

Mean, Median, Variance, and MAD of Normal Mixtures

Yclept Nemo

March 1, 2022

Abstract

Given a mixture of normal variables derive the mean, median, variance, and median absolute deviation (MAD) of the mixture in terms of the component mean and variance.

1 Mean and Variance

Let X be the random variable of a mixture of n random variables with weights p_1, \dots, p_n such that $\sum_i p_i = 1$. Given the PDF f_i of each component random variable where f_i represents any distribution, the PDF of the mixture is defined as:

$$f_X(x) = \sum_i p_i f_i(x) \quad (1)$$

Mean is the first raw moment (i.e., moment about zero) and variance the second central moment (i.e., moment about the mean). Let $\mu^{(n')}$ be the n^{th} raw moment of $f_X(x)$, and $\mu_i^{(n')}$ be the n^{th} raw moment of $f_i(x)$. We use the definition of moment to define $\mu^{(n')}$ in terms of each component $\mu_i^{(n')}$:

$$\begin{aligned} \mu^{(n')} &= E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^n \sum_i p_i f_i(x) dx \\ &= \sum_i p_i \int_{-\infty}^{\infty} x^n f_i(x) dx \\ &= \sum_i p_i \mu_i^{(n')} \end{aligned} \quad (2)$$

Therefore mean is simply the weighted sum of the component means. Mean is often notated as μ rather than $\mu^{(1')}$, and both are equivalent:

$$\begin{aligned} \mu^{(1')} &= E[X] = \sum_i p_i \mu_i^{(1')} \\ \mu &= E[X] = \sum_i p_i \mu_i \end{aligned} \quad (3)$$

The second central moment $\mu^{(2)}$ about c is related to the second raw moment $\mu^{(2')}$ for any random variable X with a PDF $f_X(x)$ as follows:

$$\begin{aligned} \mu^{(2)} &= \int_{-\infty}^{\infty} (x - c)^2 f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2c \int_{-\infty}^{\infty} x f_X(x) dx + c^2 \int_{-\infty}^{\infty} f_X(x) dx \\ &\stackrel{1}{=} \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2c \int_{-\infty}^{\infty} x f_X(x) dx + c^2 \\ &\stackrel{2}{=} \mu^{(2')} - 2c\mu^{(1')} + c^2 \end{aligned} \quad (4)$$

When $c = \mu^{(1')}$:

$$\begin{aligned}\text{Var}(X) &= \mu^{(2')} - 2\mu^{(1')}\mu^{(1')} + \left(\mu^{(1')}\right)^2 \\ &= \mu^{(2')} - \left(\mu^{(1')}\right)^2\end{aligned}\tag{5}$$

Where:

- $\stackrel{1}{=}$ The integral of a PDF is 1
- $\stackrel{2}{=}$ Substituting the raw first and second moments

Variance can also be defined and expanded in terms of a random variable. This process can be generalized to derive the Law of Total Cumulance. For whichever method is used the results are equivalent.

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2\end{aligned}\tag{6}$$

The variance of the mixture X can be related to the raw first and second moments of its components by substituting 5 with 2:

$$\begin{aligned}\text{Var}(X) &= \mu^{(2')} - \left(\mu^{(1')}\right)^2 \\ &= \sum_i p_i \mu_i^{(2')} - \left(\sum_i p_i \mu_i^{(1')}\right)^2\end{aligned}\tag{7}$$

From 5 we know $\mu^{(2')} = \mu^{(2)} + \left(\mu^{(1')}\right)^2$ for any PDF. This allows the variance of the mixture X to be related to the central second moment:

$$\begin{aligned}\text{Var}(X) &= \sum_i p_i \mu_i^{(2')} - \left(\sum_i p_i \mu_i^{(1')}\right)^2 \\ &= \sum_i p_i \left(\mu_i^{(2)} + \left(\mu_i^{(1')}\right)^2\right) - \left(\sum_i p_i \mu_i^{(1')}\right)^2 \\ &= \sum_i p_i \mu_i^{(2)} + \sum_i p_i \left(\mu_i^{(1')}\right)^2 - \left(\sum_i p_i \mu_i^{(1')}\right)^2\end{aligned}\tag{8}$$

Which can be rewritten using the more common terminology for variance $\sigma_i^2 = \mu_i^{(2)}$ and mean $\mu_i = \mu_i^{(1')}$:

$$\text{Var}(X) = \sum_i p_i \sigma_i^2 + \sum_i p_i \mu_i^2 - \left(\sum_i p_i \mu_i\right)^2\tag{9}$$

2 Median and MAD

We have now derived the mean and variance for a mixture of any distributions and turn to the median and MAD. These two statistics are nonlinear and cannot be discussed without also defining the distribution. A distribution with a mean μ and variance σ^2 may have any median and MAD. If the distribution is normal then $\text{Med}(X) = \mu$ and $\text{MAD}(X) = \sigma\sqrt{2}\text{erf}^{-1}(1/2)$, because the normal distribution is symmetric. We will only derive the median and MAD for a mixture of normal distributions. These statistics do not have a convenient closed-form solution but can be solved numerically as demonstrated.

