Parameter Estimation and Reverse Martingales

Tomas Björk

Department of Finance

Stockholm School of Economics

Box 6501

S-113 83 Stockholm SWEDEN

Björn Johansson

Mathematical Statistics

Stockholm University

S-106 91 Stockholm SWEDEN

Published in

Stochastic Processes and their Applications,

Vol. **63**, 235-263. (1996)

Abstract

Within the framework of transitive sufficient processes we investigate identifiability properties of unknown parameters. In particular we consider unbiased parameter estimators, which are shown to be closely connected to time reversal and to reverse martingales. One of the main results is that, within our framework, every unbiased estimator process is a reverse martingale, thus automatically giving us strong consistency results. We also study structural properties of unbiased estimators, and it is shown that the existence of an unbiased parameter estimator is equivalent to the existence of a solution to an inverse boundary value problem. We give explicit representation formulas for the estimators in terms of Feynman-Kac type representations using complex valued diffusions, and we also give Cramér-Rao bounds for the estimation error.

Keywords: Parameter estimation, reverse martingales, martingale theory, diffusions, time reversal.

1 Basic Definitions

We consider a statistical model, i.e. a family Π of probability measures on a measurable space $(\Omega, \mathcal{F})$. Given is also a k-dimensional stochastic process $X = \{X(t);\ t \geq 0\}$ on $(\Omega, \mathcal{F})$, and for any subset I of R_+, we denote by $\mathcal{F}_I^X$ the σ-field generated by the random variables $\{X(t);\ t \in I\}$. We assume that the measures in Π are equivalent on $\mathcal{F}_t = \mathcal{F}_{[0,t]}^X$ for each $t > 0$. The intuitive interpretation of this model is that the stochastic process X is governed by some measure $P \in \Pi$, but we do not know exactly which P. The family Π formalizes the a priori information available to us, and to obtain further information we are allowed to observe the process X over time. Thus the information accesssible to us through observations is represented by the filtration $\underline{\mathcal{F}} = \{\mathcal{F}_t\}_{t \geq 0}$. A number of concrete examples will be given below.

We will study various estimation problems in connection with the model above, and in particular we will be concerned with unbiased parameter estimation problems and their relations to the theory of reverse martingales. Our work is to a large extent inspired by Lauritzen [Laur 88]. The results in the present paper are also closely related to the theory of adaptive prediction developed by the authors in an earlier paper [Bjö & Joh, 92].

We begin by giving precise definitions of the concepts of *parameter*, *unbiased estimator*, *sufficiency*, *transitivity*, and *completeness*.

Definition 1.1 *Consider the model* $(\Omega, \mathcal{F}, \Pi, X, \underline{\mathcal{F}})$.

1. *We endow the family* Π *with the* σ*-algebra* Σ_π *defined as*

$$\Sigma_\pi = \sigma\{F_A; A \in \mathcal{F}_\infty\},$$

where, for each $A \in \mathcal{F}_\infty$ *the mapping* $F_A : \Pi \to \mathbb{R}$ *is defined by*

$$F_A(P) = P(A).$$

2. *A* **parameter** Φ *is a measurable mapping*

$$\Phi : \Pi \to R$$

.

The value of $\Phi(P)$ is to be interpreted as "an aspect of the measure P", and a parameter is thus something very much like a local coordinate on a manifold. We could of course also consider parameters taking values in more general spaces, but for us R will do nicely.

Our main goal in the sequel will be that of estimating parameter values, based on observations of the process X.

Definition 1.2 *Consider a fixed parameter* Φ.

1. *A* t-**estimator** *is any* $\mathcal{F}_t$*-measurable random variable* V*.*
2. *A* t*-estimator* V *is said to be an* **unbiased** t*-estimator of* Φ *if the following conditions are satisfied*

$$V \in L^1(\Omega, \mathcal{F}_t, P), \text{ for all } P \in \Pi.$$

$$E_P[V] = \Phi(P), \text{ for all } P \in \Pi.$$

3. *An optional process* Y *is called an* **unbiased estimator process** *of* Φ *on the time interval* $[T, \infty)$ *if* $Y(t)$ *is unbiased in the sense of 2 above for every* $t \geq T$*.*
4. *An optional process* Y *is said to be a* **consistent** *estimator of* Φ *if*

$$\lim_{t \to \infty} Y(t) = \Phi(P), \ P - a.s. \text{ for all } P \in \Pi.$$

5. *A parameter is said to be* **identifiable** *if there exists an* $\mathcal{F}_\infty$*-measurable stochastic variable* V*, satisfying*

$$V = \Phi(P), \ P - a.s. \ for \ all \ P \in \Pi.$$

It is obvious that if a parameter Φ can be consistently estimated, then Φ is identifiable. It is however an open question whether identifiability of a parameter implies the existence of a consitent estimator. We believe that this is not the case (see Section 7).

The main questions which we will try to answer in this paper are the following.

- Which parameters have unbiased estimators?
- Which parameters can be consistently estimated?
- Which parameters can be identified?

In the sequel we will work within a framework of so called transitive sufficient processes, for which we now give the definitions.

Definition 1.3

1. *An* $\mathcal{F}_t$*-measurable random* k*-vector* V *is said to be* **sufficient** *for* Π *restricted to* $\mathcal{F}_t$*, if for each bounded random variable* U *in* $\mathcal{F}_t$ *there exists a Borel function* $f : R^k \to R$ *such that*

$$E_P[U|V] = f(V), \quad \text{for all } P \in \Pi.$$

2. *An* $\underline{\mathcal{F}}$*-optional process* Z *is called* **sufficient for** Π *if* $Z(t)$ *is sufficient for* Π *restricted to* $\mathcal{F}_t$*, for each* $t \geq 0$*.*

We will often use the well known fact that a random vector V is sufficient for Π restricted to $\mathcal{F}_t$ if, for some $P_0 \in \Pi$ the Radon-Nikodym derivative

$$\left.\frac{dP}{dP_0}\right| \mathcal{F}_t$$

s is $\sigma\{V\}$-measurable for each $P \in \Pi$.

Definition 1.4 *An $\underline{\mathcal{F}}$-optional process Z is called* **transitive** *if it is a $(P, \underline{\mathcal{F}})$-Markov process for each $P \in \Pi$, i.e. if the σ-fields $\mathcal{F}_t$ and $\mathcal{F}^Z_{[t,\infty)}$ are P-conditionally independent given $Z(t)$.*

The notion of a transitive sequence of statistics was introduced by Bahadur [Baha, 54]. Definition 1.4 is equivalent to stating that for each $P \in \Pi$, each $t \geq 0$ and and each bounded random variable U in $\mathcal{F}_t$, we have

$$E_P\left[U \,|Z(t)\,\right] = E_P\left[U \left| \mathcal{F}^Z_{[t,\infty)}\right.\right].$$

Definition 1.5 *An $\underline{\mathcal{F}}$-optional process X is called* **complete** *if, for every t and every Borel function g, the condition*

$$E_P\left[g(X(t))\right] = 0, \quad \textit{for all } P \in \Pi,$$

implies

$$g(X(t)) = 0, \quad \Pi - a.s.$$

To construct a transitive sufficient process in a Markovian case we will typically use the following "algorithm".

1. For every fixed t we apply the Girsanov Theorem to find a sufficient statistic for Π restricted to $\mathcal{F}_.$
2. If no finite-dimensional statistic exists then we are stuck and cannot apply the theory below.
3. If there exists a finite dimensional sufficient process Y, then Y itself is not necessarily transitive (it need not be Markovian). In most cases, however, the extended process (X, Y) will be transitive and sufficient.
4. Now we regard (X, V) as our basic process instead of X.

For a discussion of some concrete examples of this technique see [Bjö & Joh, 92] p.194.

To illustrate technique and ideas we will use three simple concrete scalar models. In all cases Ω is the space $C[0, \infty)$ and X is the coordinate process on Ω.

The Wiener model: For this model the a priori family $\Pi = \Pi_W$ is defined by $\Pi_W = \{P_\alpha; \alpha \in R\}$, where X under P_α has the dynamics

$$dX(t) = \alpha dt + dW(t), \ X(0) = 0,$$

and where W is a standard Wiener process.

The L^2-model: Here we let X have the representation

$$dX(t) = Z^* dt + dW(t), \ X(0) = 0, \tag{1}$$

where W is a Wiener process. The a priori family, denoted by Π_L, is now defined as the class of probability measures such that

- Z^* and W are independent.
- Z^* has finite second moment.

We may thus identify Π_L by the following family of distributions F (of Z^*) on the real line

$$\Pi_L = \left\{ F; \int_R z^2 dF(z) < \infty \right\}.$$

Observe that (1) is the semimartingale representation for X with respect to the filtration $\underline{\mathcal{G}}$, defined by

$$\mathcal{G}_t = \mathcal{F}_t \vee \sigma\{Z^*\}; \ t \geq 0. \tag{2}$$

It is perhaps not clear from (1) that X is a Markov process relative to the $\mathcal{F}^X$-filtration, but it is fairly easy to show that the semimartingale representation of X with respect to the $\mathcal{F}^X$-filtration is given by:

$$dX(t) = \mu(t, X(t))dt + dW(t)$$

where μ is given by

$$\mu(t, X(t)) = \frac{\int_R z \cdot \exp(zx - \frac{1}{2}z^2 t) dF(z)}{\int_R \exp(zx - \frac{1}{2}z^2 t) dF(z)}, \tag{3}$$

and F as above is the distribution function of Z^*.

