Stochastic Processes and their Applications 48 (1993) 341–356
North-Holland

# A utility based approach to information for stochastic differential equations

Nicholas G. Polson

*University of Chicago, IL, USA*

Gareth O. Roberts

*Statistical Laboratory, Cambridge, UK*

Received 4 April 1991
Revised 10 August 1992

A Bayesian perspective is taken to quantify the amount of information learned from observing a stochastic process, $X_t$, on the interval $[0, T]$ which satisfies the stochastic differential equation, $\mathrm{d}X_t = S(\theta, t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t$. Information is defined as a change in expected utility when the experimenter is faced with the decision problem of reporting beliefs about the parameter of interest $\theta$. For locally asymptotic mixed normal families we establish an asymptotic relationship between the Shannon information of the posterior and Fisher's information of the process. In particular we compute this measure for the linear case $(S(\theta, t, X_t) = \theta S(t, X_t))$, Brownian motion with drift, the Ornstein–Uhlenbeck process and the Bessel process.

Bayesian inference * local asymptotic normality * Jeffreys prior * Shannon information * Fisher information * entropy

## 1. Introduction

In this paper we consider the amount of information gain from observing a continuous time Markov process. An asymptotic relationship between the Shannon information of the posterior distribution of the parameter of interest $\theta \in \Theta$ and Fisher's information of the process is derived for classes of processes obeying suitable asymptotic normality properties. A utility based approach to quantifying information is taken and we apply it to the case where we have observations from a (possibly time-inhomogeneous) diffusion process. As examples of our approach we consider Brownian motion with drift, the Ornstein–Uhlenbeck process and a Bessel process.

Suppose that we observe a set of observations from a stochastic process on the time interval $[0, T]$ with observation set $X^T = \{X_t, 0 \leqslant t \leqslant T\}$ and natural filtration $\mathscr{F}_T = \sigma\{X_s,$

*Correspondence to*: Dr. G.O. Roberts, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB2 1SB, UK.

$0 \leqslant s \leqslant T\}$. Let $p(\theta)$ and $p(\theta|X^T)$ denote the prior and posterior densities for the parameter of interest. We begin by considering the one-dimensional case although the results will be proved in a multivariate setting. The likelihood function at time $T$, denoted by $L_T(\theta)$, of the SDE

$$\mathrm{d}X_t = S(\theta, t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \tag{1}$$

is given by Girsanov's formula (see, for example, Oksendal, 1985),

$$L_T(\theta) = \exp\left(\int_0^T \frac{S(\theta, t, X_t)}{\sigma^2(t, X_t)}\,\mathrm{d}X_t - \frac{1}{2}\int_0^T \left(\frac{S(\theta, t, X_t)}{\sigma(t, X_t)}\right)^2 \mathrm{d}t\right)$$

with respect to the martingale measure ($\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t$). We restrict ourselves to SDE's with unit diffusion coefficient due to the fact that the diffusion coefficient is determined exactly given $\mathscr{F}_T$.

Theorem 4.5 provides the main result of the paper and shows that, under suitable regularity conditions,

$$\lim_{T\to\infty}\left(E_{X^T,\theta}\left[\log\left(\frac{p(\theta|X^T)}{p(\theta)}\right)\right] - E_{X^T,\theta}[\log|i_T(\theta)|^{1/2}]\right)$$
$$= H(p) - \tfrac{1}{2}\log(2\pi \mathrm{e}), \tag{2}$$

where

$$i_T(\theta) = \left\langle \frac{\mathrm{d}}{\mathrm{d}\theta}\log L_T(\theta)\right\rangle \quad \text{and} \quad H(p) = -\int p(\theta)\log p(\theta)\,\mathrm{d}\theta.$$

Here $E_{X^T,\theta}$ denotes expectation with respect to the joint density of $(X^T, \theta)$, $i_T(\theta)$ is the observed Fisher information or quadratic variation of the score function, and $H(p)$ is the entropy functional. The expected Fisher information, $I_T(\theta)$, is defined as

$$I_T(\theta) = E_{X^T|\theta}[i_T(\theta)] .$$

We consider both the local asymptotic normal case (LAN) (see, for example, Le Cam, 1986) and local asymptotic mixed normal (LAMN) (see, for example, Jeganathan, 1983) families. When our class of process is LAN, the ratio of the observed to expected Fisher information, $i_T(\theta)/I_T(\theta)$, converges in probability to unity and hence $i_T(\theta)$ and $I_T(\theta)$ are interchangeable in (2).

The rest of the paper is outlined as follows. Section 2 introduces the utility based approach to information (Bernardo, 1979a; DeGroot, 1986) that provides the basis for our preposterior analysis. A logarithmic utility leads to the Shannon information of the posterior as a measure of information provided by an experiment, see Lindley (1956). Applications of this measure include: design of linear models (Stone, 1959; Smith and Verdinelli, 1980; design of nonlinear models (Parmigiani and Polson, 1992); characterisation of likelihoods (Polson, 1988); and noninformative priors (Bernardo, 1979b).

Section 3 provides intuition for the asymptotic behaviour of the Shannon information of the posterior in the LAN and LAMN cases. An exact decomposition is also given in the linear case $S(\theta,t,X_t)=\theta S(t,X_t)$. Crucial to our approach is a decomposition of the Shannon information of the posterior in terms of the Le Cam–Ibragimov Z-process (Le Cam, 1986). Section 4 formally derives the asymptotic relationship between the information gain and observed Fisher information for LAMN families. Section 5 illustrates our results with Brownian motion with drift, the Ornstein–Uhlenbeck process and the Bessel process.

## 2. Information

Suppose that an experimenter is faced with a decision problem with decision space $\mathscr{D}$ and utility function $U\colon \Theta\times\mathscr{D}\to\mathbb{R}$. The experimental wishes to maximise expected utility. Let $Y$ be a set of random observations on $(\Omega,\mathscr{F},P)$. We wish to compare the maximum expected utility with and without information about $Y$. Without observing data the optimal decision and the associated expected utility is given by $\sup_{d\in D}E_\theta[U(\theta,d)]$. Having observed data, the relevant maximisation is $\sup_{d\in D}E_{\theta|Y}[U(\theta,d)]$ where $E_{\theta|Y}$ denotes posterior expectation. Therefore, from a preposterior perspective, the expected change in expected utility, $E_Y[I(Y)]$, can be defined by (DeGroot, 1986)

$$E_Y[I(Y)]=E_Y\left[\sup_{d\in D}E_{\theta|Y}[U(\theta,d)]\right]-\sup_{d\in D}E_\theta[U(\theta,d)]\ , \tag{3}$$

where $E_\theta$ and $E_Y$ denote expectation with respect to $p(\theta)$ and $p(Y)=\int f(Y|\theta)p(\theta)\,\mathrm{d}\theta$, respectively. From a Bayesian perspective (3) quantifies information.

