# Approximate computation of expectations: A canonical Stein's density approach

Christophe Ley, Gesine Reinert and Yvik Swan*

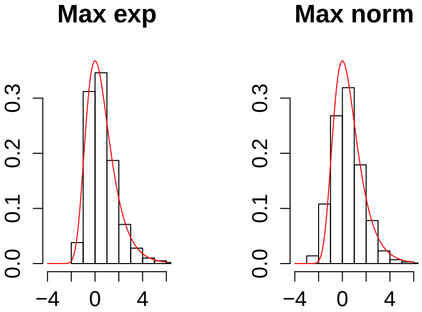
Université de Liège, Belgium

May 22, 2015

# Prologue

Consider two random objects $W$ and $Z$ and suppose that

$$\mathcal{L}(W) \approx \mathcal{L}(Z)$$




There are many ways of measuring "closeness" depending on what features one wishes to capture (see e.g. Gibbs and Su 2002).

We will focus on distances of the form

$$d_{\mathcal{H}}(W,Z) = \sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)|$$

with $\mathcal{H}$ some measure generating class.

This includes

$$\begin{aligned} TV(W,Z) &= \sup_{A\subset\mathbb{R}} |P(W\in A) - P(Z\in B)| \\ \mathcal{W}(W,Z) &= \sup_{h\in Lip(1)} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \\ \mathsf{Kol}(W,Z) &= \sup_{z\in\mathbb{R}^d} |P(W\le z) - P(Z\le z)| \end{aligned}$$

and several more.

Even in the simplest cases we can rarely assess $d_{\mathcal{H}}(W, Z)$ explicitely.

**Objective** : provide good/precise/computable bounds

$$L_1 \leq d_{\mathcal{H}}(W, Z) \leq L_2.$$

The bounds are

$$L_i = L_i(\mathcal{H}, W, Z, *), \quad i = 1, 2$$

# The Chen-Stein method

## The Chen-Stein method

Step 1 : an operator

Step 2 : a Stein equation

Step 3 : a transfer principle

Suppose that we dispose of a linear operator $f(x) \mapsto \mathcal{A}_Z f(x)$ such that

$$W \stackrel{\mathcal{L}}{=} Z \text{ if and only if } \mathbb{E}\left[\mathcal{A}_Z f(W)\right] = 0 \qquad (*)$$

for all functions $f \in \mathcal{F}(\mathcal{A}_Z)$ some class of functions.

If $(*)$ holds we say that $(\mathcal{A}_Z, \mathcal{F}(\mathcal{A}_Z))$ characterizes the law of $Z$.

Such pairs are not unique.

In the sequel we concentrate on "Stein operators".

**Gaussian operators :**

- $\mathcal{A}_Z f(x) = f'(x) - xf(x)$ (Standard Stein operator)
- $\mathcal{A}_Z f(x) = f^{(m)}(x) - H_m(x)f(x)$ with $H_m(\cdot)$ the $m$th Hermite polynomial (Goldstein Reinert 2005)
- $\mathcal{A}_Z f(x) = (1+x^2)f'(x) - (x^3 - x)f(x)$ (Ley and Swan 2015)

**Poisson operators :**

- $\mathcal{A}_Z f(x) = \lambda f(x+1) - xf(x)$
- $\mathcal{A}_Z f(x) = C_\lambda^m(x)f(x) - \alpha\Delta^m f(x)$ with $C_\lambda^m(x)$ the $m$th Charlier polynomial and $\Delta$ the forward difference operator (Goldstein Reinert 2005)

Many operators are now known for many distributions :

Exponential, chi-squared, gamma, Semi-circle, Variance gamma, Marchenko Pastur law, Multinomial, Beta distribution, Geometric, Compound geometric, Binomial, Negative binomial, Multivariate normal, Elliptical distributions (multivariate), Extreme distributions, Laplace distribution, Distributions involving hypergeometric functions, Fragility distributions, Half normal distribution, ArcSine distribution, Rank distribution of random matrices, The Conway-Maxwell-Poisson distribution, Symmetric $\alpha$-Stable distributions.

See e.g.

https ://sites.google.com/site/yvikswan/about-stein-s-method

for a list of references.

