Contents

Notations

Use the template *acronym.tex* together with the Springer document class SVMono (monograph-type books) or SVMult (edited books) to style your list(s) of abbreviations or symbols in the Springer layout.

Lists of abbreviations, symbols and the like are easily formatted with the help of the Springer-enhanced `description` environment.

ABC Spelled-out abbreviation and definition
BABI Spelled-out abbreviation and definition
CABR Spelled-out abbreviation and definition
I indicator function
$\mathcal{N}$ normal distribution (Gaussian distribution)
$\|\cdot\|_p$ $(E|\cdot|)^{1/p}$

Chapter 1
Fundamental Mathematics

1.1 Operators

1.1.1 vec

For any $p \times q$ matrix A, we call $\mathrm{vec}(A)$ the vector got by putting a_{ij} in row $(j+1)p+i$.

1.1.2 $\otimes$

For matrices $A^{(1)}$ $(p_1 \times q_1)$ and $A^{(2)}$ $(p_2 \times q_2)$, the Kroneker product $A^{(1)} \otimes A^{(2)}$ is defined by the components $a^{(1)}_{ij} a^{2}_{kl}$ in row $(i-1)p_2 + k$ $(1 \leq (i-1)p_2 + k \leq p_1 p_2)$, column $(j-1)q_2 + l$ $(1 \leq (i-1)p_2 + k \leq q_1 q_2)$.

1.2 Algebra

Let $(\Omega, \mathcal{F}, P)$ be a probability-measure space. The expectation operator E is defined as an integration on Ω, that is,

$$EX = \int_\Omega X(\omega) P(d\omega). \tag{1.1}$$

E is well know to be a linear map. Suppose X and Y is defined on Ω and a is a constant.

$$E(X+Y) = E(X) + E(Y), \tag{1.2}$$

$$E(aX) = aE(X). \tag{1.3}$$

The covariance between two jointly distributed real-valued random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y). \tag{1.4}$$

When $X = Y$ almost surely, we call it the variance of X.

The random variables are denoted by the capital letters and the constants are denoted by the small letters. The algebra of covariance is given by

$$\operatorname{Cov}(X, X) = \operatorname{Var}(X), \tag{1.5}$$
$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X), \tag{1.6}$$
$$\operatorname{Cov}(aX, bY) = ab \operatorname{Cov}(X, Y), \tag{1.7}$$
$$\operatorname{Cov}(X + a, Y + b) = \operatorname{Cov}(X, Y). \tag{1.8}$$

In general, we have

$$\operatorname{Cov}(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j \operatorname{Cov}(X_i, Y_j), \tag{1.9}$$
$$\operatorname{Var}(\sum_i a_i X_i) = \sum_i \sum_j a_i a_j \operatorname{Cov}(X_i, X_j). \tag{1.10}$$

1.2.1 Geometric Series

A famous formula for the sum of geometric series is given by

$$\sum_{t=1}^{N} z^t = z \frac{1 - z^N}{1 - z}. \tag{1.11}$$

Example 1.1 *Let* $\lambda_k = \frac{2\pi k}{n}$, $k = 0, \pm 1, \ldots, \pm(n-1)$. *Then*

$$\sum_{t=1}^{n} e^{it\lambda_k} = \begin{cases} n & \text{for } k = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

The example shows the finite Fourier transform of any random variables at nonzero natural frequencies is invariant to centering.

1.3 Calculus

Suppose that the real functions f and g are defined on an interval T. Then on this interval,

$$\inf_T (f + g) \geq \inf_T f + \inf g, \tag{1.13}$$
$$\sup_T (f + g) \leq \sup_T f + \sup g. \tag{1.14}$$

Furthermore, if f and g are also nonnegative on T, then

$$\inf_T (f \cdot g) \geq \inf_T f \cdot \inf g, \tag{1.15}$$

$$\sup_T (f \cdot g) \leq \sup_T f \cdot \sup g. \tag{1.16}$$

Example 1.2 *Set $f(x) = 1/x$ and $g(x) = x$ for $x \in T = [1, 2]$.*

1.3.1 Inequalities

There are 2 basic equations:

$$(a+b)^k \geq a^k + b^k \quad \text{if } k \geq 1, \tag{1.17}$$

$$(a+b)^k \leq a^k + b^k \quad \text{if } 0 \leq k \leq 1. \tag{1.18}$$

1.3.2 Examples

There is a condition, called Gordin's condition

$$\sum_{k=0}^{\infty} \left\{ \sum_{j=k}^{\infty} \alpha(j)^2 \right\}^{1/2} < \infty \tag{1.19}$$

in time series analysis. This condition implies

$$\sum_{j=0}^{\infty} |\alpha(j)| < \infty, \tag{1.20}$$

and

$$\sum_{j=0}^{\infty} \alpha(j)^2 < \infty, \tag{1.21}$$

and is implied by

$$\sum_{j=0}^{\infty} j|\alpha(j)| < \infty. \tag{1.22}$$

In fact, for any $k \geq 0$,

$$|\alpha(k)| = (\alpha(k)^2)^{1/2} \leq \left(\sum_{j=k}^{\infty} \alpha(j)^2 \right)^{1/2}, \tag{1.23}$$

which implies

$$\sum_{k=0}^{\infty} |\alpha(k)| \leq \sum_{k=0}^{\infty} \left(\sum_{j=k}^{\infty} \alpha(j)^2 \right)^{1/2}. \tag{1.24}$$

Also,

$$\Big(\sum_{j=0}^{\infty}\alpha(j)^2\Big)^{1/2} < \Big(\sum_{j=0}^{\infty}(j+1)\alpha(j)^2\Big)^{1/2} \tag{1.25}$$

$$= \Big(\sum_{j=0}^{\infty}\sum_{k=0}^{j}\alpha(j)^2\Big)^{1/2} \tag{1.26}$$

$$= \Big(\sum_{k=0}^{\infty}\sum_{j=k}^{\infty}\alpha(j)^2\Big)^{1/2} \tag{1.27}$$

$$\leq \sum_{k=0}^{\infty}\Big(\sum_{j=k}^{\infty}\alpha(j)^2\Big)^{1/2} \tag{1.28}$$

On the other hand, we can see that

$$\begin{aligned}\sum_{k=0}^{\infty}\left\{\sum_{j=k}^{\infty}\alpha(j)^2\right\}^{1/2} &\leq \sum_{k=0}^{\infty}\sum_{j=k}^{\infty}|\alpha(j)| \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{j}|\alpha(j)| \\ &= \sum_{j=0}^{\infty}(j+1)|\alpha(j)|.\end{aligned}$$

1.4 Measure Theory

First, we give Levesgue's monotone convergence theorem.

Theorem 1.3 (monotone convergence theorem) *Let (S, Σ, μ) be a measurable space. Let f_1, f_2, ... be a pointwise non-decreasing non-negative sequence of measurable functions, i.e.,*

$$0 \leq f_k(x) \leq f_{k+1}(x) \quad \text{for any } k \geq 1 \text{ and any } x \in S. \tag{1.29}$$

If the sequence converges pointwise to a function f, then f is measurable and

$$\lim_{k\to\infty}\int f_k\, d\mu = \int f\, d\mu. \tag{1.30}$$

One of the most important theorems in measure theory is Lebesgue's dominated convergence theorem.

Theorem 1.4 (Lebesgue's dominated convergence theorem) *Let $\{f_n\}$ be a sequence of real-valued measurable functions on a measure space (S, Σ, μ). Suppose that the sequence converges pointwise to a function f and is dominated by some integrable function g, i.e.,*

$$|f_n(x)| \leq g(x) \tag{1.31}$$

for all numbers n and all $x \in S$. Then f is integrable and

$$\lim_{n\to\infty} \int_S |f_n - f| \, d\mu = 0. \tag{1.32}$$

Another important theorem is that the integration is absolutely continuous with respect to the measure space.

Theorem 1.5 *Suppose that f is integrable. Then,*

$$\mu(A) \to 0 \quad \Rightarrow \quad \int_A f \, d\mu \to 0. \tag{1.33}$$

Proof. Let f_n be defined by

$$f_n := \begin{cases} f & \text{if } |f| \leq n, \\ n & \text{if } |f| > n. \end{cases} \tag{1.34}$$

By monotone convergence theorem, we have

$$\int_S |f| \, d\mu = \lim_{n\to\infty} \int_S |f_n| \, d\mu. \tag{1.35}$$

Therefore, the conclusion is implied by the inequality

$$\Big| \int_A f \, d\mu \Big| \leq \int_A |f| - |f_n| \, d\mu + \int_A |f_n| \, d\mu \leq \int_S |f| - |f_n| \, d\mu + n\mu(A). \tag{1.36}$$

1.4.1 Karamata's Theorem

Definition 1.6 (Regular Varying) *A measurable function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at ∞ with index ρ if for $x > 0$,*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho. \tag{1.37}$$

ρ is called the exponent of variation. If $\rho = 0$ then U is called slowly varying. Slowly varying functions are generically denoted by $L(x)$.

Note that $U(x)$ is regularly varying at ∞ if and only if $U(x^{-1})$ is regularly varying at 0.

Theorem 1.7 (i) *If $\rho \geq -1$ then $U \in \mathrm{RV}_\rho$ implies*

$$\int_0^x U(t)dt \in \mathrm{RV}_{\rho+1} \tag{1.38}$$

and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \rho + 1. \tag{1.39}$$

If $\rho < -1$ or if $\rho = -1$ with $\int_x^\infty U(s)ds < \infty$, then $U \in \mathrm{RV}_\rho$ implies $\int_x^\infty U(t)dt$ is finite,

$$\int_x^\infty U(t)dt \in \mathrm{RV}_{\rho+1} \tag{1.40}$$

and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = -\rho - 1. \tag{1.41}$$

(ii) *If U satisfies*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \lambda \in (0, \infty), \tag{1.42}$$

then $U \in \mathrm{RV}_{\lambda-1}$. If $\int_x^\infty U(t)dt < \infty$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = \lambda \in (0, \infty), \tag{1.43}$$

then $U \in \mathrm{RV}_{-\lambda-1}$.

1.5 Functional Analysis

Consider

$$f(\omega) = \sum_{n=0}^{\infty} (A_n \cos n\omega + B_n \sin n\omega),$$

where $\sum_{n=0}^\infty (|A_n| + |B_n|) < \infty$. The values of f determined on any interval of length 2π. A standard choice is the interval $\mathbb{T} = (-\pi, \pi]$, where we identify 2π-periodic functions on $\mathbb{R}$ with functions on $\mathbb{T}$. The alternative way to representate it is to rewrite it in the complex form

$$f(\omega) = \sum_{n=-\infty}^{\infty} C_n e^{in\omega}. \tag{1.44}$$

Theorem 1.8 *Suppose that $\sum_{n\in\mathbb{Z}} |C_n| < \infty$. Then f defined by (1.44) is a continuous function on $\mathbb{T}$. The coefficients are obtained as*

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{-in\omega} d\omega, \quad n \in \mathbb{Z}. \tag{1.45}$$

If g is any other L^1 function on $\mathbb{T}$, we have the Fourier reciprocity formula

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) g(\omega) d\omega = \sum_{n\in\mathbb{Z}} C_n D_{-n}$$

where D_n is the Fourier coefficient of g, defined by (1.45) with f replaced by g. In particular we have Parseval's identity

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\omega)|^2 d\omega = \sum_{n\in\mathbb{Z}}|C_n|^2.$$

[Inverse Fourier transformation]

Note that if $f(\omega)$ has a peak at λ, then C_n repeat itself on average after $2\pi/\lambda$. By the inverse Fourier transform, we have

$$C_{k+2\pi/\lambda} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{-i(k+2\pi/\lambda)\omega)}\,d\omega \tag{1.46}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{-}ik\omega e^{-i}\,d\omega \tag{1.47}$$

$$\tag{1.48}$$

Proposition 1.9 *Suppose that* $\sum_{n\in\mathbb{Z}}|n^k C_n| < \infty$ *for some* $k = 2, 3, \ldots$. *Then* $f(\omega) := \sum_{-\infty}^{\infty} C_n e^{in\omega}$ *is a* k*-times defferentiable function with* $f^{(k)}(\omega) = \sum_{n\in Z}(in)^k C_n e^{in\omega}$ *a continuous function.*

Corollary 1.10 *The convolution of an absolutely convergent trigonometric series* f *with an arbitrary* L^1 *function* g *has the representation*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)g(\lambda-\omega)d\omega = \sum_{n\in\mathbb{Z}} C_n D_n e^{in\lambda}.$$

1.5.1 Factorial and Bessel Functions

Let $C_n = 0$ for $n \le 0$ and $C_n = r^n/n!$ where $r \ge 0$ and $n = 1, 2, \ldots$. Then we have

$$f(\omega) = \sum_{n=0}^{\infty}\frac{r^n}{n!}e^{in\theta} = \sum_{n=0}^{\infty}\frac{(re^{i\theta})^n}{n!} = \exp(re^{i\theta}),$$

and then

$$\frac{r^n}{n!} = \frac{1}{2\pi}\int_{-\pi}^{\pi} exp(re^{\theta})\exp(-in\omega)d\omega, \quad r \ge 0,\ n = 0, 1, \ldots.$$

Here we define $I(2r)$ as

$$I(2r) = \sum_{n=0}^{\infty}\left(\frac{r^n}{n!}\right)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} exp(2r\cos\omega)d\omega, \quad r \ge 0.$$

1.5.2 integration

Whether $m = n$ or $m \neq n$

$$\int_{-\pi}^{\pi} \cos(m\omega)\sin(n\omega)\,d\omega = 0 \tag{1.49}$$

When $m \neq n$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = 0 \tag{1.50}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = 0 \tag{1.51}$$

$$\int_{0}^{\pi} \sin m\omega \sin n\omega\,d\omega = 0 \tag{1.52}$$

When $m = n(\neq 0)$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = \pi \tag{1.53}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = \pi \tag{1.54}$$

When $m = n = 0$ then

$$\int_{-\pi}^{\pi} \cos(m\omega)\cos(n\omega)\,d\omega = 2\pi \tag{1.55}$$

$$\int_{-\pi}^{\pi} \sin(m\omega)\sin(n\omega)\,d\omega = 0 \tag{1.56}$$

1.6 The Hermite polynomials

We introduce the Hermite polynomials used in the area of probability and statistics. They have a little different definitions in the area of physics. The Hermite polynomials $H_n(x)$ are defined by the relations

$$\left(\frac{d}{dx}\right)^n e^{-\frac{x^2}{2}} = (-1)^n H_n(x) e^{-\frac{x^2}{2}} \quad (n = 0, 1, \dots). \tag{1.57}$$

As a note, the Hermite polynomials $\tilde{H}_n(x)$ in physics are defined by

$$\tilde{H}_n(x) = 2^{n/2} H_n(\sqrt{2}x).$$

$H_n(x)$ is a polynomial of degree n, and we have

$$
\begin{aligned}
&H_0(x) = 1, && (1.58)\\
&H_1(x) = x, && (1.59)\\
&H_2(x) = x^2 - 1, && (1.60)\\
&H_3(x) = x^3 - 3x, && (1.61)\\
&H_4(x) = x^4 - 6x^2 + 3, && (1.62)\\
&H_5(x) = x^5 - 10x^3 + 15x, && (1.63)\\
&H_6(x) = x^6 - 15x^4 + 45x^2 - 15, && (1.64)\\
&\quad \cdots
\end{aligned}
$$

Note that

$$\frac{d^{k+1}}{dx^{k+1}} H_k(x) \equiv 0, \quad \text{for } k = 0, 1, \ldots,$$

we obtain by repeated partial integration

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) d\Phi(x) = \begin{cases} n! & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases} \tag{1.65}$$

Therefore, $\{\frac{1}{\sqrt{n!}} H_n(x)\}$ is the sequence of orthogonal polynomials associated with the normal distribution. With the idea of exponential generating function, we have

$$\sum_{k=0}^{\infty} \frac{H_k(x)}{k!} t^k = e^{-\frac{t^2}{2} + tx}, \tag{1.66}$$

and

$$\sum_{k=0}^{\infty} \frac{H_k(x)}{k!} \frac{H_k(y)}{k!} t^k = \frac{1}{\sqrt{1-t^2}} e^{-\frac{t^2x^2 + t^2y^2 - 2txy}{2(1-t^2)}}, \tag{1.67}$$

1.7 The Appell polynomials

An extension of the Hermite polynomials is the class of the Appell polynomial. The Appell polynomials are defined by

$$\sum_{k=0}^{\infty} \frac{A_k(x)}{k!} t^k = \frac{e^{tx}}{Ee^{tX}}, \quad z \in \mathbb{C}, \tag{1.68}$$

and then

$$\frac{d}{dx} A_j(x) = j A_{j-1}(x). \tag{1.69}$$

1.8 Probability

Suppose the random variables X and Y are independent with distribution function

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y). \tag{1.70}$$

Evaluation of $EX1(Y + \alpha X < 0)$ with $EX = 0$ is interesting in time series analysis.

$$EX1(Y + aX < 0) = \int_{\mathbb{R}^2} x1(y + \alpha x < 0)dF_X(x)dF_Y(y) \tag{1.71}$$
$$= \int_{\mathbb{R}} xP_Y(Y < -\alpha x)dF_X(x) \tag{1.72}$$
$$= E_X XF_Y(-\alpha X). \tag{1.73}$$

Since the distribution function is nondecreasing,

$$F_Y(-\alpha X1(X > 0)) \geq F_Y(-\alpha X1(X \leq 0)) \quad \text{a.s.} \quad \text{if } \alpha < 0, \tag{1.74}$$
$$F_Y(-\alpha X1(X > 0)) \leq F_Y(-\alpha X1(X \leq 0)) \quad \text{a.s.} \quad \text{if } \alpha > 0. \tag{1.75}$$

Thus $E_X XF_Y(-\alpha X) \neq 0$ if $\alpha \neq 0$.

