

Maths \Rightarrow IA \Rightarrow Maths

Ivan Nourdin

Uni. du Luxembourg

Masterclass – Strasbourg, 23-24 avril 2026

- 1 **Méthode de Stein et approximation normale**
- 2 **Réseaux de neurones profonds**
- 3 **Les LLMs comme assistants de recherche en mathématiques**

Partie 1

Méthode de Stein et approximation normale

Approximation normale : dimension 1

Soit F une variable aléatoire quelconque.

- $F \sim N(0, 1) \iff E[e^{itF}] = e^{-\frac{t^2}{2}} \quad \forall t \in \mathbb{R}$

En extrapolant : $F \simeq N(0, 1) \iff E[e^{itF}] \simeq e^{-\frac{t^2}{2}} \quad \forall t \in \mathbb{R}.$

Théorème

Soit (F_n) une suite de variables aléatoires. Alors :

$$F_n \xrightarrow{\text{loi}} N(0, 1) \iff E[e^{itF_n}] \rightarrow e^{-\frac{t^2}{2}} \quad \forall t \in \mathbb{R}.$$

- $F \sim N(0, 1) \stackrel{(\text{Stein})}{\iff} E[F\phi(F)] = E[\phi'(F)] \quad \text{“}\forall\text{”} \phi : \mathbb{R} \rightarrow \mathbb{R}$
(‘ \Rightarrow ’ i.p.p. ‘ \Leftarrow ’ $g(t) = E(e^{itF})$ vérifie une équation différentielle)

En extrapolant : $F \simeq N(0, 1) \iff E[F\phi(F)] \simeq E[\phi'(F)]$ pour une large classe de fonctions ϕ .

Lemme de Stein

Soit $N \sim N(0, 1)$. Soit $h : \mathbb{R} \rightarrow [0, 1]$ continue. Alors

$$\begin{aligned}\phi(x) &= e^{\frac{x^2}{2}} \int_{-\infty}^x (h(a) - E(h(N))) e^{-\frac{a^2}{2}} da \\ &= -e^{\frac{x^2}{2}} \int_x^{\infty} (h(a) - E(h(N))) e^{-\frac{a^2}{2}} da\end{aligned}$$

est C^1 , vérifie $\phi'(x) = x\phi(x) + h(x) - E(h(N))$ et est telle que $|\phi'(x)| \leq 2$ $\forall x \in \mathbb{R}$.

Preuve. Pour $x > 0$: $|x\phi(x)| \leq xe^{\frac{x^2}{2}} \int_x^{\infty} e^{-\frac{a^2}{2}} da \leq e^{\frac{x^2}{2}} \int_x^{\infty} a e^{-\frac{a^2}{2}} da = 1$.
Idem pour $x < 0$ en utilisant l'autre représentation. \square

Théorème de Stein

Soit $N \sim N(0, 1)$ et soit F une variable aléatoire de carré intégrable. Alors

$$\begin{aligned}d_{TV}(F, N) &:= \sup_{A \in \mathcal{B}(\mathbb{R})} |P(F \in A) - P(N \in A)| \\ &\leq \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2}} |E[\phi'(F)] - E[F\phi(F)]| =: \Delta.\end{aligned}$$

Est-ce une 'bonne' borne?

Notons

$$\Delta = \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2}} |E[\phi'(F)] - E[F\phi(F)]|.$$

1) Soit $h : \mathbb{R} \rightarrow [0, 1]$ continue. Soit ϕ donnée dans le lemme de Stein.

Lemme de Stein

Soit $N \sim N(0, 1)$. Soit $h : \mathbb{R} \rightarrow [0, 1]$ continue. Alors $\phi(x) = (\dots)$ est C^1 , vérifie $\phi'(x) = x\phi(x) + h(x) - E(h(N))$ et est telle que $|\phi'(x)| \leq 2 \forall x \in \mathbb{R}$.

On a $|E[h(F)] - E[h(N)]| = |E[\phi'(F)] - E[F\phi(F)]| \leq \Delta$.

1) Soit $h : \mathbb{R} \rightarrow [0, 1]$ continue. Soit ϕ donnée dans le lemme de Stein. On a $|E[h(F)] - E[h(N)]| = |E[\phi'(F)] - E[F\phi(F)]| \leq \Delta$.

2) Soit $B \in \mathcal{B}(\mathbb{R})$ borné. Soit $M > 0$ tel que $B \subset [-M, M]$. Par Lusin avec $\mu = (\text{Leb} + P_F)|_{[-M, M]}$ (qui est une mesure finie sur $[-M, M]$), il existe $h_n : \mathbb{R} \rightarrow [0, 1]$ continue, $\text{supp } h_n \subset [-M, M]$, telle que

$\mathbf{1}_B(x) = \lim_{n \rightarrow \infty} h_n(x)$ μ -p.p. (c'est-à-dire Leb-p.p. et P_F -p.p.)

Par convergence dominée : $E[h_n(F)] \rightarrow P(F \in B)$ et

$E[h_n(N)] \rightarrow P(N \in B)$. D'où (par **1**) $|P(F \in B) - P(N \in B)| \leq \Delta$.

3) Soit $A \in \mathcal{B}(\mathbb{R})$ quelconque. Par **2**), on a, pour tout $m \in \mathbb{N}$:

$$|P(F \in A \cap [-m, m]) - P(N \in A \cap [-m, m])| \leq \Delta.$$

Par convergence monotone, il vient $|P(F \in A) - P(N \in A)| \leq \Delta$. □

Si (F_n) est une suite de variables aléatoires, il est bien entendu plus fort d'avoir $F_n \xrightarrow{d_{TV}} F_\infty$ que d'avoir $F_n \xrightarrow{\text{loi}} F_\infty$.

Contre-exemple facile : $F_n = \frac{X_1 + \dots + X_n}{\sqrt{n}}$, où les

$X_i = \begin{cases} 1 & \text{prob. } \frac{1}{2} \\ -1 & \text{"} \end{cases}$ sont indépendantes.

Alors $F_n \xrightarrow{\text{loi}} N \sim N(0, 1)$ (TCL). Mais $P(F_n \in \mathbb{Z}/\sqrt{n}) = 1$ tandis que $P(N \in \mathbb{Z}/\sqrt{n}) = 0$, si bien que $d_{TV}(F_n, N) = 1 \forall n$.

Et si F_n et F_∞ ont une densité ?

Théorème (Scheffé)

Si $F \sim f(x) dx$ et $G \sim g(x) dx$, alors :

$$d_{TV}(F, G) = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx.$$

Preuve du théorème de Scheffé: $d_{TV}(F, G) = \frac{1}{2} \int |f - g|$

Preuve. Posons $B = \{f \geq g\}$. Soit $A \in \mathcal{B}(\mathbb{R})$. On a

$$\int_A (f - g) = \underbrace{\int_{A \cap B} (f - g)}_{\geq 0} + \underbrace{\int_{A \cap B^c} (f - g)}_{\leq 0} \leq \int_B (f - g) = \int (f - g)_+$$

Or
$$\left. \begin{aligned} 0 &= \int (f - g) = \int (f - g)_+ - \int (f - g)_- \\ \int |f - g| &= \int (f - g)_+ + \int (f - g)_- \end{aligned} \right\} \text{d'où}$$

$$\int (f - g)_+ = \frac{1}{2} \int |f - g|.$$

Ainsi $\int_A (f - g) \leq \frac{1}{2} \int |f - g|$. De manière symétrique,
 $\int_A (g - f) \leq \frac{1}{2} \int |f - g|$. D'où $|\int_A (f - g)| \leq \frac{1}{2} \int |f - g|$, et ainsi

$$d_{TV}(F, G) = \sup_{A \in \mathcal{B}(\mathbb{R})} \left| \int_A (f - g) \right| \leq \frac{1}{2} \int |f - g|.$$

Preuve du théorème de Scheffé (fin)

Par ailleurs, $\int_B (f - g) = \int (f - g)_+ = \frac{1}{2} \int |f - g|$, d'où

$$d_{TV}(F, G) = \frac{1}{2} \int |f - g|.$$



Corollaire 1 (immédiat)

Soit $(F_n)_{n \in \mathbb{N} \cup \{\infty\}}$ une suite de variables aléatoires à densité (notées $f_n : \mathbb{R} \rightarrow \mathbb{R}$). Si $f_n \rightarrow f_\infty$ dans $L^1(\mathbb{R})$ alors $d_{TV}(F_n, F_\infty) \rightarrow 0$, et on a en particulier convergence en loi.

Corollaire 2

Soit $(F_n)_{n \in \mathbb{N} \cup \{\infty\}}$ une suite de variables aléatoires à densité (notées $f_n : \mathbb{R} \rightarrow \mathbb{R}$). Si $f_n \rightarrow f_\infty$ p.p. alors $d_{TV}(F_n, F_\infty) \rightarrow 0$, et on a en particulier convergence en loi.

Preuve. On va montrer que $f_n \xrightarrow{L^1(\mathbb{R})} f_\infty$, ce qui suffit d'après le Corollaire 1. On a déjà vu que

$$\int |f_n - f_\infty| = 2 \int (f_\infty - f_n)_+.$$

Or $0 \leq (f_\infty - f_n)_+ \leq f_\infty$ avec $f_\infty \in L^1(\mathbb{R})$ [si $x, y \geq 0$ alors $(x - y)_+ \leq x$]. La conclusion suit par convergence dominée. \square

Un contre-exemple moins facile

Soient $F_n \sim \frac{2}{\pi} \cos^2(nx) \mathbf{1}_{[0,\pi]}(x) dx$ et $F_\infty \sim \frac{1}{\pi} \mathbf{1}_{[0,\pi]}(x) dx$.

Pour tout $\phi : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée :

$$\begin{aligned} E[\phi(F_n)] &= \frac{2}{\pi} \int_0^\pi \phi(x) \cos^2(nx) dx \\ &= \frac{1}{\pi} \int_0^\pi \phi(x) dx + \frac{1}{\pi} \int_0^\pi \phi(x) \cos(2nx) dx \\ &= E[\phi(F_\infty)] + o(1) \quad (\text{Riemann-Lebesgue}) \end{aligned}$$

D'où $F_n \xrightarrow{\text{loi}} F_\infty \sim \mathcal{U}_{[0,\pi]}$.

Or $d_{TV}(F_n, F_\infty) = \frac{1}{2} \int |f_n - f_\infty| = \frac{1}{2\pi} \int_0^\pi |\cos(2nx)| dx =$
 $\frac{1}{4\pi n} \int_0^{2n\pi} |\cos(u)| du = \frac{1}{4\pi} \int_0^{2\pi} |\cos(u)| du$ (ne dépend pas de n).

Théorème (Dini)

On considère $g_n : [a, b] \rightarrow \mathbb{R}$ pour $n \in \mathbb{N} \cup \{\infty\}$, avec $a < b$ et $a, b \in \mathbb{R} \cup \{\pm\infty\}$.

(Si $a = -\infty$, on pose $g_n(a) = \lim_{x \rightarrow -\infty} g_n(x)$ en supposant que la limite existe pour tout n . Idem pour $b = +\infty$.)

On suppose :

- (1) $\forall x \in [a, b], g_n(x) \xrightarrow[n \rightarrow \infty]{} g_\infty(x)$ (convergence simple)
- (2) $\forall x, y \in [a, b], \forall n \in \mathbb{N} : x \leq y \Rightarrow g_n(x) \leq g_n(y)$ (croissant)
- (3) g_∞ est continue.

Alors : $\sup_{x \in [a, b]} |g_n(x) - g_\infty(x)| \xrightarrow[n \rightarrow \infty]{} 0$ (convergence uniforme)

Application. Si $X_n \xrightarrow{\text{loi}} X_\infty$ avec X_∞ à densité (e.g. $X_\infty \sim N(0, 1)$), alors $d_{\text{Kol}}(X_n, X_\infty) := \sup_{x \in \mathbb{R}} |P[X_n \leq x] - P[X_\infty \leq x]| \rightarrow 0$.

On suppose $F : \Omega \rightarrow \mathbb{R}$ à densité f , avec $f(x) > 0 \forall x \in \mathbb{R}$, $xf(x) \rightarrow 0$ quand $|x| \rightarrow \infty$ et $f \in C^1(\mathbb{R})$ (pas optimal!)

Définition (score)

Le score de F est la fonction

$$s_F(x) = \frac{f'(x)}{f(x)}, \quad x \in \mathbb{R}$$

Proposition

Le score est l'unique fonction $s : \mathbb{R} \rightarrow \mathbb{R}$ satisfaisant

$$E\{\phi'(F)\} = -E\{s(F)\phi(F)\} \quad \forall \phi \in C_c^1(\mathbb{R})$$

Preuve. $E\{\phi'(F)\} = \int_{\mathbb{R}} \phi'(x)f(x) dx \stackrel{\text{i.p.p.}}{=} - \int_{\mathbb{R}} \phi(x)f'(x) dx = - \int_{\mathbb{R}} \phi(x) \frac{f'(x)}{f(x)} f(x) dx = -E\{s_F(F)\phi(F)\}.$

Si $E\{s(F)\phi(F)\} = E\{s_F(F)\phi(F)\} \quad \forall \phi$ alors

$$\int (s(x) - s_F(x))\phi(x)f(x) dx = 0 \quad \forall \phi \text{ et } s = s_F. \quad \square$$

Intuition. Si $s_F(F) \approx -F$ alors $F \approx N(0, 1).$

Définition (information de Fisher)

L'information de Fisher de F est :

$$I(F) = \int_{\mathbb{R}} \frac{f'(x)^2}{f(x)} dx$$

Remarque. On a $E\{s_F(F)\} = \int_{\mathbb{R}} f'(x) dx = 0$ et $I(F) = E\{s_F(F)^2\} = \text{Var}\{s_F(F)\}$.

Proposition

$$I(F + a) = I(F) \quad \forall a \in \mathbb{R} \quad I(aF) = \frac{1}{a^2} I(F)$$

Preuve. $I(F + a) = \int \frac{f'(x - a)^2}{f(x - a)} dx = I(F)$.

$$I(aF) = \int_{\mathbb{R}} \frac{\left(\frac{1}{a} f'\left(\frac{x}{a}\right)\right)^2}{\frac{1}{a} f\left(\frac{x}{a}\right)} dx = \frac{1}{a^2} \int_{\mathbb{R}} \frac{f'\left(\frac{x}{a}\right)^2}{f\left(\frac{x}{a}\right)} \frac{dx}{a} = \frac{1}{a^2} I(F).$$



Théorème (Cramér-Rao)

$$I(F) \geq \frac{1}{\text{Var}(F)} \quad (*)$$

Preuve. 1) Supposons le résultat déjà montré quand F est centrée réduite. Alors le résultat est vrai pour toute F . Pour le voir: si $\mu = EF$ et

$$\sigma^2 = \text{Var}F \text{ alors } I(F) = I(F - \mu) = \frac{1}{\sigma^2} I\left(\frac{F - \mu}{\sigma}\right) \geq \frac{1}{\sigma^2}.$$

2) On peut donc supposer F centrée réduite. Calculons :

$$E\{(F + s_F(F))^2\} = 1 + 2 \underbrace{E\{Fs_F(F)\}}_{=-1} + I(F) = I(F) - 1 \geq 0.$$

□

Théorème

$$I(F) = \frac{1}{\text{Var}F} \iff F \text{ est gaussienne.}$$

Preuve. 1) Sans perte de généralité, on peut supposer que F est centré réduite.

2) $I(F) = 1 \xrightarrow{\text{preuve précédente}} F = -s_F(F) \text{ p.s.}$

$$\Rightarrow x = -s_F(x) \forall x \Rightarrow \ln f(x) = -\frac{x^2}{2} + \text{cst} \Rightarrow f(x) = \text{cst} \cdot e^{-x^2/2}.$$

□

Théorème (Stein & Fisher)

Si $EF = 0$ et $\text{Var}F = 1$ (pour simplifier), alors :

$$d_{TV}(F, N) \leq \sqrt{\frac{\pi}{2}} \sqrt{I(F) - 1}.$$

Lemme de Stein (rappel)

Soit $N \sim N(0, 1)$ et $F \in L^2(\Omega)$. Alors :

$$d_{TV}(F, N) \leq \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2}} |E[\phi'(F)] - E[F\phi(F)]|$$

En fait, on a mieux :

$$d_{TV}(F, N) \leq \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2 \\ |\phi|_\infty \leq \sqrt{\pi/2}}} |E[\phi'(F)] - E[F\phi(F)]| \quad (*)$$

Preuve de la borne $|\phi|_\infty \leq \sqrt{\pi/2}$

La démonstration est la même, une fois observé que

$$\begin{aligned}\phi(x) &= e^{\frac{x^2}{2}} \int_{-\infty}^x (h(a) - E(h(N))) e^{-\frac{a^2}{2}} da \\ &= -e^{\frac{x^2}{2}} \int_x^\infty (h(a) - E(h(N))) e^{-\frac{a^2}{2}} da\end{aligned}$$

vérifie $|\phi|_\infty \leq \sqrt{\pi/2}$ pour toute $h : \mathbb{R} \rightarrow [0, 1]$ continue.

$$\text{En effet : } |\phi(x)| \leq \begin{cases} e^{\frac{x^2}{2}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \\ e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{y^2}{2}} dy \end{cases} \quad \text{d'où } \phi(x) \leq e^{\frac{x^2}{2}} \int_{|x|}^\infty e^{-\frac{y^2}{2}} dy.$$

Or $\lambda : x \mapsto e^{\frac{x^2}{2}} \int_{|x|}^\infty e^{-\frac{y^2}{2}} dy$ est paire et vérifie, pour $x \geq 0$:

$$\lambda'(x) = x e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{y^2}{2}} dy - 1 \leq e^{\frac{x^2}{2}} \int_x^\infty y e^{-\frac{y^2}{2}} dy - 1 = 0.$$

D'où :

$$\begin{array}{c|ccc} x & -\infty & 0 & +\infty \\ \hline \lambda(x) & \nearrow & & \searrow \end{array} \quad \text{avec} \quad \lambda(0) = \int_0^\infty e^{-\frac{y^2}{2}} dy = \sqrt{\frac{\pi}{2}}.$$

Preuve.

$$\begin{aligned}d_{TV}(F, N) &\leq \sup_{|\phi|_\infty \leq \sqrt{\pi/2}} |E\{\phi'(F)\} - E\{F\phi(F)\}| \\&= \sup_{|\phi|_\infty \leq \sqrt{\pi/2}} |E\{\phi(F)(F + s_F(F))\}| \\&\leq \sqrt{\frac{\pi}{2}} E|F + s_F(F)| \leq \sqrt{\frac{\pi}{2}} \sqrt{E\{(F + s_F(F))^2\}} \\&= \sqrt{\frac{\pi}{2}} \sqrt{I(F) - 1}\end{aligned}$$



Exemple d'application. Cas où F a la forme $F_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$, où $X_k : \Omega \rightarrow \mathbb{R}$ i.i.d. à densité qui va bien, $E(X_k) = 0$, $\text{Var}(X_k) = 1$.

On aimerait montrer que $d_{TV}(F_n, N) = O\left(\frac{1}{\sqrt{n}}\right)$.

Peut-on l'obtenir de Stein & Fisher ?

