Before we prove Thm 9.3.4, need a lemma.

Lemma 9.3.5: Let Z be a RV with $E[Z] < 0$ s.t.
$M_Z(s) < \infty$ for $-a < s < b$, for some $a, b > 0$. Then

$$P(Z \geq 0) \leq \rho \quad \text{where} \quad \rho = \inf_{0<s<b} M_Z(s) < 1.$$

Pf: For any $s \in (0,b)$ the function $f(x) = e^{sx}$ is increasing, so by Markov's Ineq.

$$P(Z \geq 0) = P(e^{sZ} \geq 1) \leq \frac{E[e^{sZ}]}{1} = M_Z(s)$$

Now take infimum over $0 < s < b$:

$$P(Z \geq 0) \leq \inf_{0<s<b} M_Z(s) = \rho.$$

Lastly, since $M_Z(0) = 1$ and $M_Z'(0) = E[Z] < 0$ (by assumption), it follows that $M_Z(s) < 1$ for all positive s sufficiently close to 0, i.e. $\rho < 1$. (calculus)

21 1a

Pf of Thm 9.3.4: Let $Y_i = X_i - \mu - \varepsilon$ s.t. $E[Y_i] = -\varepsilon < 0$.

For $-a < s < b$, we have that

$$M_{Y_i}(s) = E[e^{sY_i}] = e^{-s(\mu+\varepsilon)} E[e^{sX_i}] = e^{-s(\mu+\varepsilon)} M_{X_i}(s) < \infty$$

(since $M_{X_i}(s) < \infty \ \forall i$ by assumption).

Then by Lemma 9.3.5,

$$P\left(\frac{X_1 + \cdots + X_n}{n} \geq \mu + \varepsilon\right) = P\left(\frac{Y_1 + \cdots + Y_n}{n} \geq 0\right)$$

$$= P\left(Y_1 + \cdots + Y_n \geq 0\right)$$

$$\leq \inf_{0<s<b} M_{Y_1 + \cdots + Y_n}(s) \underset{\substack{\text{since } Y_i\text{'s} \\ \text{are indep}}}{=} \inf_{0<s<b} \left(M_{Y_1}(s)\right)^n = \rho^n$$

where $\rho = \inf_{0<s<b} M_{Y_1}(s) = \inf_{0<s<b} \left(e^{-s(\mu+\varepsilon)} M_{X_1}(s)\right)$.

Lastly, $\rho < 1$ by Lemma 9.3.5. $\blacksquare$

Weak Convergence

(Ref: R § 10, B § 25)

def: Given Borel probability distributions $\mu, \mu_1, \mu_2, \ldots$ on $\mathbb{R}$, $\{\mu_n\}$ converges weakly to μ if

$$\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$$ for all bounded continuous functions $f: \mathbb{R} \to \mathbb{R}$.

Denoted by $\boxed{\mu_n \Rightarrow \mu}$.

Recall: The boundary of a set $A \subseteq \mathbb{R}$ is

$$\partial A = \{x \in \mathbb{R} : \forall \varepsilon > 0,\ A \cap (x-\varepsilon, x+\varepsilon) \neq \phi,\ A^c \cap (x-\varepsilon, x+\varepsilon) \neq \phi\}.$$

Thm: The following are equivalent:

(1) $\mu_n \Rightarrow \mu$ (weak convergence)

(2) $\mu_n(A) \to \mu(A)$ $\forall$ measurable sets A s.t. $\mu(\partial A) = 0$.

(3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ $\forall x \in \mathbb{R}$ s.t. $\mu(\{x\}) = 0$.

(4) [Skorohod's Thm]: $\exists$ RVs $Y, Y_1, Y_2, \ldots$ defined jointly on some $(\Omega, \mathcal{F}, P)$ with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ $\forall n \in \mathbb{N}$ s.t. $Y_n \to Y$ with prob. 1.

(5) $\int_{\mathbb{R}} f\, d\mu_n \to \int_{\mathbb{R}} f\, d\mu$ $\forall$ bounded Borel-measurable functions

$f: \mathbb{R} \to \mathbb{R}$ s.t. $\mu(D_f) = 0$ where D_f is the set of points where f is discontinuous.

Pf: Exercise to the reader! (Rosenthal)

Example 1: Let μ be Lebesgue measure on $[0,1]$ & let $\mu_n\left(\frac{i}{n}\right) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$.

μ = continuous

μ_n = discrete

Note that $\mu(\mathbb{Q}) = 0$ while $\mu_n(\mathbb{Q}) = 1$ for each n.

We do have $\mu_n \Rightarrow \mu$ as $n \to \infty$. why?

For any $0 \le x \le 1$, $\mu((-\infty, x]) = x$

while $\mu_n((-\infty, x]) = \frac{\lfloor nx \rfloor}{n}$.