In particular, suppose that the decision is to report a probability density for the random variable $\theta$. We take $D=\mathscr{P}(\theta)$, the space of probability measures on the set $\Theta$, and assume that the utility function is honest, that is

$$\sup_{d\in\mathscr{P}}E_\theta[U(\theta,d)]=E_\theta[U(\theta,p(\theta))]\ .$$

In other words, before observing the data the optimal reported density is your a priori beliefs, $p(\theta)$. Similarly, after seeing data the optimal decision (that is attains $\sup_{d\in D}E_{\theta|Y}[U(\theta,d)]$) is to report your posterior $p(\theta|Y)$. One might view this as the definition of inference from a Bayesian perspective (de Finetti, 1979). The honesty property, together with a local property (that is, $U(\theta,d(\cdot))=U(\theta,d(\theta))$), characterises the family of logarithmic utility functions, $U(\theta,d(\cdot))=A\log d(\theta)+B(\theta)$, where $d(\cdot)$ now denotes the reported density, see Seidler (1958), Bernardo (1979a). The logarithmic utility, therefore, plays a natural role in Bayesian inference. The change in expected utility (3) is then given by

$$E_{Y,\theta}\left[\log\left(\frac{p(\theta|Y)}{p(\theta)}\right)\right], \tag{4}$$

that is, the expected Kullback–Leibler distance (Kullback, 1959) between the posterior, $p(\theta|Y)$ and the prior, $p(\theta)$.

The purpose of this paper is to consider the case where $Y=X^T$, where $X_t$ satisfies (1) and to derive the properties of (4) for large $T$. Let $I(X^T;\Theta)$ denote the information, so

$$I(X^T;\ \Theta)=E_{X^T,\theta}\left[\log\left(\frac{p(\theta|X^T)}{p(\theta)}\right)\right]. \tag{5}$$

For conditions on the measures $P_{X^T,\theta}$, $P_{X^T}$, $P_\theta$ for $I(X^T;\ \Theta)$ to be well-defined see Liptser and Shiryayev (1977).

To give some intuition into the limiting behaviour of this measure of information we consider the usual statistical setup of a sample of size $n$. Suppose that we have a sequence $X^{(n)}=(X_1,\ \ldots,\ X_n)$ of observations such that $X_i=\theta+\varepsilon_i$, where $\varepsilon_i$ are independent and identically distributed and $\theta\in\mathbb{R}$. Then, defining $I(X^{(n)};\ \Theta)$ in a similar fashion to (6), we have, under mild regularity conditions that (Ibragimov and H'asminsky, 1973)

$$I(X^{(n)};\ \Theta)=\tfrac{1}{2}\log\left(\frac{\phi(n)}{2\pi e}\right)+\int p(\theta)\log\left(\frac{I_f^{1/2}}{p(\theta)}\right)\mathrm{d}\theta+\mathrm{o}(1)$$

as $n\to\infty$, where $I_f=\int((f\,')^2/f)\,\mathrm{d}\mu$ is Fisher's information for one observation and $\phi(n)$ is a suitable normalisation sequence, typically $\phi(n)=n$. This result generalises to the multivariate case (Ibragimov and H'asminsky, 1973), and to non-identically distributed observations (Polson, 1992). For example, consider the nonlinear model, $y_i=\eta(x_i,\ \theta)+\varepsilon_i$, where $\varepsilon_i$ are independent and identically distributed, $x_i$ are design points and $\theta\in\mathbb{R}^k$. Then, under suitable regularity conditions,

$$I(X^{(n)};\ \Theta)=\int p(\theta)\log\left(\frac{|\sum_{i=1}^n I_i(\theta)|^{1/2}}{p(\theta)}\right)\mathrm{d}\theta-\tfrac{1}{2}k\log(2\pi e)+\mathrm{o}(1)$$

as $n\to\infty$, where $|\sum_{i=1}^n I_i(\theta)|$ is the determinant of Fisher's information and $I_i(\theta)$ is Fisher's information for the $i$th observation.

For stochastic processes, however, one may learn at different rates about $\theta$, depending on where we are in the state space. Heuristically, in the regular case, where Fisher's information grows linearly in time, we are merely interested in the average rate at which we learn about $\theta$, so that if our stochastic process is stationary and ergodic, it is intuitive that we might asymptotically learn at a rate that is an average taken over the stationary distribution for the process. However, we will also be concerned with the important non-ergodic case.

## 3. Asymptotic Shannon information

Let $\{(\Omega,\ \mathscr{F},\ (\mathscr{F}_t,\ t\geqslant 0),\ P_{X|\theta});\ \theta\in\Theta\}$ be a class of probability spaces indexed by $\theta\in\Theta\subset\mathbb{R}^k$. On each of these spaces define a (possibly time-inhomogeneous) $d$-dimensional diffusion process, $\boldsymbol{X}=\{X_t,\ t\geqslant 0\}$ by the SDE

$$\mathrm{d}\boldsymbol{X}_t=\boldsymbol{S}(\theta,\ t,\ \boldsymbol{X}_t)\,\mathrm{d}t+\mathrm{d}\boldsymbol{B}_t,$$

where $\{\boldsymbol{B}_t,\ t\geqslant 0\}$ is a $d$-dimensional standard Brownian motion. Denote by $P_{X^T|\theta}$ the restriction of $P_{X|\theta}$ to $\mathscr{F}_T$. We impose regularity conditions on the drift function $S(\cdot,\cdot,\cdot)$ to ensure that the $P_{X^T|\theta}$'s are all absolutely continuous with respect to each other for each $t\geqslant 0$. Therefore, assume that

$$P_{X^T|\theta}\left[\int_0^T \|S(\theta,t,X_t)\|^2\,\mathrm{d}t<\infty\right]=1$$

for all finite $T$ and for all $\theta\in\Theta$. For SDE's, we define the likelihood of a process as the Radon–Nikodym derivative of its law, $P_{X^T|\theta}$, with respect to a dominating measure, the natural choice being Wiener measure, $P_W$. The diffusions we consider all have laws that are absolutely continuous with respect to Wiener measure and, under the usual regularity conditions, the likelihood is given by Girsanov's formula

