantelli lemma, is a partial converse of the first Borel-Cantelli lemma. The lemma states:

If the events E_n are independent and the sum of the probabilities of the E_n diverges to infinity, i.e.,

$$\sum_{n=1}^{\infty} \Pr(E_n) = \infty$$

then the probability that infinitely many of them occur is 1, i.e.,

$$\Pr(\limsup_{n\to\infty} E_n) = 1.$$

The assumption of independence can be weakened to pairwise independence, but in that case the proof is more difficult.

- Examples: We have a deck of cards, on the face of each card an integer is printed; thus the cards are $\{1, 2, 3...\}$.

Q1. At the nth round of this game, the first n^2 cards are taken, they are shuffled. You pick one of them.

  * If your pick is 1, you win that round. Let $\mathcal{A}_n$ denote the event that you win the nth round.
  * The complement $\mathcal{A}_n^c$ of $\mathcal{A}_n$ will represent that you lose the nth round.
  * The event $\mathcal{A}(\triangleq \limsup_{n\to\infty} \mathcal{A}_n)$ represents those scenarios in which you win infinitely many rounds.
  * The complement of this event is $\mathcal{A}^c(\triangleq \liminf_{n\to\infty} \mathcal{A}_n^c)$, and this represents those scenarios in which you ultimately lose all of the rounds, i.e., you win only finitely many rounds.
  * Find $P(\mathcal{A}) = 1 - P(\mathcal{A}^c)$.

– A2. $P(\limsup_{n\to\infty} \mathcal{A}_n) = 0$ and $P(\liminf_{n\to\infty} \mathcal{A}_n^c) = 1$. Thus, a player of this game will deterministically experience that there comes a time, after which he never wins.

Q2. At the nth round of this game, the first n cards are taken, they are shuffled. You pick one of them.

* If your pick is 1, you win that round. Let $\mathcal{B}_n$ denote the event that you win the nth round.
* The complement $\mathcal{B}_n^c$ of $\mathcal{B}_n$ will represent that you lose the nth round.
* The event $\mathcal{B}(\triangleq \limsup_{n\to\infty} \mathcal{B}_n)$ represents those scenarios in which you win infinitely many rounds.
* The complement of this event is $\mathcal{B}^c(\triangleq \liminf_{n\to\infty} \mathcal{B}_n^c)$, and this represents those scenarios in which you ultimately lose all of the rounds, i.e., you win only finitely many rounds.
* Find $P(\mathcal{B}) = 1 - P(\mathcal{B}^c)$.

– A2. $P(\limsup_{n\to\infty} \mathcal{B}_n) = 0$ and $P(\liminf_{n\to\infty} \mathcal{B}_n^c) = 1$. Thus, a player of this game will always win infinitely many times.

*Unequal Distributions

- A sufficient condition for the theorem's validity

$$\sigma_{X_i}^2 > B_1 > 0 \qquad i = 1, 2, ..., n \tag{7.4-1a}$$

$$\mathbf{E}[|X_i - \mu_i|^3] < B_2 \qquad i = 1, 2, ..., n \tag{7.4-1b}$$

- No one random variable in the sum dominates.
- The CLT guarantees only that the *distribution* of the sum becomes gaussian.
- not necessarily the pdf
- quite accurate for relatively small values of n in the central region of the gaussian curve near the mean

*Equal Distributions

- independent and identically distributed (i.i.d.) random variables
- continuous
- We shall work with the zero-mean, unit-variance random variable

$$
\begin{aligned}
W_n &= (Y_n - \bar{Y_n})/\sigma_{Y_n} = \sum_{i=1}^{n}(X_i - \mu_i) \Bigg/ \left[\sum_{i=1}^{n} \sigma_{X_i}^2\right]^{1/2} \\
&= \frac{1}{\sqrt{n}\sigma_X}\sum_{i=1}^{n}(X_i - \bar{X}) \qquad (7.4\text{-}2)
\end{aligned}
$$

$$\mu_i = \bar{X} \qquad \text{all } i \qquad (7.4\text{-}3)$$

$$\sigma_{X_i}^2 = \sigma_X^2 \qquad \text{all } i \qquad (7.4\text{-}4)$$

- We show the characteristic function of W_n converges to that of a zero-mean, unit-variance gaussian random variable, i.e.,

$$\lim_{n\to\infty} \Phi_{W_n}(\omega) = \exp\left(-\omega^2/2\right). \qquad (7.4\text{-}5)$$

