Physics A: Experimental Methods Summary Notes

Shikang Ni

1 Chapter 1: Measuring and Amplifying

1.1 Measurement

Main aims: don't want transducer to affect measurement, tracking signal, dealing with noise.

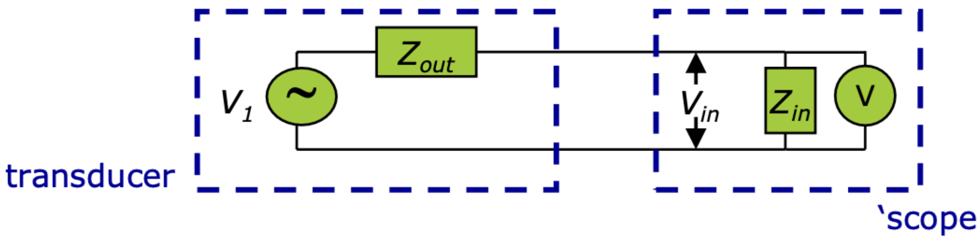


Figure 1: Thevenin's theorem states that any linear circuit can be replaced by just one single voltage in series with a single resistance

Since current in Z_{out} = current in Z_{in},

$$V_{in} = \frac{Z_{in}}{Z_{in} + Z_{out}} V_1$$

So, for voltage measurement:

1. A transducer with low Z_{out}
2. A measurement device with high Z_{in} $(1 - 10M\,\Omega)$. However, there is little power going into the oscilloscope, resulting in a weak signal. This requires the use of amplifiers.

Scope probes - compensation at high frequency

In real scopes:

$$Z_{in} = \left(\frac{1}{R_{in}} + i\omega C_{in}\right)^{-1}$$

$$V_{out} = V_{in} \frac{Z_{in}}{Z_{total}}$$

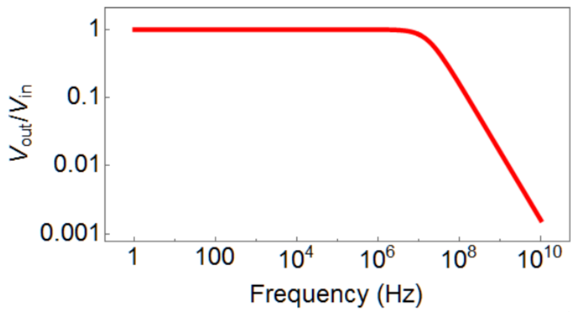
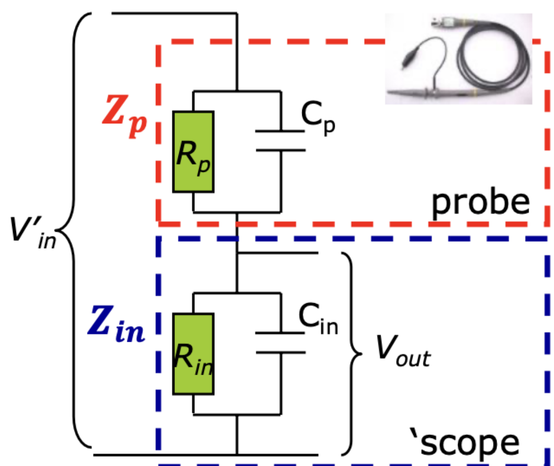


Figure 2: V_{out} is not accurate at high frequency



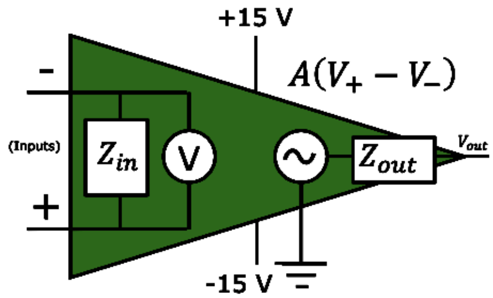
If $C_{in} R_{in} = C_p R_p$,

$$V_{out} = \frac{R_{in}}{R_{in} + R_p} V_{in}$$

C_{in} is compensated, though V_{out} is still not the same as V_{in}.

On the other hand, for current measurement: want low Z_{in} so the measurement device to take all the current.

