# RKHS in ML: Comparing a Sample and a Model

Arthur Gretton

Gatsby Computational Neuroscience Unit,
University College London

December 1, 2021

# Before: comparing two samples

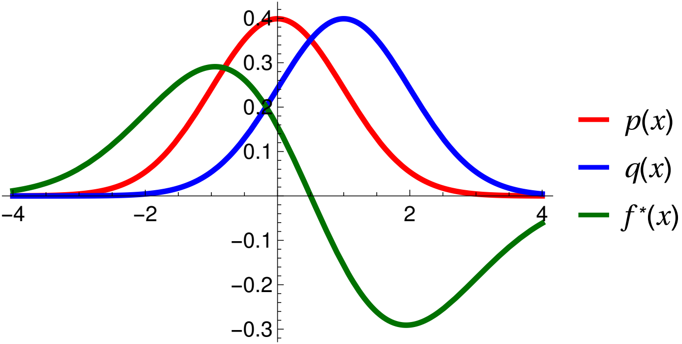
- Given: Samples from unknown distributions $P$ and $Q$.
- Goal: do $P$ and $Q$ differ?

$$\sim P$$

$$\sim Q$$

# Now: statistical model criticism

$$MMD(P, Q) = \sup_{\|f\|_{\mathcal{F}} \leq 1} [E_q f - E_p f]$$




Can we compute MMD with samples from $Q$ and a model $P$?

Remark: assume $P$ has prob. density $p$, known up to normalization.

Problem: usualy can't compute $E_p f$ in closed form.

# Stein idea

To get rid of $E_p f$ in

$$\sup_{\|f\|_{\mathcal{F}} \leq 1} [E_q f - E_p f]$$

we define the Stein operator

$$[T_p f](x) = \frac{1}{p(x)} \frac{d}{dx} (f(x) p(x))$$

Then

$$E_p T_p f = 0$$

subject to appropriate boundary conditions. (Oates, Girolami, Chopin, 2016)

# Stein idea: proof

$$E_p\left[T_p f\right] = \int \left[\frac{1}{p(x)}\frac{d}{dx}\left(f(x)p(x)\right)\right] p(x)\,dx$$

$$\int \left[\frac{d}{dx}\left(f(x)p(x)\right)\right] dx$$

$$= \left[f(x)p(x)\right]_{-\infty}^{\infty}$$

$$= 0$$

# Stein idea: proof

$$E_{p}\left[T_{p}f\right] = \int \left[\frac{1}{\cancel{p(x)}}\frac{d}{dx}\left(f(x)p(x)\right)\right]\cancel{p(x)}dx$$

$$\int \left[\frac{d}{dx}\left(f(x)p(x)\right)\right]dx$$

$$= \left[f(x)p(x)\right]_{-\infty}^{\infty}$$

$$= 0$$

# Stein idea: proof

$$E_p\left[T_p f\right] = \int \left[\frac{1}{\cancel{p(x)}}\frac{d}{dx}\left(f(x)p(x)\right)\right]\cancel{p(x)}dx$$

$$\int \left[\frac{d}{dx}\left(f(x)p(x)\right)\right]dx$$

$$= \left[f(x)p(x)\right]_{-\infty}^{\infty}$$

$$= 0$$

# Stein idea: proof

$$E_p\left[T_p f\right] = \int \left[\frac{1}{\cancel{p(x)}}\frac{d}{dx}\left(f(x)p(x)\right)\right]\cancel{p(x)}dx$$

$$\int \left[\frac{d}{dx}\left(f(x)p(x)\right)\right] dx$$

$$= \left[f(x)p(x)\right]_{-\infty}^{\infty}$$

$$= 0$$

# Stein idea: proof

$$
\begin{aligned}
E_p\left[T_p f\right] &= \int \left[\frac{1}{\cancel{p(x)}}\frac{d}{dx}\left(f(x)p(x)\right)\right]\cancel{p(x)}dx \\
&\int \left[\frac{d}{dx}\left(f(x)p(x)\right)\right]dx \\
&= \left[f(x)p(x)\right]_{-\infty}^{\infty} \\
&= 0
\end{aligned}
$$

# Kernel Stein Discrepancy

Stein operator

$$T_p f = \frac{1}{p(x)} \frac{d}{dx} (f(x) p(x))$$

Kernel Stein Discrepancy (KSD)

$$KSD(p, q, \mathcal{F}) = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g - E_p T_p g$$

# Kernel Stein Discrepancy

Stein operator

$$T_p f = \frac{1}{p(x)} \frac{d}{dx} \left( f(x) p(x) \right)$$

Kernel Stein Discrepancy (KSD)

$$KSD(p, q, \mathcal{F}) = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g - \cancel{E_p T_p g} = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g$$

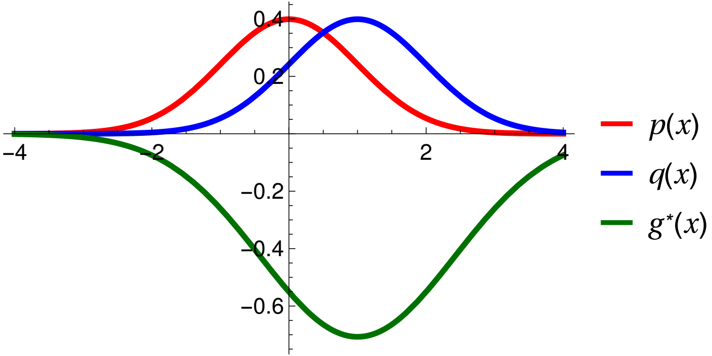
# Kernel Stein Discrepancy

Stein operator

$$T_p f = \frac{1}{p(x)} \frac{d}{dx}\left(f(x)p(x)\right)$$

Kernel Stein Discrepancy (KSD)

$$KSD(p, q, \mathcal{F}) = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g - \cancel{E_p T_p g} = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g$$