As intuition consider a mixture of two non-overlapping normal distributions N_1 and N_2 such that $\mu_1 < \mu_2$, $p_1 = 0.7$, and $p_2 = 0.3$. Visualize the mean of the mixture when shifting the center of either distribution and notice it follows the weighted sum of the component means as in 3. Arithmetically this equates to the finding the mean of a list by first subdividing and finding the mean of each sub-list. Now decrease or increase the variance of either distribution and notice the mean remains unaffected. Because the normal distribution is symmetric, for each component shifted towards the mean there exists a single component shifted away from the mean by an equal amount, and thus on balance the sums are negated and the mean remains unaffected. This demonstrates that the mixture mean depends only on the component means.

As the distributions are moved apart the variance increases. Increase a component variance and the mixture variance also increases. The mixture is not symmetric about its mean so when squared further distances are given more weight. And so the variance can be viewed as the dispersion of the means plus a mixture of variances.

The median can be attributed to a point from a single distribution given solely by the component weights. As long as the distributions do not overlap and order is preserved the median follows that point as both mean and variance fluctuate. This shows that the median depends on both mean and variance; given the above conditions exclusively on the mean and variance of a single distribution.

The MAD is the (sorted) median of the absolute distance from the median. As such it depends on the mean and variance of the intervening distributions.

There is not a 1:1 correspondence between median and MAD, and mean and variance. Mixtures with the differing mean and variance may share the same median and MAD. Mixtures with differing median and MAD may share the same mean and variance. This is not unexpected considering how median and MAD are robust against outliers. As seen below, this means that systems of relations between these statistics may have multiple solutions.

The CDF F_X of the random variable X is defined as the integral of the corresponding PDF f_X over $[-\infty, x]$. By substituting the mixture PDF from 1 with component PDFs f_i of any distribution, the resulting CDF of the mixture is shown to be the weighted sum of the component CDFs F_i of any distribution:

$$\begin{aligned}
F_X(x) &= \int_{-\infty}^x f_X(t) dt \\
&= \int_{-\infty}^x \sum_i p_i f_i(t) dt \\
&= \sum_i p_i \int_{-\infty}^x f_i(t) dt \\
&= \sum_i p_i F_i(x)
\end{aligned} \tag{10}$$

The CDF of the normal distribution is:

$$F_i(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \right] \tag{11}$$

When each component of the mixture is normally distributed the resulting CDF becomes:

$$F_X(x) = \frac{1}{2} \sum_i p_i \left[1 + \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \right] \tag{12}$$

For any distribution X where F_X^{-1} is the inverse CDF, the median is by definition:

$$\operatorname{Med}(X) = F_X^{-1} \left(\frac{1}{2} \right) \tag{13}$$

In other words the median is the point x at which the CDF is $\frac{1}{2}$; the point at which $P(X \leq x) = P(X \geq x) = \frac{1}{2}$. Since the inverse of a normal mixture X does not have a closed-form expression,

we instead express its median as the solution in x to:

$$\begin{aligned}
\frac{1}{2} &= \frac{1}{2} \sum_i p_i \left[1 + \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \right] \\
1 &= \sum_i p_i \left[1 + \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \right] \\
1 &= \sum_i p_i + \sum_i p_i \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \\
1 &\stackrel{1}{=} 1 + \sum_i p_i \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right) \\
0 &= \sum_i p_i \operatorname{erf} \left(\frac{x - \mu_i}{\sigma_i \sqrt{2}} \right)
\end{aligned} \tag{14}$$

Where:

$$\stackrel{1}{=} \text{By definition (1), } \sum_i p_i = 1$$

The function $\operatorname{erf}(x)$ is a strictly increasing function with a range of $(-1, 1)$. Given the general form $a \cdot \operatorname{erf}(bx + c)$, a is the vertical scale factor, b the horizontal scale factor and c the horizontal shift. A positive c results in a left shift; a negative c results in a right shift. A positive b results in a strictly increasing function; a negative b results in a strictly decreasing function. The resulting range is $(-a, a)$. As a sum of strictly increasing functions is itself strictly increasing we see that there is only one solution to 14, the median of X . This is good, as there can only be a single median.

Given any distribution X with median $\operatorname{Med}(X)$, define a new distribution representing the absolute deviation from the median:

$$Y = |X - \operatorname{Med}(X)| \tag{15}$$

Then by definition the median absolute deviation is the median of Y :

$$\begin{aligned}
\operatorname{MAD}(X) &= \operatorname{Med}(|X - \operatorname{Med}(x)|) \\
&= \operatorname{Med}(Y) \\
&= F_Y^{-1} \left(\frac{1}{2} \right)
\end{aligned} \tag{16}$$

When X is a normal mixture the inverse function F_Y^{-1} will not have a closed expression. But first we must derive the CDF F_Y and the PDF f_Y of the distribution Y . Let X be any distribution. Given a point m let $Y = |X - m|$, where m represents $\operatorname{Med}(X)$. The CDF of Y can be found as follows:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(|X - m| \leq y) \\
&\stackrel{1}{=} P(-y \leq X - m \leq y) \\
&= P(m - y \leq X \leq y + m) \\
&\stackrel{2}{=} F_X(y + m) - F_X(m - y)
\end{aligned} \tag{17}$$

