The van Trees inequality

0.1 Introduction

Suppose $p(x, \theta)$ is a probability density indexed by a subset $\Theta$ of the real line. For a real-valued statistic $T$ with $E_\theta T(x) = \tau(\theta)$, the information inequality asserts that

$$\text{var}_\theta(T) \geq \frac{\tau(\theta)^2}{I_p(\theta)}$$

where $I_p(\theta) = \text{var}_\theta \left( \frac{\partial \log p(x, \theta)}{\partial \theta} \right) = \int \frac{p(x, \theta)^2}{p(x, \theta)} dx$.

The van Trees (VT) inequality, due to van Trees (1968, page 72), is a Bayesian analog of the information inequality. (Actually it is just the information inequality applied to a cunningly chosen joint density.) Gill and Levit (1995) have shown how the VT inequality can be applied to a variety of statistical problems.

For a suitably chosen (prior) density $q$ on $\Theta$ and any real valued function $\psi$ on $\Theta$, the one-dimension version of the VT inequality is

$$\int_\Theta E_\theta(T(x) - \psi(\theta))^2 q(\theta) d\theta \geq \left( \int \psi(\theta) q(\theta) d\theta \right)^2 I_q + \int I_p(\theta) q(\theta) d\theta$$

Here $I_p(\theta)$ denotes the Fisher information function and $I_q = \int \frac{\psi(\theta)^2}{q(\theta)} d\theta$.

For this note I consider only the case where $\psi(\theta) = \theta$, so that the numerator in $<1>$ becomes 1:

$$\int_\Theta E_\theta(T(x) - \theta)^2 q(\theta) d\theta \geq \frac{1}{I_q + \int I_p(\theta) q(\theta) d\theta}$$

Remark. In keeping with my convention of writing $t$ instead of $\theta$ when treating $\theta$ as a dummy variable, I could have written $<2>$ as an integral with respect to $t$, replacing every $\theta$ by a $t$.

0.2 Proof of the VT inequality

Create a new family of joint densities by treating $\theta$ itself as random,

$$\gamma_h(x, t) = q(t + h)p(x, t + h)$$

for $x \in X$ and $t \in \Theta$,

where $-\delta < h < \delta$ for some small $\delta$. 
Remark. The definition makes sense only when \( q(t+h) \) is well defined. Typically \( q \) is chosen to be a smooth function with compact support: it is assumed to be as differentiable as we need and it is \( > 0 \) only in some small neighborhood of a particular \( \theta \). Take \( q(t) = 0 \) outside the neighborhood, so that \( q \) is well defined and differentiable on the whole real line, not just on \( \Theta \).

Notice that \( \gamma_h \) is nonnegative and \( \iint \gamma_h(x, t) \, dx \, dt = 1 \); it is a probability density. To avoid confusion with \( E_\theta \) as an integral over just the \( x \), write \( E_h \) for integrals with respect to both variables:

\[
E_h F(x, t) = \iint F(x, t) \gamma_h(x, t) \, dx \, dt.
\]

For example,

\[
G(h) := E_h (T(x) - t))
= \iint (T(x) - t)) q(t+h)p(x, t+h) \, dx \, dt
= \iint (T(x) - s + h)) q(s)p(x, s) \, dx \, ds \quad \text{change of variable}
= E_0 (T(x) - t)) + h.
\]

That is, \( G(h) - G(0) = h \).

For the information inequality we needed to show that the function \( \Delta_h(x, t) := \frac{p(x, t+h) - p(x, t)}{p(x, t)} \)
satisfied \( E_t \Delta_h(x, t) = 0 \) and

\[
E_t \Delta_h(x, t) (T(x) - \tau(t)) = \tau(t+h) - \tau(t) \quad \text{for each fixed } t.
\]

For the joint densities \( \gamma_h \) a similar role is played by

\[
D_h(x, t) := \frac{\gamma_h(x, t) - \gamma_0(x, t)}{\gamma_0(x, t)}.
\]

As before,

\[
E_0 D_h(x, t) = \iint (\gamma_h(x, t) - \gamma_0(x, t)) \, dx \, dt = 1 - 1 = 0.
\]
but now

\[ E_0 D_h(x, t)(T(x) - t) = E_h(T(x) - t) - E_0(T(x) - t) \]
\[ = G(h) - G(0) = h. \]

Once again Cauchy-Schwarz gives

\[ h^2 = |E_0 D_h(x, t)(T(x) - t)|^2 \leq (E_0 D_h^2(x, t)) (E_0(T(x) - t)^2), \]

The second term on the right-hand side equals

\[ \int q(t) E_t (T(x) - t)^2 dt, \]

which is the expression on the left-hand side of <2>. Expansion of the quadratic \( D_h^2(x, t) \) gives

\[ E_0 D_h(x, t)^2 = \iint \frac{\gamma_h(x, t)}{\gamma_0(x, t)} \left( -2\gamma_h(x, t) + \gamma_0(x, t) \right) dx \, dt \]

so that

\[ 1 + E_0 D_h(x, t)^2 = \iint \frac{\gamma_h(x, t)^2}{\gamma_0(x, t)} \, dx \, dt. \]

With a similar expansion followed by Taylor for small \(|h|\) we have

\[ \int \frac{p(x, t + h)^2}{p(x, t)} \, dx = 1 + \iint \frac{(p(x, t + h) - p(x, t))^2}{p(x, t)} \, dx \approx 1 + h^2 \Pi_p(t) \]
\[ \int \frac{q(t + h)^2}{q(t)} \, dt = 1 + \iint \frac{(q(t + h) - q(t))^2}{q(t)} \, dt \approx 1 + h^2 \Pi_q. \]

Combine the last three equalities to deduce, for small \(|h|\), that

\[ 1 + E_0 D_h(x, t)^2 = \iint \frac{q(t + h)^2 p(x, t + h)^2}{q(t)p(x, t)} \, dx \, dt \]
\[ = \int \frac{q(t + h)^2}{q(t)} \left( \int \frac{p(x, t + h)^2}{p(x, t)} \, dx \right) \, dt \]
\[ \approx \int \frac{q(t + h)^2}{q(t)} \left( 1 + h^2 \Pi_p(t) \right) \, dt \]
\[ \approx \int \frac{q(t + h)^2}{q(t)} \, dt + h^2 \int \frac{q(t + h)^2}{q(t)} \Pi_p(t) \, dt. \]
That is,
\[
\frac{\varepsilon_0 D_h(x, t)^2}{h^2} \approx I_q + \int \frac{q(t + h)^2}{q(t)} I_p(t) \, dt.
\]

Finally, note that the last term is changed by an order $|h|$ quantity if we reduce $q(t + h)^2/q(t)$ to $q(t)$. In the limit as $h$ tends to zero we get the expression in the denominator of the right-hand side of $<2>$.

References
