A general parametric Stein characterization

Christophe Ley[*] and Yvik Swan[†]

Ghent University and Université de Liège

Abstract

We present a general characterization theorem for parametric probability distributions in terms of a differential operator akin to the so-called Stein operators from the literature on Stein's method.

1 Introduction

Stein's method is a collection of tools allowing to compute sharp rates of convergence in stochastic approximation problems. In short, the method consists of transforming the problem of bounding a suitable distance between two probability measures (specifically distances which can be represented as integral probability metrics) into that of bounding a suitable linear operator over a well-chosen class of functions. The foundations of the method were laid down in Charles Stein's classic paper (Stein, 1972) and book (Stein, 1986). For a more modern overview of the method we refer the reader to the monograph (Chen et al., 2011) or the review article (Ross, 2011).

Contributions to the literature on Stein's method are generally of three types :

- Applications of the method to new settings (e.g. in network analysis (Franceschetti and Meester, 2006), random matrices (Mackey et al., 2014) or even quantum physics (McKeague and Levin, 2015))

- Development of new versions of the method to tackle new specific targets or dependency structures (e.g. Pearson statistics and χ^2 approximation (Gaunt et al., 2015), spin glasses and distributions from statistical physics (Chatterjee and Shao, 2011), Pólya urns and beta distribution (Goldstein and Reinert, 2013))

- Extensions of various aspects of the method to more abstract and general settings (e.g. for Pearson distributions (Afendras et al., 2011), general densities (Ley and Swan, 2013), invariant measures of diffusions (Kusuoka and Tudor, 2013)).

---

[*]Ghent University, Department of Applied Mathematics, Computer Science and Statistics, Campus Sterre, Krijgslaan 281, S9, B-9000 Ghent

[†]Université de Liège, Département de mathématique, zone polytech 1, 12 allée de la découverte, B47, Sart Tilman, B-4000 Liège

The present paper falls within the last category.

Let P be some probability distribution with respect to which one aims to perform Stein's method (that is, P is the target distribution). A Stein characterization for P is a result of the form

$$X \sim P \text{ if and only if } E\left[\mathcal{A}_P f(X)\right] = 0 \text{ for all } f \in \mathcal{F}(\mathcal{A}_P) \tag{1}$$

with $\mathcal{A}_P$ a suitable (usually linear) operator (called Stein operator for P) and $\mathcal{F}(\mathcal{A}_P)$ a well chosen class of functions (called Stein class). There exist many different ways of obtaining such operators, see (Ley et al., 2014) for a general approach.

In (Ley and Swan, 2016) we propose a tool for deriving identities such as (1), this time with the added advantage of also providing a statistical interpretation to the resulting quantities. The main result of that paper, however, is only valid in a one-parameter setting under the strong assumption that the target has support which does not depend on the parameter of interest. As will be shown in Section 2, such restrictions are not necessary and the difficulties entailed by a general setting can, if handled with care, be dealt with in a uniform and elegant manner. The main result of this paper, Theorem 2.1, therefore contributes a new tool to the literature on Stein's method, in the form of a most general parametric Stein characterization.

2 Characterizations in terms of a parameter of interest

2.1 Notations and definitions

Throughout, we let $k, p \in \mathbb{N}_0$ and consider two measure spaces $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, m_{\mathcal{X}})$ and $(\Theta, \mathcal{B}_{\Theta}, m_{\Theta})$, where $\mathcal{X}$ is either $\mathbb{R}^k$ or $\mathbb{Z}^k$, where Θ is a subset of $\mathbb{R}^p$ whose interior is non-empty, where $m_{\mathcal{X}}$ is either the Lebesgue measure or the counting measure, depending on the nature of $\mathcal{X}$, where m_{Θ} is the Lebesgue measure, and where $\mathcal{B}_{\mathcal{X}}$ and $\mathcal{B}_{\Theta}$ are the corresponding σ-algebras. In this setup we disregard the case of discrete parameter spaces (as in, e.g., the discrete uniform).

