Reparametrization trick Notes

Federico Errica
federico.errica@phd.unipi.it

March 6, 2020

Please contact me if you find errors or have doubts. There is always room for improvement and learning.

Problem Definition

Assume we have a distribution $q_\phi(\mathbf{z}|\mathbf{x})$, where ϕ are free parameters, with its own *pdf* $h_\phi(x)$. It follows that its *CDF* $H_\phi(x) = P(X \leq x)$ is absolutely continuous. Therefore we can write

$$\frac{\mathrm{d}}{\mathrm{d}x} H_\phi(x) = h_\phi(x) \tag{1}$$

Given $\varepsilon \sim P(\varepsilon)$ and $\mathbf{z} = g_\phi(\varepsilon, \mathbf{x})$, we want to find "situations" in which these 2 expectations coincide

$$\int q_\phi(\mathbf{z}|\mathbf{x}) f(\mathbf{z}) \mathrm{d}\mathbf{z} = \int P(\varepsilon) f(\mathbf{z}) \mathrm{d}\varepsilon \tag{2}$$

by carefully choosing the family of distributions q, P and the function g. By performing the change of variable from left to right $\mathbf{z} = g_\phi(\varepsilon, \mathbf{x})$ (note that $\mathbf{x}$ is fixed) we have

$$\int q_\phi(\mathbf{z}|\mathbf{x}) f(\mathbf{z}) \mathrm{d}\mathbf{z} = \int q_\phi(g_\phi(\varepsilon, \mathbf{x})|\mathbf{x}) f(g_\phi(\varepsilon, \mathbf{x})) g'_\phi(\varepsilon, \mathbf{x}) \mathrm{d}\varepsilon \tag{3}$$

At this point it suffices to show that

$$q_\phi(g_\phi(\varepsilon, \mathbf{x})|\mathbf{x}) g'_\phi(\varepsilon, \mathbf{x}) = P(\varepsilon) \tag{4}$$

Case 1: (Tractable) Inverse CDF

Assume the inverse *CDF* $H^{-1}_{\phi,x}(u)$ of $q_\phi(\mathbf{z}|\mathbf{x})$ exists and is tractable. If we choose

$$g_\phi(\varepsilon, \mathbf{x}) = H^{-1}_{\phi,x}(\epsilon) \tag{5}$$

we can show that eq. (4) holds. Thanks to eq. (1) we can write

$$q_\phi(g_\phi(\varepsilon, \mathbf{x})|\mathbf{x}) = \frac{\mathrm{d}}{\mathrm{d}g_\phi(\varepsilon, \mathbf{x})} H(g_\phi(\varepsilon, \mathbf{x})) = \frac{\mathrm{d}\varepsilon}{\mathrm{d}g_\phi(\varepsilon, \mathbf{x})} \tag{6}$$

Where we have used the fact that g_ϕ is the inverse of H. Therefore

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}g_\phi(\varepsilon, \mathbf{x})} \frac{\mathrm{d}g_\phi(\varepsilon, \mathbf{x})}{\mathrm{d}\varepsilon} = \mathbf{1} = P(\epsilon). \tag{7}$$

We choose $\epsilon \sim Unif[\mathbf{0}, \mathbf{I}]$, so that $P(\epsilon) = \mathbf{1}$.

Case 2: Family of location-scale distributions

I want to start from $P(\varepsilon)$ this time. Assume

$$\varepsilon \sim LocScale(location = \mathbf{0}, scale = \mathbf{I}) \tag{8}$$

For example $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Assume that q belongs to the location-scale family with *location* $= \mu$ and *scale* $= \sigma$. We can construct a $g(\varepsilon, \mathbf{x})$ such that eq. (4) holds (by transforming the left-hand side of the equation in the right-hand side). In other words

$$LocScale_{\mu,\sigma}(g_{\mu,\sigma}(\varepsilon, \mathbf{x})|\mathbf{x}) g'_{\mu,\sigma}(\varepsilon, \mathbf{x}) = LogScale_{\mathbf{0},\mathbf{I}}(\varepsilon) \tag{9}$$