The Gaussian mixture model: Here we consider a parameterized a priori family Π_G given by

$$\Pi_G = \{P_{\alpha,\beta}; \ \alpha \in \mathbb{R}, \ \beta \geq 0\},$$

where X under $P_{\alpha,\beta}$ has the dynamics

$$dX(t) = \frac{\alpha + \beta X(t)}{1 + \beta t} dt + dW(t).$$

Typical examples of parameters Φ one may want to estimate in the models above are the following.

The Wiener model:

$$\begin{aligned}\Phi(P_\alpha) &= \alpha \\ \Phi(P_\alpha) &= g(\alpha),\end{aligned}$$

where g is some given function.

The L^2-model:

$$\begin{aligned}\Phi(P) &= E_P\left[Z^*\right], \\ \Phi(P) &= Var_P[Z^*], \\ \Phi(P) &= E_P\left[(Z^*)^2\right],\end{aligned}$$

The Gaussian mixture model:

$$\begin{aligned}\Phi(P_{\alpha,\beta}) &= \alpha, \\ \Phi(P_{\alpha,\beta}) &= \beta, \\ \Phi(P_{\alpha,\beta}) &= g(\alpha,\beta),\end{aligned}$$

where g is some given function.

2 Unbiased Estimators and Reverse Martingales

The central theme of this paper is the connection between reverse martingale theory and unbiased parameter estimation. We now present a series of results which highlights this theme, and the results may be summarized by saying that in a transitive sufficient and complete model, unbiased estimators are *characterized* by being reverse martingales. The results are all extremely easy to prove, and they immediately give us very powerful tools for studying the asymptotic behaviour of unbiased estimators.

Proposition 2.1 *Consider a fixed parameter Φ in the model $(\Omega, \mathcal{F}, \Pi, \underline{\mathcal{F}})$. Assume that there exists an unbiased estimator process Y of Φ having the form*

$$Y(t) = f(t, X(t)),$$

for some $\underline{\mathcal{F}}$-optional process X which is sufficient, transitive and complete. Then Y is an $\underline{\mathcal{T}}^X$ - reverse martingale for all $P \in \Pi$, where the filtration $\underline{\mathcal{T}}^X$ is defined by

$$\mathcal{T}_t^X = \mathcal{F}_{[t,\infty)}^X, \quad t \geq 0.$$

Proof. Let $0 \le s \le t$. Since Y is an unbiased estimator process we have, for all $P \in \Pi$,

$$E_P[Y(t)] = E_P[Y(s)] = E_P[E_P[Y(s)\,|X(t)]].$$

Since X is sufficient, $E_P[Y(s)\,|X(t)]$ does not depend on the choice of P, and so, by the completeness of X,

$$Y(t) = E_P[Y(s)\,|X(t)]. \tag{4}$$

Using the transitivity of X we have

$$E_P[Y(s)\,|X(t)] = E_P\left[Y(s)\,\middle|\mathcal{T}_t^X\right],$$

which, together with (4) gives us the martingale property

$$Y(t) = E_P\left[Y(s)\,\middle|\mathcal{T}_t^X\right]. \quad \blacksquare$$

To assume that $Y(t) = f(t, X(t))$ is, by the Rao-Blackwell Theorem, not a severe restriction. In the sequel we will thus always assume that all unbiased estimators are of the form 2.1.

Proposition 2.2 *Fix a measure $P_0 \in \Pi$ and assume that X and Y are two $\underline{\mathcal{F}}$-optional processes such that*

1. *X is sufficient for Π.*
2. *X is a $(P_0, \underline{\mathcal{F}})$-Markov process.*
3. *Y is a $\left(P_0, \underline{\mathcal{T}}^X\right)$-reverse martingale on $[T, \infty)$ for some $T \ge 0$.*
4. *$Y(t) \in L^1(\Omega, \mathcal{F}_t, P)$ for all $t \ge T$, $P \in \Pi$.*

Now define the mapping $\Phi : \Pi \to R$ by

$$\Phi(P) = E_P[Y(T)]$$

Then Y is an unbiased estimator of Φ on $[T, \infty)$.

Proof. By the sufficiency and transitivity of X we have, for $t \in [T, \infty)$,

$$\begin{aligned}
\Phi(P) &= E_P[Y(T)] = E_P[E_P[Y(T)\,|X(t)]] = && (5)\\
&= E_P[E_{P_0}[Y(T)\,|X(t)]] = E_P\left[E_{P_0}\left[Y(T)\,\middle|\mathcal{T}_t^X\right]\right] = && (6)\\
&= E_P[Y(t)] \quad \blacksquare && (7)
\end{aligned}$$

Proposition 2.1 permits us to draw very strong conclusions concerning the consistency of unbiased estimators.

Proposition 2.3 *Assume that the process X is sufficient, transitive and complete. Then the following hold.*

1. *Every unbiased estimator process Y of the form $Y(t) = f(t, X(t))$ converges Π-a.s. to some limiting stochastic variable $Y(\infty)$ as $t \to \infty$.*

2. *If the tail-σ-field $\mathcal{T}_\infty^X = \bigcap_{t \geq 0} \mathcal{F}_{[t,\infty)}^X$ is P-trivial for every $P \in \Pi$ then every unbiased estimator is consistent, i.e.*

$$\lim_{t \to \infty} Y(t) = \Phi(P), \quad P - a.s \;\; \textit{for all } P \in \Pi.$$

Proof. Fix any $P \in \Pi$. By Proposition 2.1 every unbiased estimator process is a reverse martingale, so the first part of the proposition follows immediately from the reverse martingale convergence theorem. Thus the limiting variable $Y(\infty)$ always exists, and we have $Y(\infty) \in \mathcal{T}_\infty^X$. If the tail σ-algebra is trivial then $Y(\infty)$ must be a deterministic constant, and thus $Y(\infty)$ has to equal its expected value $\Phi(P)$ P-a.s. ∎

The main point of the results above can thus be paraphrazed as follows: In a complete transitive sufficient model unbiasedness *implies* the reverse martingale property, thus guaranteeing consistency for the case when the tail field $\mathcal{T}_\infty^X$ is trivial.

The triviality of the tail-σ-field of a sufficient transitive process has been studied in general by Lauritzen [Laur 88]. To prove triviality in a concrete case is typically a very hard problem, but once such a result is established, the question of consistency of any unbiased estimator is thereby completely resolved.

To take a simple example consider the Wiener model of section 1. It is easily seen that X is transitive and sufficient, and it is well known that the model is complete. For a proof of the triviality of the tail σ-algebra see e.g. [Bjö & Joh, 93]. Now consider the parameter $\Phi(P_\alpha) = \alpha$. Then it is of course trivial to check that

$$Y_t = \frac{X_t}{t} \tag{8}$$

is an unbiased estimator of Φ, and we see from Proposition 2.3 above that Y converges P_α-a.s. to α. The reverse martingale property of Y in equation (8) is of course well-known but our point here is that Y is a reverse martingale *because* it is unbiased.

Proposition 2.2 suggests that you may find unbiased estimators by looking for reverse martingales. This may not seem to be a very promising approach, but later we will give an important example.

3 The Mixing Theorem and Sufficient Generating Martingales

Throughout this section we assume that X is transitive and sufficient.

The main technical tool in the sequel is the so called Mixing Theorem, which we now proceed to describe. We will then need to consider the various probability measures which are generated by "pinning" $X(t)$ at a fixed point x.

Definition 3.1 *For any probability measure P, not necessarily in Π, on $(\Omega, \mathcal{F}_\infty)$ we define $P^{t,x}$ on $(\Omega, \mathcal{F}_t)$ by*

$$P^{t,x}(A) = P\left(A \,|X(t) = x\right).$$

By transitive sufficiency we note that there exists a fixed family of probability measures $Q_{t,x}$ on $(\Omega, \mathcal{F}_t)$ such that, for every $P \in \Pi$ and every $(t, x) \in R_+ \times R^k$, we have $Q_{t,x} = P^{t,x}$.

Definition 3.2 *The maximal family, $\mathcal{M}$, generated by Π is the class of all probability measures P on $(\Omega, \mathcal{F}_\infty)$ with the following properties.*

1. *For every $(t, x) \in R_+ \times R^k$ we have $P^{t,x} = Q_{t,x}$,*
2. *For each t the σ-algebras $\sigma\{X(s); s \leq t\}$ and $\sigma\{X(s); s \geq t\}$ are conditionally P-independent given $X(t)$.*

The maximal family $\mathcal{M}$ is easily seen to be a convex set, and we denote the set of its extremal points by $\mathcal{E}$. It can be shown ([Laur 88] p.196, Proposition IV.I.I) that $\mathcal{M}$ is in fact an infinite dimensional simplex in the sense that any point in $\mathcal{M}$ can be written as a unique convex combination of points in $\mathcal{E}$. This fact, henceforth called the *mixing theorem*, is one of the cornerstones in the sequel.

Theorem 3.1 *For each $P \in \Pi$ there exists a unique probability measure ν_P on $\mathcal{E}$ such that*

$$P = \int_{\mathcal{E}} Q \nu_P(dQ),$$

in the sense that, for each $A \in \mathcal{F}_\infty$ we have

$$P(A) = \int_{\mathcal{E}} Q(A) \nu_P(dQ),$$

The measure ν_P is called the mixing measure corresponding to P.

Many of the results below will be connected to the tail-σ-algebra $\mathcal{T}^X_\infty$, and $\mathcal{T}^X_\infty$ is connected to $\mathcal{E}$ by the following result (see [Laur 88]).