Theorem 1.11 *Suppose that for each u, $X_{un} \rightsquigarrow X_u$ as $n \to \infty$, and that $X_u \rightsquigarrow X$ as $u \to \infty$. Suppose further that*

$$\lim_{u\to\infty} \limsup_{n\to\infty} P\{\rho(X_{un}, Y_n) \geq \epsilon\} = 0 \tag{1.76}$$

for each posetive ϵ. Then $Y_n \rightsquigarrow X$ as $n \to \infty$.

The quantiles of X is defined by quantile function

$$\xi_0(\tau) := \inf\{x : P(X \leq x) \geq \tau\}. \tag{1.77}$$

1.8.1 Conditional Expectation

Theorem 1.12 (Jensen's Inequality) *Let ϕ be a convex function. Then for any random variable X and σ-field $\mathcal{H}$,*

$$\phi(E(X|\mathcal{H})) \leq E(\phi(X)|\mathcal{H}). \tag{1.78}$$

Theorem 1.13 (Tower property) *If σ-field $\mathcal{H} \subset \mathcal{G}$, then*

$$E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H}) = E(E(X|\mathcal{H})|\mathcal{G}). \tag{1.79}$$

Example 1.14 *Suppose $X \in \mathcal{L}^p$ with σ-field $\mathcal{G}$. Then*

$$E|X|^p \geq E(E(|X|\,|\,\mathcal{G})^p) \geq (E|X|)^p. \tag{1.80}$$

For the variance of conditional expectation, the following equation is well known:

$$V(X) = E(V(X|Y)) + V(E(X|Y)). \tag{1.81}$$

1.8.2 Generalized Domain of Attraction

Let X be a real Banach space, that is, X is a real linear, normed, complete space, with norm $\|\cdot\|$. By X^* we denote its *topological dual Banach*, that is, $x^* \in X^*$ are continuous linear functionals on X, and $\langle \cdot, \cdot \rangle$ is the dual pair between X^* and X. When the norm in X is given by a scalar product, X is called a *Hilbert* space. In that case, X^* is isomorphic to X and the dual pair is simply the scalar product. Furthermore, all real separable Hilbert spaces are isomorphic to l_2, the space of all real square-summable sequences with

$$\langle x, y \rangle := \sum_i x_i y_i, \quad \|x\| := \langle x, x \rangle^{1/2}.$$

The collection $L(X, Y)$ of all bounded linear operators from X into Y, using the operator norm, is also Banach space. Here, the assumption that A is bounded and linear is equivalent to A being continuous and linear form X to Y, where the topologies are given by the norms. When $X = Y$, $L(X, Y)$ is denoted by $\mathrm{End}(X)$; in which case, we also have that the product of two operators in $\mathrm{End}(X)$ is a continuous linear operator: if $A, B \in \mathrm{End}(X)$, then $AB : X \to X$ is given by $(AB)x = A(Bx)$ for $x \in X$. Moreover, $\|AB\| \leq \|A\|\|B\|$ for all $A, B \in \mathrm{End}(X)$. With this multiplication of operators, $\mathrm{End}(X)$ becomes a topological semigroup. By $\mathrm{Aut}(X)$, we denote the set of all invertible operators in $\mathrm{End}(X)$. These inverse are also continuous and linear, so $\mathrm{Aut}(X)$ is a topological group.

Theorem 1.15 *Let ξ_n, ξ be $\mathbb{R}^d$-valued random variables. Then ξ_n converges in distribution to ξ in $\mathbb{R}^d$ if and only if for every $a \in \mathbb{R}^d$, $\langle a, \xi_n \rangle$ converges in distribution to $\langle a, \xi \rangle$ in $\mathbb{R}^1$.*

Lemma 1.1 *Consider symmetrization of μ, i.e. $\mu^0 := \mu * \mu^-$. Then the characteristic function of μ^0 is real-valued.*

Lemma 1.2 *In the case of a separable metric space,* suppμ *always exists.* suppμ = {x ∈ X : *for every open G containing x, $\mu(G) \neq 0$*}.

Proposition 1.16 *Let μ, $\nu \in \mathcal{P}(X)$. Then*

$$\mathrm{supp}(\mu * \nu) = \overline{(\mathrm{supp}\,\mu + \mathrm{supp}\,\nu)}.$$

A more general proposition is given as follows:

Proposition 1.17 *Let μ be a probability on the topological space S_1 and let $f : S_1 \to S_2$ be a continuous mapping into the topological space S_2. Then*

$$f\mu = \overline{f\mathrm{supp}\,\mu}$$

In particular, for Banach spaces X and Y, probability μ on X, and a bounded linear operator $A : X \to Y$, we obtain

$$\mathrm{supp}(A\mu) = A(\mathrm{supp}\mu).$$

Proposition 1.18 *Let $\mu \in \mathcal{P}(X)$. Then $(\mathrm{supp}\,\mu)^\perp = \{x^* \in X^* : \hat{\mu}(tx^*) = 1$ for all $t \in \mathbb{R}^1\}$.*

1.8.3 Infinitely Divisible and Stable

Definition 1.19 *A probability μ on a Banach space X is said to be infinitely divisible if for each integer $n \geq 2$ there exists an element $\mu_n \in \mathcal{P}(X)$ such that $\mu_n^n = \mu$, where the nth power of a probability is taken in the sense of convolution.*

Definition 1.20 *A measure $\mu \in \mathcal{P}$ is called operator-stable if there are a measure $\nu \in \mathcal{P}$, a sequence $\{A_n\}$ of linear operators, and a sequence $\{a_n\}$ of vectors such that*

$$A_n \nu^n \delta(a_n) \Rightarrow \mu.$$

1.8.4 Basic Concepts

We say that a measure μ on $\mathbb{R}^d$ is *full* if its support is not contained in any proper hyperplane of $\mathbb{R}^d$, that is, for any x in $\mathbb{R}^d$ and any subspace W of $\mathbb{R}^d$ with $\dim W < d$, we have $\mu(W + x) < 1$.

1. The idea of fullness is the natural extension of nondegeneracy on $\mathbb{R}^1$.
2. It is shown that the set of all full measures is an open subsemigroup of $\mathcal{P}(\mathbb{R}^d)$.

Generally, the set of all full measures on $\mathbb{R}^d$ is denoted by $\mathcal{F}(\mathbb{R}^d)$. Also the set $H(\mu)$ is defined as

$$H(\mu) = \{y \in \mathbb{R}^d; \hat{\mu}(y) = 1\}.$$

Proposition 1.21 *The following statements are equivalent.*

1. *μ is full.*
2. *μ^0 is full.*
3. *$H(\mu^0)$ does not contain any one-dimensional subspace.*
4. *For each $y \neq 0$, the measure $\Pi_y \mu$ is nondegenerate on $\mathbb{R}$ where $\Pi_y(x) = \langle x, y \rangle$ for $x \in \mathbb{R}^d$.*

Corollary 1.22 *Let $A \in \mathrm{End}(\mathbb{R}^d)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then $A\mu$ is full if and only if A is invertible and μ is full.*

For the Banach space X, let $\mathcal{A}(X)$ denote the set of all *affine* transformations on X, that is, each $\alpha \in \mathcal{A}(X)$ is given by an operator $A \in \mathrm{End(X)}$ and a vector $a \in X$, $\alpha := \langle A; a \rangle$, in the following way:

$$\alpha x := Ax + a.$$

In the same way, $r\alpha := \langle rA; ra \rangle$. The set $\mathcal{A}(X)$ is equipped by the norm

$$\|\alpha\| := \max\{\, \|A\|, \|a\|\, \},$$

then it is a Banach space.

Corollary 1.23 *Let $\alpha_n, \alpha \in \mathcal{A}(X)$, and assume $\alpha_n x \to \alpha x$ for all $x \in X$. Then $\mu_n \Rightarrow \mu$ in $\mathcal{P}(X)$ implies that $\alpha_n \mu_n \Rightarrow \alpha\mu$.*

By $\mathcal{A}_I$, we denote the set of all invertible affine transformations on $\mathbb{R}^d$.

Corollary 1.24 *If $\alpha_n\mu_n \Rightarrow \mu$ with $\mu_n \in \mathcal{P}$, $\mu \in \mathcal{F}$, and $\alpha_n \in \mathcal{A}$, then $\alpha_n \in \mathcal{A}_I$ and μ_n is full for all sufficiently large n.*

Next, we introduce the concept which is called "conditionally compact". (The concept is called "relatively compact" in some books.)

Definition 1.25 *A subset Γ of $\mathcal{P}(S)$ is called conditionally compact if every sequence $\{\mu_n\}$ in Γ contains a subsequence which is weakly convergent in $\mathcal{P}(S)$; the limit probability need not be in Γ.*

Definition 1.26 *A subset Γ of $\mathcal{P}(S)$ is called tight if for every $\epsilon > 0$, there is a compact set K such that $\mu(K) > 1 - \epsilon$ for all $\mu \in \Gamma$.*

Theorem 1.27 (Prohorov's Theorem) *For a metric space S, every tight set Γ in $\mathcal{P}(S)$ is conditionally compact. When S is separable and complete, Γ being conditionally compact implies that Γ is tight.*

Lemma 1.3 *If $\mu_n \Rightarrow \mu$ with μ full and if $\{\alpha_n\mu_n\}$ is tight, where $\alpha_n \in \mathcal{A}$, then $\sup\|\alpha_n\| < \infty$, that is, $\{\alpha_n\}$ is conditionally compact in $\mathcal{A}$.*

In the convergence of types theorems, a fundamental role is played by the set of operators having the property that the limit measure μ is unchanged by the action of one of these operators. More formally, we define *the invariant semigroup* of μ, $\mathrm{Inv}(\mu)$, to be

$$\mathrm{Inv}(\mu) = \{\alpha \in \mathcal{A} : \mu = \alpha\mu\}.$$

Theorem 1.28 *If μ is full, then $\mathrm{Inv}(\mu)$ is a compact subgroup of $\mathcal{A}_I$. Conversely, if μ is nonfull, then $\mathrm{Inv}(\mu)$ is neither a group nor compact.*

Lemma 1.4 *Let $\mu \in \mathcal{P}$ and $\alpha \in \mathcal{A}_I$. Then*

$$\mathrm{Inv}(\alpha\mu) = \alpha(\mathrm{Inv}(\mu))\alpha^{-1}.$$

Definition 1.29 *Two measures μ and ν are of the same operator type provided there is $\alpha \in \mathcal{A}$ such that $\mu = \alpha\nu$.*

Theorem 1.30 *Assume that $\beta_n\mu_n \Rightarrow \mu$, where $\beta_n \in \mathcal{A}$, $\mu_n \in \mathcal{P}$, and μ full. In order that $\alpha_n\mu_n \Rightarrow \nu$, with $\alpha_n \in \mathcal{A}$ and ν full, it is necessary and sufficient that $\nu = \alpha\mu$ for some $\alpha \in \mathcal{A}_I$, that is, μ and ν are of the same operator type, and, for all sufficiently large n,*

$$\alpha_n = \alpha\eta_n\gamma_n\beta_n,$$

where $\eta_n \to \eta_0 = \langle I; 0\rangle$ and $\gamma_n \in \mathrm{Inv}(\mu)$.

1.8.5 Notations and Assumptions

In the sequent subsection, we assume that $X, X_1, X_2, X_3, \ldots$ are i.i.d on R^d with common distribution μ and that μ belongs to the strict generalized domain of attraction of some full operator stable law ν on $\mathbb{R}^d$ with no normal component. If X belongs to the generalized domain of attraction of Y, then there exist linear operator A_n and nonrandom vectors a_n such that

$$A_n(X_1 + X_2 + \cdots + X_n) - a_n \Rightarrow Y.$$

For A_n, we say a sequence of linear operators on R^d is regularly varying with index $(-E)$ if

$$A_{[tn]}A_n^{-1} \to t^{-E},$$

for all $t > 0$. As in the Meerschaert and Scheffler (1999), the notation t^{-E} means $t^{-E} = \exp(-E \log t)$ where $\exp(A) = I + A + A^2/2! + \cdots$ is the usual exponential operator. S_n is used to be the sum of the sample,

$$S_n = X_1 + \cdots + X_n,$$

while M_n is used to represent the sample covariance matrix, i.e.

$$M_n = \sum_{i=1}^{n} X_i X_i'.$$

We give three lemmas from Meerschaert and Scheffler (1999) below.

Lemma 1.5 *Suppose that μ is regularly varying with exponent E and*

$$nA_n\mu \to \phi$$

holds. If every eigenvalue of E has real part exceeding $1/2$ then

$$A_n M_n A_n^* \Rightarrow W$$

where W is infinitely divisible on M_s^d with Lévy representation $[C, 0, T\phi]$.

Lemma 1.6

$$(A_n S_n, A_n M_n A_n^*) \Rightarrow (Y, W).$$

Lemma 1.7 *If $A_n S_n \Rightarrow Y$ and $A_n M_n A_n^* \Rightarrow W$ hold with $A_n = a_n^{-1} I$ then $M_n^{-1/2} S_n \Rightarrow W^{-1/2} Y$.*

If α of marginal distribution of X_1 are different, it is easy to see that we can take

$$A_n = \mathrm{d}iag\,(n^{-1/\alpha_1}, \ldots, n^{-1/\alpha_d}),$$

and E becomes

$$E = \mathrm{d}iag\,(1/\alpha_1, \ldots, 1/\alpha_d).$$

Here, we only think the case that $\alpha_i = \alpha$ for $i = 1, \ldots, d$.

We give three general lemmas, which is examined by Meerschaert and Scheffler (1999).

Lemma 1.8 *Suppose that μ is regularly varying with exponent E and*

$$nA_n\mu \to \phi$$

holds. If every eigenvalue of E has real part exceeding $1/2$ then

$$A_n M_{n,Z} A_n^* \Rightarrow M$$

where M is infinitely divisible on M_s^d with Lévy representation $[C, 0, T\phi]$. Furthermore, the limit M is operator stable with exponent where $\xi M = EM + ME^*$.

Lemma 1.9

$$(A_n S_n, A_n M_{n,Z} A_n^*) \Rightarrow (Y, M).$$

Lemma 1.10 *If $A_n S_n \Rightarrow Y$ and $A_n M_{n,Z} A_n^* \Rightarrow M$ hold with $A_n = a_n^{-1} I$ then $M_{n,Z}^{-1/2} S_n \Rightarrow M^{-1/2} Y$.*

1.9 Statistics

1.9.1 Useful Results

Prohorov's Theorem can be simplified to the following form if the random vectors X_n are in $\mathbb{R}^k$.

Definition 1.31 *A set of random vectors $\{X_\lambda \; \lambda \in \Lambda\}$ is called uniformly tight if for every $\epsilon > 0$, there exists a constant M such that*

$$\sup_{\lambda \in \Lambda} P(\|X_\lambda\| > M) < \epsilon. \tag{1.82}$$

Corollary 1.32 *Let X_n be random vectors in $\mathbb{R}^k$. Then*

(i) *If $X_n \rightsquigarrow X$, then $\{X_n : n \in \mathbb{N}\}$ is uniformly tight;*
(ii) *If X_n is uniformly tight, then there exists a subsequence with $X_{n_j} \rightsquigarrow X$ as $j \to \infty$ for some X.*

1.9.2 Inequalities in Statistics

Suppose a random variable X has finite mean μ and finite variance σ^2. Then for any real $\epsilon > 0$,

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \tag{1.83}$$

It is easy to see that for i.i.d random variables X_i's, we have

$$\text{Var} \sum_{i=1}^{n} X_i \leq \sum_{i=1}^{n} EX_i^2. \tag{1.84}$$

However, if X_i's are not mutually independent, then an easy calculus leads to

$$\text{Var} \sum_{i=1}^{n} X_i \leq n \sum_{i=1}^{n} EX_i^2. \tag{1.85}$$

The inequality is not useful since it is not sharp enough. In stead of (1.85),

$$ES_n^2 = nEX^2 + 2\sum_{k=1}^{n-1}(n-k)EX_1X_k, \tag{1.86}$$

if the mean of X is 0. If the second term in the right hand side of (1.86) is absolutely summable, S_n can be evaluated by

$$\frac{1}{n}ES_n^2 = EX^2 + 2\sum_{k=1}^{n-1}(n-k)EX_1X_k, \tag{1.87}$$

that is, $1/nES_n^2$ is asymptotically $EX^2 + \sum_{k=1}^{n-1} EX_1X_k$.

