VIII. Signal Estimation

Reading:
Event-Related Potentials (ERPs) are electrical potentials emitted by the brain in relation to a particular paradigm event.

A simple example of an ERP is a Sensory Evoked Potential (SEP) which is a potential that occurs as a direct result of sensory stimulation.

The ERP is time-locked because it always occurs at a particular time in a paradigm. For example, a SEP is time-locked to the sensory stimulus.
The recorded time series is considered to be the sum of the ERP (signal) which arises in the brain in direct conjunction with the paradigm event and noise which comes from a number of sources and is not related to the event.

The need for signal estimation techniques arises because the amplitude of ERPs is usually similar to or less than that of the noise.
Noise is defined as that part of the data that is not the signal (ERP in this case).
Noise comes from many sources. Every stage of data collection contributes noise to the data.
In the case of scalp recording, noise consists of:

1. cephalic noise -- potentials from parts of the brain other than that generating the signal.
2. extracephalic cranial noise -- potentials from scalp & neck muscles, skin, eye movements, tongue movements, etc.
3. extracranial physiological noise -- potentials from parts of the body below the head, e.g. EKG.
4. thermal noise -- random fluctuation at the electrode.
5. movement artifacts -- noise introduced by subject movement.
6. electronic noise -- fluctuations introduced by the electronics of the amplifier and A-D converter.
7. environmental noise -- e.g. 60-Hz line noise.
8. quantization noise -- noise due to conversion of analog signal to amplitude discrete signal.
Superimposition was one of the earliest techniques for detecting the ERP signal. It can show the similarity and variability of successive data traces. The signal is emphasized by the line-to-line correlation of the traces. Superimposition is still a useful preliminary step in data analysis.
Averaging for Signal Estimation

Averaging consists of adding together the set of recorded waveforms associated with each repetition of the paradigm. The sum is divided by the number of realizations.

The most direct way of estimating the (ERP) signal $s(t)$ is by averaging over a sample of $N$ data waveforms to produce the sample mean.
Intuitive explanation for why averaging provides an estimate of the signal

Averaging provides an estimate of the signal because the signal (which is always the same sign on each realization) does not cancel in averaging, whereas the noise (because its distribution has 0 mean) tends to cancel on average. In averaging, the sum of the signal grows in direct proportion to the number of repetitions.
When this sum across realizations at a time point is divided by N, the signal is estimated. The sum of the noise increases less rapidly due to cancellation effects. When the noise sum is divided by N, the averaged noise term falls below the original noise level. In other words, averaging will attenuate the noise, but not the signal. Therefore, the average will approximate the signal with an error due to the residual noise. A more formal explanation follows.
Fig. 2. Superimposition of 16 traces on Y-T planer; average shown below. This and all subsequent figures (except Fig. 4) were produced by the LENC 8 computer at the Barion Neurological Institute, Bristol (Great Britain).

Fig. 3. A large signal-to-noise improvement is obtained when the noise is composed of brief transients. a: Isolated sine wave and noise. b: Superimposition of 9 traces. c: Average of 9 traces (from Cooper et al. 1969).
Fig. 4. A threefold improvement in signal-to-noise is obtained when noise is continuous. 

a: Isolated sine wave signal. 
b: Continuous sine wave ("noise"). 
c: Superimposition of 9 traces of signal and noise. 
d: Average of 9 traces (from Cooper et al. 1969).
Statistical Model of Signal + Noise

To estimate the signal in the presence of noise, we make the following assumptions:

1. At any time point, we have a sample set, consisting of $N$ repetitions of the data time series, $\{x_i(t)\}$, which we assume is taken from a larger set of realizations generated by a stochastic process.

