Entropic Fourth Moment Theorem

Ivan Nourdin (Université de Lorraine)
Giovanni Peccati (Université du Luxembourg)
Yvik Swan (Université du Luxembourg)

Warning!

I could have tried to speak about free Lévy process (which is a nice topic by the way), but...

Warning!

I could have tried to speak about free Lévy process (which is a nice topic by the way), but...

I am sorry to tell that this talk is going to be Lévy free!

What we want to do

Let F be a given random variable with, say, mean zero and variance 1. Let $N \sim N(0,1)$.

In many situations of interest, we may expect the law of F to be close of that of N. How to formalize this ?

Total variation distance

Let X, Y be two random variables. We define the *total variation distance* between (the laws of) X and Y as:

$$d_{TV}(X,Y) \;=\; \sup_{A \in \mathcal{B}(\mathbb{R})} \Big| P(X \in A) - P(Y \in A) \Big|.$$

Stein's method

Let F be a random variable with mean zero and unit variance. Let $N \sim N(0,1)$.

Then

$$d_{TV}(F,N) \leq \sup_{\substack{\|g\|_\infty \leq \sqrt{\frac{2}{\pi}} \\ \|g'\|_\infty \leq 2}} \Big| E[g'(F)] - E[Fg(F)] \Big|.$$

Nota: only valid in dimension 1 !

Score function and Fisher information

Assume that F has a density $p_F : \mathbb{R} \to [0, \infty)$. Provided it exists, define the *score function* of F as

$$s_F(x) = (\log p_F)'(x) = \frac{p_F'(x)}{p_F(x)}.$$

Then, for all test function $g : \mathbb{R} \to \mathbb{R}$, one calculates

$$E[g'(F)] \; = \; \int g' p_F = -\int g p_F' = -E[g(F) s_F(F)].$$

At this stage, it is useful to observe that, if for some W one has

$$E[g'(F)] = -E[g(F)W] \quad \text{for all test function } g : \mathbb{R} \to \mathbb{R},$$

then

$$s_F(x) = E[W|F = x].$$

(What is nice here is that one can express s_F without knowing the density of F.)

Let us go back to the Stein's inequality:

$$
\begin{aligned}
d_{TV}(F,N) &\leq \sqrt{\frac{2}{\pi}}\, E|F + s_F(F)| \\
&\overset{(*)}{\leq} \sqrt{\frac{2}{\pi}}\sqrt{J(F)-1}\ \left(= \sqrt{\frac{2}{\pi}}\sqrt{J(F)-J(N)} \right),
\end{aligned}
$$

where

$$
J(F) = E[s_F(F)^2] = \int \frac{(p_F')^2}{p_F}
$$

is the *Fisher information* of F.

(Inequality $(*)$ holds true since $E[F s_F(F)] = \int x p_F' = -\int p_F = -1$ so that $J(F) - 1 = E[(s_F(F) + F)^2]$.)

Stein's factor

Provided it exists, define the *Stein's factor* of F as

$$\tau_F(x) = \frac{1}{p_F(x)} \int_x^\infty y p_F(y) dy.$$

Then, for all test function $g : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} E[\tau_F(F) g'(F)] &= \int_{-\infty}^{+\infty} g'(x) \left(\int_x^\infty y p_F(y) dy \right) dx = \int x g(x) p_F(x) dx \\ &= E[F g(F)]. \end{aligned}$$

In particular, $E[\tau_F(F)] = E[F^2] = 1$, implying in turn $\mathrm{Var}(\tau_F(F)) = E[(1 - \tau_F(F))^2]$.

Exactly as for the score function, it is useful to observe here that, if for some W one has

$$E[Fg(F)] = E[g'(F)W] \quad \text{for all test function } g : \mathbb{R} \to \mathbb{R},$$

then

$$\tau_F(x) = E[W|F = x].$$

(What is nice here again is that one can express τ_F without knowing the density of F.)

