Deriving Bayesian and frequentist estimators from time-invariance estimating equations: a unifying approach

Antonietta Mira
Università della Svizzera italiana, Lugano and University of Insubria, Como

Adrian Baddeley
Curtin University, Perth

BFF4, Harvard University, May 2, 2017

INDEX

- Estimating equations
- Time-invariance estimating equations
- Examples of **frequentist** time-invariance estimators
- Examples of **Bayesian** time-invariance estimators
- Godambe-Heyde optimality criterion
- Optimality of MLE and Bayesian estimators
- Discussion

Consider a system of random variables $X = (X_i : i = 1, \cdots, M)$ with a complex joint distribution, $f(x|\theta)$, governed by parameter θ

Our aim is to estimate θ given data x and possibly prior $\nu(\theta)$

Maximum likelihood

$\hat{\theta} = \text{argmax}_\theta f(x, \theta)$

Method of moments

solve for θ in $\mathbb{E}_{X|\theta} S(X) = S(x)$

Bayesian estimator with absolute error loss function

$\hat{\theta} = \text{argmax}_\theta \pi(\theta|x)$

Bayesian estimator with squared error loss function

$\hat{\theta} = \mathbb{E}_{\theta|x}[\theta]$

Estimating equations

solve for θ in $g(x, \theta) = 0$

UNBIASED ESTIMATING EQUATIONS

We call

$$g(x, \theta) = 0$$

an unbiased **frequentist** estimating equation if

$$\mathbb{E}_{X|\theta}[g(X, \theta)] = 0, \qquad \forall \theta$$

the estimator is not necessarily unbiased but is usually consistent

Similarly, define an unbiased **Bayesian** estimating equation if

$$\mathbb{E}_{\theta|x}[g(x, \theta)] = 0, \qquad \forall x$$

Finally, define an unbiased **general** estimating equation if

$$\mathbb{E}_{\theta,x}[g(x, \theta)] = 0$$

Example:

In the **method of moments** we solve

$$\mathbb{E}_{X|\theta} S(X) = S(x)$$

This is equivalent to solve the E.E.: $g(x, \theta) = 0$ where

$$g(x, \theta) = \mathbb{E}_{X|\theta} S(X) - S(x)$$

This is an **unbiased frequentist E.E.**:

$$\begin{aligned}
\mathbb{E}_{X|\theta}\{g(X,\theta)\} &= \mathbb{E}_{X|\theta}\{\mathbb{E}_{X|\theta}[S(X)] - S(X)\} \\
&= \mathbb{E}_{X|\theta}[S(X)] - \mathbb{E}_{X|\theta}[S(X)] \\
&= 0
\end{aligned}$$

Example:

The **Bayesian estimator** for $h(\theta)$:

$$\hat{h}(\theta) = \mathbb{E}_{\theta|x} h(\theta)$$

can be derived from an E.E.: $g(x, \theta) = 0$ where

$$g(x, \theta) = \mathbb{E}_{\theta|x} h(\theta) - h(\theta)$$

This is an **unbiased Bayesian E.E.**:

$$\begin{aligned} \mathbb{E}_{\theta|x}\{g(x, \theta)\} &= \mathbb{E}_{\theta|x}\{\mathbb{E}_{\theta|x}[h(\theta)] - h(\theta)\} \\ &= \mathbb{E}_{\theta|x}[h(\theta)] - \mathbb{E}_{\theta|x}[h(\theta)] \\ &= 0 \end{aligned}$$

Example:

MLE for Exponential Family

Assume the probability density f has the form

$$f(x|\theta) = Z(\theta) \exp\{\theta^{\mathsf{T}} V(x)\}, \quad x \in \mathcal{X}$$

$\theta \in \Theta \subseteq \mathbb{R}^p$ is the canonical parameter

$V : \mathcal{X} \to \mathbb{R}^p$ is the canonical sufficient statistics

Given data x, under regularity conditions, MLE is the solution of

$$U(\hat{\theta}, x) = 0$$

where $U(\theta, x) = \frac{\partial}{\partial \theta} \log f(\theta, x)$ is the *score function*

For a regular exponential family

$$U(\theta, x) = V(x) - \mathbb{E}_{X|\theta}[V(X)]$$

MLE is the solution of

$$\mathbb{E}_{X|\theta}[V(X)] = V(x)$$

i.e. method of moments applied to V and the score function is an unbiased frequentist estimating function:

$$\mathbb{E}_{X|\theta}[U(\theta, X)] = 0$$

CHALLENGE

The normalizing constant $Z(\theta)$ may be an intractable function of θ so that the likelihood cannot be maximized explicitly:

$$\begin{aligned}\frac{1}{Z(\theta)} &= \int \exp\{\theta^T V(x)\} dx \\ \mathbb{E}_{X|\theta}[V(X)] &= \frac{\partial}{\partial\theta} \log Z(\theta)\end{aligned}$$

and the score function cannot be evaluated

Possible remedies

approximate $Z(\theta)$ analytically
approximate $Z(\theta)$ by simulation
replace score by another estimating function