Then $\left| \mu_n((-\infty, x]) - \mu((-\infty, x]) \right| = \left| \frac{\lfloor nx \rfloor}{n} - x \right|$

$\le \frac{1}{n} \to 0$ as $n \to \infty$

Note that $\partial(\mathbb{Q} \cap [0,1]) = [0,1]$ so $\mu(\partial(\mathbb{Q} \cap [0,1])) \neq 0$

so doesn't contradict part (2).

Connections to other types of convergence

Prop: If $X_n \to X$ in probability, then $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ as $n \to \infty$.

Remark: Sometimes write $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ as $n \to \infty$ as $X_n \Rightarrow X$ as $n \to \infty$. We say that $\{X_n\}$ converges weakly to X, or in distribution.

$$\boxed{\begin{array}{l} X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{P} X \quad \Longrightarrow \quad X_n \Rightarrow X \ (\text{in dist'n}) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)) \\ \qquad \Rightarrow \exists \text{ RVs } Y_n \ \& \ Y \text{ having the same laws} \\ \qquad\quad \text{s.t. } Y_n \xrightarrow{a.s.} Y \qquad (\text{Thm}, (4)) \end{array}}$$

Note: The converse of the proposition above is FALSE.

$$\mathcal{L}(X_n) = \mathcal{L}(X) \not\Rightarrow X_n \xrightarrow{P} X \text{ as } n \to \infty,$$

Example: $X, X_1, X_2, \ldots$ i.i.d. each equal to ± 1 with prob. $\frac{1}{2}$, then $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ but

$$P(|X - X_n| \geq 2) = \frac{1}{2} \not\to 0 \text{ as } n \to \infty$$

either $X = 1$ and $X_n = -1$ $\to$ prob $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
or $X = -1$ and $X_n = +1$ $\to$ " " $= \frac{1}{4}$
$\}$ sum these to get $\frac{1}{2}$.

so X_n does NOT converge to X in probability or w/prob. 1.

[However, if X is constant then the converse does hold.]

Lastly, we can use Skorohod's Thm to translate results involving convergence w/prob. 1 to results involving weak convergence (or convergence in prob. by using Prop 10.2.1 R.)

For example,

<u>Prop</u>: Suppose $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ as $n \to \infty$, with $X_n \geq 0$.

Then $E[X] \leq \liminf_n E[X_n]$.

[Pf uses Fatou's Lemma].

e.g. If $X = 0$ and $P(X_n = n) = \frac{1}{n}$, $P(X_n = 0) = 1 - \frac{1}{n}$,

then $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ and

$$0 = E[X] \leq \liminf_n E[X_n] = 1$$

Characteristic Functions

(Ref: R § 11, B § 26) — this will not be on Exam 2, yes on the Final

def: Given a random variable X, its characteristic function (aka Fourier Transform) is given by

$$\boxed{\phi_X(t) = E\left[e^{itX}\right]} = E\left[\cos(tX)\right] + i\,E\left[\sin(tX)\right], \quad t \in \mathbb{R}$$

- Function from $\mathbb{R}$ to $\mathbb{C}$
- By Change of Variable Thm, $\phi_X(t)$ only depends on the distribution of X

"Similar" to MGF $\left(M_X(s) = E\left[e^{sX}\right]\right)$ but the appearance of the imaginary number $i = \sqrt{-1}$ is a significant difference:

$$\left|e^{itX}\right| = 1 \quad \text{for any real } t \text{ and } X$$

$e^{it} = \cos t + i \sin t$ for $t \in \mathbb{R}$

$|e^{it}| = \sqrt{\cos^2 t + \sin^2 t} = 1$

$$|\phi_X(t)| \le 1 < \infty \quad \forall t, X \quad \text{whereas}$$

Always exists since it is bounded

$M_X(s)$ can be infinite for some $s \neq 0$.

More details:

A RV X with distribution μ has characteristic function

$$\phi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx}\, d\mu(x) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx$$

If X has PDF f_X that is Riemann integrable

Properties

- The characteristic function uniquely determines the distribution (as in MGFs).
- Pointwise convergence of characteristic functions implies weak convergence of the corresponding distributions. → Continuity Thm, used to prove CLT

Moments & Derivatives

As with MGFs,

- $\phi_X(0) = 1$ for any RV X
- X & Y indep RVs $\Rightarrow$ $\phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$

Let $\mu = \mathcal{L}(X)$. Then

$$\left|\phi_X(t+h) - \phi_X(t)\right| = \left|\int \left(e^{i(t+h)x} - e^{itx}\right) d\mu(x)\right|$$

Lec 21 (3b)

~~Lecture 6~~

1 Characteristic Functions

For weak convergence of probability measures on $\mathbb{R}^d$, or equivalently weak convergence of $\mathbb{R}^d$-valued random variables, an important *convergence determining class* of functions are functions of the form $f(x) = e^{it\cdot x}$, with $t \in \mathbb{R}^d$. They lead to

