Lecture 10
Empirical Probability Distributions

As before, let $\{X_n, n \geq 1\}$ be a sequence of independent identically distributed random variables.

$\forall \omega \in \Omega$, $X_n(\omega)$ are observed values chosen from $\mathcal{P}$ or equivalently $\mathcal{F}$, the probability distribution function.

For each n and each $\omega \in \Omega$, $\{X_j(\omega), 1 \leq j \leq n\}$ are n real numbers which can be rearranged in ascending order

$$Y_{n1}(\omega) \leq Y_{n2}(\omega) \leq \cdots \leq Y_{nn}(\omega)$$

called the order statistics.

Definition: Define $F_n(\cdot, \omega)$ as follows

$$F_n(x, \omega) = 0 \text{ if } x < Y_{n1}(\omega)$$

$$F_n(x, \omega) = \frac{k}{n} \text{ if } Y_{nk}(\omega) \leq x < Y_{nk+1}(\omega)$$

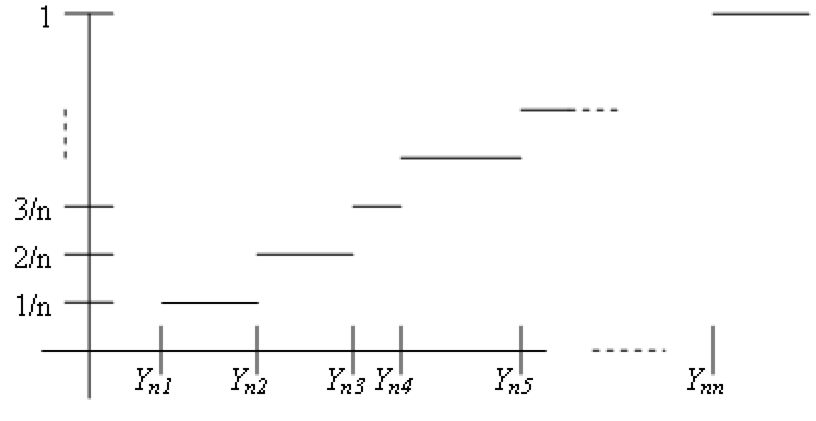
$$F_n(x, \omega) = 1 \text{ if } x \geq Y_{nn}(\omega)$$

F_n is called the empirical distribution function.
Note that $F_n(\cdot, \omega)$ is a random function.

Suppose

$$I_j(x, \omega) = \begin{cases} 1 & \text{if } X_j(\omega) \leq x \\ 0 & \text{if } X_j(\omega) > x. \end{cases}$$

I_j is the indicator of whether $X_j \leq x$ or not.



F_n is a step function. Notice that $nF_n(x)$ just counts the # of observation less than or equal to x.

$$nF_n(x,\omega) = \sum_{j=1}^{n} I_j(x,\omega).$$

But because the X_j are i.i.d.; and $I_j(x,\omega)$ depends only on $X_j(\omega)$, then I_j are independent also. Moreover, they are Bernouli random variables. That is

$$I_j = \begin{cases} 1 & \text{with probability } p = F(x) \\ 0 & \text{with probability } q = 1 - F(x) \end{cases}$$

This is easy to see because

$$p = P[I_j = 1] = P[X_j \leq x] = F(x)$$
$$q = P[I_j = 0] = P[X_j > x] = 1 - P[X_j \leq x] = 1 - F(x)$$

By the strong Law of large numbers

$$F_n(x,\omega) = \frac{1}{n}\sum_{j=1}^{n} I_j(x,\omega) \rightarrow \underline{F}[I_j(x,\omega)]$$
$$= P[I_j = 1] \cdot 1 + P[I_j = 0] \cdot 0 = p = F(x) \text{ a.s.}$$

Thus $F_n(x,\omega) \overset{n\rightarrow\infty}{\rightarrow} F(x)$ a.s.

This convergence is pointwise. Unfortunately we really want convergence to be uniform.

We want to make a global statement about the convergence.

$$F_n(x,\omega) \rightarrow F(x) \text{ a.s.}$$

means for each x $\exists$ a null set $N(x) \ni P[N(x)] = 0$

$$F_n(x,\omega) \rightarrow F(x)\ \forall \omega \in \Omega \setminus N(x).$$

Clearly if we take the countable union of null sets they will still be a null set.

