Geometry of Exponential Models

Salem Said

Nonlinear Signal Processing Lab
University of Melbourne

Differential and Information Geometry 2011

Outline

Introduction
- Motivation
- Definition and examples

Statistical properties
- Natural form
- Sampling

Geometry of MLE
- Example
- Duality and efficiency

Curvature and asymptotics
- Dually flat connections
- Asymptotics

Summary

bibliography

Outline

Introduction
- Motivation
- Definition and examples

Statistical properties
- Natural form
- Sampling

Geometry of MLE
- Example
- Duality and efficiency

Curvature and asymptotics
- Dually flat connections
- Asymptotics

Summary

bibliography

Why exponential models

- Analytically tractable in terms of a small number of quantities
- Statistical properties
  - Stable with respect to sampling
  - Efficient estimators
  - Conjugate priors $\rightsquigarrow$ Linear finite dimensional filters
- Include famous models

  Normal, Poisson, Binomia, Beta, Gamma, chi-2, etc
- Q.M.D : Any sufficiently smooth model is locally (curved) exponential

Definition

— $(X, \mathcal{B})$ sample space, $\Theta \subset \mathbb{R}^d$ parameter space,
— $\eta_k : \Theta \to \mathbb{R}$, $T_k : X \to \mathbb{R}$, $h : X \to \mathbb{R}_+$ (“locally integrable")

$$p_\theta(x) = \exp\left[\sum_{k=1}^{s} \eta_k(\theta) T_k(x) - B(\theta)\right] h(x) \qquad (1)$$

Density w.r.t common σ-finite μ

Binomial $X = \{1, \ldots, n\}$, $\Theta = [0, 1]$,

$$p_\theta = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$p_\theta(x) = \exp\left[x \log\left(\frac{\theta}{1-\theta}\right) - n \log\left(\frac{1}{1-\theta}\right)\right] \binom{n}{x}$$

Examples

Exponential tilt

—$X = \mathbb{R}^d$, $\mu = dP$ probability measure
— MGF: $\phi(\theta)$ defined over $\Theta \subset \mathbb{R}^d$

$$p_\theta(x) = \exp[(x, \theta) - \Lambda(\theta)] \qquad \Lambda(\theta) = \log[\phi(\theta)]$$

Θ is convex set and Λ a convex function

Exponential Fourier model

— $X = [0, 2\pi]$, $n \geq 1$ given, $\Theta = \mathbb{R}^{2n}$, s=2n
—$l = 1, \ldots, n$, $T_{2l}(x) = \sqrt{2}\cos(lx)$, $T_{2l-1}(x) = \sqrt{2}\sin(lx)$

$$p_\theta(x) = \exp\left[\sum_{k=1}^{2n} \theta_k T_k(x) - \psi(\theta)\right]$$

Used to approximate a target density in the KL-divergence

Examples

Exponential tilt

—$X = \mathbb{R}^d$, $\mu = dP$ probability measure

— MGF: $\phi(\theta)$ defined over $\Theta \subset \mathbb{R}^d$

$$p_\theta(x) = \exp[(x, \theta) - \Lambda(\theta)] \qquad \Lambda(\theta) = \log[\phi(\theta)]$$

Θ is convex set and Λ a convex function

Exponential Fourier model

— $X = [0, 2\pi]$, $n \geq 1$ given, $\Theta = \mathbb{R}^{2n}$, s=2n

—$l = 1, \ldots, n$, $T_{2l}(x) = \sqrt{2}\cos(lx)$, $T_{2l-1}(x) = \sqrt{2}\sin(lx)$

$$p_\theta(x) = \exp\left[\sum_{k=1}^{2n} \theta_k T_k(x) - \psi(\theta)\right]$$

Used to approximate a target density in the KL-divergence

Outline

Introduction
Motivation
Definition and examples
Statistical properties
Natural form
Sampling
Geometry of MLE
Example
Duality and efficiency
Curvature and asymptotics
Dually flat connections
Asymptotics
Summary
bibliography

Canonical form

Natural parameters A bigger model, parameters $\eta_1, \ldots, \eta_s$

$$p(x; \eta) = \exp\left[\sum_{k=1}^{s} \eta_k T_k(x) - \Lambda(\eta)\right] h(x) \tag{2}$$