Still using Stein's inequality, one deduce this time that

$$\begin{aligned} d_{TV}(F,N) &\leq 2\,E|1-\tau_F(F)| \quad \Big(= 2\,E|\tau_N(N)-\tau_F(F)| \Big) \\ &\leq 2\sqrt{\mathsf{Var}(\tau_F(F))}. \end{aligned}$$

Fourth Moment Theorem

Let F have the following form:

$$F = \sum_{k_1,\dots,k_p \geq 1} f(k_1,\dots,k_p) G_{k_1} \dots G_{k_p}$$

where $f : \mathbb{N}^p \to \mathbb{R}$ vanishes on diagonals, has compact support (and is symmetric with respect to its arguments), and where the $G_k \sim N(0,1)$ are independent. As before, we also assume that $E[F^2] = 1$. Then F has a density (Shigekawa) and the following inequality holds true (Nourdin-Peccati)

$$d_{TV}(F,N) \leq \frac{2}{\sqrt{3}} \sqrt{\left| E[F^4] - 3 \right|}. \qquad (\$)$$

(Rk1: as such, one recovers the celebrated Fourth Moment Theorem of Nualart and Peccati. Rk2: ($) may be improved.)

Proof of ($) consists in applying Stein's inequality, after observing that

$$\tau_F(F) = E\Big[\frac{1}{p}\|DF\|^2|F\Big]$$

implying in turn that

$$\mathsf{Var}(\tau_F(F)) \leq \frac{1}{p^2}\mathsf{Var}(\|DF\|^2) \leq \frac{1}{3}\Big(E[F^4] - 3\Big).$$

(Nota: we don't know any exact expression for the density of F; but it is actually not an issue, since we don't need it ;-)

(Relative) entropy

Let us go back to the general case, where F is any centered random variable with unit variance and a density p_F. Define the *entropy* of F as

$$\mathsf{Ent}(F) = -\int p_F \log p_F.$$

A straightforward calculation gives, with $N \sim N(0,1)$,

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) = \int p_F \log\left(\frac{p_F}{\phi}\right) \geq 0 \quad \text{(relative entropy } F\text{)}.$$

(This is because $\log\phi(x) = -x^2/2 - \log(\sqrt{2\pi})$, so that $\int \phi \log \phi = \int p_F \log \phi$.)

Pinsker inequality

$$2d_{TV}(F,N) \stackrel{\text{Scheffe}}{=} \int |p_F - \phi| \;\le\; \sqrt{2\Big(\mathsf{Ent}(N) - \mathsf{Ent}(F)\Big)}.$$

Valid in any dimension! The proof is nothing but an immediate consequence of the following inequality:

$$3(x-1)^2 \le (4+2x)(x\log x - x + 1), \qquad x > 0.$$

Indeed:

$$\Big|\frac{p_F}{\phi} - 1\Big| \le \frac{1}{\sqrt{3}}\sqrt{4 + 2\frac{p_F}{\phi}}\sqrt{\frac{p_F}{\phi}\log(\frac{p_F}{\phi}) - \frac{p_F}{\phi} + 1}$$

implying in turn, by Cauchy-Schwarz,

$$\int |p_F - \phi| = \int \Big|\frac{p_F}{\phi} - 1\Big|\phi \le \frac{1}{\sqrt{3}}\sqrt{\int (4 + 2\frac{p_F}{\phi})\phi}\sqrt{\int (\frac{p_F}{\phi}\log(\frac{p_F}{\phi}) - \frac{p_F}{\phi} + 1)\phi\,.}\Bigg)$$

de Bruijn identity

Set $F_t = \sqrt{t}F + \sqrt{1-t}N$ with $N \sim N(0,1)$ independent of F. Then

$$\mathrm{Ent}(N) - \mathrm{Ent}(F) = \int_0^1 \frac{J(F_t) - 1}{2t} dt.$$

On the other hand, the Fisher information satisfies a kind of convexity property: $J(F_t) \leq tJ(F) + (1-t)J(N) = tJ(F) + 1 - t$.

Combining the two, one gets

$$\mathrm{Ent}(N) - \mathrm{Ent}(F) \leq \frac{1}{2}(J(F) - 1).$$

A natural question

Do we have

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) \leq \mathsf{cst}\,\mathsf{Var}(\tau_F(F)) \quad ?$$

(This question is not for free! If it holds true, it would be particularly useful in dimension *strictly* greater than 1.)

