Experimental Methods

Alexander Byrne

Michaelmas 2020

1 Impedances

In the linear regime, all *real* voltage sources (transducers) can be represented by an *ideal* voltage source in series with an output impedance Z_o. This can be deduced experimentally by joining the terminals of the source when it produces a known voltage V and measuring the current I, which will be equal to V/Z_o.

An oscilloscope is modelled as an input impedance Z_i in parallel with a perfect voltmeter, reading V_i. Measuring the output voltage V_o of a transducer with a scope is like a circuit with the two in series; the one in the slides. We then obtain:

$$V_i = V_o \frac{Z_i}{Z_i + Z_o}$$

So if we want the oscilloscope to accurately record the output of the transducer, we require $Z_i \gg Z_o$. The Z_i is more accurately modelled as a resistance in parallel with a capacitance, which causes a problem at high frequencies. This is compensated for by using a scope probe, which removes the frequency dependence at the cost of decreasing the voltage by a certain factor, often 10 or 100.

For a more general "black box", we can define the input impedance of it as the input voltage divided by the input current. For a buffer connection, the input current is 0A, because op-amps draw essentially no current (see later), so their input impedance is very large; their output impedance turns out to be very low. This is why buffers are sometimes used to connect two circuits but ramp up the impedance for some reason.

If instead we wish to measure the current, we instead want Z_i to be *low*. If we wish to transfer as much power as possible, we want $Z_i = Z_o$ (impedance matching).

Because for an oscilloscope Z_i is so high, very little power is transferred to it. In order for signals to be detected, therefore, they first must be amplified...

2 Op-amps

For an ideal op-amp, the terminals draw no current, and the terminals are at the same voltage. You can then solve all the circuits with feedback built into them. Depending on what components are used in an op-amp circuit, they can function as inverters, integrators, differentiators, filters, summers, and more.

Non-ideal op-amps have an amplification (A) which is finite, complex, and frequency-dependent, a finite r_i (so draws a lil current), and a non-zero R_o (so there's a slight voltage drop going out). In the limit of quite large A, r_i, and quite small R_o, we obtain similar results, finding that $Z_o \propto R_o/A$ and $Z_i \propto r_i A$, so the approximation of an ideal op-amp is a good one.

Op-amps' amplifications have a gain inversely proportional to frequency at high frequency. This is because at high frequencies, capacitative components might cause a voltage to be inverted, effectively swapping negative feedback to positive. As most normal things (e.g. square wave input) have high-frequency components, this is very likely to happen, so the open-loop gain A is damped so that for high frequencies it is safely less than 1 and saturation does not occur.

In general, feedback is written:

$$\begin{aligned} V_o &= A(V_i + \beta V_o) \\ \Rightarrow V_o &= \frac{A}{1 - A\beta} V_i \end{aligned}$$

At high A, this is becomes independent of any small changes in A. If β depends on frequency, then any component at the frequency where $\beta = A^{-1}$ will cause the op-amp to quickly saturate, generating a square wave as the input at the special frequency flips back and forth.

3 Errors etcetera

All measurements have an *error*, equal to the *measured* value minus the *true* value, which is never known.

Random errors have an average value of 0. The measured values can be used to estimate the true value, which is what stats was invented for. *Systematic errors* are all the other errors, i.e. ones that are constant or drift over time. These can only be removed by good experimental design, and have a variety of sources so are difficult to eliminate.