Proposition 3.1 *For any $P \in \mathcal{M}$ the following are equivalent*

1. $P \in \mathcal{E}$
2. $\mathcal{T}_\infty^X$ *is* P*-trivial.*

In order to obtain concrete formulas below we will sometimes need more than the rather abstract mixing theorem above. The situation turns out to be particularly nice when there exists a finite dimensional stochastic vector Z which acts as a sufficient statistic at $t = \infty$. This is in fact often the case, and by studying concrete examples we have been led to the following definition.

Definition 3.3 *A* k*-dimensional optional process* Z *is said to be a sufficient generating martingale (SGM) if the following conditions hold.*

1. $\sigma\{X_t\} = \sigma\{Z_t\}$, *for all* $t \geq 0$.
2. Z *is a reverse martingale for all* $P \in \mathcal{E}$.
3. $\sigma\{Z_\infty\} = \mathcal{T}_\infty^X, P$ *-a.s. for all* $P \in \mathcal{M}$

We note that since Z is a reverse martingale for the extremal family it will converge $\mathcal{E}$-almost surely to a limiting variable Z_∞. By the mixing theorem this convergence will also take place $\mathcal{M}$-almost surely, so Z_∞ in (3) above is indeed well defined. An SGM can be viewed as a "normalized" version of the observation process X, and the main advantage of working with an SGM is seen by the following result.

Proposition 3.2 *Suppose that the model admits a sufficient generating martingale* Z*. For each* $z \in R^k$ *and any* $P \in \mathcal{M}$*, let the measure* P_z *be defined by*

$$P_z(A) = P(A \,|\, Z_\infty = z). \tag{9}$$

Then, by sufficiency, P_z *is independent of the particular choice of* P *and we have*

$$\mathcal{E} = \{P_z; z \in R^k\} \tag{10}$$

Proof. We obviously have, for each $A \in \mathcal{F}_\infty$ and each $P \in \mathcal{M}$,

$$P(A) = \int_{R^k} P_z(A) dF_P(z),$$

where F_P is the P-distribution of Z_∞. Thus the convex hull of the P_z-measures equals $\mathcal{M}$, and it is easy to see that all the P_z-measures are extremal. ∎

It is now natural to ask if an SGM always exists. Generally speaking this is still an open question, but our conjecture is that every transitive sufficient complete model where X has continuous trajectories possesses an SGM. We now give some partial results. First we see that to check the reverse martingale property it is (almost) enough to check it for one single P.

Proposition 3.3 *Suppose that a process Z has the properties that*

1. $\sigma\{X_t\} = \sigma\{Z_t\}$, *for all* $t \geq 0$.
2. *There exists some* $P_0 \in \mathcal{M}$ *such that*

$$E_{P_0}\left[Z_s \left| Z_t = z\right.\right] = z, \text{ for all } z \text{ and for all } s, t \text{ with } s < t. \tag{11}$$

Then Z is a P-reverse martingale for all $P \in \mathcal{M}$ such that $Z \in L^1(\Omega, P)$.

Proof. If (11) is satisfied for some $P_0 \in \mathcal{M}$ it will be satisfied for all $P \in \mathcal{M}$ since, by (1), Z is sufficient. Suppose furthermore that P is such that $Z \in L^1(\Omega, P)$. Then (11) implies that

$$E_P\left[Z_s \left| Z_t\right.\right] = Z_t, \; P - a.s.$$

Using the Markov property (in backward time) of Z this gives us

$$E_P\left[Z_s \left| \mathcal{T}_\infty^X\right.\right] = E_P\left[Z_s \left| Z_t\right.\right] = Z_t, \; P - a.s.$$

which shows that Z is a reverse P-martingale. ∎

It is an annoying fact that the condition (11) alone does not even guarantee that Z is a reverse P_0-martingale. The problem is that (11) does not imply that Z_s is P_0-integrable, it only tells us that Z_s belongs to $L^1(\Omega, P_Z^{t,z})$, where the tied-down measure $P_Z^{t,z}$ is defined by

$$P_Z^{t,z}(A) = P\left(A \left| Z_t = z\right.\right), \; A \in \mathcal{F}_t.$$

We now turn to the existence of an SGM. For models where there exists a transition density for the prediction sufficient process X we have a promising candidate.

Proposition 3.4 *Suppose that, for some $P \in \mathcal{M}$, X has a transition density $p(s, y; t, x)$, i.e. $P\left(X_t \in dx \left| X_s = y\right.\right) = p(s, y; t, x)dx$, and suppose furthermore that p is continuously differentiable in the y-variable. Let $P_{s,y}$ denote the measure generated by starting X in the state y at time s and define, for fixed (s, y) the process $Z_t; t \geq s$ by*

$$Z_t = \nabla_y \log p\left(s, y; t, X_t\right). \tag{12}$$

Suppose that Z defined by (12) is an integrable process with respect to $P_{s,y}$. Then Z is a reverse martingale on $[s, \infty)$ with respect to $P_{s,y}$.

Proof.

$$E\left[Z_t\left|X_T=x\right.\right]=E\left[\frac{p_y(s,y;t,X_t)}{p(s,y;t,X_t)}\left|X_T=x\right.\right]=$$

$$\int\frac{p_y(s,y;t,\xi)}{p(s,y;t,\xi)}p(T,x;t,\xi)d\xi\ =\int\frac{p_y(s,y;t,X_t)}{p(s,y;t,X_t)}\frac{p(t,\xi;T,x)p(s,y;t,\xi)}{p(s,y;T,x)}d\xi\ =$$

$$\int\frac{p_y(s,y;t,\xi)p(t,\xi;T,x)}{p(s,y;T,x)}d\xi\ =\frac{1}{p(s,y;T,x)}\int p_y(s,y;t,\xi)p(t,\xi;T,x)d\xi\ =$$

$$\frac{1}{p(s,y;T,x)}\nabla_y\int p_y(s,y;t,\xi)p(t,\xi;T,x)d\xi\ =\frac{1}{p(s,y;T,x)}\nabla_y p(s,y;T,x)=$$

$$=\nabla_y\log p(s,y;T,x),$$

which, again using the Markov property of X in backward time, gives us

$$E_{s,y}\left[Z_t\left|\mathcal{F}_{[T,\infty)}\right.\right]=E_{s,y}\left[Z_t\left|X_T\right.\right]=Z_T.\blacksquare$$

A natural candidate as SGM is thus given by

$$Z_t=\nabla_y\log p\left(0,X_0;t,X_t\right)\tag{13}$$

where we can choose any P in $\mathcal{M}$ to compute the transition density. We have so far been unable to give a really nice set of a priori conditions which will guarantee that Z above in fact is an SGM. In a concrete case we thus have to check the defining properties of the SGM, and then the following remarks can be helpful.

1. It follows from (3.4) that Z will be a reverse martingale with respect to any $P\in\mathcal{M}$ for which Z is an integrable process.

2. To qualify as an SGM the process Z also has to generate the same filtration as X, which is equivalent to the statement that, for each t, the mapping

$$\begin{aligned}&H_t:R^k\rightarrow R^k\\&x\ \mapsto\ \nabla_y\log p\left(0,x_0;t,x\right),\end{aligned}$$

is a global bijection. A necessary condition for this property is of course that H_t is locally invertible i.e. that the Jacobian $G(t,x)$ of H_t, is an invertible matrix for each (t,x), where

$$G_{i,j}(t,x)=\frac{\partial^2}{\partial x_i\partial x_j}\log p\left(0,x_0;t,x\right),\quad i,j=1,...,k$$

and if the mapping H_t also is proper then H_t will in fact be a global bijection.

3. Finally Z_∞ has to generate the tail sigma field $\mathcal{T}_\infty^X$. In practice this seems to be the hardest condition to check, and one result in this direction is the following proposition. The process Z below need not be the one defined by equation 13.

Proposition 3.5 *Suppose that Z is a scalar reverse martingale with continuous trajectories, such that $\sigma\{X_t\} = \sigma\{Z_t\}$ for all $t \geq 0$. Suppose furthermore that its quadratic variation satisfies*

$$d\langle Z\rangle_t = h(t)dt.$$

Then $\sigma\{Z_\infty\} = \mathcal{T}_\infty^X$.

Proof. By introducing the deterministic time-transformation $T(t) = 1/t$, defining Y by $Y(t) = Z(T(t))$ and defining the filtration $\underline{\mathcal{G}}$ by

$$\mathcal{G}_t = \sigma\{X_s; s \geq 1/t\}, \text{ for } t > 0,$$
$$\mathcal{G}_0 = \sigma\{Z_\infty\},$$

the proof boils down to that of showing that $\mathcal{G}_0 = \mathcal{G}_{0+}$. By a stochastic time transformation we may turn the Y-process, into standard Brownian Motion, and the result now follows from the right continuity of the Brownian filtration. ∎

We now consider our three test models in the light of the theory above.

The Wiener model: The natural base measure is the Wiener measure P_0, and from Girsanov's Theorem we have

$$L_t^\alpha = \exp\left\{\alpha X_t - \frac{\alpha^2 t}{2}\right\}.$$

It now follows from the factorization theorem that X itself is sufficient, and since X is Markovian it is indeed a transitive sufficient process. The model is easily seen to be complete. It can be shown (see e.g. [Bjö & Joh, 93]) that this model is in fact extremal, i.e.

$$\Pi_W = \mathcal{E}_W$$

This means that for a measure P in the maximal family the process X has the representation

$$dX(t) = Z^* dt + dW(t) \tag{14}$$

where Z^* is a stochastic variable which is independent of W. For a given P the distribution F_p of Z^* can be identified with the mixing measure corresponding to P. Thus $\mathcal{E}_W$ can be identified with the real line, and $\mathcal{M}_W$ can be identified with the set of all probability distributions on R.

