Instructions for Paper Submissions to AISTATS 2021: Supplementary Materials

1 FORMATTING INSTRUCTIONS

To prepare a supplementary pdf file, we ask the authors to use `aistats2021.sty` as a style file and to follow the same formatting instructions as in the main paper. The only difference is that the supplementary material must be in a *single-column* format. You can use `supplement.tex` in our starter pack as a starting point, or append the supplementary content to the main paper and split the final PDF into two separate files.

Note that reviewers are under no obligation to examine your supplementary material.

2 Derivations of density corrections

2.1 Simplex softmax parameterization

We define the transformation $\phi : \mathbb{R}^{n-1} \to \Delta^n : z \mapsto x_-$, where $x = \begin{pmatrix} x_- \\ \frac{1}{r} \end{pmatrix}$, $x_i = \frac{1}{r} e^{z_i}$ for $i \in \{1, \dots, n-1\}$, and $r = 1 + \sum_{i=1}^{n-1} e^{z_i}$.

First we compute the scalar derivatives:

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}z_j} &= e^{z_j} = r x_j \\ \frac{\mathrm{d}x_i}{\mathrm{d}z_j} &= \delta_{ij} \frac{1}{r} e^{z_i} - \frac{1}{r^2} e^{z_i} \frac{\mathrm{d}r}{\mathrm{d}z_j} = \delta_{ij} x_i - x_i x_j, \quad 1 \le i \le n-1 \end{aligned}$$

where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ is the Kronecker delta.

If $\operatorname{diag}(x)$ is the diagonal matrix whose diagonal are the elements of x, then the Jacobian is

$$J = (I_{n-1} - x_- \mathbf{1}_{n-1}^\top) \operatorname{diag}(x_-),$$

where $\mathbf{1}_n$ is the n-vector of ones.

Using Sylvester's determinant theorem, $|I_{n-1} - x_- \mathbf{1}_{n-1}^\top| = 1 - \mathbf{1}_{n-1}^\top x_- = 1 - \sum_{i=1}^{n-1} x_i = x_n$, so

$$\text{correction} = |J| = x_n \prod_{i=1}^{n-1} x_i = \prod_{i=1}^{n} x_i = \exp\left(\sum_{i=1}^{n-1} z_i\right) \left(1 + \sum_{i=1}^{n-1} e^{z_i}\right)^{-n}$$

2.2 Simplex augmented softmax parameterization

We define the transformation $\phi : \mathbb{R}^n \backslash \{0\} \to \Delta^{n-1} \times \mathbb{R}_{>0} : z \mapsto (x, r)$, where $r = \sum_{i=1}^n e^{z_i}$ and $x_i = \frac{1}{r} e^{z_i}$.

Similar to above, to compute the density correction we use the bijective transformation $f : z \mapsto (x_-, r)$. First we compute the scalar derivatives:

$$\begin{aligned} \frac{\mathrm{d}r}{\mathrm{d}z_j} &= e^{z_j} = r x_j \\ \frac{\mathrm{d}x_i}{\mathrm{d}z_j} &= \delta_{ij} \frac{1}{r} e^{z_i} - \frac{1}{r^2} e^{z_i} \frac{\mathrm{d}r}{\mathrm{d}z_j} = \delta_{ij} x_i - x_i x_j, \end{aligned}$$

which has the Jacobian

$$J = \begin{pmatrix} I_{n-1} - x_- \mathbf{1}_{n-1}^\top & -x_- \\ r\mathbf{1}_{n-1}^\top & r \end{pmatrix} \mathrm{diag}(x),$$

For invertible A, the determinant of the block matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is $|A||D - CA^{-1}B|$. A square matrix is invertible iff its determinant is non-zero. From the previous section, $|I_{n-1} - x_- \mathbf{1}_{n-1}^\top| = x_n > 0$, so the determinant of the Jacobian is

$$|J| = x_n \left| r + r\mathbf{1}_{n-1}^\top (I_{n-1} - x_- \mathbf{1}_{n-1}^\top)^{-1} x_- \right| \prod_{i=1}^n x_i$$

Let $w = (I_{n-1} - x_- \mathbf{1}_{n-1}^\top)^{-1} x_-$. Then,

$$\begin{aligned} w - x_- \sum_{i=1}^{n-1} w_i &= x_- \\ w &= x_- \left(1 - \sum_{i=1}^{n-1} w_i\right) \\ \sum_{i=1}^{n-1} w_i &= \sum_{i=1}^{n-1} \left(x_- (1 - \sum_{i=1}^{n-1} w_i)\right) = \left(\sum_{i=1}^{n-1} x_i\right)\left(1 - \sum_{i=1}^{n-1} w_i\right) = (1 - x_n)\left(1 - \sum_{i=1}^{n-1} w_i\right) \\ \sum_{i=1}^{n-1} w_i &= \frac{1 - x_n}{x_n} = \frac{1}{x_n} - 1 \\ w &= x_- \left(1 - \left(\frac{1}{x_n} - 1\right)\right) = \frac{1}{x_n} x_- \end{aligned}$$

Then

$$|J| = x_n r \left| 1 + \frac{1}{x_n} \sum_{i=1}^{n-1} x_i \right| \prod_{i=1}^n x_i = r \prod_{i=1}^n x_i$$

To keep the target distribution proper, we must select a prior distribution $\pi(r)$ for r. If we choose $r \sim \chi_n$, then the product of the correction and the density of the prior for r is proportional to

$$\text{correction} = \pi(r)|J| = r^n e^{-r^2/2} \prod_{i=1}^n x_i = \exp\left(\sum_{i=1}^n z_i - \frac{1}{2}\left(\sum_{i=1}^n e^{z_i}\right)^2\right).$$