Sous-additivité de l'information de Fisher

On va d'abord montrer que si U et V sont indépendantes, alors pour tout $\beta \in [0, 1]$:

$$I(\sqrt{\beta} U + \sqrt{1-\beta} V) \leq \beta I(U) + (1-\beta) I(V).$$

Preuve.

$$\begin{aligned} E\{\phi'(\sqrt{\beta} U + \sqrt{1-\beta} V)\} &= \frac{1}{\sqrt{\beta}} E\{E\{\sqrt{\beta} \phi'(\sqrt{\beta} U + \sqrt{1-\beta} V)\} |_{v=V}\} \\ &= \frac{1}{\sqrt{\beta}} E\{s_U(U) \phi(\sqrt{\beta} U + \sqrt{1-\beta} V)\} \end{aligned}$$

De même :

$$E\{\phi'(\sqrt{\beta} U + \sqrt{1-\beta} V)\} = \frac{1}{\sqrt{1-\beta}} E\{s_V(V) \phi(\sqrt{\beta} U + \sqrt{1-\beta} V)\}.$$

$$\begin{aligned} D'où E\{\phi'(\sqrt{\beta} U + \sqrt{1-\beta} V)\} &= \\ E\{(\sqrt{\beta} s_U(U) + \sqrt{1-\beta} s_V(V)) \phi(\sqrt{\beta} U + \sqrt{1-\beta} V)\}, \\ \Rightarrow s_{\sqrt{\beta} U + \sqrt{1-\beta} V}(\sqrt{\beta} U + \sqrt{1-\beta} V) &= \\ E\{\sqrt{\beta} s_U(U) + \sqrt{1-\beta} s_V(V) | \sqrt{\beta} U + \sqrt{1-\beta} V\}. \end{aligned}$$

$$\begin{aligned} \Rightarrow & I(\sqrt{\beta}U + \sqrt{1-\beta}V) \\ &= E\{E\{\sqrt{\beta}s_U(U) + \sqrt{1-\beta}s_V(V) \mid \sqrt{\beta}U + \sqrt{1-\beta}V\}^2\} \\ &\leq E\{(\sqrt{\beta}s_U(U) + \sqrt{1-\beta}s_V(V))^2\} \quad (\text{Jensen}) \\ &= \text{Var}(\sqrt{\beta}s_U(U) + \sqrt{1-\beta}s_V(V)) \quad (\text{centré}) \\ &= \beta I(U) + (1-\beta)I(V). \quad \square \end{aligned}$$

En posant $U = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$, $V = \frac{1}{\sqrt{m}} \sum_{k=n+1}^{n+m} X_k$ et $\beta = \frac{n}{n+m}$:

$$n I(F_n) + m I(F_m) \geq (n+m) I(F_{n+m}).$$

Par Fekete (si $a_{n+m} \leq a_n + a_m$ alors $\lim \frac{a_n}{n}$ existe dans $\mathbb{R} \cup \{-\infty\}$), ceci implique que $(I(F_n))_{n \geq 1}$ converge ☺

Sauf que ça ne dit pas que cette limite est 1 ☹ (c'est vrai si $I(X_1) < \infty$, mais ça demande un peu – beaucoup! – plus de travail!)

Noyau de Stein : mieux adapté !

Score : $E\{\phi'(F)\} = -E\{s_F(F)\phi(F)\} \quad \forall \phi \in C_c^1(\mathbb{R}).$

Noyau de Stein

$$\tau_F(x) = \frac{\int_x^\infty u f(u) du}{f(x)} \quad (\text{OK si } F \in L^1(\Omega))$$

Proposition

Supposons $E[F] = 0$. Le noyau de Stein est l'unique fonction $\tau : \mathbb{R} \rightarrow \mathbb{R}$ satisfaisant

$$E\{F\phi(F)\} = E\{\tau(F)\phi'(F)\} \quad \forall \phi \in C_c^1(\mathbb{R})$$

Preuve. $E\{F\phi(F)\} = \int_{\mathbb{R}} x\phi(x)f(x) dx \stackrel{\text{i.p.p.}}{=} \int_{\mathbb{R}} \phi'(x)\tau_F(x)f(x) dx = E\{\tau_F(F)\phi'(F)\}$. L'unicité se traite comme pour le score. □

Théorème (Stein)

Si $E(F) = 0$, alors :

$$d_{TV}(F, N) \leq 2 E|1 - \tau_F(F)|$$

Si de plus $\text{Var}F = 1$, alors : $d_{TV}(F, N) \leq 2\sqrt{\text{Var}(\tau_F(F))}$.

Preuve. On a

$$\begin{aligned} d_{TV}(F, N) &\leq \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2}} |E[\phi'(F)] - E[F\phi(F)]| \\ &= \sup_{\substack{\phi \in C^1 \\ |\phi'|_\infty \leq 2}} |E[\phi'(F)(1 - \tau_F(F))]| \leq 2 E|1 - \tau_F(F)|. \end{aligned}$$

Pour la 2ème inégalité, il suffit d'observer que $E[\tau_F(F)] = \text{Var}(F)$ et d'appliquer Cauchy-Schwarz.

Retour à $F_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$, avec X_k i.i.d. centrée réduite

$$\begin{aligned} E\{F_n \phi(F_n)\} &= \frac{1}{\sqrt{n}} \sum_{k=1}^n E \left\{ X_k \phi \left(\frac{1}{\sqrt{n}} X_k + \underbrace{\frac{1}{\sqrt{n}} \sum_{j \neq k} X_j}_{\perp\!\!\!\perp \text{ de } X_k} \right) \right\} \\ &= \frac{1}{n} \sum_{k=1}^n E\{\tau_{X_k}(X_k) \phi'(F_n)\} \end{aligned}$$

D'où :

$$\tau_{F_n}(F_n) = E \left\{ \frac{1}{n} \sum_{k=1}^n \tau_{X_k}(X_k) \mid F_n \right\}$$

Retour à $F_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$, avec X_k i.i.d. centrée réduite

On en déduit :

$$\tau_{F_n}(F_n) - 1 = E \left\{ \frac{1}{n} \sum_{k=1}^n (\tau_{X_k}(X_k) - 1) \mid F_n \right\}$$

$$E \{ (\tau_{F_n}(F_n) - 1)^2 \} = E \{ E \{ \dots \mid F_n \}^2 \} \leq_{\text{Jensen}} E \left\{ \left(\frac{1}{n} \sum_{k=1}^n (\tau_{X_k}(X_k) - 1) \right)^2 \right\}$$

$$= \text{Var} \left(\frac{1}{n} \sum_{k=1}^n \tau_{X_k}(X_k) \right) = \boxed{\frac{1}{n}} \text{Var}(\tau_{X_1}(X_1)).$$

Grâce au théorème de Stein, on obtient notre taux en $\frac{1}{\sqrt{n}}$ à condition bien sûr que $\text{Var}(\tau_{X_1}(X_1)) < \infty$ ☺

Cas où F est de la forme $F_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k N_k$ avec X_k iid centrées réduites et $N_k \sim N(0, 1)$ iid indépendantes

“Rappel”: $E\{N\phi(N)\} = E\{\phi'(N)\}$ si $N \sim N(0, 1)$ et $\phi \in C_c^1(\mathbb{R})$.

On a :

$$\begin{aligned} E\{F_n\phi(F_n)g(X_1, \dots, X_n)\} &= \frac{1}{\sqrt{n}} \sum_{k=1}^n E\{N_k\phi(F_n)X_k g(X_1, \dots, X_n)\} \\ &= \frac{1}{n} \sum_{k=1}^n E\{\phi'(F_n)X_k^2 g(X_1, \dots, X_n)\} \\ &= E\left\{ \phi'(F_n) \underbrace{\frac{1}{n} \sum_{k=1}^n X_k^2}_{=: A} g(X_1, \dots, X_n) \right\} \end{aligned}$$

Théorème (Favaro, Hanin, Marinucci, Nourdin, Peccati, 2025)

On suppose $E[F] = 0$, $\text{Var}(F) = 1$ et qu'il existe une v.a. $A \geq 0$ telle que

$$E\{F\phi(F)\psi(A)\} = E\{A\phi'(F)\psi(A)\} \quad \forall \phi, \psi \in C_c^1(\mathbb{R}).$$

Alors :

- 1 $\text{Var}(A) = \frac{1}{3}(E(F^4) - 3)$ (\leftarrow 4^e cumulant de F)
- 2 $d_{TV}(F, N) \leq \frac{8}{3}(E(F^4) - 3)$

Digression sur les cumulants

Supposons, pour simplifier, $E|F|^m < \infty \forall m$. On a

$$E\{F^m\} = (-i)^m \frac{d^m}{dt^m} E\{e^{itF}\}.$$

Définition (cumulants)

$$\kappa_m(F) = (-i)^m \frac{d^m}{dt^m} \underbrace{\log}_{\substack{\downarrow \\ \text{branche principale} \\ \text{du log} \Rightarrow \text{pas de pb} \\ \text{autour de 0}}} E\{e^{itF}\}$$

Supposons $E[F] = 0$ pour simplifier. Alors

$$E\{e^{itF}\} = 1 - \frac{t^2}{2} E(F^2) - i \frac{t^3}{6} E(F^3) + \frac{t^4}{24} E(F^4) + \dots$$

$$\Rightarrow \log E\{e^{itF}\} \approx -\frac{t^2}{2} E(F^2) - i \frac{t^3}{6} E(F^3) + \frac{t^4}{24} (E(F^4) - 3(E(F^2))^2) + \dots$$

$$\text{(en utilisant que } \log(1+u) \approx u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} \dots \text{)}$$

Cumulants (suite)

Si $E[F] = 0$ alors :

$$\kappa_1(F) = 0, \kappa_2(F) = \text{Var}F, \kappa_3(F) = E[F^3], \kappa_4(F) = E[F^4] - 3\text{Var}(F)^2, \dots$$

Proposition

- Si $\lambda \in \mathbb{R}$: $\kappa_m(\lambda F) = \lambda^m \kappa_m(F)$
- Si $F \perp\!\!\!\perp G$: $\kappa_m(F + G) = \kappa_m(F) + \kappa_m(G)$
- Si $F \sim N(\mu, \sigma^2)$: $\kappa_1(F) = \mu$, $\kappa_2(F) = \sigma^2$, $\kappa_m(F) = 0 \forall m \geq 3$
(réciproque vraie)

Retour à $F_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k N_k$ où $N_k \sim N(0, 1)$ i.i.d., X_k i.i.d. centrée réduite, et $(X_1, \dots, X_n) \perp\!\!\!\perp (N_1, \dots, N_n)$.

$$E[F_n^4] - 3 = \kappa_4(F_n) = \frac{1}{n^2} \kappa_4\left(\sum_{k=1}^n X_k N_k\right) = \boxed{\frac{1}{n}} \kappa_4(X_1 N_1).$$

Notre théorème donne : $d_{TV}(F_n, N) = O(1/n)$.

Preuve du théorème F-H-M-N-P

1)

$$\begin{aligned}d_{TV}(F, N) &\leq 2 E\{|E(1 - A | F)|\} \quad (\text{car } E\{F\phi(F)\} = E\{A\phi'(F)\} \text{ avec } \psi = 1) \\ &= 2 E\{\text{sgn}(E(1 - A | F)) \cdot E[1 - A | F]\} \\ &\leq 2 \sup_{|g| \leq 1} E\{g(F)E[1 - A | F]\} = \sup_{|g| \leq 1} E\{g(F)(1 - A)\}\end{aligned}$$

$$\begin{aligned}2) E\{g(F)(1 - A)\} &\stackrel{E(A)=1}{=} E\{(g(F) - Eg(N))(1 - A)\} \\ &\stackrel{\phi_g \text{ défini du lemme de Stein}}{=} E\{(\phi'_g(F) - F\phi_g(F))(1 - A)\} \\ &= E\{\phi'_g(F)(1 - A)\} - E\{F\phi_g(F)(1 - A)\} \\ &\stackrel{\text{par hypothèse}}{=} E\{\phi'_g(F)(1 - A)\} - E\{A\phi'_g(F)(1 - A)\} \\ &= E\{\phi'_g(F)(1 - A)^2\}\end{aligned}$$

$\Rightarrow d_{TV}(F, N) \leq 8 E\{(1 - A)^2\} = 8 \text{Var}A \Rightarrow 2)$ du théorème.

($|\phi'_g|_\infty \leq 4$, au lieu de 2, car $g : \mathbb{R} \rightarrow [-1, 1]$)

Preuve du théorème (suite) : calcul de $\kappa_4(F) = 3 \text{Var}A$

3) Posons $\gamma(t) = E\{e^{itF} \mid A\}$.

On a alors $\gamma'(t) = i E\{Fe^{itF} \mid A\} \stackrel{\text{hyp}}{=} -At E\{e^{itF} \mid A\} = -At \gamma(t)$.

D'où $\gamma(t) = e^{-A\frac{t^2}{2}}$ et $E\{e^{itF}\} = E\{e^{-A\frac{t^2}{2}}\}$.

Ainsi :

$$\sum_{n \geq 1} \kappa_n(F) \frac{(it)^n}{n!} = \log E\{e^{itF}\} = \log E\{e^{-A\frac{t^2}{2}}\} = \sum_{m \geq 1} \kappa_m(A) \left(-\frac{t^2}{2}\right)^m \frac{1}{m!}$$

$n = 4$ et $m = 2$:

$$\frac{\kappa_4(F)}{4!} = \frac{1}{2^2 \cdot 2!} \kappa_2(A) \quad \Rightarrow \quad \kappa_4(F) = 3 \text{Var}A. \quad \square$$

Partie 2

Réseaux de neurones profonds

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} de la forme

$$W_1 x + b_1$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1)

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} de la forme

$$\sigma(W_1 x + b_1)$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1),
 $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} de la forme

$$\sigma(W_2 \sigma(W_1 x + b_1) + b_2)$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1),
 $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)
- $W_2 \in \mathcal{M}_{n_2, n_1}(\mathbb{R})$ (poids 2), $b_2 \in \mathbb{R}^{n_2}$ (biais 2)

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} de la forme

$$\sigma(\dots\sigma(W_2 \sigma(W_1 x + b_1) + b_2)\dots)$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1),
 $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)
- $W_2 \in \mathcal{M}_{n_2, n_1}(\mathbb{R})$ (poids 2), $b_2 \in \mathbb{R}^{n_2}$ (biais 2)
- ...

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} de la forme

$$\sigma(W_L \sigma(\dots \sigma(W_2 \sigma(W_1 x + b_1) + b_2) \dots) + b_L)$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1),
 $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)
- $W_2 \in \mathcal{M}_{n_2, n_1}(\mathbb{R})$ (poids 2), $b_2 \in \mathbb{R}^{n_2}$ (biais 2)
- ...
- $W_L \in \mathcal{M}_{n_L, n_{L-1}}(\mathbb{R})$ (poids L), $b_L \in \mathbb{R}^{n_L}$ (biais L)

Réseau de neurones profond (entièrement connecté)

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} vers $\mathbb{R}^{n_{L+1}}$ de la forme

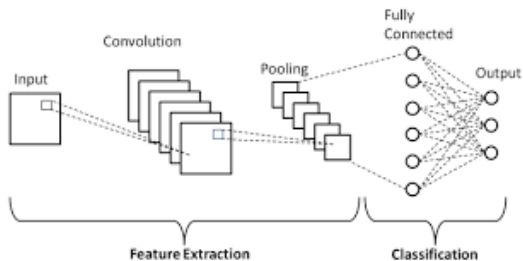
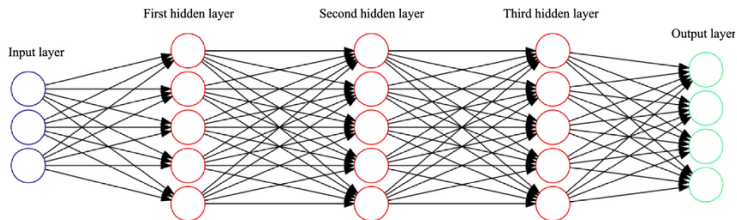
$$W_{L+1} \sigma(W_L \sigma(\dots \sigma(W_2 \sigma(W_1 x + b_1) + b_2) \dots) + b_L) + b_{L+1}$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1), $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)
- $W_2 \in \mathcal{M}_{n_2, n_1}(\mathbb{R})$ (poids 2), $b_2 \in \mathbb{R}^{n_2}$ (biais 2)
- ...
- $W_L \in \mathcal{M}_{n_L, n_{L-1}}(\mathbb{R})$ (poids L), $b_L \in \mathbb{R}^{n_L}$ (biais L)
- $W_{L+1} \in \mathcal{M}_{n_{L+1}, n_L}(\mathbb{R})$ (poids $L + 1$), $b_{L+1} \in \mathbb{R}^{n_{L+1}}$ (biais $L + 1$)

On note $\boxed{\alpha^{L+1}(x)}$ le vecteur final dans $\mathbb{R}^{n_{L+1}}$ (c'est-à-dire $\alpha^{L+1} : \mathbb{R}^{n_0} \rightarrow \mathbb{R}^{n_{L+1}}$) et on l'appelle la *sortie*. N.B. : le réseau est dit **profond** lorsque $L \geq 2$.

Pas la seule architecture, mais présente dans toutes



En pratique (apprentissage supervisé)

- 1 Collecter les données : $\mathcal{D} = \{x_i, f(x_i)\}_{i=1}^n$
- 2 Fixer l'architecture : $L, n_1, \dots, n_L, \sigma$ (\rightarrow hyperparamètres)
 $\rightsquigarrow (W_1, \dots, W_{L+1}, b_1, \dots, b_{L+1}) =: \theta$ (\rightarrow paramètres entraînaibles)
- 3 Initialiser : θ est généralement choisi aléatoirement (gaussien)
- 4 Entraîner le réseau : θ est amélioré pas à pas par descente de gradient

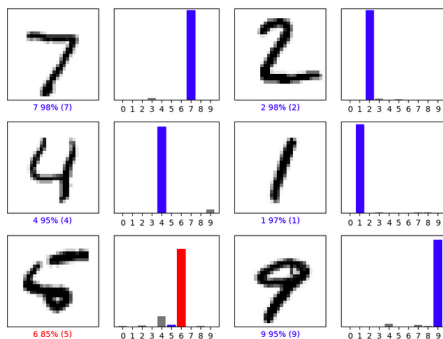
$$\theta(t+1) = \theta(t) - \eta \nabla_{\theta} \mathcal{L}(\theta(t))$$

$$\mathcal{L}(\theta) \stackrel{\text{e.g.}}{=} \frac{1}{n} \sum_{i=1}^n (\alpha^{L+1}(x_i, \theta) - f(x_i))^2$$

- 5 Tester le réseau : prendre un nouveau x et vérifier si $\alpha^{L+1}(x, \theta^*) \approx f(x)$

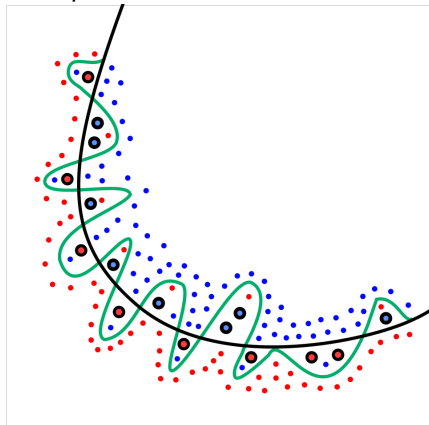
Exemple classique : la base MNIST

- Il s'agit d'une collection de (Python \rightsquigarrow <https://tinyurl.com/4kh7786w>)
 - 60 000 images d'**entraînement**
 - 10 000 images de **test**de taille $28 \times 28 = 784$ pixels.
- Chaque \mathbf{x}_i est un vecteur dans \mathbb{R}^{784} .
- $f(\mathbf{x}_i) = (0, \dots, 0, 1, 0, \dots, 0)$ avec le 1 à la position du chiffre correspondant à \mathbf{x}_i .



