Stein method for quadratic forms

Rui Wang

Thursday 20th December, 2018

1 haha2

Theorem 1. *Let* $\zeta_1, \ldots, \zeta_d$ *be iid random variables with mean* 0 *and variance* 1*, and assume* $\mu_k := \mathrm{E}(\zeta_1^k)$ *is finite for* $k \leq 8$*. Let* $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_d)^\top \in \mathbb{R}^d$*. For* $k = 1, \ldots, K$*, let* $\mathbf{Q}_k = (q_{i_j}^{(k)})$ *be a* $d \times d$ *symmetric matrix and let* $\check{\mathbf{Q}}_k = \mathrm{diag}(q_{11}^{(k)}, \ldots, q_{dd}^{(k)})$*,* $\hat{\mathbf{Q}}_k = \mathbf{I}_d - \check{\mathbf{Q}}_k$*. Define* $\hat{w}_k = \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_k \boldsymbol{\zeta}$*,* $\check{w}_k = \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_k \boldsymbol{\zeta} - \mathrm{tr}(\mathbf{Q}_k)$*, and*

$$W = \begin{pmatrix} \hat{w}_1 \\ \check{w}_1 \\ \vdots \\ \hat{w}_K \\ \check{w}_K \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_1 \boldsymbol{\zeta} \\ \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_1 \boldsymbol{\zeta} - \mathrm{tr}(\mathbf{Q}_1) \\ \vdots \\ \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_1 \boldsymbol{\zeta} \\ \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_1 \boldsymbol{\zeta} - \mathrm{tr}(\mathbf{Q}_1) \end{pmatrix} \in \mathbb{R}^{2K}.$$

Finally, let $Z \sim \mathcal{N}_{2K}(0, \mathbf{I}_{2K})$ *and* $\mathbf{V} = \mathrm{Cov}(W)$*. There is an absolute constant* $0 < C < \infty$ *such that*

$$haha$$

Proof. Let $f : \mathbb{R}^{2K} \to \mathbb{R}$ be a four-times differentiable function. From xxx, there is a $4 - times$ differentiable function $g : \mathbb{R}^{2K} \to \mathbb{R}$ satisfying the Stein identity

$$\mathrm{E}[f(W)] - \mathrm{E}[f(\mathbf{V}^{1/2} W)] = \mathrm{E}[\nabla^\top \mathbf{V} \nabla g(W) - W^\top \nabla g(W)]$$

and

$$\left| \frac{\partial^k g(\mathbf{x})}{\prod_{j=1}^k \partial x_{i_j}} \right| \leq \frac{1}{k} \left| \frac{\partial^k f(\mathbf{x})}{\prod_{j=1}^k \partial x_{i_j}} \right| \quad \text{for all } \mathbf{x} = (x_1, \ldots, x_{2K})^\top \in \mathbb{R}^{2K},\ k = 1, 2, 3, \text{ and } i_j \in \{1, \ldots, 2K\}.$$

To prove the theorem, we bound

$$S = \mathrm{E}[\nabla^\top \mathbf{V} \nabla g(W) - W^\top \nabla g(W)].$$

Next, we use exchangeability. Let $\boldsymbol{\zeta}' = (\zeta_1', \ldots, \zeta_d')^\top$ be an independent copy of ζ, and let $\underline{i} \in \{1, \ldots, d\}$ be an independent and uniformly distributed random index. Define the vector

$W' \in \mathbb{R}^{2K}$ exactly as we defined W, except that $\boldsymbol{\zeta}_{\underline{i}}$ is replaced with $\boldsymbol{\zeta}'_{\underline{i}}$ throughout. More precisely, let $e_i \in \mathbb{R}^d$ be the ith standard basis vector in $\mathcal{R}^d$ and define

$$\begin{aligned}\hat{w}'_k =&(\boldsymbol{\zeta} + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}})^\top \hat{\mathbf{Q}}_k(\boldsymbol{\zeta} + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}})\\ =&\hat{w}_k + 2(\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}}^\top \hat{\mathbf{Q}}_k\boldsymbol{\zeta},\end{aligned}$$

$$\begin{aligned}\check{w}'_k =&(\boldsymbol{\zeta} + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}})^\top \check{\mathbf{Q}}_k(\boldsymbol{\zeta} + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}}) - \mathrm{tr}(\mathbf{Q}_k)\\ =&\check{w}_k + e_{\underline{i}}^\top \check{\mathbf{Q}}_k e_{\underline{i}}((\zeta'_{\underline{i}})^2 - \zeta_{\underline{i}}^2),\end{aligned}$$