## The Chen-Stein method

Step 1 : an operator

Step 2 : a Stein equation

Step 3 : a transfer principle

Given $Z$ with characterizing $(\mathcal{A}_Z, \mathcal{F}(\mathcal{A}_Z))$ suppose that for all $h \in \mathcal{H}$ there exists a unique $f_h \in \mathcal{F}(\mathcal{A}_Z)$ such that

$$\mathcal{A}_Z f_h(x) = h(x) - \mathbb{E}h(Z)$$

Let $\mathcal{F}(Z, \mathcal{H})$ be the collection of such $f_h$.

We call such equations *Stein equations*.

## The Chen-Stein method

Step 1 : an operator

Step 2 : a Stein equation

Step 3 : a transfer principle

Then

$$\begin{aligned} d_{\mathcal{H}}(W, Z) &= \sup_{h \in \mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \\ &= \sup_{f \in \mathcal{F}(Z, \mathcal{H})} |\mathbb{E}[\mathcal{A}_Z f(W)]| \end{aligned}$$

As discovered by Stein (in Gaussian context) and Chen (in Poisson context) the quantity on the rhs is a good starting point for assessing the lhs as long as the pair $(\mathcal{A}_Z, \mathcal{F}(\mathcal{A}_Z))$ is well chosen.

What makes for a good pair $(\mathcal{A}_Z, \mathcal{F}(\mathcal{A}_Z))$ ?

Three criteria :

- $\mathcal{F}(\mathcal{A}_Z)$ needs to be large and easy to describe ;
- The functions in $\mathcal{F}(Z, \mathcal{H})$ need to have good properties (e.g. bounded with bounded derivatives) ;
- The object "$\mathbb{E}\mathcal{A}_Z f(W)$" needs to provide good handles.

In all cases cited these three criterion are met in spectacular fashion.

- After a little thought the operator $\mathcal{A}_Z$ has a nice form and the class $\mathcal{F}(\mathcal{A}_Z)$ always ends up explicit.
- Precise bounds are known on the solutions.
- There are an uncanny number of ways to tackle $\mathbb{E}\left[\mathcal{A}_Z f(W)\right]$

The literature is immense and outreach of *Stein's method* is tentacular.

See, e.g., Stein (1986), Barbour and Chen (2005), Nourdin Peccati (2011), Chen Goldstein Shao (2011), Ross (2011) or

$$https://sites.google.com/site/malliavinstein/home$$

# Comparison of operators

Let $Z$ and $W$ have pair $(\mathcal{A}_Z, \mathcal{F}(\mathcal{A}_Z))$ and $(\mathcal{A}_W, \mathcal{F}(\mathcal{A}_W))$, respectively.

Suppose that

$$\mathcal{F}(Z, \mathcal{H}) \subset \mathcal{F}(\mathcal{A}_W).$$

Then

$$\begin{aligned} \mathbb{E}[h(W)] - \mathbb{E}[h(Z)] &= \mathbb{E}[\mathcal{A}_Z f(W)] \\ &= \mathbb{E}[\mathcal{A}_Z f(W)] - \mathbb{E}[\mathcal{A}_W f(W)] \end{aligned}$$

for all $h \in \mathcal{H}$ and

$$d_{\mathcal{H}}(W, Z) = \sup_{f \in \mathcal{F}(Z,\mathcal{H})} |\mathbb{E}[(\mathcal{A}_Z - \mathcal{A}_W) f(W)]| .$$

Thus assessing the distance between the laws of $Z$ and $W$ boils down to assessing the difference between their operators.

# Comparison of operators

Comparing scores

Comparing Stein kernels

Let $X \sim p_X$ and define

$$\rho_X(x) = \frac{p'_X(x)}{p_X(x)} = (\log p_X(x))'$$

the score function of $X$.

Introduce operators of the form

$$\mathcal{A}_X f(x) = \frac{(f(x)p_X(x))'}{p_X(x)} = f'(x) + \frac{p'_X(x)}{p_X(x)} f(x).$$

Sufficient assumptions on $X$ can be identified under which $(\mathcal{A}_X, \mathcal{F}(\mathcal{A}_X))$ characterizes $X$ (see Stein et al. 2004 or Chatterjee and Shao 2011).