Consider a couple $(\mathcal{X}, \Theta)$ equipped with the corresponding σ-algebras and measures. We say that the measurable function $g : \mathcal{X} \times \Theta \to \mathbb{R}^+$ forms a family of θ-*parametric densities*, denoted by $g(\cdot; \theta)$, if $\int_{\mathcal{X}} g(x; \theta) dx = 1$ for all $\theta \in \Theta$. In this case we call θ the parameter of interest for g. When $\mathcal{X} = \mathbb{R}^k$, corresponding to the absolutely continuous case, the mapping $x \mapsto g(x; \theta)$ is, for all $\theta \in \Theta$, a probability density function evaluated at the point $x \in \mathbb{R}^k$. When $\mathcal{X} = \mathbb{Z}^k$, corresponding to the discrete case, $g(x; \theta)$ is the probability mass associated with $x \in \mathbb{Z}^k$ and $g(\cdot; \theta)$ therefore maps $\mathbb{Z}^k$ onto $[0, 1]$. This unified terminology will allow us to treat absolutely

continuous and discrete distributions in one common framework. For the sake of simplicity, we rule out mixed distributions.

Throughout this paper, the densities we shall work with all belong to the class $\mathcal{G} := \mathcal{G}(\mathcal{X}, \Theta)$ of θ-parametric densities for which the mapping $\theta \mapsto g(\cdot; \theta)$ is differentiable in the sense of distributions. We write $\nabla_\theta = \left(\partial_{\theta_1}, \ldots, \partial_{\theta_p} \right)$ the gradient operator, with $\partial_{\theta_i}, i = 1, \ldots, p$ the derivative in the sense of distributions with respect to θ_i. Such distributions have support (bounded or unbounded) possibly depending on the parameter θ; we will denote this support by $S_\theta := S_\theta(g)$, be it dependent on θ or not. We draw the attention of the reader to the fact that the notions we introduce are local, in the sense that they only need be defined in a neighborhood of some $\theta_0 \in \Theta$ (where it is understood that a set $\Theta_0 \subset \Theta$ is a neighborhood of θ_0 if Θ_0 contains an open subset of Θ which itself contains θ_0).

Definition 2.1. *Let θ_0 be an interior point of $\Theta \subset \mathbb{R}^p$ and let $g \in \mathcal{G}$. We define the class $\mathcal{F}(g; \theta_0)$ as the collection of* test functions $f : \mathcal{X} \times \Theta \to \mathbb{R}$ *such that the following three conditions are satisfied in some neighborhood $\Theta_0 \subset \Theta$ of θ_0:*

Condition (i) : there exists $c_f \in \mathbb{R}$ such that $\int_{\mathcal{X}} f(x; \theta) g(x; \theta) dm_{\mathcal{X}}(x) = c_f$ for all $\theta \in \Theta_0$.

Condition (ii) : the mapping $\theta \mapsto f(\cdot; \theta) g(\cdot; \theta)$ is differentiable in the sense of distributions over Θ_0.

Condition (iii) : there exist p $m_{\mathcal{X}}$-integrable functions $h_i : \mathcal{X} \to \mathbb{R}^+$, $i = 1, \ldots, p$, such that $|\partial_{\theta_i}(f(x; \theta) g(x; \theta))| \leq h_i(x)$ over $\mathcal{X}$ for all $i = 1, \ldots, p$ and for all $\theta \in \Theta_0$.

Definition 2.2. *Let θ_0 be an interior point of Θ. Also let g and $\mathcal{F}(g; \theta_0)$ be as above. We define the* Stein operator $\mathcal{T}_{\theta_0, g} : \mathcal{F}(g; \theta_0) \to \mathcal{X}^*$ *as*

$$ \mathcal{T}_{\theta_0, g}(f)(x) = \begin{cases} \frac{\nabla_\theta(f(x;\theta)g(x;\theta))|_{\theta=\theta_0}}{g(x;\theta_0)} & \text{if } x \in S_{\theta_0} \\ 0 & \text{otherwise.} \end{cases} \tag{2} $$

If the gradients $\nabla_\theta(f(x; \theta))|_{\theta=\theta_0}$ and $\nabla_\theta(g(x; \theta))|_{\theta=\theta_0}$ are well-defined (we stress that this is not necessarily the case) on $\mathcal{X}$ then

$$ \mathcal{T}_{\theta_0, g}(f)(x) = \left(\nabla_\theta(f(x; \theta))|_{\theta=\theta_0} + f(x; \theta_0) \frac{\nabla_\theta(g(x; \theta))|_{\theta=\theta_0}}{g(x; \theta_0)} \right) \mathbb{I}_{S_{\theta_0}}(x). $$

It should be noted that, for any test function f, the support of $\mathcal{T}_{\theta_0, g}(f)(x)$ is included in S_{θ_0}.

Among densities $g \in \mathcal{G}$, those which satisfy the following (local) regularity assumption at a given interior point $\theta_0 \in \Theta$ will play a particular role.