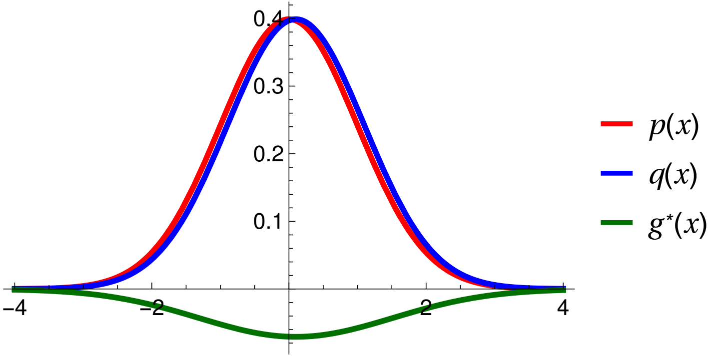
# Kernel Stein Discrepancy

Stein operator

$$T_p f = \frac{1}{p(x)} \frac{d}{dx} \left( f(x) p(x) \right)$$

Kernel Stein Discrepancy (KSD)

$$KSD(p, q, \mathcal{F}) = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g - \cancel{E_p T_p g} = \sup_{\|g\|_{\mathcal{F}} \leq 1} E_q T_p g$$



# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx}(f(x)p(x)) \\
&= \frac{1}{p(x)} \left( p(x)\frac{d}{dx}f(x) + f(x)\frac{d}{dx}p(x) \right) \\
&= \frac{d}{dx}f(x) + f(x)\frac{1}{p(x)}\frac{d}{dx}p(x) \\
&= \frac{d}{dx}f(x) + f(x)\frac{d}{dx}\log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left(\frac{d}{dx}\log p(x)\right) f(x) + \frac{d}{dx}f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx} (f(x)p(x)) \\
&= \frac{1}{p(x)} \left( p(x)\frac{d}{dx}f(x) + f(x)\frac{d}{dx}p(x) \right) \\
&= \frac{d}{dx}f(x) + f(x)\frac{1}{p(x)}\frac{d}{dx}p(x) \\
&= \frac{d}{dx}f(x) + f(x)\frac{d}{dx}\log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left(\frac{d}{dx}\log p(x)\right) f(x) + \frac{d}{dx}f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx}\left(f(x)p(x)\right) \\
&= \frac{1}{p(x)} \left( p(x)\frac{d}{dx}f(x) + f(x)\frac{d}{dx}p(x) \right) \\
&= \frac{d}{dx}f(x) + f(x)\frac{1}{p(x)}\frac{d}{dx}p(x) \\
&= \frac{d}{dx}f(x) + f(x)\frac{d}{dx}\log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left(\frac{d}{dx}\log p(x)\right) f(x) + \frac{d}{dx}f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx} (f(x)p(x)) \\
&= \frac{1}{p(x)} \left( p(x)\frac{d}{dx} f(x) + f(x)\frac{d}{dx} p(x) \right) \\
&= \frac{d}{dx} f(x) + f(x)\frac{1}{p(x)} \frac{d}{dx} p(x) \\
&= \frac{d}{dx} f(x) + f(x)\frac{d}{dx} \log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left( \frac{d}{dx} \log p(x) \right) f(x) + \frac{d}{dx} f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx}(f(x)p(x)) \\
&= \frac{1}{p(x)} \left( p(x)\frac{d}{dx}f(x) + f(x)\frac{d}{dx}p(x) \right) \\
&= \frac{d}{dx}f(x) + f(x)\frac{1}{p(x)}\frac{d}{dx}p(x) \\
&= \frac{d}{dx}f(x) + f(x)\frac{d}{dx}\log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left(\frac{d}{dx}\log p(x)\right) f(x) + \frac{d}{dx}f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Re-write stein operator as:

$$
\begin{aligned}
[T_p f](x) &= \frac{1}{p(x)} \frac{d}{dx} \left(f(x) p(x)\right) \\
&= \frac{1}{p(x)} \left( p(x) \frac{d}{dx} f(x) + f(x) \frac{d}{dx} p(x) \right) \\
&= \frac{d}{dx} f(x) + f(x) \frac{1}{p(x)} \frac{d}{dx} p(x) \\
&= \frac{d}{dx} f(x) + f(x) \frac{d}{dx} \log p(x)
\end{aligned}
$$

Can we get a dot product in feature space?

$$
\begin{aligned}
[T_p f](x) &= \left( \frac{d}{dx} \log p(x) \right) f(x) + \frac{d}{dx} f(x) \\
&=: \langle f, \xi_x \rangle_{\mathcal{F}}
\end{aligned}
$$

# Simple expression using kernels

Step 1: we need reproducing property for the derivative: for differentiable $k(x-x')$,

$$\frac{d}{dx} f(x) = \left\langle f, \frac{d}{dx} k(x,\cdot) \right\rangle_{\mathcal{F}}$$

$$\frac{d}{dx}\frac{d}{dx'} k(x-x') = \left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}}$$

Proof for $\mathcal{X} := [-\pi, \pi]$, periodic boundary conditions.

