Transform then pool or pool and then transform?

Luiz Max F. de Carvalho

September 23, 2018

Abstract

In this note I claim to give a proof that the order of operations between transforming and pooling a set of distributions does not matter **if and only if** the transform in question is invertible. I also give an example where the transform-then-pool and pool-then-transform with a non-invertible transform lead to distributions in the same family which are nonetheless distinct.

Key-words: logarithmic pooling; invertible; log-normal.

Background

Logarithmic pooling is a popular method for combining opinions on an agreed quantity, specially when these opinions can be framed as probability distributions. Let $\mathbf{F}_\theta := \{f_0(\theta), f_1(\theta), \dots, f_K(\theta)\}$ be a set of distributions representing the opinions of $K+1$ experts and let $\boldsymbol\alpha := \{\alpha_0, \alpha_1, \dots, \alpha_K\} \in \mathcal{S}^K$ be the vector of weights, such that $\alpha_i > 0\, \forall i$ and $\sum_{i=0}^K \alpha_i = 1$, i.e., $\mathcal{S}^{K+1}$ is the space of all open simplices of dimension $K+1$. The **logarithmic pooling operator** $\mathcal{LP}(\mathbf{F}_\theta, \boldsymbol\alpha)$ is defined as

$$\mathcal{LP}(\mathbf{F}_\theta, \boldsymbol\alpha) := \pi(\theta|\boldsymbol\alpha) = t(\boldsymbol\alpha) \prod_{i=0}^{K} f_i(\theta)^{\alpha_i}, \tag{1}$$

where $t(\boldsymbol\alpha) = \int_{\boldsymbol\Theta} \prod_{i=0}^K f_i(\theta)^{\alpha_i} d\theta$. This pooling method enjoys several desirable properties and yields tractable distributions for a large class of distribution families (Genest et al., 1984).

Pool then transform or transform and then pool?

Definition 1. *Let $A, B \subseteq \mathbb{R}^p$. A function $h : A \to B$ is **invertible** iff $\exists\, h^{-1} : B \to A$ with $h^{-1}(h(a)) = a \,\forall\, a \in A$. Let π_A be an arbitrary probability measure in A. If h is monotonic and differentiable we can write $\pi_B(B) = \pi(h^{-1}(A))|J|$, where $|J|$ is the absolute determinant of the Jacobian matrix with entries $J_{ik} := \partial h_k^{-1}/\partial a_i$, $i, k = 1, 2, \dots, p$.*

Suppose the we are interested in the distribution of a random variable $Y \in \mathcal{Y} \subseteq \mathbb{R}^q$ when one has a random variable $X \in \mathcal{X} \subseteq \mathbb{R}^p$ with $\phi : \mathcal{X} \to \mathcal{Y}$. Let $|J_\phi|$ be the Jacobian determinant w.r.t. ϕ. Suppose further that each expert i produces a distribution $f_i(X)$ such that we can construct the object $\mathbf{F}_X = \{f_1(X), f_2(X), \dots, f_K(X)\}$. Then one can either:

(a) **Pool-then-transform:** construct $\pi_X(X|\boldsymbol\alpha) = \mathcal{LP}(\mathbf{F}_X, \boldsymbol\alpha)$ and then apply ϕ to obtain $\pi_Y(Y|\boldsymbol\alpha) := \pi_X(\phi^{-1}(Y)|\boldsymbol\alpha)|J_\phi|$;

(b) **Transform-then-pool:** apply the transform to each component i of $\mathbf{F}_X$ to build

$$\mathbf{G}_Y := \{g_i(Y), g_2(Y), \dots, g_K(Y)\}$$

and obtain $\pi'_Y(Y|\boldsymbol\alpha) = \mathcal{LP}(\mathbf{G}_Y, \boldsymbol\alpha)$.

Remark 1. *If ϕ is invertible, then $\pi_Y(Y|\boldsymbol\alpha) \equiv \pi'_Y(Y|\boldsymbol\alpha)$.*

Proof. First,

$$\pi_Y(y|\boldsymbol\alpha) \propto \pi_X(\phi^{-1}(y))|J_\phi|, \tag{2}$$

$$= \prod_{i=0}^{K} \left[f_i(\phi^{-1}(y))\right]^{\alpha_i} |J_\phi|. \tag{3}$$