$$\frac{\mathrm{d}P_{X^T|\theta}}{\mathrm{d}P_W}(X^T)=\exp\left(\int_0^T S'(\theta,t,X_t)\,\mathrm{d}X_t-\frac{1}{2}\int_0^T \|S(\theta,t,X_t)\|^2\,\mathrm{d}t\right),$$

where $'$ denotes transpose. Define the marginal law for $X^T$, $P_{X^T}$ on $\{\Omega,\mathscr{F}_T\}$ by

$$P_{X^T}[A]=\int P_{X^T|\theta}[A]p(\theta)\,\mathrm{d}\theta$$

for all events $A\subset\mathscr{F}_T$. More generally, we will write $P_{X^T,\theta}[B]$ for possibly $\theta$ dependent sets $B\subset\mathscr{F}_T\times\Theta$.

In this section, we try to motivate the subsequent results. By Bayes theorem,

$$\frac{p(\theta|X^T)}{p(\theta)}=\frac{\mathrm{d}P_{X^T|\theta}}{\mathrm{d}P_{X^T}}.$$

Therefore, the information gain can be rewritten as

$$I(X^T;\ \Theta)=E_{X^T,\theta}\left[\log\left(\frac{\mathrm{d}P_{X^T|\theta}}{\mathrm{d}P_{X^T}}\right)\right]. \tag{6}$$

Now consider the following identity that holds for all positive definite choices of the matrix $Q_T(\theta)$:

$$\frac{\mathrm{d}P_{X^T}}{\mathrm{d}P_{X^T|\theta}}(X^T)=|Q_T(\theta)|^{-1/2}\int_{\mathbb{R}^k}\frac{\mathrm{d}P_{X^T|\theta+Q_T(\theta)^{-1/2}\alpha}}{\mathrm{d}P_{X^T|\theta}}(X^T)\times p(\theta+Q_T(\theta)^{-1/2}\alpha)\,\mathrm{d}\alpha\,, \tag{7}$$

where $|Q_T(\theta)|$ is the modulus of the determinant of $Q_T(\theta)$. This follows from the definition of $\mathrm{d}P_{X^T}$ and the fact that the transformation $\theta\mapsto\theta+Q_T(\theta)^{-1/2}\alpha$ has Jacobian $|Q_T(\theta)|^{-1/2}$. Now define

$$Z_{T,\theta}(\alpha)=\frac{L_T(\theta+Q_T(\theta)^{-1/2}\alpha)}{L_T(\theta)}=\frac{\mathrm{d}P_{X^T|\theta+Q_T(\theta)^{-1/2}\alpha}}{\mathrm{d}P_{X^T|\theta}}(X^T)\,.$$

Hence, (7) becomes

$$\frac{\mathrm{d}P_{X^T}}{\mathrm{d}P_{X^T|\theta}}(X^T) = |Q_T(\theta)|^{-1/2} \int_{\mathbb{R}^k} Z_{T,\theta}(\alpha)p(\theta+Q_T(\theta)^{-1/2}\alpha)\,\mathrm{d}\alpha\,. \tag{8}$$

This identity is of fundamental use in establishing the asymptotic properties of (6).

Heuristically, by choosing $Q_T(\theta)$ carefully we can obtain a meaningful limit of the process $Z_{T,\theta}(\alpha)$ as $T\to\infty$. Let us for the moment suppose that $Z_{T,\theta}(\alpha) \Rightarrow Z_\theta(\alpha)$, where $\Rightarrow$ denotes weak convergence, and that interchange of limits and uniformity of convergence are valid in the following: provided that $Q_T(\theta)\to\infty$, we have $p(\theta+Q_T(\theta)^{-1/2}\alpha)\to p(\theta)$ as $T\to\infty$ and hence from (8),

$$\frac{|Q_T(\theta)|^{1/2}}{p(\theta)}\frac{\mathrm{d}P_{X^T}}{\mathrm{d}P_{X^T|\theta}}(X^T) \Rightarrow \int_{\mathbb{R}^k} Z_\theta(\alpha)\,\mathrm{d}\alpha\,. \tag{9}$$

Combining (6) and (9), we informally deduce that $I(X^T;\,\Theta)$ asymptotically satisfies

$$\lim_{T\to\infty}\,(I(X^T;\,\Theta) - E_\theta[\log|Q_T(\theta)|^{1/2}])$$

$$= H(p) - E_{X,\theta}\left[\log \int_{\mathbb{R}^k} Z_\theta(\alpha)\,\mathrm{d}\alpha\right], \tag{10}$$

where $H(p)$ is the entropy functional of the prior. We now consider specific families of limiting processes $Z_\theta(\alpha)$.

### 3.1. Local asymptotic normal families

The typical scenario for the limit process $Z_\theta(\alpha)$ is the following: there exists matrices $Q_T(\theta)$ such that

$$\log Z_\theta(\alpha) = \lim_{T\to\infty}\,\log\frac{\mathrm{d}P_{X^T|\theta+Q_T^{-1/2}(\theta)\alpha}}{\mathrm{d}P_{X^T|\theta}}(X^T)$$

$$= \alpha' G^{1/2}(\theta)\Delta - \tfrac{1}{2}\alpha' G(\theta)\alpha\,, \tag{11}$$

where $'$ denotes transpose and $\Delta$ is a random vector and $G(\theta)$ is a possibly $\theta$-dependent random matrix. If $\Delta\sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$ and $\Delta$, $G(\theta)$ are independent then the family of measure $\{P_{X^T|\theta};\ \theta\in\Theta\}$ are said to be *locally asymptotically mixed normal* (LAMN) and if $G(\theta) = \boldsymbol{I}$ they are said to be *locally asymptotically normal* (LAN).