1.10 U-statistics

Let $X_1, \ldots, X_n$ be independent observations on a distribution F. Consider a parameter function $\theta = \theta(F)$ for which there is an unbiased estimator. That is,

$$\theta(F) = E_F\{h(X_1, \ldots, X_m)\}. \tag{1.88}$$

for some function $h = h(x_1, \ldots, x_m)$, called a kernel. Without loss of generality, we can assume that h is symmetric in all its arguments. If not, the kernel can be replaced by the symmetric kernel

$$\frac{1}{m!}\sum_p h(x_{i_1}, \ldots, x_{i_m}), \tag{1.89}$$

where $\sum_p$ denotes summation over the $m!$ permutations $(i_1, \ldots, i_m)$ of $(1, \ldots, m)$.

The U-statistic for estimation of θ is defined as

$$U_n = U(X_1, \ldots, X_n) = \frac{1}{\binom{n}{m}}\sum_c h(X_{i_1}, \ldots, X_{i_m}), \tag{1.90}$$

where $\sum_c$ denotes summation over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \ldots, i_m\}$ from $\{1, \ldots, n\}$. Obviously, U_n is also an unbiased estimator of θ.

We list up common U-statistics here for reference. **[For $m = 1$]**

Example 1.33 (Mean)

$$h(x) = x. \tag{1.91}$$

Remark 1.34 *There are a lot of kernels to define "mean". It is not necessary to have "mean" in the case that $m = 1$. We just give a formal way to define U-statistics. Other examples of U-statistics below are also defined in the easiest way.*

Example 1.35 (Sample distribution function)

$$h(x) = I(x \le t_0). \tag{1.92}$$

Example 1.36 (Sample kth moment)

$$h(x) = x^k \tag{1.93}$$

Example 1.37

$$h(x) = \gamma(x). \tag{1.94}$$

[For $m = 2$]

Example 1.38

$$h(x_1, x_2) = x_1 x_2. \tag{1.95}$$

Example 1.39 (Variance)

$$h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2. \tag{1.96}$$

Example 1.40 (Gini's mean difference)

$$h(x_1, x_2) = |x_1 - x_2|. \tag{1.97}$$

Example 1.41 (Wilcoxon one-sample statistic)

$$h(x_1, x_2) = I(x_1 + x_2 \leq 0). \tag{1.98}$$

[For $m = 5$]

Example 1.42 (A measure of dependence for a bivariate distribution F) *Let*

$$\psi(z_1, z_2, z_3) = I(z_2 \leq z_1) - I(z_3 \leq z_1) \tag{1.99}$$

and then the kernel h is defined as

$$h((x_1, y_1), (x_2, y_2), \ldots, (x_5, y_5)) = \frac{1}{4}\psi(x_1, x_2, x_3)\psi(x_1, x_4, x_5)\psi(y_1, y_2, y_3)\psi(y_1, y_4, y_5). \tag{1.100}$$

For any kernel h, we define h_c as

$$h_c(x_1, \ldots, x_c) = E_F\{h(x_1, \ldots, x_c, X_{c+1}, \ldots, X_m)\}, \tag{1.101}$$

for $1 \leq c \leq m - 1$ such that

$$h_c(x_1, \ldots, x_c) = E_F\{h_{c+1}(x_1, \ldots, x_c, X_{c+1})\}. \tag{1.102}$$

Define $\zeta_0 = 0$ and, for $1 \leq c \leq m$,

$$\zeta_c = \mathrm{Var}_F\{h_c(X_1, \ldots, X_c)\}. \tag{1.103}$$

It is known that U-statistic has asymptotic normality as follows:

Theorem 1.43 *If $E_F h^2 < \infty$ and $\zeta_1 > 0$, then*

$$n^{1/2}(U_n - \theta) \rightsquigarrow \mathcal{N}(0, m^2\zeta_1). \tag{1.104}$$

Chapter 2
Models in Statistics

2.1 Models for i.i.d. Samples

We give some density functions of i.i.d. random variables in this section. The possible parameters of the distribution are given after the semicolon.

Example 2.1 (Exponential distribution) *If $X_1, \dots, X_n$ are distributed as exponential distribution, then the density function is given by*

$$f(x;\lambda) = \lambda e^{-\lambda x}. \tag{2.1}$$

The mean and the variance of the distribution are given by

$$E(X_1) = \frac{1}{\lambda}, \tag{2.2}$$
$$V(X_1) = \frac{1}{\lambda^2}. \tag{2.3}$$

Example 2.2 (Weibull distribution) *If $X_1, \dots, X_n$ are distributed as Weibull distribution, then the density function is given by*

$$f(x;k,\theta) = k\theta^{-k}x^{k-1}e^{-(x/\theta)^k}. \tag{2.4}$$

The mean and the variance of the distribution are given by

$$E(X_1) = \theta\Gamma(1+k^{-1}), \tag{2.5}$$
$$V(X_1) = \theta^2(\Gamma(1+2k^{-1}) - \Gamma^2(1+k^{-1})). \tag{2.6}$$

Example 2.3 (Gamma distribution) *If $X_1, \dots, X_n$ are distributed as Gamma distribution, then the density function is given by*

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}. \tag{2.7}$$

The mean and the variance of the distribution are given by

$$E(X_1) = \alpha\beta \tag{2.8}$$
$$V(X_1) = \alpha\beta^2. \tag{2.9}$$

In warranty analysis, the renewal process is usually modeled by exponential distribution for a given failure rate λ. To distinguish the increasing failure rate and the decreasing failure rate, the Weibull distribution is considered in the field. The Weibull distribution with $0 < k < 1$ is used for the decreasing failure rate and with $k > 1$ for the increasing failure rate. We call k the index of the failure rate. For a given k, if the model with Weibull distribution has the same mean with exponential distribution, then θ is determined by

$$\theta = \frac{1}{\lambda \Gamma(1 + k^{-1})}. \tag{2.10}$$

Furthermore, the idea can be generated to Gamma distribution with the identical mean, variance and the index of the failure rate. For a given k,

$$\alpha = \frac{\Gamma^2(1 + k^{-1})}{\Gamma(1 + 2k^{-1}) - \Gamma^2(1 + k^{-1})}, \tag{2.11}$$
$$\beta = \frac{\Gamma(1 + 2k^{-1}) - \Gamma^2(1 + k^{-1})}{\lambda \Gamma^2(1 + k^{-1})}. \tag{2.12}$$

[distributions]

2.2 Models of Time Series

Consider a stationary time series X_t with auto-covariance function $\gamma_X(j) := EX_0X_j - EX_0^2$ at lag j. Define

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} \gamma_X(j) e^{ij\lambda}, \quad \lambda \in (-\pi, \pi]. \tag{2.13}$$

From Theorem 1.8, it is shown that $f_X(\lambda)$ is continuous and further symmetric about 0 if

$$\sum_{j \in \mathbb{Z}} |\gamma_X(j)| < \infty. \tag{2.14}$$

However, the problem with the usual spectral density can be found is the following examples.

Example 2.4 (Linton and Whang (2007)) *Suppose*

$$X_t = \xi_0(\tau_0) + \epsilon_t v(\epsilon_{t-1}, \dots), \tag{2.15}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. random variables with $\tau_0 = P(\epsilon_t < 0)$ and v a measurable function. The spectral density of $\{X_t\}$ is $f_X(\lambda) = \gamma_X(0)/(2\pi)$ contains no information about the process since X_t is an unicorrelated time series.

2.2.1 Nonlinear time series model

Suppose $\{X_t\}_{t\in\mathbb{Z}}$ is a nonlinear process defined by

$$X_t = Y(\epsilon_t, \epsilon_{t-1}, \dots), \tag{2.16}$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. copies of a random variable ϵ and Y is a measurable function.

Let $\{\epsilon'_i\}$ be an i.i.d. copy of $\{\epsilon_i\}$ and $\xi'_i = (\dots, \epsilon'_{i-1}, \epsilon'_i)$ the shift process of $\xi_i = (\dots, \epsilon_{i-1}, \epsilon_i)$. For $I \subset \mathbb{Z}$, define

$$\epsilon_{j,I} = \begin{cases} \epsilon'_j & \text{if } j \in I, \\ \epsilon_j & \text{if } j \notin I. \end{cases} \tag{2.17}$$

Definition 2.5 (Functional or physical dependence measure) *For $p > 0$ and $I \subset \mathbb{Z}$, let $\delta_p(I, n) = \|g(\xi_n) - g(\xi_{n,I})\|_p$ and $\delta_p(n) = \|g(\xi_n) - g(\xi_n^*)\|_p$.*

Definition 2.6 (Predictive dependence measure) *Let $p \geq 1$ and g_n be a Borel function on $\mathbb{R}^\infty \to \mathbb{R}$ such that $g_n(\xi_0) = E(X_n|\xi_0)$, $n \geq 0$. Let $\omega_p(I, n) = \|g_n(\xi_0) - g_n(\xi_{0,I})\|_p$ and $\omega_p(n) = \|g_n(\xi_0) - g_n(\xi_0^*)\|_p$.*

Remark 2.7 *The interpretation of $g_n(\cdot)$ can be seen by*

$$g_n(\xi_0) = E(X_n|\xi_0) = E(g(\xi_n)|\xi_0). \tag{2.18}$$

Definition 2.8 (p-stability) *Let $p \geq 1$. The process $\{X_n\}$ is said to be p-stable if $\Omega_p := \sum_{n=0}^\infty \omega_p(n) < \infty$, and p-strong stable if $\Delta_p := \sum_{n=0}^\infty \delta_p(n) < \infty$.*

2.2.2 Quantiles

The variable of interest for the quantile analysis is defined by

$$V_t(\tau, \xi) = \tau - 1\{X_t < \xi\}, \quad (\tau, \xi) \in (0, 1) \times \mathbb{R}. \tag{2.19}$$

Let $\xi_0(\tau)$ be defined as in (1.77). For simplicity, we define

$$V_t(\tau) = V_t(\tau, \xi_0(\tau)). \tag{2.20}$$

Note that

$$V_t(\tau) = \begin{cases} \tau - 1 & \text{if } X_t < \xi_0(\tau), \\ \tau & \text{if } X_t \geq \xi_0(\tau). \end{cases} \tag{2.21}$$

Interestingly, if the quantile τ is chosen by researcher, then $V_t(\tau)$ is obviously a random variable corresponding to X_t. If $\{X_t\}$ is weakly stationary, then $\{V_t\}$ is also weakly stationary. If we further suppose the distribution function of X_t is continuous at $\xi_0(\tau)$, then it is easy to see that

$$E\,V_t(\tau) = 0, \tag{2.22}$$
$$\operatorname{Var} V_t(\tau) = (1-\tau)\tau. \tag{2.23}$$

Since $\{V_t\}$ is a zero mean weakly stationary process, we define its spectral density as

$$f_V(\omega) = \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \gamma_V(j) e^{ij\omega}, \tag{2.24}$$

where

$$\gamma_V(j) = EV_0(\tau)V_j(\tau) = \begin{cases} (\tau-1)^2 & \text{if } X_0 < \xi_0(\tau) \text{ and } X_j < \xi_0(\tau), \\ \tau(\tau-1) & \text{if } X_0 < \xi_0(\tau) \text{ and } X_j \geq \xi_0(\tau), \text{ or } X_0 \geq \xi_0(\tau) \text{ and } X_j < \xi_0(\tau), \\ \tau^2 & \text{if } X_0 \geq \xi_0(\tau) \text{ and } X_j \geq \xi_0(\tau). \end{cases}$$
(2.25)

When we turn our attention from the usual periodogram to the quantile periodogram, we have to first estimate $\xi_0(\tau)$. The estimate $\hat{\xi}_n(\tau)$ can be achieved by the following check function

$$\hat{\xi}_n(\tau) = \min_{x\in\mathbb{R}} \sum_{t=1}^{n} \rho_\tau(X_t - x), \tag{2.26}$$

which is proposed in Koenker and Bassett (1978). Let the corresponding periodogram be defined by

$$I_{n,\tau}(\lambda) = \frac{1}{2\pi}\left| n^{-1/2} \sum_{t=1}^{n} \hat{V}_t(\tau) e^{-it\lambda} \right|^2, \tag{2.27}$$

where $\hat{V}_t(\tau) = V_t(\tau, \hat{\xi}_n(\tau))$.

For estimation of $\xi_0(\tau)$, we need the following assumptions given in Hagemann (2013). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be an i.i.d. copies of $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and suppose

$$X_t' = Y(\epsilon_t, \ldots, \epsilon_1, \epsilon_0^*, \epsilon_{-1}^*, \ldots). \tag{2.28}$$

Assumption 2.1 *For a given $\tau \in (0,1)$, there exists $\delta > 0$ and $\sigma \in (0,1)$ such that*

$$\sup_{\xi\in\mathscr{X}_r(\delta)} \|1\{X_n < \xi - 1\{X_n' < \xi\}\}\| = O(\sigma^n), \tag{2.29}$$

where $\mathscr{X}_r(\delta) = \{\xi \in \mathbb{R}; |\xi_0(\tau) - \xi| \leq \delta\}$.

Assumption 2.2 *The distribution function F_X of X_0 is Lipschitz continuous in a neighborhood of $\xi_0(\tau)$ and has a positive and continuous density at $\xi_0(\tau)$.*

Theorem 2.9 *Let $\lambda_n = 2\pi j_n/n$ with $j_n \in \mathbb{Z}$ be a sequence of natural frequencies such that $\lambda_n \to \lambda \in (0,\pi)$ with $f_V(\lambda) > 0$. Under Assumptions 2.1 and 2.2, for any fixed $k \in \mathbb{Z}$, the collection of quantile periodograms*

$$I_{n,\tau}(\lambda_n - 2\pi k/n), I_{n,\tau}(\lambda_n - 2\pi(k-1)/n), \ldots, I_{n,\tau}(\lambda_n + 2\pi k/n), \tag{2.30}$$

converges jointly in distribution to independent exponential variables with mean $f_V(\lambda)$.

Define $X_t^* = Y(\epsilon_t, \ldots, \epsilon_1, \epsilon_0^*, \epsilon_{-1}, \ldots)$.

Assumption 2.3 *For a given $\tau \in (0,1)$ and $\mathscr{X}_\tau(\delta)$ as in Assumption 2.1, there exists a $\delta > 0$ such that*

$$\sum_{t=0}^{\infty} \sup_{\xi \in \mathscr{X}_\tau(\delta)} \|1\{X_t < \xi\} - 1\{X_t^* < \xi\}\| < \infty. \tag{2.31}$$

Define

$$\mathscr{W} = \{\ w \text{ is bounded and continuous, } w(x) = w(-x) \text{ for all } x \in \mathbb{R},\ w(0) = 1,$$

$$\bar{w}(x) := \sup_{y \geq x} |w(y)| \text{ satisfies } \int_0^\infty \bar{w}(x)\, dx < \infty,\ W(\lambda) := (2\pi) \int_{-\infty}^{-\infty} w(x) e^{-ix\lambda}\, dx \text{ satisfies } \int_{-\infty}^{-\infty} |W(\lambda)|\, d\lambda < \infty.\}$$

Theorem 2.10 *Under Assumptions 2.2, 2.3, if $w \in \mathscr{W}$, $B_n \to \infty$ and $B_n = o(\sqrt{n})$, then*

$$\hat{f}_V(\lambda) \xrightarrow{\mathcal{P}} f_V(\lambda) \tag{2.32}$$

uniformly in $\lambda \in (-\pi, \pi]$.

Assumption 2.4 *There is some n^* such that for all $n > n^*$, $F_{\tilde{X}}(x) := P(\tilde{X}_0 \leq x)$ is Lipschitz continuous in a neighborhood of $\xi_0(\tau)$ and $E|X_n - X_n'| = O(\rho^n)$ for some $\rho \in (0,1)$.*

Theorem 2.11 *Under Assumptions 2.2, 2.4, if w is even and Lipschitz continuous with with support $[-1,1]$, $w(0) = 1$, $\lim_{x \to 0}(1 - w(x))/|x|^3 < \infty$, $B_n \to \infty$, $B_n = o(n^{1/4})$, $n = o(B_n^7)$, then*

$$|m_n/B_n|(\hat{f}_V(\lambda) - f_V(\lambda)) \rightsquigarrow \mathcal{N}(0, \sigma^2(\lambda)), \tag{2.33}$$

where $\sigma^2(\lambda) = (1 + h(2\lambda)) f_V(\lambda) \int_{-1}^1 w(x)^2\, dx$, and $h(\lambda) = 1$ if $\lambda = 2\pi k$ for some $k \in \mathbb{Z}$ and 0 otherwise.

2.3 Stationary, Absolutely Regular Processes

2.3.1 Regularity

Let $\mathcal{M}_a^b$ be the σ-algebra generated by $x(n)$, $a \leq n \leq b$, $\mathcal{M}_b$ if $a = -\infty$ and $\mathcal{M}_{-\infty}$ for the intersection of all $\mathcal{M}_b$.