Then using

$$\begin{aligned}\mathcal{A}_Z f(x) &= f'(x) - \rho_Z(x) f(x)\\ \mathcal{A}_W f(x) &= f'(x) - \rho_W(x) f(x)\end{aligned}$$

we get [if reasonable assumptions are satisfied]

$$\begin{aligned}d_{\mathcal{H}}(W,Z) &= \sup_{f\in\mathcal{F}(Z,\mathcal{H})} \left|\mathbb{E}\left[(\mathcal{A}_Z - \mathcal{A}_W) f(W)\right]\right|\\ &= \sup_{f\in\mathcal{F}(Z,\mathcal{H})} \left|\mathbb{E}(\rho_W(W) - \rho_Z(W)) f(W)\right|\end{aligned}$$

Need to bound this quantity.

There are two ways of bounding the rhs :

- Technique 1 :

$$d_{\mathcal{H}}(W,Z) \leq \kappa_{Z,\mathcal{H}} \mathbb{E}\left|\rho_W(W) - \rho_Z(W)\right|$$

  with

$$\kappa_{1,Z,\mathcal{H},W} = \sup_{f \in \mathcal{F}(Z,\mathcal{H})} \|f\|$$

- Technique 2 :

$$d_{\mathcal{H}}(W,Z) \leq \kappa_{Z,\mathcal{H}} \sqrt{\mathbb{E}\left(\rho_W(W) - \rho_Z(W)\right)^2}$$

  with

$$\kappa_{2,Z,\mathcal{H},W} = \sqrt{\sup_{f \in \mathcal{F}(Z,\mathcal{H})} \left|\mathbb{E}\left[f(W)^2\right]\right|}.$$

There are many ways to bound the constant.

The constants $\kappa_{1,Z,\mathcal{H},W}$ are the "magic factors" and

$$\kappa_{2,Z,\mathcal{H},W} \leq \kappa_{1,Z,\mathcal{H},W}$$

Better bounds can sometimes be obtained.

**Proposition (Ley and S. 2013)** Take

$$Z \sim p_Z(x) = ce^{-d|x|^\alpha} \mathbb{I}_S(x)$$

for some $\alpha$ and $S$ scale invariant subset of $\mathbb{R}$. Suppose that $W$ has support included in $S$. Then

$$\kappa_{Z,\mathcal{H},W} \leq \frac{\|h\|_\infty}{2^{1/\alpha}}.$$

The quantity

$$\mathbb{E}\left|\rho_W(W)-\rho_Z(W)\right| \text{ and } \mathbb{E}\left(\rho_W(W)-\rho_Z(W)\right)^2$$

have, in many cases, good properties.

- See e.g. Shimizu (1975) and Stein (1986, Lesson 6) and many others.
- See particularly Johnson and Barron (2004) for a study of

$$\mathcal{J}(W,Z)=\mathbb{E}\left(\rho_W(W)-\rho_Z(W)\right)^2$$

  the so-called FID between $W$ and $Z$.

# Comparison of operators

Comparing scores

Comparing Stein kernels

Let $X \sim p_X$ with mean $\mu$ and consider $\tau_X(x)$ such that

$$\frac{(\tau_X(x)p_X(x))'}{p_X(x)} = \mu - x$$

(called the Stein kernel/factor/coefficient of $X$).

Introduce operators of the form

$$\mathcal{A}_X f(x) = \frac{(\tau_X(x)f(x)p_X(x))'}{p_X(x)} = \tau_X(x)f'(x) + (\mu - x)f(x)$$

Sufficient assumptions on $X$ can be identified under which $(\mathcal{A}_X, \mathcal{F}(\mathcal{A}_X))$ characterizes $X$ (see Döbler (2012) or Tudor and Kusuoka (2012, 2014)).

Suppose $X$ and $Z$ have the same mean $\mu$. Then using

$$\begin{aligned} \mathcal{A}_Z f(x) &= \tau_Z(x) f'(x) + (\mu - x) f(x) \\ \mathcal{A}_W f(x) &= \tau_W(x) f'(x) + (\mu - x) f(x) \end{aligned}$$

we get [if assumptions are satisfied]

$$\begin{aligned} d_{\mathcal{H}}(W, Z) &= \sup_{f \in \mathcal{F}(Z,\mathcal{H})} \left| \mathbb{E}\left[ (\mathcal{A}_Z - \mathcal{A}_W) f(W) \right] \right| \\ &= \sup_{f \in \mathcal{F}(Z,\mathcal{H})} \left| \mathbb{E}(\tau_W(W) - \tau_Z(W)) f'(W) \right| \end{aligned}$$

with a similar discussion on the constant depending on

$$\sup_{f \in \mathcal{F}(Z,\mathcal{H})} \|f'\| \text{ or } \sqrt{\sup_{f \in \mathcal{F}(Z,\mathcal{H})} E\left[f'(W)^2\right]}$$

The quantities

$$\mathbb{E}\left|\tau_W(W) - \tau_Z(W)\right| \text{ and } \mathbb{E}\left(\tau_W(W) - \tau_Z(W)\right)^2$$

have, in certain cases, great properties.