Natural parameter space

$$\Xi = \left\{ \eta \in \mathbb{R}^s \,\middle|\, \int p(x;\eta) d\mu = 1 \right\}$$

Hölder inequality $\rightsquigarrow$ Ξ a convex set and Λ a convex function

- Full rank: Model identifiable and Ξ has nonepmty interior
- $\rightsquigarrow$ functions $1, T_1, \ldots, T_s$ are linearly independent

Sufficient statistics

Factorisation criterion (Fisher-Neyman)

$$dP_\theta(x) = f_\theta(T(x))g(x)d\mu(x) \Longrightarrow T \text{sufficient}$$

$(T_1, \ldots, T_s)$ is sufficient; if model has full rank, it is also minimal

Distribution of $(T_1, \ldots, T_s)$

$$p(t|\eta) = \exp\left[\sum_{k=1}^{s} \eta_k t_k - \Lambda(\eta)\right] \tag{3}$$

Density w.r.t image of $h.\mu$ under $(T_1, \ldots, T_s)$

It is possible to consider (3) instead of (2); (3) is called Natural form

Properties under sampling

Independent samples $X = (x_1, \ldots, x_n)$ from $p(x|\eta)$

$$p(X;\eta) = \exp\left[\sum_{k=1}^{s} \eta_k T'_k(X_i) - \Lambda'(\eta)\right] h'(X)$$

$T'_k(X) = \sum_{i=1}^n T_k(x_i)$; $\Lambda'(\eta) = n\Lambda(\eta)$; $h'(X) = \prod_{i=1}^n h(x_i)$

Density w.r.t $\mu \otimes \ldots \otimes \mu$ (n times)

Ptiman-Koopman-Darmois

$(p_\theta|\theta \in \Theta)$; $p_\theta(x)$ continuously differentiable in $x \in I$ real interval; if for $n \geq 1$

$$p_\theta(x_1) \ldots p_\theta(x_n)$$

has a k-dimensional sufficient statistic then p_θ is of the form (1) with $s \leq k$.

Outline

- Introduction
  - Motivation
  - Definition and examples
- Statistical properties
  - Natural form
  - Sampling
- Geometry of MLE
  - Example
  - Duality and efficiency
- Curvature and asymptotics
  - Dually flat connections
  - Asymptotics
- Summary
- bibliography

Exponential distribution

Average parameter (w.r.t. Lebesgue measure on $\mathbb{R}_+^*$)

$$p_\theta(x) = \frac{1}{\theta} \exp\left[\frac{-x}{\theta}\right] \qquad \theta > 0 \tag{4}$$

Natural parameter (w.r.t. Lebesgue measure on $\mathbb{R}_-^*$)

$$p(t|\eta) = \exp\left[\eta t + \log(\eta)\right] \qquad \eta > 0 \tag{5}$$

Identifiability

$$p \mapsto \eta(p) \quad ; \quad p \mapsto \theta(p) \quad ; \quad \eta(p) = \frac{1}{\theta(p)} \quad ; \quad \theta(p) = \frac{1}{\eta(p)}$$

MLE (observed point)

$$\theta^*(x) = x \quad ; \quad \eta^*(t) = -\frac{1}{t} \quad ; \quad \eta^*(x) = \frac{1}{\theta(x)} \quad ; \quad \theta^*(x) = \frac{1}{\eta(x)}$$

Exponential distribution

Average parameter (w.r.t. Lebesgue measure on $\mathbb{R}_+^*$)

$$p_\theta(x) = \frac{1}{\theta} \exp\left[\frac{-x}{\theta}\right] \qquad \theta > 0 \tag{4}$$

Natural parameter (w.r.t. Lebesgue measure on $\mathbb{R}_-^*$)

$$p(t|\eta) = \exp\left[\eta t + \log(\eta)\right] \qquad \eta > 0 \tag{5}$$

Identifiability

$$p \mapsto \eta(p) \quad ; \quad p \mapsto \theta(p) \quad ; \quad \eta(p) = \frac{1}{\theta(p)} \quad ; \quad \theta(p) = \frac{1}{\eta(p)}$$

MLE (observed point)

$$\theta^*(x) = x \quad ; \quad \eta^*(t) = -\frac{1}{t} \quad ; \quad \eta^*(x) = \frac{1}{\theta(x)} \quad ; \quad \theta^*(x) = \frac{1}{\eta(x)}$$