Definition 2.12 *The process $x(n)$ is said to be regular if $\mathcal{M}_{-\infty}$ is trivial.*

Suppose T is the automorphism of $\mathcal{M}_\infty$ induced by $x(n) \to x(n+1)$. Regularity implies weak mixing, namely

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |P(A \cap T^n B) - P(A)P(B)| \downarrow 0, \quad \text{for all } A, B \in \mathcal{M}_{-\infty}. \tag{2.34}$$

2.3.2 Stationarity, ergodicity and mixing conditions

Let $\{\xi_i, -\infty < i < \infty\}$ be a p-dimensional sequence of stochastic vectors defined on a probability space $(\Omega, \mathcal{A}, P)$. For $a \leq b$, let $\mathcal{M}_a^b$ denote the σ-algebra of events generated by $\xi_a, \ldots, \xi_b$.

Definition 2.13 *The process satisfies the ϕ-mixing condition if*

$$\phi(n) = \sup_{B \in \mathcal{M}_{-\infty}^0, A \in \mathcal{M}_n^\infty} \frac{1}{P(B)} |P(A \cap B) - P(A)P(B)| \downarrow 0 \tag{2.35}$$

The process is called uniformly mixing.

Definition 2.14 *The process is called absolutely regular, if*

$$\beta(n) = \sup_{a \in \mathbb{Z}} E\{ \sup_{A \in \mathcal{M}_{a+n}^\infty} |P\{A|\mathcal{M}_{-\infty}^a\} - P(A)|\} \downarrow 0. \tag{2.36}$$

Definition 2.15 *The process satisfies Rosenblatt's strong mixing condition if*

$$\alpha(n) = \sup_{B \in \mathcal{M}_{-\infty}^0, A \in \mathcal{M}_n^\infty} |P(A \cap B) - P(A)P(B)| \downarrow 0. \tag{2.37}$$

Simply, we call the process strong mixing.

Example 2.16 (Kolmogorov and Rozanov (1960)) *A Gaussian stationary process is strongly mixing if it has a continuous positive spectral density function.*

Example 2.17 (Andrews (1984)) *The autoregressive process is not strong mixing when the innovations are i.i.d. Bernoulli random variables.*

Definition 2.18 *The process is strictly stationary and absolutely regular, if*

$$\beta(n) = E\{\sup_{A\in\mathcal{M}_n^\infty} |P\{A|\mathcal{M}_{-\infty}^0\} - P(A)|\} \downarrow 0 \tag{2.38}$$

as $n \to \infty$.

Definition 2.19 *The process is said to be uniformly mixing if*

$$\phi(n) = \sup_{a\in\mathbb{Z}, A\in\mathcal{M}_1^a, B\in\mathcal{M}_{a+n}^\infty} \max\{|P(A|B) - P(A)|, |P(B|A) - P(B)|\} \downarrow 0. \tag{2.39}$$

Remark 2.20 *Processes which are absolutely regular maintain the property under time reveral, but this is not the case for uniformly mixing processes.*

Remark 2.21 (Relationships between the conditions for the process)
$*$*-mixing* $\Rightarrow$ *uniformly mixing in both directions of time* $\Rightarrow$ *uniformly mixing* $\Rightarrow$ *absolutely regular* $\Rightarrow$ *Rosenblatt's strong mixing*

Example 2.22 *Let* $\{\xi_t\}$ *be a m-dependent process. Then the process is uniformly mixing.*

Example 2.23 *Let* $\xi_t = a\xi_{t-1} + \epsilon_t$, *where* ϵ_t *is i.i.d.* $\mathcal{N}(0,1)$ *and* $|a| < 1$. *Then the process* $\{\xi_t\}$ *is strong mixing but not uniformly mixing.*

Proposition 2.24 *Let* $\{\xi_t\}$ *be a strictly stationary process. If the process is strong mixing, then it is ergodic.*

Proof. See Rosenblatt (1978).

Remark 2.25 *Mixing conditions is more general since it is defined for the processes that are not necessarily strictly stationary.*

Remark 2.26 *See Billingsley [pp. 182-186].*

Remark 2.27 *Measurable functions of mixing processes are mixing and of the same size. Note that whereas functions of ergodic processes retain ergodicity for any* τ, *finite or infinite, mixing is guaranteed only for finite* τ.

Suppose that $\{\xi_i\}$ is a p-dimensional strictly stationary, absolutely regular process with distribution function $F(x)$.

Let $i_1 < i_2 < \cdots < i_k$ be arbitrary integers. For any $j(1 \le j \le k-1)$, put

$$P_j^{(k)}(E^{(j)} \times E^{(k-j)}) = P(\xi_{i_1}, \ldots, \xi_{i_j}) \in E^{(j)}) P((\xi_{i_{j+1}}, \ldots, \xi_{i_k}) \in E^{(k-j)}) \tag{2.40}$$

and

$$P_0^{(k)}(E^{(k)}) = P((\xi_{i_1}, \ldots, \xi_{i_k}) \in E^{(k)}) \tag{2.41}$$

where $E^{(i)}$ is a Borel set in $\mathbb{R}^{ip}$.

Example 2.28 (Difference between $P_0^{(k)}$ and $P_j^{(k)}$) *Suppose* $X_1 = X_2$ *a.s. with each marginal distribution defined by*

$$X_1 = \begin{cases} -1, & \text{w.p. } \frac{1}{2}, \\ 1, & \text{w.p. } \frac{1}{2}, \end{cases} \quad X_2 = \begin{cases} -1, & \text{w.p. } \frac{1}{2}, \\ 1, & \text{w.p. } \frac{1}{2}. \end{cases} \tag{2.42}$$

Then

$$P(X_1 = 1, X_2 = 1) = \frac{1}{2}, \tag{2.43}$$

but

$$P(X_1 = 1)P(X_2 = 1) = \frac{1}{4}. \tag{2.44}$$

On the other hand,

$$P(X_1 = 1, X_2 = -1) = 0, \tag{2.45}$$

but

$$P(X_1 = 1)P(X_2 = -1) = \frac{1}{4}. \tag{2.46}$$

2.3.3 Basic Lemmas

Lemma 2.1 *For any $j (0 \leq j \leq k-1)$, let $h(x_1, \ldots, x_k)$ be a Borel function such that*

$$\int_{\mathbb{R}^{kp}} |h(x_1, \ldots, x_k)|^{1+\delta} dP_j^{(k)} \leq M \tag{2.47}$$

for some $\delta > 0$. Then

$$|\int_{\mathbb{R}^{kp}} h(x_1, \ldots, x_k) dP_0^{(k)} - \int_{\mathbb{R}^{kp}} h(x_1, \ldots, x_k) dP_j^{(k)}| \leq 4M^{1/(1+\delta)} \beta^{\delta/(1+\delta)} (i_{j+1} - i_j). \tag{2.48}$$

Proof. The proof of the Lemma depends on the definition of absolute regularity and the representation of Rozanov and Volkonskii (1961).

Let

$$n^{-[r]} = \{n(n-1) \cdots (n-r+1)\}^{-1}. \tag{2.49}$$

For every $c (0 \leq c \leq m)$, let

$$g_c(x_1, \ldots, x_c) = \int_{\mathbb{R}^{(m-c)p}} g(x_1, \ldots, x_m) dF(x_{c+1}) \ldots dF(x_m), \tag{2.50}$$

and

$$U_n^{(c)} = n^{-[c]} \sum_{1 \leq i_1 < \cdots < i_c \leq n} \int_{\mathbb{R}^{cp}} g_c(x_1, \ldots, x_c) \prod_{j=1}^{c} d[u(x - \xi_{i_j}) - F(x_j))] \tag{2.51}$$

where $u(v)$ is equal to one when all the p components of ν are non-negative; otherwise, $u(v) = 0$.Then

$$U_n = \theta(F) + \sum_{c=1}^{m} \binom{m}{c} U_n^{(c)}. \tag{2.52}$$

Lemma 2.2 *If there is a positive number δ such that for $r = 2 + \delta$,*

$$\mu_r = \int_{\mathbb{R}^{pm}} |g_1(x_1, \ldots, x_m)|^r dF(x_1), \ldots, dF(x_m) \leq M_0 < \infty \tag{2.53}$$

and for all integers $i_1, i_2, \ldots, i_m (i_1 < i_2 < \cdots < i_m)$,

$$\nu_r = E|g(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_m})|^r \leq M_0 < \infty \tag{2.54}$$

hold. Further, for some $\delta'(0 < \delta' < \delta)$ *and* $\beta(n) = O(n^{-(2+\delta')/\delta'})$, *then we have*

$$E(U_n^{(c)})^2 = O(n^{-1-\gamma}), \quad 2 \leq c \leq m, \tag{2.55}$$

where $\gamma = 2(\delta - \delta')/(\delta'(2+\delta)) > 0$.

Lemma 2.3 *If there is a positive number* δ *such that for* $r = 4 + \delta$,

$$\mu_r = \int_{\mathbb{R}^{pm}} |g_1(x_1, \ldots, x_m)|^r dF(x_1), \ldots, dF(x_m) \leq M_0 < \infty \tag{2.56}$$

and for all integers $i_1, i_2, \ldots, i_m (i_1 < i_2 < \cdots < i_m)$,

$$\nu_r = E|g(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_m})|^r \leq M_0 < \infty \tag{2.57}$$

hold. For some $\delta'(0 < \delta' < \delta)$ *and* $\beta(n) = O(n^{-3(4+\delta'/(2+\delta'))})$, *then we have*

$$E(U_n^{(2)})^4 = O(n^{-3-\gamma'}) \tag{2.58}$$

where $\gamma' = 6(\delta - \delta')/\{(4+\delta)(2+\delta)'\} > 0$ *and*

$$E(U_n^{(c)})^2 = O(n^{-3}), \quad 3 \leq c \leq m. \tag{2.59}$$

Lemma 2.4 *If the conditions of Lemma 2.2 are satisfied, then*

$$E(V_n^{(c)})^2 = O(n^{-1-\gamma}) \quad (1 \leq c \leq m). \tag{2.60}$$

Lemma 2.5 *If the conditions of Lemma 2.3 are satisfied, then*

$$E(V_n^{(2)})^4 = O(n^{-3-\gamma'}) \tag{2.61}$$

and

$$E(V_n^{(c)})^2 = O(n^{-3}) \quad 3 \leq c \leq m. \tag{2.62}$$

2.4 Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes

Consider a functional

$$\theta(F) = \int_{\mathbb{R}^{mp}} g(x_1, \ldots, x_m) dF(x_1) \cdots dF(x_m) \tag{2.63}$$

defined over $\mathcal{F} = \{F : |\theta(F)| < \infty$, where $g(x_1, \ldots, x_m)$ is symmetric in its m arguments. As an estimator of $\theta(F)$, we define a U-statistic

$$U_N = \binom{N}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} g(\xi_{i_1}, \ldots, \xi_{i_m}), \quad N \ge m. \tag{2.64}$$

Also, we consider a von Mises' differentiable statistical functional $\theta(F_N)$ defined by

$$\theta(F_N) = N^{-m} \sum_{i_1=1}^{N} \cdots \sum_{i_m=1}^{N} g(\xi_{i_1}, \ldots, \xi_{i_m}), \quad N \ge 1. \tag{2.65}$$

We denote by

$$\sigma_N^2 = E(\sum_{r=1}^{N} \hat{h}_1(X_t))^2 \tag{2.66}$$

the exact variance, and denote by

$$\sigma^2 = E(\hat{h}_1(X_1))^2 + 2 \sum_{t>1} Eh\hat{h}_1(X_1)\hat{h}_1(X_t) \tag{2.67}$$

its asymptotic variance if the sum converges absolutely. Here, $\sigma^2 = \lim N^{-1}\sigma_N^2$.

Theorem 2.29 *Let $g : \mathscr{X}^m \to \mathbb{R}$ be a non-degenerate kernel. Then the asymptotic distribution of $\frac{N}{m\sigma_N}(U_N(g) - \theta(F))$ is $\mathcal{N}(0,1)$ provided one of the following conditions is satisfied:*

(a) *$\{X_n\}_{n\ge1}$ is uniformly mixing in both directions of time, $\sigma_N^2 \to \infty$ and for some $\delta > 0$,*

$$\sup_{1\le t_1 < \cdots < t_m} E|g(X_1, \ldots, X_m)|^{2+\delta} < \infty. \tag{2.68}$$

(b) *$\{X_n\}_{n\ge1}$ is uniformly mixing in both directions of time with mixing coefficients $\phi(n)$ satisfying $\sum \phi(n) < \infty$, $\sigma^2 \ne 0$ and*

$$\sup_{1\le t_1 < \cdots < t_m} E(g(X_1, \ldots, X_{t_m}))^2 < \infty. \tag{2.69}$$

(c) *$\{X_n\}_{n\ge1}$ is absoulutely regular with coefficients $\beta(n)$ satisfying $\sum \beta(n)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$, $\sigma^2 \ne 0$ and*

$$\sup_{1\le t_1 < \cdots < t_m} E|g(X_{t_1}, \ldots, X_{t_m})|^{2+\delta} < \infty. \tag{2.70}$$

The same statement holds for v. Mises' functionals when the supremum in (a)-(c) is replaced by the supremum over all choices of $1 \le t_i (1 \le i \le m)$.

Theorem 2.30 *If $g : \mathscr{X}^m \to \mathbb{R}$ is a non-degenerate kernel, then*

$$\gamma(N) = \sup_{x\in\mathbb{R}} |P(\frac{N}{m\sigma_N})(U_N - \theta(F) \le x - \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\frac{t^2}{2})\, dt| \to 0 \tag{2.71}$$

under each of the following conditions:

(a) *$\{X_n\}_{n\ge1}$ is uniformly mixing in both directions of time, $\sigma_N^2 \to \infty$ and*

$$\sup_{1\leq t_1<\cdots<t_m} E|g(X_1,\ldots,X_{t_m})|^3 < \infty. \tag{2.72}$$

In this case

$$\gamma(N) = O((\lambda_N \log \lambda_N)^{1/3}) \tag{2.73}$$

where $\lambda_N = \max\{2\phi^{1/6}([N^\beta]), N^{-\alpha}\}$ *and where* $0 < \alpha < \frac{1}{5}$, $0 < \beta < 1 - 5\alpha$ *denote constants.*

(b) $\{X_n\}_{n\geq 1}$ *is uniformly mixing in both directions of time with coefficients* $\phi(n)$ *satisfying* $\phi(n) = O(q^n)$ *for some* $0 < q < 1$, $\sigma^2 \neq 0$ *and*

$$\sup_{1\leq t_1<\cdots<t_m} E|g(X_1,\ldots,X_{t_m})|^3 < \infty. \tag{2.74}$$

Here

$$\gamma(N) = O(N^{-1/3+\lambda}) \quad \textit{for each } \lambda > 0. \tag{2.75}$$

(c) $\{X_n\}$ *is absolutely regular with coefficients* $\beta(n)$ *satisfying* $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\epsilon})$ *for some* $0 < \delta \leq 1$, $0 \leq \epsilon < 1$, $\sigma^2 \neq 0$, *and*

$$\sup_{1\leq t_1<\cdots<t_m} E|g(X_1,\ldots,X_{t_m})|^{2+\delta} < \infty. \tag{2.76}$$

In this case $\gamma(N) = O(N^{-\lambda})$ *where* $\lambda = (1-\epsilon)\delta/144$.

Next, let C be the space of all continuous real-valued functions on $[0,1]$, where we give C the uniform topology. For every $n \geq m$, let $X_n = \{X_n(t),\, 0 \leq t \leq 1\}$ be a random element in C defined by

$$X_n(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq (m-1)n, \\ k(U_k - \theta(F))/(m\sigma n^{1/2}) & \text{for } t = k/n,\ m \leq k \leq n, \\ \text{linearly interpolated} & \text{for } t \in [k/n,(k+1)/n], \quad m-1 \leq k \leq n-1. \end{cases} \tag{2.77}$$

Similarly, let $X_n^* = \{X_n^*(t), 0 \leq t \leq 1\}$ be a random element in C defined by

$$X_n^*(t) = \begin{cases} 0 & \text{for } 0 \leq t = 0, \\ k(U_k - \theta(F))/(m\sigma n^{1/2}) & \text{for } t = k/n,\ 1 \leq k \leq n, \\ \text{linearly interpolated} & \text{for } t \in [k/n,(k+1)/n], \quad 0 \leq k \leq n-1. \end{cases} \tag{2.78}$$

Let $W = \{W(t), 0 \leq t \leq 1\}$ be a standard Brownian motion.

Theorem 2.31 *If there is a positive number* δ *such that for* $r = 4 + \delta$,

$$\mu_r = \int_{\mathbb{R}^{pm}} |g_1(x_1,\ldots,x_m)|^r dF(x_1),\ldots,dF(x_m) \leq M_0 < \infty \tag{2.79}$$

and for all integers $i_1, i_2, \ldots, i_m (i_1 < i_2 < \cdots < i_m)$,

$$\nu_r = E|g(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_m})|^r \leq M_0 < \infty \tag{2.80}$$

hold. Also for some $\delta'(0 < \delta' < \delta)$

$$\beta(n) = O(n^{-3(4+\delta')/(2+\delta')}) \tag{2.81}$$

then, both X_n and X_n^* converge weakly to W and $\rho(X_n, X_n^*) \to 0$ as $n \to \infty$.

