# Likelihood of mean vs replicates

**Peter Ashcroft**
Institute for Integrative Biology
ETH Zürich
peter.ashcroft@env.ethz.ch
ashcroftp.github.io

**Is the likelihood of the a set of replicate observations equivalent to the likelihood of the mean of those replicates?**

## Question formulation

Suppose we have a dataset $\{z_1, z_2, \ldots, z_n\}$, for which we have a model $\theta$ which predicts a value $\widehat{z}$ to match these observations. Assuming the $z_i$ are normally distributed, the likelihood of observing this data given the model $\theta$ is

$$\mathcal{L}(\{z_1, z_2, \ldots, z_n\}|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i - \widehat{z})^2}{2\sigma^2}\right), \tag{1}$$

where $\sigma$ is the observation error, which we assume to be given by the standard deviation of $\{z_1, z_2, \ldots, z_n\}$. The log-likelihood is given by

$$\ln \mathcal{L}(\{z_1, z_2, \ldots, z_n\}|\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(z_i - \widehat{z})^2}{2\sigma^2}.$$

Alternatively, instead of considering each individual replicate of data, we could instead fit just the mean $\overline{z} = (\sum_{i=1}^{n} z_i)/n$, with the likelihood now given by

$$\mathcal{L}(\overline{z}|\theta) = \frac{1}{\sqrt{2\pi\overline{\sigma}^2}} \exp\left(-\frac{(\overline{z} - \widehat{z})^2}{2\overline{\sigma}^2}\right), \tag{2}$$

where $\overline{\sigma}$ is the error of the mean. The log-likelihood is given by

$$\ln \mathcal{L}(\overline{z}|\theta) = -\frac{1}{2}\ln(2\pi\overline{\sigma}^2) - \frac{(\overline{z} - \widehat{z})^2}{2\sigma^2}.$$

The question I ask here is, are these two formulations equivalent?

## Solution

We can manipulate Eq. (1) to produce the following expression:

$$\mathcal{L}(\{z_1, z_2, \ldots, z_n\}|\theta) = \frac{1}{\left(2\pi n \frac{\sigma^2}{n} e\right)^{\frac{n}{2}}} \exp\left(-\frac{(\langle z\rangle - \widehat{z})^2}{2\frac{\sigma^2}{n}}\right). \tag{3}$$

where $\sigma/\sqrt{n}$ is the standard error of the mean (SEM).

We see that the exponential term of the likelihood has a direct agreement with $\mathcal{L}(\overline{z}|\theta)$ [Eq. (2)] using $\overline{\sigma} = \sigma/\sqrt{n}$. However, the prefactor is incomparable between the two formulations.

Therefore, there is no direct equivalence between the likelihood of the individual replicates and the likelihood of the mean.

## Explicit calculation

Explicitly, we use the following expressions:

$$
\begin{aligned}
\langle z \rangle &= \frac{\sum_{i=1}^{n} z_i}{n}, \\
\left\langle z^2 \right\rangle &= \frac{\sum_{i=1}^{n} z_i^2}{n}, \\
\mathrm{var(z)} &= \left\langle z^2 \right\rangle - \langle z \rangle^2 = \sigma^2,
\end{aligned}
$$

and then Eq. (1) can be treated as

$$
\begin{aligned}
\mathcal{L}(\{z_1, z_2, \ldots, z_n\}|\theta) &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(z_i - \widehat{z})^2}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(z_i^2 - 2z_i\widehat{z} - \widehat{z}^2)}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{(n\left\langle z^2 \right\rangle - 2n\langle z \rangle \widehat{z} - n\widehat{z}^2)}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{n(\mathrm{var(z)} + \langle \mathrm{z} \rangle^2 - 2\langle \mathrm{z} \rangle \widehat{\mathrm{z}} - \widehat{\mathrm{z}}^2)}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{n\,\mathrm{var(z)}}{2\sigma^2}\right) \exp\left(-\frac{n(\langle z \rangle^2 - 2\langle z \rangle \widehat{z} - \widehat{z}^2)}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{n\sigma^2}{2\sigma^2}\right) \exp\left(-\frac{n(\langle z \rangle - \widehat{z})^2}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2 e\right)^{-\frac{n}{2}} \exp\left(-\frac{(\langle z \rangle - \widehat{z})^2}{2\frac{\sigma^2}{n}}\right).
\end{aligned}
$$