Representation formulae for score functions

Ivan Nourdin, Giovanni Peccati and Yvik Swan*

Département de Mathématique, Université de Liège

July 2, 2014

Scoooores

Let X be a centered d-random vector with covariance $B > 0$.

Definition

The Stein kernel of X is a $d \times d$ matrix $\tau_X(X)$ such that

$$E\left[\tau_X(X)\nabla\varphi(X)\right] = E\left[X\varphi(X)\right]$$

for all $\varphi \in C_c^\infty(\mathbb{R}^d)$.

Definition

The score of X is the $d \times 1$ vector $\rho_X(X)$ such that

$$E\left[\rho_X(X)\varphi(X)\right] = -E\left[\nabla\varphi(X)\right]$$

for all $\varphi \in C_c^\infty(\mathbb{R}^d)$.

In the Gaussian case $Z \sim \mathcal{N}_d(0, C)$ the Stein identity

$$E\left[Z\varphi(Z)\right] = E\left[C\nabla\varphi(Z)\right]$$

gives

$$\rho_Z(Z) = -C^{-1}Z \text{ and } \tau_Z(Z) = C.$$

Intuitively, a measure of proximity

$$\rho_X(X) \approx -B^{-1}X$$

and

$$\tau_X(X) \approx B$$

should provide an assessment of "Gaussianity".

Definition

The standardised Fisher information of X is

$$J_{st}(X) = BE\left[\left(\rho_X(X) + B^{-1}X\right)\left(\rho_X(X) + B^{-1}X\right)^T\right].$$

A simple computation gives

$$J_{st}(X) = BJ(X) - Id$$

with $J(X) = E\left[\rho_X(X)\rho_X(X)^T\right]$ the Fisher information matrix.

Definition

The Stein discrepancy is

$$S(X) = E\left[\|\tau_X(X) - B\|^2_{H.S.}\right].$$

Control on $J_{st}(X)$ and $S(X)$ provides control on several distances (Kullback-Leibler, Kolmogorov, Wasserstein, Hellinger, Total Variation, ...) between the law of X and the Gaussian.

Controlling $J_{st}(X)$:

- Johnson and Barron through careful analysis of the score function (PTRF, 2004)
- Artstein, Ball, Barthe, Naor through "variational tour de force" (PTRF, 2004)

Controlling $S(X)$:

- Cacoullos Papathanassiou and Utev (AoP 1994) in a number of settings
- Nourdin and Peccati through their infamous Malliavin/Stein fourth moment theorem (PTRF, 2009)
- Extension to abstract settings (Ledoux, AoP 2012)

Let Z be centered Gaussian with density $\phi = \phi_d(\cdot; C)$.

Definition

The relative entropy between X and Z is

$$D(F \,||\, Z) = E\left[\log(f(X)/\phi(X))\right] = \int_{\mathbb{R}^d} f(x) \log\left(\frac{f(x)}{\phi(x)}\right) dx.$$

The Pinsker-Csiszar-Kullback inequality yields

$$2TV(X, Z) \leq \sqrt{2D(X \,||\, Z)}.$$

In other words

$$D(X \,||\, Z) \Rightarrow TV(X, Z)^2.$$

Usefulness of $J_{st}(X)$ can be seen via the de Bruijn identity.

Let $X_t = \sqrt{t}X + \sqrt{1-t}Z$ and $\Gamma_t = tB + (1-t)C$. Then

$$D(X \,||\, Z) = \int_0^1 \frac{1}{2t} \mathrm{tr}\left(C\Gamma_t^{-1} J_{st}(X_t)\right) dt$$

$$+ \frac{1}{2}\left(\mathrm{tr}\left(C^{-1}B\right) - d\right) + \int_0^1 \frac{1}{2t} \mathrm{tr}\left(C\Gamma_t^{-1} - I_d\right) dt$$

In other words

$$J_{st}(X_t) \Rightarrow D(X \,||\, Z) \Rightarrow (TV(X,Z))^2.$$

Usefulness of $S(X)$ can be seen via Stein's method.