Exponential distribution

Average parameter (w.r.t. Lebesgue measure on $\mathbb{R}_+^*$)

$$p_\theta(x) = \frac{1}{\theta} \exp\left[\frac{-x}{\theta}\right] \qquad \theta > 0 \qquad (4)$$

Natural parameter (w.r.t. Lebesgue measure on $\mathbb{R}_-^*$)

$$p(t|\eta) = \exp\left[\eta t + \log(\eta)\right] \qquad \eta > 0 \qquad (5)$$

Identifiability

$$p \mapsto \eta(p) \;\; ; \;\; p \mapsto \theta(p) \;\; ; \;\; \eta(p) = \frac{1}{\theta(p)} \;\; ; \;\; \theta(p) = \frac{1}{\eta(p)}$$

MLE (observed point)

$$\theta^*(x) = x \;\; ; \;\; \eta^*(t) = -\frac{1}{t} \;\; ; \;\; \eta^*(x) = \frac{1}{\theta(x)} \;\; ; \;\; \theta^*(x) = \frac{1}{\eta(x)}$$

Exponential distribution

Average parameter (w.r.t. Lebesgue measure on $\mathbb{R}_+^*$)

$$p_\theta(x) = \frac{1}{\theta} \exp\left[\frac{-x}{\theta}\right] \qquad \theta > 0 \tag{4}$$

Natural parameter (w.r.t. Lebesgue measure on $\mathbb{R}_-^*$)

$$p(t|\eta) = \exp\left[\eta t + \log(\eta)\right] \qquad \eta > 0 \tag{5}$$

Identifiability

$$p \mapsto \eta(p) \;\; ; \;\; p \mapsto \theta(p) \;\; ; \;\; \eta(p) = \frac{1}{\theta(p)} \;\; ; \;\; \theta(p) = \frac{1}{\eta(p)}$$

MLE (observed point)

$$\theta^*(x) = x \;\; ; \;\; \eta^*(t) = -\frac{1}{t} \;\; ; \;\; \eta^*(x) = \frac{1}{\theta(x)} \;\; ; \;\; \theta^*(x) = \frac{1}{\eta(x)}$$

Exponential distribution

Fisher information

$$I(\eta) = \frac{1}{\eta^2} \qquad I(\theta) = \frac{1}{\theta^2}$$

$$I(\theta(p)) = \frac{1}{I(\eta(p))}$$

Efficiency $\theta = \eta^{-1}$, $\psi(\eta) = -\log(\eta)$

$$\mathbb{E}_\theta(\theta^* - \theta)^2 = \psi''(\eta) \qquad I(\eta) = \psi''(\eta)$$

	θ^*	η^*
Unbiased	yes	no
Efficient	yes	no

Exponential distribution

Fisher information

$$I(\eta) = \frac{1}{\eta^2} \qquad I(\theta) = \frac{1}{\theta^2}$$

$$I(\theta(p)) = \frac{1}{I(\eta(p))}$$

Efficiency $\theta = \eta^{-1}$, $\psi(\eta) = -\log(\eta)$

$$\mathbb{E}_\theta(\theta^* - \theta)^2 = \psi''(\eta) \qquad I(\eta) = \psi''(\eta)$$

	θ^*	η^*
Unbiased	yes	no
Efficient	yes	no

Exponential distribution

Fisher information

$$I(\eta) = \frac{1}{\eta^2} \qquad I(\theta) = \frac{1}{\theta^2}$$

$$I(\theta(p)) = \frac{1}{I(\eta(p))}$$

Efficiency $\theta = \eta^{-1}$, $\psi(\eta) = -\log(\eta)$

$$\mathbb{E}_\theta(\theta^* - \theta)^2 = \psi''(\eta) \qquad I(\eta) = \psi''(\eta)$$

	θ^*	η^*
Unbiased	yes	no
Efficient	yes	no

Dual coordinates

- ψ is convex, C^2. Assume ψ strictly convex
- Assume model has full rank ((2) and (3))

Legendre transform $\theta, \eta \in \mathbb{R}^s$

$$\phi(\theta) = \max_{\eta}\{\theta\eta - \psi(\eta)\} \qquad \psi(\eta) = \max_{\eta}\{\eta\theta - \phi(\theta)\} \tag{6}$$

