

Andy Jones

Publications CV Technical blog

   ajones788@gmail.com

Scale mixtures of normals

Here, we discuss two distributions which arise as scale mixtures of normals: the Laplace and the Student- t .

Introduction

Mixture models are typically first introduced in the context of discrete mixtures. For example, the Gaussian mixture model (GMM) is often the canonical mixture model. In particular, if we assume x follows a GMM with K mixture components, we can write the model as follows.

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x; \mu_k, \sigma_k)$$

where π_1, \dots, π_K are the mixture weights that must satisfy $\sum_k \pi_k = 1$.

However, we can also consider continuous mixtures. Consider, for example, a continuous mixture of Gaussians defined by the following hierarchical model:

$$\begin{aligned} x &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &\sim \mathcal{N}(0, \tau^2) \end{aligned}$$

Now, we can write the marginal distribution of x :

$$p(x) = \int_{-\infty}^{\infty} p(x|\mu)p(\mu)d\mu$$

We call this a continuous mixture of Gaussians. Notice that this mixture has a similar form to the discrete mixture above. In particular, $p(\mu)$ here plays a similar role as $\pi_{k=1}^K$ above as the “mixing distribution”.

In the Gaussian setting, we can compute this marginal in closed form (see appendix for full derivation) to get:

$$x \sim \mathcal{N}(0, \tau^2 + \sigma^2)$$

Intuitively, we can think of the Gaussian prior on μ as adding extra variability to x , as compared to a model with a fixed μ .

Here, the mixing distribution was specified for the mean parameter μ , but we can also specify it for the variance σ^2 , as we’ll see next.

Scale mixtures

Consider the following hierarchical model:

$$\begin{aligned} x &\sim \mathcal{N}(\mu, \sigma^2) \\ \sigma^2 &\sim p(\sigma^2) \end{aligned}$$

Assuming a constant μ for now, the marginal distribution of x is then

$$p(x) = \int_0^{\infty} p(x|\sigma^2)p(\sigma^2)d\sigma^2$$

Below, we consider two choices for $p(\sigma^2)$ and discuss the implications for the marginal density of x .

Laplace

Consider placing an exponential prior on σ^2 such that $\sigma^2 \sim \text{Exp}(2\lambda^2)$. Recall that the PDF of the exponential distribution is

$$p(\sigma^2; \lambda) = 2\lambda^2 \exp(-2\lambda^2 \sigma^2)$$

We can then compute the marginal of x as follows:

$$\begin{aligned} p(x) &= \int_0^\infty p(x|\sigma^2) p(\sigma^2) d\sigma^2 \\ &= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} x^2\right) 2\lambda^2 \exp(-2\lambda^2 \sigma^2) d\sigma^2 \\ &= 2\lambda^2 \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-2\lambda^2 \sigma^2 - \frac{1}{2\sigma^2} x^2\right) d\sigma^2 \\ &= 2\lambda^2 \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{2\lambda^2}{\sigma^2} \left((\sigma^2)^2 + \frac{1}{4\lambda^2} x^2\right)\right) d\sigma^2 \\ &= 2\lambda^2 \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{2\lambda^2}{\sigma^2} \left(\sigma^2 - \frac{1}{2\lambda} x\right)^2 + 2\sigma^2 \frac{1}{2\lambda} x\right) d\sigma^2 \\ &= 2\lambda^2 \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{2\lambda^2}{\sigma^2} \left(\sigma^2 - \frac{x}{2\lambda}\right)^2 - 2\lambda x\right) d\sigma^2 \\ &= 2\lambda^2 \exp(-2\lambda x) \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{2\lambda^2 (x/2\lambda)^2}{\sigma^2 (x/2\lambda)^2} \left(\sigma^2 - \frac{x}{2\lambda}\right)^2\right) d\sigma^2 \\ &= 2\lambda^2 \exp(-2\lambda x) \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{2\lambda^2 \frac{x^2}{4\lambda^2}}{\sigma^2 \frac{x^2}{4\lambda^2}} \left(\sigma^2 - \frac{x}{2\lambda}\right)^2\right) d\sigma^2 \\ &= 2\lambda^2 \exp(-2\lambda x) \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2 (x/2\lambda)^2} \left(\sigma^2 - \frac{x}{2\lambda}\right)^2\right) d\sigma^2 \\ &= 2\lambda^2 \exp(-2\lambda x) \frac{1}{x} \int_0^\infty \sigma^2 \sqrt{\frac{x^2}{2\pi\sigma^3}} \exp\left(-\frac{x^2}{2\sigma^2 (x/2\lambda)^2} \left(\sigma^2 - \frac{x}{2\lambda}\right)^2\right) d\sigma^2 \end{aligned}$$

Now let $\tau := x^2$ and $\mu = \frac{x}{2\lambda}$.