Fix $d = 1$. Then, given $h : \mathbb{R} \to \mathbb{R}$ such that $\|h\|_\infty \leq 1$ seek g_h solution of the Stein equation to get

$$\begin{aligned} E\left[h(X)\right] - E\left[h(Z)\right] &= E\left[g_h'(X) - Xg_h(X)\right] \\ &= E\left[(1 - \tau_X(X))g_h'(X)\right] \end{aligned}$$

so that

$$\begin{aligned} TV(X, Z) &= \frac{1}{2} \sup_{\|h\|_\infty \leq 1} \left|E\left[h(X)\right] - E\left[h(Z)\right]\right| \\ &\leq \left(\frac{1}{2} \sup_{\|h\|_\infty \leq 1} \|g_h'\|\right) \sqrt{S(X)}. \end{aligned}$$

In other words

$$S(X) \Rightarrow TV(X, Z)^2.$$

If h is not smooth there is no way of obtaining sufficiently precise estimates on the quantity "∇g_h" in dimension greater than 1.

For the moment Stein's method only works in dimension 1 for total variation distance.

The IT approach via de Bruijn's identity does not suffer from this "dimensionality issue".

We aim to mix the Stein method approach and the IT approach.

To this end we need one final ingredient : a representation formulae for the score in terms of the Stein kernel.

Theorem

Let $X_t = \sqrt{t}X + \sqrt{1-t}Z$ *with* X *and* Z *independent. Then*

$$\rho_t(X_t) + C^{-1}X_t = -\frac{t}{\sqrt{1-t}}E\left[(I_d - C^{-1}\tau_X(X))Z \,|\, X_t\right] \qquad (1)$$

for all $0 < t < 1$.

Proof when $d = 1$ and $C = 1$.

$$E\left[E\left[(1-\tau_X(X))Z \,|\, X_t\right]\phi(X_t)\right] = E\left[(1-\tau_X(X))Z\phi(X_t)\right]$$
$$= \sqrt{1-t}E\left[\phi'(X_t)\right] - \sqrt{1-t}E\left[\tau_X(X)\phi'(X_t)\right]$$
$$= \sqrt{1-t}E\left[\phi'(X_t)\right] - \sqrt{\frac{1-t}{t}}E\left[X\phi(X_t)\right]$$
$$= \sqrt{1-t}E\left[\phi'(X_t)\right] - \frac{\sqrt{1-t}}{t}E\left[X_t\phi(X_t)\right] + \frac{1-t}{t}E\left[Z\phi(X_t)\right]$$
$$= \sqrt{1-t}E\left[\phi'(X_t)\right] - \frac{\sqrt{1-t}}{t}E\left[X_t\phi(X_t)\right] + \frac{1-t}{t}\sqrt{1-t}E\left[\phi'(X_t)\right]$$
$$= -\frac{\sqrt{1-t}}{t}\left(E\left[\phi'(X_t)\right] - E\left[X_t\phi(X_t)\right]\right)$$

This formula provides a nearly one-line argument.

Define

$$\Delta(X,t) = E\left[(I_d - C^{-1}\tau_X(X))Z \,|\, X_t\right].$$

Take $d = 1$ and all variances set to 1. Then

$$J_{st}(X_t) = E\left[(\rho_t(X_t) + X_t)^2\right] = \frac{t^2}{1-t} E\left[\Delta(X,t)^2\right]$$

so that

$$D(X \,||\, Z) = \frac{1}{2}\int_0^1 \frac{t}{1-t} E\left[\Delta(X,t)^2\right] dt.$$

Also,

$$E\left[\Delta(X,t)^2\right] \leq E\left[(1-\tau_X(X))^2\right] = S(X).$$

This yields

$$D(X||Z) \leq \frac{1}{2} S(X) \int_0^1 \frac{t}{1-t} dt$$

which is useless.

There is hope, nevertheless :

$$\int_0^1 \frac{t}{1-t} dt$$

is barely infinity.

Recall $X_t = \sqrt{t}X + \sqrt{1-t}Z$. Then

$$\Delta(X,t) = E\left[(1-\tau_X(X))Z \,|\, X_t\right]$$

is such that

$$\Delta(X,0) = \Delta(X,1) = 0 \text{ a.s.}$$

Hence we need to identify conditions under which

$$\frac{t}{1-t} E\left[\Delta(X,t)^2\right]$$

is integrable at $t = 1$.

The behaviour of $\Delta(X, t)$ around $t \approx 1$ is central to the understanding of the law of X.