In order to find an SGM we use the measure P_0 to compute the transition densities, which are given by

$$p(s, y; t, x) = \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{(x-y)^2}{2t}\right\}.$$

Using formula 13 we obtain the following well known process

$$Z_t = \frac{X_t}{t}, \tag{15}$$

and it is easy to see that Z is in fact an SGM.

The L^2-model: By the results above we see that this model is a submodel of the maximal family for the Wiener model. Thus X itself is again a prediction sufficient process and the extremal and maximal families, denoted by $\mathcal{E}_L$ and $\mathcal{M}_L$ respectively, are given by $\mathcal{E}_L = \mathcal{E}_W$, $\mathcal{M}_L = \mathcal{M}_W$. An SGM is again given by equation (15).

The Gaussian mixture model: This model is in fact a submodel of the L^2-model. Using formula (3) it is easy to see that the measure $P_{\alpha,\beta}$ corresponds to a Gaussian mixing measure (= distribution of Z^*) with mean α and variance β. From the results above we see that the extremal and maximal families $\mathcal{E}_G$ and $\mathcal{M}_G$ are given by

$$\begin{aligned} \mathcal{E}_G &= \{P_{\alpha,0}; \alpha \in R\} = \mathcal{E}_W, \\ \mathcal{M}_G &= \mathcal{M}_W, \end{aligned}$$

and again an SGM is given by formula (15).

An example in discrete time: Although the theory above has been presented only in continuous time the basic ideas work as well in discrete time. We illustrate with the following example, where $\Omega = R^\infty$.
Let X be the coordinate process on Ω and let $P_{\mu,\sigma}$ denote the measure under which $X_1, X_2, \ldots$ are i.i.d. and $N(\mu, \sigma)$. A sufficient statistic for the family $\{P_{\mu,\sigma}; \mu \in R,\ \sigma \in (0, \infty)\}$ at time n is (Y_n, V_n) where

$$Y_n = \sum_{k=1}^{n} X_k, \qquad V_n = \sum_{k=1}^{n} X_k^2.$$

The natural base measure is $P_{0,1}$, under which (Y, V) is a Markov process with transition densities $p^*(m, x, u; n, y, v) = p(n-m, y-x, v-u)$, where

$$p(n, y, v) = \frac{\left(v - y^2/n\right)^{(n-3)/2}}{T\sqrt{n\pi}\left(\frac{n-1}{2}\right)\frac{2^n}{2}}.$$

Following the construction in equation 13 we define the functions f_1 and f_2 by

$$\begin{aligned} f_1(n,y,v) &= \nabla_x \log p^*(0,0,0;n,y,v) = \\ &= \frac{n-3}{n}\frac{y}{v-y^2/n} \\ f_2(n,y,v) &= \nabla_u \log p^*(0,0,0;n,y,v) = \\ &= \frac{1}{2}\Big(1-\tfrac{1}{v-y^2/n}\Big). \end{aligned}$$

Now we define the process $Z=(Z^1,Z^2)$, which is our candidate as an SGM, by

$$Z_n^1 = f_1(n,Y_n,V_n) = \frac{n-3}{n-1}\frac{\overline{X}_n}{S_n^2} \tag{16}$$

$$Z_n^2 = f_2(n,Y_n,V_n) = \frac{1}{2}\Big(1-\tfrac{n-3}{n-1}\tfrac{1}{S_n^2}\Big) \tag{17}$$

and we know from Proposition 3.4 that Z in fact is a reverse martingale. Furthermore the mapping $(y,v)\mapsto(f_1(n,y,v),f_2(n,y,v))$ is obviously one to one, and it only remains to show that Z_∞ generates the tail sigma algebra $\mathcal{J}$. This however follows immediately from the fact that $\mathcal{J}$ is trivial (see [Mar 70]), and thus Z is an SGM.

To connect these result with our results in Section 2 and to classical theory we recall the well known fact that the process

$$\xi=\big\{\big(\overline{X}_n,S_n^2\big),n\geq 2\big\}$$

is an unbiased estimator of the parameter (μ,σ^2). Since this model is complete, it follows from Proposition 2.1 the process ξ must be a reverse martingale (a fact which does not seem to be known), and since in this case the tail sigma field is known to be trivial, we immediately obtain the consistency result

$$\lim_{n\to\infty}\xi_n=(\mu,\sigma^2),\;\; P_{\mu,\sigma}-a.s.$$

For this model we have already constructed another SGM, namely the process Z defined by equations (16) - (17). Since Z is a reverse martingale it is thus the unbiased estimator of *something* and a fairly easy calculation shows that in fact

$$\begin{aligned} E_{\mu,\sigma}\left[Z_n^1\right] &= \frac{\mu}{\sigma^2}, \\ E_{\mu,\sigma}\left[Z_n^2\right] &= \frac{1}{2}\big(1-\tfrac{1}{\sigma^2}\big). \end{aligned}$$

Here we recognize the "canonical parameters" of the model in the sense of the theory of exponential families. This fact, together with other concrete examples as the Wiener model above, strongly indicates that the gradlog construction in formula 12 is not of an ad hoc nature but rather that this construction in some sense is canonical.

4 Asymptotic normality of unbiased estimators

The fact that unbiased estimators are reverse martingales suggests the possibility of deriving asymptotic normality from the central limit theorem for reverse martingales. We shall not attempt a general discussion at this point, but instead illustrate the idea for a familiar example.
The sample space is R^∞, $X = X_1, X_2, ...$ is the coordinate process on this space and $\mathcal{F} = \sigma(X_1, X_2, ...)$. Let $P_{\mu,\nu}$ denote the probability measure for which $X = X_1, X_2, ...$ are independent and normally distributed with mean μ and variance ν. It is well known that the pair

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad W_n = \frac{1}{n-1}\sum_{k=1}^{n} X_k^2$$

is a minimal sufficient statistic for the model $\{P_{\mu,\nu}; \mu \in R, \nu > 0\}$. It is also well known that

$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad U_n = \frac{1}{n-1}\sum_{k=1}^{n} (X_k - T_k)^2$$

are unbiased estimators of μ and ν and that the model is complete. Thus $\{T_n; n \geq 2\}$ and $\{U_n; n \geq 2\}$ are reverse martingales with respect to any measure $P_{\mu,\nu}$ and the filtration

$$\mathcal{J}_n = \sigma\{(S_n, W_n), (S_{n+1}, W_{n+1}), ...\}, \; n \geq 2.$$

Let us now define the probability measure P_θ on $(\Omega, \mathcal{F})$ by

$$P_\theta(A) = \int P_{\mu,\nu}(A) F_\theta(d\mu, d\nu), \quad A \in \mathcal{F}, \tag{18}$$

where

$$F_\theta(d\mu, d\nu) = \epsilon_\theta(d\mu)\nu^{-2}e^{-1/\nu}$$

(This is Example 1, p.2 in Basawa and Scott [Bas & Sco, 83]). It now follows that $\{T_n\}$ and $\{U_n\}$ are reverse martingales with respect to P_θ as well. Furthermore the joint density of $X = X_1, ..., X_n$ under P_θ is given by

$$f(x_1, ..., x_n) = (2\pi)^{-n/2}\Gamma(n/2+1)\left(1 + \tfrac{1}{2}\textstyle\sum_{k=1}^{n} (x_k - \theta)^2\right)^{-n/2-1}.$$

The estimator T_n is an unbiased estimator of θ and it is also the maximum likelihood estimator. An argument for constructing confidence intervals for θ goes as follows. Let $Z_n = (T_n - \theta)/\sqrt{U_n/n}$. Then one can show that

$$P_\theta(Z_n \leq z, U_n \leq u) \to \Phi(z) P_\theta(U_\infty \leq u), \tag{19}$$

where Φ is the standard normal distribution and $U_\infty = \lim_{n\to\infty} U_n$ has the density $\nu^{-2}e^{-1/\nu}$. Since U_∞ has the same distribution for all θ, the statistic U_n

is approximately ancillary, and the (asymptotic) conditionality principle yields that confidence intervals for θ should be based on the asymptotic normality of Z_n. The purpose of this section is to show how the basic convergence result (19) follows from the central limit theorem for reverse martingales and the mixing formula (18).
We will use the following central limit theorem for reverse martingales (c.f. [Hall & Hey, 80], [Eag & Web, 78]).
Let $\{\mathcal{J}_n, n \geq 1\}$ be a decreasing sequence of σ-fields and let $\{T_n, n \geq 1\}$ be a reverse martingale with respect to $\{\mathcal{J}_n, n \geq 1\}$, such that $E\left[T_n^2\right] < \infty, n \geq 1$. Set

$$Y_n = T_n - T_{n-1}, \quad V_n^2 = \sum_{k=n}^{\infty} E\left[Y_k^2 \,|\mathcal{J}_{k+1}\right], \quad n \geq 1.$$

