The t-distribution [2]

Qi Zhao

August 15, 2018

The second significant problem with using the normal distribution [4] to describe visual data [1] is that it is not robust: the height of the normal pdf falls off very rapidly as it moves into the tails. The effect of this is that outliers (unusually extreme observations) drastically affect the estimated parameters (Figure 1). The t-distribution is a closely related distribution in which the length of the tails is parameterized.

Figure 2. As well as the mean mu and scaling parameter σ^2, the t-distribution has a parameter ν which is termed the degrees of freedom. As ν decreases, the tails of the distribution become longer and the model becomes more robust. b) This is seen more clearly on a log scale.

The multivariate t-distribution has pdf in Equation **??**.

$$Pr(x) = Stud_x[\mu, \Sigma, \nu]$$
$$= \frac{\Gamma(\frac{\nu+D}{2})}{(\nu\pi)^{D/2}|\Sigma|^{1/2}\Gamma[\nu/2]}(1 + \frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{\nu})^{-(\nu+D)/2}$$

(1)

where D is the dimensionality of the space, μ is a d X 1 mean vector, Σ is a DXD positive definite scale matrix, and $\nu \in [0, \infty]$ is the degrees of freedom. As for the multivariate normal distribution, the scale matrix can take full, diagonal or spherical forms. The covariance of the distribution is given by $\Sigma \nu / (\nu - 2)$ for $\nu > 2$ and is infinite if $0 \leq \nu \leq 2$.
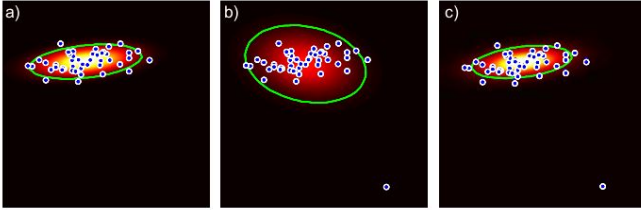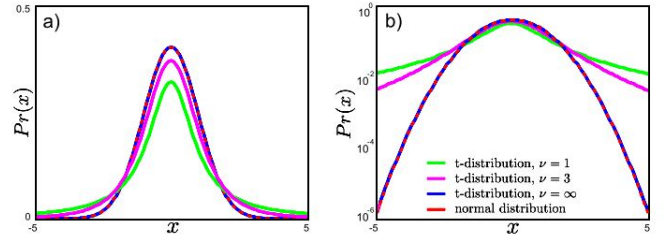
Figure 1. Motivation for t-distribution. a) The multivariate normal model fit to data. b) Adding a single outlier completely changes the fit. c) With the multivariate t-distribution the outlier does not have such a drastic effect.

The univariate t-distribution (Figure 2) has probability density [3] function where μ is the mean and σ^2 is the scale parameter. The degrees of freedom $\nu \in (0, \infty]$ controls the length of the tails: when ν is small there is considerable weight in the tails. For example, with $mu = 0$ and $\sigma^2 = 1$ a data point at x = -5 is roughly 10^4 = 10000 times more likely under the t-distribution with $\nu = 1$ than under the normal distribution. As ν tends to infinity, the distribution approximates a normal more and more closely and there is less weight in the tails. The variance of the distribution is given by $\sigma\nu / (nu - 2)$ for $\nu > 2$ and infinite if $0 < \nu \leq 2$.

References

[1] G. L. Andrienko and N. V. Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4):355–374, 1999. 1

[2] J. Geweke. Efficient simulation from the multivariate normal and student t-distributions subject to linear constraints. *Journal of Dermatologic Surgery & Oncology*, 11(4):420–423, 1991. 1

[3] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962. 1

[4] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9(6):1135–1151, 1981. 1