1.2 Operational Amplifiers*



Open loop gain, A, is the gain without feedback:

$$V_A = A(V_+ - V_-)$$

An ideal amplifier:

1. $A = \infty$
2. $Z_{in} = \infty$
3. $Z_{out} = 0$

A non-ideal amplifier:

1. $|A| \approx 10^4 - 10^6$, is not real and a function of frequency
2. Z_{in} is high
3. Z_{out} is non-zero
4. Finite slew rate, $\frac{dV_{out}}{dt}$

Negative feedback

Letting the op-amp sense a fraction of the output at the -ve input. As the output rises, the difference between the +ve and -ve inputs i.e. $V_+ - V_-$ decreases so V_{out} decreases. An equilibrium is reached that stops the output from saturating.

Solving Op-Amp problems (Nodal analysis)

This equilibrium state can be encapsulated by the 2 golden rules:

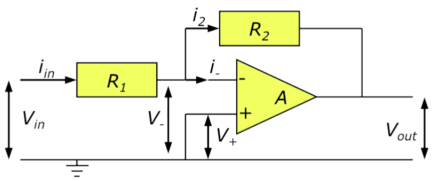
1. Inputs draw no current:
$$I_- = I_+ = 0$$
2. The voltages on the +ve and -ve inputs are the same, provided that V_{out} is not saturated, $V_{out} \neq \pm 15$:
$$V_+ = V_-$$

Closed loop gain, G, is the gain of the circuit with feedback:

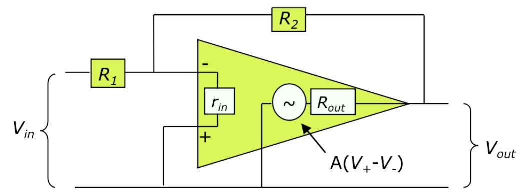
$$G = \frac{V_{out}}{V_{in}}$$

Use the **Golden Rules** and **Ohm's law** (for an ideal co-amp).

Inverting amplifiers



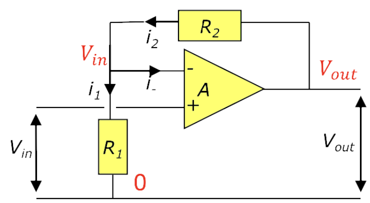
(a) Inverting ideal amplifier



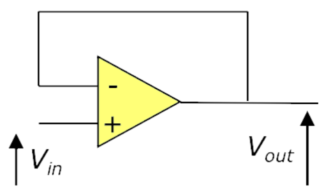
(b) Inverting non-ideal amplifier

$$\boxed{G = -\frac{R_2}{R_1}}$$

Non-inverting amplifiers



(a) Non-inverting ideal amplifier



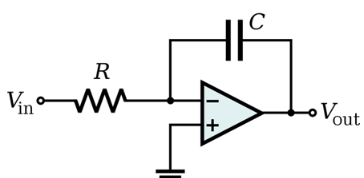
(b) A buffer circuit

$$\boxed{G = 1 + \frac{R_2}{R_1}}$$

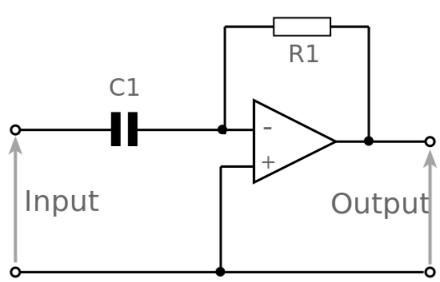
$$V_{out} = V_{in} \text{ with } \frac{R_2}{R_1} = 0$$

Integrators, differentiators, filters

By replacing R_2 with a combination of resistors, capacitors and inductors, we can make all kinds of other useful circuit.



