The van Trees inequality

0.1 Introduction

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

$$\operatorname{var}_\theta(T) \ge \frac{\dot{\tau}(\theta)^2}{\mathbb{I}_p(\theta)} \qquad \text{where } \mathbb{I}_p(\theta) = \operatorname{var}_\theta\left(\frac{\partial \log p(x,\theta)}{\partial\theta}\right) = \int \frac{\dot{p}(x,\theta)^2}{p(x,\theta)}dx.$$

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 Θ and any real valued function ψ on Θ, the one-dimension version of the VT inequality is

<1> $$\int_\Theta \mathbb{E}_\theta(T(x)-\psi(\theta))^2 q(\theta)\,d\theta \ge \frac{\left(\int \dot{\psi}(\theta)q(\theta)\,d\theta\right)^2}{\mathbb{I}_q + \int \mathbb{I}_p(\theta)q(\theta)\,d\theta}$$

Here $\mathbb{I}_p(\theta)$ denotes the Fisher information function and $\mathbb{I}_q = \int \dot{q}(\theta)^2/q(\theta)\,d\theta$. For this note I consider only the case where $\psi(\theta)=\theta$, so that the numerator in <1> becomes 1:

<2> $$\int_\Theta \mathbb{E}_\theta(T(x)-\theta)^2 q(\theta)\,d\theta \ge \frac{1}{\mathbb{I}_q + \int \mathbb{I}_p(\theta)q(\theta)\,d\theta}$$

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

0.2 Proof of the VT inequality

Create a new family of joint densities by treating θ itself as random,

<3> $$\gamma_h(x,t) = q(t+h)p(x,t+h) \qquad \text{for } x\in\mathcal{X} \text{ and } t\in\Theta,$$

where $-\delta < h < \delta$ for some small δ.

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 θ. Take $q(t)=0$ outside the neighborhood, so that q is well defined and differentiable on the whole real line, not just on Θ.

Notice that γ_h is nonnegative and $\iint \gamma_h(x,t)\,dx\,dt = 1$; it is a probability density. To avoid confusion with $\mathbb{E}_\theta$ as an integral over just the x, write $\mathcal{E}_h$ for integrals with respect to both variables:

$$\mathcal{E}_h F(x,t) = \iint F(x,t)\gamma_h(x,t)\,dx\,dt.$$

For example,

$$\begin{aligned} G(h) &:= \mathcal{E}_h\left(T(x)-t\right) \\ &= \iint \left(T(x)-t\right) q(t+h)p(x,t+h)\,dx\,dt \\ &= \iint \left(T(x)-s+h\right) q(s)p(x,s)\,dx\,ds \qquad \text{change of variable} \\ &= \mathcal{E}_0\left(T(x)-t\right)+h. \end{aligned}$$

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 $\mathbb{E}_t\Delta_h(x,t)=0$ and

$$\mathbb{E}_t\Delta_h(x,t)\left(T(x)-\tau(t)\right) = \tau(t+h)-\tau(t) \qquad \text{for each fixed } t.$$

For the joint densities γ_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,

$$\mathcal{E}_0 D_h(x,t) = \iint \left(\gamma_h(x,t)-\gamma_0(x,t)\right)\,dx\,dt = 1-1=0,$$

but now

$$\begin{aligned}\mathcal{E}_0 D_h(x,t)(T(x)-t) &= \mathcal{E}_h(T(x)-t) - \mathcal{E}_0(T(x)-t)\\ &= G(h) - G(0) = h.\end{aligned}$$

Once again Cauchy-Schwarz gives

<4> $$h^2 = |\mathcal{E}_0 D_h(x,t)(T(x)-t)|^2 \le \left(\mathcal{E}_0 D_h^2(x,t)\right)\left(\mathcal{E}_0(T(x)-t)^2\right).$$

The second term on the right-hand side equals

$$\int q(t)\mathbb{E}_t\left(T(x)-t\right)^2 dt,$$

which is the expression on the left-hand side of <2>.

Expansion of the quadratic $D_h^2(x,t)$ gives

$$\mathcal{E}_0 D_h(x,t)^2 = \iint \frac{\gamma_h(x,t)^2}{\gamma_0(x,t)} - 2\gamma_h(x,t) + \gamma_0(x,t)\,dx\,dt$$

so that

$$1 + \mathcal{E}_0 D_h(x,t)^2 = \iint \frac{\gamma_h(x,t)^2}{\gamma_0(x,t)}\,dx\,dt.$$

With a similar expansion followed by Taylor for small $|h|$ we have

$$\begin{aligned}\int \frac{p(x,t+h)^2}{p(x,t)}\,dx &= 1 + \iint \frac{\left(p(x,t+h)-p(x,t)\right)^2}{p(x,t)}\,dx \approx 1 + h^2\mathbb{I}_p(t)\\ \int \frac{q(t+h)^2}{q(t)}\,dt &= 1 + \int \frac{\left(q(t+h)-q(t)\right)^2}{q(t)}\,dt \approx 1 + h^2\mathbb{I}_q.\end{aligned}$$

Combine the last three equalities to deduce, for small $|h|$, that

$$\begin{aligned}1 + \mathcal{E}_0 D_h(x,t)^2 &= \iint \frac{q(t+h)^2 p(x,t+h)^2}{q(t)p(x,t)}\,dx\,dt\\ &= \int \frac{q(t+h)^2}{q(t)}\left(\int \frac{p(x,t+h)^2}{p(x,t)}\,dx\right)dt\\ &\approx \int \frac{q(t+h)^2}{q(t)}\left(1 + h^2\mathbb{I}_p(t)\right)\,dt\\ &\approx \int \frac{q(t+h)^2}{q(t)}\,dt + h^2\int \frac{q(t+h)^2}{q(t)}\mathbb{I}_p(t)\,dt.\end{aligned}$$

That is,

$$\frac{\mathcal{E}_0 D_h(x,t)^2}{h^2} \approx \mathbb{I}_q + \int \frac{q(t+h)^2}{q(t)} \mathbb{I}_p(t)\, dt.$$

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

Gill, R. and B. Levit (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli 1*, 59–79.

van Trees, H. L. (1968). *Detection, Estimation and Modulation Theory, Part 1.* Wiley & Sons.