Example:

Pseudo-Likelihood Estimator

Suppose $X = (X_1, \cdots, X_M)$ is a vector of discrete random variables
Define

$$\mathrm{PL}(\theta, x) = \prod_{i=1}^{M} \mathbb{P}_\theta\{X_i = x_i | X_{-i} = x_{-i}\}$$

Pseudo-Likelihood estimator:

$$\tilde{\theta} = \mathrm{argmax}_{\theta \in \Theta}\, \mathrm{PL}(\theta, x)$$

Rationale:

- PLHD good approximation to the LHD if X_i approx. independent
- PLHD usually very simple and tractable
- MPLE is faster to compute than MLE by a factor of $10^2 - 10^4$
- MPLE is consistent and asymptotically normal
- MPLE can be biased and inefficient in small samples

Example:

Pseudo-Likelihood Estimator for Exponential Family

Suppose $X = (X_1, \cdots, X_M)$ has regular exponential family density
Then the MPLE normal equations are:

$$\frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{X|\theta}[V(X)|X_{-i} = x_{-i}] = V(x)$$

This is a frequentist unbiased EE

Example:

Bayesian Estimator for Exponential Family

If the posterior belongs to the exponential family

$$\pi(\theta|t(x)) = Z(t)^{-1} \exp\{t^T \phi(\theta)\}$$

ϕ = sufficient statistic
$t = t(x)$ = canonical parameter
$W(\theta, t)$ = derivative of log-posterior $= \frac{\partial}{\partial t} log\ \pi(\theta, t)$ = B-Score
For a regular exponential family:

$$W(\theta, t) = \phi(\theta) - \mathbb{E}_{\theta|x}[\phi(\theta)]$$

BAYES ESTIMATOR of $\phi(\theta)$: solution of $W(\theta, t) = 0$
$W(\theta, t)$ is an unbiased Bayesian estimating function:

$$\mathbb{E}_{\theta|X}[W(\theta, t(x))] = 0$$

Time-invariance Estimating Equations

(frequentist)

Baddeley (1995), Kessler & Sørensen (1995), Hansen & Scheinkman (1995)

A natural way to understand and simulate a highly structured random process X, with sample space $\mathcal{X}$, is as the equilibrium distribution of a Markov process $\mathbf{Y} = (Y_n, n = 1, 2 \cdots)$ with states in $\mathcal{X}$

X	$\mathbf{Y}$
random pack of cards	shuffling process
Poisson distribution	birth-and-death process
Markov random field	Gibbs sampler
Gibbs point process	spatial birth-death process

- $\mathbf{Y}_n$ is a mathematical fiction in the original context
- The "time" variable, n, is not part of the original system X
- There are many alternative choices of $\mathbf{Y}$ for each X (MCMC)

INFINITESIMAL GENERATOR of Markov Chains:

Given a Markov chain Y_n with values on $\mathcal{X}$, the generator $\mathcal{A}$ of Y_n is

$$\begin{aligned}(\mathcal{A}S)(x) &= \mathbb{E}[S(Y_{n+1}) - S(Y_n) \mid Y_n = x] \\ &= \mathbb{E}[S(Y_{n+1}) \mid Y_n = x] - S(x)\end{aligned}$$

for any $S : \mathcal{X} \to \mathbb{R}^d$

If $\mathcal{X}$ is finite and Y has transition probabilities $P(x, y)$:

$$\begin{aligned}(\mathcal{A}S)(x) &= \sum_{y \in \mathcal{X}} [S(y) - S(x)] P(x, y) \\ &= \sum_{y \in \mathcal{X}} [S(y)] P(x, y) - S(x)\end{aligned}$$

So $\mathcal{A} = P - I$

STANDARD RESULT:

If $X \sim \pi$ (the stationary distribution of the MC) then

$$\begin{aligned}
\mathbb{E}_\pi[(\mathcal{A}S)(X)] &= \mathbb{E}_\pi[\mathbb{E}[S(Y_{n+1}) - S(Y_n)|Y_n = X]] \\
&= \mathbb{E}_\pi[\mathbb{E}[S(Y_{n+1})|Y_n = X]] - \mathbb{E}_\pi[S(X)] \\
&= \mathbb{E}_\pi[S(Y_{n+1})] - \mathbb{E}_\pi[S(Y_n)] \\
&= 0
\end{aligned}$$

for essentially all S

If $\mathcal{X}$ is finite and π is stationary distribution for Y_n i.e. $\pi P = \pi$ then

$$\pi\mathcal{A} = \pi(P - I) = \pi P - \pi = 0$$

NEW (unifying) ESTIMATION FRAMEWORK

For each θ, represent the distribution of X under θ, $f(x|\theta)$, as the equilibrium distribution of some $\mathbf{Y}^{(\theta)} = (Y_n^{(\theta)})$
Let $\mathcal{A}_\theta$ = infinitesimal generator of $\mathbf{Y}^{(\theta)}$
Choose a statistic $S = S(X)$
Given observed data $X = x$, estimate θ as the solution (if $\exists$!) $\hat{\theta}_T$ of

$$(\mathcal{A}_\theta\, S)\,(x) = 0$$

This is an unbiased frequentist estimating equation for θ
$\hat{\theta}_T$ = **TIME-INVARIANCE ESTIMATOR**

Degrees of freedom: Y and S

X	$\mathbf{Y}$	$S(x)$	$\hat{\theta}_T$
anything	i.i.d.	any	M-of-M
exp. family	i.i.d.	$V(x)$	MLE
exp. family	Gibbs sampler	$V(x)$	MPLE
exp. family	Langevin Diff	$V(x)$	MLE
Poisson dist	Birth-Death	x	MLE
continuous r.v.	OU diffusion	any	Variational Estimator

X	$\mathbf{Y}$	$S(x)$	$\hat{\theta}_T$
anything	i.i.d.	any	**M-of-M**
exp. family	i.i.d.	$V(x)$	**MLE**
exp. family	Gibbs sampler	$V(x)$	MPLE
exp. family	Langevin Diff	$V(x)$	MLE
Poisson dist	Birth-Death	x	MLE
continuous r.v.	OU diffusion	any	Variational Estimator

Example: Method of moments and MLE

Let X be a random element of any space $\mathcal{X}$
Let $\mathbf{Y}_n$ be **i.i.d.** copies of X_n
Then $\mathbf{Y}$ has infinitesimal generator

$$(\mathcal{A}S)(x) = \mathbb{E}[S(X_{n+1}) - S(X_n) \mid X_n = x] = \mathbb{E}\left[S(X)\right] - S(x)$$

and $\hat{\theta}_T$ is the solution of

$$\mathbb{E}_\theta\left[S(X)\right] = S(x)$$

i.e. this leads to the method-of-moments estimator

If X has a distribution from an exponential family
take S = canonical sufficient statistics
and Y = iid sampling from $f(x|\theta)$
then time-invariance EE gives the normal equations for the MLE

X	**Y**	$S(x)$	$\hat{\theta}_T$
anything	i.i.d.	any	M-of-M
exp. family	i.i.d.	$V(x)$	MLE
exp. Family	Gibbs sampler	$V(x)$	**MPLE**
exp. family	Langevin Diff	$V(x)$	MLE
Poisson dist	Birth-Death	x	MLE
continuous r.v.	OU diffusion	any	Variational Estimator

Example: Maximum Pseudo-LHD for Exp. Family

Given a finite system of r.v. $X = (X_1, \cdots, X_M)$
Take Y to be single site random scan Gibbs sampler
The generator of Y is:

$$(\mathcal{A}S)(x) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{X|\theta}[S(X)|X_{-i} = x_{-i}] - S(x)$$

If we choose $S = V$ and
$f(x|\theta)$ belongs to the exponential family,
then the time-invariance EE is equivalent to the
maximum pseudo-likelihood normal equations

X	$\mathbf{Y}$	$S(x)$	$\hat{\theta}_T$
anything	i.i.d.	any	M-of-M
exp. family	i.i.d.	$V(x)$	MLE
exp. family	Gibbs sampler	$V(x)$	MPLE
exp. family	Langevin Diff	$V(x)$	MLE
Poisson dist	Birth-Death	x	**MLE**
continuous r.v.	OU diffusion	any	Variational Estimator

Example: MLE and MoM for Poisson distribution

Let $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, unknown
Thus $\mathcal{X} = \mathbb{N} = \{0, 1, 2, \ldots\}$ and $\Theta = \mathbb{R}_+$
We observe a single realisation x of X
Let Y_n be the **birth-death process** with transition rates

$$r(x,\, x+1) = \lambda \qquad \text{and} \qquad r(x,\, x-1) = x$$

Then for any $S : \mathbb{N} \to \mathbb{N}$

$$(\mathcal{A}_\lambda S)(x) = \lambda\,[S(x+1) - S(x)] + x\,[S(x-1) - S(x)]$$

Take $S(x) \equiv x$; then $\qquad (\mathcal{A}_\lambda S)(x) = \lambda - x$
and the time-invariance estimator is

$$\hat{\lambda}_T = x$$

which is also the MLE and method-of-moments estimator

X	$\mathbf{Y}$	$S(x)$	$\hat{\theta}_T$
M. random field	Gibbs sampler	$V(x)$	MPLE
M. random field	block-update Gibbs	$V(x)$	Generalized MPLE
point process	spatial birth-death	$V(x)$	MPLE
point process	multiple birth-death	$V(x)$	Generalized MPLE
point process	spatial birth-death	any	Takacs-Fiksel
random censoring	random lifetime	e.d.f.	Reduced sample estimator