Questions (sans réponse)

- Pourquoi $L \rightarrow \infty$ améliore-t-il autant les performances ?
- Pourquoi ReLU surpasse-t-elle de nombreuses autres fonctions d'activation ?
- Pourquoi les réseaux de neurones ne sur-apprennent-ils pas ?



Dans ce qui suit :

- on **fixe** $L \geq 1$ (= nombre de couches cachées) et $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (= fonction d'activation) ;
- on fait tendre n_1, \dots, n_L (= tailles des couches cachées) **vers l'infini** ;
- on part d'une initialisation aléatoire **gaussienne** i.i.d. (Plus précisément : pour tout $1 \leq l \leq L + 1$, on suppose que $(W_l)_{i,j} \sim N(0, \frac{c_w}{n_{l-1}})$ et $(b_l)_i \sim N(0, c_b)$ sont tous indépendants.)

L'objectif est de comprendre le **comportement asymptotique** du **vecteur de sortie** pour un vecteur d'entrée (déterministe) donné.

Théorème central limite

Remarque : ici seulement, on ne suppose pas que les poids et les biais sont nécessairement gaussiens. On suppose qu'ils sont i.i.d. (avec la bonne variance), comme dans le TCL classique.

Théorème (Neal '96, ..., Hanin '23)

Quand $n_1, \dots, n_L \rightarrow \infty$, on a, en loi dans $C^r(\mathbb{R}^{n_0}, \mathbb{R}^{n_{L+1}})$

$$\left(\underbrace{\alpha^{L+1}(\mathbf{x})}_{\in \mathbb{R}^{n_{L+1}}} \right)_{\mathbf{x} \in \mathbb{R}^{n_0}} \longrightarrow \left(G_1^{(L+1)}(\mathbf{x}), \dots, G_{n_{L+1}}^{(L+1)}(\mathbf{x}) \right)_{\mathbf{x} \in \mathbb{R}^{n_0}}$$

Ici, r est lié à la régularité de σ , les $G_i^{(L+1)}$ sont des copies indépendantes et

$$\begin{aligned} \mathbb{E} [G_1^{(L+1)}(\mathbf{x}) G_1^{(L+1)}(\mathbf{y})] &= c_b + c_w \mathbb{E} [\sigma(G_1^{(L)}(\mathbf{x})) \sigma(G_1^{(L)}(\mathbf{y}))] \\ &= \dots \quad (\text{récurrence}). \end{aligned}$$

[Home](#) > [Probability Theory and Related Fields](#) > [Article](#)

Quantitative CLTs in deep neural networks

Published: 04 February 2025

(2025) [Cite this article](#)



[Probability Theory and Related Fields](#)

[Aims and scope](#) →

[Submit manuscript](#) →

[Download PDF](#) ↓

Access provided by [Université de Luxembourg](#)

[S. Favaro](#), [B. Hanin](#) ✉, [D. Marinucci](#), [I. Nourdin](#) & [G. Peccati](#)

44 Accesses [Explore all metrics](#) →

[Use our pre-submission checklist](#) →

Avoid common mistakes on your manuscript.



- TCL **quantitatif** dans le régime $n_1, \dots, n_L \approx n \rightarrow \infty$.
- Plusieurs résultats :
 - 1 Dimension 1 : taux en n^{-1} ;
 - 2 Multidimensionnel : taux en $n^{-1/2}$,
 - 3 Fonctionnel : taux en $n^{-1/8}$.
- Outil principal : **la méthode de Stein**

Rappel: Une version plus restrictive mais améliorée de Stein

Favaro, Hanin, Marinucci, Nourdin, Peccati (PTRF '25)

- Supposons que $F \sim N(0, V)$ conditionnellement à V , pour un certain $V : \Omega \rightarrow \mathbb{R}_+$ d'espérance 1. (Autrement dit : $\mathbb{E}[e^{itF}] = \mathbb{E}[e^{-\frac{1}{2}t^2V}]$ avec $\mathbb{E}[V] = 1$.)
- Alors, pour $N \sim N(0, 1)$ (où $1 = \text{Var}(F) = \mathbb{E}[V]$),

$$d_{TV}(F, N) \leq 8 \text{Var}(V) = \frac{8}{3}(\mathbb{E}[F^4] - 3).$$

- Lorsque $F \sim N(0, V)$ conditionnellement à V pour un certain $V : \Omega \rightarrow \mathbb{R}$, on a encore

$$\mathbb{E}[F\phi(F)] \stackrel{(*)}{=} \mathbb{E}[V\phi'(F)] \quad \text{pour toute } \phi : \mathbb{R} \rightarrow \mathbb{R}.$$

- L'énoncé précédent donnait $d_{TV}(F, N) \leq 2\sqrt{\text{Var}(V)}$ sous (*).

Réseau de neurones : toute fonction de \mathbb{R}^{n_0} vers $\mathbb{R}^{n_{L+1}}$ de la forme

$$W_{L+1} \sigma(W_L \sigma(\dots \sigma(W_2 \sigma(W_1 x + b_1) + b_2) \dots) + b_L) + b_{L+1}$$

avec

- $x \in \mathbb{R}^{n_0}$ (entrée), $W_1 \in \mathcal{M}_{n_1, n_0}(\mathbb{R})$ (poids 1), $b_1 \in \mathbb{R}^{n_1}$ (biais 1),
 $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (fonction d'activation – appliquée composante par composante)
- $W_2 \in \mathcal{M}_{n_2, n_1}(\mathbb{R})$ (poids 2), $b_2 \in \mathbb{R}^{n_2}$ (biais 2)
- ...
- $W_L \in \mathcal{M}_{n_L, n_{L-1}}(\mathbb{R})$ (poids L), $b_L \in \mathbb{R}^{n_L}$ (biais L)
- $W_{L+1} \in \mathcal{M}_{n_{L+1}, n_L}(\mathbb{R})$ (poids $L + 1$), $b_{L+1} \in \mathbb{R}^{n_{L+1}}$ (biais $L + 1$)

On note $\boxed{\alpha^{L+1}(x)}$ le vecteur final dans $\mathbb{R}^{n_{L+1}}$ (c'est-à-dire $\alpha^{L+1} : \mathbb{R}^{n_0} \rightarrow \mathbb{R}^{n_{L+1}}$) et on l'appelle la *sortie*. N.B. : le réseau est dit **profond** lorsque $L \geq 2$.

- Soit $F = \alpha_1^{L+1}(x) (\in \mathbb{R})$ avec $L \geq 1$ et $x \in \mathbb{R}^{n_0}$ fixés.
- On a $\alpha_1^{L+1}(x) = \sum_{j=1}^{n_L} (W_{L+1})_{1j} \alpha_j^L(x) + (b_{L+1})_1$, de sorte que $F \sim N(0, V)$ avec $V = c_b + \frac{c_w}{n_L} \|\alpha^L(x)\|^2$.
- On note $\bar{F} = F / \sqrt{\text{Var}(F)}$ la version normalisée de F .
- Par la version améliorée de Stein, on déduit que $d_{TV}(\bar{F}, N) \leq \frac{8}{3} \kappa_4(\bar{F})$.
- Hanin (AAP'23) a calculé le comportement asymptotique de tous les cumulants de F : $\kappa_m(F) \approx n^{1-\frac{m}{2}}$ (à une constante multiplicative près). En particulier, $\text{Var}(F) \approx 1$ et $\kappa_4(F) \approx \frac{1}{n}$.
- On en déduit $d_{TV}(\bar{F}, N) = O\left(\frac{1}{n}\right)$.

Borne optimale

- Soit $F = \alpha_1^{L+1}(x) (\in \mathbb{R})$ avec $L \geq 1$ et $x \in \mathbb{R}^{n_0}$ fixés.
- On a $F \sim N(0, V)$, donc $\bar{F} = F/\sqrt{\text{Var}(F)}$ vérifie $\bar{F} \sim N(0, \frac{V}{\text{Var}(F)})$.
- On calcule : $\mathbb{E}[\cos \bar{F}] - \mathbb{E}[\cos N] = \mathbb{E}[e^{-\frac{V}{2\text{Var}(F)}}] - e^{-\frac{1}{2}}$.
- Par Taylor, on déduit $\mathbb{E}[\cos \bar{F}] - \mathbb{E}[\cos N]$

$$\begin{aligned} &= \frac{e^{-\frac{1}{2}}}{8} \mathbb{E} \left[\left(\frac{V}{\text{Var}(F)} - 1 \right)^2 \right] - \frac{e^{-\frac{1}{2}}}{48} \mathbb{E} \left[\left(\frac{V}{\text{Var}(F)} - 1 \right)^3 \right] \\ &\quad + O \left\{ \mathbb{E} \left[\left(\frac{V}{\text{Var}(F)} - 1 \right)^4 \right] \right\} \\ &= \frac{e^{-\frac{1}{2}}}{8} \frac{\text{Var}(V)}{\text{Var}(F)^2} - \frac{e^{-\frac{1}{2}}}{48} \frac{\kappa_3(V)}{\text{Var}(F)^3} + O \left\{ \frac{\kappa_4(V) + 3\kappa_2(V)^2}{\text{Var}(F)^4} \right\} \\ &= \frac{e^{-\frac{1}{2}} \kappa_4(F)}{24 \text{Var}(F)^2} + O \left(\max \left\{ \left(\frac{\kappa_4(F)}{\text{Var}(F)^2} \right)^2, \frac{|\kappa_6(F)|}{\text{Var}(F)^3}, \frac{|\kappa_8(F)|}{\text{Var}(F)^4} \right\} \right) = \frac{\text{cst.}}{n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Partie 3

Les LLMs comme assistants de recherche en mathématiques

Mon parcours personnel avec l'IA générative

- **Découverte de l'IA générative** : décembre 2022
- **Fondation de GrewIA** : février 2023

Comme tout le monde, j'ai commencé par expérimenter. Très vite, j'ai réalisé que c'était quelque chose de **révolutionnaire**.

Les premiers bénéfices : **coder plus facilement** (Python, HTML), **aide aux textes** (traduction, reformulation).

Mais pour les mathématiques ?



241 - (-241) + 1



241 - (-241) + 1 is equivalent to $241 + 241 + 1$, which simplifies to $483 + 1$. So 241 - (-241) + 1 is equal to 484.





1. A simple lemma

Let $f : [0, T] \rightarrow \mathbb{R}$ be nondecreasing and suppose

$$f(t) > f(0) \quad \text{for all } t > 0. \quad (*)$$

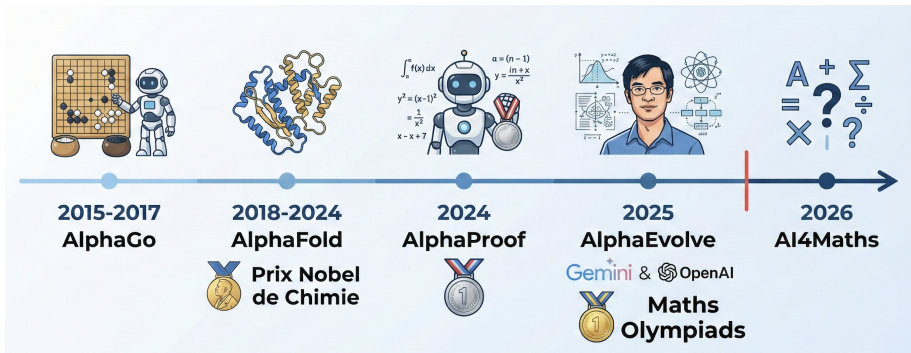
Then f is *strictly* increasing on $[0, T]$:

- Take $0 \leq t_1 < t_2$.
- By nondecreasingness, $f(t_2) \geq f(t_1) \geq f(0)$.
- If $f(t_2) = f(t_1)$, then $f(t_1) = f(t_2) \geq f(0)$.
But $(*)$ says $f(t_1) > f(0)$ since $t_1 > 0$,
contradiction.
- Hence $f(t_2) > f(t_1)$.

So to get strict increase you *don't* need a direct inequality between every pair $t_1 < t_2$; it is enough to know:

1. nondecreasing on \mathcal{J}^+ , and
2. strictly larger than its value at 0 for all $t > 0$.





Ligne rouge verticale dans la frise chronologique : **Bubeck, août 2025** (j'expliquerai cela plus tard).

Même aujourd'hui, les LLMs les plus perfectionnés peuvent faire des **erreurs mathématiques** sur des problèmes très simples, tout en étant **très bon** pour résoudre des problèmes très complexes.

Prompt : Generate a cute educational poster in pastel colors for children, showing a multiplication table



FrontierMath est un benchmark développé par EpochAI pour évaluer la capacité des systèmes d'IA à résoudre des problèmes mathématiques complexes.

Combinatorics in the Tsirelson space

Tier 3

[Problem](#) [Solution](#)

Background. The Tsirelson space is a Banach space introduced in 1974 as a counterexample to a longstanding conjecture in Banach space theory. It is a subspace of ℓ^∞ described as follows.

Let P_k denote the operator on ℓ^∞ that modifies each $(x_n : n \in \mathbb{N}) \in \ell^\infty$ by setting the first k coordinates equal to 0 (without changing any other coordinates). Given $\vec{x} = (x_n : n \in \mathbb{N}) \in \ell^\infty$, the support of \vec{x} is $\text{supp}(\vec{x}) = \{n \in \mathbb{N} : x_n \neq 0\}$. The Tsirelson space is the span of the smallest pointwise-closed subset K of the unit ball in ℓ^∞ satisfying the following two properties:

- For every $j \in \mathbb{N}$, the unit vector \vec{e}_j and all its multiples $\lambda \vec{e}_j$ with $|\lambda| \leq 1$ belong to K .
- Suppose $\vec{x}_1, \dots, \vec{x}_k$ is a finite sequence of members of K , each with finite support, such that $\max(\text{supp}(\vec{x}_i)) < \min(\text{supp}(\vec{x}_j))$ for all $i < j \leq k$. Then $\frac{1}{2}P_k(\vec{x}_1 + \dots + \vec{x}_k) \in K$.

Problem.

Let N denote the largest natural number such that there is a member $(x_n : n \in \mathbb{N})$ of the set K described above with $x_3 = 1/2$, $x_N > 1/20$, and $x_i > x_j$ whenever $3 \leq i < j \leq N$. It turns out that N is too large to write down in a reasonable amount of space, so let us ask instead: what is the largest power of 2 dividing N ? In other words, find the integer $p \geq 0$ such that $N = m2^p$ for some odd number m .

Combinatorics in the Tsirelson space

Tier 3

[Problem](#) [Solution](#)

Answer: 402653211.

Solution: Let us begin by introducing some helpful shorthand. In the first half of this explanation, we will deal with quite a few sequences of the form $(0, 0, \dots, 0, 0, b, b - \varepsilon, b - 2\varepsilon, \dots, a + 2\varepsilon, a + \varepsilon, a, 0, 0, \dots)$ where $0 < a < b < 1$. For convenience, let $(a, b, (m, n))$ denote the sequence (x_1, x_2, x_3, \dots) such that $x_i = 0$ if $i \leq m$ or $i > n$, and $x_i = a + (n - i) \frac{b - a}{n - m - 1}$ if $i \in (m, n]$.

Observation 1. Let $\frac{1}{4} < a < b \leq \frac{1}{2}$. For any $m \in \mathbb{N}$, $(a, b, (m, 2m)) \in K$.

Proof. For each $i \in (m, 2m]$ and each $c \in [a, b]$, we have $2c\vec{e}_i \in K$ by rule 1 in the definition of K . Therefore

$$(a, b, (m, 2m)) = \frac{1}{2}P_m(2b\vec{e}_{m+1} + \dots + 2a\vec{e}_{2m}) \in K$$

by rule 2.

Observation 2. Let $\frac{1}{8} < a < b \leq \frac{1}{4}$. For any $m \in \mathbb{N}$, $(a, b, (m, m2^m)) \in K$.

Proof. Fix $m \in \mathbb{N}$. By the previous observation, we have $(2a_1, 2b_1, (m, 2m)), (2a_2, 2b_2, (2m, 4m)), \dots, (2a_m, 2b_m, (m2^{m-1}, m2^m)) \in K$ for any numbers

FrontierMath est un benchmark développé par EpochAI pour évaluer la capacité des systèmes d'IA à résoudre des problèmes mathématiques complexes.

An optimization problem in BMO space

Tier 4

[Problem](#) [Solution](#)

Let

$$c = \sup_f \int_0^1 (f(t)^3 + |f(t)|) dt$$

where supremum is taken over all integrable functions $f : [0, 1] \rightarrow \mathbb{R}$ such that $\int_0^1 f(t) dt = -10$, $\int_0^1 f^2(t) dt = 100 + \frac{1}{12}$, and

$$\sup_{J \subset [0, 1]} \frac{1}{|J|} \int_J |f(t) - \frac{1}{|J|} \int_J f(s) ds|^2 dt \leq \frac{1}{12},$$

where supremum is taken over all subintervals $J \subset [0, 1]$. Here $|J|$ denotes the length of the interval.

Find $c + \frac{1985}{2}$.

An optimization problem in BMO space

Tier 4

[Problem](#) [Solution](#)

Answer: $\frac{\sqrt{3}}{36} + \frac{\sqrt{3}}{6} e^{-20\sqrt{3}-\frac{1}{3}}$

Let us consider the following optimization problem

$$B(x_1, x_2; \varepsilon) \stackrel{\text{def}}{=} \sup_{f \in L^1([0, 1])} \left\{ \int_0^1 F(f(t)) dt : \int_0^1 f(t) dt = x_1, \int_0^1 f^2(t) dt = x_2, \sup_{J \subset [0, 1]} \frac{1}{|J|} \int_J \left| f(t) - \frac{1}{|J|} \int_J f(s) ds \right|^2 dt \leq \varepsilon \right\}$$

where $F(t) = t^3 + |t|$. Then clearly $B(-10, 100 + \frac{1}{12}; \frac{1}{12}) = c$. It is not difficult to see that B is defined in the following infinity parabolic strip

$$\Omega_\varepsilon = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 \leq x_2 \leq x_1^2 + \varepsilon^2\},$$

with the boundary condition $B(x_1, x_2^*) = F(x_1)$. Our goal will be to find $B(x_1, x_2; \varepsilon)$ for all $(x_1, x_2) \in \Omega_\varepsilon$. I will say that to the best of my knowledge the difficulty of recovering the single value $B(-10, 100 + \frac{1}{12}; \frac{1}{12})$ is the same as recovering the values of $B(x_1, x_2; \varepsilon)$ at all points in Ω_ε .

Où en sommes-nous ?

FrontierMath est un benchmark développé par EpochAI pour évaluer la capacité des systèmes d'IA à résoudre des problèmes mathématiques complexes.

FrontierMath: Open Problems

A collection of unsolved mathematics problems that have resisted serious attempts by professional mathematicians. AI solutions would meaningfully advance the state of human mathematical knowledge. This work was supported by Schmidt Sciences.