Where:

$$\begin{aligned}
&\stackrel{1}{=} \text{Expanding the cases of the absolute value} \\
&\stackrel{2}{=} \text{By definition of the CDF}
\end{aligned}$$

For the PDF f_Y take the derivative of the CDF F_Y . Remember that F_X is a definite integral of f_X and use the fundamental theorem of calculus and the chain rule to find:

$$\begin{aligned}
f_Y(y) &= F_Y'(y) \\
&= 1 \cdot f_X(y + m) - (-1) \cdot f_X(m - y) \\
&= f_X(y + m) + f_X(m - y)
\end{aligned} \tag{18}$$

We can verify this using the change-of-variable technique for the case of a two-to-one function such as $Y = |X - m|$. Let $Y = g(X)$ be a continuous two-to-one function which can be split into two invertible functions $X_1 = g_1^{-1}(Y)$ and $X_2 = g_2^{-1}(Y)$. The PDF of Y is related to the PDF of X as:

$$f_Y(y) = f_X(g_1^{-1}(y)) \cdot \left| \frac{d}{dy} g_1^{-1}(y) \right| + f_X(g_2^{-1}(y)) \cdot \left| \frac{d}{dy} g_2^{-1}(y) \right| \quad (19)$$

Let's split our two-to-one function into two invertible functions:

$$\begin{aligned} Y &= |X - m| \\ &= \begin{cases} X - m, & X \geq m \\ m - X, & X \leq m \end{cases} \end{aligned} \quad (20)$$

The inverse of Y can be found by solving each piecewise equation and substituting the bounds:

$$X = \begin{cases} Y + m, & X \geq m \quad \wedge \quad Y \geq 0 \\ m - Y, & X \leq m \quad \wedge \quad Y \geq 0 \end{cases} \quad (21)$$

By applying the change-of-variable technique we derive the same PDF:

$$\begin{aligned} f_Y(y) &= f_X(g_1^{-1}(y)) \cdot \left| \frac{d}{dy} g_1^{-1}(y) \right| + f_X(g_2^{-1}(y)) \cdot \left| \frac{d}{dy} g_2^{-1}(y) \right| \\ &= f_X(y + m) \cdot \left| \frac{d}{dy} (y + m) \right| + f_X(m - y) \cdot \left| \frac{d}{dy} (m - y) \right| \\ &= f_X(y + m) \cdot |1| + f_X(m - y) \cdot |-1| \\ &= f_X(y + m) + f_X(m - y) \end{aligned} \quad (22)$$

It is important to note the domain of Y found via substitution (21). Even though the domain of the resulting PDF f_Y and CDF F_Y may differ, the supported domain of these functions is at most the domain Y . This is termed the *support* of the distribution; for Y it is $[0, \infty]$. As an example, let X be the normal distribution $X \sim N(\mu, \sigma^2)$ and m be the median of X , $m = \mu$. The PDF f_X is by definition:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (23)$$

By substituting f_X using 22 or 18 the PDF f_Y of Y is:

$$\begin{aligned} f_Y(y) &= f_X(y + m) + f_X(m - y), \quad y \geq 0 \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y+m-\mu}{\sigma}\right)^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{m-y-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \left(e^{-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{-y}{\sigma}\right)^2} \right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} 2e^{-\frac{1}{2}\left(\frac{y^2}{\sigma^2}\right)} \\ &= \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma^2}\right)}, \quad y \geq 0 \end{aligned} \quad (24)$$

This is the well-known *half-normal* distribution. Although the domain of f_Y is $[-\infty, \infty]$, the support of f_Y is $[0, \infty]$. The CDF F_Y of this half-normal distribution is found by substituting the CDF F_X of the normal distribution (11) using 17. The CDF F_Y has the same support, $[0, \infty]$. Because erf is an odd function, that is:

$$-\text{erf}(x) = \text{erf}(-x) \quad (25)$$

The CDF F_Y can be simplified further:

$$\begin{aligned}
F_Y(y) &= F_X(y+m) - F_X(m-y), \quad y \geq 0 \\
&= \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{y+m-\mu}{\sigma\sqrt{2}} \right) \right] - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{m-y-\mu}{\sigma\sqrt{2}} \right) \right] \\
&= \frac{1}{2} \operatorname{erf} \left(\frac{y}{\sigma\sqrt{2}} \right) - \frac{1}{2} \operatorname{erf} \left(\frac{-y}{\sigma\sqrt{2}} \right) \\
&= \operatorname{erf} \left(\frac{y}{\sigma\sqrt{2}} \right), \quad y \geq 0
\end{aligned} \tag{26}$$

The median of Y given a normal X is known to be $\sigma\sqrt{2} \operatorname{erf}^{-1}(1/2)$, by 16 this is also the MAD of X . Let's see if we are in agreement:

$$\begin{aligned}
\frac{1}{2} &= \operatorname{erf} \left(\frac{y}{\sigma\sqrt{2}} \right), \quad y \geq 0 \\
\frac{y}{\sigma\sqrt{2}} &= \operatorname{erf}^{-1} \left(\frac{1}{2} \right), \quad y \geq 0 \\
y &= \sigma\sqrt{2} \operatorname{erf}^{-1} \left(\frac{1}{2} \right), \quad y \geq 0
\end{aligned} \tag{27}$$