TYPE I time-invariance EE:

$$(\mathcal{A}_\theta\, S)\,(x) = \mathbb{E}[S(X_{n+1}) - S(X_n) \mid X_n = x]$$

where $S : \mathcal{X} \to \mathbb{R}^d$

TYPE II time-invariance EE:

From the connection with Stein-Chen method of distributional approximation we can define

$$(\mathcal{F}T)(x) = \mathbb{E}[T(X_n, X_{n+1}) | X_n = x] \qquad (1)$$

where $T : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is *antisymmetric*: $T(x,y) = -T(y,x)$ and the Markov chain is in discrete time and reversible wrt π

Then we have

$$\mathbb{E}_\pi(\mathcal{F}T)(X) = 0$$

so this is another source of frequentist unbiased EE

Type II is a generalization of Type I: take $T(x,y) = S(y) - S(x)$

Missing observations

observable data $= X$
unobservable data $= Z$
full data $= (X, Z)|\theta \sim f(X, Z|\theta)$
the time-invariance type I and II EE becomes:

$$\begin{aligned}(\mathcal{A}_\theta S)(x,z) &= 0 \\ (\mathcal{F}_\theta T)(x,z) &= 0\end{aligned}$$

$x =$ **observed** realization of X
$z =$ **unobservable** realization of Z

These equations give rise to usable estimators
only if *the left hand sides do not depend on z*

This can be obtained taking Y **to be a Gibbs sampler**:

$$(x, z) \xrightarrow{\mathbb{P}(z' | x)} (x, z') \xrightarrow{\mathbb{P}(x' | z')} (x', z')$$

$$\begin{aligned} (\mathcal{A}_\theta S)(x, z) &= 0 \\ \int_{\mathcal{Z}} \int_{\mathcal{X}} \left[S(x', z') - S(x, z) \right] \mathbb{P}_\theta(\,\mathrm{d}x' \mid z')\, \mathbb{P}_\theta(\,\mathrm{d}z' \mid x) &= 0 \\ \mathbb{E}_\theta[U_\theta(Z) \mid x] - S(x, z) &= 0 \end{aligned}$$

where $\qquad U_\theta(z) = \mathbb{E}_\theta[S(X, z) \mid z]$
This is an unbiased EE wrt the joint: $(X, Z)|\theta$
If $S(x, z) = V(x)$ in an exp. fam. then the EE becomes:

$$\mathbb{E}_\theta[U_\theta(Z) \mid x] = V(x)$$

and it is an unbiased EE wrt the marginal: $X|\theta$

We may also want to predict Z from X

Then take $S(x, z) = V(z)$
It follows that:

$$U_\theta(z) = \mathbb{E}_\theta[S(X, z) \mid z] = V(x)$$

The EE becomes:

$$\mathbb{E}_\theta[V(Z) \mid x] = V(z)$$

and is an unbiased EE wrt the conditional: $Z|x, \theta$

Bayesian Time-Invariance E.E. on Θ:

Take the process Y have **state space** Θ, instead of $\mathcal{X}$
and be **stationary wrt** $\pi(\theta|x)$, instead of $f(x|\theta)$

TYPE I EE:
Let Y be **i.i.d. sampling** from the posterior
Take:

$$S(\theta) = g(\theta) \text{ the function we want to estimate}$$

Then the time-invariance E.E. gives Bayesian estimator for $g(\theta)$

TYPE II EE:
Let Y be an **independence MH** with proposal $=$ prior
Take:

$$T(\theta, \theta') = \frac{g(\theta')\pi(\theta|x)}{\nu(\theta)\alpha(\theta', \theta)} - \frac{g(\theta)\pi(\theta'|x)}{\nu(\theta')\alpha(\theta, \theta')}$$

Then the time-invariance E.E. gives Bayesian estimator for $g(\theta)$

General Time-Invariance Estimating Equations:

Suppose X and θ are both random:

$$\begin{array}{ccc} \theta & \sim & \nu(\theta) \\ X \mid \theta & \sim & f(x \mid \theta) \\ \theta \mid x & \sim & \pi(\theta \mid x) \end{array}$$