Expressing the relative entropy by means of the Stein's factor

Theorem (Nourdin, Peccati, Swan) Set $F_t = \sqrt{t}F + \sqrt{1-t}N$ with $N \sim N(0,1)$ independent of F. Then

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) = \frac{1}{2}\int_0^1 \frac{t}{1-t} E\Big[E[N(1-\tau_F(F))|F_t]^2\Big] dt.$$

Corollary. One has

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) \leq \frac{1}{2}\mathsf{Var}(\tau_F(F)) \int_0^1 \frac{t}{1-t} dt.$$

Proof of the theorem

One wants to prove that

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) = \frac{1}{2}\int_0^1 \frac{t}{1-t} E\Big[E[N(1-\tau_F(F))|F_t]^2\Big]dt.$$

But don't forget (de Bruijn identity) that

$$\mathsf{Ent}(N) - \mathsf{Ent}(F) = \int_0^1 \frac{J(F_t)-1}{2t}dt$$

as well as

$$J(F_t) - 1 = E[(s_{F_t}(F_t) + F_t)^2].$$

Hence, to proof the theorem it suffices to check that

$$s_{F_t}(F_t) + F_t = -\frac{t}{\sqrt{1-t}} E[N(1-\tau_F(F))|F_t].$$

Proof continued

Let $g : \mathbb{R} \to \mathbb{R}$ be a test function. One has

$$
\begin{aligned}
& E[N(1-\tau_F(F))g(\sqrt{t}F+\sqrt{1-t}N)] \\
= \; & \sqrt{1-t}E[(1-\tau_F(F))g'(\sqrt{t}F+\sqrt{1-t}N)] \\
= \; & \sqrt{1-t}\Big\{E[g'(\sqrt{t}F+\sqrt{1-t}N)] - \frac{1}{\sqrt{t}}E[Fg(\sqrt{t}F+\sqrt{1-t}N)]\Big\} \\
= \; & \sqrt{1-t}\Big\{E[g'(F_t)] - \frac{1}{t}E[(F_t-\sqrt{1-t}N)g(F_t)]\Big\} \\
= \; & \sqrt{1-t}\Big\{E[g'(F_t)] - \frac{1}{t}E[F_tg(F_t)] + \frac{1-t}{t}E[g'(F_t)]\Big\} \\
= \; & \frac{\sqrt{1-t}}{t}\Big\{E[g'(F_t)] - E[F_tg(F_t)]\Big\} \\
= \; & -\frac{\sqrt{1-t}}{t}\,E[(s_{F_t}(F_t)+F_t)g(F_t)]. \qquad \blacksquare
\end{aligned}
$$

Going back to the silly corollary: a deeper analysis is needed!

Corollary. We let the previous notation prevail. Assume further that

$$E[|\tau_F(F)|^3] < \infty$$

and that there exist $c, \delta > 0$ such that

$$\forall t \in (0,1] : \quad E\Big[\Big|E[N(1-\tau_F(F))|F_t]\Big|\Big] \leq \frac{c}{t}(1-t)^\delta.$$

Then, provided that $\mathsf{Var}(\tau_F(F)) \leq 1$,

$$\begin{aligned} \mathsf{Ent}(N) - \mathsf{Ent}(F) &\leq \frac{1}{\delta}\mathsf{Var}(\tau_F(F))|\log(\mathsf{Var}(\tau_F(F))| \\ &+ \frac{2}{\delta}\sqrt{3c(1+E[|\tau_F(F)|^3])}\,\mathsf{Var}(\tau_F(F)). \end{aligned}$$

Proof

First step. One has, using Hölder as well as the convexity property of the cube function,

$$E\Big[E[N(1-\tau_F(F))|F_t]^2\Big] \leq \frac{2\sqrt{3c(1+E[|\tau_F(F)|^3])}}{t}(1-t)^{\delta/2}.$$

