# Some Essential Probability Theory:

## Sigma-fields and Filtrations

Recall: a $\sigma$-field is a collection of random events having a special mathematical structure, so that probabilities can be assigned to all members.

Let the *filtered* probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$, represent the distribution of the process $\{X_t, \ -\infty < t < \infty\}$ – past, present and future.

Here,

$$\mathbf{F} = \{\mathcal{F}_t, -\infty < t < \infty\}$$

where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ represents the $\sigma$-field generated by $\{X_s, s \le t\}$: the 'history of the process up to date $t$'.

- **F** is called a *filtration*.

# Points

- If we observe $X_t$, then we know whether or not all the random events of the form $X_t < x$ have occurred, for all $x$. Also, whether or not $x_1 < X_t < x_2$, for all pairs of numbers $x_1 < x_2$, etc., etc.
- The collection of all the events of this form, relating to all $X_{t-j}$ for $j \geq 0$, including all complements and countable unions, constitutes $\mathcal{F}_t$.
- The sequence $\{\mathcal{F}_t\} = \ldots, \mathcal{F}_{t-1}, \mathcal{F}_t, \mathcal{F}_{t+1}, \ldots$ is *nondecreasing* – assuming information not forgotten. Write

$$\ldots \mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \ldots \subseteq \mathcal{F}$$

- More generally, $\{\mathcal{F}_t\}$ can represent the history of *many* sequences. $X_t$ is called $\mathcal{F}_t$*-measurable* if is predictable exactly from the information in $\mathcal{F}_t$.
- If $X_t$ is $\mathcal{F}_t$-measurable it is also $\mathcal{F}_{t+1}$ -measurable, $\mathcal{F}_{t+2}$-measurable, etc.

# Conditional Expectations.

Given a random variable $Y$ distributed on $(\Omega, \mathcal{F}, P, \mathbf{F})$,

$$E(Y|\mathcal{F}_t)$$

is a $\mathcal{F}_t$ -measurable random variable, representing the best prediction of $Y$ from the standpoint of an observer who possesses the information in $\mathcal{F}_t$.

(Example: $Y = X_{t+1}$ - the 1-step-ahead forecast)

**Properties**:

1. The Law of Iterated Expectations (LIE):

$$E[E(Y|\mathcal{F}_t)] = E(Y)$$

2. If $X_t$ is a $\mathcal{F}_t$-measurable random variable, then

$$E(X_tY|\mathcal{F}_t) = X_tE(Y|\mathcal{F}_t) \text{ a.s.}$$

- Note, a.s. means “almost surely”, i.e., with probability 1. This tag is needed to qualify statements/equalities involving random variables.
  Note, if $Y = 0$ a.s. then $E(Y) = 0$.

- To rule out exceptions of probability zero is difficult, because it is known that different a.s.-equal ‘versions’ of a conditional expectation exist.

3. $E(Y|\mathcal{F}_t)$ is the MMSE forecast.

By the LIE

$$\begin{aligned} E[(Y - E(Y|\mathcal{F}_t))(Z_t - E(Y|\mathcal{F}_t))] &= E[E(Y - E(Y|\mathcal{F}_t)|\mathcal{F}_t)(Z_t - E(Y|\mathcal{F}_t))] \\ &= 0 \end{aligned}$$

since $Z_t - E(Y|\mathcal{F}_t)$ is an $\mathcal{F}_t$-measurable random variable and .

$$E[E(Y - E(Y|\mathcal{F}_t))|\mathcal{F}_t)] = E(Y) - E(Y) = 0 \text{ (LIE).}$$

Hence, for all $\mathcal{F}_t$ -measurable r.v.s $Z_t$,

$$\begin{aligned} E(Y - Z_t)^2 &= E[Y - E(Y|\mathcal{F}_t) + E(Y|\mathcal{F}_t) - Z_t]^2 \\ &= E(Y - E(Y|\mathcal{F}_t))^2 + E(E(Y|\mathcal{F}_t) - Z_t)^2 \geq E(Y - E(Y|\mathcal{F}_t))^2. \end{aligned}$$