$$p(x) = 2\lambda^2 \exp(-2\lambda x) \frac{1}{x} \int_0^\infty \sigma^2 \sqrt{\frac{\tau}{2\pi\sigma^3}} \exp\left(-\frac{\tau}{2\sigma^2 \mu^2} (\sigma^2 - \mu)^2\right) d\sigma^2$$

Recognizing the integrand (past the first σ^2) as a inverse-Gaussian distribution with mean μ and scale parameter τ , we can notice that the integral is an expectation of an inverse-Gaussian-distributed random variable σ^2 . Thus,

$$\begin{aligned} p(x) &= 2\lambda^2 \exp(-2\lambda x) \frac{1}{x} \mathbb{E}[\sigma^2] \\ &= 2\lambda^2 \exp(-2\lambda x) \frac{1}{x} \mu \\ &= 2\lambda^2 \exp(-2\lambda x) \frac{1}{x} \frac{x}{2\lambda} \\ &= \lambda \exp(-2\lambda x) \end{aligned}$$

Notice that this has the form of a Laplace distribution with scale parameter $b = \frac{1}{2\lambda}$. To be precise, I should have started using $|x|$ instead of x starting in line (14). Inserting that now, we have a real Laplace density: $p(x) = \lambda \exp(-2\lambda|x|)$.

Laplace priors (or, equivalently, a scale mixture of normals with exponential mixing distribution) are often used as a Bayesian analogue to the LASSO. Intuitively, the correspondence with the LASSO can be seen in the log-likelihood, which will penalize large values of $|x|$.

Student-t

Consider the following hierarchical model:

$$x \sim \mathcal{N}(0, \sigma^2)$$

$$\sigma^2 \sim \text{Inv-Gamma}(\nu/2, \nu/2)$$

We can again find the marginal distribution of σ^2 :

$$\begin{aligned}
 p(x) &= \int_0^\infty p(x|\sigma^2)p(\sigma^2)d\sigma^2 \\
 &= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} (\sigma^2)^{-\nu/2-1} \exp\left(-\frac{\nu/2}{\sigma^2}\right) d\sigma^2 \\
 &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{x^2 + \nu}{2\sigma^2}\right) (\sigma^2)^{-\frac{\nu+1}{2}-1} d\sigma^2 \\
 &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)\sqrt{2\pi}} \Gamma\left(\frac{\nu+1}{2}\right) \left(\frac{x^2 + \nu}{2}\right)^{-\frac{\nu+1}{2}}
 \end{aligned}$$

With a little more algebra, this can be put in the form of a Student-t distribution with ν degrees of freedom.

References

- Gelman, Andrew. “Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper).” Bayesian analysis 1.3 (2006): 515-534.
- Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott. “Handling sparsity via the horseshoe.” Artificial Intelligence and Statistics. 2009.
- Kenneth Tay’s [derivation of the Laplace distribution](#)
- John Cook’s [derivation of the Student-t](#)

Appendix

Derivation of marginal for continuous Gaussian mixture

$$\begin{aligned}
p(x) &= \int_{-\infty}^{\infty} p(x|\mu)p(\mu)d\mu \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{1}{2\tau^2}\mu^2\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 - \frac{1}{2\tau^2}\mu^2\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}x^2 - \frac{1}{\sigma^2}2x\mu + \frac{1}{\sigma^2}\mu^2 + \frac{1}{\tau^2}\mu^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\mu^2 - \frac{1}{\sigma^2}2x\mu + \frac{1}{\sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\mu^2 - \frac{1}{\sigma^2}2x\mu + \frac{1}{\sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}2x\mu + \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}2x\mu + \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right) + \left(\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right) - \left(\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}2x\mu + \left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2\right) - \left(\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right)\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\mu^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2\right) d\mu \\
&= \frac{1}{\sigma\tau 2\pi} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\frac{\tau^2}{\tau^2 + \sigma^2}x\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right) \sqrt{2\pi} \sqrt{\frac{\sigma^2\tau^2}{\tau^2 + \sigma^2}} \\
&= \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2)}} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\left(\frac{\tau^2}{\tau^2 + \sigma^2}\right)^2 x^2 - \frac{\tau^2}{\tau^2 + \sigma^2}x^2\right)\right) \\
&= \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2)}} \exp\left(-\frac{1}{2}\left(\frac{\tau^2 + \sigma^2}{\sigma^2\tau^2}\right)\left(\left(\left(\frac{\tau^2}{\tau^2 + \sigma^2}\right)^2 - \frac{\tau^2}{\tau^2 + \sigma^2}\right)x^2\right)\right) \\
&= \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2)}} \exp\left(-\frac{1}{2(\tau^2 + \sigma^2)}x^2\right)
\end{aligned}$$