Assumption A : there exists a rectangular bounded neighborhood $\Theta_0 \subset \Theta$ of $\theta_0 = (\theta_0^1, \ldots, \theta_0^p)$ such that

A1. there exists a $m_{\mathcal{X}}$-integrable function $h : \mathcal{X} \to \mathbb{R}^+$ such that $g(x; \theta) \leq h(x)$ over $\mathcal{X}$ for all $\theta \in \Theta_0$;

A2. there exists $\epsilon > 0$ such that $\int_{S_\theta} g(x; u) dm_{\mathcal{X}}(x) \geq \epsilon$ for all $u, \theta \in \Theta_0$.

Assumption A1 is satisfied by densities which are bounded both as a function of $x \in \mathcal{X}$ and of $\theta \in \Theta_0$, in which case one can take $h(x) = \sup_{\theta \in \Theta_0} g(x; \theta)$ which is necessarily integrable. Assumption A2 is satisfied as soon as the support of $g(\cdot; \theta)$ does not depend on θ. Let $X_u \sim g(\cdot; u)$ for some $u \in \Theta$. Assumption A2 implies $P(X_u \in S_\theta) \geq \epsilon$ for all u and all $\theta \in \Theta_0$.

Example 2.1.

- *The Arcsine distribution with $\theta \in \mathbb{R}$ a location parameter*

$$g(x; \theta) = \left(\pi \sqrt{(x - \theta)(1 - x + \theta)} \right)^{-1} \mathbb{I}(\theta \leq x \leq \theta + 1)$$

  does not satisfy Assumption A1.

- *The Student distribution t_ν with $\theta \in \mathbb{R}$ a location parameter*

$$g(x; \theta) \propto \left(1 + \frac{(x - \theta)^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$

  satisfies Assumption A at $\theta_0 = 0$.

- *The exponential distribution with $\theta \in \mathbb{R}$ a location parameter*

$$g(x; \theta) = e^{-(x-\theta)} \mathbb{I}(x \geq \theta)$$

  satisfies Assumption A1 at $\theta_0 = 0$. To see why it also satisfies Assumption A2 at $\theta_0 = 0$ we note that

$$\int_{S_\theta} g(x; u) dm_{\mathcal{X}}(x) = \int_\theta^{+\infty} e^{-(x-u)} \mathbb{I}(x \geq u) dm_{\mathcal{X}}(x) = e^{-(\theta - u)^+} \geq \epsilon$$

  with $\epsilon = \inf_{\theta, u \in \Theta_0} e^{-(\theta - u)^+}$ (which is necessarily positive).

Theorem 2.1. *Let $g \in \mathcal{G}$, let Z_θ be distributed according to $g(\cdot; \theta)$, and let X be a random vector taking values on $\mathcal{X}$. Fix an interior point $\theta_0 \in \Theta$. Then the following two assertions hold:*

(1) *If $X \overset{\mathcal{L}}{=} Z_{\theta_0}$, then $\mathrm{E}[\mathcal{T}_{\theta_0, g}(f)(X)] = 0$ for all $f \in \mathcal{F}(g; \theta_0)$.*

(2) *If g also satisfies Assumption A at θ_0 and if $P(X \in S_{\theta_0}) > 0$ then*

$$X \mid X \in S_{\theta_0} \overset{\mathcal{L}}{=} Z_{\theta_0}$$

  if $\mathrm{E}[\mathcal{T}_{\theta_0, g}(f)(X)] = 0$ for all $f \in \mathcal{F}(g; \theta_0)$.

4

Proof. (1) Since Condition (iii) allows for differentiating w.r.t. θ under the integral in Condition (i) and since differentiating w.r.t. θ is allowed thanks to Condition (ii), the claim follows immediately.

(2) First suppose that $p = 1$, and fix $\Theta_0 \subset \Theta$, a bounded (rectangular) neighborhood of θ_0 on which g satisfies Assumption A at θ_0. Define, for $A \in \mathcal{B}_\mathcal{X}$, the mapping

$$f_A : \mathcal{X} \times \Theta_0 \to \mathbb{R} : (x, \theta) \mapsto \frac{1}{g(x;\theta)} \int_{\theta_0}^\theta l_A(x;u,\theta)g(x;u)dm_\Theta(u) \qquad (3)$$

with

$$l_A(x;u,\theta) := (\mathbb{I}_A(x) - \mathrm{P}(Z_u \in A \,|\, Z_u \in S_\theta))\,\mathbb{I}_{S_\theta}(x),$$