Construct a process on $(\mathcal{X}, \Theta)$ stationary wrt $\nu(\theta) f(x \mid \theta)$

For example consider the Gibbs sampler:

$$(x, \theta) \xrightarrow{\pi(\theta' \mid x)} (x, \theta') \xrightarrow{f(x' \mid \theta')} (x', \theta')$$

Then **Type I EE**

$$(\mathcal{A}S)(x, \theta) = \int_\Theta \int_\mathcal{X} S(x', \theta')\, f(x' \mid \theta')\, \pi(\theta' \mid x)\, \mathrm{d}x'\, \mathrm{d}\theta' - S(x, \theta)$$

which is an unbiased EE wrt the joint of $(X, \Theta) : \nu(\theta) f(x \mid \theta)$

Unifying framework

Y process on the $\mathcal{X}$ space	$\longrightarrow$	TI frequentist EE
Y process on the Θ space	$\longrightarrow$	TI Bayesian EE
Y process on the $(\mathcal{X}, \Theta)$ space	$\longrightarrow$	TI general EE

Y could be used to study properties of the resulting estimators

Eaton's Θ-chain + Hobert & Robert's $\mathcal{X}$-chain: studied to determine whether the formal Bayes estimate is ν-admissible
If either of these chains is null recurrent the Bayes estimate is ν-admissible
The $\mathcal{X}$, the Θ and the $(\mathcal{X}, \Theta)$-chains are always either all (positive) recurrent or all transient
$m(x)$ is proper $\Longleftrightarrow$ prior is proper

ADMISSIBILITY $\iff$ RECURRENCE

Brown (1971) showed that the (least squares) best invariant estimator of the mean of a multivariate normal distribution is admissible if and only if an associated diffusion is recurrent

Johnstone (1984, 1986) related the admissibility of estimators of Poisson means to the recurrence of birth-death Markov chains

Table: Updating regimes on Θ and resulting estimators

Stat. dist.	Markov process	Function	Type	Estimator
$\pi(\theta\|x)$	i.i.d. from $\pi(\theta\|x)$	$S(\theta) = g(\theta)$	I	**BE**
$\pi(\theta\|x)$	M-H with prop. $\nu(\theta)$	$T(\theta, \theta') = (2)$	II	**BE**
$\nu(\theta)$ Eaton's chain	$f(x'\|\theta)\pi(\theta'\|x')$	$T(\theta, x) =$ $h(x) - g(\theta)$	I	**MM**

$$T(\theta, \theta') = \frac{g(\theta')\pi(\theta|x)}{\nu(\theta)\alpha(\theta', \theta)} - \frac{g(\theta)\pi(\theta'|x)}{\nu(\theta')\alpha(\theta, \theta')} \qquad (2)$$

These EE are unbiased wrt the marginal of θ

Table: Updating regimes on $\mathcal{X}$ and resulting estimators

Stat. dist.	Markov process	Function	Type	Estimator
$f(x\|\theta)$	i.i.d. from $f(x\|\theta)$	$S(x) = h(x)$	I	**MM**
$f(x\|\theta)$	M-H with prop. $m(x)$	$T(x, x') = (3)$	II	**MM**
$m(x)$ Hobert's chain	$\pi(\theta'\|x)f(x'\|\theta')$	$T(\theta, x) = g(\theta) - h(x)$	I	**BE**

$$T(x, x') = \frac{h(x')f(x|\theta)}{m(x)\alpha(x', x)} - \frac{h(x)f(x'|\theta)}{m(x')\alpha(x, x')} \qquad (3)$$

These EE are unbiased wrt the marginal of X

Table: Updating regimes on $(\mathcal{X}, \Theta)$ and resulting estimators

Stat. dist.	Markov process	Function	Estimator
$f(x\|\theta)\nu(\theta)$	$\pi(\theta'\|x)f(x'\|\theta')$	$S(x,\theta) = g(\theta)$	**BE**
$\pi(\theta\|x)m(x)$	$f(x'\|\theta)\pi(\theta'\|x')$	$S(x,\theta) = h(x)$	**MM**

These EE are unbiased wrt the joint (X, θ)