Fourier transforms:

$$f(x) = \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell \exp(\imath \ell x), \qquad \hat{f}_\ell = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) \exp(-\imath \ell x)\, dx.$$

Fourier series representation of derivative:

$$\frac{d}{dx} f(x) \xrightarrow{\mathcal{F}} \left\{ (\imath \ell) \hat{f}_\ell \right\}_{\ell=-\infty}^{\infty}.$$

# Simple expression using kernels

Step 1: we need reproducing property for the derivative: for differentiable $k(x - x')$,

$$\frac{d}{dx} f(x) = \left\langle f, \frac{d}{dx} k(x, \cdot) \right\rangle_{\mathcal{F}}$$

$$\frac{d}{dx} \frac{d}{dx'} k(x - x') = \left\langle \frac{d}{dx} k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}}$$

Proof for $\mathcal{X} := [-\pi, \pi]$, periodic boundary conditions.

Fourier transforms:

$$f(x) = \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell \exp(\imath \ell x), \qquad \hat{f}_\ell = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-\imath \ell x)\, dx.$$

Fourier series representation of derivative:

$$\frac{d}{dx} f(x) \xrightarrow{\mathcal{F}} \left\{ (\imath \ell) \hat{f}_\ell \right\}_{\ell=-\infty}^{\infty}.$$

# Simple expression using kernels

Step 1: we need reproducing property for the derivative: for differentiable $k(x-x')$,

$$\frac{d}{dx}f(x) = \left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}}$$

$$\frac{d}{dx}\frac{d}{dx'}k(x-x') = \left\langle \frac{d}{dx}k(x,\cdot), \frac{d}{dx'}k(x',\cdot)\right\rangle_{\mathcal{F}}$$

Proof for $\mathcal{X} := [-\pi, \pi]$, periodic boundary conditions.

Fourier transforms:

$$f(x) = \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell \exp(\imath \ell x), \qquad \hat{f}_\ell = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\exp(-\imath \ell x)\, dx.$$

Fourier series representation of derivative:

$$\frac{d}{dx}f(x) \xrightarrow{\mathcal{F}} \left\{(\imath \ell)\hat{f}_\ell\right\}_{\ell=-\infty}^{\infty}.$$

# Simple expression using kernels

Proof of $\frac{d}{dx}f(x) = \left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}}$:

Define

$$g(y) := \frac{d}{dx}k(x-y) = \sum_{\ell=-\infty}^{\infty} (\imath\ell)\hat{k}_\ell \exp\left(\imath\ell(x-y)\right).$$

$g(y)$ real so

$$g(y) = \bar{g}(y) = \sum_{\ell=-\infty}^{\infty} -(\imath\ell)\hat{k}_\ell \exp\left(\imath\ell(y-x)\right),$$

since $\bar{\hat{k}}_\ell = \hat{k}_\ell$.

Fourier coefficients of $g(y)$:

$$\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$$

# Simple expression using kernels

Proof of $\frac{d}{dx}f(x) = \left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}}$:

Define

$$g(y) := \frac{d}{dx}k(x-y) = \sum_{\ell=-\infty}^{\infty} (\imath\ell)\hat{k}_\ell \exp\left(\imath\ell(x-y)\right).$$

$g(y)$ real so

$$g(y) = \bar{g}(y) = \sum_{\ell=-\infty}^{\infty} -(\imath\ell)\hat{k}_\ell \exp\left(\imath\ell(y-x)\right),$$

since $\bar{\hat{k}}_\ell = \hat{k}_\ell$.

Fourier coefficients of $g(y)$:

$$\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$$

# Simple expression using kernels

Proof of $\frac{d}{dx} f(x) = \left\langle f, \frac{d}{dx} k(x, \cdot) \right\rangle_{\mathcal{F}}$:

Define

$$g(y) := \frac{d}{dx} k(x-y) = \sum_{\ell=-\infty}^{\infty} (\imath \ell) \hat{k}_\ell \exp\left(\imath \ell (x-y)\right).$$

$g(y)$ real so

$$g(y) = \bar{g}(y) = \sum_{\ell=-\infty}^{\infty} -(\imath \ell) \hat{k}_\ell \exp\left(\imath \ell (y-x)\right),$$

since $\bar{\hat{k}}_\ell = \hat{k}_\ell$.

Fourier coefficients of $g(y)$:

$$\hat{g}_\ell = -(\imath \ell) \hat{k}_\ell \exp(-\imath \ell x)$$

# Simple expression using kernels

From previous slide, $\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$

We can write

$$\left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}} = \langle f, g(\cdot)\rangle_{\mathcal{F}}$$

$$= \frac{\left(\hat{f}_\ell\right)\left(\bar{\hat{g}}_\ell\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left(\hat{f}_\ell\right)\left(\overline{-\imath\ell\hat{k}_\ell \exp(-\imath\ell x)}\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} (\imath\ell)\left(\hat{f}_\ell\right)(\exp(\imath\ell x)) = \frac{d}{dx}f(x).$$

Also true more generally: see Steinwart and Christmann, Ch. 4.3 (proof via mean value theorem).

# Simple expression using kernels

From previous slide, $\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$

We can write

$$\left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}} = \langle f, g(\cdot)\rangle_{\mathcal{F}}$$

$$= \frac{\left(\hat{f}_\ell\right)\left(\bar{\hat{g}}_\ell\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left(\hat{f}_\ell\right)\left(\overline{-\imath\ell\hat{k}_\ell \exp(-\imath\ell x)}\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} (\imath\ell)\left(\hat{f}_\ell\right)(\exp(\imath\ell x)) = \frac{d}{dx}f(x).$$

Also true more generally: see Steinwart and Christmann, Ch. 4.3 (proof via mean value theorem).