(a) An integrator



(b) A differentiator

$$V_{out} = -\frac{1}{RC}\int_0^t V_{in}dt$$

$$V_{out} = RC\,\frac{dV_{in}}{dt}$$

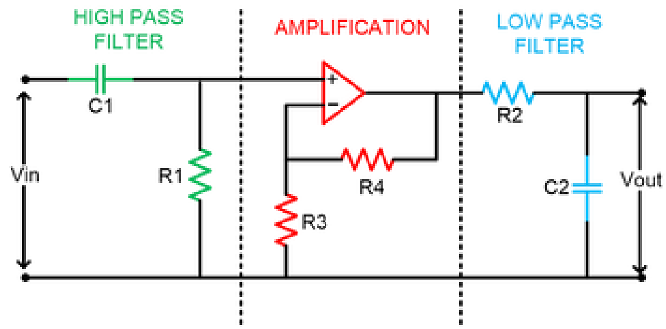


Figure 6: A band-pass filter

High-pass: take voltage across R. Low pass: take voltage across C.
Time constant of RC circuit: $\tau = RC$

Note: Bode plot is a log-log plot.

Derivation for integrator [2019 Tripos Paper 1]
The capacitor voltage, V_C is given by $V_C = \frac{Q}{C}$. Applying the golden rule for op-amp, we have $V_+ = V_- \Rightarrow V_x = 0$. By considering potential difference across C, we obtain:

$$V_C = V_x - V_{out} = -V_{out}$$

Differentiating with respect to time,

$$-\frac{dV_{out}}{dt} = \frac{1}{C}\frac{dQ}{dt}$$

Recognising that $\frac{dQ}{dt}$ is equals to the electric current through capacitor, I_f and using another golden rule for op-amp, where $I_+ = I_- = 0$, we get:

$$I_{in} = I_f \Rightarrow \frac{V_{in}}{R_{in}} = \frac{dV_{out}C}{dt}$$

$$\therefore \frac{V_{in}}{V_{out}} \times \frac{dt}{R_{in}C} = 1$$

$$V_{out} = -\frac{1}{R_{in}C}\int_0^t V_{in} dt$$

.
Possible problem: V_{out} diverges for D.C. input voltage, resulting in saturation.
Solution: Add a large resistor parallel to the capacitor, known as the shunt resistor, to allow the capacitor to discharge slowly.

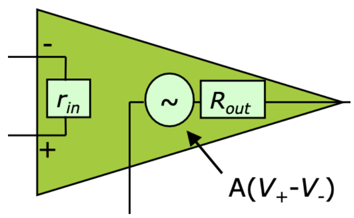
Non-ideal amplifier

Input impedance not infinite, output impedance non-zero and A is not infinite and not real (since A is not real, a phase lag will occur).

$$\boxed{V_{out} = V_{in} \cdot A - I_{out} \cdot Z_{out}}$$

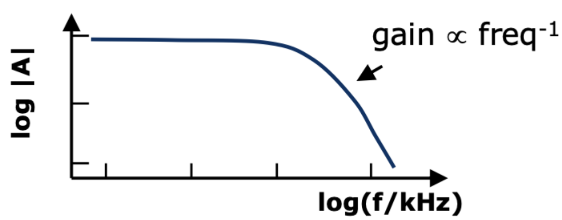
Key idea: a non-ideal amplifier approximates to an ideal one provided A is high.

When solving non-ideal amplifiers, don't analyze with the Golden Rules, but just conserve current and use Ohm's law, by nodal analysis.



Approach: find I_{in}, I_{out} and the relationship between them.

Frequency dependence of gain



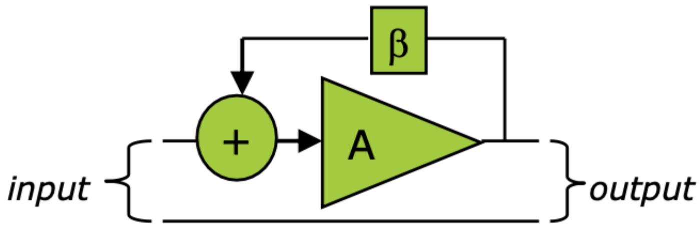
Higher frequency leads to higher and higher phase shift. Amplifier drops in gain at high frequency.

The feedback resistor usually has complex impedance. So, over a big enough frequency range, this can lead to a **phase shift** of π. This will change –ve feedback into +ve feedback, resulting in saturation/oscillation rather than amplification.

Consequence of A being not real: at high frequency, negative feedback turn into unstable positive feedback.

So, real op-amps are designed such that $|A| \to 0$ as $\omega \to \infty$.

