11.1

Intro

In this exercise, we will derive Equation 11.61 (i.e. decompose the Student t distribution as an infinite mixture of Gaussians). We will restrict ourselves to the 1D case, for simplicity. Before starting, let's think about what this result tell us. The Student t distribution is an alternative to the Gaussian that handles outliers better. Nevertheless, when making the degrees of freedom ν much bigger than 5, we have essentialy a Gaussian again.

So, we know that the two distributions (Gaussian and Student) have some point in common. The Student distribution has a mean μ and a scale parameter Σ, just like the Gaussian. So if we could decompose the Student as a sum of Gaussians, it would make sense for them to share those parameters. But, we must remember that the Student handles outliers better than the Gaussian. So how about we share just μ and the **shape** of Σ? That is the nuts and bolts of how the mixture is built. We have an infinite mixture of Gaussian with the same mean and the same shape of covariance matrices. However, each covariance matrix is scaled by a factor z_i and the contribution of each component is weighted by a gamma distribution. Notice that the hyperparameter ν of our Gamma distribution will determine which scales will have more impact in the final distribution. In particular, for $\nu \leq 5$, we have relevant weights in the interval $z \in (0, 1)$ which is are the scales that can make $\frac{\Sigma}{z_i}$ handle better outliers that $\mathcal{N}(\mu, \Sigma)$! This explains why $\nu \gg 5$ makes the Student distribution as bad as the Gaussian to handle outliers.

Solution

We will start at the infinite mitxute, make the computations and show that it becomes the Student t distribution. First of all, notice that the infinite mixture (more formally known as compound distribution) is non-negative and normalized to 1. Now, we can ignore all the terms that do not depent on x and perform the calculations until we arrive at the functional form of the Student distribution. Let's start by the integrand.

$$\begin{aligned}
&\mathcal{N}(\mu, \frac{\sigma^2}{z}) = \frac{1}{Z_N} z \exp\left[-\frac{z}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \\
&Ga(z|\frac{\nu}{2}, \frac{\nu}{2}) = \frac{1}{Z_G} z^{\frac{\nu}{2}-1} \exp\left[-\frac{\nu}{2}z\right] \\
&\mathcal{N}(\mu, \frac{\sigma^2}{z}) Ga(z|\frac{\nu}{2}, \frac{\nu}{2}) = \frac{1}{Z} z^{\frac{\nu}{2}} \exp\left[-u(x)z\right] \\
&u(x) = \frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2} + \nu\right]
\end{aligned} \tag{1}$$

So our integral with respect to z becomes:

$$\int_0^\infty \frac{1}{Z} z^{\frac{\nu}{2}} \exp\left[-u(x)z\right] dz = \frac{1}{Z}\int_0^\infty z^{\frac{\nu}{2}} \exp\left[-u(x)z\right] dz \tag{2}$$

Now we have to solve the integral of the product of a monomial and an exponential. Fortunatly, thanks to the limits of the integral, this can be expressed as a gamma function.

$$
\begin{aligned}
&\int_0^\infty z^{\frac{\nu}{2}} \exp\left[-u(x)z\right] dz \\
&t = uz \\
&\int_0^\infty \left(\frac{t}{u}\right)^{\frac{\nu}{2}} \exp(-t)\frac{1}{u} dt = u(x)^{-\left(1+\frac{\nu}{2}\right)} \int_0^\infty t^{\frac{\nu}{2}} \exp(-t) dt = \\
&u(x)^{-\left(1+\frac{\nu}{2}\right)} \Gamma(\frac{\nu}{2}+1)
\end{aligned} \tag{3}
$$

Therefore, the total expression is given by:

$$
\begin{aligned}
&\frac{1}{Z}\Gamma(\frac{\nu}{2}+1)u(x)^{-\left(1+\frac{\nu}{2}\right)} = \frac{1}{Z}\Gamma(\frac{\nu}{2}+1)\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}+\nu\right]^{-\left(1+\frac{\nu}{2}\right)} = \\
&\frac{1}{Z_t}\left[1+\frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}\right]^{-\left(1+\frac{\nu}{2}\right)}
\end{aligned} \tag{4}
$$

The second term is the functional form of the Student t distribution and Z_t is just its normalization constant.

Conclusion

In this exercise, we proved that the Student t distribution can be expressed as an infinite mixture of Gaussians. In my opinion, the solution is not that hard once you realize which terms you can ignore in each passage and have some familiarity with the Gamma distribution.