Now let X be a mixture of normal distributions while Y remains unchanged. That is, let $Y = |X - m|$ where m is a point representing $\operatorname{Med}(X)$. Given the PDF of a normal distribution in 23, by substitution with 1 the PDF of the normal mixture X is:

$$f_X(x) = \sum_i p_i \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_i}{\sigma_i} \right)^2} \tag{28}$$

Therefore the PDF f_Y of Y is:

$$\begin{aligned}
f_Y(y) &= f_X(y+m) + f_X(m-y), \quad y \geq 0 \\
&= \sum_i p_i \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{y+m-\mu_i}{\sigma_i} \right)^2} + \sum_i p_i \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{m-y-\mu_i}{\sigma_i} \right)^2} \\
&= \sum_i p_i \frac{1}{\sigma_i \sqrt{2\pi}} \left[e^{-\frac{1}{2} \left(\frac{y+m-\mu_i}{\sigma_i} \right)^2} + e^{-\frac{1}{2} \left(\frac{m-y-\mu_i}{\sigma_i} \right)^2} \right], \quad y \geq 0
\end{aligned} \tag{29}$$

And given the CDF of the normal mixture X in 12, the CDF F_Y of Y is:

$$\begin{aligned}
F_Y(y) &= F_X(y+m) - F_X(m-y), \quad y \geq 0 \\
&= \frac{1}{2} \sum_i p_i \left[1 + \operatorname{erf} \left(\frac{y+m-\mu_i}{\sigma_i \sqrt{2}} \right) \right] - \frac{1}{2} \sum_i p_i \left[1 + \operatorname{erf} \left(\frac{m-y-\mu_i}{\sigma_i \sqrt{2}} \right) \right] \\
&= \frac{1}{2} \sum_i p_i + \frac{1}{2} \sum_i p_i \operatorname{erf} \left(\frac{y+m-\mu_i}{\sigma_i \sqrt{2}} \right) \\
&\quad - \frac{1}{2} \sum_i p_i - \frac{1}{2} \sum_i p_i \operatorname{erf} \left(\frac{m-y-\mu_i}{\sigma_i \sqrt{2}} \right) \\
&= \frac{1}{2} \sum_i p_i \operatorname{erf} \left(\frac{y+m-\mu_i}{\sigma_i \sqrt{2}} \right) - \frac{1}{2} \sum_i p_i \operatorname{erf} \left(\frac{m-y-\mu_i}{\sigma_i \sqrt{2}} \right) \\
&= \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y+m-\mu_i}{\sigma_i \sqrt{2}} \right) - p_i \operatorname{erf} \left(\frac{-y+m-\mu_i}{\sigma_i \sqrt{2}} \right) \right], \quad y \geq 0
\end{aligned} \tag{30}$$

When $m = \operatorname{Med}(X)$ it follows from 16 that the median absolute deviation of X is the median of Y , since Y describes the absolute deviation of X from m . The inverse function F_Y^{-1} has no

closed-form solution and therefore both $\text{MAD}(X)$ and $\text{Med}(Y)$ are the solution in y to:

$$\begin{aligned} \frac{1}{2} &= \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y + m - \mu_i}{\sigma_i \sqrt{2}} \right) - p_i \operatorname{erf} \left(\frac{-y + m - \mu_i}{\sigma_i \sqrt{2}} \right) \right], \quad y \geq 0 \\ \frac{1}{2} &\stackrel{1}{=} \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y + (m - \mu_i)}{\sigma_i \sqrt{2}} \right) + p_i \operatorname{erf} \left(\frac{y - (m - \mu_i)}{\sigma_i \sqrt{2}} \right) \right], \quad y \geq 0 \\ 1 &= \sum_i \left[p_i \operatorname{erf} \left(\frac{y + (m - \mu_i)}{\sigma_i \sqrt{2}} \right) + p_i \operatorname{erf} \left(\frac{y - (m - \mu_i)}{\sigma_i \sqrt{2}} \right) \right], \quad y \geq 0 \end{aligned} \quad (31)$$

Where:

$$\stackrel{1}{=} \text{ Because erf is an odd function, see } 25$$

Let us examine the shape of the graph of the right-hand side. Since a sum of strictly increasing functions is itself strictly increasing, there is only one solution. Since erf is an odd function, $\operatorname{erf}(x) + \operatorname{erf}(-x) = 0$. Therefore every component has the point $(0, 0)$ and so does the resulting sum. Every component is the sum of two erf functions each representing the \pm distance from the component mean to the median, ensuring the resulting function is also odd. A positive distance results in the characteristic erf bump in $\mathbb{R}_{<0}$ due to the left shift; the second function ensures the bump is mirrored to $\mathbb{R}_{>0}$. The sum of two erf functions doubles the range from $(-1, 1)$ to $(-2, 2)$; the $1/2$ compensates for that. In fact for $y \geq 0$ this function can be approximated as:

$$F_{Y_a}(y) \approx \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y - |m - \mu_i|}{\sigma_i \sqrt{2}} \right) + 1 \right], \quad y \geq 0 \quad (32)$$