The behaviour of

$$E\left[\Delta(X,t)^2\right] \text{ at } t \approx 1$$

is closely connected to the so-called MMSE dimension studied by the IT community.

This quantity revolves around the remarkable "MMSE formula"

$$\frac{d}{dr} I(X; \sqrt{r}X + Z) = E\left[(X - E[X \mid \sqrt{r}X + Z])^2\right]$$

due to Guo, Shamai and Verdu (IEEE, 2005)

The connexion is explicitly stated in NPSb (IEEE, 2014).

In NPSa (JFA, 2014) we suggest the following IT alternative to Stein's method.

First cut the integral :

$$
\begin{aligned}
& 2D(X||Z) \\
& \leq E\left[(1-\tau_X(X))^2\right] \int_0^{1-\epsilon} \frac{t}{1-t} dt + \int_{1-\epsilon}^1 \frac{t}{1-t} E\left[\Delta(X,t)^2\right] dt \\
& \leq E\left[(1-\tau_X(X))^2\right] |\log \epsilon| + \int_{1-\epsilon}^1 \frac{t}{1-t} E\left[\Delta(X,t)^2\right] dt.
\end{aligned}
$$

Next suppose that when t is close to 1 we have

$$
E\left[\Delta(X,t)^2\right] \leq C_\kappa t^{-1}(1-t)^\kappa \tag{2}
$$

for some $\kappa > 0$.

We deduce

$$2D(X \,||\, Z) \leq S(X)|\log \epsilon| + C_\eta \int_{1-\epsilon}^{1} (1-t)^{-1+\kappa} dt$$
$$= S(X)|\log \epsilon| + \frac{C_\kappa}{\kappa} \epsilon^\kappa.$$

The optimal choice is $\epsilon = E\left[(1-\tau_X(X))^2\right]^{1/\kappa}$ which leads to

$$D(X \,||\, Z) \leq \frac{1}{2\kappa} S(X) \log S(X) + \frac{C_\kappa}{2\kappa} S(X)$$

which provides a bound on the total variation distance in terms of $S(X)$ which is of the correct order up to a logarithmic factor.

Under what conditions do we have (2)?

It is relatively easy to show (via Hölder's inequality) that

$$E\left[|\tau_X(X)|^{2+\eta}\right] < \infty \text{ and } E\left[|\Delta(X,t)|\right] \leq ct^{-1}(1-t)^{\delta} \quad (3)$$

implies (2).

It now remains to identify under which conditions we have (3).

Lemma (Poly's first lemma)

Let X be an integrable random variable and let Y be a $\mathbb{R}^d$-valued random vector having an absolutely continuous distribution. Then

$$E\left|E\left[X \mid Y\right]\right| = \sup E\left[Xg(Y)\right],$$

where the supremum is taken over all $g \in C_c^1$ such that $\|g\|_\infty \leq 1$

Thus

$$E\left|E\left[Z(1-\tau_X(X))\,|\,X_t\right]\right| = \sup E\left[Z(1-\tau_X(X))g(X_t)\right].$$

Now choose $g \in C_c^1$ such that $\|g\|_\infty \leq 1$. Then

$$\begin{aligned}
&E\left[Z(1-\tau_X(X))g(X_t)\right] \\
&= E\left[Zg(X_t)\right] - E\left[Zg(X_t)\tau_X(X)\right] \\
&= E\left[Zg(X_t)\right] - \sqrt{1-t}E\left[\tau_X(X)g'(X_t)\right] \\
&= E\left[Z\left(g(X_t)-g(X)\right)\right] - \sqrt{\frac{1-t}{t}}E\left[g(X_t)X\right]
\end{aligned}$$

and thus

$$\left|E\left[Z(1-\tau_X(X))g(X_t)\right]\right| \leq \left|E\left[Z\left(g(X_t)-g(X)\right)\right]\right| + t^{-1}\sqrt{1-t}.$$

Also

$$\begin{aligned}
&\sup \left| E\left[Z\left(g(X_t) - g(X) \right) \right] \right| \\
&= \sup \left| \int_{\mathbb{R}} x E\left[g(\sqrt{t}X + \sqrt{1-t}x) - g(X) \right] \phi_1(x) dx \right| \\
&\leq 2 \int_{\mathbb{R}} |x| \, TV(\sqrt{t}X + \sqrt{1-t}x, X) \phi_1(x) dx.
\end{aligned}$$