We know that, when X is a random variable with zero mean and unit variance, another random variable Y can be espressed as

$$Y \overset{d}{=} \mu_Y + \sigma_Y X \tag{10}$$

Furthermore, their *pdfs* are related in this way

$$f_Y(y) = f_X(\frac{(y - \mu_Y)}{\sigma_Y X}) \frac{1}{\sigma_Y} \tag{11}$$

Therefore, if we define $\mathbf{z} = g_{\mu,\sigma}(\varepsilon, \mathbf{x}) = \mu + \sigma\varepsilon$ we have (using eq. (11))

$$\begin{aligned} LocScale_{\mu,\sigma}(\mathbf{z}|\mathbf{x})\sigma &= LogScale_{\mathbf{0},\mathbf{I}}(\frac{(\mathbf{z} - \mu)}{\sigma}) \\ LocScale_{\mu,\sigma}(\mathbf{z}|\mathbf{x}) &= LogScale_{\mathbf{0},\mathbf{I}}(\varepsilon)\frac{1}{\sigma} \end{aligned} \tag{12}$$

where we have used the fact that $g'_{\mu,\sigma}(\varepsilon, \mathbf{x}) = \sigma$.

1 Integral solution, Appendix B

We want to solve this integral, where $\mathbf{z}, \mu, \sigma \in \mathbb{R}^J$

$$\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2\mathbf{I}) \log \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I}) d\mathbf{z} \tag{13}$$

By definition

$$\begin{aligned}
&\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2\mathbf{I})\big(-\frac{1}{2}\mathbf{z}^T\mathbf{z} - \log\sqrt{(2\pi)^J}\big)d\mathbf{z} \\
&= -\frac{1}{2}\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2\mathbf{I})||\mathbf{z}||^2 d\mathbf{z} - \log\sqrt{(2\pi)^J}\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2\mathbf{I})d\mathbf{z}
\end{aligned}$$

Since the right integral sums to 1 we end up with

$$-\frac{1}{2\sqrt{(2\pi)^J det(\sigma^2\mathbf{I})}}\int e^{-\frac{1}{2}(\mathbf{z}-\mu)^T(\sigma^2\mathbf{I})^{-1}(\mathbf{z}-\mu)}||\mathbf{z}||^2 d\mathbf{z} - \frac{J}{2}\log 2\pi \tag{14}$$

Now we work out the integral that has been left, and we will plug the result back in eq. (14). But before we do that, I think a recap of some known integrals is useful.

- The Gaussian integral $\int e^{-a(x+b)^2} = \sqrt{\frac{\pi}{a}}$
- $\int xe^{-x^2} = 0$ (since xe^{-x^2} is an odd function it suffices to show that the integral from 0 to ∞ is finite, in particular it is $\frac{1}{2}$)

Also, notice that the function inside the integral in eq. (14) is non-negative, therefore we can apply the Fubini-Tonelli's theorem and compute the integral one z_i at a time.

$$\begin{aligned}
&\int e^{-\frac{1}{2}(\mathbf{z}-\mu)^T(\sigma^2\mathbf{I})^{-1}(\mathbf{z}-\mu)}||\mathbf{z}||^2 d\mathbf{z} \\
&= \int\int..\int e^{-\frac{1}{2}(\mathbf{z}-\mu)^T(\sigma^2\mathbf{I})^{-1}(\mathbf{z}-\mu)}||\mathbf{z}||^2 dz_1 dz_2..dz_J \\
&= \int\int..\int e^{-\frac{1}{2}(\mathbf{z}-\mu)^T(\sigma^2\mathbf{I})^{-1}(\mathbf{z}-\mu)}(\sum_k^J z_k^2) dz_1 dz_2..dz_J
\end{aligned}$$

we can rewrite the exponent as a product of exponents, one for each component z_i. Then we can carry out the constants in this way

$$= \int e^{-\frac{1}{2}\frac{(z_J-\mu_J)^2}{\sigma_J^2}}\int e^{-\frac{1}{2}\frac{(z_{J-1}-\mu_{J-1})^2}{\sigma_{J-1}^2}}..\int e^{-\frac{1}{2}\frac{(z_1-\mu_1)^2}{\sigma_1^2}}(\sum_k^J z_k^2) dz_1 dz_2..dz_J \tag{15}$$

