Osama Mahmood

Central Limit Theorem

Assumptions from the previous presentation...

Theorem 1: Suppose X is a rv with $E(|X|^k) < \infty$. Then for $0 \le j \le k$, ϕ_X has finite jth derivative, given by $\phi_X^{(j)}(t) = E[(iX)^j e^{itX}] \Rightarrow \phi_X^{(j)}(0) = i^j E(X^j)$.

Theorem 2: (Fourier Uniqueness Theorem) Let X and Y be random variables. Then $\phi_X(t) = \phi_Y(t)$ for all $t \in R$ if and only if $L(X) = L(Y)$, that is X and Y have the same distribution.

Theorem 3: (Helly Selection Principle) Let $\{F_n\}$ be a sequence of cummulative distribution functions (i.e. $F_n(x) = \mu_n((-\infty, x])$ for some probability distribution μ_n). Then there is a subsequence $\{F_{n_k}\}$ and a non-decreasing right-continuous function F with $0 \le F \le 1$, such that $\lim_{k \to \infty} F_{n_k} = F(x)$ for all $x \in R$ such that F is continuous at x.

Theorem 4: (Weak Convergence). The following are equivalent...

(1) $\mu_n \Rightarrow \mu$ (as $\{\mu_n\}$ converges weakly to μ)

(2) $\mu_n(A) \to \mu(A)$ for all measurable sets A : $\mu(\partial A) = 0$

(3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in R$ such that $\mu\{x\} = 0$

(4) (Skorohod's Theorem) There are random variables $Y_1, Y_2, Y_3 \ldots$ defined jointly on some probability triple with $L(Y) = \mu$ and $L(Y_n) = \mu_n$ for each $n \in N$ such that $Y_n \to Y$ with probability 1.

(5) $\int_R f \, d\mu_n \to \int f \, d\mu$ for all Bounded borel-measurable functions $f: R \to R$ such that $\mu(D_f) = 0$ where f is discontinuous at D_f.

Def. We say that a collection of probability measures $\{\mu_n\}$ on $\mathbb{R}$ is tight if for all all $\epsilon > 0$ there are $a < b$ with $\mu_n([a,b]) \geq 1-\epsilon$ for all n. That all of the measures give most of their mass to the same finite interval.

Theorem 5: If $\{\mu_n\}$ is a tight sequence of probability measures then there is a subsequence $\{\mu_{n_k}\}$ and a probability measure μ, such that $\mu_{n_k} \Rightarrow \mu$, $\{\mu_{n_k}\}$ converges weakly to μ.

Proof. Let $F_n = \mu_n((-\infty, x])$. Then by Helly Selection Prin there is a subsequence F_{n_k} and a function F such that $F_{n_k}(x) \longrightarrow F(x)$ all points of continuity of F $(0 \leq F \leq 1)$.
Now we claim that F is a p.d.f. Let $\epsilon > 0$ we can find points $a < b$ ~~that~~ that are continuity points of F such that

$\mu_n([a,b]) \geq 1-\epsilon$ for all n, but

$$
\begin{aligned}
\lim_{n\to\infty} F(x) - \lim_{x\to-\infty} F(x) &\geq F(b) - F(a) \\
&= \lim_k [F_{n_k}(b) - F_{n_k}(a)] \\
&= \lim_k \mu_{n_k}((a,b]) \\
&\geq 1-\epsilon
\end{aligned}
$$

Since this is for all $\epsilon > 0$ we have $\lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) = 1$, proving the claim. Since F is a p.d.f we can define probability measure μ by $\mu((a,b]) = F(b) - F(a)$, for $a < b$, then by weak convergence $\mu_{n_k} \Rightarrow \mu$.

Theorem 6: Let $\{\mu_n\}$ be a tight sequence of probability distribution on $\mathbb{R}$. Suppose that μ is the only possible weak limit of $\{\mu_n\}$, in the sense that whenever $\mu_{n_k} \Rightarrow \nu$ then $\nu = \mu$. Then $\mu_n \Rightarrow \mu$. Hence if a subsequence converges to a point then the full sequence converge to that point as well.

