# Weighted Means and an Introduction to Maximum Likelihood Estimation

## I. INTRODUCTION

Hughes and Hase don't talk about the origin of the formula for the weighted mean. It's easy to derive the formula, and the derivation provides a simple introduction to the ideas behind the curve fitting that will be the next topic in our data analysis classes, so I'll do the derivation here. But before getting to that I'm going to do start with a simpler question: Why does averaging results give us the best estimate of the mean of the parent distribution? It seems intuitive, but there is a mathematical justification.

## II. SAMPLE MEAN AS MAXIMUM-LIKELIHOOD ESTIMATOR

**Assumption**: The data, $x_1, x_2, \ldots, x_N$, are random samples from a normal distribution with known standard deviation $\alpha$, and unknown mean $\bar{x}$ that we are trying to determine from the data.

If the mean is $\bar{x}$, the probability of obtaining a value of $x$ in the range $x_i$ to $x_i + dx$

$$p(x_i|\bar{x}) = \frac{1}{\alpha\sqrt{2\pi}} \exp\left[-\frac{(x_i - \bar{x})^2}{2\alpha^2}\right] dx, \tag{1}$$

and the probability of getting the whole data set, if the mean is $\bar{x}$, is proportional to

$$\begin{aligned} P(x_1, \ldots, x_N) &= p(x_1|\bar{x}) \times p(x_2|\bar{x}) \times \cdots \times p(x_N|\bar{x}) \\ &= \frac{1}{\alpha\sqrt{2\pi}} \exp\left[-\frac{(x_1 - \bar{x})^2}{2\alpha^2}\right] \times \cdots \times \frac{1}{\alpha\sqrt{2\pi}} \exp\left[-\frac{(x_N - \bar{x})^2}{2\alpha^2}\right], \\ &= \left(\frac{dx}{\alpha\sqrt{2\pi}}\right)^N \exp\left[-\frac{dx}{2\alpha^2}\sum_{i=1}^{N}(x_i - \bar{x})^2\right] \end{aligned} \tag{2}$$

The *maximum-likelihood estimate* of $\bar{x}$ is that value of $\bar{x}$ that makes the probability $P$ of Eq. (2) the largest. The constants out front are irrelevant, so this is the value of $\bar{x}$ that makes the magnitude of the argument of the exponential the smallest; you can see why this is called a *least squares* estimate.

To minimize the exponent, take a derivative with respect to $\bar{x}$, and set it equal to zero:

$$\frac{dP}{d\bar{x}} = 0 \qquad \longrightarrow \frac{dP}{d\bar{x}} = \left(\frac{dx}{\alpha\sqrt{2\pi}}\right)^N \exp\left[-\frac{dx}{2\alpha^2}\sum_{i=1}^{N}(x_i - \bar{x})^2\right] \left(-\frac{dx}{2\alpha^2}\sum_{i=1}^{N}2(x_i - \bar{x})(-1)\right) = 0 \tag{3}$$

It follows that

$$\sum_{i=1}^{N}(x_i - \bar{x}) = 0 \qquad \longrightarrow \qquad \sum_{i=1}^{N} x_i - \bar{x}\sum_{i=1}^{N} 1 = 0 \tag{4}$$

Solving this gives

$$(\bar{x})_{\rm MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i = \text{sample mean.} \tag{5}$$

The maximum-likelihood estimate of the parent mean is the sample mean. (Technically I should demonstrate that the extremum is a minimum, but I'll leave that up to you.)

## III. WEIGHTED MEAN AS MAXIMUM-LIKELIHOOD ESTIMATOR

**Assumption**: The data $x_1, x_2, \ldots, x_N$, are random samples from normal distributions with the same unknown mean $\bar{x}$ that we are trying to determine from the data, but the normal distributions do not have the same standard deviation. The standard deviations $\alpha_1$, $\alpha_2$, etc. are known.