By direct evaluation, we can compute the final term in (10) under (11) as follows:

$$\int_{\mathbb{R}^k} Z_\theta(\alpha)\,\mathrm{d}\alpha = \int_{\mathbb{R}^k} \exp(-\tfrac{1}{2}\alpha' G(\theta)\alpha + \alpha' G^{1/2}(\theta)\Delta)\,\mathrm{d}\alpha$$

$$= (2\pi)^{k/2}|G(\theta)|^{-1/2}\exp(\tfrac{1}{2}\Delta'\Delta)\,.$$

Therefore,

$$E_{X,\theta}\left[\log \int_{\mathbb{R}^k} Z_\theta(\alpha)\,\mathrm{d}\alpha\right] = E_{X,\theta}[\log|G(\theta)|^{-1/2}] + \tfrac{1}{2}E_{X,\theta}[\Delta'\Delta] - \tfrac{1}{2}k\log(2\pi) .$$

If $\Delta \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$ then

$$E_{X,\theta}\left[\log \int_{\mathbb{R}^k} Z_\theta(\alpha)\,\mathrm{d}\alpha\right] = E_{X,\theta}[\log|G(\theta)|^{-1/2}] + \tfrac{1}{2}k\log(2\pi\mathrm{e}) ,$$

which implies from (10) that

$$\lim_{T\to\infty} (I(X^T;\, \Theta) - E_\theta[\log|Q_T(\theta)|^{1/2}])$$

$$= H(p) + E_{X,\theta}[\log|G(\theta)|^{1/2}] - \tfrac{1}{2}k\log(2\pi\mathrm{e}) . \qquad (12)$$

In the LAN case, this reduces to

$$\lim_{T\to\infty} (I(X^T;\, \Theta) - E_\theta[\log|Q_T(\theta)|^{1/2}]) = H(p) - \tfrac{1}{2}k\log(2\pi\mathrm{e}) . \qquad (13)$$

The formal asymptotics of the above are derived in Section 4.

### *3.2. The linear case $S(\theta, t, X_t) = \theta S(t, X_t)$*

The expected information $I(X^T;\, \Theta)$ in the linear case $S(\theta, t, X_t) = \theta S(t, X_t)$ can be evaluated explicitly for all $T$ when a priori $\theta \sim \mathrm{N}(\theta_0, \sigma_0^2)$ where $\theta \in \mathbb{R}$. The likelihood can be taken to be

$$\frac{\mathrm{d}P_{X^T|\theta}}{\mathrm{d}P_W}(X^T) = \exp\left(\theta \int_0^T S'(t, X_t)\,\mathrm{d}X_t - \tfrac{1}{2}\theta^2 \int_0^T \|S(t, X_t)\|^2\,\mathrm{d}t\right).$$

The posterior is directly computable, via Bayes Theorem, as

$$p(\theta|X^T) \sim \mathrm{N}\left(A^{-1}(T)\left(\int_0^T S'(t, X_t)\,\mathrm{d}X_t + \theta_0\tau_0^2\right), A^{-1}(T)\right), \qquad (14)$$

where $A(T) = \int_0^T \|S(t, X_t)\|^2\,\mathrm{d}t + \tau_0^2$ and $\tau_0^2 = \sigma_0^{-2}$. The information $I(X^T;\, \Theta)$ is analytically computable, for all values of $T$, as follows: first, note that

$$I(X^T;\, \Theta) = -E_{X^T}[H(p(\cdot\,|X^T))] + H(p) .$$

Secondly, the entropy of a univariate normal with arbitrary mean and variance $s_1^2$, is given by $\frac{1}{2}\log(2\pi\mathrm{e}) + \log s_1$. Following (14), we obtain

$$I(X^T;\, \Theta) = \tfrac{1}{2}E_{X^T,\theta}\left[\log\left(\int_0^T \|S(t, X_t)\|^2\,\mathrm{d}t + \tau_0^2\right)\right] + H(p) - \tfrac{1}{2}\log(2\pi\mathrm{e}) \qquad (15)$$

for all $T$. Here the observed Fisher information is $i_T(\theta)=\int_0^T \|S(t, X_t)\|^2 \, dt$. The decomposition is equivalent to (12) as long as $i_T(\theta)/I_T(\theta) \to G(\theta)$ and $I_T(\theta) \to \infty$ as $T \to \infty$.

### 3.3. A continuous time analogue of Jeffreys prior

We can re-express (12) as follows:

$$I(X^T;\ \Theta) = \int_\Theta p(\theta) \log\left(\frac{|Q_T(\theta)|^{1/2}}{p(\theta)}\right) d\theta + \int_\Theta p(\theta)\, E_{X|\theta}[\log |G(\theta)|^{1/2}] + o(1) \tag{16}$$

as $T \to \infty$. It is interesting to note that when $|Q_T(\theta)|^{1/2} \exp(E_{X|\theta}[\log|G(\theta)|^{1/2}]) \in L^1(\Theta)$, the right hand side is bounded by

$$\log \int |Q_T(\theta)|^{1/2} \exp(E_{X|\theta}[\log|G(\theta)|^{1/2}])\, d\theta$$

with equality if and only if

$$p^\dagger(\theta) \propto |Q_T(\theta)|^{1/2} \exp(E_{X|\theta}[\log |G(\theta)|^{1/2}]) \,. \tag{17}$$

Therefore, the experimenter who expects to learn the most from experimentation has prior beliefs given by (17). We will see that the natural choice for $Q_T(\theta)$ is Fisher information $I_T(\theta)$. Moreover, in the LAN case $|G(\theta)|=1$ and in this case we obtain a continuous time analogue of Jeffreys prior, that is $p^\dagger(\theta) \propto |I_T(\theta)|^{1/2}$. It should be noted that the use of such automatic rules for prior specification should be taken with great caution and that Jeffreys never proposed their use in the context of continuous time models.

## 4. Formal asymptotics

**Theorem 4.1.** *Suppose that $Z_{T,\theta}(\alpha)$ converges weakly to $Z_\theta(\alpha)$ in $P_{X|\theta}$ measure for all $\theta \in \Theta$ and that the following conditions hold:*

(A1) *The prior density $p(\theta)$ is uniformly Holder continuous; there exist positive constants $k_1$, $\varepsilon_1$ such that*

$$|p(\theta_1) - p(\theta_2)| \leqslant k_1 |\theta_1 - \theta_2|^{\varepsilon_1} \quad \forall \theta_1, \theta_2 \in \Theta \,.$$

(A2) *The following uniformity conditions are satisfied:*

(i) *There exists a constant $k_2$ such that $\forall \theta$, $\theta + Q_T(\theta)^{-1/2}\alpha_1$, $\theta + Q_T(\theta)^{-1/2}\alpha_2 \in \Theta$,*

$$E_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1} \left( \int_0^T \|S(\theta + Q_T(\theta)^{-1/2}\alpha_2, t, X_t) - S(\theta + Q_T(\theta)^{-1/2}\alpha_1, t, X_t)\|^2 \, dt \right) \leqslant k_2 |\alpha_2 - \alpha_1|^2 \,.$$