Random errors are often assumed to be Gaussian-distributed; the uncertainty is quoted as the standard distribution σ of the distribution. The mean

is estimated by:

$$\bar{X} = \frac{1}{N}\sum_i X_i$$

The variance of any set is the mean square deviation from the mean:

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

Where $\mu = \mathrm{E}(X)$. The variance of two variables is simply the sum of the variances:

$$\begin{aligned}\mathrm{Var}(X_1 + X_2) &= \mathrm{E}\left[(X_1 + X_2 - \mu_1 - \mu_2)^2\right] \\ &= \mathrm{E}\left[(X_1-\mu_1)^2 + 2(X_1-\mu_1)(X_2-\mu_2) + (X_2-\mu_2)^2\right] \\ &= \mathrm{Var}(X_1) + 2\mathrm{Cov}(X_1, X_2) + \mathrm{Var}(X_2) \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)\end{aligned}$$

where the last line follows from the independence of the variables, so that the two have zero covariance. Similarly, the variance of N times a random variable is:

$$\begin{aligned}\mathrm{Var}(NX) &= \mathrm{E}\left[(NX - N\mu)^2\right] \\ &= N^2\mathrm{E}\left[(X-\mu)^2\right] \\ &= N^2\mathrm{Var}(X)\end{aligned}$$

Now consider taking the variance of the sum of N variables drawn independently from the same distribution (i.e. $\mathrm{E}(X_i) = \mathrm{E}(X)$, $\mathrm{Var}(X_i) = \mathrm{Var}(X)$, $i \in \{1, ..., N\}$). This quantity is:

$$\begin{aligned}\mathrm{Var}\left(\sum_i X_i\right) &= \sum_i \mathrm{Var}(X_i) \\ \mathrm{Var}(N\bar{X}) &= N\mathrm{Var}(X) \\ N^2\mathrm{Var}(\bar{X}) &= N\mathrm{Var}(X) \\ \mathrm{Var}(\bar{X}) &= \frac{1}{N}\mathrm{Var}(X)\end{aligned}$$

The mean of $\bar{X}$ and X is the same (μ), so taking the mean of many measurements just reduces the variance by a factor of N, and the standard deviation by a factor of $\sqrt{N}$.

Given a set of data, it is quite easy to calculate the mean. One might assume that the best way to calculate the variance is:

$$s^2 = \frac{1}{N}\sum_i (X_i - \bar{X})^2$$

but because $\bar{X}$ is calculated *for the purpose of being close to the* X_i, $(X_i - \bar{X})^2$ will *obviously* be quite low - and the deviations from the true mean are likely to be higher. Consider the expression:

$$\sum_i (X_i - \bar{X})^2$$

Its expected value is given by:

$$\begin{aligned}
\mathrm{E}\sum_i (X_i - \bar{X})^2 &= \mathrm{E}\sum_i (X_i - \mu + \mu - \bar{X})^2 \\
&= \mathrm{E}\sum_i \left[(X_i - \mu)^2 + 2(\mu - \bar{X})(X_i - \mu) + (\mu - \bar{X})^2\right] \\
&= \sum_i \mathrm{E}\left[(X_i - \mu)^2\right] - 2\mathrm{E}\left[(\bar{X} - \mu)\sum_i (X_i - \mu)\right] + \sum_i \mathrm{E}\left[(\mu - \bar{X})^2\right] \\
&= \sum_i \mathrm{Var}(X) - 2\mathrm{E}\left[N(\bar{X} - \mu)^2\right] + \sum_i \mathrm{Var}(\bar{X}) \\
&= N\mathrm{Var}(X) - 2N\mathrm{Var}(\bar{X}) + N\mathrm{Var}(\bar{X}) \\
&= (N-1)\mathrm{Var}(X)
\end{aligned}$$

so s^2 would be off by a factor of $(N-1)/N$. The best estimate from a sample of the true variance of an underlying distribution is therefore given by:

$$\sigma^2 = \frac{1}{N-1}\sum_i (X_i - \bar{X})^2$$

since the expected value of this estimator is $\mathrm{Var}(X)$. The variance in the mean is then given by:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N} = \frac{1}{N(N-1)}\sum_i (X_i - \bar{X})^2 = \frac{1}{N-1}s^2$$

3.1 Error propagation

If we have measured Gaussian uncertainties in $x, y, ...$, and we wish to measure the uncertainty of a function $f(x, y, ...)$, it is given by:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$$

Important special cases of this are given here;

$$f = x + y \qquad\qquad \Rightarrow \sigma_f^2 = \sigma_x^2 + \sigma_y^2$$

$$f = x^n \qquad \Rightarrow \frac{\sigma_f}{f} = |n|\frac{\sigma_x}{x}$$

$$f = xy \text{ or } \frac{x}{y} \qquad \Rightarrow \frac{\sigma_f^2}{f^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}$$

For particularly complicated functions, just evaluate $\sigma_x(\partial f/\partial x)$ "empirically", i.e. as $f(x + \sigma_x) - f(x)$.

