

The goal of this document is to derive an analytical expression for a bimodal probability density function that is characterized entirely in terms of three variables

- the mean,  $\mu$ ,
- the standard deviation,  $\sigma$ , and
- the mean-squared,  $m$ .

In order to calculate each of these from a given probability density function,  $P(x)$ , we use the expressions

$$\mu = \int_{-\infty}^{\infty} xP(x)dx \quad (1)$$

$$m = \int_{-\infty}^{\infty} x^2P(x)dx \quad (2)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 P(x)dx \quad (3)$$

In order to build up this probability density function, I choose to model it as the sum of two normal distributions. Each of these has the form:

$$P_n(x, \mu, \sigma) = Ae^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (4)$$

so that the overall PDF, composed of two of these has the form

$$P(x, \mu_1, \sigma_1, \mu_2, \sigma_2) = P_n(x, \mu_1, \sigma_1) + P_n(x, \mu_2, \sigma_2) \quad (5)$$

$$= Ae^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + Ae^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}. \quad (6)$$

So our first task is to figure out the normalization constant,  $A$ . To calculate this, we require that

$$1 = \int_{-\infty}^{\infty} P(x, \mu_1, \sigma_1, \mu_2, \sigma_2)dx \quad (7)$$

Solving this in Mathematica yields:

$$A = \frac{1}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} \quad (8)$$

So the complete PDF has the form

$$P(x, \mu_1, \sigma_1, \mu_2, \sigma_2) = \frac{1}{\sqrt{2\pi}(\sigma_1 + \sigma_2)} \left( e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \right) \quad (9)$$

So our goal here is to relate these parameters, which describe individual distributions, to the parameters in the bulleted list above, which describe the composite overall distribution. Scott Stickels told me that “mode separation is approximately  $\sqrt{m - \mu^2 - \sigma}$ ”. I interpret this to mean that:

$$\mu_1 = \mu - \frac{1}{2}\sqrt{m - \mu^2 - \sigma} \quad (10)$$

$$\mu_2 = \mu + \frac{1}{2}\sqrt{m - \mu^2 - \sigma} \quad (11)$$

Therefore, we can simplify our expression for  $P$  using this, but we also need to relate  $\sigma_1$  and  $\sigma_2$  to  $\sigma$ . In order to do that, we can assume that the distribution is symmetric (ie  $\sigma_1 = \sigma_2 = s$ ), substitute in the expressions above for  $\mu_1$ , and  $\mu_2$ , and calculate the variance

$$\sigma^2 = \frac{1}{4}(m + 4s^2 - \mu^2 - \sigma) \quad (12)$$

Therefore, we can solve this expression for the standard deviation of the individual normal distributions,  $s$ . The result is two solutions:

$$s = -\frac{1}{2}\sqrt{-m + \mu^2 + \sigma + 4\sigma^2} \quad (13)$$

$$s = \frac{1}{2}\sqrt{-m + \mu^2 + \sigma + 4\sigma^2} \quad (14)$$

$$(15)$$

Since the standard deviation should be positive, I choose the second one. Therefore, the total PDF is

$$P(x, \mu, \sigma, m) = \frac{1}{\sqrt{2\pi}\sqrt{-m + \mu^2 + \sigma + 4\sigma^2}} \left( e^{-\frac{1}{2}\frac{(2x-2\mu+\sqrt{m-\mu^2-\sigma})^2}{-m+\mu^2+\sigma+4\sigma^2}} + e^{-\frac{1}{2}\frac{(-2x+2\mu+\sqrt{m-\mu^2-\sigma})^2}{-m+\mu^2+\sigma+4\sigma^2}} \right) \quad (16)$$