$\Rightarrow$ for all $x \in Q$ a countable set, e.g. rationals
$\Rightarrow \exists N_Q \ni P(N_Q) = 0$ and

$$F_n(x,\omega) \rightarrow F(x)\ \forall \omega \in \Omega \setminus N_Q$$

Because Q $(\overset{cardinality}{rationals})$ are dense, each element of N_Q corresponding to an x rational, it follows that.

$$F_n(x,\omega) \rightarrow F(x) \forall x \in \mathbb{R} \text{ and } \forall \omega \in \Omega \setminus N_Q$$

Lemma: Let $F_n(x)$ and $F(x)$ be right continuous distribution functions with Q being the rationals and

$$J = \{x \in \mathbb{R} : F(x) \text{ has a jump at } x\}.$$

J is countable. Suppose

$$F_n(x) \to F(x) \; \forall x \in Q$$

and

$$F_n(x) - F_n(x-) \to F(x) - F(x-) \; \forall x \in J$$

then $F_n(x)$ converges uniformly to $F(x)$, $x \in \mathbb{R}$.

Proof: Note uniform convergence means given $\epsilon > 0 \exists n_0$ not dependent on $x \ni$

$$\text{for } n \geq n_0, |F_n(x) - F(x)| \leq \epsilon \forall x.$$

Alternatively for $n \geq n_0$

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \leq \epsilon.$$

Suppose that the convergence is not uniform.

$\Rightarrow \exists \epsilon > 0$ and a sequence n_k of integers, $n_k \to \infty$ and a sequence $\{X_k\}$ in $\mathbb{R} \ni$

$$\forall k |F_{n_k}(x_k) - F(x_k)| \geq \epsilon > 0.$$

Suppose $x_k \to \infty$. $\Rightarrow \exists x \ni P[X(\omega) \leq x] \geq 1 - \delta \; \forall \delta \leq 1$.

$$F_{n_k}(x) \to P[X \leq x] \geq 1 - \delta.$$

also $F(x) \geq 1 - \delta$.

$$\Rightarrow |F_{n_k}(x_k) - F(x_k)| \leq |F_{n_k}(x) - F(x)| \leq 2\delta$$

Choose $\delta < \frac{\epsilon}{2}$,
then

$$|F_{n_k}(x_k) - F(x_k)| \leq 2\delta < 2\frac{\epsilon}{2} = \epsilon.$$

This is contrary to assumption $|F_{n_k}(x_k) - F(x_k)| \geq \epsilon \forall k$. Hence $x_k \nrightarrow \infty$. Similarly $x_k \nrightarrow -\infty$.

Thus $\limsup x_k < \infty$, say $x_0 = \limsup x_k$.

We can thus take a subsequence of x_k, which we again relabel as $x_k \ni$

$$x_k \to x_0.$$

Suppose then $r_1, r_2 \in Q \ni r_1 < x_0 < r_2$.

1. Consider $x_k \uparrow x_0, x_k < x_0 \Rightarrow r_1 < x_k$ eventually

$$\epsilon \le F_{n_k}(x_k) - F(x_k) \le F_{n_k}(x_0-) - F(r_1)$$
$$\le F_{n_k}(x_0-) - F_{n_k}(x_0) + F_{n_k}(r_2) - F(r_2) + F(r_2) - F(r_1)$$

Let $k \to \infty$

$$\epsilon \le F(x_0-) - F(x_0) + F(r_2) - F(r_2) + F(r_2) - F(r_1)$$
$$= F(x_0-) - F(x_0) + F(r_2) - F(r_1)$$

Let $r_2 \downarrow x_0$, $r_1 \uparrow x_0$.

$$\epsilon \le F(x_0-) - F(x_0) + F(x_0) - F(x_0-) = 0.$$

This is contradiction.

2. Suppose then $x_k \uparrow x_0$, $r_1 < x_k < x_0$ eventually, but

$$\epsilon \le F(x_k) - F_{n_k}(x_k) \le F(x_0-) - F_{n_k}(r_1)$$
$$\le F(x_0-) - F(r_1) + F(r_1) - F_{n_k}(r_1)$$

Let $k \to \infty$

$$\epsilon \le F(x_0-) - F(r_1) + F(r_1) - F(r_1)$$
$$= F(x_0-) - F(r_1)$$

Let $r_1 \uparrow x_0$

$$\epsilon \le F(x_0-) - F(x_0-) = 0. \text{ Again contradiction}$$

3. Suppose $x_k \downarrow x_0$, $x_k \ge x_0$, $r_1 < x_0 \le x_k < r_2$ eventually

$$\epsilon \le F(x_k) - F_{n_k}(x_k) \le F(r_2) - F_{n_k}(x_k)$$
$$\le F(r_2) - F(r_1) + F(r_1) - F_{n_k}(r_1) + F_{n_k}(x_0-) - F_{n_k}(x_0)$$

Let $k \to \infty$

$$\epsilon \le F(r_2) - F(r_1) + F(r_1) - F(r_1) + F(x_0-) - F(x_0)$$

Let $r_2 \downarrow x_0$, $r_1 \uparrow x_0$

$$\epsilon \le F(x_0) - F(x_0-) + F(x_0-) - F(x_0) = 0$$

Again contradiction.