Feedbacks in general



$$\text{output} = A(\text{input} + \beta \times \text{output})$$

where β is the fraction fed back, and it is negative for -ve feedback.

Closed loop gain is:

$$G = \frac{\text{output}}{\text{input}} = \frac{A}{1 - A\beta}$$

For large A,

$$G \approx -\frac{1}{\beta}$$

it is finite and independent of A.
Hence, provided $|A\beta| \gg 1$, negative feedback removes all problems of instability of A in real experiments.

Positive feedback - Oscillator

For $\beta > 0$, output will be reinforced by the feedback and the output can become unstable or saturate.

This is useful if $\beta > 0$ and $A\beta = 1$ for only one frequency, resulting in a system that oscillate at that frequency.

2 Chapter 2: Errors

An estimate of error in the experimental value allow us to know if the experimental value is significantly different from the theoretical prediction.

Random error

Definition: the error is equally likely to be +ve or −ve such that the average error is zero. Hence, can be reduced by techniques of repeated measurements and averaging.

Systematic error

Definition: all the errors that are not random. It is approximately constant in time or it might drift. As a result, no combination of measurement can reduce it. Can only be avoided with good experimental design.

Example: miscalibrated oscilloscope time-base, background signal

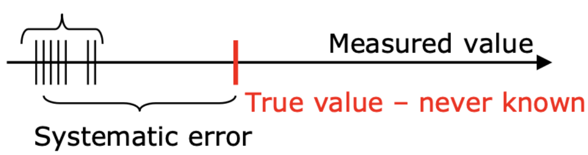
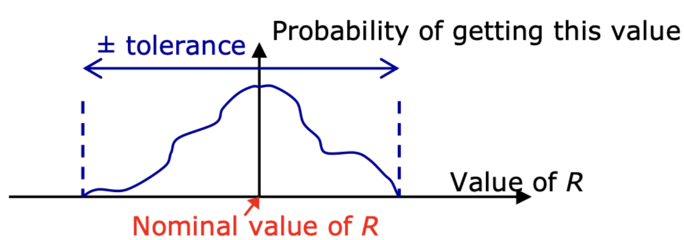


Figure 7: Diagram showing the effect of both random and systematic error

2.1 Defining random error

2.1.1 Tolerance

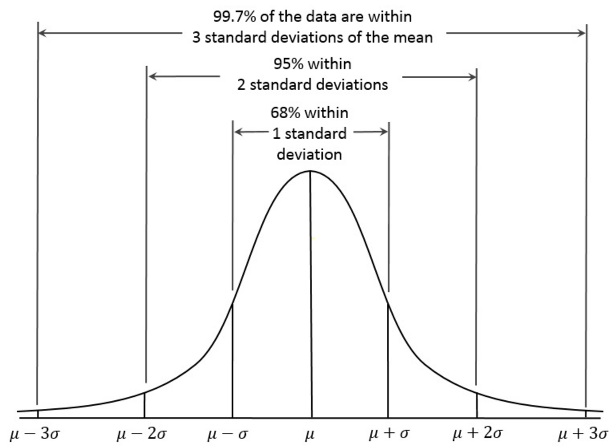
Definition: the full range within which a value may lie. One is 100% confident that R lies within the range nominal ± tolerance.



Example: resistor color code

2.1.2 Standard deviation

Definition: a measure of spread of a set of values (assumed Gaussian). Can be sample standard deviation or unbiased estimate of population standard deviation. σ^2 is calculated by the mean squared deviation of each data points from their average.



If we know the population standard deviation, we can derive a confidence interval.

2.2 Statistics

$$\text{Mean: } \overline{x} = \frac{\sum_i x_i}{N}$$

** Mean of the measurements is an unbiased estimator of the true value of the quantity.*

$$\text{Standard deviation: } \sigma = \sqrt{\frac{1}{N}\sum_i (x_i - \overline{x})^2}$$

$$\text{Error in single datum: } \Delta x = \sqrt{\frac{1}{N-1}\sum_i (x_i - \overline{x})^2} = \sigma\sqrt{\frac{N}{N-1}}$$

** There is a factor of N-1, since we have had to use the data themselves to estimate what the mean value is.*

$$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y) \text{ for X and Y are independent}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$\text{In general: } Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y) + 2ab\,Cov(X, Y)$$

$$\text{Uncertainty in mean: } \Delta x_{mean} = \boxed{\frac{\Delta x}{\sqrt{N}}} = \frac{\sigma}{\sqrt{N-1}}$$

Example: Measuring a small patch of the 2.7 K cosmic microwave background radiation for 100 times longer should decrease the measurement error by a factor of 10.