Wrapping up we get

$$\begin{aligned}
&E\left| E\left[Z(1 - \tau_X(X)) \,|\, X_t \right] \right| \\
&\leq 2E\left[|Z| \, TV(\sqrt{t}X + \sqrt{1-t}Z, X) \right] + t^{-1}\sqrt{1-t}.
\end{aligned}$$

It therefore all boils down to a condition on

$$TV(\sqrt{t}X + \sqrt{1-t}x, X).$$

Recall that we want

$$E\left|E\left[Z(1-\tau_X(X))\,|\,X_t\right]\right| \leq ct^{-1}(1-t)^{\delta}. \tag{3}$$

As it turns out, in view of previous results, a sufficient condition for (3) is

$$TV(\sqrt{t}X + \sqrt{1-t}x, X) \leq \kappa(1+|x|)t^{-1}(1-t)^{\alpha}.$$

This condition – and its multivariate extension – is satisfied by a wide family of random vectors including those for which they can apply their fourth moment bound

$$S(X) \leq c(E\left[X^4\right] - 3).$$

Theorem (Entropic CLTs on Wiener chaos)

Let $d \geq 1$ and $q_1, \ldots, q_d \geq 1$ be fixed integers. Consider vectors

$$F_n = (F_{1,n}, \ldots, F_{d,n}) = (I_{q_1}(h_{1,n}), \ldots, I_{q_d}(h_{d,n})), \quad n \geq 1,$$

with $h_{i,n} \in \mathfrak{H}^{\odot q_i}$. Let C_n denote the covariance matrix of F_n and let $Z_n \sim \mathcal{N}_d(0, C_n)$ be a centered Gaussian random vector in $\mathbb{R}^d$ with the same covariance matrix as F_n.

Let

$$\Delta_n := E[\|F_n\|^4] - E[\|Z_n\|^4],$$

Assume that $C_n \to C > 0$ and $\Delta_n \to 0$, as $n \to \infty$.

Then, the random vector F_n admits a density for n large enough, and

$$D(F_n \| Z_n) = O(1)\, \Delta_n |\log \Delta_n| \quad \text{as } n \to \infty, \tag{4}$$

where $O(1)$ indicates a bounded numerical sequence depending on $d, q_1, ..., q_d$, as well as on the sequence $\{F_n\}$.

Let X_i, $i = 1, \dots, n$ be independent random vectors with Stein kernels $\tau_i(X_i)$ and score functions $\rho_i(X_i)$, $i = 1, \dots, n$.

For all $t = (t_1, \dots, t_n) \in [0,1]^d$ such that $\sum_{i=1}^n t_i = 1$ we define

$$W_t = \sum_{i=1}^n \sqrt{t_i} X_i$$

and denote Γ_t the corresponding covariance matrix. Then

$$\rho_t(W_t) + \Gamma_t^{-1} W_t = \sum_{i=1}^n \frac{t_i}{\sqrt{t_{i+1}}} E\left[\left(Id - \Gamma_t^{-1}\tau_i(X_i)\right)\rho_{i+1}(X_{i+1})|W_t\right]$$

where we identify $X_{n+1} = X_1$ and $t_{n+1} = t_1$.

Lemma (Poly's second lemma)

Let X and Y be square-integrable random variables with mean $E[X] = 0$. Then

$$E\left[(E\left[X \mid Y\right])^2\right] = \sup_{\varphi \in \mathcal{H}(Y)} \left(E\left[X\varphi(Y)\right]\right)^2,$$

where the supremum is taken over the collection $\mathcal{H}(Y)$ of functions φ such that $E[\varphi(Y)] = 0$ and $E\left[\varphi(Y)^2\right] \leq 1$.

Theorem

Let $W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ where the X_i are independent random variables with Stein factor $\tau_i(X_i)$ and score function $\rho_i(X_i)$. Then

$$J_{st}(W_n) = \sup_{\varphi \in \mathcal{H}(W_n)} \left(E\left[\varphi'(W_n) - W_n\varphi(W_n)\right]\right)^2.$$

There seem to be many applications of the last formula.

For instance the difference

$$J_{st}(W_{n+1}) - J_{st}(W_n)$$

can be studied in quite some detail.