- the uniqueness of Fourier transforms

$$
\begin{aligned}
\Phi_{W_n}(\omega) &= \mathbf{E}[e^{j\omega W_n}] = E\left[\exp\left\{\frac{j\omega}{\sqrt{n}\sigma_X}\sum_{i=1}^{n}(X_i - \bar{X})\right\}\right] \\
&= \left\langle E\left\{\exp\left[\frac{j\omega}{\sqrt{n}\sigma_X}(X_i - \bar{X})\right]\right\}\right\rangle^n \qquad (7.4\text{-}6)
\end{aligned}
$$

$$
\begin{aligned}
& E\left\{\exp\left[\frac{j\omega}{\sqrt{n}\sigma_X}(X_i-\bar{X})\right]\right\} \\
&= E\left\{1+\left(\frac{j\omega}{\sqrt{n}\sigma_X}\right)(X_i-\bar{X})+\left(\frac{j\omega}{\sqrt{n}\sigma_X}\right)^2\frac{(X_i-\bar{X})^2}{2}+\frac{R_n}{n}\right\} \\
&= 1-(\omega^2/2n)+\mathbf{E}[R_n]/n
\end{aligned}
\tag{7.4-7}
$$

- $\mathbf{E}[R_n]$ approaches zero as $n \to \infty$.

$$\ln[\Phi_{W_n}(\omega)] = n\ln\{1-(\omega^2/2n)+\mathbf{E}[R_n]/n\} \tag{7.4-8}$$

- Since

$$\ln(1-z) = -\left[z+\frac{z^2}{2}+\frac{z^3}{3}+\cdots\right] \qquad |z|<1 \tag{7.4-9}$$

$$\ln[\Phi_{W_n}(\omega)] = -(\omega^2/2)+\mathbf{E}[R_n]-\frac{n}{2}\left[\frac{\omega^2}{2n}-\frac{\mathbf{E}[R_n]}{n}\right]^2+\cdots \tag{7.4-10}$$

$$\lim_{n\to\infty}\{\ln[\Phi_{W_n}(\omega)]\} = \ln\left\{\lim_{n\to\infty}\Phi_{W_n}(\omega)\right\} = -\omega^2/2 \tag{7.4-11}$$