About Open Problems

Submit a problem

Interested in purchasing access to solution verifiers? Contact ai@epoch.ai

Problems solved by AI

Problem	Notability	Solved
MODERATELY INTERESTING 0/2 SOLVED		
SOLID RESULT 0/0 SOLVED		
MAJOR ADVANCE 0/0 SOLVED		
BREAKTHROUGH 0/0 SOLVED		
Problem 11		
Ramsey Numbers for Book Graphs Prove a tight lower bound on Ramsey numbers for a class of off-diagonal book graphs.	Moderately interesting	No
A Ramsey-style Problem on Hypergraphs Construct hypergraphs as large as possible that do not have a certain easy-to-check, difficult-to-find property.	Moderately interesting	No

Ramsey Numbers for Book Graphs

Prove a tight lower bound on Ramsey numbers for a class of off-diagonal book graphs.

Notability: Moderately interesting

Solved: No

Field: Combinatorics

Contributor: William J. Wesley

SEW Visiting Assistant Professor, University of California, San Diego

On this page:

[About the problem](#) • [Attempts by AI](#) • [AI prompts](#) • [Mathematician survey](#)

About the problem

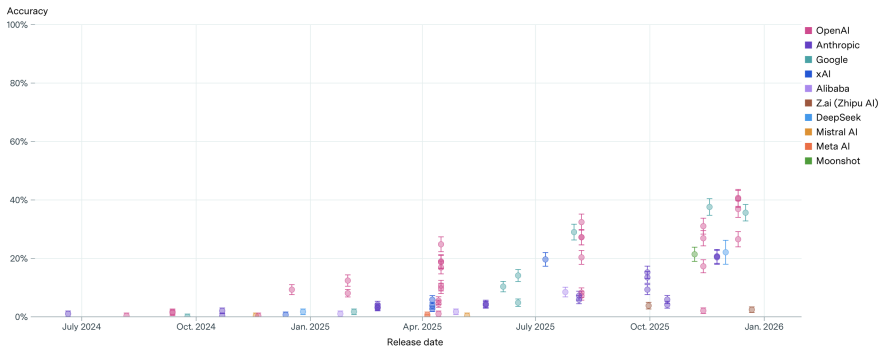
This problem aims to determine the Ramsey numbers $R(B_{n_1}, \dots, B_{n_r})$, where B_{n_i} denotes a triangular [book graph](#) on n_i vertices. The upper bound $R(B_{n_1}, \dots, B_{n_r}) \leq 4n - 1$ was established in 1978, and was more recently shown to be sharp for an infinite family of n as well as for all $n \geq \dots$

Given these results, it seems likely that the upper bound is sharp for all n . The goal of this problem is to prove this, by finding an algorithm that, given n , produces a graph witnessing

Où en sommes-nous ?

FrontierMath est un benchmark développé par EpochAI pour évaluer la capacité des systèmes d'IA à résoudre des problèmes mathématiques complexes.

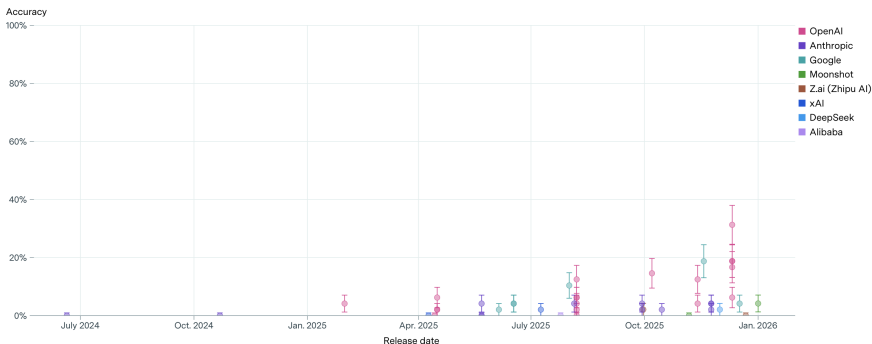
AI Model Performance on FrontierMath



Où en sommes-nous ?

FrontierMath est un benchmark développé par EpochAI pour évaluer la capacité des systèmes d'IA à résoudre des problèmes mathématiques complexes.

AI Model Performance on FrontierMath



Mathematical research with GPT-5: a Malliavin-Stein experiment

Charles-Philippe Diez* Luís da Maia* Ivan Nourdin*

Abstract

On August 20, 2025, GPT-5 was reported to have solved an open problem in convex optimization. Motivated by this episode, we conducted a controlled experiment in the Malliavin-Stein framework for central limit theorems. Our objective was to assess whether GPT-5 could go beyond known results by extending a *qualitative* fourth-moment theorem to a *quantitative* formulation with explicit convergence rates, both in the Gaussian and in the Poisson settings. To the best of our knowledge, the derivation of such quantitative rates had remained an open problem, in the sense that it had never been addressed in the existing literature. The present paper documents this experiment, presents the results obtained, and discusses their broader implications.

Keywords: Gaussian analysis, Poisson approximation, Malliavin calculus, Probability theory, Stochastic processes, Artificial intelligence in research

AMS 2010 Subject Classification: 60G15, 60G55, 60H07, 60F05, 68T50



Greg Brockman @gdb · Sep 7

gpt-5 pro for novel mathematics — in partnership with a math professor:



Ethan Mollick @emollick · Sep 6

We are starting to see some nuanced discussions of what it means to work with advanced AI

In this case, GPT-5 Pro was able to do novel math, but only when guided by a math professor (though the paper also noted the speed of ...

Mathematical research with GPT-5:
Malliavin-Stein experiment

Charles-Philippe Diez* Luís da Maia* Ivan Nourdin*

with probably now also lost the ability to combine ideas coming from different sources. At that point, the result appears useful for an individual (like, looking at new ideas). In practice, however, it may also be significant to look for possibly really interesting problems to be solved and eventually guide us on their to reach correct to solutions.

The main idea was to use with this technology, in the current state, to find a well-defined and already had to a proliferation of mathematical results produced with AI. This could represent the already technology with technical content that will be developed, eventually, eventually, leading to higher level of research work to be done on. The situation is reminiscent of other cultural domains already transformed by more generative technologies, a field of research, especially, but not only, that often creates and raises the bar level.

Greg Brockman

Article [Talk](#)

From Wikipedia, the free encyclopedia

Gregory Brockman (born November 29, 1987) is an American [entrepreneur](#), [investor](#) and [software engineer](#) who is a co-founder^[2] and currently the president of [OpenAI](#).^[3] He

Un mois après notre prépublication (oct. 2025)



Paata Ivanisvili @PI010101 · Oct 5



GPT-5 Pro found a counterexample to the NICD-with-erasures majority optimality (Simons list, p.25).

simons.berkeley.edu/sites/default/...

At $p=0.4$, $n=5$, $f(x) = \text{sign}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$ gives $\mathbb{E}|f(x)| = 0.43024$ vs best majority 0.42904 .

active Correlation Distillation with Erasures
 $\rightarrow \{-1, 1\}$ be an unbiased function. Let $\mathbf{z} \sim$
ordinate z_i is (independently) ± 1 with prob
 $p < 1/2$ and n odd, is it true that $\mathbb{E} z_i$
Here we identify f with its multilinear expansion

ected that $\mathbb{E} z_i |f(\mathbf{z})|$ is maximized when
onnell and Wright [OW12].

at if f 's influences are assumed at most τ then
 $\text{Maj}_n(\mathbf{z}) + o_\tau(1)$.

25

crete counterexample (explicit function)
 $n=5$ and $p=0.40$. Define
 $f(x_1, x_2, x_3, x_4, x_5) = \text{sign}(x_1 - 3x_2 + x_3 - x_4 + 3x_5)$
out this f :
a linear threshold function (LTF) with weight vector $w = (1, -3, 1, -1, 3)$ and zero threshold
the threshold is 0 and no assignment yields $w \cdot x = 0$, $f(-x) = \text{sign}(-w \cdot x) = -$
odd and hence unbiased.
not any majority Maj_k (for $k = 1, 3, 5$), in fact its Hamming distances from $\text{Maj}_1, \text{Maj}_3,$
 $(-1, 1)^5$ are 12, 16, 14, respectively.
ues at $p = 0.40$:
 $\mathbb{E}|f| = 0.43024$.
rging the three majority candidates,
 $\Phi_{0.40}(\text{Maj}_1) = 0.40000$, $\Phi_{0.40}(\text{Maj}_3) = 0.42400$, $\Phi_{0.40}(\text{Maj}_5) = 0.42904$
 $\Phi_{0.40}(f) = \max\{\Phi_{0.40}(\text{Maj}_1), \Phi_{0.40}(\text{Maj}_3), \Phi_{0.40}(\text{Maj}_5)\} = 0.00120 > 0$,
ctly outperforms every majority at $p = 0.40$.
withness, the same f also beats Maj_3 at $p = 0.30$ and $p = 0.40$: $\Phi_{0.30}(f) = 0.3541$
 $\Phi_{0.30}(\text{Maj}_3) = 0.35365$ and $\Phi_{0.40}(f) = 0.4031239999 > \Phi_{0.40}(\text{Maj}_5) = 0.40296812477$
is a valid counterexample

56

277

1.5K

719K



<https://simons.berkeley.edu/sites/default/files/openprobsmerged.pdf>

Un mois après notre prépublication (oct. 2025)



Quel sera l'impact de l'IA sur les mathématiques dans les prochaines années ? - Timothy Gowers



Collège de France
133 k abonnés

S'abonner



167



Partager



Enregistrer



9,4 k vues il y a 1 mois

Colloque de rentrée 2025 : Formes de l'intelligence : IA, connaissance, déduction, apprentissage

Conférence du 16 octobre 2025 : Quel sera l'impact de l'IA sur les mathématiques dans les prochaines années ?

Un mois après notre prépublication (oct. 2025)



Terence Tao

@tao@mathstodon.xyz

I experimented with using a more powerful AI (ChatGPT5 Pro) to attempt an open problem mathoverflow.net/questions/425... which is not quite in my area of expertise (specifically, differential geometry): chatgpt.com/share/68e85cba-722.... The outcome was surprisingly nuanced, and illustrates a point I made in an earlier post mathstodon.xyz/deck/@tao/11450... that one has to measure the usefulness of a tool at multiple scales. In summary: AI assistance was very useful at small scales, somewhat unhelpful at medium scales, but somewhat helpful again at large scales; the problem remains unsolved, but I understand it much better now.

Mais reprenons depuis le début



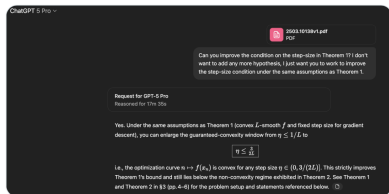
Sebastien Bubeck
@SebastienBubeck



Claim: gpt-5-pro can prove new interesting mathematics.

Proof: I took a convex optimization paper with a clean open problem in it and asked gpt-5-pro to work on it. It proved a better bound than what is in the paper, and I checked the proof it's correct.

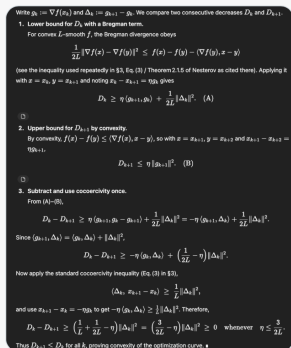
Details below.



6:05 PM · Aug 20, 2025 · 6.7M Views



Sebastien Bubeck @SebastienBubeck · Aug 20
By the way this is the proof it came up with:





2503.10138v1.pdf
PDF


Can you improve the condition on the step-size in Theorem 1? I don't want to add any more hypothesis, I just want you to work to improve the step-size condition under the same assumptions as Theorem 1.

Request for GPT-5 Pro

Reasoned for 17m 35s

Yes. Under the *same* assumptions as Theorem 1 (convex L -smooth f and fixed step size for gradient descent), you can enlarge the guaranteed-convexity window from $\eta \leq 1/L$ to

$$\eta \leq \frac{3}{2L}$$

i.e., the optimization curve $n \mapsto f(x_n)$ is convex for any step size $\eta \in (0, 3/(2L)]$. This strictly improves Theorem 1's bound and still lies below the non-convexity regime exhibited in Theorem 2. See Theorem 1 and Theorem 2 in §3 (pp. 4–6) for the problem setup and statements referenced below. 

Sébastien Bubeck

 [Add languages](#) ▼

[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#) ▼

From Wikipedia, the free encyclopedia

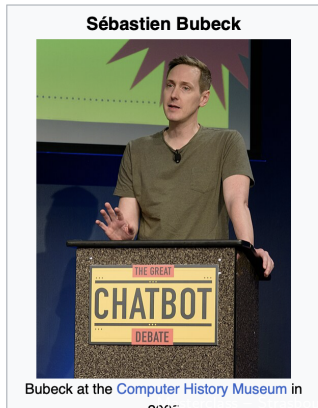


This biographical article **is written like a résumé**. Please [help improve it](#) by revising it to be [neutral](#) and [encyclopedic](#). *(August 2025)*

Sébastien Bubeck (born April 16, 1985)^[1] is a French-American [computer scientist](#). He was [Microsoft's](#) Vice President of Applied Research, Distinguished Scientist, and led the Machine Learning Foundations group at [Microsoft Research Redmond](#). Bubeck was formerly professor at [Princeton University](#) and a researcher at the [University of California, Berkeley](#).^[2] He is known for his contributions to [online learning](#), [optimization](#) and more recently studying [deep neural networks](#), and in particular [transformer models](#). Since 2024, he works for [OpenAI](#).

Work [[edit](#)]

Bubeck's work spans a wide variety of topics in [machine learning](#), [theoretical computer science](#) and [artificial](#)



Bubeck at the [Computer History Museum](#) in 2025

Gaussian fluctuation for Gaussian Wishart matrices of overall correlation ☆

Ivan Nourdin , Fei Pu  

Show more 

 Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.spl.2021.109269>

[Get rights and content](#) 

Abstract

In this note, we study the Gaussian fluctuations for the Wishart matrices $d^{-1} \mathcal{X}_{n,d} \mathcal{X}_{n,d}^T$, where $\mathcal{X}_{n,d}$ is a $n \times d$ random matrix whose entries are jointly Gaussian and correlated with row and column covariance functions given by r and s respectively such that $r(0) = s(0) = 1$. Under the assumptions $s \in \ell^{4/3}(\mathbb{Z})$ and $\|r\|_{\ell^1(\mathbb{Z})} < \sqrt{6}/2$, we establish the $\sqrt{n^3/d}$ convergence rate for the Wasserstein distance between a normalization of $d^{-1} \mathcal{X}_{n,d} \mathcal{X}_{n,d}^T$ and the corresponding Gaussian ensemble. This rate is the same as the optimal one computed in Bubeck et al. (2016), Bubeck and Ganguly (2018) and Jiang and Li (2015) for the total variation distance, in the particular case where the Gaussian entries of $\mathcal{X}_{n,d}$ are independent. Similarly, we obtain the $\sqrt{n^{2p-1}/d}$ convergence rate for the Wasserstein distance in the setting of symmetric

Une vue plus nuancée



Ernest Ryu @ErnestRyu

Aug 21

This is really exciting and impressive, and this stuff is in my area of mathematics research (convex optimization).

I have a nuanced take. 🗨️ (1/9)



Sebastien Bubeck @SebastienBubeck

Aug 20

Claim: gpt-5-pro can prove new interesting mathematics.

Proof: I took a convex optimization paper with a clean open problem in it and asked gpt-5-pro to work on it. It proved a better bound than what is in the paper, and I checked the proof it's correct.

Details below.

ChatGPT 5 Pro -

2503.10138v1.pdf
PDF

Can you improve the condition on the step-size in Theorem 1? I don't want to add any more hypothesis, I just want you to work to improve the step-size condition under the same assumptions as Theorem 1.

Request for GPT-5 Pro
Reasoned for 17m 35s

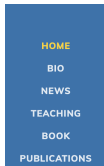
Yes. Under the same assumptions as Theorem 1 (convex L -smooth f and fixed step size for gradient descent), you can enlarge the guaranteed-convexity window from $\eta \leq 1/L$ to

$$\eta \leq \frac{1}{2L}$$

i.e., the optimization curve $n \mapsto f(x_n)$ is convex for any step size $\eta \in (0, 3/(2L)]$. This strictly improves Theorem 1's bound and still lies below the non-convexity regime exhibited in Theorem 2. See Theorem 1 and Theorem 2 in §3 (pp. 4-6) for the problem setup and statements referenced below. ☐

Aug 21, 2025 · 6:00 AM UTC

🗨️ 38 🔄 310 🗨️ 52 ❤️ 3,183



Ernest K. Ryu

Assistant Professor
Department of Mathematics
University of California, Los Angeles

B = mouvement brownien

Intégrale de Wiener

$$I(f) := \int_{\mathbb{R}} f(u) dB_u, \quad f \in L^2(\mathbb{R})$$

B = mouvement brownien

Intégrale de Wiener-Itô

$$I_d(f) := \int_{\mathbb{R}^d} f(u_1, \dots, u_d) dB_{u_1} \dots dB_{u_d}, \quad f \in L^2_S(\mathbb{R}^d), \quad d \geq 1$$

Théorème du quatrième moment

Théorème du quatrième moment (Nualart, Peccati, 2005)

Cadre : Soient $d \geq 2$ et $F_n = I_d(f_n)$ avec $(f_n) \subset L^2_s(\mathbb{R}^d)$.

Hypothèse : Supposons $\mathbb{E}[F_n^2] \rightarrow \sigma^2$ lorsque $n \rightarrow \infty$.

Conclusion : Alors

$$F_n \xrightarrow{\text{law}} N(0, \sigma^2) \iff \mathbb{E}[F_n^4] \rightarrow 3\sigma^4.$$

Malliavin-Stein (Nourdin, Peccati, 2007)

Soit $F : \Omega \rightarrow \mathbb{R}$ dans $\mathbb{D}^{1,2}$ avec $\mathbb{E}[F] = 0$ et $\text{Var}(F) = 1$.

Alors

$$\sup_{A \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(F \in A) - \mathbb{P}(N(0,1) \in A)| \leq 2 \mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle|.$$

Version quantitative du TMQ

Soit $F = I_d(f)$ avec $f \in L^2_5(\mathbb{R}^d)$ et $\text{Var}(F) = 1$.