Let $C_0(\subset C)$ be the space of continuous functions on $[0,1]$ vanishing at 0, with the uniform topology and for each $\omega \in \Omega$, define the functions $Y_n(t,\omega)$ and $Y_n^*(t,\omega)$ in C_0 as follows:

$$Y_n(t,\omega) = \frac{X_n(t,\omega)}{(2\log\log n\sigma^2)^{frac12}}, \quad n \geq \max(m, 3/\sigma^2) \tag{2.82}$$

and

$$Y_n(t,\omega) = \frac{X_n^*(t,\omega)}{(2\log\log n\sigma^2)^{frac12}}, \quad n \geq 3/\sigma^2. \tag{2.83}$$

Furthermore, we denote by K the subset of C_0 consisting of all functions $h(t)$ absolutely continuous with respect to Lebesgue measure such that

$$\int_0^1 \dot{h}^2(t)\, dt \leq 1, \tag{2.84}$$

where $\dot{h}(t)$ stands for the Radon-Nikodym derivative of h.

Theorem 2.32 *If the conditions in Theorem 2.31 are satisfied, then for almost all $\omega \in \Omega$, the sequence of functions $\{Y_n(t,\omega), n \geq \max m, 3/\sigma^2\}$ and $\{Y_n^*(t,\omega), n \geq 3/\sigma^2\}$ are preconpact in C_0 and their derived sets coincides with the set K*

Exercise 2.1 *Show* (2.22) *and* (2.23).

2.4.1 GARCH process

A GARCH(1,1) process is given by the equations

$$X_t = \sigma_t Z_t, \quad t \in^Z, \tag{2.85}$$

where (Z_t) is an i.i.d. sequence with $EZ = 0$ and $\mathrm{Var}(Z) = 1$, and

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + \sigma_{t-1}^2 C_{t-1}, \quad C_t = \alpha_1 Z_t^2 + \beta_1. \tag{2.86}$$

Chapter 3
Convergence in Probability

3.1 Convergence in Probability

Lemma 3.1 (Convexity Lemma) *Let $\{\lambda_n(\theta) : \theta \in \Theta\}$ be a sequence of random convex functions defined on a convex, open subset Θ of $\mathbb{R}^d$. Suppose $\lambda(\cdot)$ is a real-valued function on Θ for which $\lambda_n(\theta) \xrightarrow{\mathcal{P}} \lambda(\theta)$ for each $\theta \in \Theta$. Then for each compact subset K of Θ,*

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{\mathcal{P}} 0. \tag{3.1}$$

The function $\lambda(\cdot)$ is necessarily convex on Θ.

3.2 Almost Sure Convergence

Let $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$.

Lemma 3.2 (Kolmogorov's strong law of large numbers) *Suppose X_i's are independent and*

$$\sum_{i=0}^{\infty} E(X_i - EX_i)^2 / i^2 < \infty, \tag{3.2}$$

then

$$\frac{1}{n}(S_n - ES_n) \to 0 \quad a.s. \tag{3.3}$$

If we suppose X_i's are independent and identically distributed, then we have the following result.

Corollary 3.1 *Suppose $\{X_i\}$ are independent and identically distributed. If $E|X_i| < \infty$, then*

$$\frac{1}{n}(S_n - nEX_1) \to 0 \quad a.s. \tag{3.4}$$

If, in addition, $E|X_1|^p < \infty$ for some $1 < p < 2$, then

$$\frac{1}{n^{1/p}}(S_n - nEX_1) \to 0 \quad a.s. \tag{3.5}$$

3.3 Weak Convergence

Let $\{X_i, i = 1, \dots\}$ be a stationary Gaussian sequence with $EX_i = 0$ and $EX_i^2 = 1$. Let $G(X_i)$ have mean 0 and finite variance. Consider

$$Z_N(t) = \frac{1}{d_N} \sum_{i=1}^{[Nt]} G(X_i), \tag{3.6}$$

where $0 \le t \le 1$ and d_N^2 is asymptotically proportional to Var $\sum_{i=1}^{N} G(X_i)$. The weak convergence is understood to hold in $D[0,1]$.

For a probability measure P on $(D, \mathcal{D})$, let T_P consist of those t in $[0,1]$ for which the projection π_t is continuous except at points forming a set of P-measure 0. The points 0 and 1 always lie in T_P. If $0 < t < 1$, then $t \in T_P$ if and only if $P(J_t) = 0$, where

$$J_t = \{x; x(t) \ne x(t-)\}. \tag{3.7}$$

Define

$$w''_x(\delta) = \sup_{\substack{t_1 \le t \le t_2 \\ t_2 - t_1 \le \delta}} \min\{|x(t) - x(t_1)|, |x(t_2) - x(t)|\}. \tag{3.8}$$

Theorem 3.2 (Billingsley (1968)'s Theorem 15.4) *Suppose that*

$$P_n \pi^{-1}_{t_1 \cdots t_k} \Rightarrow P \pi^{-1}_{t_1 \cdots t_k} \tag{3.9}$$

holds whenever $t_1, \dots, t_k$ all lie in T_P. Suppose further that $P(J_1) = 0$. Suppose finally that, for each positive ϵ and η, there exist a δ, $0 < \delta < 1$, and an integer n_0 such that

$$P_n\{x : w''_x(\delta) \ge \epsilon\} \le \eta, \quad n \ge n_0. \tag{3.10}$$

Then $P_n \Rightarrow P$.

Theorem 3.3 (Billingsley (1968)'s Theorem 15.6) *Suppose that*

$$(X_n(t_1), \dots, X_n(t_k)) \rightsquigarrow (X(t_1), \dots, X(t_k)) \tag{3.11}$$

holds whenever $t_1, \dots, t_k$ all lie in T_P; that $P(J_1) = 0$; and that

$$P\{|X_n(t) - X_n(t_1)| \ge \lambda, |X_n(t_2) - X_n(t)| \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}}(F(t_2) - F(t_1))^{2\alpha} \tag{3.12}$$

for $t_1 \le t \le t_2$ and $n \ge 1$, where $\gamma \ge 0$, $\alpha > \frac{1}{2}$, and F is nondecreasing, continuous function on $[0,1]$. Then $X_n \rightsquigarrow X$ in $D[0,1]$.

3.3.1 *The Hermite Rank m*

Let X denote a stand normal random variable and define

$$\mathcal{G} = \{G; EG(X) = 0, EG^2(X) < \infty\}. \tag{3.13}$$

$\mathcal{G}$ is then a subset of

$$\boldsymbol{L}^2(\mathbb{R}, \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}) = \{G; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\infty} G^2(x) \exp(-\frac{x^2}{2}) dx < \infty\}. \tag{3.14}$$

Note that the Hermite polynomials form a complete orthogonal system of functions in $\boldsymbol{L}^2(\mathbb{R}, \frac{e^{-x^2/2}}{\sqrt{2\pi}})$.

Introduce the projection $J(q)$ as

$$J(q) = EG(X)H_q(X), \tag{3.15}$$

and define the Hermite rank of G as

$$m = \min_{q \in \mathbb{N}}(q; J(q) \neq 0). \tag{3.16}$$

For example, odd powers of X have Hermite rank 1. Even powers of X with their mean subtracted have Hermite rank 2. The Hermite polynomial H_m has Hermite rank m.

3.3.2 *Fractinoal Brownian Motion and The Rosenblatt Process*

Let $\{X_k\}$ be a normalized stationary Gaussian sequence, and let $r(k) \equiv EX_iX_{i+k}$, $k = 1, 2, \ldots$, be its correlation kernel. Suppose $0 < H < 1$.

Definition 3.4 (The class $\mathcal{G}_m$)

$$\mathcal{G}_m = \{G; G \in \mathcal{G}, G \text{ has Hermite rank } m\}. \tag{3.17}$$

Note that

$$\mathcal{G} = \mathcal{G}_\infty \cup \mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots, \tag{3.18}$$

with $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$, if $i \neq j$, and where $\mathcal{G}_\infty \equiv \{G(x) \equiv 0\}$.

Definition 3.5 (The class $(m)(D, L(\cdot))$) *For any positive integer m, $\{X_i\} \in (m)(D, L(\cdot))$ if $r(k) \sim k^{-D}L(k)$ as $k \to \infty$ with $0 < D < \frac{1}{m}$ and L slowly varying.*

Note that

$$(m_2)(D, L(\cdot)) \subset (m_1)(D, L(\cdot)), \quad m_2 > m_1. \tag{3.19}$$

Definition 3.6 (The class $(m)'(H, L(\cdot))$) *For any positive integer m, $\{X_i\} \in (m)'(H, L(\cdot))$ if*

(i) $\lim_{k \to 0} r(k) = 0$,
(ii) $\sum_{i=1}^N \sum_{j=1}^N (r(i-j))^m \sim N^{2H} L(N) \quad$ *as $N \to \infty$,*

(iii) $\sum_{i=1}^{N}\sum_{j=1}^{N}|r(i-j)|^m = O(N^{2H}L(N)) \quad N\to\infty.$

Lemma 3.3

$$\{X_i\}\in(m)(D,L(\cdot)) \quad \text{implies} \quad \{X_i\}\in(m)'(1-\frac{mD}{2},\frac{2L^m(\cdot)}{(1-mD)(2-mD)}). \tag{3.20}$$

Conversely, suppose that $r(k)$ is monotone decreasing for large k. Then

$$\{X_i\}\in(m)'(H,L(\cdot)) \quad \text{implies} \quad \{X_i\}\in(m)(\frac{2-2H}{m},[H(2H-1)L(\cdot)]^{1/m}). \tag{3.21}$$

Theorem 3.7 *Let $G\in\mathcal{G}_m$ for some $m>1$.*

(i) If $\{X_i\}\in(m)'(H,L(\cdot))$, then

$$\text{Var}(\sum_{i=1}^{N}G(X_i)) \sim \frac{J^2(m)}{m!}N^{2H}L(N), \quad N\to\infty, \frac{1}{2}<H<1 \tag{3.22}$$

where $J(m)=EG(X)H_m(X)$.

(ii) If the sequence $r(k)$ is non-negative for large k and converges as $k\to\infty$, then (3.22) *entails $\{X_i\}\in(m)'(H,L(\cdot))$.*

Suppose

$$Z_{N,m}(t)=\frac{1}{d_N}\sum_{i=1}^{[Nt]}H_m(X_i). \tag{3.23}$$

Definition 3.8 (Properties $\Pi(H)$ of $\bar{Z}(t)$) (i) $\bar{Z}(0)=0$ *a.s.*

(ii) *$\bar{Z}(t)$ has strictly stationary increments, that is the random function $M_h(t)=\bar{Z}(t+h)-\bar{Z}(t)$, $h\geq 0$, is strictly stationary.*

(iii) *$\bar{Z}(t)$ is semi-stable of order H, that is*

$$P\{\bar{Z}(ct_1)\leq x_1,\dots,\bar{Z}(ct_p)\leq x_p\} = P\{c^H\bar{Z}(t_1)\leq x_1,\dots,c^H\bar{Z}(t_p)\leq x_p\} \tag{3.24}$$

(iv) *$E\bar{Z}(t)=0$ and $E|\bar{Z}(t)|^\gamma<\infty$ for $\gamma\leq\frac{1}{H}$.*

(v) *$\bar{Z}(t)$ is separable and a.s. continuous.*

Note that Properties $\Pi(H)$ are scale-invariant.

Theorem 3.9 *Let $G\in\mathcal{G}_m$ for some $m\geq 1$ and suppose $\{X_i\}\in(m)'(H,L(\cdot))$. If $d_N^2\sim N^{2H}L(N)$ and the finite-dimensional distribution of $Z_{N,m}(t)$ converge, then $Z_N(t)$ converges weakly in $D[0,1]$ to some process $\frac{J(m)}{m!}\bar{Z}_m(t)$ endowed with the properties $\Pi(H)$.*

Definition 3.10 (the fractional Brownian motion process) *The fractional Brownian motion process $B_H(t)$, defined for $0<H<1$, is a Gaussian process endowed with the properties $\Pi(H)$. In particular, $EB_H(t)=0$ and $EB_H^2(t)=t^{2H}$.*

In the case $m=1$, the limiting process $\bar{Z}(t)$ is the fractional Brownian motion process $B_H(t)$. In the case $m=2$, the limiting process $\bar{Z}(t)$ is called the Rosenblatt process.

Chapter 4
Central limit theorems

4.1 The Classical Central Limit Thorems

4.1.1 Introduction

The foundation of asymptotic statistics is *central limit theorem* (CLT). A sophisticated statistics are necessarily has a ideal limit of errors of the decision or inference. The definition of loss function is of course crucial because it has an large effect on the statistics. However, in the common case, we will set the square error as our loss function, and at the same time, the error of the inference will be evaluated as the variance between the statistics and the real value.

The initial inference may date back to that of mean. The sample mean $\bar{X}$, which is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{consider } X = (X_i)_{i=1,\dots,n} \text{ is independent identical distributed (i.i.d).}$$

are statistics of mean μ of X's underlined distribution. The statistics is asymptotically normal and the variance of the error will become smaller and smaller if we increase the number of observations.

In fact, almost all statistics have the property of *asymptotically normal* (AN), since the inference not only has to be accurate, it also has to have a small error. The way to decrease error has been thought in many situations, and the smaller error does not have to has the property of accuracy, like Jack-knife, which is also studied very much. In the book, we will think the usual case that the statistics are accuracy, or we call it *unbiased*.

4.1.2 The Classical Central Limit Theorems

Theorem 4.1 (Multivariate Central Limit Theorem) *Let $\{\boldsymbol{X}_k\}$ be a sequence of i.i.d d-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Let $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Then*

$$n^{1/2}(\bar{X} - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Sigma).$$

Theorem 4.2 (Cramér-Wold Device) *Assume X_n and X are d-dimensional random vectors. Then*

$$X_n \Rightarrow X \quad \text{if and only if} \quad t'X_n \Rightarrow t'X \quad \text{for all } t \in \mathbb{R}^k. \tag{4.1}$$

Theorem 4.3 (Lindeberg Central Limit Theorem) *Suppose $X_{n,1}, \ldots, X_{n,k_n}$ are independent real-valued random variables for each n. Assume $E(X_{n,i}) = 0$ and $\sigma^2_{n,i} = E(X^2_{n,i}) < \infty$. Let $s^2_n = \sum_{i=1}^{k_n} \sigma^2_{n,i}$. Suppose for each $\epsilon > 0$,*

$$\sum_{i=1}^{k_n} \frac{1}{s_n^2} E[\, X^2_{n,i} I\{\, |X_{n,i}|\,\} > \epsilon s_n \,] \to 0 \quad n \to \infty. \tag{4.2}$$

Then

$$\sum_{i=1}^{k_n} X_{n,i}/s_n \rightsquigarrow \mathcal{N}(0,1). \tag{4.3}$$

Corollary 4.4 (Lyapounov Central Limit Theorem) *Suppose $X_{n,1}, \ldots, X_{n,k_n}$ are independent for each n. Assume that $E(X_{n,i}) = 0$ and $\sigma^2_{n,i} = E(X^2_{n,i}) < \infty$. Let $s^2_n = \sum_{i=1}^{k_n} \sigma^2_{n,i}$. Suppose*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \frac{1}{s_n^{2+\delta}} E[\, |X_{n,i}|^{2+\delta} \,] = 0 holds. \tag{4.4}$$

Then

$$\sum_{i=1}^{k_n} X_{n,i}/s_n \rightsquigarrow \mathcal{N}(0,1). \tag{4.5}$$

4.2 Central Limit Theorem for M-estimation

4.2.1 Notations

- $X_1, \ldots, X_n \sim$ i.i.d. F;
- ρ: $\mathbb{R} \to \mathbb{R}$: a continuous convex function;
- $Q(\xi) = \sum_{i=1}^n \rho(x_i - \xi)$;
- $Q^* = \inf_\xi Q(\xi)$
- $[\, T_n(x)\,] = \{\xi^* \,|\, Q(\xi^*) = Q^*\}$;
- $\psi = \rho'$;
- $\lambda(\xi) = \int \psi(t - \xi) F(dt)$.

Lemma 4.1 *Assume that*

- $\lambda(c) = 0$*;*
- $\lambda(\xi)$ *is differentiable at* $\xi = c$ *and* $\lambda'(c) < 0$*,*
- $\int \phi^2(t-\xi)F(dt)$ *si finite and continuous at* $\xi = c$*.*

Then $n^{1/2}(T_n(x) - c) \rightsquigarrow \mathcal{N}(0, V(\psi, F))$, *where*

$$V(\psi, F) = \int \psi^2(t-c)F(dt)/(\lambda'(c))^2. \tag{4.6}$$

Proof. Let

$$s^2 = \int (\psi(t - g\sigma n^{-1/2})) - \lambda(g\sigma n^{-1/2}))^2 F(dt), \tag{4.7}$$

then the $y_i = (\psi(x_i - g\sigma n^{-1/2}) - \lambda(g\sigma n^{-1/2}))/s$ are independent random variables with mean 0 and variance 1.