3.2 Systematic Errors

Some general tactics:

- Calibrate instruments by seeing if they gave a response you know they should
- Exploit symmetries (e.g. swapping leads)
- Recognise any prejudices you may have, and design experiments to avoid them
- Use null methods - it is easy to distinguish something from nothing
- Beware of changes in time - measure in the order ABCCBA rather than AABBCC etc.
- Differential measurements against a standard or known value
- Beware of selection effects - make sure your experiment doesn't exclude anything by its actual design

4 Signal Processing

It is much easier to sample digitally (at a set of fixed values) than analogue (continuously). There is a minimum rate of sampling required for a given signal, because an under-sampled high-frequency wave just appears as a low-frequency wave, a phenomenon known as *aliasing*. This rate is *double the highest-frequency component in the signal*. Otherwise, when sampling (AKA convoluting with an inverse-frequency Dirac comb in frequency space) the convolution images will overlap and frequency information is lost.

One can sample at less than the Nyquist rate if all the frequency components are within a small bandwidth at high frequency. In that case, we can just shift the frequency spectrum down to the origin, and we can then sample at just double the *bandwidth*.

Independently of this, the *spectral resolution* of a sampler (i.e. the smallest differences in frequency that can be distinguished) Δf is given by $\Delta f = 1/T$ where T is the amount of time for which we have sampled the function.

Digitisation also involves *quantisation* of the y-values. Quantisation at an infinitely fine scale is not possible, but it isn't desirable either because there is no point quantising further if all it will reveal is noise. N-bit sampling involves quantisation into 2^N levels, or bins. Quantisation also be used to reduce noise. If one samples at 4 times the Nyquist rate (i.e. *8* times the maximum-frequency component), the four values obtained can be averaged, effectively reducing the noise of the signal by a factor of 2 ($= \sqrt{4}$).

5 Noise

If the signal and the noise have non-overlapping spectra, *filters* can be used to remove noise. A filter which rises and falls quickly around the signal is ideal. However, often there is noise present at all frequencies - pink noise goes as $1/f$; shot noise and Johnson noise are white, constant at *all* frequencies. Noise is usually lowest at high frequencies, so signals at low frequencies should be encoded at high frequencies for reliable transmission. The phase-sensitive detector is a good way of doing this.

Experiments often need to be shielded from vibration, which might be caused by people moving about, seismic noise etc. We model whatever the experiment is resting on as a damped harmonic oscillator, with a certain ω_0. If the frequency of the noise (forcing) is much higher than ω_0, there will be very little response, so a first tactic would be to decrease ω_0, most easily by decreasing k, such as with an air cushion. For an air cushion 0.2m thick, ω_0 comes out to about 1Hz, so pretty much all noise will provoke almost no response.

Eliminating thermal noise is often important in quantum experiments, where one system is often very cold and must be shielded from the hotter surroundings. The first steps are to reduce evaporation (using a lid), conduction (insulate), and convection (vacuum). Reducing radiation is often more complicated. The net radiation flux between the hot area and the cold area is $\sigma\epsilon(T_h^4 - T^4)$. Often this is most easily reduced by reducing ϵ, using shiny surfaces like polished foils (which have $\epsilon \approx 0.03$), but note that ϵ is often a function of wavelength, and it only really matters what ϵ does at the peak of the black-body curve $\lambda_p = b/T$. Another option is to put a heat shield around the apparatus. In the steady state, the screen reaches a temperature T_s and the output flux is equal to the input flux, i.e.:

$$2T_s^4 = T_h^4 + T^4$$

So the net heat flux to the apparatus is then:

$$\begin{aligned}&\frac{1}{2}\sigma\epsilon(T_h^4 + T^4) - \sigma\epsilon T^4\\ &= \frac{1}{2}\sigma\epsilon(T_h^4 - T^4)\end{aligned}$$

1/2 of what it was without the screen. One can further show that introducing n screens reduces the incident radiation by a factor of $n + 1$. A final way to reduce the total incident heat is to reduce the area of the apparatus, making it spherical where possible.