Alors

$$\sup_{A \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(F \in A) - \mathbb{P}(N(0,1) \in A)| \leq \frac{2}{\sqrt{3}} (\mathbb{E}[F^4] - 3).$$

Page web Malliavin-Stein (500+ références)

The screenshot shows a web browser window with the address bar displaying "sites.google.com/site/malliavinstein/home". The page title is "malliavinstein". The main content is a list of 42 references, each starting with a number and a year in parentheses, followed by a link to a paper. The references cover various topics in probability theory and stochastic processes, including Kolmogorov bounds, monochromatic subgraphs, Gaussian fluctuations, central limit theorems for stochastic wave equations, extensions of the Stein-Malliavin-Stein method, exponential dimensional dependence, quantitative bounds for Gaussian processes, fourth-moment theorems, caveats on metrizing convergence, Gaussian approximation on Skorohod space, local asymptotic normality, irreducible central limit theorems, Berry-Esséen bounds, central limit theorems for squared increment sums, quartic variation of solutions, central limit theorems for non-linear functionals, density convergence on Markov diffusion chaos, mathematical research with GPT-5, extension of a theorem of Wschebor, normal approximation of stabilizing Poisson pair functionals, asymptotics of Yule's nonsense correlation, invariant measures for the open KPZ equation, and the fourth-moment theorem on Hilbert spaces.

malliavinstein

20. M. Al-Foraih and K. Es-Sebאי (2025): [An improved Kolmogorov bound for approximate maximum likelihood estimators for the alpha-Brownian bridge](#)
21. M. D. Andrade and B. Bhattacharya (2025): [Monochromatic Subgraphs in Randomly Colored Dense Multiplex Networks](#)
22. R. Balan and M. Salins (2025): [Gaussian fluctuations for the nonlinear stochastic heat equation with drift](#)
23. R. Balan and G. Zheng (2025): [Central limit theorem for stochastic nonlinear wave equation with pure-jump Lévy white noise](#)
24. A. Balasev-Samarski and A.-R. Mansouri (2025): [An extension of the Stein/Malliavin-Stein method to independent random variables](#)
25. A. Basse-O'Connor and D. Kramer-Bang (2025): [Exponential dimensional dependence in high-dimensional Hermite method of moments](#)
26. A. Basse-O'Connor and D. Kramer-Bang (2025): [Quantitative bounds for high-dimensional non-linear functionals of Gaussian processes](#)
27. A. Basse-O'Connor, D. Kramer-Bang and C. Svendsen (2025): [Fourth-Moment Theorems for Sums of Multiple Integrals](#)
28. F. Bassetti, S. Bourguin, S. Campese and G. Peccati (2025): [A Caveat on Metrizing Convergence in Distribution on Hilbert Spaces](#)
29. S. Bourguin and S. Campese (2025): [Gaussian approximation on the Skorohod space via Malliavin calculus and regularization](#)
30. C. Cai (2025): [Local Asymptotic Normality for Mixed Fractional Brownian Motion with \$0 < H < 3/4\$](#)
31. F. Caravenna, F. Cottini and G. Peccati (2025): [On irreducible central limit theorems](#)
32. Y. Chen, Y. Li and H. Zhou (2025): [Berry-Esséen bounds for the statistical estimators of an Ornstein-Uhlenbeck process driven by a general Gaussian noise](#)
33. N. Chenavier and C. Y. Robert (2025): [Central limit theorems for squared increment sums of fractional Brownian fields based on a Delaunay triangulation in 2D](#)
34. I. Cimpean, Y. Nachit and C.A. Tudor (2025): [Quartic variation of the solution to the semilinear stochastic heat equation: limit behavior and asymptotic independence with respect to the data](#)
35. F. Coppini and W.M. Ruszel (2025): [Central limit theorems for non-linear functionals of Gaussian fields via Wiener chaos decomposition](#)
36. T. Dang and Y. Hu (2025): [Density convergence on Markov diffusion chaos via Stein's method](#)
37. C.-P. Diez, L. da Maia and I. Nourdin (2025): [Mathematical research with GPT-5: a Malliavin-Stein experiment](#)
38. C. Donati-Martin and A. Rouault (2025): [Extension of a theorem of Wschebor to free and matrix Brownian motions](#)
39. H. Döring, A. Garin, C. Hirsch and N.N. Lundbye (2025): [Normal approximation of stabilizing Poisson pair functionals with column-type dependence](#)
40. S. Douissi, P. Ernst and F. Viens (2025): [Asymptotics of Yule's nonsense correlation for Ornstein-Uhlenbeck paths: The correlated case](#)
41. A. Dunlap, Y. Gu and T. Rosati (2025): [Invariant measures for the open KPZ equation: an analytic perspective](#)
42. M.-C. Dülker and P. Zrubavelidze (2025): [The Fourth-Moment Theorem on Hilbert Spaces](#)

FOURTH-MOMENT THEOREMS FOR SUMS OF MULTIPLE INTEGRALS

ANDREAS BASSE-O'CONNOR*, DAVID KRAMER-BANG† & CLEMENT SVENDSEN‡

ABSTRACT. Nualart & Peccati ([Nualart and Peccati, 2005, Thm 1]) established the first *fourth-moment theorem* for random variables in a fixed Wiener chaos, i.e. they showed that convergence of the sequence of fourth moments to the fourth moment of the standard Gaussian distribution is sufficient for weak convergence to the standard Gaussian. In this paper, we provide what we believe to be the first generalization to chaos expansions with more than a single term. Specifically, we show that a fourth-moment theorem holds for random variables consisting of sums of two multiple integrals of orders $p, q \in \mathbb{N}$, where p, q have different parities. Furthermore, we show that such random variables cannot themselves be Gaussian, again generalizing what is known for the fixed Wiener chaos setting. Finally, we show a fourth-moment theorem for variables with infinite Wiener chaos expansions when the terms in the expansions are independent and satisfy an additional regularity condition in terms of the Ornstein-Uhlenbeck operator.

Theorem 1.1. *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables with chaos decompositions $X_n = Y_n + Z_n$ for all $n \in \mathbb{N}$, where $(Y_n)_{n \in \mathbb{N}}$ and $(Z_n)_{n \in \mathbb{N}}$ are sequences from \mathbb{H}_p and \mathbb{H}_q respectively, and $p \in \mathbb{N}$ is odd and $q \in \mathbb{N}$ is even. Assume that $\mathbb{E}[X_n^2] = 1$ for all $n \in \mathbb{N}$. Then $\lim_{n \rightarrow \infty} \mathbb{E}[X_n^4] = 3$ if and only if $X_n \xrightarrow{d} \mathcal{N}(0, 1)$ as $n \rightarrow \infty$.*

Le candidat parfait !

Proof of Theorem 1.1. Due to the correspondence between multiple integrals and Wiener chaoses, there exist vectors $u_n \in \mathcal{H}^{\odot p}, v_n \in \mathcal{H}^{\odot q}$ for each $n \in \mathbb{N}$ such that $X_n = I_p(u_n) + I_q(v_n)$. Assume first that $\mathbb{E}[X_n^4] \rightarrow 3$ as $n \rightarrow \infty$. Since the sequence of fourth moments is bounded, we know that $(X_n)_{n \in \mathbb{N}}$ is tight, and hence by Prokhorov's theorem (see [Billingsley, 1999, Thm 3.1]) it suffices to show that if $(n_k)_{k \in \mathbb{N}}$ is a subsequence such that $(X_{n_k})_{k \in \mathbb{N}}$ has a limit in distribution then this limit is $\mathcal{N}(0, 1)$. To ease notation, we assume without loss of generality that $(X_n)_{n \in \mathbb{N}}$ itself has a limit μ in distribution. First, we show

$$(3.1) \quad \kappa_4(X_n) = \kappa_4(I_p(u_n)) + \kappa_4(I_q(v_n)) + 6\text{Cov}(I_p(u_n)^2, I_q(v_n)^2), \quad \text{for all } n \in \mathbb{N}.$$

By the binomial theorem, it follows for all $n \in \mathbb{N}$,

$$(3.2) \quad \mathbb{E}[X_n^4] = \sum_{k=0}^4 \binom{4}{k} \mathbb{E}[I_p(u_n)^k I_q(v_n)^{4-k}].$$

For $k = 1, 3$, the corresponding mixed moments in (3.2) vanish. Indeed, using the product formula (2.1) three times directly yields

$$(3.3) \quad I_p(u_n)^3 I_q(v_n) = \sum_{r_1=0}^p \sum_{r_2=0}^{(2p-2r_1) \wedge p} \sum_{r_3=0}^{(3p-2(r_1+r_2)) \wedge q} c(r_1, r_2, r_3, p, q) I_{3p+q-2(r_1+r_2+r_3)}(h(r_1, r_2, r_3, u_n, v_n))$$

for some constants $c(r_1, r_2, r_3, p, q) \in \mathbb{R}$ and vectors $h(r_1, r_2, r_3, u, v) \in \mathcal{H}^{\odot(3p+q-2(r_1+r_2+r_3))}$. The constants and vectors can be given explicitly, but as they are unimportant for this proof, we choose not to state them to simplify the expression. Recall now that multiple integrals of order $m \in \mathbb{N}_0$ are centered unless $m = 0$. Since p is odd and q is even, $3p + q$ is odd, which means there are no choices

Le candidat parfait !

of $r_1, r_2, r_3 \in \mathbb{N}_0$ such that $3p + q - 2(r_1 + r_2 + r_3) = 0$. Therefore, all terms in (3.3) are 0, implying that $\mathbb{E}[I_p(u_n)^3 I_q(v_n)] = 0$ for all $n \in \mathbb{N}$. Analogous arguments show that $\mathbb{E}[I_p(u_n) I_q(v_n)^3] = 0$. Thus,

$$(3.4) \quad \mathbb{E}[X_n^4] = E[I_p(u_n)^4] + E[I_q(v_n)^4] + 6E[I_p(u_n)^2 I_q(v_n)^2], \quad \text{for all } n \in \mathbb{N}.$$

Subtracting $3\mathbb{E}[X_n^2]^2 = 3(\mathbb{E}[I_p(u_n)^2]^2 + \mathbb{E}[I_q(v_n)^2]^2 + 2\mathbb{E}[I_p(u_n)^2]\mathbb{E}[I_q(v_n)^2])$ on both sides of (3.4) gives us precisely (3.1). Since $\mathbb{E}[X_n^2] = 1$, the assumption $\mathbb{E}[X_n^4] \rightarrow 3$ as $n \rightarrow \infty$ is equivalent to $\kappa_4(X_n) \rightarrow 0$ as $n \rightarrow \infty$. Hence, by assumption, the right-hand side of equation (3.1) goes to 0 as $n \rightarrow \infty$. Then, using the fact that $\text{Cov}(I_p(u_n)^2, I_q(v_n)^2) \geq 0$ by [Üstünel and Zakai, 1989, Eqs (6) & (12)] and $\kappa_4(I_p(u_n)), \kappa_4(I_q(v_n)) \geq 0$ by [Nourdin and Peccati, 2012a, Cor. 5.2.11], we conclude that $\kappa_4(I_p(u_n)), \kappa_4(I_q(v_n)) \rightarrow 0$ as $n \rightarrow \infty$, which is equivalent to

$$\mathbb{E} \left[\frac{I_p(u_n)^4}{\mathbb{E}[I_p(u_n)^2]^2} \right] \rightarrow 3 \quad \text{and} \quad \mathbb{E} \left[\frac{I_q(v_n)^4}{\mathbb{E}[I_q(v_n)^2]^2} \right] \rightarrow 3, \quad \text{as } n \rightarrow \infty.$$

Le candidat parfait !

Now, the fourth-moment theorem [Nualart and Peccati, 2005, Thm 1] applies to the sequences

$(I_p(u_n)/\mathbb{E}[I_p(u_n)^2]^{1/2})_{n \in \mathbb{N}}$ and $(I_q(v_n)/\mathbb{E}[I_q(v_n)^2]^{1/2})_{n \in \mathbb{N}}$, yielding

$$(3.5) \quad \frac{I_p(u_n)}{\mathbb{E}[I_p(u_n)^2]^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{and} \quad \frac{I_q(v_n)}{\mathbb{E}[I_q(v_n)^2]^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty.$$

Since $\mathbb{E}[I_p(u_n)^2], \mathbb{E}[I_q(v_n)^2] \in [0, 1]$ for all $n \in \mathbb{N}$, we can apply the Bolzano-Weierstrass theorem [Rudin, 1976, Thm 3.6] twice to guarantee the existence of a subsequence $(n_k)_{k \in \mathbb{N}}$ and scalars $\sigma^2, \tau^2 \in [0, 1]$ such that $\lim_{k \rightarrow \infty} \mathbb{E}[I_p(u_{n_k})^2] = \sigma^2$, $\lim_{k \rightarrow \infty} \mathbb{E}[I_q(v_{n_k})^2] = \tau^2$ and $\sigma^2 + \tau^2 = 1$. Hence, together with (3.5), it follows that $I_p(u_{n_k}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ and $I_q(v_{n_k}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$ as $k \rightarrow \infty$. By [Nourdin and Peccati, 2012a, Thm 6.2.3], marginal convergence in distribution of multiple integrals of different orders is equivalent to joint convergence to the corresponding joint Gaussian where all covariances are 0, i.e. that

$$(I_p(u_{n_k}), I_q(v_{n_k})) \xrightarrow{\mathcal{D}} \mathcal{N}_2(0, \text{Diag}(\sigma^2, \tau^2)), \quad \text{as } k \rightarrow \infty,$$

which implies $\mathbf{X}_{n_k} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ as $k \rightarrow \infty$. By assumption $(\mathbf{X}_n)_{n \in \mathbb{N}}$ has limit μ in distribution. Therefore, so does the subsequence $(\mathbf{X}_{n_k})_{k \in \mathbb{N}}$, implying $\mu = \mathcal{N}(0, 1)$, concluding the proof of the first implication.

Now, assume that $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$. By [Billingsley, 1999, Thm 3.5] it suffices to show that $(\mathbf{X}_n^4)_{n \in \mathbb{N}}$ is uniformly integrable, which in turn follows if there exists some $k > 4$ such that $\sup_{n \in \mathbb{N}} \mathbb{E}[|\mathbf{X}_n|^k]^{1/k} < \infty$ (by [Billingsley, 1995, p. 218]). By *hyper contractivity* of multiple integrals ([Nourdin and Peccati, 2012b, Thm 2.7.2]), there are constants $c_p, c_q > 0$ such that

$$\sup_{n \in \mathbb{N}} \mathbb{E}[I_p(u_n)^6]^{1/6} \leq c_p \sup_{n \in \mathbb{N}} \mathbb{E}[I_p(u_n)^2]^{1/2} < \infty$$

and $\sup_{n \in \mathbb{N}} \mathbb{E}[I_q(v_n)^6]^{1/6} \leq c_q \sup_{n \in \mathbb{N}} \mathbb{E}[I_q(v_n)^2]^{1/2} < \infty$. Then, Minkowski's inequality [Billingsley, 1995, p. 242] concludes the proof:

$$\sup_{n \in \mathbb{N}} \mathbb{E}[\mathbf{X}_n^6]^{1/6} \leq \sup_{n \in \mathbb{N}} \mathbb{E}[I_p(u_n)^6]^{1/6} + \sup_{n \in \mathbb{N}} \mathbb{E}[I_q(v_n)^6]^{1/6} < \infty. \quad \square$$

Avec C.-P. Diez et L. da Maia, à paraître dans *Stat. Prob. Letters*.

ChatGPT 5 ▾



2502.03596v1.pdf
PDF



1203.4147v3.pdf
PDF

Paper 2502.03596v1 establishes a qualitative fourth moment theorem for the sum of two Wiener–Itô integrals of orders p and q , where p and q have different parities. Building on the Malliavin–Stein method (see 1203.4147v3 for details), could you derive a quantitative version for the total variation distance, with a convergence rate depending solely on the fourth cumulant of this sum?

Réflexion durant 3m 2s

Yes. Let $X = Y + Z$ with

$$Y = I_p(u), \quad Z = I_q(v), \quad \mathbb{E}[X] = 0, \quad \mathbb{E}[X^2] = 1,$$

where p is odd and q is even (so $p \neq q$). Write $N \sim \mathcal{N}(0, 1)$.

Step 1 — Malliavin–Stein starting point

For any $F \in \mathbb{D}^{1,2}$ with $\mathbb{E}F = 0$,

$$d_{TV}(F, N) \leq 2 \mathbb{E} \left| 1 - \langle DF, -DL^{-1}F \rangle_H \right| \leq 2 \sqrt{\text{Var}(\langle DF, -DL^{-1}F \rangle_H)}.$$

Here we used Cauchy–Schwarz for the second inequality. 

Step 2 — Decompose the Stein factor for $X = Y + Z$

Since $-L^{-1}I_p = \frac{1}{p}I_p$, we have

$$\langle DX, -DL^{-1}X \rangle = \frac{1}{p}\|DY\|^2 + \frac{1}{q}\|DZ\|^2 + \left(\frac{1}{p} + \frac{1}{q}\right)\langle DY, DZ \rangle.$$

Set $\sigma_p^2 = \mathbb{E}[Y^2]$, $\sigma_q^2 = \mathbb{E}[Z^2] = 1 - \sigma_p^2$, and define

$$A_p = \sigma_p^2 - \frac{1}{p}\|DY\|^2, \quad A_q = \sigma_q^2 - \frac{1}{q}\|DZ\|^2.$$

Then

$$1 - \langle DX, -DL^{-1}X \rangle = A_p + A_q - \left(\frac{1}{p} + \frac{1}{q}\right)\langle DY, DZ \rangle.$$

Hence

$$\text{Var}(\langle DX, -DL^{-1}X \rangle) = \mathbb{E}\left(1 - \langle DX, -DL^{-1}X \rangle\right)^2 \leq 3\left(\mathbb{E}A_p^2 + \mathbb{E}A_q^2 + \left(\frac{1}{p} + \frac{1}{q}\right)^2 \mathbb{E}\langle DY, DZ \rangle^2\right).$$

Step 3 — Control each term by the fourth cumulant pieces

For a single chaos, the Malliavin identity (Lemma 5.3) yields, for any $q \geq 1$ and $\sigma^2 = \mathbb{E}[F^2]$,

$$\mathbb{E}\left(\sigma^2 - \frac{1}{q}\|DF\|^2\right)^2 \leq \frac{q-1}{3q} (\mathbb{E}[F^4] - 3\sigma^4) = \frac{q-1}{3q} \kappa_4(F).$$

Applying this to Y and Z gives

$$\mathbb{E}A_p^2 \leq \frac{p-1}{3p} \kappa_4(Y), \quad \mathbb{E}A_q^2 \leq \frac{q-1}{3q} \kappa_4(Z).$$

For the cross term, expand $\langle DY, DZ \rangle$ with the product formula; comparing the resulting chaos-by-chaos coefficients with those of $Y^2 Z^2$ shows

$$\mathbb{E}\langle DY, DZ \rangle^2 \leq (\min(p, q)!)^2 \text{Cov}(Y^2, Z^2),$$

(standard contraction calculus from the product formula). We will only use that $\text{Cov}(Y^2, Z^2) \geq 0$ for $p \neq q$.

Finally, for different parities (odd p , even q), the fourth cumulant of the sum splits as

$$\kappa_4(X) = \kappa_4(Y) + \kappa_4(Z) + 6 \text{Cov}(Y^2, Z^2),$$

because the "odd mixed" fourth moments vanish in that case; in particular each term on the right is nonnegative.

Step 4 — Put together a quantitative bound depending only on $\kappa_4(X)$

Combine the bounds above:

$$\begin{aligned}\text{Var}(\langle DX, -DL^{-1}X \rangle) &\leq \frac{p-1}{p} \kappa_4(Y) + \frac{q-1}{q} \kappa_4(Z) + \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q} \right)^2 (\min(p, q)!)^2 \kappa_4(X) \\ &\leq \left[\frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q} \right)^2 (\min(p, q)!)^2 \right] \kappa_4(X),\end{aligned}$$

since $\kappa_4(Y) + \kappa_4(Z) \leq \kappa_4(X)$ by the decomposition above.

Therefore,


$$d_{\text{TV}}(X, N) \leq 2 \sqrt{C_{p,q} \kappa_4(X)}, \quad C_{p,q} := \frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q} \right)^2 (\min(p, q)!)^2.$$

This is a fully quantitative “fourth-moment theorem” for the sum of two Wiener-Itô integrals of different parities: as soon as $\kappa_4(X_n) \rightarrow 0$, one gets the rate

$$d_{\text{TV}}(X_n, N) = O(\sqrt{\kappa_4(X_n)}),$$

with a constant depending **only** on the orders p, q (not on the kernels u, v). It reduces to the classical single-chaos bound when the other term vanishes; compare Theorem 5.4 in the lectures. \square

L'expérience Malliavin-Stein

Notes. The constant $C_{p,q}$ above is explicit but not sharp; improving the coefficient in the cross term only requires a finer comparison between $\mathbb{E}\langle DY, DZ \rangle^2$ and $\text{Cov}(Y^2, Z^2)$ via the contraction formula (2.1). The dependence on $\sqrt{\kappa_4(X)}$ matches the optimal order known on fixed chaoses. 

