Transform then pool or pool and then transform?

Luiz Max F. de Carvalho

September 20, 2019

Abstract

In this note I discuss the order of operations between transforming and pooling a set of distributions in relation to whether the transform in question is invertible. I give an explicit example where both procedures lead to the same density even when the transform is not invertible and also discuss 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. The lesson to be learned is that while the order of operations usually matters, it may not, depending on the distributions under consideration.

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

TODO

- Get rid of all abuse of notation. The statement $\phi(\mathcal{LP}(\mathbf{F_X}, \boldsymbol{\alpha})) = \pi_Y(y|\boldsymbol{\alpha})$ is an abomination.

- Clearly establish the conditions under which $\phi()$ produces a pushforward measure that admits Radon-Nikodym derivatives. Without this, it becomes meaningless to talk about "induce-then-pool"; you can only pool proper probability densities and thus one needs to guarantee said densities in fact exist.

- Maybe formulate a measure-theoretic version of LP to deal with the difficulties above.

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 $\mathbf{F}_\theta := \{f_0(\theta), f_1(\theta), \ldots, f_K(\theta)\}$ be a set of distributions representing the opinions of $K + 1$ experts and let $\boldsymbol{\alpha} := \{\alpha_0, \alpha_1, \ldots, \alpha_K\} \in \mathcal{S}^K$ 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 $t(\boldsymbol{\alpha}) = \left[\int_{\boldsymbol{\Theta}} \prod_{i=0}^K f_i(\theta)^{\alpha_i} d\theta\right]^{-1}$. This pooling method enjoys several desirable properties and yields tractable distributions for a large class of distribution families (Genest et al., 1984, 1986; Carvalho et al., 2019).

Pool then transform or transform and then pool?

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 (absolute) 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 $\mathbf{F}_X = \{f_1(X), f_2(X), \ldots, f_K(X)\}$. 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

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

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

First, let us start with a couple definitions

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 $\pi_B(B) = \pi_A(h^{-1}(A))|J|$, where $|J|$ is the absolute determinant of the Jacobian matrix with entries $J_{ik} := \partial h_k^{-1}/\partial a_i$, $i, k = 1, 2, \ldots, p$.*

Definition 2. *Let f and g be probability density functions with support on $\mathcal{X} \subseteq \mathbb{R}^d$. We say $f \equiv g$ if $f(x) = g(x)$ for almost every point $x \in \mathcal{X}$, that is, almost everywhere, a.e. This is equivalent to stating that $\mu(A) = 0$ where $A := \{x \in \mathcal{X} : f(x) \neq g(x)\}$ and μ is the dominating measure.*

Remark 1. *If ϕ is invertible, then $\pi_Y \equiv \pi_Y'$.*

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. □

Corollary 1. *The log-pooled distribution from a collection of inverse-Gamma densities with parameters η_i, β_i is inverse-Gamma with parameters $\eta^\star = \sum_{i=0}^{K} \alpha_i \eta_i$ and $\beta^\star = \sum_{i=0}^{K} \alpha_i \beta_i$.*

Proof. First, recall that if $X \sim \text{Gamma}(a, b)$ then $1/X \sim \text{Inverse-Gamma}(a, b)$. Carvalho et al. (2019) show that if the f_i are Gamma densities with parameters η_i and β_i (shape/rate parametrisation) then the pooled distribution is a Gamma distribution with parameters $\eta^\star = \sum_{i=0}^{K} \alpha_i \eta_i$ and $\beta^\star = \sum_{i=0}^{K} \alpha_i \beta_i$. Since $\phi(x) = 1/x$ is invertible, the result follows from Remark 1. □

While Corollary 1 is straightforward to get by direct calculation, our goal here is to illustrate how Remark 1 could be used in situations where direct calculation is unwieldly. An interesting idea is whether Remark 1 is an "if and only if" (iff) result. As we shall see, the answer is no (Remark 2).

Remark 2. *It is possible to have $\pi_Y \equiv \pi_Y'$ even when ϕ is not invertible.*

Proof. We will show this by way of an explicit example. In general, we can define

$$g_i(y) = \sum_{\omega \in \Omega(y)} \frac{f_i(\omega)}{|\phi'(\omega)|}. \tag{8}$$

where $\Omega(y) := \{x : \phi(x) = y\}$ [1]. Notice that for any $x \in \mathcal{X}$, there exists y such that $x \in \Omega(y)$. Assume that the transform ϕ is non-invertible, i.e., that $|\Omega(y)| > 1$ for some $y \in \mathcal{Y}$, and $f_i \not\equiv f_j$, for some i, j. We