The error in this approximation is:

$$\begin{aligned} \varepsilon &= F_Y - F_{Y_a}, \quad y \geq 0 \\ &= \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y + (m - \mu_i)}{\sigma_i \sqrt{2}} \right) + p_i \operatorname{erf} \left(\frac{y - (m - \mu_i)}{\sigma_i \sqrt{2}} \right) \right] \\ &\quad - \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y - |m - \mu_i|}{\sigma_i \sqrt{2}} \right) + 1 \right] \\ &= \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y + |m - \mu_i|}{\sigma_i \sqrt{2}} \right) + p_i \operatorname{erf} \left(\frac{y - |m - \mu_i|}{\sigma_i \sqrt{2}} \right) \right. \\ &\quad \left. - p_i \operatorname{erf} \left(\frac{y - |m - \mu_i|}{\sigma_i \sqrt{2}} \right) - 1 \right] \\ &= \frac{1}{2} \sum_i \left[p_i \operatorname{erf} \left(\frac{y + |m - \mu_i|}{\sigma_i \sqrt{2}} \right) - 1 \right], \quad y \geq 0 \end{aligned} \quad (33)$$

Which for $y \geq 0$, is near the asymptote and very small.

Now to solve for $\text{Med}(X)$ (14) or $\text{MAD}(X)$ (31), assuming all other parameters are known and the index of summation i is bounded, replace each erf with its Taylor series expansion to the n th order term. Expand the expression, collect like terms, and apply series reversion using the standard reversion solution coefficients. Apply a change of variables if there is a constant term $a_0 \neq 0$ because standard series reversion requires $a_0 = 0$. This generates an approximate solution which can be refined using first-order Newton's method. The mean μ can be used as the initial point for Taylor expansion.

However we are interested in the case where the median and MAD are known, as well as the relationship between parameter sequences μ_i, σ_i such as: $\mu_2 = 2\mu_1, \mu_3 = 3\mu_1, \dots$. This leaves two non-linear equations of two unknowns.

2.1 The Shape of the Graphs

Consider the median of a normal mixture in 14 as the solution $F_x(\mu, \sigma; x) = 0$ to the following equation in two variables μ and σ where x (median) is a constant. The function represents a 3-dimensional surface; the solution will be a level curve.

$$F_x(\mu, \sigma; x) = \sum_i p_i \operatorname{erf} \left(\frac{x - f_{\mu_i}(\mu, \dots)}{f_{\sigma_i}(\sigma, \dots) \sqrt{2}} \right) \quad (34)$$

The simplest component functions are:

$$\begin{aligned} f_{\mu_i}(\mu, \dots) &= \mu_i \\ f_{\sigma_i}(\sigma, \dots) &= \sigma_i \end{aligned} \quad (35)$$

There must however be some restrictions. By definition variance must not be negative and for simplicity assume that all component functions are increasing. What is the shape of the level curve?

Now consider the MAD of a normal mixture in 31 as the solution $F_y(\mu, \sigma; x, y) = 0$ to the following equation in two variables μ and σ where x (median) and y (MAD) are both constant. Once again the function represents a 3-dimensional surface; the solution will be a level curve. With the same component functions and restrictions as the previous function, what is the shape of the level curve?

$$\begin{aligned} F_y(\mu, \sigma; x, y) &= -1 + \sum_i p_i \operatorname{erf} \left(\frac{y + x - f_{\mu_i}(\mu, \dots)}{f_{\sigma_i}(\sigma, \dots)\sqrt{2}} \right) \\ &\quad - \sum_i p_i \operatorname{erf} \left(\frac{-y + x - f_{\mu_i}(\mu, \dots)}{f_{\sigma_i}(\sigma, \dots)\sqrt{2}} \right), \quad y \geq 0 \end{aligned} \quad (36)$$

The shape of the level curves $F_x = 0$ and $F_y = 0$ can be derived from the partial derivatives of the corresponding surfaces using the implicit function theorem. For example let $\sigma = f_{\sigma_i}(\mu)$ be the set of functions between the critical points in F_x along σ . Then the derivative of each f_{x_i} is given as:

$$f'_{x_i}(\mu) = - \frac{\frac{dF_x}{d\mu}}{\frac{dF_x}{d\sigma}}, \quad \frac{dF_x}{d\sigma} \neq 0 \quad (37)$$

As a system two equations the median and MAD provide a constraint to the mean and variance of the mixture. By analysing the shape of the graph we can hope to determine the number of solutions to this system.

$$\begin{cases} F_x(\mu, \sigma; x) = 0 \\ F_y(\mu, \sigma; x, y) = 0 \end{cases} \quad (38)$$

Analysis of the level curves in this fashion quickly becomes piecewise very complicated even when restricted to the limited range of the erf function. Therefore no further work was done in this direction.

2.2 A Numerical Example

Morse code symbols must be one of three durations. Density plots of human Morse code demonstrate that each duration can be represented as a roughly normal random variable. Therefore Morse code duration can be modeled as a mixture X of three normals X_1 , X_2 , and X_3 . Interestingly the standard deviation of each component is logarithmically proportional to its mean duration, suggesting a Weber-Fechner relation. This opens the possibility that Morse code duration instead follows a log-normal mixture. This is unlikely for two reasons. The Weber-Fechner law suggests that actual outputs are logarithmically proportional to intended outputs and unrelated to deviations. Finally, log-normal distributions feature a positive skew which is not present in the data. Our intent is to build a robust model of Morse code duration using median and MAD, exploiting existing constraints to match the number of unknowns and equations. By providing the median and MAD - statistics less sensitive to outliers - we can solve the system numerically for the expected mean and standard deviation.