# Simple expression using kernels

From previous slide, $\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$

We can write

$$\left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}} = \langle f, g(\cdot)\rangle_{\mathcal{F}}$$

$$= \frac{\left(\hat{f}_\ell\right)\left(\bar{\hat{g}}_\ell\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left(\hat{f}_\ell\right)\left(\overline{-\imath\ell\cancel{\hat{k}_\ell}\exp(-\imath\ell x)}\right)}{\cancel{\hat{k}_\ell}}$$

$$= \sum_{\ell=-\infty}^{\infty} (\imath\ell)\left(\hat{f}_\ell\right)(\exp(\imath\ell x)) = \frac{d}{dx}f(x).$$

Also true more generally: see Steinwart and Christmann, Ch. 4.3 (proof via mean value theorem).

# Simple expression using kernels

From previous slide, $\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$

We can write

$$\left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}} = \langle f, g(\cdot)\rangle_{\mathcal{F}}$$

$$= \frac{\left(\hat{f}_\ell\right)\left(\bar{\hat{g}}_\ell\right)}{\hat{k}_\ell}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left(\hat{f}_\ell\right)\left(\overline{-\imath\ell\cancel{\hat{k}_\ell}\exp(-\imath\ell x)}\right)}{\cancel{\hat{k}_\ell}}$$

$$= \sum_{\ell=-\infty}^{\infty} (\imath\ell)\left(\hat{f}_\ell\right)\left(\exp(\imath\ell x)\right) = \frac{d}{dx}f(x).$$

Also true more generally: see Steinwart and Christmann, Ch. 4.3 (proof via mean value theorem).

# Simple expression using kernels

From previous slide, $\hat{g}_\ell = -(\imath\ell)\hat{k}_\ell \exp(-\imath\ell x)$

We can write

$$
\begin{aligned}
\left\langle f, \frac{d}{dx}k(x,\cdot)\right\rangle_{\mathcal{F}} &= \langle f, g(\cdot)\rangle_{\mathcal{F}} \\
&= \frac{\left(\hat{f}_\ell\right)\left(\bar{\hat{g}}_\ell\right)}{\hat{k}_\ell} \\
&= \sum_{\ell=-\infty}^{\infty} \frac{\left(\hat{f}_\ell\right)\left(\overline{-\imath\ell \cancel{\hat{k}_\ell} \exp(-\imath\ell x)}\right)}{\cancel{\hat{k}_\ell}} \\
&= \sum_{\ell=-\infty}^{\infty} (\imath\ell)\left(\hat{f}_\ell\right)\left(\exp(\imath\ell x)\right) = \frac{d}{dx}f(x).
\end{aligned}
$$

Also true more generally: see Steinwart and Christmann, Ch. 4.3 (proof via mean value theorem).

# Next step: taking expectations

We have shown:

$$
\begin{aligned}
[T_p f](z) &= \left(\frac{d}{dz}\log p(z)\right) f(z) + \frac{d}{dz} f(z) \\
&= \left\langle f, \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot) \right\rangle_{\mathcal{F}} \\
&=: \langle f, \xi_z \rangle_{\mathcal{F}} .
\end{aligned}
$$

# Next step: taking expectations

We have shown:

$$
\begin{aligned}
[T_p f](z) &= \left(\frac{d}{dz}\log p(z)\right) f(z) + \frac{d}{dz} f(z) \\
&= \left\langle f, \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot) \right\rangle_{\mathcal{F}} \\
&=: \langle f, \xi_z \rangle_{\mathcal{F}} .
\end{aligned}
$$

Step 2: show that

$$
E_{z\sim q}[T_p f] = E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} = \langle f, E_{z\sim q} \xi_z \rangle_{\mathcal{F}} .
$$

# Next step: taking expectations

We have shown:

$$
\begin{aligned}
[T_p f](z) &= \left(\frac{d}{dz}\log p(z)\right) f(z) + \frac{d}{dz} f(z) \\
&= \left\langle f, \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot) \right\rangle_{\mathcal{F}} \\
&=: \langle f, \xi_z \rangle_{\mathcal{F}} .
\end{aligned}
$$

Step 2: show that

$$
E_{z\sim q}[T_p f] = E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} = \langle f, E_{z\sim q} \xi_z \rangle_{\mathcal{F}} .
$$

Riesz theorem!

# Next step: taking expectations

Riesz theorem: need boundedness,

$$\left| E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} \right| \leq \|f\|_{\mathcal{F}} \lambda$$

for some $\lambda \in \mathbb{R}$.

By Jensen and Cauchy-Schwarz,

$$\begin{aligned} \left| E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} \right| &\leq E_{z\sim q} \left| \langle f, \xi_z \rangle_{\mathcal{F}} \right| \\ &\leq \|f\|_{\mathcal{F}} \underbrace{E_{z\sim q} \|\xi_z\|_{\mathcal{F}}}_{\text{bounded?}}. \end{aligned}$$

# Next step: taking expectations

Riesz theorem: need boundedness,

$$\left| E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} \right| \leq \|f\|_{\mathcal{F}} \, \lambda$$

for some $\lambda \in \mathbb{R}$.