for $k = 1, \dots, K$. Then $W' = (\hat{w}'_1, \check{w}'_1, \dots, \hat{w}'_K, \check{w}'_K)^\top \in \mathbb{R}^{2K}$. Its straightforward to verify that

$$\mathrm{E}(\hat{w}'_k - \hat{w}_k|\boldsymbol{\zeta}) = -\frac{2}{d}\hat{w}_k, \quad \mathrm{E}(\check{w}'_k - \check{w}_k|\boldsymbol{\zeta}) = -\frac{1}{d}\check{w}_k.$$

Then

$$\mathrm{E}(W' - W|\boldsymbol{\zeta}) = -\Lambda_K W,$$

where

$$\Lambda_1 = \begin{pmatrix} \frac{2}{d} & 0 \\ 0 & \frac{1}{d} \end{pmatrix}, \quad \Lambda_K = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Lambda_1 \end{pmatrix} \in \mathbb{R}^{2K \times 2K}.$$

By exchangeability, we have

$$\begin{aligned}0 =&\frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') + \nabla g(W))]\\ =&\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}\nabla g(W)] + \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))]\\ =&-\mathrm{E}[W^\top \nabla g(W)] + \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))].\end{aligned}$$

That is,

$$\mathrm{E}[W^\top \nabla g(W)] = \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))].$$

Apply Taylor's theorem,

$$\begin{aligned}&W^\top \nabla g(W)\\ =&\frac{1}{2}\sum_{i,j=1}^{2K} \Lambda_{K,ii}^{-1} D^{ij} g(W)(w'_i - w_i)(w'_j - w_j) + \frac{1}{4}\sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W)(w'_i - w_i)(w'_j - w_j)(w'_k - w_k)\\ &+ \frac{1}{12}\sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W)(w'_i - w_i)(w'_j - w_j)(w'_k - w_k)(w'_l - w_l)\\ =&\frac{1}{2}\,\mathrm{tr}[(W' - W)(W' - W)^\top \Lambda_K^{-\top}\nabla^2 g(W)] + \frac{1}{4}\sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W)(w'_i - w_i)(w'_j - w_j)(w'_k - w_k)\\ &+ \frac{1}{12}\sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W)(w'_i - w_i)(w'_j - w_j)(w'_k - w_k)(w'_l - w_l),\end{aligned} \tag{1}$$

where $t^* \in [0,1]$. Also by exchangeability,

$$\mathrm{E}[(W'-W)(W'-W)^\top] = 2\,\mathrm{E}[W(W-W')^\top] = 2\,\mathrm{E}[WW^\top\Lambda_K^\top] = 2\mathbf{V}\Lambda_K^\top.$$

It follows that

$$\mathrm{E}[\nabla^\top\mathbf{V}\nabla g(W)] = \mathrm{E}\,\mathrm{tr}[\mathbf{V}\nabla^2 g(W)] = \frac{1}{2}\,\mathrm{E}\,\mathrm{tr}[\mathrm{E}[(W'-W)(W'-W)^\top]\Lambda_K^{-\top}\nabla^2 g(W)]$$

Thus,

$$\begin{aligned}
S =& \mathrm{E}[\nabla^\top\mathbf{V}\nabla g(W) - W^\top\nabla g(W)] \\
=& \frac{1}{2}\,\mathrm{E}\,\mathrm{tr}[\mathrm{E}[(W'-W)(W'-W)^\top]\Lambda_K^{-\top}\nabla^2 g(W)] - \frac{1}{2}\,\mathrm{E}\,\mathrm{tr}[(W'-W)(W'-W)^\top\Lambda_K^{-\top}\nabla^2 g(W)] \\
& - \frac{1}{4}\,\mathrm{E}\sum_{i,j,k=1}^{2K}\Lambda_{K,ii}^{-1}D^{ijk}g(W)(w_i'-w_i)(w_j'-w_j)(w_k'-w_k) \\
& - \frac{1}{12}\,\mathrm{E}\sum_{i,j,k,l=1}^{2K}\Lambda_{K,ii}^{-1}D^{ijkl}g(t^*(W'-W)+W)(w_i'-w_i)(w_j'-w_j)(w_k'-w_k)(w_l'-w_l).
\end{aligned}$$

$\square$