The means of each component are specified by the Morse code standard.

$$\begin{aligned} \mu_1 &= \mu_1 \\ \mu_2 &= 3\mu_1 \\ \mu_3 &= 7\mu_1 \end{aligned} \quad (39)$$

The weights of each component can be derived by frequency analysis of Morse code symbols for the English language. See [morse_data.py:time_frequencies](#) for more information.

$$\begin{aligned} p_1 &= 0.5988 \\ p_2 &= 0.3598 \\ p_3 &= 0.0413 \end{aligned} \tag{40}$$

In the space of linear-mean to logarithmic-deviation the relation between component mean and standard deviation is linear. While a line has two unknowns, to avoid an underdetermined system we require no more than one unknown. Expecting the y-intercept to be a physiological constant I undertook an overview of research into tempo regulation and finger-tapping. Hammerschmidt, Frieler, and Wöllner [HFW21] found the standard deviation of spontaneous motor tempo (SMT) for the majority of natural inter-tap intervals (< 1300 ms) to be relatively constant, averaging 72 ms with 3.8 ms standard deviation. This is consistent with the y-intercept of my data at 79 ms and if accurate would result in the following relation. Note that the point $(0, b)$ represents a fixed standard deviation common to all components while the slope a represents the relative differences in standard deviations between components.

$$\begin{aligned} \ln \sigma &= a\mu + b_l \\ \sigma &= e^{a\mu + b_l} \\ \sigma &= e^{a\mu} e^{b_l} \\ \sigma &= b e^{a\mu} \end{aligned} \tag{41}$$

Where:

$$\begin{aligned} b_l &= \ln b \\ b &\approx 72 \end{aligned}$$

The study of tempo regulation has several competing theories each modeling different fundamental timing mechanisms which remain poorly understood [Lit00]. There are many degrees of freedom in the human sensorimotor system allowing Morse code tappers to exhibit a wide variety of behaviors [NC93]. Most theories predict either a linear standard deviation or a linear variance [RC98] which contradicts my logarithmic standard deviation. Even then it might be premature to assume that SMT standard deviation should represent the minimum (y-intercept) standard deviation. Instead of selecting the y-intercept from existing literature we select a point $p^{(1)} = (\mu^{(1)}, \sigma^{(1)})$ from our existing data. This is possible because the components do not overlap and accurate because the weight p_1 represents the majority of the data.

We can truncate a sample s of X with $|s|$ data points to a sample of X_1 using the weight p_1 .

$$\begin{aligned} i &= \lceil p_1 |s| \rceil \\ s^{(1)} &= [s_{(1)}, \dots, s_{(i)}] \end{aligned} \tag{42}$$

The truncated sample $s^{(1)}$ undoubtedly contains outliers from other components. Because X_1 is normal the median and MAD can be used as an estimate of mean and standard deviation.

$$\mu^{(1)} = \text{Med}(s^{(1)}) \tag{43}$$

$$\sigma^{(1)} = \frac{\text{MAD}(s^{(1)})}{\sqrt{2} \text{erf}^{-1}(1/2)} \tag{44}$$

Now the function of mean to standard deviation in point-slope form.

$$\begin{aligned} \ln \sigma &= a(\mu - \mu^{(1)}) + \sigma_l^{(1)} \\ \sigma &= e^{a(\mu - \mu^{(1)}) + \sigma_l^{(1)}} \\ \sigma &= e^{a(\mu - \mu^{(1)})} e^{\sigma_l^{(1)}} \\ \sigma &= e^{a(\mu - \mu^{(1)})} \sigma^{(1)} \end{aligned} \tag{45}$$

Where:

$$\sigma_l^{(1)} = \ln \sigma^{(1)}$$

There are now sufficient constraints to specialize the normal mixture median function F_x in 34 for Morse code. For f_{σ_i} we use the mapping $\mu \mapsto \sigma$ from 41 which assumes the y-intercept is a physiological constant. We then substitute the relations given in 39 for f_{μ_i} and all instances of mean.

$$\begin{aligned} F_{med}(\mu_1, a) = & + p_1 \operatorname{erf} \left(\frac{med - \mu_1}{be^{a\mu_1} \sqrt{2}} \right) \\ & + p_2 \operatorname{erf} \left(\frac{med - 3\mu_1}{be^{3a\mu_1} \sqrt{2}} \right) \\ & + p_3 \operatorname{erf} \left(\frac{med - 7\mu_1}{be^{7a\mu_1} \sqrt{2}} \right) \end{aligned} \quad (46)$$

This form uses the alternative mapping $\mu \mapsto \sigma$ from 45 in point-slope form which is derived from an existing sample.