By Jensen and Cauchy-Schwarz,

$$\begin{aligned} \left| E_{z\sim q} \langle f, \xi_z \rangle_{\mathcal{F}} \right| &\leq E_{z\sim q} \left| \langle f, \xi_z \rangle_{\mathcal{F}} \right| \\ &\leq \|f\|_{\mathcal{F}} \underbrace{E_{z\sim q} \|\xi_z\|_{\mathcal{F}}}_{\text{bounded?}}. \end{aligned}$$

# Next step: taking expectations

Compute the squared norm:

$$
\begin{aligned}
\|\xi_z\|_{\mathcal{F}}^2 &= \langle \xi_z, \xi_z \rangle_{\mathcal{F}} \\
&= \left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot), \ldots \right\rangle_{\mathcal{F}} \\
&= \underbrace{\left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot), \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) \right\rangle_{\mathcal{F}}}_{(A)} \\
&+ \underbrace{\left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \bigg|_{x=x'=z}}_{(B)= \frac{d}{dx}\frac{d}{dx'} k(x-x')\big|_{x=x'=z}} \\
&+ 2 \underbrace{\left\langle \left( \frac{d}{dx} \log p(x) \right) k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \bigg|_{x=x'=z}}_{(C)}
\end{aligned}
$$

# Next step: taking expectations

Compute the squared norm:

$$\|\xi_z\|^2_{\mathcal{F}} = \langle \xi_z, \xi_z \rangle_{\mathcal{F}}$$

$$= \left\langle \left( \frac{d}{dz} \log p(z) \right) k(z, \cdot) + \frac{d}{dz} k(z, \cdot), \ldots \right\rangle_{\mathcal{F}}$$

$$= \underbrace{\left\langle \left( \frac{d}{dz} \log p(z) \right) k(z, \cdot), \left( \frac{d}{dz} \log p(z) \right) k(z, \cdot) \right\rangle_{\mathcal{F}}}_{(A)}$$

$$+ \underbrace{\left\langle \frac{d}{dx} k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Big|_{x=x'=z}}_{(B)= \frac{d}{dx} \frac{d}{dx'} k(x-x') \big|_{x=x'=z}}$$

$$+ 2 \underbrace{\left\langle \left( \frac{d}{dx} \log p(x) \right) k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Big|_{x=x'=z}}_{(C)}$$

# Next step: taking expectations

Compute the squared norm:

$$
\begin{aligned}
\|\xi_z\|^2_{\mathcal{F}} &= \langle \xi_z, \xi_z \rangle_{\mathcal{F}} \\
&= \left\langle \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) + \frac{d}{dz}k(z,\cdot), \dots \right\rangle_{\mathcal{F}} \\
&= \underbrace{\left\langle \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot), \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) \right\rangle_{\mathcal{F}}}_{(A)} \\
&+ \underbrace{\left\langle \frac{d}{dx}k(x,\cdot), \frac{d}{dx'}k(x',\cdot) \right\rangle_{\mathcal{F}}\Bigg|_{x=x'=z}}_{(B)=\frac{d}{dx}\frac{d}{dx'}k(x-x')\big|_{x=x'=z}} \\
&+ 2\underbrace{\left\langle \left(\frac{d}{dx}\log p(x)\right) k(x,\cdot), \frac{d}{dx'}k(x',\cdot) \right\rangle_{\mathcal{F}}\Bigg|_{x=x'=z}}_{(C)}
\end{aligned}
$$

# Next step: taking expectations

Compute the squared norm:

$$
\begin{aligned}
\|\xi_z\|^2_{\mathcal{F}} &= \langle \xi_z, \xi_z \rangle_{\mathcal{F}} \\
&= \left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot), \ldots \right\rangle_{\mathcal{F}} \\
&= \underbrace{\left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot), \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) \right\rangle_{\mathcal{F}}}_{(A)} \\
&+ \underbrace{\left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}}_{(B)=\frac{d}{dx}\frac{d}{dx'} k(x-x')\big|_{x=x'=z}} \\
&+ 2 \underbrace{\left\langle \left( \frac{d}{dx} \log p(x) \right) k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}}_{(C)}
\end{aligned}
$$

# Next step: taking expectations

Compute the squared norm:

$$
\begin{aligned}
\|\xi_z\|^2_{\mathcal{F}} &= \langle \xi_z, \xi_z \rangle_{\mathcal{F}} \\
&= \left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) + \frac{d}{dz} k(z,\cdot), \ldots \right\rangle_{\mathcal{F}} \\
&= \underbrace{\left\langle \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot), \left( \frac{d}{dz} \log p(z) \right) k(z,\cdot) \right\rangle_{\mathcal{F}}}_{(A)} \\
&+ \underbrace{\left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}}_{(B) = \frac{d}{dx}\frac{d}{dx'} k(x-x') \big|_{x=x'=z}} \\
&+ 2 \underbrace{\left\langle \left( \frac{d}{dx} \log p(x) \right) k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}}_{(C)}
\end{aligned}
$$

# First two (easy) terms

First term (A):

$$(A) = \left\langle \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot), \left(\frac{d}{dz}\log p(z)\right) k(z,\cdot) \right\rangle_{\mathcal{F}}$$

$$= \left[ \left(\frac{d}{dz}\log p(z)\right)^2 \underbrace{k(z,z)}_{=c} \right]$$

# First two (easy) terms

Second term (B):

$$(B) = \left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left[-\imath\ell \hat{k}_\ell \exp(-\imath\ell x)\right] \overline{\left[-\imath\ell \hat{k}_\ell \exp(-\imath\ell x')\right]}}{\hat{k}_\ell} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} -(\imath\ell)^2 \hat{k}_\ell \underbrace{\exp\left(\imath\ell(x'-x)\right)}_{=1 \text{ when } x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \ell^2 \hat{k}_\ell =: C > 0$$