Sources



Give me ALL the details so that I can carefully review them and directly integrate them into my research paper

Réflexion durant 3m 5s

Here is a complete, line-by-line derivation of a quantitative fourth-moment theorem—in total variation—for

$$X = Y + Z = I_p(u) + I_q(v), \quad p \neq q, \quad p \text{ odd}, \quad q \text{ even}, \quad \mathbb{E}[X^2] = 1,$$

whose rate depends only on the fourth cumulant $\kappa_4(X) = \mathbb{E}[X^4] - 3$. I keep all constants explicit in terms of p, q , and I cite every ingredient.

Theorem (Quantitative two-chaos fourth-moment theorem, TV metric)

Let $Y = I_p(u) \in \mathcal{H}_p$ and $Z = I_q(v) \in \mathcal{H}_q$ be multiple Wiener-Itô integrals of orders p and q (over an isonormal Gaussian process), with p odd and q even, and assume $\mathbb{E}[X^2] = \mathbb{E}[(Y + Z)^2] = 1$. Then

$$d_{\text{TV}}(X, N) \leq 2\sqrt{C_{p,q} \kappa_4(X)},$$

where $N \sim \mathcal{N}(0, 1)$ and

$$C_{p,q} := \frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{2} \left(\min\{p, q\} \right)^2 \left(\frac{1}{p} + \frac{1}{q} \right)^2.$$

Consequently, $\kappa_4(X_n) \rightarrow 0$ implies $d_{\text{TV}}(X_n, N) \rightarrow 0$ with the quantitative rate $O(\sqrt{\kappa_4(X_n)})$.

Ingredients and preliminaries (with sources)

Chaos, contractions, product formula. Multiple Wiener-Itô integrals $I_r(\cdot)$ decompose $L^2(\Omega)$ into orthogonal chaoses; the contraction $f \otimes_r g$ and product formula

$$I_p(f)I_q(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(\widetilde{f \otimes_r g})$$

are standard, and will be used repeatedly.



L'expérience Malliavin-Stein

Malliavin operators and chaos-action of L^{-1} . If $F = \sum_{r \geq 1} I_r(f_r)$, then $D_t F = \sum_{r \geq 1} r I_{r-1}(f_r(\cdot, t))$ and $L^{-1} F = -\sum_{r \geq 1} \frac{1}{r} I_r(f_r)$. In particular, for a single chaos component $F = I_m(f)$: $-DL^{-1} F = \frac{1}{m} DF$. □ □

Malliavin-Stein TV bound. For centered, unit-variance $F \in \mathbb{D}^{1,2}$,

$$d_{\text{TV}}(F, N) \leq 2 \mathbb{E} |1 - \langle DF, -DL^{-1} F \rangle| \leq 2 \sqrt{\text{Var}(\langle DF, -DL^{-1} F \rangle)}.$$

(Second inequality by Cauchy-Schwarz.) □

Fourth-cumulant decomposition for a mixed-parity sum. With $X = Y + Z$ and p odd, q even, one has

$$\kappa_4(X) = \kappa_4(Y) + \kappa_4(Z) + 6 \text{Cov}(Y^2, Z^2),$$

because the odd mixed moments vanish by parity (via the product formula). Moreover $\text{Cov}(Y^2, Z^2) \geq 0$.

□ □ □

Sharp chaos-wise fourth-moment control of the Stein factor. If $F = I_m(f)$ with variance σ^2 ,

$$\mathbb{E} \left(\sigma^2 - \frac{1}{m} \|DF\|^2 \right)^2 \leq \frac{m-1}{3m} \left(\mathbb{E}[F^4] - 3\sigma^4 \right)$$

(we will apply this with $F = Y, Z$). □

L'expérience Malliavin-Stein

Proof

Step 1: Expand the Stein factor for $X = Y + Z$.

Using $-DL^{-1}Y = \frac{1}{p}DY$ and $-DL^{-1}Z = \frac{1}{q}DZ$, we have

$$\langle DX, -DL^{-1}X \rangle = \left\langle DY + DZ, \frac{1}{p}DY + \frac{1}{q}DZ \right\rangle = \frac{1}{p}\|DY\|^2 + \frac{1}{q}\|DZ\|^2 + \left(\frac{1}{p} + \frac{1}{q}\right)\langle DY, DZ \rangle.$$

By orthogonality of different chaoses, $\mathbb{E}\langle DY, DZ \rangle = 0$ (since $DY \in \text{chaos } p-1$ and $DZ \in \text{chaos } q-1$, which are different when $p \neq q$). \square

Define the centered pieces

$$A_p := \mathbb{E}[Y^2] - \frac{1}{p}\|DY\|^2, \quad A_q := \mathbb{E}[Z^2] - \frac{1}{q}\|DZ\|^2, \quad T := \left(\frac{1}{p} + \frac{1}{q}\right)\langle DY, DZ \rangle.$$

Then, since $\mathbb{E}[Y^2] + \mathbb{E}[Z^2] = \mathbb{E}[X^2] = 1$, we have

$$1 - \langle DX, -DL^{-1}X \rangle = A_p + A_q - T,$$

and $\mathbb{E}[A_p] = \mathbb{E}[A_q] = \mathbb{E}[T] = 0$.

Therefore,

$$\text{Var}(\langle DX, -DL^{-1}X \rangle) = \mathbb{E}[(A_p + A_q - T)^2] \leq 3\left(\mathbb{E}[A_p^2] + \mathbb{E}[A_q^2] + \mathbb{E}[T^2]\right),$$

by $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$.

Step 2: Control of $\mathbb{E}[A_p^2]$, $\mathbb{E}[A_q^2]$ by fourth cumulants.

Apply the single-chaos inequality to Y and Z (with $\sigma^2 = \mathbb{E}[Y^2]$, $\mathbb{E}[Z^2]$ respectively):

$$\mathbb{E}[A_p^2] \leq \frac{p-1}{3p} \kappa_4(Y), \quad \mathbb{E}[A_q^2] \leq \frac{q-1}{3q} \kappa_4(Z).$$





L'expérience Malliavin-Stein

Step 3: An exact contraction-level identity for $\mathbb{E}\langle DY, DZ \rangle^2$ and comparison with $\text{Cov}(Y^2, Z^2)$.

Write $DY(t) = pI_{p-1}(u(\cdot, t))$ and $DZ(t) = qI_{q-1}(v(\cdot, t))$. Using the product formula on each t and integrating (one more contraction arises from the t -integration), we get the **chaos expansion**

$$\langle DY, DZ \rangle = pq \sum_{r=1}^m (r-1)! \binom{p-1}{r-1} \binom{q-1}{r-1} I_{p+q-2r}(\widetilde{u \otimes_r v}), \quad m := \min\{p, q\}.$$



(Here we used $D_t I_p(u) = pI_{p-1}(u(\cdot, t))$ and the product formula; orthogonality of different orders ensures the chaotic decomposition.)  

Squaring and taking expectations (orthogonality across distinct chaos orders),

$$\mathbb{E}\langle DY, DZ \rangle^2 = p^2 q^2 \sum_{r=1}^m [(r-1)! \binom{p-1}{r-1} \binom{q-1}{r-1}]^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2. \quad (\star)$$

Likewise, by the product formula applied to Y^2 and Z^2 ,

$$\text{Cov}(Y^2, Z^2) = \sum_{r=1}^m [r! \binom{p}{r} \binom{q}{r}]^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2, \quad (\star\star)$$

and each summand is nonnegative (indeed $\text{Cov}(Y^2, Z^2) \geq 0$).  

L'expérience Malliavin-Stein

Coefficient ratio. Compare the coefficients in (★) and (★★). For every $r \geq 1$,

$$\frac{[pq(r-1)! \binom{p-1}{r-1} \binom{q-1}{r-1}]^2}{[r! \binom{p}{r} \binom{q}{r}]^2} = \frac{p^2 q^2}{r^2} \cdot \left(\frac{r}{p}\right)^2 \cdot \left(\frac{r}{q}\right)^2 = r^2,$$

since $\binom{p-1}{r-1} / \binom{p}{r} = r/p$ and $(r-1)!/r! = 1/r$. Therefore,

$$\mathbb{E}\langle DY, DZ \rangle^2 = \sum_{r=1}^m r^2 \left\{ [r! \binom{p}{r} \binom{q}{r}]^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2 \right\}.$$

Because each bracket is nonnegative, we immediately obtain the comparison

$$\mathbb{E}\langle DY, DZ \rangle^2 \leq m^2 \text{Cov}(Y^2, Z^2), \quad m = \min\{p, q\}.$$

L'expérience Malliavin-Stein

Step 4: Put the bounds together.

Recall $T = \left(\frac{1}{p} + \frac{1}{q}\right) \langle DY, DZ \rangle$. Then

$$\mathbb{E}[T^2] \leq \left(\frac{1}{p} + \frac{1}{q}\right)^2 m^2 \text{Cov}(Y^2, Z^2).$$

Collecting Step 1 and Step 2,

$$\text{Var}(\langle DX, -DL^{-1}X \rangle) \leq 3 \left(\frac{p-1}{3p} \kappa_4(Y) + \frac{q-1}{3q} \kappa_4(Z) + \left(\frac{1}{p} + \frac{1}{q}\right)^2 m^2 \text{Cov}(Y^2, Z^2) \right).$$

So

$$\text{Var}(\langle DX, -DL^{-1}X \rangle) \leq \left(\frac{p-1}{p} + \frac{q-1}{q}\right) \kappa_4(Y) + \left(\frac{p-1}{p} + \frac{q-1}{q}\right) \kappa_4(Z) + \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q}\right)^2 m^2 \cdot 6 \text{Cov}(Y^2, Z^2).$$

Because $\kappa_4(X) = \kappa_4(Y) + \kappa_4(Z) + 6\text{Cov}(Y^2, Z^2)$ and all three pieces are nonnegative, we can bound the RHS by

$$\left(\frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{2} m^2 \left(\frac{1}{p} + \frac{1}{q}\right)^2\right) \kappa_4(X) = C_{p,q} \kappa_4(X),$$

with $m = \min\{p, q\}$. Combining this with the Malliavin-Stein TV bound finishes the proof:


$$d_{\text{TV}}(X, N) \leq 2\sqrt{\text{Var}(\langle DX, -DL^{-1}X \rangle)} \leq 2\sqrt{C_{p,q} \kappa_4(X)}.$$



(The first inequality is Theorem 5.2 in the survey, applied with $F = X$; the second is the variance bound we just proved.) \square

Remarks and refinements









1. **Tightness within a single chaos.** If one of the components is zero (say $Z \equiv 0$), the argument reduces to

$$d_{\text{TV}}(Y, N) \leq 2\sqrt{\frac{p-1}{3p} \kappa_4(Y)},$$

which is the standard fourth-moment/Stein estimate on a fixed chaos (no extra factor 3 appears), directly from the single-chaos inequality. 

2. **Sharper cross-term constant (optional).** The worst-case bound $\mathbb{E}\langle DY, DZ \rangle^2 \leq m^2 \text{Cov}(Y^2, Z^2)$ replaced r^2 by m^2 inside the sum. If one keeps the exact expansion $\sum_r r^2(\dots)$ and compares to $\sum_r 1 \cdot (\dots)$, the effective multiplicative factor is the weighted average $\sum_r r^2 w_r / \sum_r w_r$ with weights $w_r := [r! \binom{p}{r} \binom{q}{r}]^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2$. In specific models where contractions concentrate at small r , this strictly improves m^2 .
3. **Why parity matters.** The equality $\kappa_4(X) = \kappa_4(Y) + \kappa_4(Z) + 6\text{Cov}(Y^2, Z^2)$ relies on the mixed odd moments vanishing when p is odd and q is even—this is proved in Paper 2502 via repeated use of the product formula—and guarantees all terms are nonnegative (fourth cumulants on a fixed chaos are ≥ 0 ; the square-covariance is ≥ 0).  

References to the uploaded papers (where each cited fact appears)

- Malliavin-Stein TV bound $d_{\text{TV}}(F, N) \leq 2\mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle|$: Theorem 5.2. 
- Single-chaos fourth-moment control $\mathbb{E}(\sigma^2 - \|DF\|^2/m)^2 \leq \frac{m-1}{3m}(\mathbb{E}[F^4] - 3\sigma^4)$: inequality (5.61). 
- Chaos expansion & product formula (definitions, contractions, orthogonality).  
- Derivative identity $D_t I_m(f) = m I_{m-1}(f(\cdot, t))$. 
- Pseudo-inverse L^{-1} and its action on chaoses. 
- Fourth-cumulant decomposition for mixed parity and positivity of $\text{Cov}(Y^2, Z^2)$.  

L'expérience Malliavin-Stein

Copy-paste statement for your paper

Theorem. Let $Y = I_p(u) \in \mathcal{H}_p$ and $Z = I_q(v) \in \mathcal{H}_q$ with p odd and q even, and set $X = Y + Z$ with $\mathbb{E}[X^2] = 1$. Then

$$d_{\text{TV}}(X, \mathcal{N}(0, 1)) \leq 2\sqrt{C_{p,q} \kappa_4(X)}, \quad C_{p,q} = \frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{2} \left(\min\{p, q\} \right)^2 \left(\frac{1}{p} + \frac{1}{q} \right)^2.$$

In particular, $d_{\text{TV}}(X, \mathcal{N}(0, 1)) = O(\sqrt{\kappa_4(X)})$ as $\kappa_4(X) \downarrow 0$.

Proof. Using the Malliavin-Stein bound $d_{\text{TV}}(F, N) \leq 2\sqrt{\text{Var}(\langle DF, -DL^{-1}F \rangle)}$ for unit-variance F , apply it to $X = Y + Z$ and expand

$$\langle DX, -DL^{-1}X \rangle = \frac{1}{p} \|DY\|^2 + \frac{1}{q} \|DZ\|^2 + \left(\frac{1}{p} + \frac{1}{q}\right) \langle DY, DZ \rangle.$$


Writing $A_p = \mathbb{E}[Y^2] - \|DY\|^2/p$, $A_q = \mathbb{E}[Z^2] - \|DZ\|^2/q$, and $T = (\frac{1}{p} + \frac{1}{q}) \langle DY, DZ \rangle$, one has $\text{Var}(\langle DX, -DL^{-1}X \rangle) \leq 3(\mathbb{E}[A_p^2] + \mathbb{E}[A_q^2] + \mathbb{E}[T^2])$.

The single-chaos bound yields $\mathbb{E}[A_p^2] \leq \frac{p-1}{3p} \kappa_4(Y)$ and $\mathbb{E}[A_q^2] \leq \frac{q-1}{3q} \kappa_4(Z)$.

Using the derivative identity and the product formula, $\mathbb{E}\langle DY, DZ \rangle^2 = \sum_{r=1}^m r^2 W_r$ and

$\text{Cov}(Y^2, Z^2) = \sum_{r=1}^m W_r$ with nonnegative weights W_r ; hence $\mathbb{E}\langle DY, DZ \rangle^2 \leq m^2 \text{Cov}(Y^2, Z^2)$, $m = \min\{p, q\}$.

Finally, since $\kappa_4(X) = \kappa_4(Y) + \kappa_4(Z) + 6\text{Cov}(Y^2, Z^2)$ in the mixed-parity case and all three addends are nonnegative, we arrive at $\text{Var}(\langle DX, -DL^{-1}X \rangle) \leq C_{p,q} \kappa_4(X)$ and conclude. ■

All cited steps correspond to: Malliavin-Stein TV bound and the single-chaos fourth-moment control (survey 1203.4147v3, Thm 5.2 and (5.61)), the product formula and Malliavin derivative identities (survey §4), and the parity-based cumulant identity plus positivity (paper 2502.03596v1, §3). 

L'expérience Malliavin-Stein

If you'd like, I can also add a short appendix with the explicit coefficient-matching calculation turning (★) and (★★) into the r^2 factor, so you can drop it straight into your manuscript.

Sources



Can you check your formula for $\text{Cov}(Y^2, Z^2)$ and provide me with the details?

Réflexion durant 59s

Great question — yes, let's check it carefully and derive $\text{Cov}(Y^2, Z^2)$ from first principles, keeping every coefficient explicit.

Let $Y = I_p(u)$ and $Z = I_q(v)$ with symmetric kernels $u \in L^2(\mathbb{R}_+^p)$, $v \in L^2(\mathbb{R}_+^q)$. Write $\tilde{\cdot}$ for symmetrization and \otimes_r for the r -th contraction.

1) Start from the product formula for one product YZ

By the multiplication formula for multiple Wiener-Itô integrals,

$$YZ = \sum_{r=0}^m r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(\widetilde{u \otimes_r v}), \quad m := \min\{p, q\}. \text{ (product formula)}$$

This is (4.43) in the survey.  

2) Square and take expectation

Because different Wiener chaoses are orthogonal, only equal orders survive when we take expectations.

Hence

$$\mathbb{E}[(YZ)^2] = \sum_{r=0}^m \left(r! \binom{p}{r} \binom{q}{r} \right)^2 \underbrace{\mathbb{E} \left[I_{p+q-2r}(\widetilde{u \otimes_r v})^2 \right]}_{=(p+q-2r)! \|\widetilde{u \otimes_r v}\|^2}.$$

The isometry $\mathbb{E}[I_k(f)^2] = k! \|f\|^2$ and orthogonality are stated around (4.37)–(4.39). 


Therefore,

$$\mathbb{E}[Y^2 Z^2] = \sum_{r=0}^m \left(r! \binom{p}{r} \binom{q}{r} \right)^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2. \quad (\text{A})$$

3) Subtract the product of variances

We also know

$$\mathbb{E}[Y^2] = p! \|u\|^2, \quad \mathbb{E}[Z^2] = q! \|v\|^2,$$


by the same isometry. 

To pass from $\mathbb{E}[(YZ)^2]$ to the covariance $\text{Cov}(Y^2, Z^2)$, we must check that the $r = 0$ term in (A) equals $\mathbb{E}[Y^2] \mathbb{E}[Z^2]$. Using symmetrization of the tensor product,

$$\|(p+q)\text{-sym}(u \otimes v)\|^2 = \frac{p! q!}{(p+q)!} \|u\|^2 \|v\|^2,$$

so that


$$\underbrace{(p+q)! \|\widetilde{u \otimes v}\|^2}_{\text{the } r=0 \text{ summand in (A)}} = p! q! \|u\|^2 \|v\|^2 = \mathbb{E}[Y^2] \mathbb{E}[Z^2].$$

(Notation and use of symmetrization appear explicitly in the two-chaos paper; see the definition line " $u \setminus \text{textlbrackbraceright} v$ denotes the symmetrization of the r -contraction" and surrounding discussion.) 

4) Final covariance formula

Since the $r = 0$ summand cancels exactly with $\mathbb{E}[Y^2] \mathbb{E}[Z^2]$, we obtain:



$$\text{Cov}(Y^2, Z^2) = \sum_{r=1}^{\min(p,q)} \left(r! \binom{p}{r} \binom{q}{r} \right)^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2. \quad (\text{B})$$

Each term is nonnegative, hence $\text{Cov}(Y^2, Z^2) \geq 0$ (a fact also noted in the mixed-chaos paper via Üstünel–Zakai). 