4.3 Functional Central Limit Theorem

4.3.1 Notations

- $\xi_1, \ldots, \xi_n \sim$ i.i.d. F;
- $S_n = \xi_1 + \cdots + \xi_n$;
- C: the space of continuous functions on $[0, 1]$ with uniform topology;
  D: the space of càdlàg function on $[0, 1]$ with Skorohod topology, that is, for $x, y \in D$ there exist a $\lambda \in \Lambda$ such that

$$\sup_t |\lambda t - t| \le \epsilon, \tag{4.8}$$

$$\sup_t |x(t) - y(\lambda t)| \le \epsilon. \tag{4.9}$$

  Here, Λ denote the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself.
- A random element X_n of C or D by

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}. \tag{4.10}$$

Theorem 4.5 *Suppose the random variables* ξ_n *are i.i.d* $(0, \sigma^2)$. *Then the random functions* X_n *satisfy*

$$X_n \rightsquigarrow W, \tag{4.11}$$

where W *is a Brownian Motion with*

$$EW_t = 0, \tag{4.12}$$

$$EW_s W_t = s \quad \textit{if } s \le t. \tag{4.13}$$

4.4 Central Limit Theorems on Dependent Sequence

An example by Herrndorf (1983) showed that moment conditions up to second order are not enough for a standard version of a central limit theorem. This implies that one may need more than summability of cumulants up to fourth order in the contest of spectral density estimates.

4.4.1 Notations

- $(\Omega, \mathcal{F}, P)$: a basic probability space.
- $\mathscr{G}$ and $\mathscr{H}$: Two measurable sub-σ-fields of $\mathcal{F}$,
- $\phi(\mathscr{G}, \mathscr{H}) = \sup_{G\in\mathscr{G},\, H\in\mathscr{H},\, P(G)>0} |P(H|G) - P(H)|$,
- $\alpha(\mathscr{G}, \mathscr{H}) = \sup_{G\in\mathscr{G},\, H\in\mathscr{H}} |P(H\cap G) - P(G)P(H)|$.
- $\mathcal{B}^n_{-\infty}$: the smallest collection of subsets of Ω that contains the union of the σ-fields $\mathcal{B}^n_a$ as $a \to -\infty$.
- $\mathcal{B}^\infty_{n+m}$: the smallest collection of subsets of Ω that contains the union of the σ-fields $\mathcal{B}^a_{n+m}$ as $a \to \infty$.
- $\|X\|_p := (E|X|^p)^{1/p}$.

4.4.2 Mixing Inequalities

From (1.15), we obtain

$$\alpha(\mathscr{G}, \mathscr{H}) \le \phi(\mathscr{G}, \mathscr{H}) \sup_{G\in\mathscr{G},\, H\in\mathscr{H},\, P(G)>0} |P(G)|, \tag{4.14}$$

if $P(G) \neq 0$. Therefore,

$$\phi(\mathscr{G}, \mathscr{H}) = 0 \Rightarrow \alpha(\mathscr{G}, \mathscr{H}) = 0. \tag{4.15}$$

Definition 4.6 *For a sequence of random vectors $\{X_t\}$ with $\mathcal{B}^n_{-\infty}$ and $\mathcal{B}^\infty_{n+m}$, define the mixing conefficients,*

$$\phi(m) = \sup_n \phi(\mathcal{B}^n_{-\infty}, \mathcal{B}^\infty_{n+m}), \tag{4.16}$$

$$\alpha(m) = \sup_n \alpha(\mathcal{B}^n_{-\infty}, \mathcal{B}^\infty_{n+m}). \tag{4.17}$$

If $\phi(m) \to 0$ ($\alpha(m) \to 0$) as $m \to \infty$ ($\alpha \to \infty$), we say $\{X_t\}$ is ϕ-mixing (α-mixing), respectively.

Theorem 4.7 *Let g be a measurable function into $\mathbb{R}^k$ and define $Y_t = g(X_t, \ldots, X_{t+\tau})$, where τ is finite. If the sequence of $\{X_t\}$ is ϕ-mixing, then Y_t is ϕ-mixing.*

Proof. See White and Domowitz (1984, Lemma 2.1).

As indicated in White (2001), whereas functions of ergodic processes retain ergodicity for any τ, finite of infinite, mixing is guaranteed only for finite τ.

Theorem 4.8 *Suppose that X and Y are random variables which are $\mathscr{G}$- and $\mathscr{H}$- measurable, respectively, and $\boldsymbol{X} \le C_1$, $\boldsymbol{Y} \le C_2$ a.s. Then*

$$|EXY - EXEY| \le 4C_1C_2\alpha(\mathscr{G},\mathscr{H}). \tag{4.18}$$

Corollary 4.9 *Suppose that X and Y are random variables which are $\mathscr{G}$- and $\mathscr{H}$- measurable, respectively, and that $E|X|^p < \infty$ for some $p > 1$, while $|Y| \le C$ a.s. Then*

$$|EXY - EXEY| \le 6C\|X\|_p\alpha(\mathscr{G},\mathscr{H})^{1-1/p}. \tag{4.19}$$

Corollary 4.10 *Suppose that X and Y are random variables which are $\mathscr{G}$- and $\mathscr{H}$-measurable, respectively, and that $E|X|^p < \infty$, $E|Y|^q < \infty$, where p, $q > 1$, $p^{-1} + q^{-1} < 1$. Then*

$$|EXY - EXEY| \le 8\|X\|_p\|Y\|_q\alpha(\mathscr{G},\mathscr{H})^{1-1/p-1/q}. \tag{4.20}$$

Theorem 4.11 *Suppose that X and Y are random variables which are $\mathscr{G}$- and $\mathscr{H}$- measurable, respectively, and that $E|X|^p < \infty$, $E|Y|^q < \infty$, where p, $q > 1$, $p^{-1} + q^{-1} = 1$. Then*

$$|EXY - EXEY| \le 2\|X\|_p\|Y\|_q\phi(\mathscr{G},\mathscr{H})^{1/p}. \tag{4.21}$$

Further, the result continues to hold for $p = 1$, $q = \infty$, where

$$\|Y\|_\infty = \textit{ess sup}|Y| = \inf\{C|P(|Y| > C) = 0\}. \tag{4.22}$$

4.5 Linear Proess

Definition 4.12 *$\{X_n, \mathcal{F}_n\}$ is called a mixingale sequence if, for sequences of nonnegative constants c_n and ψ_m, where $\psi_m \to 0$ as $m \to \infty$, we have*

- $\|E(X_n|\mathcal{F}_{n-m}\|_2 \le \psi_m c_n$
- $\|X_n - E(X_n|\mathcal{F}_{n+m})\|_2 \le \psi_{m+1}c_n$

Example 4.13 (Linear Process) $X_n = \sum_{i=-\infty}^{\infty} \alpha_{i-n}\xi_i$ *with* $\sum_{i=-\infty}^{\infty} \alpha_i^2 < \infty$.
Then $\{X_n, \mathcal{F}_n\}$ is a mixingale with all $c_n^2 = \sigma^2$ and $\phi_m^2 = \sum_{|i|\ge m} \alpha_i^2$.

Theorem 4.14 (Ibragimov and Linnik (1971), Taniguchi and Kakizawa (2000)) *Let $\{X_t\}$ be a linear process*

$$X_n = \sum_{i=-\infty}^{\infty} \alpha_{i-n}\xi_i, \tag{4.23}$$

with $\{\xi_i\}$ a sequence of i.i.d $(0, \sigma^2)$. As $n \to \infty$, if

$$\sigma_n^2 \equiv E(X_1 + X_2 + \cdots + X_n)^2 \to \infty, \tag{4.24}$$

then

$$\sigma^{-1}\sum_{j=1}^{n} X_j \rightsquigarrow \mathcal{N}(0,1). \tag{4.25}$$

Furthermore, the result of the linear process can be extended to a more general case. First, we define Appell rank $m^* = m^*(G)$ by

$$m^*(G) = \inf_j \{j : c_j \neq 0\}, \tag{4.26}$$

where

$$G(x) = \sum_{j=0}^{\infty} \frac{A_j(x)}{j!} c_j. \tag{4.27}$$

Then Giraitis and Surgailis (1986) gave the following the theorem.

Theorem 4.15 *Suppose $\{X_t\}$ is generated by a linear process*

$$X_n = \sum_{i=-\infty}^{\infty} \alpha_{i-n}\xi_i \tag{4.28}$$

with all moments finite. Let $G(x)$ be a polynomial with Appel rank $m^* \geq 2$, and let $S_n = \sum_{t=1}^{n} G(X_t)$. Moreover, If

$$\liminf_{n\to\infty} \frac{\mathrm{Var}(S_n)}{n} > 0, \tag{4.29}$$

and

$$\sum_{t=-\infty}^{\infty} \Big(\sum_{s=-\infty}^{\infty} |a_{t-s}a_s| \Big)^{m^*} < \infty, \tag{4.30}$$

then

$$\frac{S_n}{\{\mathrm{Var} S_n\}^{1/2}} \rightsquigarrow \mathcal{N}(0,1). \tag{4.31}$$

4.5.1 Approach in frequency domain

Consider

$$x(n) = \sum_{j=0}^{\infty} \beta(j)\epsilon(n-j), \quad E\{\epsilon(m)\epsilon(n)\} = \delta_{m,n}, \tag{4.32}$$

under the assumption that

$$\sum_{j=0}^{\infty} |\beta(j)| < \infty. \tag{4.33}$$

Easily, we can see that the spectrum of the process is

$$f(\lambda) = \frac{1}{2\pi} |\sum_{j=0}^{\infty} \beta(j) e^{ij\omega}|^2. \tag{4.34}$$

Here, we consider the regression problem of $x(n)$ on the process $y^{(N)}(n)$ with the following assumptions almost surely.

Assumption 4.1 (Grenandar's conditions)

(i) $\lim_{N\to\infty} d^2(N) = \infty, \quad$ for $d^2(N) = \sum_{n=1}^{N} |y^{(N)}(n)|^2$,
(ii) $\lim_{N\to\infty} |y^{(N)}(N)|/d(N) = 0$,
(iii) $\lim_{N\to\infty} \left\{ \sum_{m=1}^{N-n} y^{(N)}(m) y^{(N)}(m+n) \right\} / d^2(N) = \rho(n), \quad n \geq 0.$

Note that $\rho(n)$ has the following representation

$$\rho(n) = \int_{-\pi}^{\pi} e^{in\lambda} F_y(d\lambda), \tag{4.35}$$

where F_y is an even distribution function. Further, we assume that

$$\int_{-\pi}^{\pi} f(\lambda) F_y(d\lambda) > 0 \quad \text{a.s.} \tag{4.36}$$

Theorem 4.16 *Let $x(n)$ have zero mean and consider $y^{(N)}(n)$ be generated by a process independent of $x(n)$ with Grenandar's conditions. Then*

$$\sum_{n=1}^{N} \{ y^{(N)} x(n) \} / d(N) \tag{4.37}$$

converge to the normal distribution with zero mean and variance

$$2\pi \int_{-\pi}^{\pi} f(\lambda) F_y(d\lambda), \tag{4.38}$$

if we have any of the following four conditions:

(i) *$x(n)$ is regular.*
(ii) *$x(n)$ is weakly mixing and $y^{(N)}(n)$ is stationary.*
(iii) *$x(n)$ is ergodic and $y^{(N)}(n) = n^k \cos n\lambda_j(N)$ or $y^{(N)}(n) = n^k \sin n\lambda_j(N)$ where $\lambda_j(N)$ ($j = 0, 1, \ldots, m$) is one of the frequencies which satisfy $\lambda_1(N) < \cdots < \lambda_m(N)$ and are m frequencies nearest to λ_0 of the form $\omega_t = 2\pi t/N$.*
(iv) *Without (4.33), $x(n)$ is regular and $f(\lambda)$ is piecewise continuous with no discontinuous at the jumps in $F_y(\lambda)$ and the best linear predictor is the best predictor.*

4.5.2 Central Limit Theorem for spectral density estimates

Theorem 4.17 ([104]) *Let $X = \{X_n\}$ be a strictly stationary mixing process with $EX_j = 0$. Assume that the cumulant functions of order two and four are summable. Further, let the spectral density estimate $f_n(\lambda)$ have weights $w_k^{(n)}$ defined in terms of a function $a(\cdot)$ that is piecewise continuous, continuous at zero with $a(0) = 1$, symmetric about zero and is such that $xa(x)$ is bounded. Let*

$$Y_u^{(n)}(\lambda) = \sum_{k=-c(n)}^{c(n)} X_u X_{u+k} w_k^{(n)} \cos k\lambda$$

with $w_k^{(n)} = a(b(n))$ *and* $c(n) = \alpha b_n^{-1}$ *for all sufficiently large fixed* α*. Set*

$$Z_n(\lambda) = \sum_{u=1}^{m} \frac{Y_u^{(n)}(\lambda)}{(nb_n^{-1})^{1/2}}$$

with $m = m(n)$ *and* $c(n) = o(m(n), m(n) = o(n)$*. Consider the distribution function* $F_{n,m}(x)$ *of* $Z_n(\lambda)$ *and assume that*

$$\frac{n}{m(n)} \inf_{|x|>\eta} x^2 dF_{n,m(n)}(x) \to 0$$

as $n \to \infty$*,* $b_n \to 0$*,* $nb_n \to \infty$ *for each* $\eta > 0$*. Then* $f_n(\lambda) - Ef_n(\lambda)$ *is asymptotically normally distributed with mean zero and variance*

$$\frac{2\pi(1+\eta(\lambda))}{nb_n} f^2(\lambda) \int W^2(\alpha) d\alpha.$$

Chapter 5
Asymptotics in Non-regular Case

5.1 Limiting Distribution for L_1 Regression Estimation under General Conditions

5.1.1 Assumptions

(A1) $\{\epsilon_i\}$ are i.i.d. random variables with median 0 with distribution function F continuous at 0.
(A2) For some positive definite matrix C,

$$\lim_{n\to\infty} \frac{1}{n} X_n^T X_n = C. \tag{5.1}$$

(A3) For each $\boldsymbol{u}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Psi_n(\boldsymbol{u}^T \boldsymbol{x}_i) = \tau(\boldsymbol{u}) \tag{5.2}$$

for some convex function $\tau(\boldsymbol{u})$ taking values in $[0.\infty]$, where $\{\Psi_n(t)\}$ is defined as

$$\Psi_n(t) = \int_0^t \sqrt{n}(F(s/a_n) - F(0))\, ds, \tag{5.3}$$

which for each n is a convex function.

5.1.2 Main Result

Theorem 5.1 *Assume*

$$Y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \tag{5.4}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$ *and* $\boldsymbol{x}_i = (1, x_{1i}, \ldots, x_{pi})^T$. *Define*

$$Z_n(\boldsymbol{u}) = \frac{a_n}{\sqrt{n}} \sum_{i=1}^{n} [|\epsilon_i - \boldsymbol{x}_i^T \boldsymbol{u}/a_n| - |\epsilon_i|]. \tag{5.5}$$

Under assumptions (A1)–(A3), for any $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$*,*

$$(Z_n(\boldsymbol{u}_1), \ldots, Z_n(\boldsymbol{u}_k)) \rightsquigarrow (Z(\boldsymbol{u}_1), \ldots, Z(\boldsymbol{u}_k)), \tag{5.6}$$

where

$$Z(\boldsymbol{u}) = -\boldsymbol{u}^T W + 2\tau(\boldsymbol{u}) \tag{5.7}$$

with W *a* $(p+1)$*-variate normal random vector with mean vector* $\mathbf{0}$ *and covariance matrix* C*.*

Chapter 6
Methods in Statistics

6.1 Bracketing Methods

6.1.1 Introduction

Bracketing arguments have been developed in empirical process theory. The idea is used to prove the Glivenko-Cantelli theorem, i.e., the uniform law of large numbers.

The empirical distribution function F_n for a sample $\xi_1, \ldots, \xi_n$ from a distribution function F on the real line is defined by

$$F_n(x) = \frac{1}{n} \sum_{i \leq n} 1_{\{\xi \leq t\}}(x) \quad \text{for each } x \in \mathbb{R}. \tag{6.1}$$

The bracketing methods control the difference between the empirical distribution and the distribution function by an interval $t_1 \leq t \leq t_2$,

$$F_n(t_1) - F(t_2) \leq F_n(t) - F(t) \leq F_n(t_2) - F(t_1). \tag{6.2}$$

The two bounds converge almost surely to $F(t_1) - F(t_2)$ and $F(t_2) - F(t_1)$. If t_2 and t_1 are close enough together, then $F_n(t) - F(t)$ is also close enough. Furthermore, if the whole line can be covered by a union of finitely many such intervals, then $\sup_t |F_n(t) - F(t)|$ is eventually small.

Definition 6.1 *A bracket $[\,l, u\,]$ of a pair of P-integrable functions $l \leq u$ on $\mathcal{X}$ is defined by*

$$[\,l, u\,] := \{g : l(x) \leq g(x) \leq u(x) \quad \textit{for all } x\}. \tag{6.3}$$

For $1 \leq q \leq \infty$, the bracketing number $N^{(q)}_{[\ ]}(\delta, \mathcal{F}, P)$ for a subclass of functions $\mathcal{F} \subset \mathcal{L}^q(P)$ is defined as the smallest value of N for which there exist brackets $[\,l_i, u_i\,]$ with $\mathbb{P}(u_i - l_i)^q \leq \delta^q$ for $i = 1, \ldots, N$ and $\mathcal{F} \subset \cup_i [\,l_i, u_i\,]$.