Error propagation, assuming Gaussian errors:

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

If f is complicated, $\frac{\partial f}{\partial x}$ can be evaluated empirically :

$$\frac{\partial f}{\partial x} = \frac{f(\overline{x} + \sigma_x, \overline{y}, \overline{z}, ...) - f(\overline{x}, \overline{y}, \overline{z}, ...)}{\sigma_x}$$

2.3 Removal of systematic errors

Strategies:

1. **Calibration of experiment and instruments.**
   - Use different instruments to check for unknown systematic effects.
   - Measure a known value to check if it is accurate.
   - Check against more accurate standard equipment.
   - Check for zero offsets.
2. **Exploit symmetry.**
   For example, reverse the voltage source to your measurement device and the magnitude of V should be unchanged.
3. **Null method measurement**, where a quantity X being measured is opposed by an adjustable quantity until an indicating device shows balance, hence allowing X to be measured without systematic error.

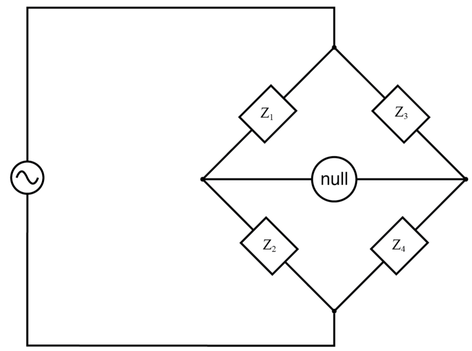


Figure 8: A bridge circuit for null method

Balance condition:

$$\frac{Z_{unknown}}{Z_2} = \frac{Z_{adjustable}}{Z_4}$$

where $Z_2, Z_{adjustable}$ and Z_4 need to be accurately known.

Benefit: the indicating device i.e. null device does not need to be linear or well calibrated

For example, $Z_{unknown}$ can be an unknown capacitor and $Z_{adjustable}$ is an adjustable capacitor.

4. **Quantity that change with time.** Minimise its effect by choosing appropriate sequence of measurements. Do not measure in order AAABBBCCC but rather ABCCBAABC for instance.
   For example, warming of apparatus.

5. **Differential measurement.** We are measuring a difference and not absolute value. Any constant error in one direction that appear in both quantities will cancel out.

   For example,

   - Thermocouple: compare output in unknown temperature of around 100 $°C$ with output in boiling water. This measures the difference in the quantity with the system in the desired state and then in a standard and similar state.
   - Micrometer: the difference in location of the spindle relative to the anvil measures the width of the sample. Relies on the fact that the frame is rigid on short timescales.

     Avoiding backlash: you get a different systematic error as you move the spindle to left or right. So, always approach from the same side.

6. **Selection effects.** Making sure we measure what we want to measure, preventing erroneous conclusions.

   For example, distant radio sources appear to be more powerful than nearby ones. This is a result of selection bias as weak distant sources are unlikely to be detected in the first place.

3 Chapter 3: Sampling theory

3.1 Why sample digitally?

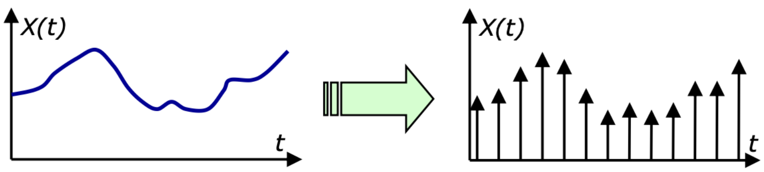


Figure 9: Picture to draw

Data is captured by sampling digitally. This means we are usually converting a continuous analogue signal into bits of digital signal. We do it because:

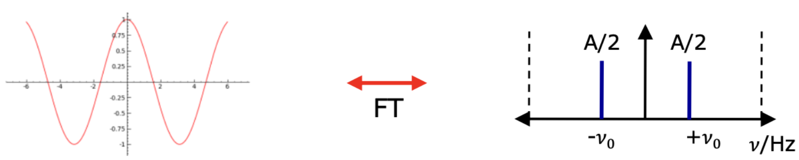
- Maintaining precision for analogue signal is difficult.
- We may not have access to the system all the time.
- Sampling reduces the data volume, making it more manageable.
- Sampled data may be easier to transmit without information loss.