$$\lim_{n\to\infty}\Phi_{W_n}(\omega) = e^{-\omega^2/2} \tag{7.4-12}$$

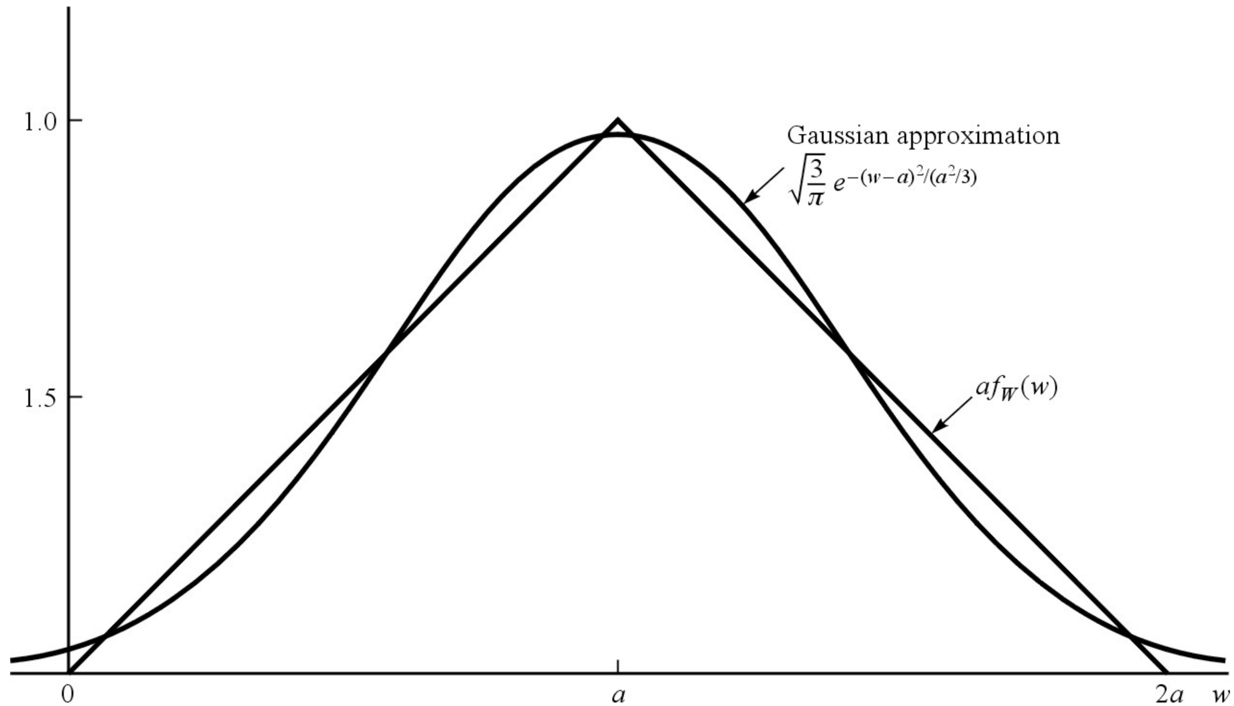
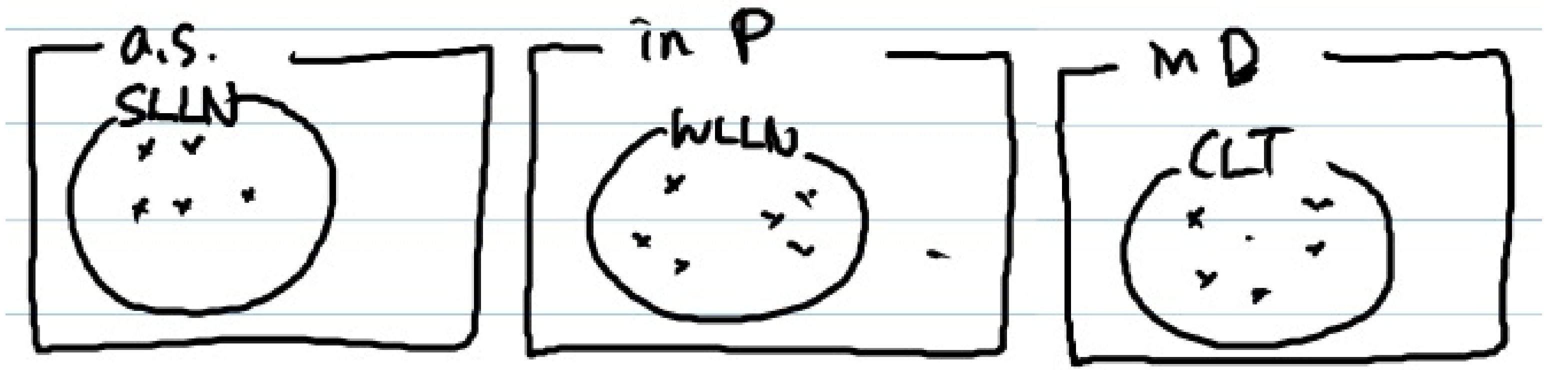