Proof: If $\mu_n \not\Rightarrow \mu$ then by ~~W.C.~~ it is not the case that $\mu_n(\infty, x] \to \mu(-\infty, x]$ for all $x \in \mathbb{R}$, with $\mu\{x\} = 0$. Hence we can find a subsequence $\{n_k\}$ with $\mu\{x\} = 0$ but with

$$\left|\mu_{n_k}((-\infty,x]) - \mu((-\infty,x])\right| \geq \epsilon, \quad k \in N \quad (\ast).$$

On the other hand $\{\mu_{n_k}\}$ is a subcollection of $\{\mu_n\}$ and hence tight, so by ~~W.C~~ Th 5 there is a further subsequence $\{\mu_{n_{k_j}}\}$ which converges weakly to some probability measure ν, but then we must have $\nu = \mu$, which contradicts $(\ast)$.

Theorem 7: Let $\{\mu_n\}$ be a sequence of probability measures on $\mathbb{R}$, with characteristic functions $\phi_n(t) = \int e^{itx} \mu_n(dx)$. Suppose there is a function g which is continuous at 0, such that $\lim_n \phi(t) = g(t)$ for each $|t| < t_0$ for some $t_0 > 0$. Then $\{\mu_n\}$ is tight.

Proof: We first note that $g(0) = \lim_{n\to\infty} \phi_n(0) = \lim_{n\to\infty} 1 = 1$. Then we compute that for $y > 0$

$$\mu_n\left(\left(-\infty, -\tfrac{2}{y}\right] \cup \left[\tfrac{2}{y}, \infty\right)\right) = \int_{|x| \geq \frac{2}{y}} 1 \; \mu_n(dx)$$

$$\leq 2 \int_{|x| \geq \frac{2}{y}} \left(1 - \frac{1}{y|x|}\right) \mu_n(dx) \; ; \quad 1 - \frac{1}{y|x|} \geq \frac{1}{2}$$

$$\leq 2 \int_{|x| \geq \frac{2}{y}} \left(1 - \frac{\sin(yx)}{yx}\right) \mu_n(dx) \; ; \quad \frac{\sin(yx)}{yx} = \frac{\sin(y|x|)}{y|x|} \leq \frac{1}{y|x|}$$

$$= \int_{|x| \geq \frac{2}{y}} \left(\frac{1}{y}\right) \int_{-y}^{y} (1 - \cos(tx))\, dt \, \mu_n(dx) \; ; \quad \int_{-y}^{y} \cos(tx)\, dt = \frac{2\sin(yx)}{x}$$

$$\leq \int_{x \in \mathbb{R}} \left(\frac{1}{y}\right) \int_{-y}^{y} (1 - \cos(tx))\, dt \, \mu_n(dx) \; ; \quad 1 - \cos(tx) \geq 0$$

$$= \int_{x \in \mathbb{R}} \left(\frac{1}{y}\right) \int_{-y}^{y} (1 - e^{itx})\, dt \, \mu_n(dx) \; ; \quad \int_{-y}^{y} \sin(tx)\, dt = 0$$

$$= \frac{1}{y} \int_{-y}^{y} (1 - \phi_n(t))\, dt \; ; \quad \text{Fubini Theorem}$$

Let $\epsilon > 0$ and since $g(0) = 1$ and g is continuous at 0 we can find y_0 with $0 < y_0 < t_0$ such that $|1 - g(t)| \leq \frac{\epsilon}{4}$, whenever $|t| \leq y_0$

$\left| \frac{1}{y_0} \int_{-y_0}^{y_0} (1 - g(t))\, dt \right| < \frac{\epsilon}{2}$. Now $\phi_n(t) \to g(t)$ for all

$|t| \leq y_0$, and $|\phi_n(t)| \leq 1$. Hence by the bounded convergence theorem we can find $n_0 \in \mathbb{N}$ such that $\left| \frac{1}{y_0} \int_{-y_0}^{y_0} (1 - \phi_n(t))\, dt \right| < \epsilon$ for all $n \geq n_0$.

Hence $\mu_n\left(-\frac{2}{y_0}, \frac{2}{y_0}\right) = 1 - \mu_n\left(\left(-\infty, -\frac{2}{y_0}\right] \cup \left[\frac{2}{y_0}, \infty\right)\right) > 1 - \epsilon$, $n \geq n_0$. so $\{\mu_n\}$ is tight.