4. The prediction errors are uncorrelated with every $\mathcal{F}_t$-measurable random variable $X_t$. By LIE,

$$E[X_t(Y - E(Y|\mathcal{F}_t))] = E(X_tY) - E[E(X_tY|\mathcal{F}_t)] = E(X_tY) - E(X_tY) = 0.$$

5. (Multi-step prediction) If $\ldots\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \ldots$... then for $j \geq 0$, $k \geq 0$,

$$E(E(X_t|\mathcal{F}_{t-j})|\mathcal{F}_{t-j-k}) = E(X_t|\mathcal{F}_{t-j-k}).$$

# Application: Martingales

A fundamental concept in time series analysis. Consider sequence of random variables

$$\{S_t\} = \ldots, S_{t-1}, S_t, S_{t+1}, \ldots$$

and an associated increasing sequence of -fields

$$\{\mathcal{F}_t\} = \ldots, \mathcal{F}_{t-1}, \mathcal{F}_t, \mathcal{F}_{t+1}, \ldots$$

such that $S_t$ is $\mathcal{F}_t$ -measurable.

- Note, $S_{t-k}$ also $\mathcal{F}_t$-measurable for all $k > 0$.
- We call $(S_t, \mathcal{F}_t)$ an *adapted* pair.

**Definition**: $\{S_t\}$ is called a martingale if it is adapted to $\{\mathcal{F}_t\}$ and for all $t$,

$$E|S_t| < \infty$$

$$E(S_t|\mathcal{F}_{t-1}) = S_{t-1} \text{ a.s.}$$

**Interpretation**: a martingale is a process whose *increments* are unpredictable one-step ahead.

- Must stipulate $E|S_t| < \infty$ to ensure $E(S_t|\mathcal{F}_{t-1})$ well-defined.

# Martingale Differences

The sequence of the increments of a martingale is called a martingale difference process (m.d.).

**Definition**: $\{X_t\}$ is a m.d. process if it is adapted to $\{\mathcal{F}_t\}$ and for all $t$, $E|X_t| < \infty$ and $E(X_t|\mathcal{F}_{t-1}) = 0$.

- *Example:* an independent process is also an m.d. process.
- Martingale differences are *serially uncorrelated*. For any $j > 0$, note that

$$\begin{aligned} E(X_t X_{t-j}) &= E(E(X_t X_{t-j}|\mathcal{F}_{t-j})) \quad \text{(LIE)} \\ &= E(E(X_t|\mathcal{F}_{t-j})X_{t-j}) \\ &= E(0.X_{t-j}) = 0 \end{aligned}$$

  However, serially uncorrelated processes are not necessarily m.d.

- The m.d. property is not *reversible* The process $\{X_{-t}, t \geq 1\}$ is not a m.d., although it is uncorrelated.
- One of the main reasons the martingale is an important concept in time series analysis is that a number of useful properties are known for m.d. processes.

# Asymptotic Theory

$$X_1, X_2, X_3, \dots, X_n, \dots = \{X_n\}.$$

$\{X_n\}$ *converges* to $X$ if for $\varepsilon > 0$, there exists $n_\varepsilon \geq 1$ such that $|X_n - X| < \varepsilon$ for all $n > n_\varepsilon$.

Denoted by $\lim_{n\to\infty} X_n = X$.

Suppose $\{X_n\}$ a random sequence with distribution function $\{F_n\}$.

**Definition** If $\{F_n(x)\}$ converges to $F(x)$ *for* all continuity points of $F$, $F_n$ is said to *converge weakly* to $F$, ($F_n \Rightarrow F$).

$\{X_n\}$ is said to *converge in distribution* to $X$, where $X$ is a random variable with distribution function $F$ ($X_n \xrightarrow{\mathrm{d}} X$).

**Definition** If $X$ is a random variable, and for all $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

then $X_n$ *converges in probability* ($X_n \xrightarrow{\mathrm{pr}} X$).