# First two (easy) terms

Second term (B):

$$(B) = \left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left[-\imath\ell \hat{k}_\ell \exp(-\imath\ell x)\right] \overline{\left[-\imath\ell \cancel{\hat{k}_\ell} \exp(-\imath\ell x')\right]}}{\cancel{\hat{k}_\ell}} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} -(\imath\ell)^2 \hat{k}_\ell \underbrace{\exp\left(\imath\ell(x'-x)\right)}_{=1 \text{ when } x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \ell^2 \hat{k}_\ell =: C > 0$$

# First two (easy) terms

Second term (B):

$$(B) = \left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \frac{\left[-\imath\ell \hat{k}_\ell \exp(-\imath\ell x)\right] \overline{\left[-\imath\ell \cancel{\hat{k}_\ell} \exp(-\imath\ell x')\right]}}{\cancel{\hat{k}_\ell}} \Bigg|_{x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} -(\imath\ell)^2 \hat{k}_\ell \underbrace{\exp\left(\imath\ell(x'-x)\right)}_{=1 \text{ when } x=x'=z}$$

$$= \sum_{\ell=-\infty}^{\infty} \ell^2 \hat{k}_\ell =: C > 0$$

# First two (easy) terms

Second term (B):

$$
\begin{aligned}
(B) &= \left\langle \frac{d}{dx} k(x,\cdot), \frac{d}{dx'} k(x',\cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z} \\
&= \sum_{\ell=-\infty}^{\infty} \frac{\left[-\imath\ell \hat{k}_\ell \exp(-\imath\ell x)\right] \overline{\left[-\imath\ell \cancel{\hat{k}_\ell} \exp(-\imath\ell x')\right]}}{\cancel{\hat{k}_\ell}} \Bigg|_{x=x'=z} \\
&= \sum_{\ell=-\infty}^{\infty} -(\imath\ell)^2 \hat{k}_\ell \underbrace{\exp\left(\imath\ell(x'-x)\right)}_{=1 \text{ when } x=x'=z} \\
&= \sum_{\ell=-\infty}^{\infty} \ell^2 \hat{k}_\ell =: C > 0
\end{aligned}
$$

# Third (slightly harder) term

Third term (C):

$$(C) = \left\langle \left( \frac{d}{dx} \log p(x) \right) k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} \frac{\left[ \hat{k}_\ell \exp(-\imath \ell x) \right] \overline{\left[ (-\imath \ell) \hat{k}_\ell \exp(-\imath \ell x') \right]}}{\hat{k}_\ell} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} (\imath \ell) \hat{k}_\ell \underbrace{\exp\left( \imath \ell (x' - x) \right)}_{=1 \text{ when } x=x'}$$

$$= 0.$$

# Third (slightly harder) term

Third term (C):

$$(C) = \left\langle \left( \frac{d}{dx} \log p(x) \right) k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} \frac{\left[ \hat{k}_\ell \exp(-\imath \ell x) \right] \overline{\left[ (-\imath \ell) \hat{k}_\ell \exp(-\imath \ell x') \right]}}{\hat{k}_\ell} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} (\imath \ell) \hat{k}_\ell \underbrace{\exp\left( \imath \ell (x' - x) \right)}_{=1 \text{ when } x=x'}$$

$$= 0.$$

# Third (slightly harder) term

Third term (C):

$$(C) = \left\langle \left( \frac{d}{dx} \log p(x) \right) k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} \frac{\left[ \hat{k}_\ell \exp(-\imath \ell x) \right] \overline{\left[ (-\imath \ell) \cancel{\hat{k}_\ell} \exp(-\imath \ell x') \right]}}{\cancel{\hat{k}_\ell}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} (\imath \ell) \hat{k}_\ell \underbrace{\exp\left(\imath \ell (x' - x)\right)}_{=1 \text{ when } x=x'}$$

$$= 0.$$

# Third (slightly harder) term

Third term (C):

$$(C) = \left\langle \left( \frac{d}{dx} \log p(x) \right) k(x, \cdot), \frac{d}{dx'} k(x', \cdot) \right\rangle_{\mathcal{F}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} \frac{\left[\hat{k}_\ell \exp(-\imath \ell x)\right] \overline{\left[(-\imath \ell) \cancel{\hat{k}_\ell} \exp(-\imath \ell x')\right]}}{\cancel{\hat{k}_\ell}} \Bigg|_{x=x'=z}$$

$$= \left( \frac{d}{dz} \log p(z) \right) \sum_{\ell=-\infty}^{\infty} (\imath \ell) \hat{k}_\ell \underbrace{\exp\left(\imath \ell (x' - x)\right)}_{=1 \text{ when } x=x'}$$

$$= 0.$$

# Putting it all together

We found:

$$\|\xi_z\|_{\mathcal{F}}^2 = C + \left(\frac{d}{dz}\log p(z)\right)^2 c,$$

Thus for boundedness, we have the condition:

$$E_{z\sim q}\|\xi_z\|_{\mathcal{F}} = E_{z\sim q}\sqrt{C + \left(\frac{d}{dx}\log p(x)\right)^2 c}$$
$$\leq \sqrt{E_{z\sim q}\left[C + \left(\frac{d}{dz}\log p(z)\right)^2 c\right]},$$