Average parameter $\nabla\phi$ and $\nabla\psi$ are inverse functions

$$\eta(\theta) = \nabla\phi(\theta) \qquad \theta(\eta) = \nabla\psi(\eta)$$

Entropy $(\partial_k\psi(\eta) = \mathbb{E}_\eta[t_k])$

$$\phi(\theta(\eta)) = \mathbb{E}_\eta\left[\sum_{k=1}^{s}\eta_k t_k - \psi(\eta)\right] \tag{7}$$

Efficiency

Fisher information

$$g_{kl}(\eta) = \partial_k \partial_l \psi(\eta) \qquad g^{kl}(\theta) = \partial_k \partial_l \phi(\theta) \tag{8}$$

$$g_{kl}(\eta) = \partial_k \theta_l(\eta) \qquad g^{kl}(\theta) = \partial_k \eta_l(\theta) \rightsquigarrow \text{Fisher information}$$

MLE (observed point)

$$\theta^*(x) = t(x) \qquad \nabla\psi(\eta^*(x)) = \theta^*(x) \tag{9}$$

Efficiency θ^* is efficient

$$\mathbb{E}_\theta[(\theta^*_k - \theta_k)(\theta^*_l - \theta_l)] = \partial_k \partial_l \psi(\eta(\theta)) = g_{kl}(\eta(\theta))$$

This relation is affine invariant

Efficiency

Fisher information

$$g_{kl}(\eta) = \partial_k \partial_l \psi(\eta) \qquad g^{kl}(\theta) = \partial_k \partial_l \phi(\theta) \tag{8}$$

$$g_{kl}(\eta) = \partial_k \theta_l(\eta) \qquad g^{kl}(\theta) = \partial_k \eta_l(\theta) \rightsquigarrow \text{Fisher information}$$

MLE (observed point)

$$\theta^*(x) = t(x) \qquad \nabla\psi(\eta^*(x)) = \theta^*(x) \tag{9}$$

Efficiency θ^* is efficient

$$\mathbb{E}_\theta[(\theta^*_k - \theta_k)(\theta^*_l - \theta_l)] = \partial_k \partial_l \psi(\eta(\theta)) = g_{kl}(\eta(\theta))$$

This relation is affine invariant

Efficiency

Fisher information

$$g_{kl}(\eta) = \partial_k \partial_l \psi(\eta) \qquad g^{kl}(\theta) = \partial_k \partial_l \phi(\theta) \tag{8}$$

$$g_{kl}(\eta) = \partial_k \theta_l(\eta) \qquad g^{kl}(\theta) = \partial_k \eta_l(\theta) \rightsquigarrow \text{Fisher information}$$

MLE (observed point)

$$\theta^*(x) = t(x) \qquad \nabla\psi(\eta^*(x)) = \theta^*(x) \tag{9}$$

Efficiency θ^* is efficient

$$\mathbb{E}_\theta[(\theta^*_k - \theta_k)(\theta^*_l - \theta_l)] = \partial_k \partial_l \psi(\eta(\theta)) = g_{kl}(\eta(\theta))$$

This relation is affine invariant

Efficiency

Fisher information

$$g_{kl}(\eta) = \partial_k \partial_l \psi(\eta) \qquad g^{kl}(\theta) = \partial_k \partial_l \phi(\theta) \tag{8}$$

$$g_{kl}(\eta) = \partial_k \theta_l(\eta) \qquad g^{kl}(\theta) = \partial_k \eta_l(\theta) \rightsquigarrow \text{ Fisher information}$$

MLE (observed point)

$$\theta^*(x) = t(x) \qquad \nabla\psi(\eta^*(x)) = \theta^*(x) \tag{9}$$

Efficiency θ^* is efficient

$$\mathbb{E}_\theta[(\theta^*_k - \theta_k)(\theta^*_l - \theta_l)] = \partial_k \partial_l \psi(\eta(\theta)) = g_{kl}(\eta(\theta))$$

This relation is affine invariant

Canonical divergence

Kullback-Liebler divergence (θ, η unrelated)

$$D(\theta|\eta) = \psi(\eta) + \phi(\theta) - \eta\theta \tag{10}$$

MLE (θ^* observed point)

$$\eta^*\theta^* - \psi(\eta^*) = \max_\eta \{\eta\theta^* - \psi(\eta)\} \Leftrightarrow D(\theta^*|\eta^*) = \min_\eta D(\theta^*|\eta)$$