where

$$\mathrm{P}(Z_u \in B) = \int_\mathcal{X} \mathbb{I}_B(x)g(x;u)dm_\mathcal{X}(x)$$

for $B \in \mathcal{B}_\mathcal{X}$. Note that Assumption A2 guarantees that the event $[Z_u \in S_\theta]$ has a non-zero probability. To see that f_A belongs to $\mathcal{F}(g;\theta_0)$, first note that

$$\begin{aligned}
\int_\mathcal{X} f_A(x;\theta)g(x;\theta)dm_\mathcal{X}(x) &= \int_\mathcal{X} \int_{\theta_0}^\theta l_A(x;u,\theta)g(x;u)dm_\Theta(u)dm_\mathcal{X}(x) \\
&= \int_{\theta_0}^\theta \int_\mathcal{X} l_A(x;u,\theta)g(x;u)dm_\mathcal{X}(x)dm_\Theta(u),
\end{aligned}$$

where the last equality follows from Fubini's theorem, which can be applied for all $\theta \in \Theta_0$, since in this case there exists a constant M such that

$$\int_{\theta_0}^\theta \int_\mathcal{X} |l_A(x;u,\theta)|g(x;u)dm_\mathcal{X}(x)dm_\Theta(u) \le 2|\theta - \theta_0| \le M$$

for all $\theta \in \Theta_0$. We also have, by definition of l_A,

$$\begin{aligned}
\int_\mathcal{X} & l_A(x;u,\theta)g(x;u)dm_\mathcal{X}(x) \\
&= \mathrm{P}(Z_u \in A \cap S_\theta) - \mathrm{P}(Z_u \in A \,|\, Z_u \in S_\theta)\,\mathrm{P}(Z_u \in S_\theta) \\
&= 0.
\end{aligned}$$

Hence f_A satisfies Condition (i). Condition (ii) is easily checked. Regarding Condition (iii), one sees that

$$\partial_t \left(f_A(x;t)g(x;t)\right)|_{t=\theta} = l_A(x;\theta,\theta)g(x;\theta) + H(x;\theta), \qquad (4)$$

with $H(x;\theta)$ given in (5). As shown there, it is easy to bound $H(x;\theta)$ uniformly in θ over Θ_0 by a $m_\mathcal{X}$-integrable function. Moreover, Assumption A guarantees that the same holds for $l_A(x;\theta,\theta)g(x;\theta)$. Hence f_A satisfies

Condition (iii). We have thus proved that $f_A \in \mathcal{F}(g; \theta_0)$. Using $H(x; \theta_0) = 0$ for all $x \in \mathcal{X}$ (see Section 3) along with the fact that

$$\mathrm{E}[\mathcal{T}_{\theta_0, g}(f_A)(X)] = \mathrm{E}[\mathbb{I}_{A \cap S_{\theta_0}}(X) - \mathrm{P}(Z_{\theta_0} \in A)\mathbb{I}_{S_{\theta_0}}(X)] = 0$$

the conclusion follows for $p = 1$.

Next suppose that $p > 1$. Let $\theta_0 := (\theta_0^1, \ldots, \theta_0^p)$ and fix $\Theta_0 := \Theta_0^1 \times \ldots \times \Theta_0^p$ a bounded (rectangular) neighborhood of θ_0 on which g satisfies Assumption A at θ_0. Define, for all $j = 1, \ldots, p$ and for all $A \in \mathcal{B}_{\mathcal{X}}$, the mappings

$$\bar{\theta}_0^j : \Theta_0^j \to \Theta_0 : u \mapsto (\theta_0^1, \ldots, \theta_0^{j-1}, u, \theta_0^{j+1}, \ldots, \theta_0^p)$$

and

$$f_A^j : \mathcal{X} \times \Theta_0 \to \mathbb{R} : (x, \theta) \mapsto \frac{1}{g(x; \theta)} \int_{\theta_0^j}^{\theta^j} l_A^j(x; u, \theta^j) g(x; \bar{\theta}_0^j(u)) dm_\Theta(u),$$

with

$$l_A^j(x; u, \theta^j) := \left(\mathbb{I}_A(x) - \mathrm{P}\left(Z_u^j \in A \mid Z_u^j \in S_{\bar{\theta}_0^j(\theta^j)} \right) \right) \mathbb{I}_{S_{\bar{\theta}_0^j(\theta^j)}}(x),$$

where

$$\mathrm{P}(Z_u^j \in B) := \int_{\mathcal{X}} \mathbb{I}_B(x) g(x; \bar{\theta}_0^j(u)) dm_{\mathcal{X}}(x)$$

for $B \in \mathcal{B}_{\mathcal{X}}$. The p-variate equivalent of the function f_A in (3) is given by $f_A^{(p)}(x; \theta) := \sum_{j=1}^p f_A^j(x; \theta)$. Along the same lines as for the special case $p = 1$, Conditions (i)-(iii) are now easily seen to be satisfied by $f_A^{(p)}$ (we draw the reader's attention to the fact that the rectangular nature of the neighborhood Θ_0 is important in order to ensure Condition (iii)). The result readily follows. $\quad\square$