- Several authors use this approach, e.g. Stein (1986, Lesson 6), Cacoullos et al. (1992), Döbler (2012) or Tudor and Kusuoka (2012, 2014).
- An important area is in the continuation of the field initiated by Nourdin and Peccati (2009).
- See also Ledoux, Nourdin and Peccati (2015) for a study of

$$\mathcal{S}(W, Z) = \mathbb{E}\left(\tau_W(W) - \tau_Z(W)\right)^2$$

the so-called *Stein discrepancy* between $W$ and $Z$.

Similar constructions can also be done in the discrete case (see Ley and S. 2013).

In many cases, however, the score $\rho$ or Stein kernel $\tau$ do not bear good properties.

For example in the exponential, Laplace, $\alpha$-stable, Kummer-$U$, Variance-Gamma, ...

then one needs second order or weirder operators.

**Question :** does there exist a systematic way of constructing Stein operators ?

## A canonical Stein density approach

# A canonical Stein density approach

Let $(\mathcal{X}, \mathcal{B}, \mu)$ be a measure space and take a linear operator

$$\mathcal{D} : dom(\mathcal{D}) \subset \mathcal{X}^{\star} \to im(\mathcal{D})$$

such that $dom(\mathcal{D}) \setminus \{0\} \neq \emptyset$.

**Example 1 :** $\mathcal{D}f(\cdot) = f'(\cdot)$

**Example 2 :** $\mathcal{D}f(\cdot) = \Delta^{+} f(\cdot) = f(\cdot + 1) - f(\cdot)$

**Assumption (product formula)** There exists a linear operator $\mathcal{D}^\star$ : $dom(\mathcal{D}^\star) \subset \mathcal{X}^\star \to im(\mathcal{D}^\star)$ and a constant $l := l_{\mathcal{X},\mathcal{D}}$ such that

$$\mathcal{D}(f(x)g(x+l)) = g(x)\mathcal{D}f(x) + f(x)\mathcal{D}^\star g(x)$$

for all $(f,g) \in dom(\mathcal{D}) \times dom(\mathcal{D}^\star)$.

**Example 1 :**

$$(fg)' = f'g + fg'$$

**Example 2 :**

$$\Delta^+(f(x)g(x-1)) = \Delta^+ f(x)g(x) + f(x)\Delta^- g(x)$$

# A canonical Stein density approach

Let $X$ have density $p$ with respect to $\mu$.

**Definition 1** The *(canonical) Stein class* $\mathcal{F}(p) \equiv \mathcal{F}(X)$ for $X$ is the collection of $p$-integrable functions $f$ such that

- $fp \in dom(\mathcal{D})$,
- $\mathcal{D}(fp)$ is integrable, and
- $\int \mathcal{D}(fp)\mathrm{d}\mu = 0$.

**Definition 2** The *(canonical) Stein operator* $\mathcal{T}_p \equiv \mathcal{T}_X$ for $p$ is the linear operator on $\mathcal{F}(X)$ defined as

$$\mathcal{T}_X f = \frac{\mathcal{D}(fp)}{p}.$$

Note how $E[\mathcal{T}_X f(X)] = 0$ for all $f \in \mathcal{F}(X)$.

## A canonical Stein density approach

**Assumption (invertibility)** Suppose that there exists $\mathcal{D}^{-1} : im(\mathcal{D}) \to dom(\mathcal{D})$ a linear operator such that

$$\mathcal{D}\left(\mathcal{D}^{-1}h\right) = h$$

for all $h \in im(\mathcal{D})$ and, for $f \in dom(\mathcal{D})$,

$$\mathcal{D}^{-1}\left(\mathcal{D}f\right)$$

is defined up to addition with an element of $ker(\mathcal{D})$.