Definition 1.1 [Characteristic Function] *Let X be a $\mathbb{R}^d$-valued random variable with distribution μ on $\mathbb{R}^d$. Then the characteristic function of X (or μ) is defined to be*

$$\phi(t) := \mathbb{E}[e^{it\cdot X}] = \int e^{it\cdot x}\mu(dx).$$

When μ has a density with respect to Lebesgue measure, i.e., $\mu(dx) = \rho(x)dx$, $\phi(\cdot)$ is just the Fourier transform of the function $\rho(\cdot)$. Therefore in general, we can think of $\phi(\cdot)$ as the Fourier transform of the measure μ.

Here are some properties which are immediate from the definition:

Proposition 1.2 [Properties of Characteristic Functions]

(i) *Since $e^{it\cdot x} = \cos tx + i\sin tx$ has bounded real and imaginary parts, $\phi(t)$ is well-defined.*

(ii) *$\phi(0) = 1$ and $|\phi(t)| \leq 1$ for all $t \in \mathbb{R}^d$.*

(iii) *ϕ is uniformly continuous on $\mathbb{R}^d$. More precisely,*

$$|\phi(t+h) - \phi(t)| = \left|\mathbb{E}[e^{i(t+h)\cdot X}] - \mathbb{E}[e^{it\cdot X}]\right| = \left|\mathbb{E}[e^{it\cdot X}(e^{ih\cdot X} - 1)]\right| \leq \mathbb{E}[|e^{ih\cdot X} - 1|],$$

which tends to 0 as $h \to 0$ by the bounded convergence theorem.

(iv) *The complex conjugate of ϕ, $\overline{\phi(t)}$, is the characteristic function of $(-X)$, and $\phi(t) \in \mathbb{R}$ for all $t \in \mathbb{R}^d$ if $X \stackrel{\text{dist}}{=} (-X)$.*

(v) *For $a \in \mathbb{R}$ and $b \in \mathbb{R}^d$, $aX + b$ has characteristic function $\varphi(t) = e^{ib\cdot t}\phi(at)$.*

(vi) *If and X and Y are two independent random variables with characteristic functions ϕ_X and ϕ_Y, then $X+Y$ has characteristic function $\phi_{X+Y}(t) = \mathbb{E}[e^{it\cdot(X+Y)}] = \phi_X(t)\phi_Y(t)$.*

Examples

Proposition 1.3 [Common Distributions and Their Characteristic Functions]

(a) *The delta measure at $a \in \mathbb{R}$: $\mu(dx) = \delta_a(dx)$. Then $\phi(t) = e^{ia\cdot t}$.*

(b) *The coin flip: $\mu(\{1\}) = \mu(\{-1\}) = \frac{1}{2}$. Then $\phi(t) = \frac{e^{it}+e^{-it}}{2} = \cos t$.*

(c) *The Poisson distribution with parameter $\lambda > 0$: $\mu(\{n\}) = e^{-\lambda}\frac{\lambda^n}{n!}$ for $n \in \{0\} \cup \mathbb{N}$. Then $\phi(t) = e^{-\lambda}\sum_{n=0}^{\infty} e^{itn}\frac{\lambda^n}{n!} = e^{-\lambda(1-e^{it})}$.*

(d) *The compound Poisson distribution: $X = \sum_{n=1}^{T} Y_i$, where T has Poisson distribution with parameter λ, and $(Y_i)_{i\in\mathbb{N}}$ is an independent sequence of i.i.d. random variables with characteristic function $\psi(\cdot)$. Then $\phi(t) = \mathbb{E}[e^{itX}] = e^{-\lambda(1-\psi(t))}$.*

$|ab| = |a||b|$

$$\leq \int \left| e^{i(t+h)x} - e^{itx} \right| d\mu(x)$$

$$= \int \underbrace{|e^{itx}|}_{=1} \, |e^{ihx} - 1| \, d\mu(x) = \int |e^{ihx} - 1| \, d\mu(x)$$

As $h \to \infty$, this decreases to 0 by the Bounded Convergence Thm (since $|e^{ihx} - 1| \leq 2$). Hence, $\phi_X(t)$ is always a uniformly continuous function.

Derivatives of $\phi_X(t)$ are straightforward.

Prop 11.0.1: Suppose X is a RV with $E[|X|^k] < \infty$. Then for $0 \leq j \leq k$, ϕ_X has finite j^{th} derivative given by

$$\boxed{\phi_X^{(j)}(t) = E\left[(iX)^j e^{itX}\right]}.$$

In particular, evaluate at $t = 0$ to get

$$\boxed{\phi_X^{(j)}(0) = i^j E[X^j]}.$$

Pf: Uses induction on j. See text for details.