So Riesz holds when $E_{z\sim q}\left(\frac{d}{dz}\log p(z)\right)^2 < \infty$

# Putting it all together

We found:

$$\|\xi_z\|_{\mathcal{F}}^2 = C + \left(\frac{d}{dz}\log p(z)\right)^2 c,$$

Thus for boundedness, we have the condition:

$$E_{z\sim q}\|\xi_z\|_{\mathcal{F}} = E_{z\sim q}\sqrt{C + \left(\frac{d}{dx}\log p(x)\right)^2 c}$$
$$\leq \sqrt{E_{z\sim q}\left[C + \left(\frac{d}{dz}\log p(z)\right)^2 c\right]},$$

So Riesz holds when $E_{z\sim q}\left(\frac{d}{dz}\log p(z)\right)^2 < \infty$

# Kernel stein discrepancy

Closed-form expression for KSD: given independent $z, z' \sim q$, then
(Chwialkowski, Strathmann, G., ICML 2016) (Liu, Lee, Jordan ICML 2016)

$$\begin{aligned} \mathrm{KSD}(p, q, \mathcal{F}) &= \sup_{\|g\|_{\mathcal{F}} \leq 1} E_{z\sim q}\left([T_p g](z)\right) \\ &= \sup_{\|g\|_{\mathcal{F}} \leq 1} E_{z\sim q} \langle g, \xi_z \rangle_{\mathcal{F}} \\ &= \sup_{\|g\|_{\mathcal{F}} \leq 1} \langle g, E_{z\sim q}\xi_z \rangle_{\mathcal{F}} = \|E_{z\sim q}\xi_z\|_{\mathcal{F}} \end{aligned}$$

Test statistic:

$$\|E_{z\sim q}\xi_z\|_{\mathcal{F}}^2 = E_{z,z'\sim q} h_p(z, z')$$

where

$$\begin{aligned} h_p(x, y) := &\, \partial_x \log p(x) \partial_y \log p(y) k(x, y) \\ &+ \partial_y \log p(y) \partial_x k(x, y) + \partial_x \log p(x) \partial_y k(x, y) \\ &+ \partial_x \partial_y k(x, y) \end{aligned}$$

Do not need to normalize $p$, or sample from it.

# Kernel stein discrepancy

Closed-form expression for KSD: given independent $z, z' \sim q$, then

(Chwialkowski, Strathmann, G., ICML 2016) (Liu, Lee, Jordan ICML 2016)

$$\begin{aligned}
\mathrm{KSD}(p, q, \mathcal{F}) &= \sup_{\|g\|_{\mathcal{F}} \le 1} E_{z\sim q}\left([T_p g](z)\right) \\
&= \sup_{\|g\|_{\mathcal{F}} \le 1} E_{z\sim q} \langle g, \xi_z \rangle_{\mathcal{F}} \\
&= \sup_{\|g\|_{\mathcal{F}} \le 1} \langle g, E_{z\sim q}\xi_z \rangle_{\mathcal{F}} = \|E_{z\sim q}\xi_z\|_{\mathcal{F}}
\end{aligned}$$

Test statistic:

$$\|E_{z\sim q}\xi_z\|_{\mathcal{F}}^2 = E_{z,z'\sim q} h_p(z, z')$$

where

$$\begin{aligned}
h_p(x, y) := &\ \partial_x \log p(x) \partial_y \log p(y) k(x, y) \\
&+ \partial_y \log p(y) \partial_x k(x, y) + \partial_x \log p(x) \partial_y k(x, y) \\
&+ \partial_x \partial_y k(x, y)
\end{aligned}$$

Do not need to normalize $p$, or sample from it.

# Constructing threshold for a statistical test

Given samples $\{z_i\}_{i=1}^n \sim q$, empirical KSD (test statistic) is:

$$\widehat{\text{KSD}}(p, q, \mathcal{F}) := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} h_p(z_i, z_j).$$

When $q = p$, obtain estimate of null distribution with wild bootstrap:

$$\widetilde{KSD}(p, q, \mathcal{F}) := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sigma_i \sigma_j h_p(z_i, z_j).$$

where $\{\sigma_i\}_{i=1}^n$ i.i.d, $E(\sigma_i) = 0$, and $E(\sigma_i^2) = 1$

- Consistent estimate of the null distribtion when $q = p$
- Consistent test (Type II error goes to zero) under a rich class of alternatives Chwialkowski, Strathmann, G., ICML 2016

# Does the Riesz condition matter?

Consider the standard normal,

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right).$$

Then

$$\frac{d}{dx} \log p(x) = -x.$$

If $q$ is a Cauchy distribution, then the integral

$$E_{z\sim q}\left(\frac{d}{dz}\log p(z)\right)^2 = \int_{-\infty}^{\infty} z^2 q(z)\, dz$$

is undefined.