(ii) *There exists positive constants* $\varepsilon_2$, $k_3$ *such that* $Q_T(\theta)^{1/2} \geqslant k_3 T^{\varepsilon_2}\mathbf{1}\mathbf{1}'$ $\forall T \geqslant 0$, $\theta \in \Theta$, *where* **1** *is a row vector of ones.*

(A3) *For each* $\theta$, *one of the following holds: either,*

$$A^T \sup_{|\alpha| \geqslant A} Z_{T,\theta}(\alpha) \to 0$$

*in* $P_{X^T|\theta}$*-measure as* $A \to \infty$ *for all* $T > 0$, *or*

$$\sup_{\theta \in \Theta} p(\theta) < \infty\,, \quad \textit{and} \quad \int_{|\alpha| > A} Z_{T,\theta}(\alpha)\, \mathrm{d}\alpha \to 0\,,$$

*in* $P_{X^T|\theta}$*-measure as* $A \to \infty$.

The proof of this result requires the following lemma.

**Lemma 4.1.** *Suppose* (A2) *holds, then* $Z_{T,\theta}(\alpha)$ *satisfies the uniform Lipschitz condition,*

$$E_{X^T|\theta}[\,|Z_{T,\theta}(\alpha_1) - Z_{T,\theta}(\alpha_2)|\,] \leqslant c\,|\alpha_2 - \alpha_1|$$

*for some constant* $c$, *uniformly in* $\theta$, $T$ *and* $\alpha_1$, $\alpha_2 \in \mathbb{R}^k$.

**Proof.** Writing $J(\theta, \alpha, T) = \log Z_{T,\theta}(\alpha)$, then $E_{X^T|\theta}[\,|Z_{T,\theta}(\alpha_1) - Z_{T,\theta}(\alpha_2)|\,]$ can be rewritten as

$$E_{X^T|\theta}[(\mathrm{e}^{J(\theta,\alpha_1,T)} - \mathrm{e}^{J(\theta,\alpha_2,T)})I(A_1)] + E_{X^T|\theta}[(\mathrm{e}^{J(\theta,\alpha_1,T)} - \mathrm{e}^{J(\theta,\alpha_2,T)})I(\bar{A}_1)]\,,$$

where $I(A_1)$ is the indicator of the set $A_1 = \{Z_{T,\theta}(\alpha_1) - Z_{T,\theta}(\alpha_2) < 0\}$ and $\bar{A}_1$ is the complement of $A_1$. This can be bounded by

$$\begin{aligned} &E_{X^T|\theta}[\mathrm{e}^{J(\theta,\alpha_1,T)}(J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T))^+] \\ &\quad + E_{X^T|\theta}[\mathrm{e}^{J(\theta,\alpha_2,T)}(J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T))^+]\,, \end{aligned}$$

where $(\cdot)^+$ denotes positive part. Therefore,

$$\begin{aligned} &E_{X^T|\theta}[\,|Z_{T,\theta}(\alpha_1) - Z_{T,\theta}(\alpha_2)|\,] \\ &\quad \leqslant E_{X^T|\theta}[(Z_{T,\theta}(\alpha_1) + Z_{T,\theta}(\alpha_2))\,|J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T)|\,]\,. \end{aligned} \tag{18}$$

Since $Z_{T,\theta}(\alpha_1) = \mathrm{d}P_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1} / \mathrm{d}P_{X^T|\theta}$ we have

$$\begin{aligned} &E_{X^T|\theta}[Z_{T,\theta}(\alpha_1)\,|J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T)|\,] \\ &\quad = E_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1}[\,|J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T)|\,]\,. \end{aligned}$$

Similarly for the second term on the right hand side of (18). We will prove the uniform Lipschitz condition for these two terms. Now,

$$\begin{aligned} &E_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1}[\,|J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T)|\,] \\ &\quad \leqslant \tfrac{1}{2} E_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1}[K(\theta, \alpha_1, \alpha_2, T)] + E_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha_1} \end{aligned}$$

$$\times \left[ \left| \int_0^T (S'(\theta + Q_T(\theta)^{-1/2}\alpha_2, t, X_t) - S'(\theta + Q_T(\theta)^{-1/2}\alpha_1, t, X_t)) \, \mathrm{d}B_t \right| \right],$$

where

$$K(\theta, \alpha_1, \alpha_2, T)$$
$$= \int_0^T \| S(\theta + Q_T(\theta)^{-1/2}\alpha_2, t, X_t) - S(\theta + Q_T(\theta)^{-1/2}\alpha_1, t, X_t) \|^2 \, \mathrm{d}t .$$

In turn, by Cauchy–Schwarz and the isometry for Brownian integrals,

$$E_{X^T | \theta + Q_T(\theta)^{-1/2}\alpha_1} [ | J(\theta, \alpha_1, T) - J(\theta, \alpha_2, T) | ]$$
$$\leqslant (E_{X^T | \theta + Q_T(\theta)^{-1/2}\alpha_1} [K(\theta, \alpha_1, \alpha_2, T)])^{1/2}$$
$$+ \tfrac{1}{2} E_{X^T | \theta + Q_T(\theta)^{-1/2}\alpha_1} [K(\theta, \alpha_1, \alpha_2, T)] .$$

However, using (A2) (i),

$$E_{X^T | \theta + Q_T(\theta)^{-1/2}\alpha_1} [K(\theta, \alpha_1, \alpha_2, T)] \leqslant k_2 | \alpha_1 - \alpha_2 |^2 .$$

Hence, from (18), we have

$$E_{X^T | \theta} [ | Z_{T,\theta}(\alpha_1) - Z_{T,\theta}(\alpha_2) | ] \leqslant (\tfrac{1}{2} k_2 + k_2^{1/2}) | \alpha_2 - \alpha_1 |$$

for $|\alpha_2 - \alpha_1| < 1$. By the triangle inequality the above identity holds globally as well, completing the proof. □

**Proof of Theorem 4.1.** It remains to show that

$$\int Z_{T,\theta}(\alpha) \frac{p(\theta + Q_T(\theta)^{-1/2}\alpha)}{p(\theta)} \, \mathrm{d}\alpha \Rightarrow \int Z_\theta(\alpha) \, \mathrm{d}\alpha ,$$

and that the random variables

$$\log \left( \int Z_{T,\theta}(\alpha) \frac{p(\theta + Q_T(\theta)^{-1/2}\alpha)}{p(\theta)} \, \mathrm{d}\alpha \right)$$

are uniformly integrable, that is for the sets

$$B_k = \left\{ \left| \int Z_{T,\theta}(\alpha) \frac{p(\theta + Q_T(\theta)^{-1/2}\alpha)}{p(\theta)} \, \mathrm{d}\alpha \right| > k \right\}$$

we have

$$E_{X^T | \theta} \left( \log \left( \int Z_{T,\theta}(\alpha) \frac{p(\theta + Q_T(\theta)^{-1/2}\alpha)}{p(\theta)} \, \mathrm{d}\alpha \right) I(B_k) \right)$$

going uniformly to zero as $k \to \infty$ for $\theta \in \Theta$, $T \in [0, \infty)$.