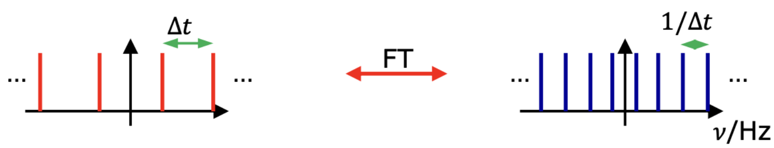
The parameters for sampling are:

1. **Sampling rate**: Nyquist's criterion - minimum $2f$ [How fast?]
2. **Sampling duration**: 1/bandwidth [How long?]
3. **Sampling accuracy**: quantisation [How accurate?]

3.2 Fourier transform

Looking at the frequency content of the system.





$$g(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

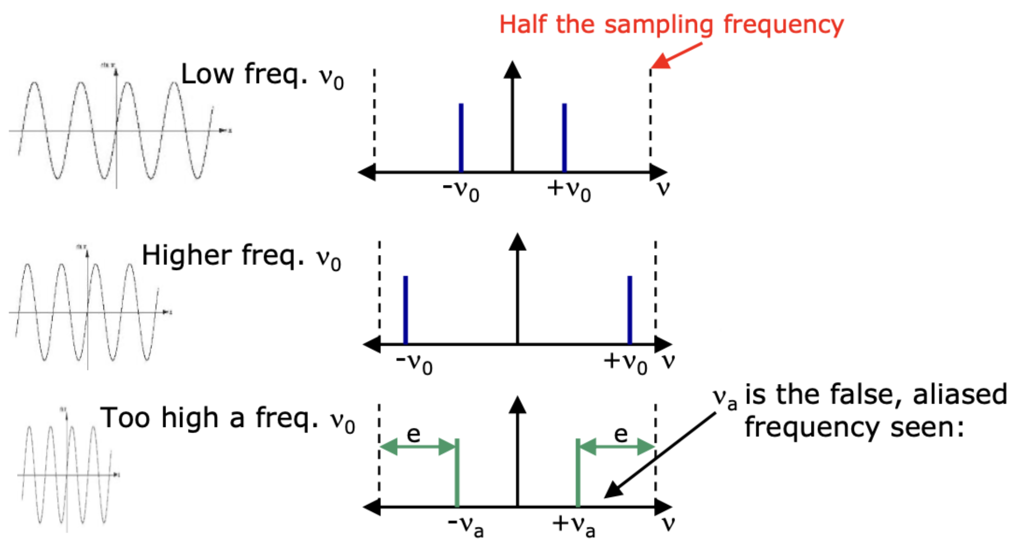
3.3 Nyquist's criterion

> For a band-limited function, sampling frequency needs to be minimally twice of the highest frequency component present in the signal.
> If the sampling is noiseless, you can recover the signal perfectly from its samples.

When you under-sampling and don't obey Nyquist's criterion - Aliasing

$$\text{True freq} = \frac{\text{Sampling freq}}{2} + e$$

where e is the frequency gap between $\frac{1}{2}$ sampling frequency and the aliased frequency.



Fourier transform representation

Sampled signal is the product of the true signal and the comb of impulse δ-functions.

$$s(t) = x(t) \times c(t)$$

$$\widetilde{s(\nu)} = \widetilde{x(\nu)} \ * \ \delta\text{-comb with spacing } \Delta\nu = 1/T$$

In order for the convolution images to not overlap, requires $\frac{1}{T} >$ width of spectrum, Δf

Recovering the original signal from the sampled signal

First,

$$\widetilde{x(\nu)} = \widetilde{s(\nu)} \times \text{top-hat with width of 1/T}$$

then,

$$x(t) = FT^{-1}\left[\widetilde{x(\nu)}\right] = s(t) * \text{sinc}\left(\frac{\pi t}{T}\right)$$

This only works if the convolution images do not overlap if the sampling frequency, $f_{sampling} = \frac{1}{T}$ satisfies:

$$f_{sampling} \geq 2f_{\max}$$

3.4 Sub-Nyquist Sampling

For the case where $f_{min} > 0$ with a known bandwidth B. It is still OK with:

$$f_{sampling} \geq 2B$$

If there is overlap:

1. Sample faster.
2. Shift the signal to lower frequency, while preserving information. This will not change B, but it may be easier to sample faster at these lower signal frequencies.