---

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

have

$$\pi_Y(y|\boldsymbol{\alpha}) = T(\boldsymbol{\alpha}) \sum_{\omega \in \Omega(y)} \frac{\prod_{i=0}^{K} f_i(\omega)^{\alpha_i}}{|\phi'(\omega)|} \quad \text{and} \tag{9}$$

$$\pi_Y'(y|\boldsymbol{\alpha}) = T'(\boldsymbol{\alpha}) \prod_{i=0}^{K} g_i(y)^{\alpha_i} = T'(\boldsymbol{\alpha}) \prod_{i=0}^{K} \left[\sum_{\omega \in \Omega(y)} \frac{f_i(\omega)}{|\phi'(\omega)|} \right]^{\alpha_i}, \tag{10}$$

with

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

Now let us construct an explicit example. Let X be a quantity of interest and let $\boldsymbol{F}_X$ be such that each f_i is a Gaussian density with zero mean[2] and variance v_i, i.e.,

$$f_i(x) = \frac{1}{\sqrt{2\pi v_i}} \exp\left(-\frac{x^2}{2v_i}\right)$$

Now consider $\phi(x) = x^2$ and $Y = \phi(X)$. Clearly, $\Omega(y) = \{\omega_0, \omega_1\} = \{-\sqrt{y}, \sqrt{y}\}$ and hence

$$g_i(y) = \frac{f_i(\omega_0)}{|2\omega_0|} + \frac{f_i(\omega_1)}{|2\omega_1|}, \tag{11}$$

$$= \frac{f_i(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi v_i y}} \exp\left(-\frac{y}{2v_i}\right). \tag{12}$$

where the second line follows by using the symmetry of the Gaussian density around zero. We are now prepared to write

$$\pi_Y(y|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi v^* y}} \exp\left(-\frac{y}{2v^*}\right). \tag{13}$$

whence $v^* = \left[\sum_{i=0}^{K} \frac{\alpha_i}{v_i}\right]^{-1}$. This calculation is very similar to the pool of log-normals – see the Appendix of Carvalho et al. (2019) for the derivation. An explicit computation gives π_Y':

$$\pi_Y'(y|\boldsymbol{\alpha}) \propto \prod_{i=0}^{K} g_i(y)^{\alpha_i}, \tag{14}$$

$$\propto \prod_{i=0}^{K} \left(\frac{1}{\sqrt{2\pi v_i y}} \right)^{\alpha_i} \exp\left(-\frac{\alpha_i y}{2v_i}\right), \tag{15}$$

$$= \frac{1}{\sqrt{2\pi v^* y}} \exp\left(-\frac{y}{2v^*}\right). \tag{16}$$

which establishes $\pi_Y(y|\boldsymbol{\alpha}) = \pi_Y'(y|\boldsymbol{\alpha})$ for any $y \in \mathcal{Y}$. $\qquad \square$

See Figure 1 for graphical depiction of the example given above. We could have chosen any distribution on $(-\infty, \infty)$ symmetric about 0. This result shows that it is indeed possible to obtain the same density even when ϕ is not one-to-one.

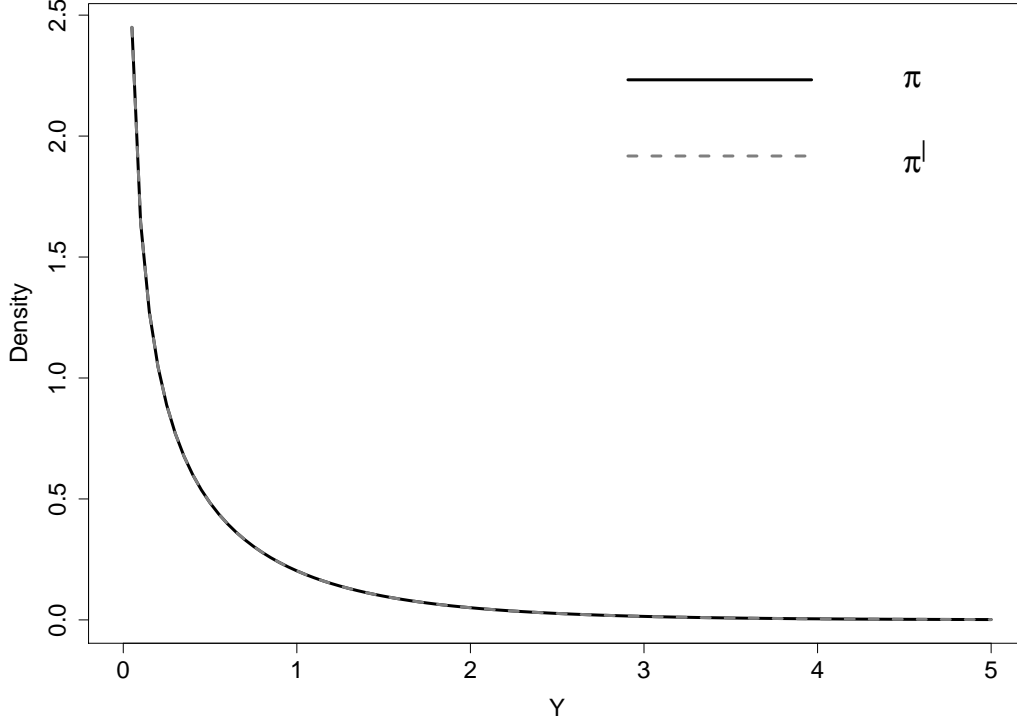
[2]This is important so we can employ a symmetry argument later on.