1.0
1.5
0
a
2a
w
Gaussian approximation
$\sqrt{\frac{3}{\pi}}\, e^{-(w-a)^2/(a^2/3)}$
$af_W(w)$

Figure 7.4-1


a.s.
SLLN
in P
WLLN
in D
CLT

7.5 Random Numbers: Meaning and Generation

- (Monte Carlo integration) Suppose we want to compute

$$I = \int_0^1 g(x)dx,$$

  where

  - it is not difficult to compute $g(x_0)$ for $0 \le x_0 \le 1$
  - but it is hard to directly compute the integration.
  - The trick is to view I as

$$I = \mathbf{E}\{g(X)\},$$

    where X is a uniform random variable.
  - Now, we invoke an SLLN, which says

$$\frac{\sum_{i=1}^{n} g(X_i)}{n} \longrightarrow \mathbf{E}\{g(X)\}.$$

  - Thus, we can claim that

$$I \approx \frac{\sum_{i=1}^{n} g(X_i)}{n}$$

    for sufficiently large n.
  - Therefore, instead of direct integration, we can perform a repeated trial with i.i.d. uniform random variables and obtain a good estimate of I.

- Such a technique to replace an integration is called **a Monte Carlo integration**.

- (BER of BPSK in a Rayleigh fading channel) Q. Find a good estimate of

$$\mathbf{E}\left[Q\left(\sqrt{\frac{2E_b}{N_0}}X\right)\right],$$

  where X is a Rayleigh random variable with unit mean.

  – Actually, a closed-form expression can be obtained by direct integration. (Try it!)

  – However, a Monte Carlo integration can also be used to obtain an estimate. (Try it!)

- The trouble in using a Monte Carlo integration is on that there is few random number generator.

  – ERNIE (Electronic Random Number Indicating Equipment) 1 was a hardware random number generator based on quantum physics. It weas developled in 1957.

  – Mostly, people use a soft-ware pseudo-random number generator.

- (p. 287) The most general algorithm for generating a random number sequence is to use a nonlinear recursion.

  – MATLAB also uses pseudo-random number generators.

- methods to obtain arbitrary distribution

  – percentile transformation method

– rejection method

– mixing method

a is chosen to make $a\frac{f_Y(x)}{f_X(x)} \le 1 \quad \forall x$

$\{U \le a\frac{f_Y(X)}{f_X(X)}\}$

Random Number Generation.

– percentile method.

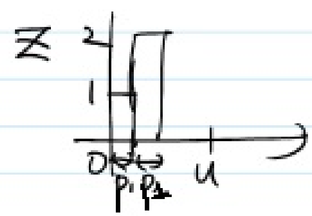
– rejection method. independent U, X, $f_X(x|M) = f_Y(x)$.

– mixing method $f_Y(y) = p_1 f_{X_1}(y) + p_2 f_{X_2}(y) + \cdots + p_m f_{X_m}(y)$

where $1 \ge p_i \ge 0$

$\sum_{i=1}^{m} p_i = 1$

– ⋮



- Markov Chain Monte Carlo (MCMC)

7.6 Sampling and Some Limit Theorems

- sampling
  - continuous-time to discrete-time conversion = analog to digital conversion
  - continuous-time to vector conversion
  - discrete-time to discrete-time conversion
  - discrete-time to vector conversion
- Here, sampling has a different meaning.