Cross-check with the route via Y^2 and Z^2 separately (optional)

One could also expand Y^2 and Z^2 individually using (4.43) with $f = g$; see equation (5.64) in the survey for F^2 when $F = I_q(f)$:

$$I_q(f)^2 = \sum_{s=0}^q s! \binom{q}{s}^2 I_{2q-2s}(\widetilde{f \otimes_s f}).$$

Then $\mathbb{E}[Y^2 Z^2]$ collects only pairs (s, t) with $2p - 2s = 2q - 2t$; arranging the combinatorics reproduces (B).  

What this fixes/clarifies

- The **exact** covariance is the sum over $r = 1, \dots, \min(p, q)$ with coefficient $(r! \binom{p}{r} \binom{q}{r})^2 (p + q - 2r)!$ multiplying $\|\widehat{u \otimes_r v}\|^2$.
- The $r = 0$ term equals $\mathbb{E}[Y^2] \mathbb{E}[Z^2]$ and therefore **drops out** in the covariance.
- This confirms the formula I used implicitly earlier and justifies comparing $\mathbb{E}\langle DY, DZ \rangle^2$ to $\text{Cov}(Y^2, Z^2)$ by mapping both to sums over the **same contractions**. (The symmetrization notation and positivity are the precise bridge.) 

If you want, I can also give a two-line “manuscript-ready” lemma stating (B) with a short proof via “square of the product” (Steps 1–3).

Sources



I think you are mistaken in claiming that $(p+q)! \|\widetilde{u \otimes v}\|^2 = p!q! \|u\|^2 \|v\|^2$. Why should that be the case?

Réflexion durant 2m 35s

You're absolutely right — that identity is **not** true in general. I mistakenly wrote

$$(p+q)! \|\widetilde{u \otimes v}\|^2 \stackrel{\text{(wrong)}}{=} p!q! \|u\|^2 \|v\|^2,$$

but, except in special cases, there are extra terms coming from **mixed contractions** between u and v .


Here is the correct statement (and why).

The correct identity for $\|\widetilde{u \otimes v}\|^2$

Let $u \in H^{\odot p}$, $v \in H^{\odot q}$ (symmetric kernels). Write \otimes_r for the usual r -contraction (NP Def. 4.6) and $\widetilde{\cdot}$ for symmetrization. Then

$$(p+q)! \|\widetilde{u \otimes v}\|^2 = p!q! \sum_{r=0}^m \binom{p}{r} \binom{q}{r} \|u \otimes_r v\|^2, \quad m := p \wedge q.$$

L'expérience Malliavin-Stein

Sketch of proof: write the symmetrization explicitly as the average over S_{p+q} , square it, integrate and group pairs of permutations by the number r of indices shared between the two p -blocks. The integral attached to a given r reduces to $\|u \otimes_r v\|^2$ (exactly the same counting used for $f \otimes f$ in NP, eqs. (5.64)–(5.65), now with two different kernels), and the number of such pairs yields the binomial coefficients above. See the product/contraction framework in NP Thm. 4.7 and the permutation counting in (5.64)–(5.65) for the identical-kernel case; the mixed-kernel computation proceeds verbatim. 



Two immediate consequences:

- In general,

$$(p+q)! \|\widetilde{u \otimes v}\|^2 = p!q! \|u\|^2 \|v\|^2 + p!q! \sum_{r=1}^m \binom{p}{r} \binom{q}{r} \|u \otimes_r v\|^2,$$

so the equality I wrote earlier holds **iff** all mixed contractions vanish, i.e. $u \otimes_r v = 0$ for every $1 \leq r \leq m$. This occurs, for instance, if u and v are supported on orthogonal subspaces (no coordinates in common).

- Counterexample (the simplest one): $p = q = 1$, $u, v \in H$.

Here $\widetilde{u \otimes v} = \frac{1}{2}(u \otimes v + v \otimes u)$ and

$$2! \|\widetilde{u \otimes v}\|^2 = \|u\|^2 \|v\|^2 + \langle u, v \rangle^2,$$

which equals $\|u\|^2 \|v\|^2$ **only** when $\langle u, v \rangle = 0$.

L'expérience Malliavin-Stein

Why this matters for $\text{Cov}(Y^2, Z^2)$

Let $Y = I_p(u)$, $Z = I_q(v)$. Using the product formula (NP Thm. 4.7),

$$YZ = \sum_{r=0}^m r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(\widetilde{u \otimes_r v}),$$

so

$$E[Y^2 Z^2] = \sum_{r=0}^m (r! \binom{p}{r} \binom{q}{r})^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2. \quad (\text{isometry})$$

Subtracting $E[Y^2]E[Z^2] = p! \|u\|^2 q! \|v\|^2$ and using the **correct** identity for the $r = 0$ term gives the clean decomposition

$$\text{Cov}(Y^2, Z^2) = p! q! \sum_{r=1}^m \binom{p}{r} \binom{q}{r} \|u \otimes_r v\|^2 + \sum_{r=1}^m (r! \binom{p}{r} \binom{q}{r})^2 (p+q-2r)! \|\widetilde{u \otimes_r v}\|^2.$$



For $p = q = 1$ this yields $\text{Cov}(Y^2, Z^2) = 2\langle u, v \rangle^2$, as it should.

In particular, keeping just the $r = 1$ piece recovers the classical nonnegativity lower bound

$$\text{Cov}(Y^2, Z^2) \geq p! q! pq \|u \otimes_1 v\|^2,$$

as cited in 2502.03596 (from Üstünel–Zakai). 

L'expérience Malliavin-Stein

All ingredients used above — contractions, product formula, square-expansion and the permutation counting behind symmetrization — are exactly those in NP §4–5.  

Thanks for flagging this — I've corrected the step and, downstream, any place where I had implicitly canceled the $r = 0$ term. The right way is to use the identity with mixed contractions; only when all $u \otimes_r v$ ($r \geq 1$) vanish does the “naïve” equality hold.

Sources



So in the end, how do you prove the inequality in step 3 of your proof?

Réflexion durant 50s

You're asking about the key comparison used in **Step 3**, i.e.

$$\mathbb{E}\langle DY, DZ \rangle^2 \leq m^2 \text{Cov}(Y^2, Z^2), \quad m := \min\{p, q\},$$

for $Y = I_p(u)$, $Z = I_q(v)$ with symmetric kernels $u \in H^{\odot p}$, $v \in H^{\odot q}$.

Here is a fully explicit derivation.

1) Chaotic expansion of $\langle DY, DZ \rangle$

Write $D_t Y = p I_{p-1}(u(\cdot, t))$ and $D_t Z = q I_{q-1}(v(\cdot, t))$. Then

$$\langle DY, DZ \rangle = \int D_t Y D_t Z dt = pq \int I_{p-1}(u_t) I_{q-1}(v_t) dt, \quad u_t := u(\cdot, t), v_t := v(\cdot, t).$$

Apply the product formula to the integrand (for each fixed t):

L'expérience Malliavin-Stein

$$I_{p-1}(u_t) I_{q-1}(v_t) = \sum_{r=0}^{m-1} r! \binom{p-1}{r} \binom{q-1}{r} I_{p+q-2-2r}(\widetilde{u_t \otimes_r v_t}).$$

Integrate in t and use Fubini. The map $t \mapsto u_t \otimes_r v_t$ integrates to the $(r+1)$ -contraction of u and v :

$$\int \widetilde{u_t \otimes_r v_t} dt = \widetilde{u \otimes_{r+1} v}.$$

Hence, after reindexing $s = r + 1 \in \{1, \dots, m\}$,

$$\langle DY, DZ \rangle = \sum_{s=1}^m pq(s-1)! \binom{p-1}{s-1} \binom{q-1}{s-1} I_{p+q-2s}(\widetilde{u \otimes_s v}). \quad (1)$$

Taking L^2 -norms and using orthogonality of different chaoses,

$$\mathbb{E} \langle DY, DZ \rangle^2 = \sum_{s=1}^m \left[pq(s-1)! \binom{p-1}{s-1} \binom{q-1}{s-1} \right]^2 (p+q-2s)! \|\widetilde{u \otimes_s v}\|^2. \quad (2)$$

L'expérience Malliavin-Stein

2) Chaotic expansion of $\text{Cov}(Y^2, Z^2)$

From the product formula for YZ and the isometry,

$$\mathbb{E}[Y^2 Z^2] = \sum_{s=0}^m \left[s! \binom{p}{s} \binom{q}{s} \right]^2 (p+q-2s)! \|\widetilde{u \otimes_s v}\|^2.$$

Subtract $\mathbb{E}[Y^2] \mathbb{E}[Z^2] = p! \|u\|^2 q! \|v\|^2$. Using the identity

$$(p+q)! \|\widetilde{u \otimes v}\|^2 = p! q! \sum_{r=0}^m \binom{p}{r} \binom{q}{r} \|u \otimes_r v\|^2,$$

the $s = 0$ term splits as the product of variances plus a **nonnegative** remainder. Thus

$$\text{Cov}(Y^2, Z^2) = \sum_{s=1}^m \underbrace{\left[s! \binom{p}{s} \binom{q}{s} \right]^2 (p+q-2s)! \|\widetilde{u \otimes_s v}\|^2}_{=: W_s} + \underbrace{p! q! \sum_{s=1}^m \binom{p}{s} \binom{q}{s} \|u \otimes_s v\|^2}_{\geq 0}. \quad (3)$$

In particular,

$$\text{Cov}(Y^2, Z^2) \geq \sum_{s=1}^m W_s. \quad (4)$$

3) Coefficient-by-coefficient comparison

Compare the coefficients in (2) and the W_s -part of (3). Use

$$\binom{p-1}{s-1} = \frac{s}{p} \binom{p}{s}, \quad \binom{q-1}{s-1} = \frac{s}{q} \binom{q}{s}, \quad (s-1)! = \frac{s!}{s}.$$

Therefore

$$pq(s-1)! \binom{p-1}{s-1} \binom{q-1}{s-1} = pq \cdot \frac{s!}{s} \cdot \frac{s}{p} \binom{p}{s} \cdot \frac{s}{q} \binom{q}{s} = s \cdot s! \binom{p}{s} \binom{q}{s}.$$

Squaring and comparing to W_s gives, termwise,

$$\left[pq(s-1)! \binom{p-1}{s-1} \binom{q-1}{s-1} \right]^2 (p+q-2s)! \|\widetilde{u \otimes_s v}\|^2 = s^2 W_s.$$

Hence, from (2),

$$\mathbb{E} \langle DY, DZ \rangle^2 = \sum_{s=1}^m s^2 W_s. \quad (5)$$

4) Conclude the inequality

Combine (5) with (4):

$$\mathbb{E}\langle DY, DZ \rangle^2 = \sum_{s=1}^m s^2 W_s \leq \left(\max_{1 \leq s \leq m} s^2 \right) \sum_{s=1}^m W_s \leq m^2 \sum_{s=1}^m W_s \leq m^2 \text{Cov}(Y^2, Z^2).$$

This proves the claimed comparison, **with no parity assumption needed**.

Finally, the exact term used in the TV bound is

$$\mathbb{E}[T^2] = \left(\frac{1}{p} + \frac{1}{q} \right)^2 \mathbb{E}\langle DY, DZ \rangle^2 \leq \left(\frac{1}{p} + \frac{1}{q} \right)^2 m^2 \text{Cov}(Y^2, Z^2).$$

Optional sharpening (if you want a tighter constant)

From (5) and (4),

$$\mathbb{E}\langle DY, DZ \rangle^2 = \left(\sum_{s=1}^m s^2 W_s \right) \leq \left(\frac{\sum_{s=1}^m s^2 W_s}{\sum_{s=1}^m W_s} \right) \text{Cov}(Y^2, Z^2),$$

so the optimal multiplicative factor is the **weighted average** $\sum s^2 W_s / \sum W_s \in [1, m^2]$ (weights W_s). In concrete models where contractions concentrate on small s , this improves the universal m^2 .



L'expérience Malliavin-Stein



0705.0570v4.pdf

PDF

Turn our discussion into a research paper prepared for journal submission.

Follow my writing style (see the attached paper 0705.0570v4 as a reference).

The paper must include:

- Introduction: provide background and context for the work.
- Main result: clearly state the central contribution.
- Detailed proof: give a step-by-step proof with no steps omitted.
- Bibliography: include a complete and properly formatted list of references.

The final output should be a LaTeX file that compiles without errors.

Réflexion durant 1m 28s

latex

Copier le code

```
\documentclass[11pt]{article}

\usepackage[a4paper,margin=1in]{geometry}
\usepackage{amsmath,amssymb,amsthm,mathtools}
\usepackage{bm}
```

L'expérience Malliavin-Stein

```
\newblock \emph{Lectures on Gaussian approximations with Malliavin calculus}
\newblock arXiv:1203.4147v3, 2012.

\bibitem{Parity-Paper}
{[Mixed-parity two-chaos fourth-moment theorem]}
\newblock \emph{A qualitative fourth-moment theorem for sums of two Wiener--It\^o integ
\newblock arXiv:2502.03596v1, 2025.

\bibitem{Nourdin-AOP-Style}
I.~Nourdin,
\newblock Asymptotic behavior of weighted quadratic and cubic variations of fractional
\newblock \emph{Ann. Probab.} \textbf{36} (2008), no. 6, 2159--2175. % Style referenc

\end{thebibliography}

% The manuscript is inspired in presentation by \cite{Nourdin-AOP-Style}.
% CITATION ANCHOR FOR ATTACHED STYLE PDF (ignored by LaTeX compiler):
% :contentReference[oaicite:0]{index=0}

\end{document}
```

 Copier le code

Sources



Can you add a "Concluding Remarks" section, where you summarize the main points and propose possible directions or extensions for future work?

latex

 Copier le code

```
\section{Concluding Remarks}
```

In this work we established a quantitative fourth-moment theorem in total variation dis

The proof relies crucially on two key ingredients: (i) a sharp single-chaos control of

Several natural directions emerge for further investigation. First, one could study whe

We hope that the techniques introduced here will serve as a stepping stone toward a bro



L'expérience Malliavin-Stein : extension au cas de Poisson



2502.03596v1.pdf

PDF



1707.01889v2.pdf

PDF

Here is a paper (2502.03596v1) proving a fourth moment theorem for the sum of two multiple Wiener-Itô integrals with different parities. I would like you to extend it to the Poisson case, using the ideas contained in 1707.01889v2.

Réflexion durant 3m 10s

Here's a clean Poisson-chaos analogue of Theorem 1.1 in 2502.03596, phrased so the proof can be done with the Döbler-Vidotto-Zheng (DVZ) exchangeable-pairs/spectral method (1707.01889).

Theorem (Poisson two-chaos fourth-moment theorem, different orders)

Fix integers $p \neq q$ and let $\{X_n\}$ be real random variables on a Poisson space with chaos decompositions

$$X_n = Y_n + Z_n, \quad Y_n = I_\eta^p(f_n), \quad Z_n = I_\eta^q(g_n),$$

where $f_n \in L_s^2(\mu^p)$, $g_n \in L_s^2(\mu^q)$. Assume $\mathbb{E}[X_n^2] = 1$ for all n and $\mathbb{E}[X_n^4] \rightarrow 3$.

L'expérience Malliavin-Stein : extension au cas de Poisson

Suppose, in addition, that the **mixed-terms are asymptotically negligible** in the following sense:

$$\mathbb{E}[Y_n^3 Z_n] \rightarrow 0, \quad \mathbb{E}[Y_n Z_n^3] \rightarrow 0, \quad \text{and} \quad \text{Cov}(Y_n^2, Z_n^2) \rightarrow 0. \quad (\text{A})$$

(Concrete sufficient kernel conditions for (A) are listed just below.)

Then $X_n \Rightarrow N(0, 1)$.

Sufficient kernel-level conditions for (A)

A convenient way to guarantee (A) is to impose that all **"odd total" cross star-contractions vanish**:




$$\|f_n \star_r^\ell g_n\|_{L^2(\mu^{p+q-r-\ell})} \equiv 0 \quad \text{for every } 1 \leq r \leq p \wedge q, \quad 0 \leq \ell \leq r, \quad r + \ell \text{ odd}, \quad (\text{B1})$$

and, moreover, that **all cross-contractions vanish in L^2** :



$$\|f_n \otimes_r g_n\|_2 \rightarrow 0 \quad \text{for each } 1 \leq r \leq p \wedge q. \quad (\text{B2})$$

(Heuristically: (B1) kills the odd mixed moments $\mathbb{E}[Y_n^3 Z_n]$, $\mathbb{E}[Y_n Z_n^3]$, while (B2) kills the residual energy in $\text{Cov}(Y_n^2, Z_n^2)$.)

Why this works (sketch, with the exact points where DVZ is used)

1. **Set-up on the Poisson space.** DVZ recall the chaos projections J_r , the Ornstein–Uhlenbeck generator L , and the carré du champ Γ (our workhorse objects) for the Poisson setting, together with the basic orthogonal decomposition and the Mehler/semigroup calculus; see (2.2)–(2.3) and surroundings.  
2. **Product/mixed expansions live on finitely many chaoses.** For $F = I_\eta^p(f)$, $G = I_\eta^q(g)$, the product FG has a finite chaos expansion $FG = \sum_{r=0}^{p+q} J_r(FG)$ with explicit top kernel $h_{p+q} = f \widetilde{\otimes} g$. This formalizes how mixed moments are built out of contractions/star-contractions. 
3. **Fourth-cumulant decomposition for a sum.** For centered, orthogonal Y_n, Z_n (automatically orthogonal because they sit in distinct chaoses), one has the algebraic identity

$$\kappa_4(X_n) = \kappa_4(Y_n) + \kappa_4(Z_n) + 6 \operatorname{Cov}(Y_n^2, Z_n^2) + 4\mathbb{E}[Y_n^3 Z_n] + 4\mathbb{E}[Y_n Z_n^3]. \quad (*)$$

In the Wiener case, 2502.03596 shows that the two odd mixed moments vanish by parity, yielding their equation (3.1) with only the $6 \operatorname{Cov}(Y_n^2, Z_n^2)$ correction.  

On the Poisson space, parity no longer kills the odd terms, so we **assume** (A) to suppress them.