<u>Theorem 8:</u> (Continuity Theorem) Let $\mu_1, \mu_2, \mu_3 \ldots$ be probability measures, with corresponding characteristic functions $\phi, \phi_1, \phi_2, \ldots$ Then $\mu_n \Rightarrow \mu$ if and only if $\phi_n(t) \Rightarrow \phi(t)$ for all $t \in R$. In words, probability measures $\{\mu_n\}$ converge weakly to μ if and only if their characteristic functions converge pointwise to that of μ.

<u>Proof:</u>

First suppose that $\mu_n \Rightarrow \mu$. Since $\cos(tx)$ and $\sin(tx)$ are bounded and continuous we have as $t \to \infty$ for each $t \in R$ that

$$\phi_n(t) = \int \cos(tx)\, \mu_n(dx) + i \int \sin(tx)\, \mu_n(dx)$$

$$\to \int \cos(tx)\, \mu(dx) + i \int \sin(tx)\, \mu(dx)$$

$$= \phi(t)$$

Oppositely suppose that $\phi_n(t) \to \phi(t)$ for each $t \in R$. Then by Theorem 7, the $\{\mu_n\}$ are tight. Now suppose that we have $\mu_{n_k} \Rightarrow \nu$ for some subsequence $\{\mu_{n_k}\}$ and some measure ν.
Then from the previous part we have $\phi_{n_k}(t) \to \phi_\nu(t)$ for all t, where $\phi_\nu(t) = \int e^{itx} \nu(dx)$.
On the other hand we know that $\phi_{n_k}(t) \to \phi(t)$ ~~hence we~~ for all t, hence we must have $\phi_\nu \to \phi$. But from Theorem 2, this implies th $\nu = \mu$.

Now since μ is the only possible weak limit of $\{\mu_n\}$, we must have $\mu_n \to \mu$ by Theorem 6.

Lemma 9. If $X \sim N(0,1)$ then $\phi_X(t) = e^{-t^2/2}$ for all $t \in \mathbb{R}$.

Proof: We can differentiate under integral sign to obtain,

$$\phi_X'(t) = \int_{-\infty}^{\infty} ix e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= \int_{-\infty}^{\infty} i e^{itx} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2}\, dx$$

Integrating by parts gives

$$\phi_X'(t) = \int_{-\infty}^{\infty} i(it) e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= -t\, \phi_X(t)$$

Hence $\phi_X'(t) = -t\, \phi_X(t)$

$$\frac{d}{dt} \log \phi_X(t) = -t.$$

We know that $\log \phi_X(0) = \log 1 = 0$

Hence, we must have $\log \phi_X(t) = \int_0^t (-s)\, ds = -t^2/2$,

$$\Longrightarrow \phi_X(t) = e^{-t^2/2}$$

Theorem 9. (Central Limit Theorem).

Let $X_1, X_2 \ldots$ be i.i.d

Let $S_n = \sum X_i$

$$L\left(\frac{S_n - nm}{\sqrt{vn}}\right) \Rightarrow \mathcal{N}$$

Proof: we replace X_i by $\frac{X_i - m}{\sqrt{v}}$, $m=0$, $v=1$

Let $\phi_n(t) = E(e^{it\frac{S_n}{\sqrt{n}}})$ be characterist function of $\frac{S_n}{\sqrt{n}}$

By continuity theorem and Th 8, it suffices to show that $\lim_{n\to\infty} \phi_n(t) = e^{-t^2/2}$ for each fixed $t \in \mathbb{R}$.

Now set $\phi(t) = E(e^{itX_1})$. As $n \to \infty$ using Taylor expansion.

$$\phi_n(t) = E(e^{it(X_1+X_2+\ldots+X_n)/\sqrt{n}})$$

$$= \phi(t/\sqrt{n})^n$$

$$= \left(1 + \frac{it}{\sqrt{n}}E(X_1) + \frac{1}{2!}\left(\frac{it}{\sqrt{n}}\right)^2 E((X_1)^2) + o(\tfrac{1}{n})\right)^n$$

$$= \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n \longrightarrow e^{-t^2/2}.$$

$o(1/n)$ is a quantity q_n such that $\frac{q_n}{(1/n)} \to 0$ as $n \to \infty$.

The limit holds since for any $\epsilon > 0$, and for sufficiently large n we have $a_n \geq -\epsilon/n$ and also $a_n \leq \epsilon/n$, so that

$$\liminf (\ldots) \geq e^{-(t^2/2)-\epsilon}$$

and

$$\limsup (\ldots) \leq e^{-(t^2/2)+\epsilon}$$