$X$ is the *probability limit* (plim) of $X_n$.

**Definition** If

$$\lim_{n\to\infty} E(X_n - X)^2 = 0$$

$X_n$ *converges in mean square*, ($X_n \xrightarrow{\text{ms}} X$).

**Definition** If

$$X_n(\omega) \to X(\omega)$$

for each $\omega \in C$, where $P(C) = 1$, then $X_n$ *converges almost surely to* $X$ ($X_n \xrightarrow{\text{as}} X$, or $X_n \longrightarrow X$ w.p.1)

- If $X_n \xrightarrow{\text{ms}} X$ then $X_n \xrightarrow{\text{pr}} X$.
- If $X_n \xrightarrow{\text{as}} X$ then $X_n \xrightarrow{\text{pr}} X$.
- If $X_n \xrightarrow{\text{pr}} X$ then $X_n \xrightarrow{\text{d}} X$.

The reverse implications don’t hold.

# Essential Results on Stochastic Convergence

## Slutsky's Theorem

If $X_n \overset{\text{pr}}{\to} a$, and $g(\bullet)$ is continuous at $a$, then $\operatorname{plim} g(X_n) = g(a)$.

## Cramér's Theorem

If $Y_n \xrightarrow{\text{d}} Y$ and $X_n \xrightarrow{\text{pr}} a$ then

(i) $X_n + Y_n \xrightarrow{\text{d}} a + Y$

(ii) $X_n Y_n \xrightarrow{\text{d}} aY$

(iii) $\dfrac{Y_n}{X_n} \xrightarrow{\text{d}} \dfrac{Y}{a}$ when $a \neq 0$.

## Continuous Mapping Theorem

If $X_n \xrightarrow{\text{d}} X$ and $g(\bullet)$ is continuous, then $g(X_n) \xrightarrow{\text{d}} g(X)$.

# Orders of Magnitude

If $\{X_n\}$ a non-random sequence and $\{X_n/n^r\}$ remains bounded in the limit as $n \to \infty$, write $X_n = O(n^r)$.

**Definition** If $\{X_n\}$ is a random sequence and for $\varepsilon > 0$ there exists $B_\varepsilon < \infty$ such that $P\left(\frac{|X_n|}{n^r} > B_\varepsilon\right) < \varepsilon$ for all $n \geq 1$, write $X_n = O_p(n^r)$. If $\operatorname{plim} X_n/n^r = 0$, write $X_n = o_p(n^r)$.

If $Z_n = O_p(n^{r_1})$ and $W_n = O_p(n^{r_2})$ then

$$Z_n + W_n = O_p(n^{\max(r_1,r_2)})$$

$$Z_n W_n = O_p(n^{r_1+r_2}).$$

**Definition** ‘*Uniformly in n*’

“$X_n < \infty$ for every $n \geq 1$” is satisfied by (e.g.) $X_n = n$,

The statement “$X_n < \infty$ uniformly in $n$” is equivalent to

“$X_n \leq B$ for every $n \geq 1$ where $B < \infty$”.

# Weak Laws of Large Numbers (WLLN)

A “weak law” asserts convergence in probability.

A “strong law” asserts convergence with probability 1 (a.s.).

- The ergodic theorem is an example of a strong law.

**Khinchine’s Theorem** If $\{x_t\}$ is i.i.d. and $E(x_t) = \mu < \infty$. then

$$\bar{x}_n \xrightarrow{\text{pr}} \mu.$$

**Chebyshev’s Theorem** If $E(x_t) = 0$ and $\lim_{n\to\infty} E(\bar{x}_n^2) \to 0$, then

$$\bar{x}_n \xrightarrow{\text{pr}} 0.$$

- In an uncorrelated sequence with $E(x_t^2) = \sigma^2$ and $E(x_t x_s) = 0$ for $t \neq s$,

$$\bar{x}_n = O_p(n^{-1/2}).$$

- Since martingale differences are uncorrelated, Chebyshev’s theorem applies to these sequences.