2. The signal (ERP) and noise time series linearly sum to produce the data time series.
3. The signal time series, s(t), is deterministic, and is therefore identical for each repetition.
4. The noise time series are considered to be statistically independent realizations that are generated by a noise stochastic process having mean ($\mu_n$) equal to zero.
These assumptions are expressed as the following mathematical model:

\[ x_i(t) = s(t) + n_i(t) \quad i = 1, 2, \ldots, N \]
\[ 0 \leq t \leq T \]

1. \( x_i(t) \) is the data time point on the \( i \)-th realization
2. \( s(t) \) is the signal
3. \( n_i(t) \) is the noise during the time epoch of the \( i \)-th realization
4. \( N \) is the number of realizations
5. \( T \) is the duration of the time epoch over which each waveform is averaged
An important implication of the model:

Since the signal itself does not vary from one realization to the next, the variance of the data process is entirely due to the noise variance.
By averaging the data time series, we are computing the sample mean to estimate the (true) population mean. This statistic is defined as:

\[ s(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) = s(t) + \frac{1}{N} \sum_{i=1}^{N} n_i(t) \]

We see that the size of the sample mean depends on 2 terms: the magnitude of \( s(t) \) and the magnitude of the average of the noise waveforms.
We want to know whether computing the sample mean is a good way to estimate the true population mean, \( s(t) \). In other words, is the sample mean a good estimator of the population mean? We first consider the expected value of the sample mean:

\[
E\left[\hat{s}(t)\right] = E\left[\frac{1}{N} \sum_{i=1}^{N} x_i(t)\right]
\]

\[
E\left[ s(t) \right] = E\left[ s(t) \right] + E\left[ \frac{1}{N} \sum_{i=1}^{N} n_i(t) \right]
\]
The expected value of the signal is the signal itself:

\[ E[s(t)] = E \left[ \frac{1}{N} \sum_{i=1}^{N} s(t) \right] = s(t) \]

The noise process has a mean of zero. Therefore:

\[ E \left[ \frac{1}{N} \sum_{i=1}^{N} n_i(t) \right] = \frac{1}{N} \sum_{i=1}^{N} E[n_i(t)] = \frac{1}{N} \sum_{i=1}^{N} 0 = 0 \]

for all \( i \)
It follows that:

\[
E \left[ \hat{s}(t) \right] = s(t)
\]

This equation demonstrates that the sample mean is an unbiased estimate of the population mean, which is the signal \( s(t) \). The sample mean equals \( s(t) \) only as \( N \) goes to infinity. For finite \( N \), the sample mean will approximate the true signal, with an error due to residual noise.
Variability of the Sample Mean

In order to determine the degree of reliance that can be placed on the sample mean waveshape, one needs to obtain a measure of its variability. If the variability is low then more confidence can be placed in the detail of the waveform, whereas if it is high then only the general shape can be considered due to the signal.
A qualitative method for estimating the variability is to subdivide the set of available waveforms into subsets, compute the average for each subset, and compare them by superimposition. The degree of reproducibility of the averages is a measure of background noise.

However, a quantitative method is required to obtain a statistical measure of variability. A quantitative measure of this variability is provided by the variance of the distribution of sample means.
The associated standard deviation of the distribution of sample means, called the standard error of the sample mean, measures the standard distance between the sample mean and the population mean of the stochastic process, which is the signal $s(t)$. This distance can also be considered to measure the degree of error of the sample mean in estimating the (true) population mean.
The variance of the distribution of sample means is the square of the expected value of the difference between the sample mean and the population mean:

\[
\text{Var} \left( \hat{s}(t) \right) = \sigma_{\hat{s}}^2(t) = E \left[ \left( \hat{s}(t) - E \left[ \hat{s}(t) \right] \right)^2 \right] \\
E \left[ \hat{s}(t) \right] = s(t) \rightarrow \sigma_{\hat{s}}^2(t) = E \left[ \left( \hat{s}(t) - s(t) \right)^2 \right] \\
\hat{s}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) = s(t) + \frac{1}{N} \sum_{i=1}^{N} n_i(t) \rightarrow \rightarrow 
\]
\[ \sigma_s^2(t) = E \left[ \left( \frac{1}{N} \sum_{i=1}^{N} n_i(t) \right)^2 \right] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E \left[ n_i(t) n_j(t) \right] \]

where we must consider the product of noise values for different realizations \( i \) and \( j \).
However, since the values of the different realizations \((i\) and \(j)\) are uncorrelated:

\[
\sigma_s^2(t) = \frac{1}{N^2} \sum_{i=1}^{N} \left[ E \left[ n_i(t)^2 \right] \right]
\]
The variance of the noise distribution, $\sigma_n^2(t)$, is defined as $E[(n_i(t) - \mu_n)^2]$, but since the mean of the noise process ($\mu_n$) is zero, the noise variance ($\sigma_n^2(t)$) equals $E[n_i(t)^2]$. Thus, substituting and moving $\sigma_n^2(t)$ before the summation:

$$\sigma_{s}^2(t) = \frac{1}{N^2} \sum_{i=1}^{N} E[n_i(t)^2] = \frac{1}{N^2} \sigma_n^2(t) \sum_{i=1}^{N} 1 = \frac{\sigma_n^2(t)}{N}$$

$$\sigma_{s}^2(t) = \frac{\sigma_n(t)}{\sqrt{N}}$$
This result indicates that the averaging process attenuates the noise to a residual value that is directly proportional to the intensity of the background noise, and inversely proportional to the square root of the number of trials (N).

In theory, the residual noise can be made arbitrarily small by increasing the number of realizations. In practice, there are limits on the number of repetitions of the paradigm that can be performed, so there is always a standard error to deal with.
Confidence Intervals

Since the sample mean estimate can be closer or farther from the true signal, depending on the noise variance and the number of samples, we would like to have a statement about our confidence in estimating the signal (i.e. the true mean). We can do this by computing a range of values within which we can state that the true mean value lies with a given probability.
The distribution of sample means has an unknown mean equal to the true mean, and a standard deviation which is the standard error.

We can use the distribution of sample means to determine a confidence interval for the true mean, $s(t)$.

If we assume that the noise distribution is normal (Gaussian), then the data distribution is also normal, and so is the distribution of sample means. [If the noise distribution is not normal, the Central Limit Theorem states that the distribution of sample means will still be normal if the sample size $N$ is at least equal to 30].
Because the distribution of sample means is normal, we can convert the sample mean to a standard z-score. The z value is determined by subtracting the population mean from the sample mean, and dividing by the population standard error:

\[
Z = \frac{\left(\hat{s} - s\right)}{\sigma_{\hat{s}}} = \frac{\hat{s} - s}{\left(\frac{\sigma}{\sqrt{N}}\right)}
\]
Since $z$ is a normal random variable with zero mean and unity standard deviation, its probability density, $f(z)$ is known exactly:

$$f(z) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{-z^2/2}$$

This function gives the probability that $z$ lies within any range. The probability will be the integral of $f(z)$ using those values as limits. Suppose, for example, that we want the limits of the interval which has 0.95 probability of containing $z$. We obtain this by defining a confidence interval around 0 equal to $4\sigma$ ($\pm 2\sigma$ covers the middle 95% of the $z$ distribution).
The range of $z$ for which the area under the unit normal curve is 0.95 is bounded by $z = -1.96$ and $z = 1.96$. Thus:

$$p(-1.96 \leq z \leq 1.96) = \int_{-1.96}^{1.96} f(z)\,dz = 0.95$$
We can now substitute the definition of \( z \) in this equation to find the 95% confidence limits for the sample mean:

\[
Z = \frac{(\hat{s} - s)}{\sigma_{\hat{s}}} = \frac{(\hat{s} - s)}{\left(\frac{\sigma_n}{\sqrt{N}}\right)}
\]

\[
p \left[ -1.96 \leq \frac{(\hat{s} - s)}{\left(\frac{\sigma_n}{\sqrt{N}}\right)} \leq 1.96 \right] = 0.95
\]
Rearranging, we get:

\[
p \left[ \left( \hat{s} - 1.96 \frac{\sigma}{\sqrt{N}} \right) \leq s \leq \left( \hat{s} + 1.96 \frac{\sigma}{\sqrt{N}} \right) \right] = 0.95
\]

This is the 95% confidence interval for s. If the confidence interval is computed for each time point, the result is a pair of waveforms between which the true ERP signal is expected to lie with a specified probability.
Estimating the Noise Variance

\[
p\left[\left(\hat{s} - 1.96 \frac{\sigma_n}{\sqrt{N}}\right) \leq s \leq \left(\hat{s} + 1.96 \frac{\sigma_n}{\sqrt{N}}\right)\right] = 0.95
\]

The problem with this equation is that the standard deviation of the noise process, \(\sigma_n\), is not usually known.