Last Case
4. Suppose $x_k \downarrow x_0$, $x_k \geq x_0 \Rightarrow r_1 < x_0 \leq x_k < r_2$ event

$$\epsilon \leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(r_2) - F(x_0)$$

Let $k \to \infty$

$$\epsilon \leq F(r_2) - F(x_0)$$

Let $r_2 \downarrow x_0$

$$\epsilon \leq F(x_0) - F(x_0) = 0$$

Contradiction.

Thus $|F_{n_k}(x_k) - F(x_k)| \geq \epsilon > 0\ \forall k$ cannot bet true.
$\therefore$ Convergence is uniform.

Theorem: Glivenko-Cantelli

$$\sup_{-\infty<x<\infty} |F_n(x,\omega) - F(x)| \to 0 \text{ a.s.}$$

Proof: Let J be the countable set of jumps of F. $\forall x \in J$, define

$$\eta_j(x,\omega) = \left\{ \begin{array}{ll} 1 & X_j(\omega) = x \\ 0 & X_j(\omega) \neq x \end{array} \right. .$$

Then for $x \in J$,

$$F_n(x+,\omega) - F_n(x-,\omega) = \frac{1}{n}\sum_{j=1}^{n} \eta_j(x,\omega)$$

$\Rightarrow$ for $x \in J$, $\exists N(x) \ni \omega \in \Omega \setminus N(x)$

$$\Rightarrow F_n(x+,\omega) - F_n(x-,\omega) \to F(x+) - F(x-).$$

Let

$$N_1 = \bigcup_{x \in Q \cup J} N(x).$$

Then N_1 is a null set and if $\omega \in \Omega \setminus N_1$, then

$$F_n(x+,\omega) - F_n(x-,\omega) \to F(x+) - F(x-)\ \forall x \in J$$

and

$$F_n(x) \to F(x)\ \forall x \in Q.$$

By the Lemma

$$F_n(x) \to F(x) \text{ uniformly in } x\ \forall \omega \in \Omega \setminus N_1$$

i.e.

$$\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0 \text{ a.s.}$$

Characteristic Functions

Let X be a random variable on $(\Omega, \mathcal{A}, P)$, ν be the induced probability measure, and F the corresponding probability distribution function

$$\psi(t) = E\left[e^{itX}\right] = \int_\Omega e^{itX(\omega)} P(d\omega)$$
$$= \int_{\mathbb{R}} e^{itx} \nu(dx) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

is the characteristic function. Notice that $\psi(t)$ is a complex valued function of a real variable $t \ni$

$$R(\psi(t)) = \int_\Omega \cos(tX(\omega)) P(d\omega) = \int_{-\infty}^{\infty} \cos(tx) dF(x)$$

and

$$I(\psi(t)) = \int_\Omega \sin(tX(\omega)) P(d\omega) = \int_{-\infty}^{\infty} \sin(tx) dF(x).$$

ψ is associated with ν, F not X.
ψ is the Fourier-Stiettjes transform of ν.

Properties of ψ.

$$i)\ |\psi(t)| = \left|\int_{-\infty}^{\infty} e^{itx} dF(x)\right| \le \int_{-\infty}^{\infty} |e^{itx}| dF(x)$$
$$= \int_{-\infty}^{\infty} dF(x) = 1 = \psi(0).$$
$$\psi(-t) = \int_{-\infty}^{\infty} e^{i(-t)x} dF(x) = \int_{-\infty}^{\infty} e^{(-i)tx} dF(x)$$
$$= \overline{\int_{-\infty}^{\infty} e^{itx} dF(x)} = \overline{\psi(t)}.$$

i.e. $|\psi(t)| \le 1 = \psi(0)$
$\psi(-t) = \overline{\psi(t)}$.

ii) ψ is uniformly continuous in $\mathbb{R}$.

$$\psi(t+h) - \psi(t) = \int \left(e^{i(t+h)x} - e^{itx}\right) dF(x)$$

$$\Rightarrow |\psi(t+h) - \psi(t)| \le \int |e^{itx}||e^{ihx} - 1| dF(x)$$

$$= \int |e^{ihx} - 1| dF(x)$$

$$\left|e^{ihx} - 1\right| \leq \left|e^{ihx}\right| + |1| = 2.$$

$$\Rightarrow |\psi(t+h) - \psi(t)| \leq \int 2dF(x) = 2.$$

also as $h \to 0$, $\left|e^{ihx} - 1\right| \to 0$

By bounded convergence theorem

$$0 \leq |\psi(t+h) - \psi(t)| \leq \lim_{h\to 0} \int \left|e^{ihx} - 1\right| dF(x) \to 0$$