Table: Updating regimes and resulting estimators on $(\mathcal{X}, \Theta)$

Stat. dist.	Markov process	Function	Estimate
$\pi(\underline{\theta}\|\underline{x})m(\underline{x})$	$x_i' \sim f(x_i'\|x_{-i}, \underline{\theta})$ $\underline{\theta}' \sim \pi(\underline{\theta}'\|\underline{x}')$	$S(\underline{x}, \underline{\theta}) = h(\underline{x})$	Pseudo-LHD
$f(\underline{x}\|\underline{\theta})\nu(\underline{\theta})$	$\theta_j' \sim \pi(\theta_j'\|\theta_{-j}, \underline{x})$ $\underline{x}' \sim f(\underline{x}'\|\underline{\theta}')$	$S(\underline{x}, \underline{\theta}) = g(\underline{\theta})$	Pseudo-POST
$\pi(\underline{\theta}\|\underline{x})m(\underline{x})$	$\underline{x}' \sim$ block Gibbs sampler $\underline{\theta}' \sim \pi(\underline{\theta}'\|\underline{x}')$	$S(\underline{x}, \underline{\theta}) = h(\underline{x})$	General Pseudo-LHD
$f(\underline{x}\|\underline{\theta})\nu(\underline{\theta})$	$\underline{\theta} \sim$block Gibbs sampler $\underline{x} \sim f(\underline{x}'\|\underline{\theta}')$	$S(\underline{x}, \underline{\theta}) = g(\underline{\theta})$	General Pseudo-POST

OPEN QUESTION

This approach yields many different estimators (old and new) depending on the choices of $\mathbf{Y}$ and S

WHICH CHOICE OF $\mathbf{Y}$ AND S IS "BEST"?

INTUITION to ANSWER

(frequentist)
Many of these estimators can be interpreted as maximising a variant of the likelihood

If the chain Y

has good performance as an MCMC sampler
delivers efficient MCMC estimators
explores the state space rapidly

then $(\mathcal{A}S)(x)$ should be close to efficient score
$U(x,\theta) = \mathbb{E}_\theta S(X) - S(x)$
so that the time-invariance estimator should be close to the MLE

Godambe-Heyde asymptotic variance optimality criterion

First order approximation to the asymptotic variance of the TI estimator for θ, obtained using $\mathcal{A}_\theta$ and the statistic V:

$$GH(\theta) = \frac{\mathrm{Var}_\theta\left[(\mathcal{A}_\theta V)(X)\right]}{\left(\mathbb{E}_\theta \frac{\partial}{\partial\theta}(\mathcal{A}_\theta V)(X)\right)^2} \quad \text{in general}$$

$$= \frac{\mathbb{E}_\theta\left[(\mathcal{A}_\theta V)(X)^2\right]}{\mathbb{E}_\theta\left[V(X)(\mathcal{A}_\theta V)(X)\right]^2} \quad \text{for exp. fam.}$$

GODAMBE-HEYDE ORDERING

Take $f(x|\theta) \in$ exponential family with parameter space Θ
Let $S = V =$ canonical sufficient statistics
Let Y and Z be two stochastic processes stationary wrt $f(x|\theta)$
We say that $\boxed{\widehat{\theta}_{\mathbf{Y}} \quad \succeq_{GH} \quad \widehat{\theta}_{\mathbf{Z}}}$ if

$$GH(\theta)_{\mathbf{Y}} \leq GH(\theta)_{\mathbf{Z}}, \qquad \forall\theta \in \Theta$$

By Cauchy-Schwarz:

$$GH(\theta) \quad \geq \quad \frac{1}{\mathrm{Var}_{\theta}[V(X)]}$$

MLE achieves the bound since, for iid sampling:

$$\mathbb{E}[(\mathcal{A}V)(X)^2] = \mathrm{Var}[V(X)] = \mathbb{E}\left[V(X)\,(\mathcal{A}V)(X)\right]$$

and hence

$$GH(\theta) = \frac{1}{\mathrm{Var}[V(X)]}$$

Thus **MLE = GH-OPTIMAL** amongst all time-invariance estimators for exponential families derived from $V =$ sufficient statistics

Bayesian estimators achieves the GH lower bound since can be obtained by taking Y to be iid sampling from the posterior

Thus **BAYES estimator = GH-OPTIMAL** amongst all time-invariance estimators for exponential families derived from $V =$ sufficient statistics

ORDERINGS for MCMC

EFFICIENCY ORDERING

$\mathbb{P}$ is **uniformly** more efficient than $\mathbb{Q}$ for MCMC purposes if

$$\boxed{\mathbb{P} \succeq_E \mathbb{Q}} \Leftrightarrow \boxed{\text{Var}(S,\mathbb{P}) \leq \text{Var}(S,\mathbb{Q}) \qquad \forall S \in L^2(\pi)}$$

where $\text{Var}(S,\mathbb{P})$ = asympt. variance of MCMC estimator for $E_\pi S$

COVARIANCE ORDERING Mira and Geyer (2000)

$\mathbb{P}$ better than $\mathbb{Q}$ **in covariance ordering** if:

$$\boxed{\mathbb{P} \succeq_C \mathbb{Q}} \Leftrightarrow \boxed{\text{Cov}_\pi[S(X_0), S(X_1)] \leq \text{Cov}_\pi[S(Y_0), S(Y_1)] \; \forall S \in L^2(\pi)}$$