# The Central Limit Theorem (CLT)

**Lindeberg–Lévy Theorem**: If $x_t \sim iid(0,\sigma^2)$, then

$$\sqrt{n}\,\frac{\bar{x}_n}{\sigma} = \frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n} x_t \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Also write "$\underset{\text{asy}}{\sim}$" as alternative notation to "$\xrightarrow{\mathrm{d}}$"

The CLT is also a fundamental property of martingale differences.

**Theorem – Martingale CLT**:

Let $y_t$ be a stationary ergodic m.d. process, with mean 0 and variance $\sigma^2 < \infty$. Then,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t \xrightarrow{\mathrm{d}} N(0,\sigma^2)$$

- Note: The CLT does *not* hold for uncorrelated sequences, without additional restrictions on dependence.
- By contrast, the m.d. property is a sufficient restriction on dependence.

# Vectors

**Cramér–Wold Theorem** (or "Cramer-Wold Device")

$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ if and only if, $\boldsymbol{\lambda}'\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\lambda}'\boldsymbol{X}$ for all conformable fixed vectors $\boldsymbol{\lambda}$.

Thence,

**Lindeberg–Lévy Theorem for vectors**

If $\{\mathbf{x}_t\}$ is a sequence of $m$-vectors, $\mathbf{x}_t \sim \mathrm{i.i.d.}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\bar{\mathbf{x}}_n = n^{-1}\sum_{t=1}^{n}\mathbf{x}_t$, then

$$\sqrt{n}\,\bar{\mathbf{x}}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

**Cramér's Theorem for vectors**

If $\boldsymbol{X}_n \xrightarrow{d} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $(n \times 1)$ and $\operatorname{plim}\boldsymbol{A}_n = \boldsymbol{A}$ $(m \times n)$ then

$$\boldsymbol{A}_n\boldsymbol{X}_n \xrightarrow{d} \mathrm{N}(\mathbf{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')(m \times 1).$$

# Vector Martingale Differences

Suppose that $\{\boldsymbol{x}_t\}$ is a vector of processes adapted to $\{\mathcal{F}_t\}$.

**Definition**: If, for any constant vector $\boldsymbol{a}$, the process $\boldsymbol{a}'\boldsymbol{x}_t$ is a m.d., such that

$$E(\boldsymbol{a}'\boldsymbol{x}_t|\mathcal{F}_{t-1}) = 0 \text{ a.s.}$$

the vector $\boldsymbol{x}_t$ is called a vector m.d.

- Not the same thing as a vector of m.d.s. A counter-example is

$$\boldsymbol{x}_t = \begin{bmatrix} x_{1t} \\ x_{1,t+1} \end{bmatrix}$$

where $x_{1t}$ is an m.d.

- The elements of a v.m.d. are individually m.d.s with respect to $\mathcal{F}_t$.

# Nonstationarity

One of the powerful features of the martingale approach: stationarity is not necessary for the important results.

1. **The Weak Law of Large Numbers**

Suppose $\mathrm{Var}(x_t) = \sigma_t^2 \leq B < \infty$. We cannot rely on the ergodic property, which requires stationarity. However, if $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$ , then note that

$$E(\bar{x}) = 0$$

$$\mathrm{Var}(\bar{x}) = \frac{1}{n^2}\sum_{t=1}^{n}\sigma_t^2 = O(n^{-1})$$

since the process is uncorrelated.

By Chebyshev's inequality, this condition is sufficient for

$$\bar{x} \xrightarrow{\mathrm{pr}} 0.$$

In fact, the sequence of variances can be diverging!

- $\sigma_t^2 = O(t^{1-\delta})$ for $\delta > 0$ is allowed

## 2. **The Central Limit Theorem**

There is a version of the CLT for non-ergodic/stationary martingale differences, with variances $\{\sigma_t^2\}$.

The conditions are technical, but the main difference is that we must *additionally* assume

$$n^{-1}\sum_{t=1}^{n}(x_t^2-\sigma_t^2)\xrightarrow{\text{pr}}0.$$