Figure 1: **Equivalent densities under a non-invertible transform**. Solid line displays $\pi_Y(Y|\boldsymbol{\alpha})$, obtained by first pooling the distributions on X and then computing the induced distribution on Y. Dashed line displays the logarithmic pooling of individual distributions g_i, $\pi'_Y(Y|\boldsymbol{\alpha})$. For this example $K = 5$, $\boldsymbol{\alpha} = \{0.31, 0.08, 0.39, 0.08, 0.15\}$, $\boldsymbol{v} = \{0.25, 0.75, 1.00, 1.00, 0.50\}$.

An example where the two approaches are NOT the same

The order of operations, i.e., whether one adopts procedure (a) or procedure (b), will likely affect the resulting densities, however. In this section we will discuss an example where both procedures lead to densities in the same family, namely the log-normal family of distributions, which are nonetheless distinct.

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

$$f_{iU}(u|\mu_{iU}, \sigma_{iU}^2) = \frac{1}{u\sqrt{2\pi\sigma_{iU}^2}} \exp\left(-\frac{(\ln u - \mu_{iU})^2}{2\sigma_{iU}^2}\right),$$

$$g_{iV}(v|\mu_{iV}, \sigma_{iV}^2) = \frac{1}{v\sqrt{2\pi\sigma_{iV}^2}} \exp\left(-\frac{(\ln v - \mu_{iV})^2}{2\sigma_{iV}^2}\right).$$

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 $\pi_U(U|\boldsymbol{\alpha}) := \mathcal{LP}(\mathbf{F}_U, \boldsymbol{\alpha}) =$ log-normal(μ_U^*, v_U^*) with

$$\mu_U^* := \frac{\sum_{i=0}^{K} w_{iU}\mu_{iU}}{\sum_{i=0}^{K} w_{iU}}, \tag{17}$$

$$v_U^* := \frac{1}{\sum_{i=0}^{K} w_{iU}}, \tag{18}$$

$$w_{iU} := \frac{\alpha_i}{\sigma_{iU}^2}. \tag{19}$$

4

See Carvalho et al. (2019) for a proof. Analogously, $\pi_V(V|\boldsymbol{\alpha}) := \mathcal{LP}(\mathbf{G}_V, \boldsymbol{\alpha}) = \text{log-normal}(\mu_V^*, v_V^*)$. Then $\pi_Z(Z|\boldsymbol{\alpha}) = \text{log-normal}(\mu_Z^*, v_Z^*)$, with

$$\mu_Z^* = \mu_U^* - \mu_V^*,$$

$$= \frac{\sum_{i=0}^K w_{iU}\mu_{iU}}{\sum_{i=0}^K w_{iU}} - \frac{\sum_{i=0}^K w_{iV}\mu_{iV}}{\sum_{i=0}^K w_{iV}} \quad \text{and} \tag{20}$$

$$v_Z^* = v_U^* + v_V^*,$$

$$= \frac{1}{\sum_{i=0}^K w_{iU}} + \frac{1}{\sum_{i=0}^K w_{iV}}. \tag{21}$$

Now let us consider case (b). Since $r_{iZ} = \text{log-normal}(\mu_{iU} - \mu_{iV}, \sigma_{iU}^2 + \sigma_{iV}^2)$, we arrive at $\pi_Z'(Z|\boldsymbol{\alpha}) = \text{log-normal}(\mu_Z^{**}, v_Z^{**})$,

$$\mu_Z^{**} := \frac{\sum_{i=0}^K w_{iZ}\mu_{iU}}{\sum_{i=0}^K w_{iZ}} - \frac{\sum_{i=0}^K w_{iZ}\mu_{iV}}{\sum_{i=0}^K w_{iZ}}, \tag{22}$$