Proposition 4.1 *Assume that there exists a random variable η such that*

$$\frac{V_n^2}{E\left[V_n^2\right]} \xrightarrow{P} \eta^2, \quad \text{as } n \to \infty \tag{20}$$

and, for all $\epsilon \geq 0$:

$$\frac{1}{E\left[V_n^2\right]} \sum_{k=n}^{\infty} E\left[Y_k^2 I_{\left\{Y_k^2 > \epsilon E\left[V_n^2\right]\right\}} \,|\mathcal{J}_{k+1}\right] \xrightarrow{P} 0, \quad \text{as } n \to \infty \tag{21}$$

Then

$$\frac{T_n - T_\infty}{V_n} \xrightarrow{d} N(0,1)$$

Note that the Lindeberg condition (21) is implied by the Lyapunov condition

$$\frac{1}{E\left[V_n^2\right]} \sum_{k=n}^{\infty} E\left[|Y_k|^r \,|\mathcal{J}_{k+1}\right] \xrightarrow{P} 0, \quad \text{for some } r > 2 \tag{22}$$

We shall now use this result to show that for *any* P in the maximal family,

$$\frac{T_n - T_\infty}{\sqrt{U_n/n}} \xrightarrow{d} N(0,1). \tag{23}$$

To begin with, let $P = P_{\mu,\nu}$ for some $\mu \in R,\ \nu > 0$. Then one can show that

$$E\left[(T_n - T_{n+1})^2 \,|\mathcal{J}_{k+1}\right] = \frac{1}{n(n+1)} U_{n+1}, \tag{24}$$

$$E\left[(T_n - T_{n+1})^4 \,|\mathcal{J}_{k+1}\right] = \frac{1}{n(n+1)^2(n+2)} U_{n+1}^2, \tag{25}$$

The calculations are elementary but somewhat lengthy, and we omit them. From (23) we get

$$V_n^2 := \sum_{k=n}^{\infty} E\left[(T_n - T_{n+1})^2 \,|\mathcal{J}_{k+1}\right] = \sum_{k=n}^{\infty} \frac{1}{k(k+1)} U_{k+1}. \tag{26}$$

Using the relation

$$\sum_{k=n}^{\infty} \frac{1}{k(k+1)} = \frac{1}{n}$$

we get

$$E\left[V_n^2\right] = \sum_{k=n}^{\infty} \frac{\nu}{k(k+1)} = \frac{\nu}{n}.$$

Furthermore

$$\frac{1}{n} \inf_{k \geq n+1} U_k \leq V_n^2 \leq \sup_{k \geq n+1} U_k,$$

and since $U_n \to \nu,\ P_{\mu,\nu} - a.s.$, we have

$$\frac{V_n^2}{E\left[V_n^2\right]} \to 1; \quad P_{\mu,\nu} - a.s.$$

Thus the condition (20) is satisfied. Next we check that condition (22) is satisfied with $r = 4$. From (24) we get

$$\left(E\left[V_n^2\right]\right)^{-2} \sum_{k=n}^{\infty} E\left[(T_n - T_{n+1})^4 \,|\mathcal{J}_{k+1}\right] = (\nu/n)^{-2} \sum_{k=n}^{\infty} \frac{1}{k(k+1)(k+2)} U_{k+1}^2.$$

Arguing as above, it suffices to show that

$$n^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)^2(k+2)} \to 0.$$

But

$$n^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)^2(k+2)} \leq \sum_{k=n}^{\infty} \frac{1}{k(k+1)} = \frac{1}{n},$$

and so (22) holds.
Since $nV_n^2/U_n \to 1,\ P_{\mu,\nu} - a.s.$ we conclude that (23) holds under $P_{\mu,\nu}$. But then it follows that for any P in the maximal family we have

$$\begin{aligned} P\big(n(T_n - T_\infty/\sqrt{U_n} \leq z\big) &= \int P_{\mu,\nu}\big(n(T_n - T_\infty/\sqrt{U_n} \leq z\big) F_P(d\mu, d\nu) \\ &\to \int \Phi(z) F_P(d\mu, d\nu) = \Phi(z). \end{aligned}$$

To conclude the proof of (19), note that

$$\begin{aligned}&|P_\theta(Z_n \le z,\ U_n \le u) - P_\theta(Z_n \le z,\ U_\infty \le u)|\\ &\le E_\theta\left[|I_{\{U_n \le u\}} - I_{\{U_\infty \le u\}}|\right] \to 0,\end{aligned}$$

and so it is enough to prove that

$$P_\theta(Z_n \le z,\ U_\infty \le u) \to \Phi(z) P_\theta(U_\infty \le u).$$

But the mixing formula (18) gives us

$$\begin{aligned}& P_\theta(Z_n \le z,\ U_\infty \le u) = \\ = \quad & \int_0^\infty P_{\theta,\nu}(Z_n \le z, U_\infty \le u)\nu^{-2}e^{-1/\nu}d\nu \\ = \quad & \int_0^u P_{\theta,\nu}(Z_n \le z)\nu^{-2}e^{-1/\nu}d\nu \\ \to \quad & \int_0^u \Phi(z)\nu^{-2}e^{-1/\nu}d\nu \\ = \quad & \Phi(z)P_\theta(U_\infty \le u).\end{aligned}$$

5 The structure of unbiased estimators

In this section we look for conditions on a parameter Φ which are necessary for the existence of an unbiased estimator process. Let us therefore consider a fixed parameter Φ for the model $(\Omega, \mathcal{F}, \Pi, X, \underline{\mathcal{F}})$.

Definition 5.1 *For any stochastic variable V we define a family $\mathcal{M}[V]$, by*

$$\mathcal{M}[V] = \left\{P \in \mathcal{M}; V \in L^1(\Omega, P)\right\}$$

If V is an unbiased estimator of Φ relative to Π , then we extend the domain of Φ from Π to $\mathcal{M}[V]$, by

$$\Phi(P) = E_P\left[V\right],\ P \in \mathcal{M}[V]. \tag{27}$$

Lemma 5.1 *Suppose that, for some t, there exists an unbiased t-estimator V of Φ, and suppose that $\mathcal{E} \subset \mathcal{M}[V]$. Then, extending Φ as above it must hold that*

$$\Phi(P) = \int_{\mathcal{E}} \Phi(P_\epsilon)\nu_p(dP\epsilon), \tag{28}$$

where ν_p is the mixing measure for P.

Proof. We have, using the unbiasedness of V,

$$\begin{aligned}\Phi(P) &= E_P[V] = \int_\Omega V(\omega)P(d\omega) = \int_\Omega\int_{\mathcal{E}} V(\omega)P_\epsilon(d\omega)\nu_p(dP_\epsilon) = \\ &= \int_{\mathcal{E}}\int_\Omega V(\omega)P_\epsilon(d\omega)\nu_p(dP_\epsilon) = \int_{\mathcal{E}} \Phi(P_\epsilon)\nu_p(dP_\epsilon). \blacksquare\end{aligned}$$

In the case when we have access to an SGM the structure can be simplified even further.

Definition 5.2 *Consider a model possessing a sufficient generating martingale Z, and let Φ be a parameter defined on the whole of $\mathcal{E}$. Then the* **structure function**, *$\varphi : R^k \to R$, is defined by*

$$\varphi(z) = \Phi(P_z),$$

where P_z is defined by equation (9).

Proposition 5.1 *Suppose that the model possesses an SGM denoted by Z. Consider a fixed parameter Φ, and suppose that there exists some unbiased estimator of Φ. Then it must hold that*

$$\Phi(P) = \int_{R^k} \varphi(z)F_P(dz), \quad \textit{for all } P \in \mathcal{M}[V], \tag{29}$$

where F_P as usual is the P-distribution of Z_∞.

Proof. Follows immediately from the lemma above. $\blacksquare$

The point of the results above is of course that the structure of a parameter having an unbiased estimator is uniquely determined by its structure function. We may now restate Proposition 5.1 in order to give preliminary necessary and sufficient conditions for the existence of unbiased estimators.

Proposition 5.2 *Consider a model $(\Omega, \mathcal{F}, \Pi, X, \underline{\mathcal{F}})$, and a fixed parameter Φ, where we assume that $\mathcal{E} \subset \Pi$. Then there exists an unbiased estimator of Φ if and only if the following conditions hold.*

1. *Φ has an unbiased estimator on $\mathcal{E}$.*
2. *For all $P \in \Pi$ it holds that*

$$\Phi(P) = \int_{R^k} \varphi(z)F_P(dz). \tag{30}$$

Proof. Obvious. ∎

Thus the problem of estimating Φ on Π is replaced by the much easier problem of estimating Φ on $\mathcal{E}$. The latter problem is however by no means trivial and we shall return to it below.

Even as it stands, Proposition 5.2 has a number of nontrivial consequences, so let us look at some of our examples.

The Wiener model: Since, in this model, we have $\Pi = \mathcal{E}$, we do not get anything interesting from Proposition 5.2. We recall, however, that for the simple parameter $\Phi(P_\alpha) = \alpha$ we have the trivial unbiased estimator $Y(t) = \frac{X(t)}{t}$. This process is also an SGM, and the structure function of the parameter is of course given by $\varphi(z) = z$.