$$\begin{aligned} F_{med}(\mu_1, a) = & + p_1 \operatorname{erf} \left(\frac{med - \mu_1}{\sigma^{(1)} e^{a(\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \\ & + p_2 \operatorname{erf} \left(\frac{med - 3\mu_1}{\sigma^{(1)} e^{a(3\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \\ & + p_3 \operatorname{erf} \left(\frac{med - 7\mu_1}{\sigma^{(1)} e^{a(7\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \end{aligned} \quad (47)$$

We similarly specialize the normal mixture MAD function F_y in 36 for Morse code. As with F_{med} in 46 this form uses the mapping $\mu \mapsto \sigma$ from 41 (y-intercept).

$$\begin{aligned} F_{mad}(\mu_1, a) = & -1 + p_1 \operatorname{erf} \left(\frac{mad + (med - \mu_1)}{be^{a\mu_1} \sqrt{2}} \right) + p_1 \operatorname{erf} \left(\frac{mad - (med - \mu_1)}{be^{a\mu_1} \sqrt{2}} \right) \\ & + p_2 \operatorname{erf} \left(\frac{mad + (med - 3\mu_1)}{be^{3a\mu_1} \sqrt{2}} \right) + p_2 \operatorname{erf} \left(\frac{mad - (med - 3\mu_1)}{be^{3a\mu_1} \sqrt{2}} \right) \\ & + p_3 \operatorname{erf} \left(\frac{mad + (med - 7\mu_1)}{be^{7a\mu_1} \sqrt{2}} \right) + p_3 \operatorname{erf} \left(\frac{mad - (med - 7\mu_1)}{be^{7a\mu_1} \sqrt{2}} \right) \end{aligned} \quad (48)$$

This form uses the alternative point-slope mapping $\mu \mapsto \sigma$ from 45 and is intended to be used with F_{med} from 47.

$$\begin{aligned} F_{mad}(\mu_1, a) = & -1 + p_1 \operatorname{erf} \left(\frac{mad + (med - \mu_1)}{\sigma^{(1)} e^{a(\mu_1 - \mu^{(1)})} \sqrt{2}} \right) + p_1 \operatorname{erf} \left(\frac{mad - (med - \mu_1)}{\sigma^{(1)} e^{a(\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \\ & + p_2 \operatorname{erf} \left(\frac{mad + (med - 3\mu_1)}{\sigma^{(1)} e^{a(3\mu_1 - \mu^{(1)})} \sqrt{2}} \right) + p_2 \operatorname{erf} \left(\frac{mad - (med - 3\mu_1)}{\sigma^{(1)} e^{a(3\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \\ & + p_3 \operatorname{erf} \left(\frac{mad + (med - 7\mu_1)}{\sigma^{(1)} e^{a(7\mu_1 - \mu^{(1)})} \sqrt{2}} \right) + p_3 \operatorname{erf} \left(\frac{mad - (med - 7\mu_1)}{\sigma^{(1)} e^{a(7\mu_1 - \mu^{(1)})} \sqrt{2}} \right) \end{aligned} \quad (49)$$

Together either form represents a system of two equations which can be solved numerically using Newton's method, though this requires an initial estimate. As before let the truncated samples $s^{(1)}, s^{(2)}, s^{(3)}$ represent the respective components X_1, X_2, X_3 . Let $s^{(xy)}$ represent the concatenation of samples $s^{(x)} \# s^{(y)}$. There are many methods to find an initial estimate,

$$x^{(1)} = (\mu_1^{(1)}, a^{(1)}) \quad (50)$$

From $p^{(1)}$ we know $\mu_1^{(1)} = \mu^{(1)}$. The simplest method to find the slope $a^{(1)}$ is from two points, $p^{(1)}$ and a second point $p^{(2)}$ describing $s^{(2)}$. As in 42 we can truncate a sample s to a sample $s^{(2)}$ of X_2 , then use 43 to generate $p^{(2)} = (\mu^{(2)}, \sigma^{(2)})$.

$$\begin{aligned} i_1 &= \lceil p_1 |s| + 1 \rceil \\ i_2 &= \lceil (p_1 + p_2) |s| \rceil \\ s^{(2)} &= [s_{(i_1)}, \dots, s_{(i_2)}] \end{aligned} \quad (51)$$

Rearranging either 41 or 45 yields the slope.

$$a = \frac{\ln \sigma - \ln \sigma^{(1)}}{\mu - \mu^{(1)}} \quad (52)$$

Substituting the points yields the initial estimate.

$$x^{(1)} = (\mu_1^{(1)}, a^{(1)}) = \left(\mu^{(1)}, \frac{\ln \sigma^{(2)} - \ln \sigma^{(1)}}{\mu^{(2)} - \mu^{(1)}} \right) a^{(1)} \quad (53)$$

We could have also solved for $p^{(2)}$ using the variance of the normal mixture $s^{(12)}$ and its components. Substituting all known quantities into 9 yields a relation of a single unknown. The same approach for $s^{(23)}$ yields a relation of two unknowns which can be further reduced by substituting the mapping $\mu \mapsto \sigma$ and solving directly for $a^{(1)}$. These methods are however sensitive to outliers.