Several strategies are available for reducing noise in electrical circuits. Twisting wires means that wires travel essentially along the same path, meaning that stray E and B fields have no net effect. Avoiding earth loops reduces noise from varying magnetic fields. Faraday cages lead to no E fields being present within, and high-μ shields can reduce B fields within by a factor of about 100.

6 Presenting your work or something

Hopefully won't come up on the exam

7 Probability Distributions

7.1 Binomial

When a trial with two possible outcomes (with probabilities p, $1 - p$) is repeated N times, how many of each outcome occurs in total?

$$P(r) = \frac{N!}{r!(N-r)!}p^r(1-p)^{N-r}$$

This distribution is normalised:

$$\begin{aligned}\sum_{r=0}^{N} P(r) &= \sum_{r=0}^{N} \frac{N!}{r!(N-r)!}p^r(1-p)^{N-r}\\ &= (p + (1-p))^N = 1\end{aligned}$$

has a mean of Np:

$$\langle r\rangle = \sum_{r=0}^{N} rP(r)$$

$$
\begin{aligned}
&= \sum_{r=0}^{N} r \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=1}^{N} r \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=1}^{N} \frac{N!}{(r-1)!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=0}^{N-1} \frac{N!}{r!(N-1-r)!} p^{r+1} (1-p)^{N-1-r} \\
&= Np \sum_{r=0}^{N-1} \frac{(N-1)!}{r!(N-1-r)!} p^r (1-p)^{N-1-r} \\
&= Np(p+(1-p))^{N-1} = Np
\end{aligned}
$$

and a variance of $Np(1-p)$:

$$
\begin{aligned}
\langle r^2 \rangle &= \sum_{r=0}^{N} r^2 P(r) \\
&= \sum_{r=0}^{N} r^2 \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=1}^{N} r^2 \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=1}^{N} r \frac{N!}{(r-1)!(N-r)!} p^r (1-p)^{N-r} \\
&= \sum_{r=0}^{N-1} (r+1) \frac{N!}{r!(N-1-r)!} p^{r+1} (1-p)^{N-1-r} \\
&= \sum_{r=0}^{N-1} r \frac{N!}{r!(N-1-r)!} p^{r+1} (1-p)^{N-1-r} + \sum_{r=0}^{N-1} \frac{N!}{r!(N-1-r)!} p^{r+1} (1-p)^{N-1-r} \\
&= \sum_{r=1}^{N-1} \frac{N!}{(r-1)!(N-1-r)!} p^{r+1} (1-p)^{N-1-r} + Np \\
&= \sum_{r=0}^{N-2} \frac{N!}{r!(N-2-r)!} p^{r+2} (1-p)^{N-2-r} + Np \\
&= N(N-1)p^2 + Np
\end{aligned}
$$

$$\begin{aligned}\Rightarrow \mathrm{Var}(r) &= N(N-1)p^2 + Np - (Np)^2 \\ &= -Np^2 + Np = Np(1-p)\end{aligned}$$

and so standard deviation $\sqrt{Np(1-p)}$, largest for $p = 0.5$.