$$v_Z^{**} := \frac{1}{\sum_{i=0}^K w_{iZ}}, \tag{23}$$

$$w_{iZ} := \frac{\alpha_i}{\sigma_{iU}^2 + \sigma_{iV}^2}. \tag{24}$$

Clearly, $v_Z^* \leq v_Z^{**}$ and hence $\mu_Z^{**} \leq \mu_Z^* \, \forall \, \boldsymbol{\alpha}$. [NOPE. THIS AIN'T TRUE AT ALL. NEEDS REVISION].
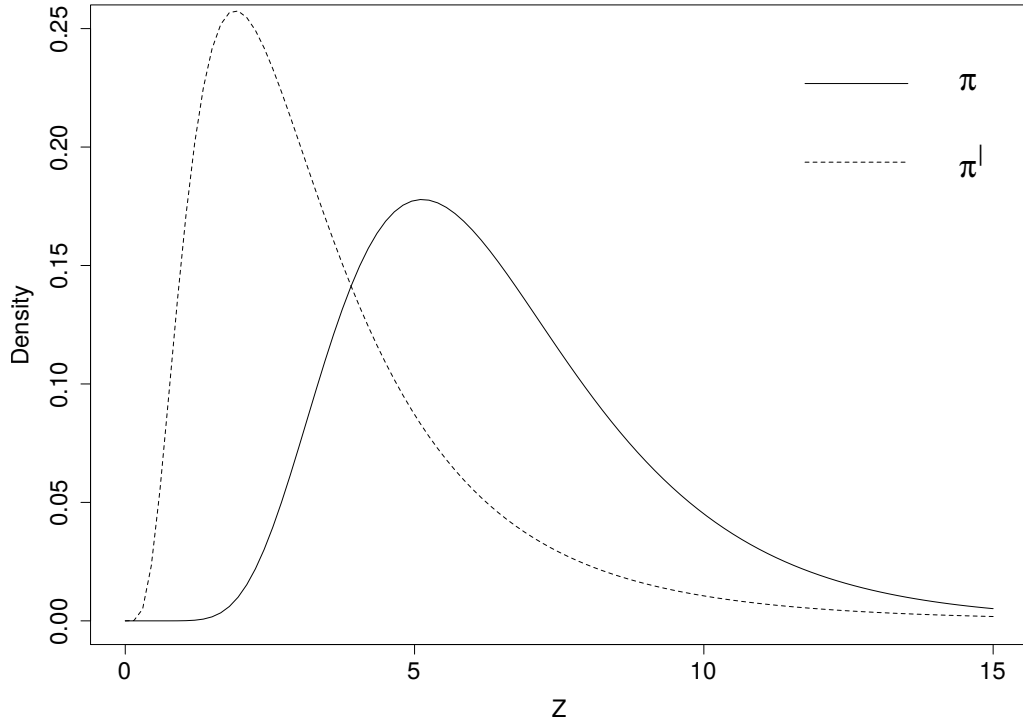

Figure 2: **Log-normal example**. Solid line displays $\pi_Z(Z|\boldsymbol{\alpha})$, obtained by first pooling the distributions on U and V and then computing the induced distribution on Z. Dashed line displays the logarithmic pooling of individual distributions r_{iZ}, $\pi_Z'(Z|\boldsymbol{\alpha})$.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

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, $\phi(p) = \log\left(p/(1-p)\right)$. 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 Carvalho et al. (2019), 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), $\phi(\mathcal{LP}(\mathbf{F}_\mathbf{X}, \boldsymbol{\alpha})) = \pi_Y(y|\boldsymbol{\alpha})$. We aim at solving the problem

$$L(\boldsymbol{\alpha}) = \sum_{i=0}^{K} \mathrm{KL}(\pi_Y || h_i), \tag{25}$$
$$\hat{\boldsymbol{\alpha}} := \arg\min L(\boldsymbol{\alpha})$$

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 $\phi(\cdot)$. When $\phi()$ is invertible, Remark 1 tells us that this procedure is the same as minimising the KL divergence between π_X and each f_i, since the Kullback-Leibler divergence is invariant under invertible differentiable transformations. Thus, in the probability/log-odds example mentioned above, the decision maker would not need to worry about the distributions on the log-odds induced by ϕ, since it is invertible.

Acknowledgements

References

Carvalho, L. M., Villela, D. V., Coelho, F. C., and Bastos, L. S. (2019). On the choice of weights for the logarithmic pooling of probability distributions. *In preparation*.

Genest, C., McConway, K. J., and Schervish, M. J. (1986). Characterization of externally bayesian pooling operators. *The Annals of Statistics*, pages 487–501.

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