Remark 2.1. *Nowhere in the proof did we need to specify whether the random vector X is univariate (for $k = 1$) or multivariate (for $k > 1$).*

Remark 2.2. *When $p > 1$, the (vectorial) operator $\mathcal{T}_{\theta_0, g}(f)$ contains, in a sense, p different characterizations of the $\theta = (\theta^1, \ldots, \theta^p)$-parametric density g at θ_0. The requirements (in this formulation of the result) on the test functions f are, perhaps, unnecessarily stringent. Indeed, setting $\theta^{(q)} := (\theta^{i_1}, \ldots, \theta^{i_q})$ for $1 \le i_1 \le \ldots \le i_q \le p$, we can obviously consider g as a $\theta^{(q)}$-parametric density. The corresponding q-dimensional sub-vector of $\mathcal{T}_{\theta_0, g}(f)$ also gives rise to a (vectorial) Stein operator for which the conclusions of Theorem 2.1 also hold at θ_0, this time with a possibly larger class of test functions f (thanks to the weakening of the requirements imposed by Condition (iii)). In particular, taking $q = 1$, we obtain p distinct one-dimensional characterizations of g at θ_0.*

Remark 2.3. *Note that both implications in Theorem 2.1 are obtained at fixed $\theta_0 \in \Theta$. We attract the reader's attention to the fact that all our calculations and manipulations, as well as all the conditions on the functions at play, are consequently local around θ_0.*

3   Proof of equality (4)

First note that

$$
\begin{aligned}
\partial_t \left(f_A(x;t)g(x;t) \right)\big|_{t=\theta} & \\
&= \partial_t \left(\int_{\theta_0}^t l_A(x;u,t)g(x;u)dm_\Theta(u) \right)\bigg|_{t=\theta} \\
&= l_A(x;\theta,\theta)g(x;\theta) + \int_{\theta_0}^{\theta} \partial_t \left(l_A(x;u,t) \right)\big|_{t=\theta}\, g(x;u)dm_\Theta(u).
\end{aligned}
$$

Now we have

$$
\begin{aligned}
\partial_t \left(l_A(x;u,t) \right)\big|_{t=\theta} &= \partial_t \left(\mathbb{I}_{S_t}(x) \right)\big|_{t=\theta} \left(\mathbb{I}_A(x) - \mathrm{P}(Z_u \in A \mid Z_u \in S_\theta) \right) \\
&\quad - \partial_t \left(\mathrm{P}(Z_u \in A \mid Z_u \in S_t) \right)\big|_{t=\theta}\, \mathbb{I}_{S_\theta}(x).
\end{aligned}
$$

On the one hand, setting

$$
\begin{aligned}
&H_1(x;\theta) \\
&:= \partial_t \left(\mathbb{I}_{S_t}(x) \right)\big|_{t=\theta} \int_{\theta_0}^{\theta} \left(\mathbb{I}_A(x) - \mathrm{P}(Z_u \in A \mid Z_u \in S_\theta) \right) g(x;u)dm_\Theta(u)
\end{aligned}
$$

we immediately get

$$
|H_1(x;\theta)| \le 2h(x) \left| \partial_t \left(\mathbb{I}_{S_t}(x) \right)\big|_{t=\theta} \right| m_\Theta([\theta_0,\theta])
$$

and is thus bounded uniformly in θ over Θ_0 by a $m_\mathcal{X}$-integrable function and satisfies $H_1(x;\theta_0) = 0$. On the other hand, we have (thanks to Assumption A2)