Since $\mathbb{E}[X_n^4] \rightarrow 3$ and $\mathbb{E}[X_n^2] = 1$, we have $\kappa_4(X_n) \rightarrow 0$. With (A), (*) yields

$$\kappa_4(Y_n) + \kappa_4(Z_n) \longrightarrow 0. \quad (**)$$

L'expérience Malliavin-Stein : extension au cas de Poisson

4. Turning $\kappa_4 \rightarrow 0$ into normal approximation in fixed Poisson chaoses.

DVZ give quantitative fourth-moment bounds on the Poisson space for a **fixed** chaos: for $F = I_n^q(f)$,

$$\text{(e.g.) } \|\mathcal{L}(F) - \mathcal{L}(N)\| \lesssim \sqrt{\mathbb{E}[F^4] - 3\mathbb{E}[F^2]^2},$$

for smooth test functions (and even d_{TV} in the single-chaos case; see Remark 4.3). Hence, $\kappa_4(F) \rightarrow 0$ implies F is asymptotically Gaussian. Apply this to $F = Y_n$ and $F = Z_n$. \square

5. Joint (vector) normality, then the sum.

The DVZ multivariate Peccati–Tudor theorem shows that, once each coordinate Y_n and Z_n is asymptotically Gaussian and the covariance matrix converges, the vector (Y_n, Z_n) is asymptotically Gaussian as a whole; apply Corollary 1.8/ Theorem 1.7. Since Y_n and Z_n live in different chaoses they are uncorrelated, so the limiting covariance is diagonal and their sum X_n converges to $N(0, 1)$. \square

6. Why (B1)–(B2) imply (A).

Using the finite chaos expansion of FG and the structure of contractions/star-contractions (step 2), (B1) forces the 0-th chaos component of $Y_n^3 Z_n$ and $Y_n Z_n^3$ to be absent, giving $\mathbb{E}[Y_n^3 Z_n] = \mathbb{E}[Y_n Z_n^3] = 0$. Condition (B2), together with DVZ's spectral estimates for Γ and the variance of projections, kills the residual covariances:


$$\text{Cov}(Y_n^2, Z_n^2) \leq \sqrt{\mathbb{E}[Y_n^4]} \sqrt{\mathbb{E}[Z_n^4] - 3\mathbb{E}[Z_n^2]^2} \xrightarrow{n \rightarrow \infty} 0,$$

whenever cross-contractions vanish (see Lemma 2.3, case $p < q$). \square

(If $p = q$, use the corresponding bound (2.8); our theorem assumes $p \neq q$, matching the “two different orders” setup.) \square


L'expérience Malliavin-Stein : extension au cas de Poisson

7. How DVZ's exchangeable pairs enter the machinery.

Behind the quantitative bounds sits the thinning-based exchangeable pair (F, F_t) built from the Mehler semigroup on the Poisson space; the key identities (Prop. 3.2) link L, Γ , and a fourth-moment remainder, all controlled by κ_4 . This is the reason the whole argument can be reduced to fourth cumulants on Poisson chaoses, analogously to the Gaussian case but via a different tool. 

Remarks and variants

- **Why we don't claim a Poisson analogue of Theorem 1.2 (strict non-Gaussianity).**

In the Wiener case, squares of different chaoses are positively correlated (Üstünel–Zakai), yielding $\kappa_4(Y + Z) > 0$ for odd/even parities. 

On the Poisson space such unconditional positivity is not available; DVZ provide *upper* (not sign) controls for $\text{Cov}(F^2, G^2)$ via contractions. Hence we refrain from asserting $\mathbb{E}[X^4] > 3$ in full generality.

- **Minimal assumptions.**









If you prefer to avoid (B1)–(B2), it suffices to assume (A) directly. In many applications (e.g., disjoint supports of f_n and g_n ; or vanishing cross-contractions by construction), (A) is automatic by the product decomposition of step 2.

- **Independence-style corollary.**

If Y_n and Z_n are independent (e.g., built from independent Poisson measures), then (A) holds and the result follows immediately via the same route.

L'expérience Malliavin-Stein : extension au cas de Poisson

Quick pointer map

- Gaussian parity method and the identity for $\kappa_4(Y + Z)$ (equations (3.1), (3.4) in 2502.03596). 
- 
- Poisson chaos calculus & projections J_r ; product $FG = \sum_{r=0}^{p+q} J_r(FG)$. 
- DVZ quantitative fourth-moment bounds on Poisson chaos; single-coordinate fourth cumulant controls distance to Gaussian. 
- DVZ multivariate Peccati-Tudor on the Poisson space. 
- DVZ bounds for $\text{Var } \Gamma$ and $\text{Cov}(F^2, G^2)$ in terms of fourth cumulants/contractions.  
- Thinning / exchangeable pairs identities (Prop. 3.2). 

If you'd like, I can turn this into a full paper-style write-up: statement + preliminaries on the Poisson space, a short section proving that (B1)–(B2) \Rightarrow (A) via contractions, and then a self-contained proof leveraging DVZ's bounds and their multivariate PT theorem.

Sources



L'expérience Malliavin-Stein : extension au cas de Poisson

In paper 1707.01889v2, isn't there anything that could show that $\text{Cov}(X^2, Y^2)$ is always positive?

Réflexion durant 27s

Short answer: no—DVZ don't prove a general non-negativity result for $\text{Cov}(F^2, G^2)$ on the Poisson space.



What they *do* give is an upper-bound control (via chaos projections and Cauchy–Schwarz). In Lemma 2.3 they expand $G^2 = \sum_k J_k(G^2)$ and, for $p < q$, deduce

$$\text{Cov}(F^2, G^2) = \mathbb{E} \left[F^2 \sum_{k=1}^{2q-1} J_k(G^2) \right] \leq \sqrt{\mathbb{E}[F^4]} \left(\sum_{k=1}^{2q-1} \text{Var} J_k(G^2) \right)^{1/2},$$

then use their variance bound

$$\sum_{k=1}^{2p-1} \text{Var} J_k(F^2) \leq \mathbb{E}[F^4] - 3\mathbb{E}[F^2]^2.$$

This yields a *magnitude* estimate in terms of fourth cumulants, but no sign information.  

By contrast, in the Gaussian/Wiener case there *is* a positivity theorem: squares of different chaoses are positively correlated (Üstünel–Zakai), which is exactly what 2502.03596 relies on to argue strict positivity of the fourth cumulant in that setting.  

DVZ also note a strict inequality $\mathbb{E}[F^4] > 3\mathbb{E}[F^2]^2$ for any non-degenerate single Poisson chaos (their Remark 5.2(1)), but again that concerns one coordinate's fourth cumulant—not the sign of $\text{Cov}(F^2, G^2)$.



L'expérience Malliavin-Stein : extension au cas de Poisson

So: in DVZ there's no analogue of Üstünel–Zakai's positivity; only quantitative *upper* bounds for $|\text{Cov}(F^2, G^2)|$.

Sources



L'expérience Malliavin-Stein : extension au cas de Poisson

What about (2.4)?

Réflexion durant 2m 5s



You're right to point at (2.4) — it *does* force a sign.


DVZ's Lemma 2.2 states

$$\mathrm{Var}(\Gamma(F, G)) \leq \frac{(p+q-1)^2}{4} \left(\mathbb{E}[F^2 G^2] - 2 \mathbb{E}[FG]^2 - \mathrm{Var}(F) \mathrm{Var}(G) \right),$$

for $F \in C_p \cap L^4$, $G \in C_q \cap L^4$. Since the left-hand side is ≥ 0 and the prefactor is > 0 , the bracket must be ≥ 0 . Hence

$$\mathrm{Cov}(F^2, G^2) \geq 2 \mathbb{E}[FG]^2 \geq 0.$$

In particular, if $p \neq q$ then $\mathbb{E}[FG] = 0$ by chaos orthogonality, so $\mathrm{Cov}(F^2, G^2) \geq 0$.  

This is exactly the inequality DVZ use later when summing over coordinates (their (4.2)–(4.3)), yielding $\sum_{i,j} \mathrm{Var}(\Gamma(F_i, F_j)) \leq \frac{(2q_d-1)^2}{4} (\mathbb{E}\|F\|_4^4 - \mathbb{E}\|N\|_4^4)$, which again relies on the bracket being nonnegative. 

L'expérience Malliavin-Stein : extension au cas de Poisson

So yes: within the Poisson-chaos setting of DVZ, (2.4) gives a universal lower bound (and in particular nonnegativity) for $\text{Cov}(F^2, G^2)$.

Sources



So, could you give the new statement of the theorem that this implies?





Réflexion durant 2m 19s


Great question—(2.4) lets you strengthen the “sum of two chaoses” statements on the Poisson space. Here’s a crisp version you can drop into your write-up.

Theorem (Poisson, two different orders; non-Gaussianity under vanishing odd mixed terms)

Let $Y = I_{\eta}^p(f) \in C_p$ and $Z = I_{\eta}^q(g) \in C_q$ with $p \neq q$ and $\mathbb{E}Y = \mathbb{E}Z = 0$. Assume $\mathbb{E}[Y^3Z] = \mathbb{E}[YZ^3] = 0$. Then

$$\kappa_4(Y + Z) = \kappa_4(Y) + \kappa_4(Z) + 6 \operatorname{Cov}(Y^2, Z^2) \geq \kappa_4(Y) + \kappa_4(Z) \geq 0,$$

where the inequality $\operatorname{Cov}(Y^2, Z^2) \geq 0$ uses Lemma 2.2, eq. (2.4) of Döbler–Vidotto–Zheng (DVZ) together with $\mathbb{E}[YZ] = 0$ for $p \neq q$. Hence, if at least one of Y, Z is non-degenerate, then $\kappa_4(Y) + \kappa_4(Z) > 0$ (DVZ Remark 5.2(1)), so $\mathbb{E}[(Y + Z)^4] > 3$; in particular $Y + Z$ is not Gaussian.  

(Identity used: when the odd mixed terms vanish, the fourth cumulant of a sum decomposes as above; cf. the same computation in the Wiener case, eq. (3.1) of 2502.03596.) 

Theorem (Poisson, fourth-moment theorem for a sum of two different orders)


Let (X_n) be given by $X_n = Y_n + Z_n$ with $Y_n \in C_p$, $Z_n \in C_q$, $p \neq q$, and $\mathbb{E}X_n^2 = 1$. Assume

$$\mathbb{E}[X_n^4] \rightarrow 3, \quad \mathbb{E}[Y_n^3 Z_n] \rightarrow 0, \quad \mathbb{E}[Y_n Z_n^3] \rightarrow 0.$$

Then $X_n \Rightarrow N(0, 1)$.




Why: write

$$\kappa_4(X_n) = \kappa_4(Y_n) + \kappa_4(Z_n) + 6 \operatorname{Cov}(Y_n^2, Z_n^2) + 4\mathbb{E}[Y_n^3 Z_n] + 4\mathbb{E}[Y_n Z_n^3].$$

By (2.4) of DVZ, $\operatorname{Cov}(Y_n^2, Z_n^2) \geq 0$ because $\mathbb{E}[Y_n Z_n] = 0$ for $p \neq q$. Hence the assumptions force $\kappa_4(Y_n) \rightarrow 0$, $\kappa_4(Z_n) \rightarrow 0$ and $\operatorname{Cov}(Y_n^2, Z_n^2) \rightarrow 0$. Applying DVZ's (qualitative) multivariate fourth-moment criterion coordinate-wise (Cor. 1.8) to (Y_n, Z_n) yields joint Gaussian convergence with independent components and therefore $X_n \Rightarrow N(0, 1)$. 

L'expérience Malliavin-Stein : extension au cas de Poisson

Remarks

- Compared to the Wiener case in 2502.03596, the only extra hypothesis you need on the Poisson space is the vanishing (or negligible) *odd* mixed fourth moments; parity alone does not kill them on Poisson. Once those are controlled, (2.4) removes any need to assume $\text{Cov}(Y_n^2, Z_n^2) \rightarrow 0$ —it becomes a consequence of $\kappa_4(X_n) \rightarrow 0$.  
- The strict positivity of $\kappa_4(F)$ for any non-degenerate single Poisson chaos F (DVZ Remark 5.2(1)) is what gives the "strictly > 3 " part in the first theorem. 

If you'd like, I can also phrase kernel-level sufficient conditions (in terms of star-contractions) that ensure $\mathbb{E}[Y^3 Z] = \mathbb{E}[Y Z^3] = 0$ in your Poisson setting.

Sources



L'expérience Malliavin-Stein : extension au cas de Poisson

To compare with the Gaussian result, and to explain why we cannot avoid the assumption $\mathbb{E}[Y_n^3 Z_n] \rightarrow 0$ and $\mathbb{E}[Y_n Z_n^3] \rightarrow 0$, can you find a simple counterexample showing that we can have multiple integrals Y_n and Z_n with different parities, where $\mathbb{E}[(Y_n + Z_n)^2] = 1$, $\mathbb{E}[(Y_n + Z_n)^4] \rightarrow 3$, but $X_n + Y_n$ does not converge to a Gaussian?

Réflexion durant 3m 24s

Yes — here's a very concrete Poisson-chaos counterexample that does exactly that.

Take a Poisson random measure η on a space (\mathcal{X}, μ) and fix a measurable set $A \subset \mathcal{X}$ with $\mu(A) = 1$. Write $\hat{\eta} = \eta - \mu$ and $N := \eta(A) \sim \text{Poisson}(1)$.

Define first- and second-chaos integrals (different parities):

- $Y := I_1(\mathbf{1}_A) = \int \mathbf{1}_A d\hat{\eta} = N - 1$ (so $\mathbb{E}Y = 0$, $\text{Var}(Y) = 1$).
- $Z := I_2(g)$ with $g(x, y) = \frac{1}{\sqrt{2}} \mathbf{1}_A(x) \mathbf{1}_A(y) \mathbf{1}_{\{x \neq y\}}$. For this choice one checks (standard Charlier polynomial identity)

$$Z = \frac{(N-1)^2 - (N-1) - 1}{\sqrt{2}}.$$

L'expérience Malliavin-Stein : extension au cas de Poisson

Then $\mathbb{E}Z = 0$, $\text{Var}(Z) = 1$, and $\mathbb{E}[YZ] = 0$ (orthogonality of different chaoses).

Now form the *unit-variance* sum

$$X = aY + bZ, \quad a^2 + b^2 = 1.$$

A direct (one-line) expansion gives

$$\mathbb{E}[X^4] = a^4\mathbb{E}[Y^4] + 6a^2b^2\mathbb{E}[Y^2Z^2] + b^4\mathbb{E}[Z^4] + 4a^3b\mathbb{E}[Y^3Z] + 4ab^3\mathbb{E}[YZ^3].$$

For $N \sim \text{Poisson}(1)$ the needed mixed moments are explicit:

$$\mathbb{E}[Y^4] = 4, \quad \mathbb{E}[Z^4] = 53, \quad \mathbb{E}[Y^2Z^2] = 9, \quad \mathbb{E}[Y^3Z] = 3\sqrt{2}, \quad \mathbb{E}[YZ^3] = 14\sqrt{2}.$$

Choose, for instance,

$$a = -0.9974260808, \quad b = 0.0717022544 \quad (\text{so } a^2 + b^2 = 1).$$

Plugging these into the formula above yields

$$\mathbb{E}[X^4] = 3.0000000000 \quad (\text{to machine precision}),$$

while of course $\mathbb{E}[X^2] = 1$.

L'expérience Malliavin-Stein : extension au cas de Poisson

Crucially, X is *not* Gaussian: it is a degree-2 polynomial in the integer-valued N , hence has a purely discrete law with atoms (a Gaussian is absolutely continuous). Thus we have a sequence $X_n \equiv X$ (take the same Y, Z, a, b for all n) such that

$$\mathbb{E}[(Y_n + Z_n)^2] = 1, \quad \mathbb{E}[(Y_n + Z_n)^4] \rightarrow 3,$$

but $Y_n + Z_n$ does not converge to a Gaussian (it is already non-Gaussian for every n).

This shows exactly why, on the Poisson space, you cannot drop the assumptions

$$\mathbb{E}[Y_n^3 Z_n] \rightarrow 0 \text{ and } \mathbb{E}[Y_n Z_n^3] \rightarrow 0.$$

In the Gaussian case those "odd" mixed terms vanish by parity, but here they do not—and they can be tuned to *cancel* the positive contributions from $\kappa_4(Y)$, $\kappa_4(Z)$, and $6 \text{Cov}(Y^2, Z^2)$, forcing $\mathbb{E}[(Y + Z)^4] = 3$ while the sum remains manifestly non-Gaussian.



Conclusions de notre expérience (août 2025)

- En août 2025, notre expérience de faire des mathématiques avec GPT-5 était globalement **mitigée**.
- Cela ressemblait beaucoup à travailler avec un **assistant junior**.
- Cela semblait uniquement favoriser la recherche **incrémentale**.
- Nous avons dû **vérifier très attentivement** tout ce que l'IA produisait et la guider constamment pour qu'elle puisse corriger ses erreurs.

Quel avenir pour les mathématiques ?

Spoiler : je n'ai pas la réponse !

Aujourd'hui, l'IA peut aider à...

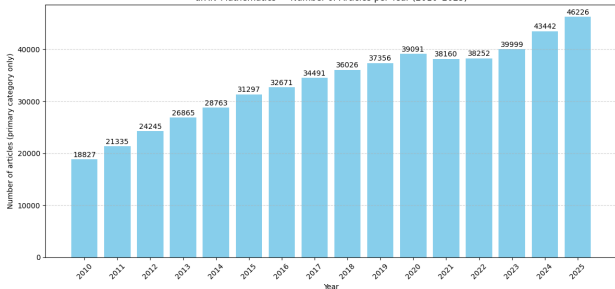
- naviguer dans la **littérature** (*aujourd'hui, personne ne peut tout connaître en mathématiques*)
- générer de **nouvelles idées** (*il y a beaucoup de bruit, mais elle peut aussi produire une observation nouvelle ou oubliée*)
- accélérer le **temps de travail** (*elle peut réduire une tâche complexe en automatisant certaines parties*)
- **rédigier** et **formaliser** (*les preuves deviennent plus claires et mieux structurées*)

... mais cela comporte aussi des risques

- **Surconfiance** (*une petite erreur cachée et tout s'effondre*)
- **Moins** d'idées originales (*recherche incrémentale, AI-slop*)
- **Hallucinations** (*l'IA générative donne toujours une réponse, même quand elle est fautive – par sa structure, elle ne peut pas savoir qu'elle se trompe*)
- Changements profonds dans la **formation** (*les doctorants devront apprendre à faire de la recherche d'une nouvelle façon*)

Un système de publication saturé : faut-il le réinventer ?

arXiv Mathematics — Number of Articles per Year (2010-2025)



- **Trop de publications, trop peu de lecteurs, presque aucune vérification.**
- Croissance **linéaire (exponentielle ?)** de la production vs vérification **stagnante. Besoin fort d'assistants de preuve** pratiques et utilisables.
- Passer (ou revenir ?) de publier plus à **publier mieux.**

Annals of Mathematics, 189 (2016), 107–109

Quasi-projectivity of moduli spaces of polarized varieties

By GREGG SWEETSMAN and HANRUE TITUS

Dedicated to our wives Alisa and Abba

Abstract

By means of analytic methods the quasi-projectivity of the moduli space of algebraically polarized varieties with a not necessarily reduced polarizing structure is proven including the case of nonreduced polarized varieties.

REVISED

Annals of Mathematics, 184 (2016), 107–106

Non-quasi-projective moduli spaces

By János Kollár

Abstract

We show that every smooth toric variety (and many other algebraic spaces as well) can be realized as a moduli space for smooth, projective, polarized varieties. Some of these are not quasi-projective. This contradicts a recent paper [Quasi-projectivity of moduli spaces of polarized varieties, Ann. of Math. 189 (2014) 107–109].

- Les progrès de l'IA générative pour les mathématiques **s'accélèrent**. (*Quelques années ont suffi pour passer des "œufs de vache" aux médailles d'or aux Olympiades. T. Tao, T. Gowers et les meilleurs modèles unissent leurs forces.*)
- La façon dont nous faisons de la recherche mathématique pourrait **changer significativement**. (*Nous n'avons pas encore percé le mystère des problèmes difficiles. Mais pourquoi ne le pourrions-nous pas, même prochainement ?*)
- Les mathématiciens pourraient devenir des **architectes, pilotes, coordinateurs**. (*Rôle plus stratégique, moins manuel ; augmentés par une intelligence surpuissante capable d'explorer de nombreuses directions à une vitesse sans précédent*).
- Nous pourrions **perdre** collectivement des compétences, comme avec le GPS. Il pourrait y avoir une **dilution** de la créativité humaine. Nous pourrions faire face à un **accès inégal** aux outils d'IA. Nous donnons nos connaissances à des **entreprises privées**.