But the l.h.s. does not involve $t \Rightarrow$ convergence is uniform in t.

iii) If ψ_x is the characteristic function corresponding to X (F_X is distribution of X) then

$$\psi_{aX+b}(t) = \int e^{it(aX+b)} dF_X(x)$$

$$= e^{itb} \int e^{a(itX)} dF_X(x)$$

$$= e^{itb} \int e^{i(at)X} dF_X(x) = e^{itb} \psi_X(at)$$

iv) If $\{\psi_n, n \geq 1\}$ are characteristic functions, and $\lambda_n \geq 0$ with

$$\sum_{n=1}^{\infty} \lambda_n = 1, \text{ then}$$

$$\sum_{n=1}^{\infty} \lambda_n \psi_n \text{ is also a characteristic function.}$$

If $\{\nu_n, n \geq 1\}$ are the corresponding probability measures then $\sum \lambda_n \nu_n$ is the measure corresponding to $\sum \lambda_n \psi_n$

Let $\psi = \sum \lambda_n \psi_n$
and $\nu = \sum \lambda_n \nu_n$

$$\psi(t) = \sum \lambda_n \psi_n = \sum \lambda_n \left(\int e^{itx} \nu_n(dx) \right)$$

$$= \sum \int e^{itx} \lambda_n \nu_n(dx) = \int e^{itx} \sum \lambda_n \nu_n(dx) = \int e^{itx} \nu(dx).$$

$v)$ If $\{\psi_j, 1 \le j \le n\}$ are characteristic functions then

$$\prod_{j=1}^{n} \psi_j \text{ is also a characteristic function}$$

Let ν_j be the probability measure corresponding to ψ_j. By previous theorem $\exists$ independent random variables X_j corresponding to ν_j.
Let

$$S_n = \sum_{j=1}^{n} X_j$$

$$E\left[e^{itS_n}\right] = E\left[\prod_{j=1}^{n} e^{itX_j}\right]$$

$$= \prod_{j=1}^{n} E\left[e^{itX_j}\right] = \prod_{j=1}^{n} \psi_j$$

Thus

$$\prod_{j=1}^{n} \psi_j \text{ is the characteristic function of } S_n.$$

Definition: The convolution of two probability distribution functions $F_1 \& F_2$ is the distribution function

$$F(x) = \int_{-\infty}^{\infty} F_1(x-y)dF_2(y)$$

written $F = F_1 * F_2$.

Theorem: Let X_1 and X_2 be independent random variables with distribution functions F_1 & F_2.

Then $F_1 * F_2$ is the distribution function of $X_1 + X_2$.

Proof: Show $P(X_1 + X_2 \le x) = (F_1 * F_2)(x)$

Define

$$I(x_1, x_2) = \begin{cases} 1 & x_1 + x_2 \le x \\ 0 & x_1 + x_2 > x \end{cases}$$

$$P[X_1 + X_2 \le x] = \int_{\Omega} I(x_1, x_2) dP$$

$$= \int\int_{\mathbb{R}^2} I(x_1, x_2)\nu^2(dx_1, dx_2)$$

$$= \int_{\mathbb{R}} \nu_2(dx_2) \int_{\mathbb{R}} I(x_1, x_2)\nu_1(dx_1)$$

$$= \int_{\mathbb{R}} \nu_2(dx_2) \int_{(-\infty, x-x_2]} \nu_1(dx_1)$$

$$= \int_{\mathbb{R}} \nu_2(dx_2)F_1(x - x_2)$$

$$= \int_{\mathbb{R}} F_1(x - x_2)dF_2(x_2) = (F_1 * F_2)(x)$$

Notice we could just as well written

$$\int\int_{\mathbb{R}^2} I(x_1, x_2)\nu^2(dx_1, dx_2)$$

$$= \int_{\mathbb{R}} \nu_1(dx_1) \int_{\mathbb{R}} I(x_1, x_2)\nu_2(dx_2)$$

$$= \int_{\mathbb{R}} \nu_1(dx_1) \int_{(-\infty, x-x_1]} \nu_2(dx_2)$$

$$= \int_{\mathbb{R}} \nu_1(dx_1)F_2(x - x_1)$$

$$= \int_{\mathbb{R}} F_2(x - x_1)dF_1(x) = (F_2 * F_1)(x)$$

$$\Rightarrow F_1 * F_2 = F_2 * F_1$$

as we would want because $X_1 + X_2 = X_2 + X_1$

Corollary: The operation $*$ is commutative and also associative.