# Does the Riesz condition matter?

Consider the standard normal,

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right).$$

Then

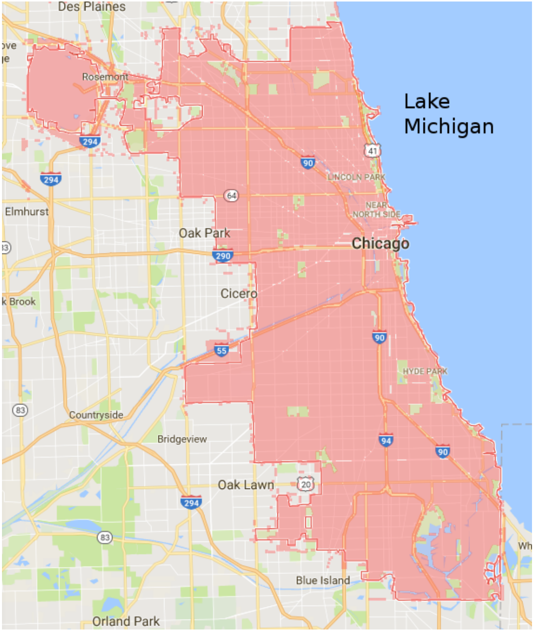
$$\frac{d}{dx} \log p(x) = -x.$$

If $q$ is a Cauchy distribution, then the integral

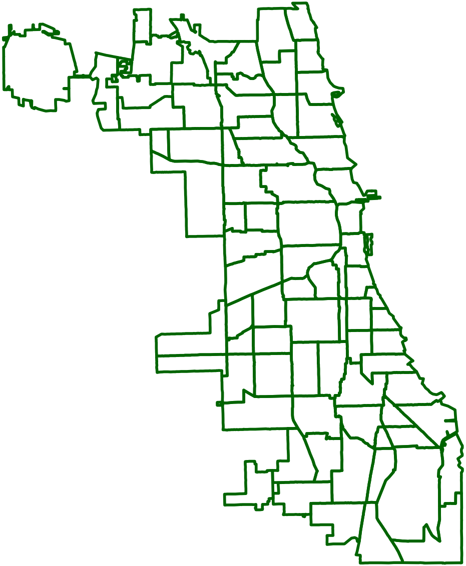
$$E_{z\sim q}\left(\frac{d}{dz}\log p(z)\right)^2 = \int_{-\infty}^{\infty} z^2 q(z)\, dz$$

is undefined.

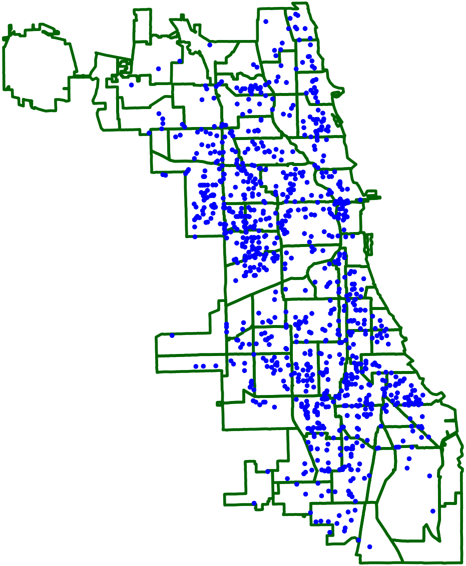
# Model Criticism



# Model Criticism

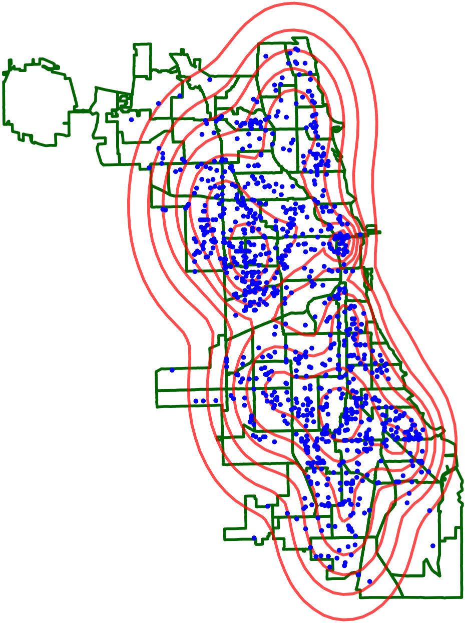


# Model Criticism



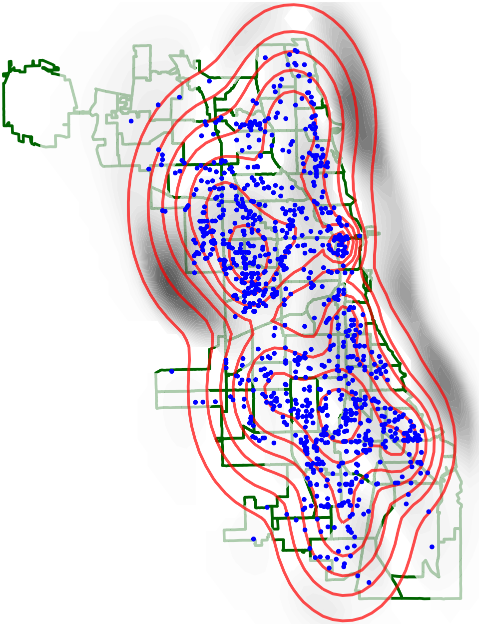
Data = robbery events in Chicago in 2016.

# The witness function: Chicago Crime



Model $p$ = 10-component Gaussian mixture.

# The witness function: Chicago Crime



Witness function $g$ shows mismatch

# Kernel stein discrepancy

## Further applications:

- Evaluation of approximate MCMC methods.
  (Chwialkowski, Strathmann, G., ICML 2016; Gorham, Mackey, ICML 2017)

## What kernel to use?

- The inverse multiquadric kernel,

$$k(x, y) = \left(c + \|x - y\|_2^2\right)^\beta$$

for $\beta \in (-1, 0)$.

arXiv.org > stat > arXiv:1703.01717

Statistics > Machine Learning

**Measuring Sample Quality with Kernels**

Jackson Gorham, Lester Mackey ICML 2017

*(Submitted on 6 Mar 2017 (v1), last revised 3 Aug 2017 (this version, v6))*