Divergence

- $D(\theta|\eta) \geq 0$
- $D(\theta|\eta) = 0$ *iff* $\theta = \theta(\eta)$

Outline

Introduction
- Motivation
- Definition and examples

Statistical properties
- Natural form
- Sampling

Geometry of MLE
- Example
- Duality and efficiency

Curvature and asymptotics
- Dually flat connections
- Asymptotics

Summary

bibliography

Triangle relation

MLE to be understood as a "projection"

Triangle relation

$$D(p|q) + D(q|r) - D(p|r) = (\eta(p) - \eta(q))(\theta(r) - \theta(q)) \quad (11)$$

Curved model

$$M = \{p(t|\eta(u))|u \in O\} \qquad O \subset \mathbb{R}^{s'} \text{ open} \quad (12)$$

(s, s')-model $(s' \leq s)$

MLE

$$D(\theta^*|\eta(u^*)) = \min_u D(\theta^*|\eta(u)) \quad (13)$$

Triangle relation

MLE to be understood as a "projection"

Triangle relation

$$D(p|q) + D(q|r) - D(p|r) = (\eta(p) - \eta(q))(\theta(r) - \theta(q)) \quad (11)$$

Curved model

$$M = \{p(t|\eta(u))|u \in O\} \qquad O \subset \mathbb{R}^{s'} \text{ open} \quad (12)$$

(s, s')-model ($s' \leq s$)

MLE

$$D(\theta^*|\eta(u^*)) = \min_u D(\theta^*|\eta(u)) \quad (13)$$

Triangle relation

MLE to be understood as a "projection"

Triangle relation

$$D(p|q) + D(q|r) - D(p|r) = (\eta(p) - \eta(q))(\theta(r) - \theta(q)) \qquad (11)$$

Curved model

$$M = \{p(t|\eta(u))|u \in O\} \qquad O \subset \mathbb{R}^{s'} \text{ open} \qquad (12)$$

(s, s')-model $(s' \leq s)$

MLE

$$D(\theta^*|\eta(u^*)) = \min_u D(\theta^*|\eta(u)) \qquad (13)$$

Triangle relation

MLE to be understood as a "projection"

Triangle relation

$$D(p|q) + D(q|r) - D(p|r) = (\eta(p) - \eta(q))(\theta(r) - \theta(q)) \quad (11)$$

Curved model

$$M = \{p(t|\eta(u))|u \in O\} \qquad O \subset \mathbb{R}^{s'} \text{ open} \quad (12)$$

(s, s')-model $(s' \leq s)$

MLE

$$D(\theta^*|\eta(u^*)) = \min_u D(\theta^*|\eta(u)) \quad (13)$$

e- and m-projections

η e-coordinate θ m-coordinate

Dual connections

$$\Gamma^{e}_{ijk}(\eta) = 0 \qquad \Gamma^{e}_{ijk}(\theta) = \partial_i \partial_j \partial_k \phi(\theta)$$

$$\Gamma^{m}_{ijk}(\theta) = 0 \qquad \Gamma^{m}_{ijk}(\eta) = \partial_i \partial_j \partial_k \psi(\eta)$$

m projection

u^* (when it exists) is the m projection of θ^* on M

For $a = 1, \ldots, s'$

$$\partial_a D(\theta^* | \eta(u^*)) = \sum_k \partial_a \eta_k(u^*)(\theta_k(u^*) - \theta^*_k) = 0 \qquad (14)$$

m-geodesic from θ^* orthogonal to M at u

e- and m-projections

η e-coordinate $\qquad$ θ m-coordinate

Dual connections

$$\Gamma^{e}_{ijk}(\eta) = 0 \qquad \Gamma^{e}_{ijk}(\theta) = \partial_i \partial_j \partial_k \phi(\theta)$$

$$\Gamma^{m}_{ijk}(\theta) = 0 \qquad \Gamma^{m}_{ijk}(\eta) = \partial_i \partial_j \partial_k \psi(\eta)$$

m projection

u^* (when it exists) is the m projection of θ^* on M

For $a = 1, \ldots, s'$

$$\partial_a D(\theta^* | \eta(u^*)) = \sum_k \partial_a \eta_k(u^*)(\theta_k(u^*) - \theta^*_k) = 0 \qquad (14)$$