Eq. (15) can be written as a sum of different multiple integrals that differ from the z_i^2 component. One can write, for a specific index k of the summation over $\mathbf{z}$'s components

$$\begin{aligned}
&= \int e^{-\frac{1}{2}\frac{(z_k-\mu_k)^2}{\sigma_k^2}} z_k^2 \int e^{-\frac{1}{2}\frac{(z_J-\mu_J)^2}{\sigma_J^2}} .. \int e^{-\frac{1}{2}\frac{(z_1-\mu_1)^2}{\sigma_1^2}} dz_1..dz_J dz_k \\
&= \sqrt{(2\pi)^{J-1} det([\sigma_1^2, , \sigma_{k-1}^2, \sigma_k^2, , \sigma_J^2]^T \mathbf{I})} \int e^{-\frac{1}{2}\frac{(z_k-\mu_k)^2}{\sigma_k^2}} z_k^2 dz_k
\end{aligned} \tag{16}$$

where we have applied the Gaussian integral multiple times and multplied the results all together (in particular, the determinant is equal to $\prod_i \sigma_i^2$). We are left with (dropping indexes to ease the notation)

$$\int e^{-\frac{1}{2}\frac{(z-\mu)^2}{\sigma^2}} z^2 dz \tag{17}$$

We perform a change of variable, $z - \mu = y$, $\frac{dy}{dz} = 1$, $z = y + \mu$ then

$$\begin{aligned}
\int e^{-\frac{1}{2}\frac{(z-\mu)^2}{\sigma^2}} z^2 dz &= \int e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} (y+\mu)^2 dy \\
&= \int y^2 e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy + \mu^2 \int e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy + \cancel{2\mu \int y e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy} \\
&= \int y^2 e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy + \mu^2 \sqrt{2\pi\sigma^2} dy
\end{aligned} \tag{18}$$

the last term has been cancelled because of one of the second property mentioned above. The very last integral is solved with integration by parts

$$\begin{aligned}
&\int y y e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy \\
&= \cancel{-\sigma^2 [y e^{-\frac{1}{2}\frac{y^2}{\sigma^2}}]_{-\infty}^{\infty}} - \int -\sigma^2 e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy \\
&= \sigma^2 \sqrt{2\pi\sigma^2}
\end{aligned} \tag{19}$$

Plugging the result back in eq. (18) yields (with index k)

$$(\sigma_k^2 + \mu_k^2)\sqrt{2\pi\sigma^2} \tag{20}$$

Eq. (16) becomes

$$\sqrt{(2\pi^J) det(\sigma^2 \mathbf{I})} (\sigma_k^2 + \mu_k^2) \tag{21}$$

Again, Eq.(15) was the sum over k of eq. (16). Therefore eq (15) can be written as

$$\sqrt{(2\pi^J) det(\sigma^2 \mathbf{I})} \sum_{k}^{J} (\sigma_k^2 + \mu_k^2) \tag{22}$$

Finally, plugging this back into eq. (14) yields the final result

$$\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2 \mathbf{I}) \log \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I}) d\mathbf{z} = -\frac{J}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} (\mu_j^2 + \sigma_j^2) \tag{23}$$

1.1 Another integral

Similarly (easier!) we can also show that

$$\int \mathcal{N}(\mathbf{z}; \mu, \sigma^2 \mathbf{I}) \log \mathcal{N}(\mathbf{z}; \mu, \sigma^2 \mathbf{I}) d\mathbf{z} = -\frac{J}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} (1 + \sigma_j^2) \tag{24}$$