The L^2-model: We consider the parameters

$$\Phi(P) = E_P[Z^*], \tag{31}$$

$$\Phi(P) = Var_P[Z^*], \tag{32}$$

$$\Phi(P) = E_P\left[(Z^*)^2\right], \tag{33}$$

The corresponding structure functions are (recall that P_z is a point mass at z).

$$\begin{aligned}
\varphi(z) &= \Phi(P_z) = E_{P_z}[Z^*] = z, \\
\varphi(z) &= \Phi(P_z) = Var_{P_z}[Z^*] = 0, \\
\varphi(z) &= \Phi(P_z) = E_{P_z}\left[(Z^*)^2\right] = z^2.
\end{aligned}$$

We now check if the parameters satisfy the integral condition of Proposition 5.1. For the parameter $\Phi(P) = E_P[Z^*]$ we have

$$\int_{R^k} \varphi(z) F_P(dz) = \int_{R^k} z F_P(dz) = E_P[Z^*] = \Phi(P),$$

so this parameter satisfies equation (29). Furthermore we know that the extremal points of this model equals the a priori family for the Wiener model, so Φ possesses an unbiased estimator process on $\mathcal{E}$, namely $Y(t) = X(t)/t$. Thus it follows from Proposition 5.2 that Φ indeed has an unbiased estimator on Π. (In this simple example one could of course prove this directly by simply checking the obvious candidate $X(t)/t$.)

For the parameter $\Phi(P) = Var_P[Z^*]$ we have

$$\int_{R^k} \varphi(z) F_P(dz) = \int_{R^k} 0 F_P(dz) = 0,$$

which is not equal to $\Phi(P)$ for any measure outside the extremal family. It follows that for this parameter no unbiased estimator exists (relative to Π).

For the case $\Phi(P) = E_P\left[(Z^*)^2\right]$ we have

$$\int_{R^k} \varphi(z) F_P(dz) = \int_{R^k} z^2 F_P(dz) = E_P\left[(Z^*)^2\right] = \Phi(P).$$

Proposition 5.2 now tells us that it is possible that the parameter can be estimated unbiasedly. Whether this really is the case depends on if one can estimate Φ on the extremal points. This is a harder problem and we will come back to it below.

The Gaussian mixture model: We consider the parameters

$$\Phi(P_{\alpha,\beta}) = \alpha, \tag{34}$$
$$\Phi(P_{\alpha,\beta}) = \beta, \tag{35}$$
$$\Phi(P_{\alpha,\beta}) = g(\alpha,\beta). \tag{36}$$

We recall that this is a submodel of the L^2-model, and that Z^* under $P_{\alpha,\beta}$ is normally distributed with mean α and variance β. Thus we already know from the L^2-model that α can be estimated unbiasedly, whereas β can not.
The case $\Phi(P_{\alpha,\beta}) = g(\alpha,\beta)$ is more interesting. Since the extremal family is $\{P_{\alpha,0}; \alpha \in R\}$ we see that the structure function is given by

$$\varphi(z) = \Phi(P_{\alpha,\beta}) = g(z,0).$$

The integral condition of Proposition 5.2 becomes

$$\Phi(P_{\alpha,\beta}) = g(\alpha,\beta) = \int_{R^k} \varphi(z) F_{P_{\alpha,\beta}}(dz) = \int_{R^k} g(z,0)\Psi(z;\alpha,\beta)dz, \tag{37}$$

where $\Psi(z;\alpha,\beta)$ is the Gaussian density with mean α and variance β. Thus we see that a necessary condition for g is that it satisfies the integral equation

$$g(\alpha,\beta) = \int_{R^k} g(z,0)\Psi(z;\alpha,\beta)dz. \tag{38}$$

Furthermore, since Ψ as a function of α and β satisfies the Fokker-Planck equation, the above equation implies that this must also be the case with the function g. Summing up we have the following result.

Proposition 5.3 *The function g has an unbiased estimator if and only if the following conditions hold.*

1.
$$\frac{\partial g}{\partial \beta} = \frac{1}{2}\frac{\partial^2 g}{\partial \alpha^2}, \quad (\alpha,\beta) \in R \times R_+. \tag{39}$$

2. *The parameter* $\Phi(P_{\alpha,\beta}) = g(\alpha, 0)$ *has an unbiased estimator on* $\mathcal{E}$.

Using Proposition 5.3 it is easy to check necessary conditions. We see at once that for $g(\alpha, \beta) = \beta$, corresponding to the parameter in (32), equation (39) is not satisfied, so this parameter has no unbiased estimator. On the other hand the function $g(\alpha, \beta) = \beta + \alpha^2$ corresponding to the parameter in (33), works nicely.

6 Boundary value problems

In Section 5 we saw that the task of finding an unbiased estimator splits into two separate problems:

1. The easy task of checking if the parameter satisfies one of the "structural equations" (28) ,(29) or (30).
2. The hard task of constructing an estimator for the restriction of the parameter to extremal family.

In this section we will derive equations for the actual construction of unbiased estimators (on the extremal family). We start with a lemma and we use the standing assumption that a sufficient martingale exists for the model.

Lemma 6.1 *Suppose that* Z *is an SGM. A given optional process* Y *is an unbiased estimator of* Φ *relative to the extremal family* $\mathcal{E}$ *on the interval* $[T, \infty)$ *if and only if the following equation is satisfied for all* $P \in \mathcal{E}$.

$$E_P \left[Y(t) \,|Z_\infty = z \right] = \varphi(z), \quad t \in [T, \infty), \quad PZ_\infty^{-1} - a.s. \tag{40}$$

Proof. Another way of writing equation 40 is

$$E_{P_z} \left[Y(t) \right] = \Phi(P_z),$$

and since the P_z-measures constitute the extremal family we are finished. ∎

Still another way of writing equation (40) is as

$$E_P \left[Y(t) \,|Z_\infty \right] = \varphi(Z_\infty),$$

which shows that unbiased parameter estimation can be viewed as a limiting case of unbiased adaptive prediction in the sense of Johansson-Björk (1992) [Bjö & Joh, 92].
It is not at all clear from Lemma 6.1 how one is to find a process Y satisfying eqauation (40). If we want to minimize expected square error then, because of predictive sufficiency, we only have to consider estimators of the form

$$Y(t) = f(t, X(t)). \tag{41}$$

If, as we have assumed, we have an SGM, then it turns out to be much more convenient to consider estimators of the form

$$Y(t) = f(t, Z(t)), \tag{42}$$

and, since by definition X and Z generate the same filtration, the forms (41) and (42) are logically equivalent. Thus we may as well look for estimators of the form (42), and we can now use the reverse martingale characterization of unbiased estimators (Theorem 2.1) to obtain a more familiar problem.

Definition 6.1 *Let $R(s)$ denote the infinitesimal operator for the process Z in reverse time, at (ordinary) time s, and let $D(S)$ denote its domain. In other words*

$$\{R(s)g\}(z) = \lim_{t \uparrow s} \frac{E_P\left[g(Z_t)\,|Z_s = z\right]}{s-t}.$$

Proposition 6.1 *Suppose that for some $T > 0$ there exists a function*

$$f : (T, \infty] \times R^k \to R$$

with

$$f(s, \cdot) \in D(s), \quad \textit{for all } s \geq T,$$

such that f solves the boundary value problem

$$\frac{\partial f}{\partial t}(t, z) = [R(t)f(t, \cdot)](z), \quad (t, z) \in (T, \infty) \times R^k, \tag{43}$$

$$f(\infty, z) = \varphi(z), \quad z \in R^k \tag{44}$$

Then the process Y defined by $Y(t) = f(t, Z(t))$ is an unbiased estimator of Φ on $\mathcal{E}$. If the model is complete then Y is also the unique mean square optimal unbiased estimator.

Proof. By Dynkin's formula, equation (43) simply says that Y is a reverse martingale. Thus it is an unbiased estimator of something, and the boundary value shows that it is actually an estimator of Φ. ∎

We may of course instead look for estimators of the form $f(t, X_t)$, and this will also lead to an equation of the form (43). The drawback of this approach is that no nice boundary values are at hand.
It is important to notice that equation (43) is the *inverse* problem of an ordinary Kolmogorov-type backward equation (where the natural boundary conditions would be given at $t = 0$).In other words we are tying to invert the semigroup generated by Z. The inverse nature of the problem also explains why unbiased estimators are so rare. If for example Z is a diffusion process, then (43) will be (a version of) the heat equation solved backwards in time. In forward time the

heat equation has a very regularizing effect on boundary data, so solving equation (43) will have an extremely irregularizing effect on the boundary function φ. Thus the structure function φ has to be very smooth for a solution to exist. Suppose now that Z is a diffusion in forward time. Then under fairly mild technical conditions it will also be a diffusion in backward time (see [Haus & Pard 86]). Since Z by definition is a reverse martingale it will thus have the reverse time representation

$$d\overline{Z}_t = \sigma(t, \overline{Z}_t) d\overline{W}_t,$$

where $\overline{W}$ is a Wiener process in reverse time, and σ is the diffusion term (which is the same in forward and in reverse time). We may now obtain a stochastic representation formula for the solution of equation (43), a fact which in view of the inverse nature of the problem is somewhat surprising.

Proposition 6.2 *Suppose that f is a solution to equation (43) and suppose also that there exists a time $T > 0$ such that the following hold.*

1. *As a function of z σ can be extended to an analytic function in the whole complex plane.*

2. *For every (t, z) with $t \geq T$ there exists a solution to the complex-valued SDE*

$$\begin{cases} dZ_s^c &= i\sigma(s, Z_s^c), \\ Z_t^c &= z. \end{cases} \tag{45}$$

on the closed interval $[T, \infty]$.

3. *For all $(t, z) \in (T, \infty) \times R^k$ the process Z^c in equation 45 and the function f satisfies the condition*

$$\int_t^\infty E_{t,z}\left[\|(\nabla_x f)(s, Z_s^c)\sigma(s, Z_s^c)\|^2 \, ds \right] < \infty \tag{46}$$

Then we have the representation formula

$$f(t, z) = E_{t,z}\left[\varphi(Z_\infty^c)\right] \tag{47}$$

where the indices in the expectation operator denotes integration with respect to the measure induced by equation (45).