For situation (b) we have:

$$g_i(y) = f_i(\phi^{-1}(y))|J_\phi|. \tag{4}$$

And,

$$\pi'_Y(y|\boldsymbol{\alpha}) \propto \prod_{i=0}^{K} g_i(y)^{\alpha_i} \tag{5}$$
$$= \prod_{i=0}^{K} \left[f_i(\phi_x^{-1}(y))|J_\phi|\right]^{\alpha_i} \tag{6}$$
$$= \prod_{i=0}^{K} \left[f_i(\phi^{-1}(y))\right]^{\alpha_i} |J_\phi|, \tag{7}$$

as claimed. □

An interesting idea is whether Remark 1 is an iff result. Let $\eta : \mathcal{X} \to \mathcal{Y}$ be a surjective non-injective differentiable function, which is not invertible on the whole of $\mathcal{Y}$, but instead is **piece-wise invertible**. Let $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ be a partition of $\mathcal{Y}$, i.e., $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, $\forall i \neq j \in \{1, 2, \ldots, T\}$ and $\bigcup_{t=1}^{T} \mathcal{Y}_t = \mathcal{Y}$. Then define the inverse functions $\eta_t^{-1}(y) : \mathcal{Y}_t \to \mathcal{X}$, $t \in \{1, 2, \ldots, T\}$. Lastly, let $|J_t|$ be the Jacobian of $\eta_t^{-1}(\cdot)$. Then we are prepared to write:

$$\pi_Y(y|\boldsymbol{\alpha}) \propto \sum_{t=1}^{T} \left(\prod_{i=0}^{K} f_i(\eta_t^{-1}(y))^{\alpha_i} \right) |J_t| \quad \text{and} \tag{8}$$
$$\pi'_Y(y|\boldsymbol{\alpha}) \propto \prod_{i=0}^{K} \left[\sum_{t=1}^{T} f_i(\eta_t^{-1}(y))|J_t| \right]^{\alpha_i} \tag{9}$$

which, I claim, will only be equal if $T = 1$, i.e. if $\eta(\cdot)$ is invertible in the usual sense. Below I try to establish a result in general for any surjective non-injective mapping, not just piece-wise invertible ones.

Remark 2. *$\pi_Y(Y|\boldsymbol{\alpha}) \equiv \pi'_Y(Y|\boldsymbol{\alpha})$ if and only if ϕ is invertible.*

Proof. In general, we can define $\Omega(y) := \{x : \eta(x) = y\}$ and thus[1]

$$g_i(y) = \sum_{x \in \Omega(y)} f_i(x). \tag{10}$$

Notice that for any $x \in \mathcal{X}$, there exists y and $\Omega(y)$ such that $x \in \Omega(y)$. Assume $|\Omega(y)| > 1$ for some $y \in \mathcal{Y}$ and $f_i \not\equiv f_j$, for some i, j. We have

$$\pi_Y(y|\boldsymbol{\alpha}) \propto \sum_{x \in \Omega(y)} \left(\prod_{i=0}^{K} f_i(x)^{\alpha_i} \right) \quad \text{and} \tag{11}$$
$$\pi'_Y(y|\boldsymbol{\alpha}) \propto \prod_{i=0}^{K} \left[\sum_{x \in \Omega(y)} f_i(x) \right]^{\alpha_i}. \tag{12}$$

Define

$$T = \int_{\mathcal{Y}} \sum_{x \in \Omega(y)} \left(\prod_{i=0}^{K} f_i(x)^{\alpha_i} \right) dy \quad \text{and} \quad T' = \int_{\mathcal{Y}} \prod_{i=0}^{K} \left[\sum_{x \in \Omega(y)} f_i(x) \right]^{\alpha_i} dy.$$

It is not hard to show[2] that $T = \left(\int_{\mathcal{X}} \prod_{i=0}^{K} f_i(x)^{\alpha_i} dx \right)^{-1}$. Since for any given y and $x \in \Omega(y)$ we have $\sum_{\omega \in \Omega(y)} f_i(\omega) > f_i(x)$ *a.e.*, it follows that $T' < T$ and hence the densities would have different normalising constants, which is impossible if $\pi_Y(y|\boldsymbol{\alpha}) = \pi'_Y(y|\boldsymbol{\alpha}) \, \forall \, y \in \mathcal{Y}$. □

There probably exists a measure-theoretic proof that is way more elegant, but this should suffice.

[1] Notice there is no guarantee that $|\Omega(y)| < \infty$.

[2] This hinges on the fact that $\int_{\mathcal{Y}} g_i(y)dy = 1 \, \forall \, i$.