**Example 1 :** $\mathcal{D}^{-1}f(\cdot) = \int^{\cdot} f$

**Example 2 :** $(\Delta^{+})^{-1}f(\cdot) = \sum^{\cdot} f$

**Definition 3** The *(canonical) inverse Stein operator* is

$$\mathcal{T}_p^{-1}(h) = \frac{\mathcal{D}^{-1}(hp)}{p}$$

**Proposition (Döbler 12, Ley, Reinert, S. 15)** Let $X$ have mean $\mu$ and density $p$ with respect to the Lebesgue measure satisfying [assumptions]. Let

$$g_h = \frac{\mathcal{T}_X^{-1}(h - \mathbb{E}\left[h(Z)\right])}{\mathcal{T}_X^{-1}(\mu - Id)}.$$

Then

$$\|g_h\| \leq \|h'\|.$$

# A canonical Stein density approach

In the continuous case : for all $f \in \mathcal{F}(X)$,

$$\begin{aligned}
\mathbb{E}\left[g'(X)f(X)\right] &= \int g'(x)f(x)p(x)dx \\
&= -\int g(x)\frac{(fp)'(x)}{p(x)}p(x)dx \\
&= -\mathbb{E}\left[g(X)\mathcal{T}_X f(X)\right]
\end{aligned}$$

for all differentiable functions $g$ such that

- $\int (gfp)' dx = 0$, and
- $\int |g' fp| dx < \infty$.

We collect all such $g$ in a class $dom((\cdot)', X, f)$.

**Theorem 1 (Ley, Reinert and S. 2015)** For all $f \in \mathcal{F}(X)$

$$\mathbb{E}\left[\mathcal{D}g(X)\, f(X)\right] = -\mathbb{E}\left[g(X)\, \mathcal{T}_X f(X)\right]$$

for all $g \in dom(\mathcal{D}, X, f)$.

**Theorem 2 (Ley, Reinert and S. 2015)** For all $h$ with $p$ mean 0

$$\mathbb{E}\left[\mathcal{T}_X^{-1}h(X)\, \mathcal{D}^{\star}g(X)\right] = -\mathbb{E}\left[g(X)\, h(X)\right]$$

for all $g \in dom(\mathcal{D}, X, \mathcal{T}_X^{-1}h)$.

**Theorem 3 (Ley, Reinert and S. 2015)** Under reasonable assumptions we can show that these relations characterize the law of $X$ either by fixing $f$ or by fixing $g$.

# A canonical Stein density approach

The Stein operator satisfies the product rule

$$\begin{aligned}\mathcal{T}_X(fg(\cdot + l)) &= \frac{\mathcal{D}(fg(\cdot + l)p)}{p} \\ &= f\mathcal{D}^\star g + g\frac{\mathcal{D}(fp)}{p} \\ &= f\mathcal{D}^\star g + g\mathcal{T}_X f\end{aligned}$$

so that

$$\mathcal{T}_X(fg(\cdot + l)) = f\mathcal{D}^\star g + g\mathcal{T}_X f$$

with $f \in \mathcal{F}(X)$ and $g \in dom(\mathcal{D}, X, f)$.

The Stein equation becomes

$$h - \mathbb{E}h(X) = f\mathcal{D}^{\star}g + g\mathcal{T}_X f$$

whose solution $(f, g)$ is a pair of functions $f \in \mathcal{F}(X)$ and $g \in dom(\mathcal{D}, X, f)$.

To solve the Stein equation we can

- fix $f \in \mathcal{F}(X)$ and choose $g$ accordingly
- fix $g \in \bigcap_{f \in \mathcal{F}(X)} dom(\mathcal{D}, X, f)$ and choose $f$ accordingly
- let $f$ and $g$ vary simultaneously.

## Note

All the structure of the problem is hidden within the definition of the classes.

Requiring $f \in \mathcal{F}(X)$ imposes many conditions if $X$ has a complicated distribution.

For example :

- in the exponential case we need $f(0) = 0$ ;
- in the Laplace case we need $f(x)e^{-|x|}$ differentiable at 0.

In practice one seeks to construct operators whose expression does not require any complicated assumptions on the test functions.

**Example 1 :** for $X$ standard normal,

$$\mathcal{A}_X(f,g)(x) = f(x)g'(x) + (f'(x) - xf(x))g(x)$$

and there are virtually no conditions on $f$ or $g$.