More interesting is to use the relation between the median of $s^{(12)}$ and its components given by 14. Rearranging for the standard deviation of a two-component mixture yields the following relation. Unfortunately the solution lies near a vertical asymptote, greatly magnifying the lack of accuracy in the sample. A similar approach is possible for $s^{(23)}$ after substituting the mapping $\mu \mapsto \sigma$ and solving directly for $a^{(1)}$; this too is equally inaccurate.

$$\sigma_2 = \frac{med - \mu_2}{\text{erf}^{-1} \left[-\frac{p_1}{p_2} \text{erf} \left(\frac{med - \mu_1}{\sigma_1 \sqrt{2}} \right) \right] \sqrt{2}} \quad (54)$$

It is also possible to repeat the current process and solve for $p^{(2)}$ using a system of equations describing the median (14) and MAD (31) of the normal mixture $s^{(23)}$. As this first requires extracting $p^{(2)}$, solving this system would be redundant.

Now that we have the initial estimate we can write the system of equations as a vector-valued function. Newton's method provides a numerical approximation to $g(\mu_1, a) = 0$.

$$g(\mu_1, a) = \begin{bmatrix} F_{med}(\mu_1, a) \\ F_{mad}(\mu_1, a) \end{bmatrix} \quad (55)$$

The Jacobian matrix of g at a point is given by the following function.

$$J_g(\mu_1, a) = \begin{bmatrix} \frac{dF_{med}}{d\mu_1}(\mu_1, a) & \frac{dF_{med}}{da}(\mu_1, a) \\ \frac{dF_{mad}}{d\mu_1}(\mu_1, a) & \frac{dF_{mad}}{da}(\mu_1, a) \end{bmatrix} \quad (56)$$

The linear approximation of g at a point $x^{(k)}$ is given by the following function.

$$g_k(x) = g(x^{(k)}) + J_g(x^{(k)})(x - x^{(k)}) \quad (57)$$

We set the linear approximation to zero then substitute and solve for the next iterate $x^{(k+1)}$.

$$\begin{aligned} 0 &= g(x^{(k)}) + J_g(x^{(k)})(x^{(k+1)} - x^{(k)}) \\ x^{(k+1)} &= x^{(k)} - [J_g(x^{(k)})]^{-1} g(x^{(k)}) \end{aligned} \quad (58)$$

Notice that the Jacobian is evaluated before inverting. Due to the inefficiency in inverting large matrices the linear approximation would typically be solved using Gaussian elimination. In this case the Jacobian is only a 2×2 matrix so we use the standard matrix inversion formula $AA^{-1} = I$ for 2×2 matrices.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad (59)$$

The inverse of A is therefore:

$$\begin{aligned} A^{-1} &= \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, & \det A \neq 0 \\ &= \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, & ad - bc \neq 0 \end{aligned} \quad (60)$$

The linear approximation is iterated until some criterion is met. If the output of the function falls to within some absolute distance threshold ε_A of the solution, stop. If the input step size falls to some absolute distance threshold ε_A , stop. If the input step size falls to some distance threshold ε_R relative to the previous step size, stop. If computational time is restricted, stop after a limited number of iterations. It is common to use some combination of these - and possible other - criteria.

$$\|g(x^{(k+1)})\| \leq \varepsilon_A \quad (61)$$

$$\|x^{(k+1)} - x^{(k)}\| \leq \min(\varepsilon_A, \varepsilon_R \|x^{(k)} - x^{(k-1)}\|) \quad (62)$$

$$k + 1 \geq K \quad (63)$$

Every solution point of the system represents a constraint between the median and MAD of the normal mixture and the mean and variance of each component. From any constraint the mean of each component can be derived using 39 then the standard deviation using either 41 or 45. In case of multiple solutions, select the solution which best represents the constraint between the mean and variance of the mixture and each component using 3 and 9. While the numerical solution demonstrated is generally applicable, the constraints on Morse code allow the initial estimate to be unusually precise. Depending on the level of precision a solution to the system may not be necessary.

This model of Morse code is limited by its constraints. The further a sample deviates from expected Morse input, the more inaccurate the solution becomes. For an adaptive solution consider instead using an expectation-maximization algorithm.

References

- [NC93] Karl M. Newell and Daniel M. Corcos, eds. *Variability and motor control*. Champaign IL: Human Kinetics Publishers, 1993. ISBN: 9780873224246.
- [RC98] David A. Rosenbaum and Charles E. Collyer, eds. *Timing of behavior: neural, psychological, and computational perspectives*. Cambridge, Mass: MIT Press, 1998. ISBN: 9780262181884.
- [Lit00] Arthur Little. “Regulating Tempo in a Continuation Finger-Tapping Task”. PhD thesis. Kingston, RI: University of Rhode Island, 2000. DOI: [10.23860/diss-little-arthur-2000](https://doi.org/10.23860/diss-little-arthur-2000). URL: https://digitalcommons.uri.edu/oa_diss/928 (visited on 02/24/2022).
- [HFW21] David Hammerschmidt, Klaus Frieler, and Clemens Wöllner. “Spontaneous Motor Tempo: Investigating Psychological, Chronobiological, and Demographic Factors in a Large-Scale Online Tapping Experiment”. In: *Frontiers in Psychology* 12 (June 2021), p. 677201. ISSN: 1664-1078. DOI: [10.3389/fpsyg.2021.677201](https://doi.org/10.3389/fpsyg.2021.677201). URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.677201/full> (visited on 02/24/2022).