However, these conditions will be satisfied by a straightforward modification of Theorem 3.1 of Ibragimov and H'asminsky (1973), if we can show the uniform Lipschitz condition

$$E_\theta[\,|Z_{T,\theta}(\alpha_1)-Z_{T,\theta}(\alpha_2)|\,]\leqslant C|\alpha_1-\alpha_2|\ .$$

However, this is assured by Lemma 4.1. The full proof will not be duplicated here. □

### *4.1. Local asymptotic normality*

**Definition 4.1** (*local asymptotic normality* (LAN)). The set of measures $\{P_{X^T|\theta},\ \theta\in\Theta\}$ is said to be locally asymptotically normal at $\theta\in\Theta$ if there exists matrices $Q_T(\theta)$ such that the following holds:

$$\log Z_{T,\theta}(\alpha)=\log\frac{\mathrm{d}P_{X^T|\theta+Q_T(\theta)^{-1/2}\alpha}}{\mathrm{d}P_{X^T|\theta}}(X^T)$$

$$=\alpha'\Delta_T(\theta,X_T)-\tfrac{1}{2}\alpha'\alpha+\beta_T(\alpha,X_T,\theta)\ ,$$

where $\beta_T(\alpha,X_T,\theta)\to0$ in $P_{X^T|\theta}$-measure and

$$\Delta_T(\theta,X_T)\Rightarrow\Delta\sim\mathrm{N}(\mathbf{0},\boldsymbol{I})$$

as $T\to\infty$ .

To make sense of LAN in the present context, we need the following notion of differentiation, necessary for the linearisation of $Z_{T,\theta}(\alpha)$.

**Definition 4.2.** A collection of random functions $f(v,t)$ indexed by $t\in[0,T]$, is said to be differentiable in $v$ in probability in $L^2[0,T]$, if there exists a function $\dot f(v,t)$, $t\in[0,T]$ such that

$$\int_0^T\left(\frac{f(v+\delta,t)-f(v,t)}{\delta}-\dot f(v,t)\right)^2\mathrm{d}t\to0$$

in probability.

A vector random function with a vector parameter is said to be differentiable in probability with respect to the parameter, if each component of the random function is differentiable in probability with respect to each component of the parameter. Denote by $\dot S(\theta,t,X_t)$ the $k\times d$ matrix of derivatives in probability.

Suppose that the following additional conditions hold:

(A4) $S(\theta,t,X_t)$ is differentiable in $\theta$ in $P_{X|\theta}$-measure in $L^2[0,T]$, for all $\theta\in\Theta$.

(A5) There exists a class of matrices $\{Q_T(\theta),\ T\geqslant0\}$ of order $k\times k$ such that

$$Q_T(\theta)^{-1/2}\left(\left(\int_0^T\dot S(\theta,t,X_t\right)\dot S'(\theta,t,X_t)\,\mathrm{d}t\right)Q_T(\theta)^{-1/2}\to\boldsymbol{I}$$

in $P_{X|\theta}$-measure as $T \to \infty$, where $\boldsymbol{I}$ is the $k \times k$ identity matrix.

Following the statistical case, it is necessary to introduce the score function. Define, the score function, $U_T(\theta)$, by

$$U_T(\theta) = \frac{\partial}{\partial\theta} \log L_T(\theta) = \int_0^T \dot{\boldsymbol{S}}'(\theta, t, \boldsymbol{X}_t)\, \mathrm{d}\boldsymbol{B}_t\,,$$

where the partial derivative is the derivative in probability as in Definition 4.2. The scaling factor $Q_T(\theta)$ will be taken to be the quadratic variation of the score function:

$$Q_T(\theta) = \langle U_T(\theta) \rangle = E_{X^T|\theta} \int_0^T \dot{\boldsymbol{S}}(\theta, t, \boldsymbol{X}_t)\dot{\boldsymbol{S}}'(\theta, t, \boldsymbol{X}_t)\, \mathrm{d}t\,,$$

that is, $Q_T(\theta) = I_T(\theta)$. The score function $U_T(\theta)$ is a $P_{X|\theta}$-martingale (see Feigin, 1976), heuristically acting as Brownian motion in the time scale of Fisher's information.

We can rewrite $Z_{T,\theta}(\alpha)$ (defined in (11)) in the following way:

$$\begin{aligned}\log Z_{T,\theta}(\alpha) = \alpha' Q_T(\theta)^{-1/2} &\int_0^T \dot{\boldsymbol{S}}'(\theta, t, \boldsymbol{X}_t)\, \mathrm{d}\boldsymbol{B}_t \\ &- \tfrac{1}{2} \int_0^T \|\boldsymbol{S}(\theta + Q_T(\theta)^{-1/2}\alpha, t, \boldsymbol{X}_t) - \boldsymbol{S}(\theta, t, \boldsymbol{X}_t)\|^2\, \mathrm{d}t \\ &+ \int_0^T (\boldsymbol{S}'(\theta + Q_T(\theta)^{-1/2}\alpha, t, \boldsymbol{X}_t) \\ &\qquad - \boldsymbol{S}'(\theta, t, \boldsymbol{X}_t) - \alpha' Q_T(\theta)^{-1/2}\dot{\boldsymbol{S}}'(\theta, t, \boldsymbol{X}_t))\, \mathrm{d}\boldsymbol{B}_t\,,\end{aligned}$$

where $\boldsymbol{B}$ is the $X | \theta$ Brownian motion. Write

$$\log Z_{T,\theta}(\alpha) = J_{T,1}(\theta) + J_{T,2}(\theta) + J_{T,3}(\theta)\,.$$

Now, by (A1) and (A2), $J_{T,3}(\theta)$ tends to 0 as $T \to \infty$, and $J_{T,2}(\theta)$ converges to $\frac{1}{2}\|\alpha\|^2$ in $P_{X|\theta}$-measure. The interesting term is $J_{T,1}(\theta)$. By definition of the score function,

$$J_{T,1}(\theta) = \alpha' Q_T(\theta)^{-1/2} U_T(\theta)\,.$$

The condition (A2) is exactly the right condition to allow the martingale central limit theorem to apply to $J_{T,1}(\theta)$. We give a version applicable to SDE's.