- Take $g = 1$ to get

  $$\mathcal{A}_X(f) = f'(x) - xf(x)$$

  with no stringent condition on $f$
- Take $f(x) = H_n(x)$ then

  $$\mathcal{A}_X(g) = H_n(x)g'(x) - H_{n+1}(x)g(x)$$

  with no stringent condition on $g$.

**Example 2 :** for $X$ exponential

$$\mathcal{A}_X(f, g) = f(x)g'(x) + (f'(x) - f(x))g(x)$$

with virtually no conditions on $g$ and $f \in \mathcal{F}(X)$.

- Take $g = 1$ and $f(x) = \tilde{f}(x) - \tilde{f}(0)$ yielding

$$\mathcal{A}_X(f) = \tilde{f}'(x) - \tilde{f}(x) + \tilde{f}(0)$$

  with virtually no condition on $\tilde{f}$.

- Take $f(x) = x$ yielding

$$\mathcal{A}_X(g) = xg'(x) + (1 - x)g(x)$$

  with virtually no condition on $g$.

**Example 3 :** for $X$ Laplace

$$\mathcal{A}_X(f,g) = f(x)g'(x) + (f'(x) - \mathsf{sign}(x)f(x))g(x)$$

with virtually no conditions on $g$ and $f \in \mathcal{F}(X)$.

- Take $g = 1$ and $f(x) = \frac{(x\tilde{f}(x)e^{|x|})'}{e^{|x|}}$ yielding

  $$\mathcal{A}_X(f) = x\tilde{f}''(x) + 2\tilde{f}' - x\tilde{f}$$

  with virtually no condition on $\tilde{f}$.
- Take $f(x) = x$ yielding

  $$\mathcal{A}_X(g) = xg'(x) + (1 - |x|)g(x)$$

  with virtually no condition on $g$.

# A general comparison result

Let $X_1$ and $X_2$ be two random variables with Stein construction $(\mathcal{T}_i, \mathcal{F}_i, dom(\mathcal{D}_i, X_i, f))$, $i = 1, 2$.

**Theorem (Ley, Reinert and S. 2015)** We have

$$\begin{aligned} \mathbb{E}h(X_2) - \mathbb{E}h(X_1) &= \mathbb{E}\left[f(X_2)\left(\mathcal{D}_1^\star g(X_2) - \mathcal{D}_2^\star g(X_2)\right)\right. \\ &\qquad \left. + g(X_2)\left(\mathcal{T}_1 f(X_2) - \mathcal{T}_2 f(X_2)\right)\right]. \end{aligned}$$

for all $(f, g) \in \mathcal{F}_1 \times dom(\mathcal{D}_1, X_1)$ such that $\mathcal{T}_1(fg) = h - \mathbb{E}h(X_1)$.

**Remarks :**

- $\mathcal{D}_1^\star$ and $\mathcal{D}_2^\star$ can be different !
- You can choose to fix $g$ and optimize in $f$ or the other way around
- You can optimize the constants

## Application 1 : Binomial approximation

Let $X_1 \sim \mathrm{Bin}(n,p)$ and $X_2 = \sum_{i=1}^n I_i$ with $I_i \sim \mathrm{Bin}(1,p_i)$, $i = 1,\ldots,n$, i.i.d. and $np = \sum_{i=1}^n p_i$.

It is easy to show

$$\tau_{X_1}(x) = (1-p)x \text{ and } \tau_{I_i}(x) = (1-p_i)x$$

also a variation on the score yields

$$\rho_{X_1}(x) = \frac{np-x}{1-p} \text{ and } \rho_{I_i}(x) = \frac{p_i - x}{1-p_i}$$

Then (Ehm, 1991)

$$|\mathbb{E}h(X) - \mathbb{E}h(W)| \leq \min\left( ||\Delta^+ g_h||_\infty, \frac{2\|g_h\|_\infty}{1-p} \right) \sum_{i=1}^n |p_i - p| p_i$$

## Application 2 : difference between Student and Gauss

Set $X_1 = Z$ standard Gaussian and $X_2 = W_\nu$ a Student $t$ random variable with $\nu > 2$ degrees of freedom.