Sampling and Estimation

- measuring some quantity
- We *sample* to get our measurement.
- sample = an observation
- an unknown parameter
- an estimate
- How to optimally estimate the parameter using the observation? $->$ statistical estimation theory $\subset$ statistical signal processing

Estimation of Mean, Power, and Variance

- Statistical Estimation Problems
  - A statistical estimation problem consists of four components.
    (i) a random observation: Usually, the observation is modeled by a random variable, a random vector, a random process, etc. The statistical property of the observation is not perfectly known.
    (ii) a set of unknown parameters: The statistical property of the random observation is fully or partly determined by a set of parameters that are unknown. For example, the mean and the variance of a gaussian random observation are unknown. We wish to estimate all or some of the unknown parameters.
    (iii) an estimate: guessed value of the unknown parameters to be estimated
    (iv) an estimator: an estimator is a mapping rule from the observation to an estimate. The output of an estimator is to be analyzed and/or the estimator is to be designed in a statistical estimation problem.
  - Def. An estimator is *unbiased* if the mean of the estimator equals the quantity being estimated.
  - Def. A sequence of estimators is *consistent* if it converges to the quantity being estimated.
    * strongly consistent: almost sure convergence

  * mean-square consistent (consistent in quadratic mean): L^2 convergence
  * weakly consistent: convergence in probability
  * asymptotically unbiased: convergence of the mean

- An example of a statistical estimation problem

  (i) observation: The observation is a sequence of uncorrelated random variables having the same mean μ and the same variance σ^2.

  * The n samples x_i represent values of pairwise uncorrelated random variable $X_i, i = 1, 2, ..., n$.

  (ii) unknown parameters:

  * They have the same mean value $\bar{X}$ and variance σ_X^2, but both unknown.

  (iii), (iv) Find the minimum variance unbiased estimator (MVUE) and verify the estimator is unbiased and/or consistent.

  * Problem 1: We wish to estimate the mean value.
  * Problem 2: We wish to estimate the variance.

- Answer to Problem 1

  – The average of the sample values gives a number which we call an *estimate* of the mean of the random variables.

$$\hat{\bar{x}}_n = \text{estimate of average of } n \text{ samples} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (7.6\text{-}1)$$

– an *estimator*

$$\hat{\hat{X}}_n = \frac{1}{n}\sum_{i=1}^{N} X_i \tag{7.6-2}$$

– The above estimator is called the *sample mean*.

– It produces a specific *estimate* of $\bar{X}$ for a specific set of samples.

– the mean and variance of the estimator

* mean

$$\mathbf{E}[\hat{\hat{X}}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \bar{X}, \qquad \text{any } n \tag{7.6-3}$$

* variance

$$\begin{aligned}
\mathbf{E}[(\hat{\hat{X}}_n - \bar{X})^2] = \sigma^2_{X_n} &= \mathbf{E}[\hat{\hat{X}}^2_n - 2\bar{X}\hat{\hat{X}}_n + \bar{X}^2] \\
&= \mathbf{E}[\hat{\hat{X}}^2_n] - \bar{X}^2 = -\bar{X}^2 + E\left[\frac{1}{n}\sum_{i=1}^{n} X_i \frac{1}{n}\sum_{j=1}^{n} X_j\right] \\
&= -\bar{X}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i X_j]
\end{aligned} \tag{7.6-4}$$

$$\begin{aligned}
\sigma^2_{X_j} &= -\bar{X}^2 + \frac{1}{n^2}[nE(X^2) + (n^2 - n)\bar{X}^2] \\
&= \frac{1}{n}[E(X^2) - \bar{X}^2] = \sigma^2_X/n
\end{aligned} \tag{7.6-5}$$

* The sample mean is an unbiased estimator.
* The sample mean is mean-square and weakly consistent.

$$P\{|\hat{\bar{X}}_n - \bar{X}| < \varepsilon\} \geq 1 - (\sigma_{X_n}^2/\varepsilon^2) = 1 - \frac{\sigma_X^2}{n\varepsilon^2} \tag{7.6-6}$$

- Answer to Problem 2
  - an estimator for the second-order non-central moment

$$\widehat{X_n^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{7.6-7}$$

  - an estimator for the second-order central moment, the variance

$$\widehat{\sigma_X^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\bar{X}}_n)^2 \tag{7.6-8}$$

  - $\widehat{\sigma_X^2}$ is unbiased. Why? If the factor $1/(n-1)$ is changed to $1/n$, then biased.