**Theorem 4.2.** *Let $\boldsymbol{Y}_T$ be the Brownian martingale given by, $\boldsymbol{Y}_T = \int_0^T \sigma(t, \boldsymbol{Y}_t)\, \mathrm{d}\boldsymbol{B}_t$, and suppose that there exists matrices $\{Q_T, T \geqslant 0\}$ such that*

$$Q_T^{-1/2} \left( \int_0^T \sigma(t, \boldsymbol{Y}_t)\sigma'(t, \boldsymbol{Y}_t)\, \mathrm{d}t \right) Q_T^{-1/2} \to \boldsymbol{I}$$

*in probability as $T \to \infty$. Then,*

$$Q_T^{-1/2} Y_T \Rightarrow \mathrm{N}_d(\mathbf{0}, \boldsymbol{I}) \quad \text{as } T \to \infty . \qquad \square$$

Recalling condition (A5), we have proved the following:

**Corollary 4.1.** *Under the conditions* (A4) *and* (A5) *the family of measures is LAN with* $Q_T(\theta) = I_T(\theta)$. □

**Theorem 4.3.** *Suppose that* (A1)–(A5) *hold. Then*

$$\lim_{T \to \infty} \left( I(X^T;\, \Theta) - E_\theta[\log |I_T(\theta)|^{1/2}] \right) = H(p) = \tfrac{1}{2} k \log(2\pi e) . \qquad \square \tag{19}$$

*4.2. Locally asymptotic mixed normal*

**Definition 4.3** (*local asymptotic mixed normality* (LAMN)). The set of measures $\{P_{X^T|\theta},\ \theta \in \Theta\}$ is said to be locally asymptotically mixed normal at $\theta \in \Theta$ if there exists matrices $Q_T(\theta)$ such that the following holds:

$$\begin{aligned} \log Z_{T,\theta}(\alpha) &= \log \frac{\mathrm{d}P_{X^T|\theta + Q_T(\theta)^{-1/2}\alpha}}{\mathrm{d}P_{X^T|\theta}} (X^T) \\ &= \alpha' G_T^{1/2}(\theta) \Delta_T(\theta, X_T) - \tfrac{1}{2} \alpha' G_T(\theta) \alpha + \beta_T(\alpha, X_T, \theta) , \end{aligned}$$

where $\beta_T(\alpha, X_T, \theta) \to 0$ in $P_{X^T|\theta}$-measure. Moreover, $G_T(\theta)$ converges a.s. to $G(\theta)$, and

$$(\Delta_T(\theta, X_T), G_T(\theta)) \Rightarrow (\Delta, G(\theta))$$

as $T \to \infty$, where $\Delta \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$ and $G(\theta)$ is a positive definite matrix random variable.

By direct computation we obtain:

**Theorem 4.4.** *If LAMN and* (A1)–(A3) *hold, then*

$$\begin{aligned} &\lim_{T \to \infty} \left( I(X^T;\, \Theta) - E_\theta[\log |I_T(\theta)|^{1/2}] \right) \\ &\quad = H(p) + E_{\theta,X}[\log |G(\theta)|^{1/2}] - \tfrac{1}{2} k \log(2\pi e) . \qquad \square \end{aligned} \tag{20}$$

We note here that typically $G(\theta)$ depends on initial conditions of the process. A useful recombination of the result can be obtained by noting that if $i_T(\theta) / I_T(\theta) \to G(\theta)$ then we have

$$\lim_{T \to \infty} \left( I(X^T;\, \Theta) - E_{X,\theta}[\log |i_T(\theta)|^{1/2}] \right) = H(p) - \tfrac{1}{2} k \log\, (2\pi e)$$

for the LAMN class. The asymptotic Shannon information gain in this case is therefore governed by observed Fisher information rather than expected Fisher information as in Theorem 4.3.

## 5. Examples

We now consider Brownian motion with drift, the Ornstein–Uhlenbeck process, and the Bessel process.

### *5.1. Brownian motion with drift*

Consider the process that is the solution of the stochastic differential equation, $\mathrm{d}X_t = \theta\,\mathrm{d}t + \mathrm{d}B_t$ where $\theta$ is given a prior density $\theta \sim \mathrm{N}(\theta_0, \sigma_0^2)$ where $\tau_0^2 = 1/\sigma_0^2$. Now $X_T$ is a sufficient statistic for $\theta$. Therefore, by Bayes theorem, the posterior $p(\theta | X^T) = p(\theta | X_T) \propto p(X_T \,|\, \theta) p(\theta)$, where $X_T \sim \mathrm{N}(\theta T, T)$ and so as in (14),

$$\theta | X^T \sim \mathrm{N}\left(\frac{X_T + \theta_0 \tau_0^2}{T + \tau_0^2}, \frac{1}{T + \tau_0^2}\right).$$

By (15), we have

$$I(X^T; \Theta) = \tfrac{1}{2}\log(T + \tau_0^2) + H(p) - \tfrac{1}{2}\log(2\pi \mathrm{e})$$

and the limiting result agrees with Theorem 4.3 with $Q_T(\theta) = T$.

### *5.2. Ornstein–Uhlenbeck process*

Consider the Ornstein–Uhlenbeck (1, $\theta$) process, that is, the solution to $\mathrm{d}X_t = \theta X_t\,\mathrm{d}t + \mathrm{d}B_t$, where $B_t$ is a Brownian motion. The likelihood is given by

$$L_T(\theta) = \exp\left(\theta \int_0^T X_t\,\mathrm{d}X_t - \tfrac{1}{2}\theta^2 \int_0^T X_t^2\,\mathrm{d}t\right).$$

The score function by

$$U_T(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log L_T(\theta) = \int_0^T X_t\,\mathrm{d}B_t\,.$$

By the isometry of Brownian integral $(\int_0^T X_t\,\mathrm{d}B_t)^2 = \int_0^T X_t^2\,\mathrm{d}t$ and hence the observed Fisher information is $i_T(\theta) = \int_0^T X_t^2\,\mathrm{d}t$. Therefore, by (15) the expected information gain is given for all $T$ by

$$I(X^T; \Theta) = \tfrac{1}{2} E_{X^T,\theta}[\log(i_T(\theta) + \tau_0^2)] + H(p) - \tfrac{1}{2}\log(2\pi \mathrm{e})\,. \tag{21}$$

where $i_T(\theta)$ is observed Fisher information.