We have

$$\tau_1 = 1 \text{ and } \tau_2(x) = \frac{x^2+\nu}{\nu-1}$$

as well as

$$\rho_1(x) = -x \text{ and } \rho_2(x) = -\frac{x(1+\nu)}{x^2+\nu}$$

which yields

$$d_{\mathrm{TV}}(Z, W_\nu) \leq \min\left(\frac{4}{\nu-2}, \sqrt{\frac{\pi}{2}}\frac{-2+8\left(\frac{\nu}{1+\nu}\right)^{(1+\nu)/2}}{(\nu-1)\sqrt{\nu}B(\nu/2,1/2)}\right),$$

# Multivariate extension

Consider

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} \mathbf{F}_1(\mathbf{x}) \\ \vdots \\ \mathbf{F}_q(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} F_{11}(\mathbf{x}) & \dots & F_{1d}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ F_{q1}(\mathbf{x}) & \dots & F_{qd}(\mathbf{x}) \end{pmatrix}$$

and define the differential $\mathcal{D} = \text{div}$ through

$$\text{div}(\mathbf{F}) \stackrel{\text{not}}{=} \boldsymbol{\nabla}^T \cdot \mathbf{F} = \begin{pmatrix} \text{div}(\mathbf{F}_1) \\ \vdots \\ \text{div}(\mathbf{F}_q) \end{pmatrix} = \begin{pmatrix} \nabla^T \cdot \mathbf{F}_1 \\ \vdots \\ \nabla^T \cdot \mathbf{F}_q \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{d} \frac{\partial F_{1i}}{\partial x_i} \\ \vdots \\ \sum_{i=1}^{d} \frac{\partial F_{qi}}{\partial x_i} \end{pmatrix}.$$

Then we have the product rule

$$\mathrm{div}(\phi \mathbf{F}) = \phi\, \mathrm{div}(\mathbf{F}) + \mathbf{F}(\nabla \phi).$$

so that

$$\mathcal{D}^{\star} = \nabla$$

and everything follows.

Let $X$ be a $d$-dimensional random vector with probability density function (pdf) $p : \mathbb{R}^d \to \mathbb{R}$ with respect to the Lebesgue measure on $\mathbb{R}^d$. Let $\Omega \subset \mathbb{R}^d$ be the support of $p$ [with assumptions].

**Definition 1** The *(canonical) Stein class* for $X$ is the class $\mathcal{F}(X)$ of all $q \times d$ tensor fields $\mathbf{F}$ (for some $q \geq 1$) for which $p\mathbf{F}$ is

- differentiable,
- $\mathrm{div}(p\mathbf{F})$ is component-wise integrable on $\Omega$
- $\int_\Omega \mathrm{div}(p\mathbf{F}) = 0$.

**Definition 2** The *(canonical) Stein operator* of $X$ is the differential operator

$$\mathcal{T}_X : \mathcal{F}(X) \to (\mathbb{R}^d)^\star : \mathbf{F} \to \mathcal{T}_X \mathbf{F} = \frac{\mathrm{div}(p\mathbf{F})}{p}.$$

## Application : multivariate Gaussian operators

Taking $\mathbf{F} = \mathbf{G}\nabla f$ with $\mathbf{G}$ a symmetric $d \times d$ matrix then

$$\begin{aligned}
\mathcal{A}_X f &= \sum_{i,j=1}^{d} \partial_i (G_{ij}\partial_j f) + \sum_{i,j=1}^{d} G_{ij}\partial_j f \frac{\partial_i p}{p} \\
&= \nabla^t \cdot (\mathbf{G}\nabla f) + \nabla f^t \mathbf{G}\nabla \log p
\end{aligned}$$

In particular, if $p$ is the density of a $\mathcal{N}_d(0, \Sigma)$ random vector then

$$\nabla \log p(x) = -\Sigma^{-1} x$$

so that, taking $\mathbf{G} = \Sigma$,

$$\mathcal{A}_X f(x) = \sum_{i,j=1}^{d} \sigma_{ij}\partial_{ij} f(x) + (\nabla f(x))^t x.$$

Many authors have considered different standardizations of $\mathcal{T}_X F$ :

- Landsman and Neslehova and coauthors (2010 : 2014) in the context of elliptical distributions ;
- Chatterjee and Meckes (2008) and Reinert and Röllin (2009) for multivariate normal
- Brown et al. (2006) in the context of the heat equation
- Peccati et al. (2014 : 2015) specifically via Stein matrices
- Artstein et al. (2004 : 2014) with variational considerations in mind.