Proof. Itô's formula. ∎

Note that the representation result above provides us with a surprising duality relation between the estimator and the parameter (structure function). We have

$$\begin{aligned} \varphi(z) &= E_{P_z}\left[f(t, Z_t)\right], \quad \text{for all} z \in R^k, & (48) \\ f(t, z) &= E_{t,z}\left[\varphi(Z_\infty^c)\right]. & (49) \end{aligned}$$

Thus we see that while (by definition) the parameter is the expected value of the estimator (eq. 48), we also have the fact that the estimator itself is the expected value of the parameter (eq. 49).

We now apply the representation formulas above to our standing examples.

The Wiener model: We seek estimators for parameters of the form $\Phi(P_\alpha) = h(\alpha)$, where h is some given real valued function, and since this model is extremal we see that the structure function φ coincides with h. We already know that $Z_t = X_t/t$ is an SGM, and we see that under P_0, Z has the forward time dynamics

$$dZ_t = -\frac{1}{t}Z_t dt + \frac{1}{t}dW_t.$$

The backward dynamics, which because of sufficiency do not depend on the choice of P, are given by

$$d\overline{Z}_t = \frac{1}{t}d\overline{W}_t,$$

so the boundary value problem of Proposition 6.1 becomes

$$\begin{aligned} \frac{\partial f}{\partial t}(t,z) &= \frac{1}{2t^2}\frac{\partial^2 f}{\partial z^2}(t,z) && (50)\\ f(\infty, z) &= h(z), && (51) \end{aligned}$$

We see indeed that this is an ill posed problem and to solve it we apply the representation formula (47). Equation (45) now reads

$$\left\{ \begin{array}{rcl} dZ_s^c &=& \frac{i}{t}dW_s, \\ Z_t^c &=& z, \end{array} \right. \qquad (52)$$

and this SDE can in fact be integrated directly. We have

$$Z_\infty^c = z = i\int_t^\infty \frac{1}{s}dW_s,$$

so we see that

$$Z_\infty^c = z + iU,$$

where U is Gaussian with zero mean and variance t^{-2}. Thus we have the following representation formula for f.

$$f(t,z) = \frac{t}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(x+iy)\exp\left\{-\tfrac{1}{2}x^2t^2\right\}dx. \qquad (53)$$

As was expected we see from equation (53) that we must demand a high degree of regularity from the function h; to start with it must be analytic. Note also that all these calculations have been made under the assumption that a solution to the boundary value problem (50)-(51) actually exists. We can however turn the whole argument around and *define* the function f by (53). Taking care of integrability conditions this procedure will give us the following result.

Proposition 6.3 *Consider the Wiener model and suppose that*

1. *The function h above is an entire analytical function*
2. *There exists positive constants A and T such that*

$$|h(x+iy)| \leq A \exp\left\{T\frac{x^2+y^2}{2}\right\}, \quad (x,y) \in R^2. \tag{54}$$

Then there exists a unique mean square optimal unbiased predictor process for the parameter $\Phi(P_\alpha) = h(\alpha)$ on the interval $[T, \infty)$. The estimator process Y is given by $Y_t = f(t, Z_t)$ where $Z_t = X_t/t$ and f is defined by (53).

Proof. The growth condition on y in (54) allows us to differentiate under the integral sign in equation (53) thus showing that f defined by (53) solves the boundary value problem (50) - (51). This implies that Y is a reverse martingale provided that Y_t is integrable for all $t \geq T$, and the growth condition on x ensures that this is indeed the case. ∎

As a concrete example let us consider the parameter $\Phi(P_\alpha) = h(\alpha) = \alpha^2$. The function h satisfies the conditions of Proposition 6.3, so Φ can in fact be estimated unbiasedly. The optimal estimator process is given by equation (53) as

$$Y_t = Z_t^2 - \frac{1}{t^2} = \frac{X_t^2 - 1}{t^2}. \tag{55}$$

The Gaussian mixture model We now return to the problem of estimating a parameter of the form $\Phi(P_{\alpha,\beta}) = g(\alpha, \beta)$. One half of this problem was solved in Section 5 where we saw that g must satisfy the heat equation (39), and we are left with the problem of deciding when the parameter Φ can be estimated on the extremal family $\mathcal{E}_G$. As we already know $\mathcal{E}_G = \Pi_W$, so we are in fact faced with the problem of estimating a parameter of the form $\Psi(P_\alpha) = h(\alpha) = g(\alpha, 0)$ in the Wiener model. The solution to this latter problem is on the other hand given by Proposition 6.3 above, so we have the following result.

Proposition 6.4 *Consider the Gaussian mixture model and a parameter Φ of the form $\Phi(P_{\alpha,\beta}) = g(\alpha, \beta)$. Then Φ possesses an unbiased estimator on the interval (T, ∞) if and only if the following conditions hold.*

1. *The function g satisfies the heat equation*

$$\frac{\partial g}{\partial \beta}(\alpha, \beta) = \frac{1}{2}\frac{\partial^2 g}{\partial \alpha^2}, \quad (\alpha, \beta) \in R \times R_+. \tag{56}$$

2. *The function h defined by $h(\alpha) = g(\alpha, 0)$ is an entire analytic function.*

3. *There is a constant A such that*

$$|h(x+iy)| \leq A \exp\left\{T \cdot \tfrac{x^2+y^2}{2}\right\}, \quad (x, y) \in R^2. \tag{57}$$

If the conditions are satisfied the estimator process Y is given by $Y_t = f(t, Z_t)$, where $Z_t = X_t/t$ and f is defined by equation (53).

Proof. Use Propositions 5.3 and 6.3. ∎

As a concrete example we take the parameter

$$\Phi(P_{\alpha,\beta}) = \alpha^2 + \beta = E_{\alpha,\beta}\left[(Z^*)^2\right].$$

In this case $g(\alpha, \beta) = \alpha^2 + \beta$ and this function clearly saisfies equation (39). The function h of proposition 6.4 is given by $h(\alpha) = \alpha^2$, and it satisfies conditions 2 - 3 of proposition 6.4. Again by the same proposition we thus have an optimal estimator, given by the formula (55).
Notice the role played by equation (39). The parameter

$$\Phi(P_{\alpha,\beta}) = \alpha^2 = \{E_{\alpha,\beta}[(Z^*)]\}^2.$$

has the same structure function as $\Phi(P_{\alpha,\beta}) = \alpha^2 + \beta$ and thus the same h-function. It does not, however, possess an unbiased estimator relative to the family Π_G, since it does not satisfy equation (39).

7 Identification

In this section we will try to understand what a parameter Φ must look like in order to be identifialble in the sense of Definition 1.2. First of all we note the obvious fact that if Φ has a consistent estimator process Y then Φ is identifiable by the stochastic variable $V = \lim_{t\to\infty} Y(t)$.
It is natural to ask if there is a converse to this result, i.e. if every identifiable parameter has a consistent estimator process. Generally speaking the answer seems to be no, but there is a trivial partial converse. Suppose that there exists an SGM Z, and suppose that Φ can be identified by the $\mathcal{T}_\infty^X$ - measurable stochastic variable V. Since $Z(\infty)$ by definition generates $\mathcal{T}_\infty^X$ we then must have $V = g(Z(\infty))$ for some Borel measurable function g, and if furthermore g

is continuous we see that we can construct a consistent process Y by the definition $Y(t) = g(Z(t))$. On the other hand let us consider the Wiener model and fix a bounded discontinuous function g, e.g. the indicator of the rationals. Now define the parameter Φ by $\Phi(P_\alpha) = g(\alpha)$. Then this parameter is obviously identifiable by the variable $V = g(Z(\infty))$ (where as usual $Z(t) = X(t)/t$), but we believe that it has no consistent estimator process. This we have not proved however.
We now turn to the problem of characterizing the class of identifiable parameters. Let us therefore consider a fixed model $(\Omega, \mathcal{F}, \Pi, X, \underline{\mathcal{F}})$ and a fixed parameter Φ.

Definition 7.1 *For each $r \in R$ let $\mathcal{L}_r$ be the class of measures defined by*

$$\mathcal{L}_r = \{p \in \Pi | \Phi(P) = r\}.$$

The family $\mathcal{L}(\Phi) = \{\mathcal{L}_r | r \in R\}$ is said to be **uniformly orthogonal** *if there exists a family $\{S_r | r \in R\}$, consisting of $\mathcal{F}$- measurable subsets of Ω such that*

$$r \neq q \Rightarrow S_r \bigcap S_q = \emptyset \tag{58}$$

$$P(S_r) = 1, \quad \textit{for all } P \in \mathcal{L}_r \tag{59}$$

We now have the following simple but useful result.