6.1.2 *Some results in* L_2 *theory*

Define

$$\rho(f) = \sup_{n,i} \|f(\xi_{ni})\|_2, \tag{6.4}$$

where $\|\cdot\|_p$ is defined by $(\mathbb{P}|\cdot|^p)^{1/p}$.

Definition 6.2 *The bracketing number $N(\delta) = N(\delta, \mathcal{F})$ equals the smallest value of N for which there exist functions $f_1, \ldots, f_N$ in $\mathcal{F}$ and $b_1, \ldots, b_N$ with $\rho(b_i) \leq \delta$ for each i such that for each f in $\mathcal{F}$ there esixts an i for which $|f - f_i| \leq b_i$.*

Example 6.3 *Suppose $\mathcal{F}$ is a parametric family, i.e.,*

$$\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^k\}, \tag{6.5}$$

where Θ is a bounded subset. Suppose also the Lipschitz condition for $f(\cdot, \theta)$ such that

$$|f(x, \theta) - f(x, \theta')| \leq L(x)|\theta - \theta'|^\lambda, \tag{6.6}$$

with $\sup_{n,i}\|L(\xi_{ni})\|_2 = C < \infty$. Then for all r small enough,

$$\sup_{n,i} \mathbb{P} \sup_{\theta' \in B(\theta, r)} |f(\xi_{ni}, \theta') - f(\xi_{ni}, \theta)|^2 \leq C^2 r^{2\lambda} \quad \textit{for all } \theta, \tag{6.7}$$

where $B(\theta, r)$ is the ball of radius r around θ. To cover the bounded set Θ, we only have to set f_i at the centers of the $O(r^k)$ many balls of radius $r = (\delta/C)^{1/\lambda}$. The bracketing numbers is given by $O(\delta^{-k/\lambda})$.

Theorem 6.4 *Let $\{\xi_{ni}\}$ be a strong mixing triangular array whose mixing coefficients satisfy*

$$\sum_{d=1}^{\infty} d^{Q-2} \alpha(d)^{\gamma/(Q+\gamma)} < \infty \tag{6.8}$$

for some even integer $Q \geq 2$ and some $\gamma > 0$, and let $\mathcal{F}$ be a uniformly bounded class of real-valued functions whose bracketing numbers satisfy

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} \, dx < \infty \tag{6.9}$$

for the same Q and γ. Then for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n \to \infty} \Big\| \sup_{\rho(f-g) < \delta} |\nu_n f - \nu_n g| \Big\|_Q < \epsilon. \tag{6.10}$$

Corollary 6.5 (Functional Central Limit Theorem) *If the conditions of Theorem 6.4 are satisfied and if $(\nu_n f_1, \ldots, \nu_n f_k)$ has an asymptotic normal distribution for all choices of $f_1, \ldots, f_k$ from $\mathcal{F}$, then $\{\nu_n f : f \in \mathcal{F}\}$ converges in distribution to a Gaussian process indexed by $\mathcal{F}$ with ρ-continuous sample paths.*

Remark 6.6 *The conditions of Theorem 6.4 require a balance between the rate of decrease in the mixing coefficients and the rate of growth in the bracketing numbers. For example, if* $N(x) = O(x^{-\beta})$ *and* $\alpha(d) = O(d^{-A})$ *for some* $\beta > 0$ *and* $A > 0$*, then the requirements would be satisfied with* Q *equal to the smallest even integer greater than* 2β *and* $\gamma = 2$*, if* $A > (Q-1)(1+Q/2)$.

Chapter 7
Empirical Likelihood Methods

7.1 EL

7.2 Local Empirical Likelihood

Let $\{z_i\}_{i=1}^n = \{(y_i', x_i')'\}_{i=1}^n$ be a random sample on $\mathcal{Z} = \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^{d_y} \times \mathbb{R}^{d_x}$. Consider random variables $(\tilde{z}, \tilde{x})$ with discrete support $\mathcal{Z}_n \times \mathcal{X}_n = \{z_1, \ldots, z_n\} \times \{x_1, \ldots, x_n\}$. Let p_{ji} be the conditional probability $P\{\tilde{z} = z_j | \tilde{x} = x_i\}$. The Nadaraya-Watson kernel weight w_{ji} is defined by

$$w_{ji} = \frac{K(\frac{x_j - x_i}{b_n})}{\sum_{j=1}^n K(\frac{x_j - x_i}{b_n})}, \tag{7.1}$$

which is used to control the likelihood contribution. Here, K is a kernel function and b_n is a bandwidth parameter. The parameter defined in the model is denoted by $\alpha_0 = (\theta_0, h_0)$ in a compact set $\mathcal{A} = \Theta \times \mathcal{H}$, satisfying

$$E[g(z, \alpha_0)|x] = 0. \tag{7.2}$$

Under the settings above, the local empirical likelihood at $\tilde{x} = x_i$ is written by

$$\max_{\{p_{ji}\}_{j=1}^n} \sum_{j=1}^n w_{ji} \log p_{ji}, \quad \text{s.t. } p_{ji} \geq 0, \sum_{j=1}^n p_{ji} = 1, \sum_{j=1}^n p_{ji} g(z_j, \alpha) = 0 \tag{7.3}$$

for each $\alpha \in \mathcal{A}$.

The local conditional empirical likelihood ratio at $\tilde{x} = x_i$ is defined as

$$l_{in}(\alpha) = \sum_{j=1}^n w_{ji} \log \hat{p}_{ji} - \sum_{j=1}^n w_{ji} \log \tilde{p}_{ji}. \tag{7.4}$$

Based on the local conditional empirical likelihood ratio, the whole estimating equation is given by

$$l_n(\alpha) = \frac{1}{n} \sum_{i=1}^n 1\{x_i \in \mathcal{X}_n\} l_{in}(\alpha), \tag{7.5}$$

for each $\alpha \in \mathcal{A}$.

7.3 Penalized Empirical Likelihood

Let $J(h)$ be a penalty function to control some physical plausibility of h for smoothness or consistency. The PEL ratio and PEL estimator are defined by

$$l_n(\alpha) - \phi_n J(h) = -\frac{1}{n}\sum_{i=1}^{n} 1\{x_i \in \mathcal{X}_n\} \sum_{j=1}^{n} w_{ji} \log(1 + \lambda_i(\alpha)' g(z_j, \alpha)) - \phi_n J(h), \tag{7.6}$$

$$\hat{\alpha} = (\hat{\theta}, \hat{h}) = \arg\max_{\alpha \in \mathcal{A}} \{l_n(\alpha) - \phi_n J(h)\}, \tag{7.7}$$

where $\lambda_i(\alpha)$ is the Lagrangian corresponding to α.

Assumption 7.1 (i) *$\{z_i\}_{i=1}^{n}$ is i.i.d.*
(ii) *Support $\mathcal{X}_n = \prod_{d=1}^{d_x} [\underline{x}_d + b_n^{-\gamma_1}, \overline{x}_d - b_n^{-\gamma_1}]$ for some $\gamma_1 \in (0,1)$.*
(iii) *The density function of x is finite and bounded away from zero on $\mathcal{X}$ and is second-order differentiable on $\mathcal{X}$.*

Denote $\mathcal{N}(\epsilon, \mathcal{A}, \|\cdot\|_s)$ as the minimum number of radius ϵ covering balls of $\mathcal{A}$ under the norm $\|\cdot\|_s$. Also, define

$$\mathcal{A}(k_1, k_2) = \{\alpha \in \mathcal{A}; k_1 \le \|\alpha - \alpha_0\|_* \le 2k_1, J(h) \le k_2\} \tag{7.8}$$

for $k_1, k_2 \in (0, \infty)$.

Assumption 7.2 (i) *α_0 is the only $\alpha \in \mathcal{A}$ satisfying $E[g(z,\alpha)|x] = 0$.*
(ii) *$0 < \phi_n = o(n^{-1/2})$, $J(h_0) < \infty$, and $J(h) \ge 0$ for all $h \in \mathcal{H}$.*
(iii) *$\sup_{\alpha \in \mathcal{A}(k_1,k_2)} \|\alpha - \alpha_0\|_s \le C(k_1^2 + k_2)^{\gamma_2}$ for some C and $k_1, k_2 \in (0, \infty)$.*
(iv) *For $U = \phi_n^{1/2}(k_1^2 + k_2)^{(1+\gamma_2)/2}$ and $L = \phi_n(k_1^2 + k_2)$,*

$$\sup_{k_1 \ge 1, k_2 \ge 1} \int_L^U \{\log \mathcal{N}(u, \mathcal{A}(k_1,k_2), \|\cdot\|_s)\}^{1/2} du / L \le C_1 n^{1/2} \tag{7.9}$$

for some constant C_1.

Assumption 7.3 (i) *Each element of $g(z, \alpha)$ satisfies an envelope condition over $\alpha \in \mathcal{A}$ with order $m \in (8, \infty)$ and is Hölder continuous in $\alpha \in \mathcal{A}$.*
(ii) *Each element of $V(x, \alpha)$ is second-order differentiable on $\mathcal{X}$ and the second derivative is uniformly bounded on $(x, \alpha) \in \mathcal{X} \times \mathcal{A}$.*
(iii) *The smallest eigenvalue of $V(x, \alpha)$ is positive and the largest eigenvalue of $V(x, \alpha)$ is finite uniformly on $(x, \alpha) \in \mathcal{X} \times \mathcal{A}$.*

Assumption 7.4 (i) *$K(x)$ is a bounded and Lipschitz continuous function with bounded support, and is symmetric around the origin.*
(ii) *$K(x)$ is an rth order kernel function with $r \ge 2$.*

(iii) *As* $n \to \infty$, $nb_n^{4r} \to 0$ *and*

$$\inf_{k_1\geq 1, k_2 \geq 1}\Big\{\frac{n^{1/2}(k_1^2+k_2)^2 b_n^{d_x}}{\max\{1, n^{-1/4+1/(2m)}(k_1^2+k_2)^{1-2/m}b_n^{-2(d_x+1)/m}\}} \tag{7.10}$$

$$-\log\Big(\frac{\mathcal{N}(n^{-1/4}(k_1^2+k_2), \mathcal{A}(n^{-1/8}k_1, k_2), \|\cdot\|_s)}{(n^{-1/4}(k_1^2+k_2)b_n^{d_x+1})^{d_x}}\Big)\Big\} \to \infty. \tag{7.11}$$

Let $V(x,\alpha) = E[g(z,\alpha)g(z,\alpha)'|x]$ and $\|\alpha-\alpha_0\|_*$ be a Fisher-type norm for $\alpha \in \mathcal{A}$ defined as

$$\|\alpha-\alpha_0\|_* = \sqrt{E[E[g(z,\alpha)|x]'V(x,\alpha)^{-1}E[g(z,\alpha)|x]]}. \tag{7.12}$$

Theorem 7.1 *Suppose that Assumtions 7.2-7.4 hold. Then*

(i) $\|\hat{\alpha}-\alpha_0\|_* = o_p(n^{-1/4})$.
(ii) $P\{J(\hat{h}) \geq (1+\delta)\max\{J(h_0),1\}\} \to 0$ *for some* $\delta \in (0,1)$.

Define $D_{h(l)}(x,\alpha_0)$ by

$$D_{h(l)}(x,\alpha_0) = E\Big[\frac{dg(z,\alpha_0)}{d\theta_{(l)}}\Big|x\Big] - E\Big[\frac{dg(z,\alpha_0)}{dh}[h_{(l)}-h_0]\Big|x\Big] \tag{7.13}$$

and $h^*_{(l)}$ by

$$h^*_{(l)} = \arg\min_{h_{(l)}\in \mathrm{Var} H} E\Big[D_{h(l)}(x,\alpha_0)'V(x,\alpha_0)^{-1}D_{h(l)}(x,\alpha_0)\Big], \tag{7.14}$$

for $l = 1,\ldots,d_\theta$ with d_θ as the dimension of Θ, and $\bar{H}$ as the closure of the linear span of $\mathcal{H}$. The shrinking subset $\mathcal{B}_n$ is defined by

$$\mathcal{B}_n = \{\alpha\in\mathcal{A}; \|\alpha-\alpha_0\|_* \leq c_n, J(h) \leq (1+\delta)\max\{J(h_0),1\}\} \text{ with } c_n = o(n^{-1/4}). \tag{7.15}$$

Assumption 7.5 (i) θ_0 *is an interior point of* $\Theta \subset \mathbb{R}^{d_\theta}$.
(ii) $E[D(x,\alpha_0)'V(x,\alpha_0)^{-1}D(x,\alpha_0)]$ *is positive definite.*
(iii) $J(v_{0h}) < \infty$ *and* $J(h+\epsilon_n v_{0h}) - J(h) \leq C\epsilon_n^{\gamma_3}J(v_{0h})$ *for some* $\gamma_3 \in [1,\infty)$ *and all* $h \in \mathcal{H}$ *being a subvector of* $\alpha \in \mathcal{B}_n$ *and* $\epsilon_n = o(n^{-1/2})$.
(iv) *There exist a measurable function* $c(x)$ *and a constant* $\gamma_4 \in [1/2,\infty)$ *such that* $\|E[g(z,\alpha)|x]\| \leq c(x)\|\alpha-\alpha_0\|_*^{\gamma_4}$ *for all* $x \in \mathcal{X}$ *and* $\alpha \in \mathcal{B}_n$ *with* $c(x) < \infty$.
(v) $\alpha + tv_0 \in \mathcal{A}$ *for all small* $t \in [0,1]$ *and all* $\alpha \in \mathcal{B}_n$, $g(z,\alpha+tv_0)$ *is second-order differentiable a.s. at* $t=0$ *for all* $\alpha\in\mathcal{B}_n$, *each element of* $g_\alpha[z,v_0]$ *satisfies an envelope condition over* $\alpha\in\mathcal{B}_n$ *with order 2 and is Hölder continuous in* $\alpha\in\mathcal{B}_n$, *and each element of* $d^2g(z,\alpha+tv_0)/dt^2|_{t=0}$ *satisfies an envelope condition over* $\alpha\in\mathcal{B}_n$ *with order 2.*
(vi) $E[E[g_{\alpha_0}[z,v_0]|x]'V(x,\alpha_0)^{-1}\{E[g_{\bar{\alpha}}[z,\alpha-\alpha_0]|x] - E[g_{\alpha_0}[z,\alpha-\alpha_0]|x]\}] = o(n^{-1/2})$ *uniformly on* $\alpha,\bar{\alpha}\in\mathcal{B}_n$, *and* $E[\|E[g_\alpha[z,v_0]|x]'V(x,\alpha_0)^{-1} - E[g_{\alpha_0}[z,v_0]|x]'V(x,\alpha_0)^{-1}\|^2] = o(n^{-1/2})$ *uniformly on* $\alpha\in\mathcal{B}_n$.

Theorem 7.2 *Suppose that Assumptions - hold. Then*

$$\sqrt{n}(\hat{\theta}-\theta_0) \rightsquigarrow \mathcal{N}(0,\Sigma^{-1}), \quad \Sigma^{-1} = E[D(x,\alpha_0)'V(x,\alpha_0)^{-1}D(x,\alpha_0)]. \tag{7.16}$$

Chapter 8
Supplement

Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is absolutely continuous function, then

$$\text{total variation of } g = \int \|\nabla g\| dx. \tag{8.1}$$

Let $T_n = T_n(x_1, \ldots, x_n)$ be an estimate of the scalar parameter θ in the distribution P_θ^n for $x^n = (x_1, \ldots, x_n)$, and let $\mathrm{EFF}(T_n, P_\theta^n)$ denote a suitably defined efficiency of T_n at P_θ^n.

$$EFF() = \frac{\mathrm{Var}_{CR}(P_\theta^n)}{\mathrm{Var}_{P_\theta^n}(T_n)}, \tag{8.2}$$

where V_{CR} is the Cramer-Rao lower bound at P_θ^n.

The measure for the process $\{x_n\}_{n \geq 1}$ is denoted P_θ^∞, and the efficiency of T at F_θ is

$$EFF(T, P_\theta^\infty) = \frac{1}{i(P_\theta^\infty)V_\infty(T)}. \tag{8.3}$$

where $V_\infty(T)$ is the asymptotic variance of $\sqrt{n}T_n$ at P_θ^∞ and $i(P_\theta^\infty) = \lim_{n\to\infty} n^{-1} i(P_\theta^n)$ is the asymptotic Fisher information for θ, $i(P_\theta^n)$ being the finite sample Fisher information for θ.