THEOREM: $\boxed{\mathbb{P} \succeq_E \mathbb{Q}} \Leftrightarrow \boxed{\mathbb{P} \succeq_C \mathbb{Q}}$

$f(x|\theta)$ = exponential family with sufficient statistic V
Let $\mathbf{Y}_\theta$ and $\mathbf{Z}_\theta$ be MC reversible wrt $f(x|\theta)$
Let $\widehat{\theta}_\mathbf{Y}$ and $\widehat{\theta}_\mathbf{Z}$ be corresponding TI estimators for the statistic V
If

1. $\mathbf{Y}_\theta$ dominates $\mathbf{Z}_\theta$ in the covariance ordering, for each θ; **and**
2. $\mathrm{Var}_\theta[(\mathcal{A}_{\mathbf{Y}_\theta} V)(X)] \leq \mathrm{Var}_\theta[(\mathcal{A}_{\mathbf{Z}_\theta} V)(X)]$,

then

$$\boxed{\widehat{\theta}_\mathbf{Y} \quad \succeq_{GH} \quad \widehat{\theta}_\mathbf{Z}}$$

i.e. **Y** is better if it has both

stronger negative correlation $\mathbb{E}[V(X)(\mathcal{A}V)(X)]$
weaker positive correlation $\mathbb{E}[V(X)(\mathcal{A}^2 V)(X)]$

than the corresponding terms for **Z**

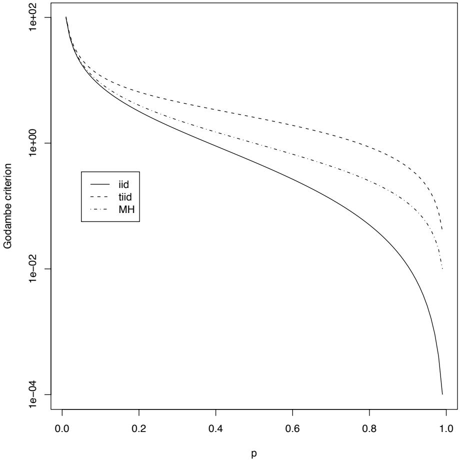
EXAMPLE: Geometric distribution

$\pi(x) = (1-p)p^{x-1}$ for $x = 1, 2, \cdots$ and $0 \leq p \leq 1$
Take $S(x) = V(x) = x$, the canonical sufficient statistic
Compare:

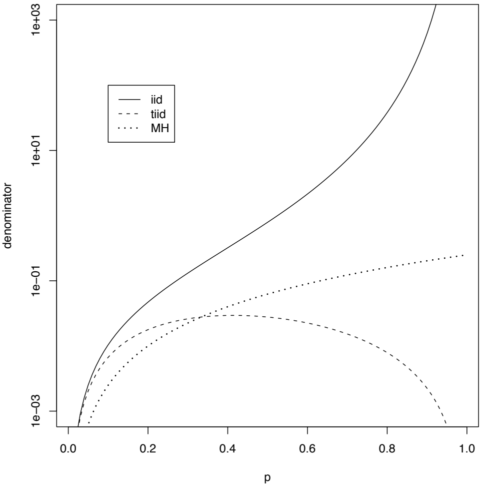
- **iid sampler**: $\hat{p}_{\text{iid}} = 1 - \frac{1}{x} = \text{MLE} = \text{MM}$
- **truncated iid sampler**: if $X_t = x$, a jump to y is proposed with probability $\pi(y)$; the proposal is accepted unless $|x - y| > c$
  For $c = 1$: $\widehat{p}_1 = 0$ if $x > 1$ and $\widehat{p} = 0, 1$ if $x = 1$
- **MH sampler** with reflecting symmetric RW proposal: $\widehat{p}_{MH} = \mathbf{1}\{x > 1\}$

$\text{iid} \succeq_P \text{tiid}_c$	$\forall c$
$\text{tiid}_c \succeq_P \text{tiid}_{c'}$	$\forall c' < c$
$\text{MH} \succeq_P \text{tiid}_1$	$p \geq 0.5$

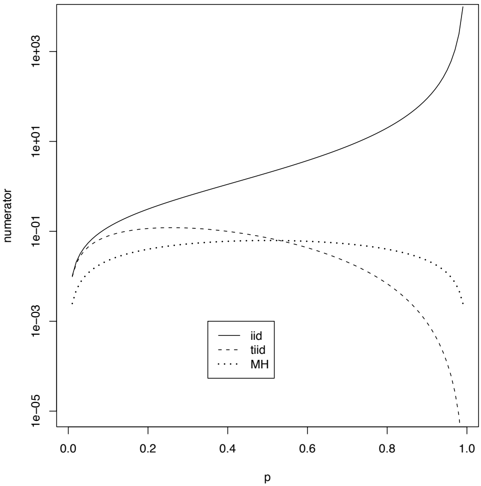
Values of Godambe-Heyde ratio plotted against p



