Student T-distribution [2] as A Marginalization

Qi Zhao

August 17, 2018

1. Introduction

As for the mixtures of Gaussians [3], it is also possible to understand the t-distribution in terms of hidden variables. It defines in Equation 1.

$$\begin{aligned} Pr(x|h) &= Norm_x[\mu, \Sigma/h] \\ Pr(h) &= Gam_h[\nu/2, \nu/2] \end{aligned} \tag{1}$$

where h is a scalar hidden variable and Gam[α, β] is the gamma distribution [4] with parameters α, β (Figure 1). The gamma distribution is a continuous probability distribution defined on the positive real axis with probability density function in Equation 2.

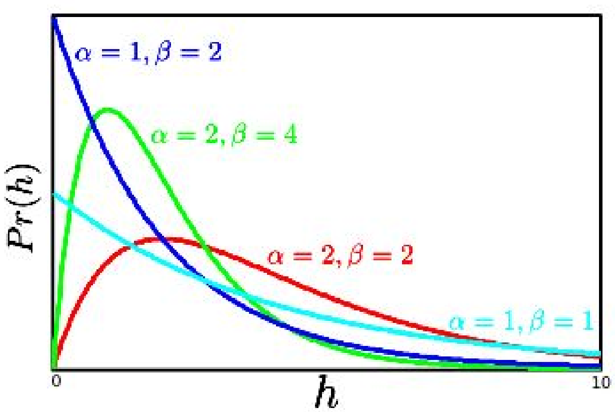

Figure 1. The gamma distribution is defined on positive real values and has two parameters α, β. The mean of the distribution is E[h] = α/β and the variance is E[(h?E[h])^2] = α/β^2 . The t-distribution can be thought of as a weighted sum of normal distributions with the same mean, but covariances that depend inversely on the gamma distribution.

$$Gam_h[\alpha, \beta] = \frac{\beta^\alpha}{\Gamma[\alpha]} exp[-\beta h] h^{\alpha-1} \tag{2}$$

where Γ[] is the gamma function.

The t-distribution is the marginalization with respect to the hidden variable h of the joint distribution between the data x and h (Figure 2).

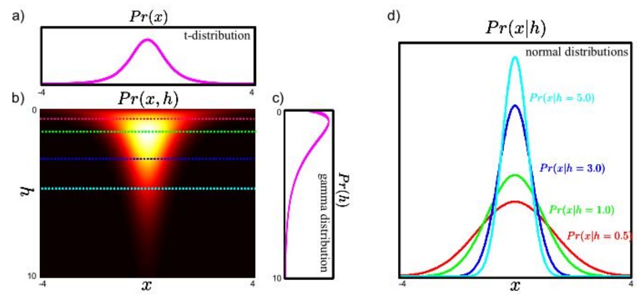

Figure 2. a) The t-distribution has a similar form to the normal distribution but longer tails. b) The t-distribution is the marginalization of the joint distribution Pr(x,h) between the observed variable x and a hidden variable h. c) The prior distribution over the hidden variable h has a gamma distribution. d) The conditional distribution Pr(x—h) is normal with a variance that depends on h. So the t-distribution can be considered as an infinite weighted sum of normal distributions [1] with variances determined by the gamma prior.

2. Conclusions

This formulation also provides a method to generate data from the t-distribution; it first generates h from the gamma distribution and then generate x from the as-sociated normal distribution Pr(x|h). Hence the hidden variable has a simple interpretation: it tells which one of the continuous family of underlying normal distributions was responsible for this data point.

References

[1] N. E. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56(3):463–474, 1969. 1

[2] M. C. Jones and M. J. Faddy. A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society*, 65(1):159–174, 2010. 1

[3] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003. 1

[4] E. W. Stacy. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33(3):1187–1192, 1962. 1