In this case the probability $P$ of obtaining the whole data set, if the mean is $\bar{x}$, is

$$\begin{aligned} P(x_1, \ldots, x_N) &= p(x_1|\bar{x}) \times p(x_2|\bar{x}) \times \cdots \times p(x_N|\bar{x}) \\ &= \frac{dx}{\alpha_1\sqrt{2\pi}} \exp\left[-\frac{(x_1-\bar{x})^2}{2\alpha_1^2}\right] \times \cdots \times \frac{dx}{\alpha_N\sqrt{2\pi}} \exp\left[-\frac{(x_N-\bar{x})^2}{2\alpha_N^2}\right], \\ &= \prod_{i=1}^{N}\left(\frac{dx}{\alpha_i\sqrt{2\pi}}\right) \times \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{\alpha_i^2}\right] \end{aligned} \tag{6}$$

Again, this probability is maximized by minimizing the exponent. In this case when we set $\frac{dP}{d\bar{x}} = 0$ we obtain

$$\begin{aligned} 0 &= \sum_{i=1} \frac{(x_i - \bar{x})}{\alpha_i^2} \\ &= \sum_{i=1}^{N}\frac{x_i}{\alpha_i^2} - \bar{x}\sum_{i=1}^{N}\frac{1}{\alpha_i^2}. \end{aligned} \tag{7}$$

Solving for $\bar{x}$ gives

$$(\bar{x})_{\rm MLE} = \frac{\sum \frac{x_i}{\alpha_i^2}}{\sum \frac{1}{\alpha_i^2}} = x_{\rm CE} = \frac{\sum w_i x_i}{\sum w_i}. \tag{8}$$

where $w_i = \frac{1}{\alpha_i^2}$. This is Eq. (4.30) in Hughes and Hase. In this case the maximum-likelihood estimate of the parent mean is the weighted mean.

## IV. GENERALIZE MAXIMUM-LIKELIHOOD ESTIMATOR

We can now generalize the insight we got from Eq. (6). So far we had on the $x$-axis the label $i$ and on the $y$-axis $x_i$ and our fit function was $\bar{x}$. Next we generalize to having data points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ and any fit function $f(x; c_1, c_2, \ldots)$, where $c_1, c_2$ etc. are the fit parameters. In the case of a line fit $f(x) = a + m\,x$, so $c_1 = a$ and $c_2 = m$. For a fit we try to determine the best fit parameters. For a line the best $a$ and $m$, and in the general case the best fit parameters $c_1, c_2, \ldots$. The argument which led to Eq.(6) is the same, we determine $P((x_1, y_1), \ldots, (x_N, y_N))$, the probability for obtaining for the data point at $x_i$ the measurement $y_i$ compared to the function value $y = f(x_i; c_1, c_2, \ldots)$. In Eq. (6) we therefore replace $x_i$ with $y_i$ and $\bar{x}$ with $f(x_i; c_1, c_2, \ldots)$

$$
\begin{aligned}
P(x_1, \ldots, x_N) &= \prod_{i=1}^{N} \left( \frac{dx}{\alpha_i \sqrt{2\pi}} \right) \times \exp\left[ -\frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - f(x_i; c_1, c_2, \ldots))^2}{\alpha_i^2} \right] \\
&= \prod_{i=1}^{N} \left( \frac{dx}{\alpha_i \sqrt{2\pi}} \right) \times \exp\left[ -\frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - y(x_i))^2}{\alpha_i^2} \right] \\
&= \prod_{i=1}^{N} \left( \frac{dx}{\alpha_i \sqrt{2\pi}} \right) \times \exp\left[ -\frac{1}{2} \chi^2 \right]
\end{aligned}
\tag{9}
$$

Therefore when we search for the parameters $(c_1, c_2, \ldots)$ for which the probability $P$ is largest, so maximizing $P$, corresponds to minimizing $\chi^2$.

Note that with the above approach you can derive the equations for the weighted least squares fit, that is Eqs. (6.3)—(6.7) of Hughes and Hase.