7.2 Poisson Distribution

Derived from the Binomial Distribution in the limit of $p \to 0$, $N \to \infty$ but $Np \to \lambda$ which is finite, and the mean. This is used to describe the number of instantaneous events happening at a uniform average rate of occurrence within a specific time-frame, e.g. lightning strikes within one hour. We have:

$$\begin{aligned}P(r) &= \lim_{\substack{p\to 0 \\ N\to\infty \\ Np\to\lambda}} \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \\ &= \lim_{N\to\infty} \frac{N!}{r!(N-r)!} \left(\frac{\lambda}{N}\right)^r \left(1 - \frac{\lambda}{N}\right)^{N-r} \\ &= \lim_{N\to\infty} \frac{N^r}{r!} \frac{\lambda^r}{N^r} \left(1 - \frac{\lambda}{N}\right)^N \\ &= \frac{\lambda^r}{r!} e^{-\lambda}\end{aligned}$$

Normalised:

$$\sum_{r=0}^{\infty} P(r) = e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = e^{-\lambda} e^{\lambda} = 1$$

Mean of λ:

$$\begin{aligned}\langle r \rangle &= \sum_{r=0}^{\infty} rP(r) = e^{-\lambda} \sum_{r=0}^{\infty} \frac{r\lambda^r}{r!} = e^{-\lambda} \sum_{r=1}^{\infty} \frac{r\lambda^r}{r!} = e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!} \\ &= e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^{r+1}}{r!} = e^{-\lambda} \lambda e^{\lambda} = \lambda\end{aligned}$$

Variance of λ as well:

$$\begin{aligned}\langle r^2 \rangle &= \sum_{r=0}^{\infty} r^2 P(r) = e^{-\lambda} \sum_{r=0}^{\infty} \frac{r^2\lambda^r}{r!} = e^{-\lambda} \sum_{r=1}^{\infty} \frac{r^2\lambda^r}{r!} = e^{-\lambda} \sum_{r=1}^{\infty} \frac{r\lambda^r}{(r-1)!} \\ &= e^{-\lambda} \sum_{r=0}^{\infty} \frac{(r+1)\lambda^{r+1}}{r!} = e^{-\lambda} \left(\sum_{r=0}^{\infty} \frac{r\lambda^{r+1}}{r!} + \sum_{r=0}^{\infty} \frac{\lambda^{r+1}}{r!} \right) \\ &= e^{-\lambda} \left(\lambda^2 e^{\lambda} + \lambda e^{\lambda} \right) = \lambda^2 + \lambda \\ \Rightarrow \mathrm{Var}(r) &= \lambda^2 + \lambda - \lambda^2 = \lambda\end{aligned}$$

and so standard deviation $\sqrt{\lambda}$.

7.3 Gaussian Distribution

Can be derived from both the Poisson distribution (small deviations from a large λ, i.e. $\lambda, r \gg 0$) and the Binomial distribution ($N \to \infty$). From Poisson, using Stirling's approximation and setting $x = r - \lambda$:

$$\begin{aligned} P(r) &= e^{-\lambda}\lambda^r e^r r^{-r}\frac{1}{\sqrt{2\pi r}} = \frac{1}{\sqrt{2\pi(\lambda+x)}}e^{-\lambda}e^{\lambda+x}\left(\frac{\lambda+x}{\lambda}\right)^{-(\lambda+x)} \\ &\approx \frac{1}{\sqrt{2\pi\lambda}}e^x\left(1+\frac{x}{\lambda}\right)^{-(\lambda+x)} = \frac{1}{\sqrt{2\pi\lambda}}e^x e^{-(\lambda+x)\ln(1+\frac{x}{\lambda})} \\ &\approx \frac{1}{\sqrt{2\pi\lambda}}e^x e^{-(\lambda+x)(\frac{x}{\lambda}-\frac{x^2}{2\lambda^2})} \approx \frac{1}{\sqrt{2\pi\lambda}}e^x e^{-x-\frac{x^2}{2\lambda}} \\ &= \frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{x^2}{2\lambda}} \end{aligned}$$

The Gaussian Distribution is so important because any quantity depending on the sum of a large number of independent variables (for instance, the mean of a set of data drawn from some distribution – the distribution of these individual data need not be Gaussian) tends to a Gaussian – this is the *Central Limit Theorem*.