Second step. One has, using the NPS representation and with $0 < \varepsilon \leq 1$,

$$\begin{aligned} & 2(\mathrm{Ent}(N)-\mathrm{Ent}(F)) = \int_0^1 \frac{t}{1-t} E\Big[E[N(1-\tau_F(F))|F_t]^2\Big]dt \\ \leq\ & \mathrm{Var}(\tau_F(F))|\log(\varepsilon)| + \int_{1-\varepsilon}^1 \frac{t}{1-t} E\Big[E[N(1-\tau_F(F))|F_t]^2\Big]dt \\ \leq\ & \mathrm{Var}(\tau_F(F))|\log(\varepsilon)| + \mathrm{cst}\ \varepsilon^{\delta/2}. \end{aligned}$$

Third and final step. One optimises in $\varepsilon = \mathsf{Var}(\tau_F(F))^{2/\delta}$ and the proof is concluded! $\blacksquare$

How to check our assumptions?

Proposition. Assume, for some $\kappa > 0$ et $0 < \delta \leq \frac{1}{2}$, that

$$\forall x \in \mathbb{R},\, t \in (0,1]: \qquad d_{TV}(\sqrt{t}F + \sqrt{1-t}x, F) \leq \frac{\kappa}{t}(1+|x|)(1-t)^{\delta}.$$

Then

$$\forall t \in (0,1]: \qquad E\Big[\Big|E[N(1-\tau_F(F))|F_t]\Big|\Big] \leq \frac{4(\kappa+1)}{t}(1-t)^{\delta},$$

implying in turn, provided that $\mathrm{Var}(\tau_F(F)) \leq 1$,

$$\begin{aligned} \mathrm{Ent}(N) - \mathrm{Ent}(F) &\leq \frac{1}{\delta}\mathrm{Var}(\tau_F(F))|\log(\mathrm{Var}(\tau_F(F))| \\ &+ \frac{4}{\delta}\sqrt{3(\kappa+1)(1+E[|\tau_F(F)|^3])}\,\mathrm{Var}(\tau_F(F)). \end{aligned}$$

Proof

First step (key!). One observes that

$$
\begin{aligned}
& E\Big[\Big|E[N(1-\tau_F(F))|F_t]\Big|\Big] \\
= \;& E\Big[E[N(1-\tau_F(F))|F_t]\,\underbrace{\mathsf{sign}(E[N(1-\tau_F(F))|F_t])}_{F_t \text{ measurable}}\Big] \\
= \;& E\Big[N(1-\tau_F(F))\,\mathsf{sign}(E[N(1-\tau_F(F))|F_t])\Big] \\
\leq \;& \sup_{\|g\|_\infty \leq 1} E\Big[N(1-\tau_F(F))g(F_t)\Big] \\
= \;& \sup_{g\in C^1:\,\|g\|_\infty \leq 1} E\Big[N(1-\tau_F(F))g(F_t)\Big] \quad \text{(Urysohn, etc.)}
\end{aligned}
$$

Proof continued

Second step. Let $g \in C^1$ satisfying $\|g\|_\infty \leq 1$. Then

$$
\begin{aligned}
& E\Big[N(1-\tau_F(F))g(F_t)\Big] = E[Ng(F_t)] - E[N\tau_F(F)g(F_t)] \\
= \;& E[Ng(F_t)] - \sqrt{1-t}\,E[\tau_F(F)g'(F_t)] \\
= \;& E[N(g(F_t)-g(F))] - \underbrace{\frac{\sqrt{1-t}}{\sqrt{t}}E[Fg(F_t)]}_{ok!\ (\leq (1-t)^\delta\, t^{-1})}.
\end{aligned}
$$

Third step. One has

$$
\begin{aligned}
|E[N(g(F_t) - g(F))]| &= \left| \int_{\mathbb{R}} xE[g(\sqrt{t}F + \sqrt{1-t}x) - g(F)]\phi(x)dx \right| \\
&\leq 2 \int_{\mathbb{R}} |x| d_{TV}(\sqrt{t}F + \sqrt{1-t}x, F)\phi(x)dx \\
&\leq \frac{2\kappa}{t}(1-t)^{\delta} \int_{\mathbb{R}} |x|(1+|x|)\phi(x)dx \leq \frac{4\kappa}{t}(1-t)^{\delta}. \qquad \blacksquare
\end{aligned}
$$