An example

Suppose $Z = U/V$ and each expert i elicits $U \sim \text{log-normal}(\mu_{iU}, \sigma^2_{iU})$, $V \sim \text{log-normal}(\mu_{iV}, \sigma^2_{iV})$, i.e.

$$\begin{aligned} f_{iU}(u|\mu_{iU}, \sigma^2_{iU}) &= \frac{1}{u\sqrt{2\pi\sigma^2_{iU}}} \exp\left(-\frac{(\ln u - \mu_{iU})^2}{2\sigma^2_{iU}}\right), \\ g_{iV}(v|\mu_{iV}, \sigma^2_{iV}) &= \frac{1}{v\sqrt{2\pi\sigma^2_{iV}}} \exp\left(-\frac{(\ln v - \mu_{iV})^2}{2\sigma^2_{iV}}\right). \end{aligned}$$

Again, let $\mathbf{F}_U = \{f_{1U}(U), f_{2U}(U), \ldots, f_{KU}(U)\}$ and $\mathbf{G}_V = \{g_{1V}(V), g_{2V}(V), \ldots, g_{KV}(V)\}$. First, let us derive $\pi_Z(Z)$ under scheme (a). It is not hard to show that $\pi_U(U|\boldsymbol{\alpha}) := \mathcal{LP}(\mathbf{F}_U, \boldsymbol{\alpha}) = \text{log-normal}(\mu^*_U, v^*_U)$ with

$$\begin{aligned} \mu^*_U &:= \frac{\sum_{i=0}^K w_{iU}\mu_{iU}}{\sum_{i=0}^K w_{iU}}, && (13) \\ v^*_U &:= \frac{1}{\sum_{i=0}^K w_{iU}}, && (14) \\ w_{iU} &:= \frac{\alpha_i}{\sigma^2_{iU}}. && (15) \end{aligned}$$

See our paper for a proof. Analagously, $\pi_V(V|\boldsymbol{\alpha}) := \mathcal{LP}(\mathbf{G}_V, \boldsymbol{\alpha}) = \text{log-normal}(\mu^*_V, v^*_V)$. Then $\pi_Z(Z|\boldsymbol{\alpha}) = \text{log-normal}(\mu^*_Z, v^*_Z)$, with

$$\begin{aligned} \mu^*_Z &= \mu^*_U - \mu^*_V, \\ &= \frac{\sum_{i=0}^K w_{iU}\mu_{iU}}{\sum_{i=0}^K w_{iU}} - \frac{\sum_{i=0}^K w_{iV}\mu_{iV}}{\sum_{i=0}^K w_{iV}} \quad \text{and} && (16) \\ v^*_Z &= v^*_U + v^*_V, \\ &= \frac{1}{\sum_{i=0}^K w_{iU}} + \frac{1}{\sum_{i=0}^K w_{iV}}. && (17) \end{aligned}$$

Now let us consider case (b). Since $r_{iZ} = \text{log-normal}(\mu_{iU} - \mu_{iV}, \sigma^2_{iU} + \sigma^2_{iV})$, we arrive at $\pi'_Z(Z|\boldsymbol{\alpha}) = \text{log-normal}(\mu^{**}_Z, v^{**}_Z)$,

$$\begin{aligned} \mu^{**}_Z &:= \frac{\sum_{i=0}^K w_{iZ}\mu_{iU}}{\sum_{i=0}^K w_{iZ}} - \frac{\sum_{i=0}^K w_{iZ}\mu_{iV}}{\sum_{i=0}^K w_{iZ}}, && (18) \\ v^{**}_Z &:= \frac{1}{\sum_{i=0}^K w_{iZ}}, && (19) \\ w_{iZ} &:= \frac{\alpha_i}{\sigma^2_{iU} + \sigma^2_{iV}}. && (20) \end{aligned}$$

Clearly, $v^*_Z \leq v^{**}_Z$ and hence $\mu^{**}_Z \leq \mu^*_Z \;\forall\, \boldsymbol{\alpha}$.

Minimising Kullback-Leibler divergence in transformed space

One might argue that procedure (b) makes little sense, given that the set of opinions $\mathbf{F}_X$ concerns only X, i.e, it was not necessarily constructed taking the transformation $\phi(\cdot)$ into account. An example is a situation where experts are asked to provide distributions on the probability p of a particular event. In general, elicitation for $f_i(p)$ will not take into account the induced distribution on the log-odds, $\phi(p) = \log p/(1-p)$. Nevertheless, the decision-maker may wish to assign the weights $\boldsymbol{\alpha}$ in a way that takes $\phi(\cdot)$ into account, e.g., by giving lower weights to experts whose distributions on the log-odds scale are unreasonable.