m-geodesic from θ^* orthogonal to M at u

e- and m-projections

η e-coordinate $\quad\quad$ θ m-coordinate

Dual connections

$$\Gamma^{e}_{ijk}(\eta) = 0 \qquad \Gamma^{e}_{ijk}(\theta) = \partial_i\partial_j\partial_k\phi(\theta)$$

$$\Gamma^{m}_{ijk}(\theta) = 0 \qquad \Gamma^{m}_{ijk}(\eta) = \partial_i\partial_j\partial_k\psi(\eta)$$

m projection

u^* (when it exists) is the m projection of θ^* on M

For $a = 1, \ldots, s'$

$$\partial_a D(\theta^* | \eta(u^*)) = \sum_k \partial_a \eta_k(u^*)(\theta_k(u^*) - \theta^*_k) = 0 \qquad (14)$$

m-geodesic from θ^* orthogonal to M at u

e- and m-projections

η e-coordinate $\qquad$ θ m-coordinate

Dual connections

$$\Gamma^{e}_{ijk}(\eta) = 0 \qquad \Gamma^{e}_{ijk}(\theta) = \partial_i \partial_j \partial_k \phi(\theta)$$

$$\Gamma^{m}_{ijk}(\theta) = 0 \qquad \Gamma^{m}_{ijk}(\eta) = \partial_i \partial_j \partial_k \psi(\eta)$$

m projection

u^* (when it exists) is the m projection of θ^* on M

For $a = 1, \dots, s'$

$$\partial_a D(\theta^* | \eta(u^*)) = \sum_k \partial_a \eta_k(u^*)(\theta_k(u^*) - \theta^*_k) = 0 \qquad (14)$$

m-geodesic from θ^* orthogonal to M at u

Asymptotics

e-flat model

If M is e-flat, it is an exponential model $\Rightarrow u^*$ efficient

Product model (i.i.d observations)

$$p(T|\eta) = \exp\left[n\left(\sum_{k=1}^{s}\eta_k \bar{t} - \psi(\eta)\right)\right] \tag{15}$$

$$T = (t_1, \ldots, t_n) \qquad \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

General model (even not exponential)

- MLE is consistent
- MLE is assymptotically normal
- MLE is assymptotically efficient

Asymptotics

CLT

$$\theta^* = \bar{t} \qquad \sqrt{n}(\theta^* - \theta(u)) \xrightarrow{d} N(0, g_{ij}(\eta(u)))$$

Estimating submanifold (for MLE)

$$A(u) = \{\theta \mid \partial_a D(\theta^* | \eta(u^*)) = 0\} \tag{16}$$

$A(u)$ is m-flat and orthogonal to M at $u = A(u) \cap M$

Second order efficiency

- M is e-curved $\rightsquigarrow$ error of order n^{-2}
- Choice of estimator given by a choice of $A(u)$
- m-flat $A(u)$ characterises MLE and is optimal

Outline

Introduction
- Motivation
- Definition and examples

Statistical properties
- Natural form
- Sampling

Geometry of MLE
- Example
- Duality and efficiency

Curvature and asymptotics
- Dually flat connections
- Asymptotics

Summary

bibliography

Summary

- Exponential families allow efficient estimation
- MLE described by Legendre pair
- Legendre pair $\Leftrightarrow$ Dual connections
- For curved families, MLE is given by m pojection
- Curved family is exponential *iff* e flat
- MLE is asymptotically efficient
- e cuvature determines second order efficiency
- e projection used with partial observation
- The EM algorithm can be understood as “em" algorithm

Outline

Introduction
- Motivation
- Definition and examples

Statistical properties
- Natural form
- Sampling

Geometry of MLE
- Example
- Duality and efficiency

Curvature and asymptotics
- Dually flat connections
- Asymptotics

Summary

bibliography

Amari & Nagaoka.
Methods of Information Geometry.
AMS, 2000.

Barndoff-Nielsen.
Information and Exponential Families in Statistical Theory.
Wiley, 1978.

Küschler & Sørensen.
Exponential Families of Stochastic Processes.
Springer, 1997.

Le Cam & Yang.
Asymptotics in Statistics, Some Basic Concepts.
Springer, 2000.