Proposition 7.1 *A necessary condition for the parameter Φ to be identifiable is that the family $\mathcal{L}(\Phi)$ is uniformly orthogonal.*

Proof. Suppose that the stochastic variable V identifies Φ. Now define S_r for each $r \in R$ by

$$S_r = \{\omega \in \Omega | V(\omega)\} = r.$$

Since $V(\omega) = \Phi(P)$, $P - a.s.$ it follows that $V(\omega) = r,\ P - a.s.$ for all $P \in \mathcal{L}_r$. Thus $P(S_r) = 1$ for all $P \in \mathcal{L}_r$, and the S_r-sets are disjoint by definition. ∎

The uniform orthogonality in Proposition 7.1 is also "almost" sufficient. Suppose indeed that the family $\mathcal{L}(\Phi)$ is uniformly orthogonal. Then it is tempting to define an identifier V by

$$V = \sum_{r \in R} r \cdot I\{\omega \in S_r\}$$

The problem is that we have no guarantee that the V we have just defined is measurable.
It is worth noticing that we really need the concept of *uniform* orthogonality

as opposed to ordinary orthogonality. Consider for example the following small submodel of the L^2-model. We define the family Π by

$$\Pi = \{N\} \bigcup \{F_\alpha \ \alpha \in R\},$$

where N is the standard normal distribution (or any diffuse distribution) and F_α is a unit mass at α. Furthermore we define the parameter by

$$\begin{cases} \Phi(F_\alpha) &= \alpha, \quad \alpha \in R, \\ \Phi(N) &= 0. \end{cases}$$

Then the family $\mathcal{L}(\Phi)$ is given by

$$\begin{cases} \mathcal{L}_r &= \{F_r\} \\ \mathcal{L}_0 &= \{F_0, N\}. \end{cases}$$

Since N has support on the real line this family is not uniformly orthogonal despite the fact that all measures at hand are pairwise orthogonal. Thus, because of Proposition 7.1, the parameter cannot be identified. Intuitively this is also quite clear since if you observe the value $Z_\infty = \alpha$, then you still do not know whether you have an observation from the distribution F_α or from N. In other words the parameter value that you would like to identify can be either α or 0. It is worth stressing that all this depend on the fact that we assume that we only observe a single trajectory of the X-process.
We now have some easy consequences of Proposition 7.1. The first shows that no nontrivial parameter can be identified if the model is so big that the the a priori family Π equals the maximal family $\mathcal{M}$.

Proposition 7.2 *Suppose that Φ is identifiable and suppose that $\Pi = \mathcal{M}$. Then Φ must be a constant.*

Proof. Suppose indeed that V identifies Φ for $\mathcal{M}$. Suppose furthermore that Φ takes more than one value, say $\Phi(P_0) = r_0$ and $\Phi(P_1) = r_1$, with $r_0 \neq r_1$ for some $P_0, P_1 \in \mathcal{M}$. Then we have

$$V = r_0, \ P_0 - a.s. \tag{60}$$
$$V = r_1, \ P_1 - a.s. \tag{61}$$

and now we define the sets S_0 and S_1 by

$$\begin{aligned} S_0 &= \{\omega \in \Omega | V(\omega) = r_0\}, \\ S_1 &= \{\omega \in \Omega | V(\omega) = r_1\}. \end{aligned}$$

Let us now define a measure P by

$$P = \frac{1}{2} \cdot (P_1 + P_2).$$

It follows from the Mixing Theorem 3.1 that P belongs to $\mathcal{M}$ and from equations (60) and (61) we obviously have

$$P(V = r_0) = P(V = r_1) = \frac{1}{2}.$$

This means however that we cannot possibly have $V = \Phi(P)$, P-a.s. so V does not in fact identify Φ, which contradicts our assumption. ∎

Thus the maximal family contains too many measures to allow us to do any identification at all. The extremal family, on the other hand, is so small that it allows us to identify any parameter, at least in the presence of an SGM. The assumption about the existence of an SGM is rather annoying, and one has a distinct feeling that it should be unnecessary.

Proposition 7.3 *Suppose that $\Pi = \mathcal{E}$ and suppose that the model possesses an SGM. Then every parameter Φ can be identified.*

Proof. Denote the SGM by Z and define the identifier V by

$$V(\omega) = \varphi(Z_\infty), \tag{62}$$

where φ is the structure function for Φ. Then, since every extremal measure is of the form P_z, where

$$P_z(\cdot) = P(\cdot|Z_\infty = z)$$

we have, for every fixed z, P_z-almost surely the equality

$$V = \varphi(Z_\infty) = \varphi(z) = \Phi(P_z). \ ∎$$

At last we will look at some of the consequences for our earlier models. We recall that in all three cases the SGM is given by $Z_t = X_t/t$.

The Wiener model Since this is an extremal model, Proposition 7.3 tells us that all parameters can be identified by equation (62).

The L^2 model A slight variation of the proof of Proposition 7.2 shows that no nontrivial parameter can be identified in this model.

The Gaussian mixture model This model is again to big to allow us to identify any nontrivial parameters. As examples the parameters $\Phi(P_{\alpha,\beta}) = \alpha$ and $\Phi(P_{\alpha,\beta}) = \alpha^2 = \beta$ can both be estimated unbiasedly, but neither can be identified.

8 Information inequalities

We will now derive a Cramér - Rao inequality for the estimation error of an unbiased parameter estimator. Let us therefore assume the existence of an SGM denoted by Z. Let us denote the restriction of the measure P_z to $\mathcal{F}_t$ by $P_{z,t}$, and assume furthermore that

$$P_{z,t} \ll m,$$

where m is some base measure. Now we define a family of Radon-Nikodym derivatives by

$$L_{z,t} = \frac{dP_{z,t}}{dm},$$

and finally we define the Fisher information matrix $I(t, z)$ by

$$I_{i,j}(t,z) = E_z\left[\left(\tfrac{\partial}{\partial z_i}\log L_{z,t}\right)\left(\tfrac{\partial}{\partial z_j}\log L_{z,t}\right)\right] = -E_z\left[\tfrac{\partial^2}{\partial z_i \partial z_j}\log L_{z,t}\right] \qquad (63)$$

where the subindex z denotes integration with respect to P_z.

Proposition 8.1 *Suppose that Y is a given square integrable unbiased t-estimator of the parameter Φ for some t. Then, given the assumptions above we have the following inequality for all $P \in \Pi$.*

$$\begin{aligned} E_P\left[\{Y - \Phi(P)\}^2\right] \quad &\geq \quad E_P\left[\nabla\varphi(Z_\infty) I(t, Z_\infty)^{-1} \nabla\varphi(Z_\infty)^*\right] + \\ &+ \quad E_P\left[\{\varphi(Z_\infty) - \Phi(P)\}^2\right], \end{aligned} \qquad (64)$$

where the information matrix $I(t, z)$ is given by equation (63) and the gradient is regarded as a row vector.

Proof. Since Y is unbiased we have , for all $z \in R$,

$$E_P\left[Y \,|Z_\infty = z\right] = E_z\left[Y\right] = \Phi(P_z) = \varphi(z),$$

i.e.

$$E_P\left[Y \,|Z_\infty\right] = \varphi(Z_\infty), \quad P - a.s.$$

Using this relation a simple calculation gives us

$$E_P\left[\{Y - \Phi(P)\}^2\right] = E_P\left[\{Y - \varphi(Z_\infty)\}^2\right] + E_P\left[\{\varphi(Z_\infty) - \Phi(P)\}^2\right].$$

Furthermore we have, by the Mixing Theorem 3.1,

$$E_P\left[\{Y - \varphi(Z_\infty)\}^2\right] = \int_{R^k} E_z\left[\{Y - \varphi(z)\}^2\right]\nu_P(dz),$$

and since $E_z[Y] = \varphi(z)$ for all z the standard Cramér-Rao inequality gives us

$$E_z\left[\{Y - \Phi(z)\}^2\right] \geq \nabla\varphi(z) I(t,z)^{-1} \nabla\varphi(z)^* \quad \blacksquare$$

The first term in the inequality (64) is due to the fact that at time t we only have access to the information $\mathcal{F}_t$, and this term gets smaller as t increases. The second term gives us a residual eror which is present even if we are allowed to observe X on the closed interval $[0, \infty]$. This term vanishes for all $P \in \Pi$ if and only if the parameter is identifiable.

References

[Baha, 54] Bahadur, R.R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Stat.* **25**, 423 – 462

[Bas & Sco, 83] Basawa, I.V. & Scott, D.J. (1983). *Asymptotic optimal inference for non-ergodic models.* Lecture Notes in Statistics *17*, Springer Verlag.

[Bjö & Joh, 92] Björk, T. & Johansson, B. (1992). Adaptive prediction and reverse martingales. *Stoch. Proc. Appl.* **43**, 191 – 222.

[Bjö & Joh, 93] Björk, T. & Johansson, B. (1993). On Theorems of de Finetti Type for Continuous Time Stochastic Processes. *Scand. J. Statist.* **20**, 289 – 312.

[Eag & Web, 78] Eagleson, G.K & Weber, N.C. (1983). Limit theorems for weakly exchangeably arrays. *Proc. Camb. Phil. Soc.* **84**, 123 – 130.

[Hall & Hey, 80] Hall, P. & Heyde, C.C. (1983). *Martingale limit theory and its applications.* Academic Press.

[Haus & Pard 86] Haussmann, U.G. & Pardoux, E. (1986). Time Reversal of Diffusions. *Ann. Probab.* **14**, 1188 – 1205.

[Küch & Laur 89] Küchler, U & Lauritzen S.L. (1989). Exponential Families, Extreme Point Models and Minimal Space-time Invariant Functions for Stochastic Processes with Stationary and Independent Increments, *Scand. J. Statist.* **16**, 237 – 261.

[Laur 88] Lauritzen S.L. (1988) *Extremal Families and Systems of Sufficient Statistics.* Springer Lecture Notes in Statistics **49**.

[Mar 70] Martin-Löf, P. (1970). Statistiska Modeller. Notes by Rolf Sundberg, Stockholm University.