$$
\begin{aligned}
&\left| \partial_t \left(\mathrm{P}(Z_u \in A \mid Z_u \in S_t) \right)\big|_{t=\theta} \right| \\
&= \left| \frac{\partial_t \left(\mathrm{P}(Z_u \in A \cap S_t) \right)\big|_{t=\theta}}{\mathrm{P}(Z_u \in S_\theta)} - \partial_t \left(\mathrm{P}(Z_u \in S_t) \right)\big|_{t=\theta} \frac{\mathrm{P}(Z_u \in A \cap S_\theta)}{\mathrm{P}(Z_u \in S_\theta)^2} \right| \\
&\le \frac{1}{\epsilon} \left(\left| \partial_t \left(\mathrm{P}(Z_u \in A \cap S_t) \right)\big|_{t=\theta} \right| + \left| \partial_t \left(\mathrm{P}(Z_u \in S_t) \right)\big|_{t=\theta} \right| \right)
\end{aligned}
$$

with

$$
\begin{aligned}
&\left| \partial_t \left(\mathrm{P}(Z_u \in A \cap S_t) \right)\big|_{t=\theta} \right| + \left| \partial_t \left(\mathrm{P}(Z_u \in S_t) \right)\big|_{t=\theta} \right| \\
&= \left| \partial_t \left(\int_{A \cap S_t} g(x;u)dm_\mathcal{X}(x) \right)\bigg|_{t=\theta} \right| + \left| \partial_t \left(\int_{S_t} g(x;u)dm_\mathcal{X}(x) \right)\bigg|_{t=\theta} \right|
\end{aligned}
$$

which is necessarily bounded by some finite constant, say $C(u, \theta)$, possibly dependent on u and θ. Hence, setting

$$H_2(x; \theta) := \mathbb{I}_{S_\theta}(x) \int_{\theta_0}^{\theta} \partial_t \left(\mathrm{P}(Z_u \in A \mid Z_u \in S_t) \right)|_{t=\theta} \, g(x; u) dm_\Theta(u)$$

we obtain

$$|H_2(x; \theta)| \leq \frac{1}{\epsilon} \mathbb{I}_{S_\theta}(x) h(x) \int_{\theta_0}^{\theta} C(u, \theta) dm_\Theta(u)$$

and H_2 is thus also bounded uniformly in θ over Θ_0 by a $m_{\mathcal{X}}$-integrable function and satisfies $H_2(x; \theta_0) = 0$. Defining

$$H(x; \theta) := H_1(x; \theta) - H_2(x; \theta) \tag{5}$$

we see that all the assertions in the proof of Theorem 2.1 hold, and, moreover, that

$$\begin{aligned} \partial_t \left(f_A(x; t) g(x; t) \right)|_{t=\theta_0} &= l_A(x; \theta_0, \theta_0) g(x; \theta_0) + H(x; \theta_0) \\ &= l_A(x; \theta_0, \theta_0) g(x; \theta_0). \end{aligned}$$

This completes the proof of Theorem 2.1.

Acknowledgements

References

Afendras, G., Papadatos, N., and Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds. *Bernoulli*, 17(2):507–529.

Chatterjee, S. and Shao, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie-Weiss model. *The Annals of Applied Probability*, 21(2):464–483.

Chen, L. H. Y., Goldstein, L., and Shao, Q.-M. (2011). *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg.

Franceschetti, M. and Meester, R. (2006). Critical node lifetimes in random networks via the Chen-Stein method. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2831–2837.

Gaunt, R. E., Pickett, A., and Reinert, G. (2015). Chi-square approximation by Stein's method with application to Pearson's statistic. *arXiv preprint arXiv:1507.01707*.

Goldstein, L. and Reinert, G. (2013). Stein's method for the Beta distribution and the Pólya-Eggenberger urn. *Journal of Applied Probability*, 50(4):1187–1205.

Kusuoka, S. and Tudor, C. A. (2013). Extension of the fourth moment theorem to invariant measures of diffusions. *Preprint arXiv:1310.3785*.

Ley, C., Reinert, G., and Swan, Y. (2014). Approximate computation of expectations: a canonical Stein operator. *arXiv preprint arXiv:1408.2998*.

Ley, C. and Swan, Y. (2013). Stein's density approach and information inequalities. *Electronic Communications in Probability*, 18(7):1–14.

Ley, C. and Swan, Y. (2016). Parametric Stein operators and variance bounds. *Brazilian Journal of Probability and Statistics*, to appear.

Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., and Tropp, J. A. (2014). Matrix concentration inequalities via the method of exchangeable pairs. *The Annals of Probability*, 42(3):906–945.

McKeague, I. W. and Levin, B. (2015). Convergence of empirical distributions in an interpretation of quantum mechanics. *Annals of Applied Probability, to appear*.

Ross, N. (2011). Fundamentals of Stein's method. *Probability Surveys*, 8:210–293.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif. Univ. California Press.

Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.