This decision process can be made more precise. In a similar spirit to the paper, one can construct $\boldsymbol{\alpha}$ so as to minimise the Kullback-Leibler divergence between each distribution in $\mathbf{F}_{\mathbf{y}}^{-1}$ and a transformation of the distribution obtained by procedure (a), $\pi_Y(y|\boldsymbol{\alpha}) = \pi_\theta(\phi^{-1}(y)|\boldsymbol{\alpha})|J_\phi|$. Let $d_i = \text{KL}(h_i(y)||\pi_Y(y|\boldsymbol{\alpha}))$.

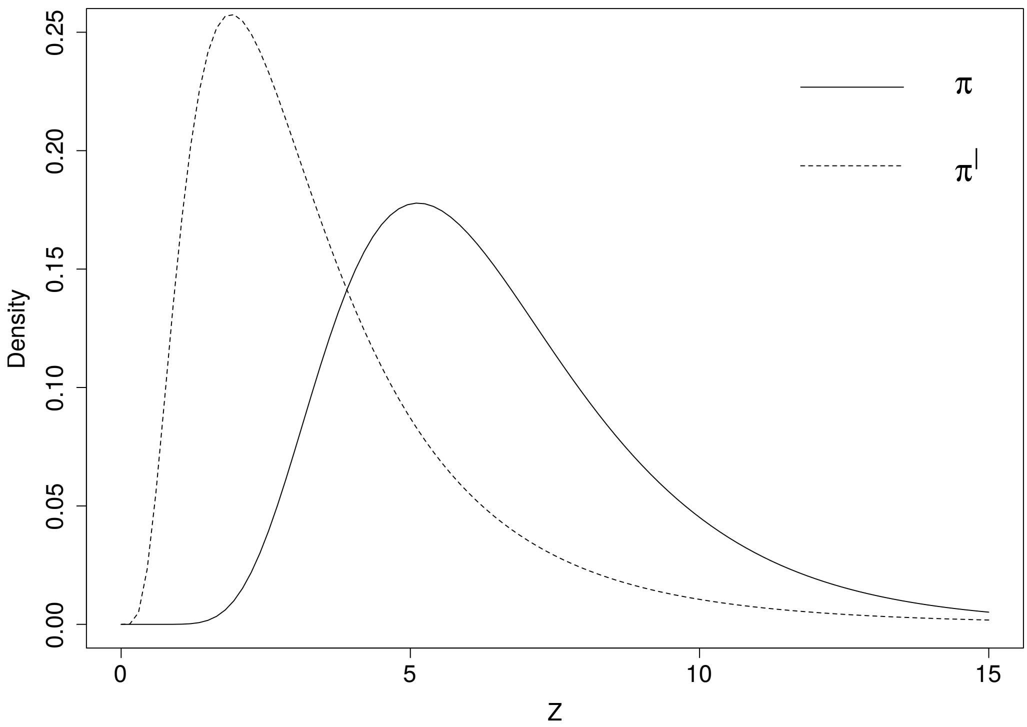

Figure 1: **Log-normal example**. Solid line displays $\pi_Z(Z|\boldsymbol{\alpha})$, obtained by first pooling the distributions on U and V and then computing the induced distribution on Z. Dashed line displays the logarithmic pooling of individual distributions r_{iZ}, $\pi'_Z(Z|\boldsymbol{\alpha})$. For this example $K = 2$, $\alpha_0 = 0.70$, $\mu_{0U} = 0.80$, $\sigma^2_{0U} = 0.40$, $\mu_{1U} = 0.5$, $\sigma^2_{1U} = 0.05$, $\mu_{0V} = -1.60$, $\sigma^2_{0V} = 0.024$, $\mu_{1V} = -1.25$ and $\sigma^2_{1V} = 0.4$.

We then aim at solving the problem

$$\begin{aligned} L(\boldsymbol{\alpha}) &= \sum_{i=0}^{K} d_i \\ \hat{\boldsymbol{\alpha}} &:= \arg\min L(\boldsymbol{\alpha}) \end{aligned} \tag{21}$$

This procedure therefore choses weights for each expert by how coherent the prior provided by each expert is with the pool-then-Transform – procedure (a) – prior in the transformed space induced by $\phi(\cdot)$.

Acknowledgements

I am grateful to Mike West (Duke) for not being impressed about the invertible case and prompting me to look at it in more detail.

References

Genest, C., Weerahandi, S., and Zidek, J. V. (1984). Aggregating opinions through logarithmic pooling. *Theory and Decision*, 17(1):61–70.