We had hoped to obtain the "entropy jump inequality" as well as the "increasingness of entropy".

There is, however, some work left before we can hooray.

Just a final word to say

thank you

to Janna, Jay and Larry for the great conference.

The key is a generalisation Carbery and Wright inequality : there is a universal constant $c > 0$ such that, for any polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most d and any $\alpha > 0$ we have

$$E[Q(X_1, \ldots, X_n)^2]^{\frac{1}{2d}} \; P(|Q(X_1, \ldots, X_n)| \leq \alpha) \leq cd\alpha^{\frac{1}{d}},$$

where $X_1, \ldots, X_n$ are independent random variables with common distribution $\mathcal{N}(0, 1)$.

Explicit conditions : fix $d, q_1, \ldots, q_d \geq 1$,

1. let $F = (F_1, \ldots, F_d)$ be a random vector such that $F_i = I_{q_i}(h_i)$ with $h_i \in \mathfrak{H}^{\odot q_i}$
2. set $N = 2d(q-1)$ with $q = \max_{1 \leq i \leq d} q_i$
3. Let C be the covariance matrix of F

Let $\Gamma = \Gamma(F)$ denote the Malliavin matrix of F, and assume that $E[\det \Gamma] > 0$ (which is equivalent to assuming that F has a density).

There exists a constant $c_{q,d,\|C\|_{H.S.}} > 0$ (depending only on q, d and $\|C\|_{H.S.}$ — with a continuous dependence in the last parameter) such that, for any $\mathbf{x} \in \mathbb{R}^d$ and $t \in [\frac{1}{2}, 1]$,

$$\begin{aligned} &\mathbf{TV}(\sqrt{t}F + \sqrt{1-t}\mathbf{x}, F) \\ &\leq c_{q,d,\|C\|_{H.S.}} \left(\beta^{-\frac{1}{N+1}} \wedge 1\right) (1 + \|\mathbf{x}\|_1) (1-t)^{\frac{1}{2(2N+4)(d+1)+2}}. \end{aligned}$$

Theorem (Entropic fourth moment bound)

Let $F_n = (F_{1,n}, ..., F_{d,n})$ be a sequence of d-dimensional random vectors such that: (i) *$F_{i,n}$ belongs to the q_ith Wiener chaos of $\mathbf{G}$, with $1 \leq q_1 \leq q_2 \leq \cdots \leq q_d$;* (ii) *each $F_{i,n}$ has variance* 1, (iii) *$E[F_{i,n}F_{j,n}] = 0$ for $i \neq j$, and* (iv) *the law of F_n admits a density f_n on $\mathbb{R}^d$. Write*

$$\Delta_n := \int_{\mathbb{R}^d} \|\mathbf{x}\|^4 (f_n(\mathbf{x}) - \phi_d(\mathbf{x}))d\mathbf{x},$$

where $\|\cdot\|$ stands for the Euclidean norm, and assume that $\Delta_n \to 0$, as $n \to \infty$. Then,

$$\int_{\mathbb{R}^d} f_n(\mathbf{x}) \log \frac{f_n(\mathbf{x})}{\phi_d(\mathbf{x})} d\mathbf{x} = O(1)\, \Delta_n |\log \Delta_n|, \tag{5}$$

where $O(1)$ stands for a bounded numerical sequence, depending on $d, q_1, ..., q_d$ and on the sequence $\{F_n\}$.

Corollary

Let $d \geq 1$ and $q_1, \ldots, q_d \geq 1$ be fixed integers. Consider vectors

$$F_n = (F_{1,n}, \ldots, F_{d,n}) = (I_{q_1}(h_{1,n}), \ldots, I_{q_d}(h_{d,n})), \quad n \geq 1,$$

with $h_{i,n} \in \mathfrak{H}^{\odot q_i}$. Let C_n denote the covariance matrix of F_n and let $Z_n \sim \mathcal{N}_d(0, C_n)$ be a centered Gaussian random vector in $\mathbb{R}^d$ with the same covariance matrix as F_n. Assume that $C_n \to C > 0$. Then, the following three assertions are equivalent, as $n \to \infty$:

(i) $\Delta_n \to 0$;

(ii) *F_n converges in distribution to $Z \sim \mathcal{N}_d(0, C)$;*

(iii) $D(F_n \| Z_n) \to 0$.