7.4 Probabilistic Noise

These probability distributions are applied to deduce how noise works in some circumstances.

Shot noise is due to the discrete nature of currents, light etc. A Poisson-based analysis of current that is not really given, gives:

$$\overline{\Delta I^2} = 2I_{\text{avg}}e\Delta\nu$$

where $\Delta\nu$ is the bandwidth of the current, I guess.

Johnson noise is due to systems thermally oscillating. In a simple circuit that they drew, we have that the mean power dissipated due to this noise $\bar{P}$ is given by:

$$\bar{P} = 4kT\Delta\nu$$

8 Inference

8.1 Bayesian Statistics

Bayes' Theorem can be used to suggest the probability that a model $\boldsymbol{a}$ (often a vector of parameters, like the gradient and intercept) is true given a set of

data:

$$P(model|data) = P(data|model)\frac{P(model)}{P(data)}$$

$P(data|model)$ is often calculable from the probability distributions above; $P(data)$ is just a normalisation, in the sense that it does not depend on what model we are testing. $P(model)$ is known as the *prior* and represents the initial likelihood of the model before any data were taken. Importantly, if the prior is very small, then some very good-fitting data would be required to give a *posterior* that is high. For instance, although CERN data once suggested that the speed of light had been exceeded ($P(data|model)$ was quite high), the fact that we previously thought such an occurrence massively unlikely (a "strong" prior) means that few people thought that this law had really been violated.

When one has no idea what scale a should be, a good initial prior is one that is uniform in log-space: $P(a) \propto 1/a\ da$

We might assume that a is equally likely in some range. In this case, we in fact have that $P(model|data) \propto P(data|model)$, and we can just calculate the best model based on the probability distribution...

8.2 Goodness of fit

We suppose that x values are measured with infinite precision, and that data points y_i are from a Gaussian distribution, with std σ_i. The $P(data|model)$ (also known as the *likelihood* L) for a model $y = f(x|\boldsymbol{a})$ is thus given by:

$$L(\boldsymbol{a}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - f(x_i|\boldsymbol{a}))^2}{2\sigma_i^2}\right)$$

This is to be maximised to find the maximum probability. We therefore set $\nabla L(\boldsymbol{a}) = \boldsymbol{0}$ and find $\boldsymbol{a}$. It is actually way easier for a Gaussian error thing to find the maxima of $\ln L$, which is:

$$\ln L = -\frac{1}{2}\sum_i \left(\frac{y_i - f(x_i|\boldsymbol{a})}{\sigma_i}\right)^2 - \frac{1}{2}\sum_i \ln(2\pi\sigma_i^2)$$

Thus to maximise the likelihood, one should minimise the quantity:

$$\chi^2 = \sum_i \left(\frac{y_i - f(x_i|\boldsymbol{a})}{\sigma_i}\right)^2$$

This can be done, for instance, for a straight line fit, where $f(x_i|\boldsymbol{a}) = a_1 x_i + a_2$. The formulae are apparently in the formula book, but some highlights

are here for the case that $\sigma_i = \sigma \ \forall i$:

$$\widehat{a_1} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$
$$\widehat{a_2} = \bar{y} - \widehat{a_1}\bar{x}$$

we see that the best fit line goes through $(\bar{x}, \bar{y})$.

If the model fits the data quite well, each point will be about σ_i away from the model value, suggesting that $\chi^2 = N$. Critical values of the χ^2 distribution are tabulated, and depend on the number of *degrees of freedom* (the dimensionality of $\boldsymbol{a}$).

Considering how this works for data around a horizontal line leads to the idea of a *weighted mean* of a sequence of numbers with different uncertainties, where rather than weighting each value equally, the values are weighted by the inverse square of their individual uncertainties:

$$\bar{y} = \frac{\sum_i y_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$$

The corresponding uncertainty in the mean is given by the root-harmonic-mean-square of the individual uncertainties:

$$\sigma_{\bar{y}}^2 = \frac{1}{\sum_i 1/\sigma_i^2}$$