A min-max robust estimate T_0 solves the problem

$$\inf_{T \in \mathcal{T}} \sup_{P \in \mathcal{P}^\infty} V(T, P^\infty). \tag{8.4}$$

The solution to this problem is usually obtained by solving the saddle point problem

$$\sup_{P^\infty} \inf_{T\in} V(T, P^\infty) = V(T_0, P_0^\infty) = \inf_{T} \sup_{P^\infty} V(T, P^\infty). \tag{8.5}$$

θ

$$C^{-1}(\varphi, \theta) \tag{8.6}$$

μ H

$$var(F)H^2(\varphi, \theta) \tag{8.7}$$

References

1. Albrecht, V.: On the convergence rate of probability of error in bayesian discrimination between two gaussian processes. In: Proc. of the Third Prague symposium on asymptotic statistics (1984)
2. Anderson, T.: The statistical analysis of time series. J. Wiley and sons (1971)
3. Andrews, D.W., Pollard, D.: An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review/Revue Internationale de Statistique pp. 119–132 (1994)
4. Bartkiewicz, K., Jakubowski, A., Mikosch, T., Wintenberger, O.: Stable limits for sums of dependent infinite variance random variables. Probability theory and related fields **150**(3-4), 337–372 (2011)
5. Basawa, I.: Neyman-le cam tests based on estimating functions. In: Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, vol. 2, pp. 811–825 (1985)
6. Basu, A., Harris, I.R., Hjort, N.L., Jones, M.: Robust and efficient estimation by minimising a density power divergence. Biometrika **85**(3), 549–559 (1998)
7. Beran, J.: On a class of m-estimators for gaussian long-memory models. Biometrika **81**(4), 755–766 (1994)
8. Beran, R.: An efficient and robust adaptive estimator of location. The Annals of Statistics **6**(2), 292–313 (1978)
9. Bernstein, S.: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Mathematische Annalen **97**(1), 1–59 (1927)
10. Bhansali, R.J.: Autoregressive and window estimates of the inverse correlation function. Biometrika **67**(3), 551–566 (1980)
11. Bickel, P., Klaassen, C., Ya'acov Ritov, J., Wellner, J.: Efficient and adaptive estimation for semiparametric models, vol. 1. Springer New York (1998)
12. Billingsley, P.: Convergence of probability measures, vol. 493. Wiley-Interscience (1968)
13. Bloomfield, P.: An exponential model for the spectrum of a scalar time series. Biometrika **60**(2), 217–226 (1973)
14. Bose, A.: Bahadur representation of M_m estimates. The Annals of Statistics **26**(2), 771–777 (1998)
15. Brillinger, D.R.: Time series: data analysis and theory, vol. 36. Siam (2001)
16. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer-Verlag (1991)
17. Can, S., Mikosch, T., Samorodnitsky, G.: Weak convergence of the function-indexed integrated periodogram for infinite variance processes. Bernoulli **16**(4), 995–1015 (2010)
18. Cleveland, W.S.: The inverse autocorrelations of a time series and their applications. Technometrics **14**(2), 277–293 (1972)
19. Danielsson, J., de Haan, L., Peng, L., de Vries, C.G.: Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate analysis **76**(2), 226–248 (2001)
20. Davis, R., Hsing, T.: Point process and partial sum convergence for weakly dependent random variables with infinite variance. The Annals of Probability **23**(2), 879–917 (1995)
21. Davis, R., Marengo, J., Resnick, S.: Extremal properties of a class of multivariate moving averages (1986)
22. Davis, R., Resnick, S.: Limit theory for moving averages of random variables with regularly varying tail probabilities. The Annals of Probability pp. 179–195 (1985)
23. Davis, R., Resnick, S.: More limit theory for the sample correlation function of moving averages. Stochastic Processes and their Applications **20**(2), 257–279 (1985)
24. Davis, R., Resnick, S.: Limit theory for the sample covariance and correlation functions of moving averages. The Annals of Statistics pp. 533–558 (1986)

25. Denker, M., Keller, G.: On u-statistics and v. mises' statistics for weakly dependent processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **64**(4), 505–522 (1983)
26. DiCiccio, T., Hall, P., Romano, J.: Empirical likelihood is bartlett-correctable. The Annals of Statistics pp. 1053–1061 (1991)
27. Dunsmuir, W.: A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. The Annals of Statistics **7**(3), 490–506 (1979)
28. Dzhaparidze, K.: Parameter estimation and hypothesis testing in spectral analysis of stationary time series. Springer (1986)
29. Dzhaparidze, K.O.: On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary gaussian process with rational spectral density. Theory of Probability & Its Applications **16**(3), 550–554 (1971)
30. Embrechts, P., Goldie, C.M.: On closure and factorization properties of subexponential and related distributions. Journal of the Australian Mathematical Society (Series A) **29**(02), 243–256 (1980)
31. Fabian, V., Hannan, J.: On estimation and adaptive estimation for locally asymptotically normal families. Probability Theory and Related Fields **59**(4), 459–478 (1982)
32. Feller, W.: An introduction to probability theory and its applications, vol. 2. John Wiley & Sons (1971)
33. Fujisawa, H., Eguchi, S.: Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis **99**(9), 2053–2081 (2008)
34. Fuller, W.A.: Introduction to Statistical Time Series, vol. 230. John Wiley & Sons (1996)
35. Garel, B., Hallin, M.: Local asymptotic normality of multivariate arma processes with a linear trend. Annals of the Institute of Statistical Mathematics **47**(3), 551–579 (1995)
36. Geyer, C.J.: On the asymptotics of convex stochastic optimization. Unpublished manuscript **37** (1996)
37. Godambe, V.P.: An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics **31**(4), 1208–1211 (1960)
38. Godambe, V.P., Thompson, M.E.: Some aspects of the theory of estimating equations. Journal of Statistical Planning and Inference **2**(1), 95–104 (1978)
39. Godambe, V.P., Thompson, M.E.: Robust estimation through estimating equations. Biometrika **71**(1), 115–125 (1984)
40. Grenander, U., Rosenblatt, M.: Statistical Analysis of Stationary Time Series. Johm Wiley & Sons, New York (1957)
41. Griffin, P.S., Mason, D.M.: On the asymptotic normality of self-normalized sums. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 109, pp. 597–610. Cambridge Univ Press (1991)
42. Hagemann, A.: Robust spectral analysis. Unpublished manuscipt, Department of Exonomics, University of Illinois, arXiv preprint arXiv:1111.1965 (2011)
43. Hagemann, A.: Stochastic equicontinuity in nonlinear time series models. The Econometrics Journal **17**(1), 188–196 (2014)
44. Hájek, J., Šidák, Z., Sen, P.: Theory of rank tests. Academic press New York (1967)
45. Hall, P.: On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society. Series B (Methodological) pp. 37–42 (1982)
46. Hall, P., Heyde, C.C.: Martingale limit theory and its application. Academic press New York (1980)
47. Hallin, M., Ingenbleek, J., Puri, M.: Linear serial rank tests for randomness against arma alternatives. The Annals of Statistics pp. 1156–1181 (1985)
48. Hallin, M., Ingenbleek, J., Puri, M.: Linear and quadratic serial rank tests for randomness against serial dependence. Journal of time series analysis **8**(4), 409–424 (1987)
49. Hallin, M., Puri, M.: Rank tests for time series analysis a survey (1991)
50. Hallin, M., Werker, B.: Optimal testing for semi-parametric ar models-from gaussian lagrange multipliers to autoregression rank scores and adaptive tests. statistics textbooks and monographs **158**, 295–350 (1999)
51. Hallin, M., Werker, B.: Semi-parametric efficiency, distribution-freeness and invariance. Bernoulli **9**(1), 137–165 (2003)
52. Hamilton, J.D.: Time series analysis, vol. 2. Princeton university press Princeton (1994)
53. Hannan, E.: Multiple Time Series. John Wiley & Sons (1970)
54. Hannan, E.J.: The asymptotic theory of linear time-series models. Journal of Applied Probability **10**(1), 130–145 (1973)
55. Hannan, E.J.: The estimation of frequency. Journal of Applied probability **10**, 510–519 (1973)
56. Hewitt, E., Stromberg, K.: Real and abstract analysis. A modern treatment of the theory of functions of a real variable. Third printing, vol. 25. Splinger-Verlag (1975)

57. Hjort, N., Pollard, D.: Asymptotics for minimisers of convex processes. Unpublished manuscript (1993)
58. Hodges, J.L., Lehmann, E.L.: Estimates of location based on rank tests. The Annals of Mathematical Statistics **34**(2), 598–611 (1963)
59. Hoeffding, W.: A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics **19**(3), 293–325 (1948)
60. Hosoya, Y.: The bracketing condition for limit theorems on stationary linear processes. The Annals of Statistics **17**(1), 401–418 (1989)
61. Hosoya, Y., Taniguchi, M.: A central limit theorem for stationary processes and the parameter estimation of linear processes. The Annals of Statistics **10**, 132–153 (1982)
62. Hsing, T., Wu, W.B.: On weighted u-statistics for stationary processes. Annals of Probability pp. 1600–1631 (2004)
63. Huber, P.: Robust statistics. Wiley, New York (2009)
64. Huber, P.J.: Robust estimation of a location parameter. The Annals of Mathematical Statistics **35**(1), 73–101 (1964)
65. Ibragimov, I.A., Linnik, Y.V.: Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen (1971)
66. Inagaki, N., Kondo, M.: Several estimators of the autocorrelation based on limiter estimating functions for a stationary gaussian process. Journal of the Japan Statistical Society **10**(1), 1–15 (1980)
67. Kanamori, T., Fujisawa, H.: Affine invariant divergences associated with proper composite scoring rules and their applications. Bernoulli **20**(4), 2278–2304 (2014)
68. Kholevo, A.: On estimates of regression coefficients. Theory of Probability & Its Applications **14**(1), 79–104 (1969)
69. Kitamura, Y., et al.: Empirical likelihood methods with weakly dependent processes. The Annals of Statistics **25**(5), 2084–2102 (1997)
70. Kley, T., Volgushev, S., Dette, H., Hallin, M.: Quantile spectral processes: Asymptotic analysis and inference. arXiv preprint arXiv:1401.8104 (2014)
71. Klüppelberg, C., Mikosch, T.: Spectral estimates and stable processes. Stochastic processes and their applications **47**(2), 323–344 (1993)
72. Klüppelberg, C., Mikosch, T.: Some limit theory for the self-normalized periodogram of stable processes. Scandinavian Journal of Statistics **21**, 485–491 (1994)
73. Klüppelberg, C., Mikosch, T.: The integrated periodogram for stable processes. The Annals of Statistics pp. 1855–1879 (1996)
74. Knight, K.: Limiting distributions for l_1 regression estimators under general conditions. The Annals of Statistics **26**(2), 755–770 (1998)
75. Koenker, R.: Quantile regression. 38. Cambridge university press (2005)
76. Koenker, R., Bassett, J.G.: Regression quantiles. Econometrica: journal of the Econometric Society **46**(1), 33–50 (1978)
77. Kreiss, J.: On adaptive estimation in stationary arma processes. The Annals of Statistics pp. 112–133 (1987)
78. Lehmann, E., Romano, J.: Testing statistical hypotheses. Springer (2005)
79. Li, T.H.: Laplace periodogram for time series analysis. Journal of the American Statistical Association **103**(482), 757–768 (2008)
80. Li, T.H.: Quantile periodograms. Journal of the American Statistical Association **107**(498), 765–776 (2012)
81. Liu, Y.: Asymptotic moments of symmetric self-normalized sums. Scientiae Mathematicae Japonicae Online **26**, 561–569 (2013)
82. Liu, Y.: Asymptotics for m-estimators in time series. Advances in Science, Technology and Environmentology **B10**, 63–77 (2014)
83. Logan, B., Mallows, C., Rice, S., Shepp, L.: Limit distributions of self-normalized sums. The Annals of Probability **1**(5), 788–809 (1973)
84. Lütkepohl, H.: New introduction to multiple time series analysis. Cambridge Univ Press (2005)
85. Magnus, W., Oberhettinger, F.: Formulas and theorems for the functions of mathematical physics. Chelsea Pub. Co. (1954)
86. Meerschaert, M., Scheffler, H.: Limit distributions for sums of independent random vectors: Heavy tails in theory and practice, vol. 321. Wiley-Interscience (2001)
87. Mikosch T., R.S., Samorodnitsky, G.: The maximum of the periodogram for a heavy-tailed sequence. The Annals of Probability **28**(2), 885–908 (2000)

88. Mikosch, T., Gadrich, T., Kluppelberg, C., Adler, R.: Parameter estimation for arma models with infinite variance innovations. The Annals of Statistics **23**(1), 305–326 (1995)
89. Monti, A.C.: Empirical likelihood confidence regions in time series models. Biometrika **84**(2), 395–405 (1997)
90. Monti, A.C., Ronchetti, E.: On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate m-estimators. Biometrika **80**(2), 329–338 (1993)
91. Niemiro, W.: Asymptotics for M-estimators defined by convex minimization. The Annals of Statistics pp. 1514–1533 (1992)
92. Nolan, J.: Stable Distributions - Models for Heavy Tailed Data. Birkhauser (2012)
93. Ogata, H., Taniguchi, M.: An empirical likelihood approach for non-gaussian vector stationary processes and its application to minimum constrast estimation. Australian and New Zealand Journal of Statistics **52**(4), 451–468 (2010)
94. Owen, A.: Empirical likelihood ratio confidence intervals for a single functional. Biometrika **75**(2), 237–249 (1988)
95. Owen, A.: Empirical likelihood ratio confidence regions. The Annals of Statistics **18**(1), 90–120 (1990)
96. Peña, V.H., Lai, T.L., Shao, Q.M.: Self-normalized processes: Limit theory and Statistical Applications. Springer (2009)
97. Petrov, V.V.: Sums of Independent Random Variables. Springer, Berlin (1975)
98. Pollard, D.: Asymptotics for least absolute deviation regression estimators. Econometric Theory **7**(02), 186–199 (1991)
99. Quinn, B.G., Hannan, E.J.: The estimation and tracking of frequency, vol. 9. Cambridge University Press (2001)
100. Quinn, B.G., Thomson, P.J.: Estimating the frequency of a periodic function. Biometrika **78**(1), 65 (1991)
101. Resnick, S.: Extreme values, regular variation, and point processes. Springer (2007)
102. Resnick, S., Stărică, C.: Tail index estimation for dependent data. The Annals of Applied Probability **8**(4), 1156–1183 (1998)
103. Rice, J.A., Rosenblatt, M.: On frequency estimation. Biometrika **75**(3), 477–484 (1988)
104. Rosenblatt, M.: Asymptotic normality, strong mixing and spectral density estimates. The Annals of Probability **12**(4), 1167–1180 (1984)
105. Rozanov, Y.A.: On a new class of statistical estimates. Soviet-Japanese Symposium, Theory of Probability pp. 239–252 (1969)
106. Rozanov, Y.A.: Some approximation problems in the theory of stationary processes. Journal of Multivariate Analysis **2**(2), 135–144 (1972)
107. Samoradnitsky, G., Taqqu, M.: Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall/CRC (1994)
108. Sen, P.K.: Limiting behavior of regular functionals of empirical distributions for stationary*-mixing processes. Probability Theory and Related Fields **25**(1), 71–82 (1972)
109. Serfling, R.J.: Approximation theorems of mathematical statistics, vol. 162. Wiley. com (1980)
110. Skowronek, S., Volgushev, S., Kley, T., Dette, H., Hallin, M.: Quantile spectral analysis for locally stationary time series. arXiv preprint arXiv:1404.4605 (2014)
111. Strang, G.: Linear Algebra and Its Applications. Cengage Learning; 4th edition (2005)
112. Swensen, A.: The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. Journal of Multivariate Analysis **16**(1), 54–70 (1985)
113. Taniguchi, M.: On estimation of parameters of gaussian stationary processes. Journal of Applied Probability **16**(3), 575–591 (1979)
114. Taniguchi, M.: On estimation of the integrals of certain functions of spectral density. Journal of Applied Probability **17**(1), 73–83 (1980)
115. Taniguchi, M.: An estimation procedure of parameters of a certain spectral density model. Journal of the Royal Statistical Society. Series B **43**, 34–40 (1981)
116. Taniguchi, M.: Robust regression and interpolation for time series. Journal of Time Series Analysis **2**(1), 53–62 (1981)
117. Taniguchi, M.: Minimum contrast estimation for spectral densities of stationary processes. Journal of the Royal Statistical Society. Series B **49**, 315–325 (1987)
118. Taniguchi, M., van Garderen, K.J., Puri, M.L.: Higher order asymptotic theory for minimum contrast estimators of spectral parameters of stationary processes. Econometric theory **19**(06), 984–1007 (2003)
119. Taniguchi, M., Hirukawa, J.: Generalized information criterion. Journal of Time Series Analysis **33**(2), 287–297 (2011)

120. Taniguchi, M., Kakizawa, Y.: Asymptotic theory of statistical inference for time series. New York: Springer-Verlag (2000)
121. van der Vaart, Wellner, J.A.: Weak Convergence. Springer (1996)
122. van der Vaart, A.W.: Asymptotic statistics, vol. 3. Cambridge university press (2000)
123. Walker, A.M.: On the estimation of a harmonic component in a time series with stationary independent residuals. Biometrika **58**(1), 21 (1971)
124. Whittle, P.: The simultaneous estimation of a time series harmonic components and covariance structure. Trabajos de estadística **3**(1-2), 43–57 (1952)
125. Whittle, P.: Some results in time series analysis. Scandinavian Actuarial Journal **1952**(1-2), 48–60 (1952)
126. Xenophontos, C.: A formula for the nth derivative of the quotient of two functions (2007)
127. Yoshihara, K.i.: Limiting behavior of u-statistics for stationary, absolutely regular processes. Probability Theory and Related Fields **35**(3), 237–252 (1976)
128. Zhang, G., Taniguchi, M.: Nonparametric approach for discriminant analysis in time series. Journaltitle of Nonparametric Statistics **5**(1), 91–101 (1995)