We now discuss the limiting properties of $i_T(\theta)$ as $T \to \infty$. It is convenient to consider the ergodic ($\theta<0$) and transient ($\theta>0$) cases separately due to the asymptotic properties of $X_T$ in each case.

(i) *Ergodic case*, $\theta<0$. The behaviour of observed Fisher information can be determined as follows: the ergodic theorem ensures that,

$$\frac{1}{T}\int_0^T X_t^2 \, \mathrm{d}t \to -\frac{1}{2\theta}$$

in $P_{X|\theta}$-measure. Therefore,

$$\lim_{T\to\infty} (\log i_T(\theta) - \log T) \to \log\left(-\frac{1}{2\theta}\right).$$

The result in (21) agrees with the asymptotic result of Theorem 4.3.

(ii) *Transient case*, $\theta>0$. We can write

$$\mathrm{e}^{-\theta T}X_T - X_0 = \int_0^T \mathrm{e}^{-\theta t} \, \mathrm{d}B_t .$$

Since $E[(\mathrm{e}^{-\theta T}X_T - X_0)^2] = (1-\mathrm{e}^{-2\theta T)}/(2\theta)$, which is bounded, $\mathrm{e}^{-\theta T}X_T$ is a uniformly integrable martingale, and so converges to $Y$, say. Moreover, $Y \sim \mathrm{N}(X_0, 1/(2\theta))$. The behaviour of $i_T(\theta)$ follows from the fact that

$$\frac{\mathrm{e}^{2\theta T}}{2\theta}\int_0^T X_t^2 \, \mathrm{d}t \to Y^2$$

in $P_{X|\theta}$-measure as $T\to\infty$. Therefore,

$$\mathrm{e}^{2\theta T}\int_0^T X_t^2 \, \mathrm{d}t \Rightarrow \chi_1^2(2\theta X_0^2)$$

where $\chi_1^2(\cdot)$ denotes a non-central chi-squared distribution.

The ratio of observed to expected Fisher information, $i_T(\theta)/I_T(\theta)$, converges weakly to

$$G(\theta) = (1+2\theta X_0^2)^{-1}\chi_1^2(2\theta X_0^2) .$$

This model falls into the LAMN class as $Y$ and $\Delta_T$ are asymptotically independent. Therefore the behaviour of asymptotic Shannon information is given by (15).

### 5.3. Bessel process

Another non-regular situation occurs with the following Bessel($k$) process which is defined as the solution of the SDE

$$\mathrm{d}X_t = \frac{k-1}{2(X_t-\theta)} \, \mathrm{d}t + \mathrm{d}B_t .$$

It is straightforward to show that the observed Fisher information is given by

$$i_T(\theta)=\tfrac{1}{4}(k-1)^2\int_0^T \frac{1}{(X_t-\theta)^4}\,\mathrm{d}t\,. \tag{22}$$

The behaviour depends on the starting value $X_0$ and whether $1<k<3$ or $k\geqslant 3$. In the case when $1<k<3$ the Bessel process is recurrent, so that when it hits $\theta$ it is easy to check that $i_T(\theta)$ becomes almost surely infinite, giving infinite information in a finite time interval. Therefore no analogue of Theorem 4.4 holds. In the case when $k\geqslant 3$ the Bessel process is transient. Provided that $X_0\neq\theta$ we have $\lim_{T\to\infty}I_T(\theta)<\infty$ and the information is bounded for all $T$ violating (A2) (ii) in Theorem 4.1.

## Acknowledgement

The authors would like to thank Dean Foster and the referee for their comments and suggestions.

## References

J.M. Bernardo, Expected information as expected utility, Ann. Statist. 7 (1979a) 686–690.

J.M. Bernardo, Reference posterior distributions for Bayesian inference (with discussion), J. Roy. Statist. Soc. Ser. B 41 (1979b) 113–148.

M.H. DeGroot, Changes in utility as information, in: L. Daboni et al., eds., Recent Developments in the Foundations of Utility and Risk Theory (Dordrecht, Reidel, 1986).

B. De Finetti, In discussion of Bernardo (1979b), J. Roy. Statist. Soc. Ser. B 41 (1979) 135.

P. Feigin, Maximum likelihood estimation for continuous time stochastic processes, Adv. Appl. Probab. 8 (1976) 712–736.

I.A. Ibragimov and R.Z. H'asminsky, On the information contained in a sample about a parameter, 2nd Internat. Symp. on Inform. Theory (1973) pp. 295–309.

P. Jeganathan, Some asymptotic properties of risk functions when the limit of the experiment is mixed normal, Indian Statist. J. A (1983) 66–87.

S. Kullback, Information Theory and Statistics (Wiley, New York, 1959).

L. Le Cam, Asymptotic Methods in Statistical Decision Theory (Springer, Berlin, 1986).

D.V. Lindley, On the measure of information provided by an experiment, Ann. Statist. 27 (1956) 986–1005.

R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes II (Springer, Berlin, 1977).

B. Oksendal, Stochastic Differential Equations (Springer, Berlin, 1985).

G. Parmigiani and N.G. Polson, Bayesian design for random walk barriers, in: J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, eds., Bayesian Statistics, Vol. 4 (1992) pp. 715–721.

N.G. Polson, Some Bayesian perspectives on statistical modelling, Ph.D. thesis, Univ. of Nottingham (Nottingham, 1988).

N.G. Polson, On the expected amount of information from a nonlinear model, J. Roy. Statist. Soc. Ser. B 54 (1992) 889–896.

A.F.M. Smith and I. Verdinelli, A note on Bayes designs for inference using a hierarchical linear model, Biometrika 67 (1980) 613–619.

M. Stone, Application of a measure of information to the design and comparison of regression experiments, J. Roy. Statist. Soc. Ser. B 21 (1959) 55–70.