However, there is a method for estimating the variance of the noise process.
An estimate of the variance of the noise process is obtained from the mean squared difference between the estimated signal (i.e. average) and each individual waveform:

\[
\hat{\sigma}_n^2(t) = \frac{1}{N-1} \sum_{i=1}^{N} \left[ x_i(t) - \hat{S}(t) \right]^2
\]

\[
= \frac{1}{N-1} \left[ \sum_{i=1}^{N} x_i(t)^2 - 2\hat{S}(t) \sum_{i=1}^{N} x_i(t) + \hat{S}(t)^2 \sum_{i=1}^{N} 1 \right]
\]

(It can be shown that the use of N-1 rather than N in this equation gives an unbiased estimate.)
We now use the fact that:

\[ \sum_{i=1}^{N} x_i(t) = N \hat{s}(t) \]

Substituting this into the previous equation gives:

\[ \hat{\sigma}_n^2(t) = \frac{1}{N-1} \left[ \sum_{i=1}^{N} x_i^2(t) - 2N \hat{s}(t)^2 + N \hat{s}(t)^2 \right] \]

\[ = \frac{\sum_{i=1}^{N} x_i^2(t) - N \hat{s}(t)^2}{N-1} \]
This equation gives an estimate of the noise variance based on quantities that derive directly from the data:

\[ \hat{\sigma}_n^2(t) = \frac{\sum_{i=1}^{N} x_i^2(t) - N\hat{s}(t)^2}{N - 1} \]

This equation gives an estimate of the noise variance based on quantities that derive directly from the data: \( N, x(t), \text{and} \hat{s}(t). \)
Now that we have an expression that estimates the noise variance, we can use it to estimate the standard error:

\[
\hat{\sigma}_s(t) = \frac{\hat{\sigma}_n(t)}{\sqrt{N}}
\]

This estimated standard error provides an estimate of the standard distance between the sample mean and the population mean.
To use this estimated standard error to get confidence intervals, we would like to use it in the expression for $z$:

$$Z = \left( \frac{\hat{s} - s}{\sigma_{\hat{s}}} \right) = \left( \frac{\hat{s} - s}{\frac{\sigma}{\sqrt{N}}} \right)$$
However, we cannot do this. We must use the estimated standard error:

\[
\hat{\sigma}_s(t) = \frac{\hat{\sigma}_n(t)}{\sqrt{N}}
\]

rather than the true standard error:

\[
\sigma_s(t) = \frac{\sigma_n(t)}{\sqrt{N}}
\]
The statistic that is based on the estimated standard error is known to conform to the Student's t distribution with N-1 degrees of freedom rather than the z distribution (as with the true standard error).
Therefore, in place of $z$:

$$
Z = \frac{(\hat{S} - S)}{\sigma_{\hat{S}}} = \frac{(\hat{S} - S)}{\left(\frac{\sigma_n}{\sqrt{N}}\right)}
$$

we use $t$:

$$
t = \frac{(\hat{S} - S)}{\hat{\sigma}_{\hat{S}}} = \frac{(\hat{S} - S)}{\left(\frac{\hat{\sigma}_n}{\sqrt{N}}\right)}
$$
We can use the t-statistic in the same way as the z-statistic to estimate the confidence interval for the population mean.

As an example, suppose \( N=30 \), and we want the 95% confidence interval.

Then, from the Student's t table:

\[
p\left(-2.045 \leq t \leq 2.045\right) = \int_{-2.045}^{2.045} f(t, 29) \, dt = 0.95
\]
Substituting for $t$ and rearranging, we obtain:

$$p\left[\left(\hat{s}(t) - 2.045 \frac{\hat{\sigma}_n(t)}{\sqrt{30}}\right) \leq s(t) \leq \left(\hat{s}(t) + 2.045 \frac{\hat{\sigma}_n(t)}{\sqrt{30}}\right)\right] = 0.95$$
We can determine the confidence interval for the population mean (the signal, s(t)) from this equation.

Of course, we must know the sample mean and the estimate of the standard deviation of the noise. But we already have expressions for these:

$$\hat{s}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t)$$

$$\hat{\sigma}_n(t)^2 = \frac{\sum_{i=1}^{N} x_i(t)^2 - N\hat{s}(t)^2}{N - 1}$$