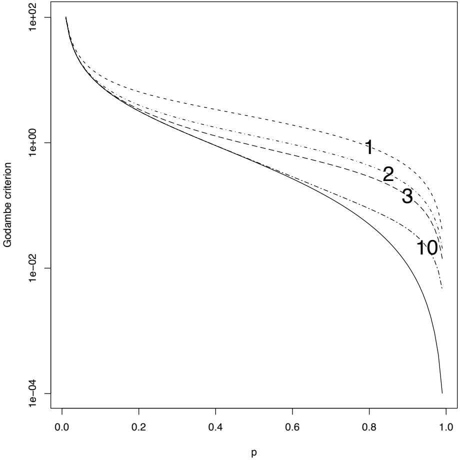
Denominators of the Godambe-Heyde ratio



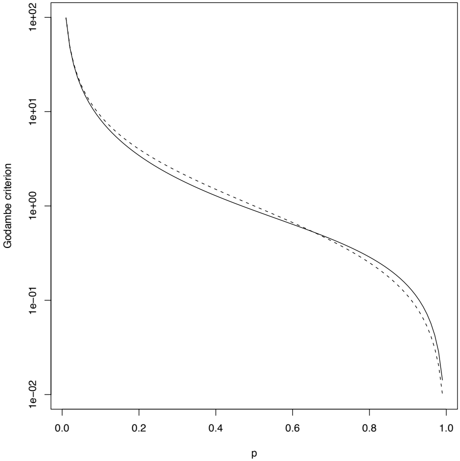
Numerators of the Godambe-Heyde criterion



Godambe-Heyde ratio for the tiid with $c = 1, 2, 3, 10, \infty$



GH ratio for tiid with $c = 3$ (solid) and the MH (dotted)



INDEPENDENCE-MH SAMPLER

closeness of the proposal $q(\cdot)$ to the target distribution
$\Downarrow$
closeness of the estimating fct $\mathcal{A}_\theta V(x)$ to the efficient score $U(x, \theta)$
$\Downarrow$
high GH-efficiency of the time-invariance estimator

MPLE is "linearisation" of MLE

A standard result:

$$\exp\{t\mathcal{A}\} = P_t$$

i.e.

$$(\exp\{t\mathcal{A}\}h)(x) = \mathbb{E}_\theta[h(Y_t) \mid Y_0 = x]$$

If $\mathbf{Y}$ converges in distribution to X then we have

$$\lim_{t\to\infty} \exp\{t\mathcal{A}\}V(x) = \mathbb{E}_\theta V(X)$$

thus the MLE is the solution of the normal equations

$$\lim_{t\to\infty} \exp\{t\mathcal{A}_\theta\}V(x) = V(x)$$

or

$$\lim_{t\to\infty} (\exp\{t\mathcal{A}_\theta\} - I)\, V(x) = 0$$

The pseudo-likelihood estimating equations are

$$\mathcal{A}_\theta V(x) = 0$$

and we can think of $\mathcal{A}_\theta$ as a first order approximation to the series $\exp\{t\mathcal{A}\} - I$

TI-EE are particularly relevant when:

- model of interest is highly structured
- likelihood or the posterior are not known analytically but can be expressed as equilibrium distributions of an associated Markov processes

This is the same setting when MCMC and MPLE are helpful

We can think of TI-EE as the "middle way"

Statistical performance: invariance, consistency, asymptotic normality and optimality discussed in Baddeley (2000)

Spatial processes: Kurtz and Li (2004) give

LLN + CLT for TI-Estimators

Conclusions

- TI-EE involve an arbitrary choice of Y and S
  - Y is a mathematical fiction
  - natural choices of S and Y often yield known estimators (BE, MLE, M-of-M, MPLE, Takacs-Fiksel,variational est...)
- TI-EE = common structure for different estimation techniques
- TI-EE = general recipe for deriving new estimators
- TI-EE = general framework for comparing estimators
- Computational complexity of $\hat{\theta}$ depends on Y transition structure
- Efficiency of time-invariance estimators is *related* to the quality of Y as an MCMC sampler

Open questions

- How can we incorporate the loss fct in Bayesian TI-EE?
- Which Y and S give TI-EE that result in Stein estimator?
- Derive Kaplan-Meier estimator as a time-invariance estimator
- How to study large sample properties of TI-EE?
- Conjecture that large sampler properties of time-invariance estimators mimic those of MLE, MPLE, BE
- How to study large-domain limit theory for spatial models?
- How to handle random effects?