Application to the 4th Moment Thm

Let (F_n) be a sequence of the form:

$$F_n = \sum_{k_1,\ldots,k_p=1}^{N_n} f_n(k_1,\ldots,k_p) G_{k_1} \ldots G_{k_p}$$

where each $f_n : \{1,\ldots,N_n\}^p \to \mathbb{R}$ vanishes on diagonals and is such that $E[F_n^2] = 1$, the $G_k \sim N(0,1)$ are independent and $N_n \to \infty$. Then [Nourdin-Poly],

$$\begin{aligned} d_{TV}(\sqrt{t}F_n + \sqrt{1-t}x, F_n) &\leq \text{cst } d_{FM}(\sqrt{t}F_n + \sqrt{1-t}x, F_n)^\delta \\ &\leq \text{cst } \left(E|(\sqrt{t}-1)F_n + \sqrt{1-t}x|\right)^\delta \\ &\leq \text{cst } (1+|x|)(1-t)^{\delta/2}. \end{aligned}$$

Suppose moreover that

$$\Delta_n := E[F_n^4] \to 3.$$

Then

$$F_n \stackrel{\text{loi}}{\to} N(0,1) \quad \text{[Nualart-Peccati]}$$

and

$$\text{Var}(\tau_{F_n}(F_n)) \leq \frac{\Delta_n}{3} \quad \text{[Nourdin-Peccati].}$$

By combining these two results, one gets that

$$\text{Ent}(N) - \text{Ent}(F_n) = \Delta_n |\log(\Delta_n)| \, O(1).$$

More generally, one has the following result.

Entropic Fourth Moment Theorem (Nourdin, Peccati, Swan): Let $d \geq 1$ and $p_1, \ldots, p_d \geq 1$ be integers. Let (F_n) be a sequence of random vectors $\mathbb{R}^d$ of the form:

$$F_{i,n} = \sum_{k_1,\ldots,k_{p_i}=1}^{N_{i,n}} f_{i,n}(k_1, \ldots, k_{p_i}) G_{k_1} \ldots G_{k_{p_i}}, \quad i = 1, \ldots, d,$$

where each $f_{i,n} : \{1, \ldots, N_{i,n}\}^{p_i} \to \mathbb{R}$ vanishes on diagonals and is symetric with respect to its arguments, the $G_k \sim N(0,1)$ are independent and $N_{i,n} \to \infty$. Suppose in addition that

$$E[F_{i,n} F_{j,n}] \to C(i,j)$$

with $C = (C(i,j))_{1 \leq i,j \leq d}$ invertible, and that

$$\Delta_n := E[\|F_n\|^4] - E[\|N\|^4] \to 0,$$

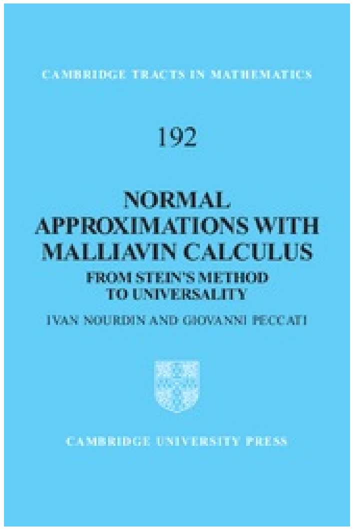
where $N \sim N_d(0, C)$. Then

$$\mathrm{Ent}(N) - \mathrm{Ent}(F_n) = \Delta_n |\log(\Delta_n)| \, O(1).$$

In particular (by Pinsker)

$$d_{TV}(F_n, N) = \sqrt{\Delta_n |\log(\Delta_n)|}\, O(1).$$

To conclude, a bit of advertisement ;-)



I. Nourdin and G. Peccati (2012): *Normal Approximations with Malliavin Calculus: from Stein's Method to Universality*. Cambridge University Press (Cambridge Tracts in Mathematics)