3.5 Quantization affects accuracy

'N' bit sampling: using 2^N quantizing bins.

Benefits of oversampling:

- **Oversample to reduce quantization error**.
  E.g. sampling at $4f_{Nyquist}$ (average 4 successive samples of the signal) improves the resolution by a factor of 2.
- **Oversample to reduce noise**.
  Sample at $N \cdot f_{nyquist}$ (average N samples) improves the Signal-to-Noise by a factor of $\sqrt{N}$.

3.6 Sampling duration

The lower the frequency we are interested in the longer we have to sample the signal. Spectral resolution:

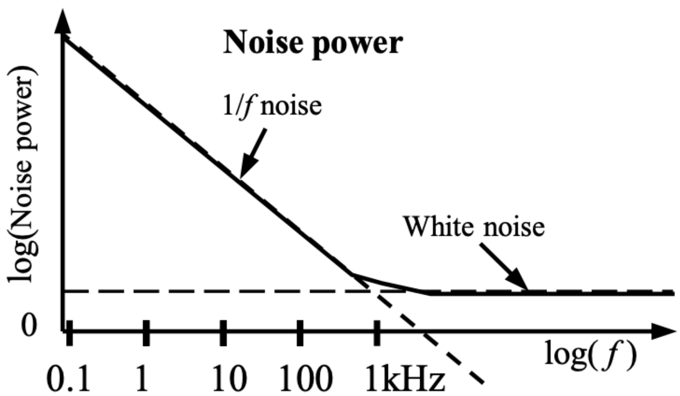
$$\Delta f = \frac{1}{T_{max}}$$

Key ideas:

- If you want to be sensitive to long-period variations you have to wait a long time to see them.
- High spectral resolution comes from sampling longer, not sampling faster.

4 Chapter 4: Coping with noise

Frequency dependence of noise



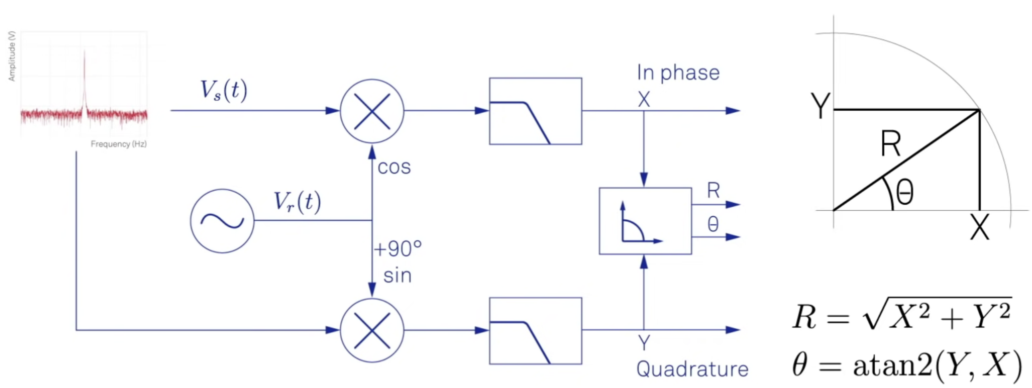
1. At low frequency, 1/f noise
2. At high frequency, white noise *e.g. shot noise, Johnson noise*

Conclusion: effect of noise is worst at low frequency or DC.

4.1 Filtering & Phase-sensitive detection

Filtering: the separation in temporal- or spatial- frequency space of the signal of interest from noise. Most effective when the signal and noise have non-overlapping spectra.

Lock-in/phase-sensitive detection: used to detect weak periodic signals buried in noise, by encoding the DC signal with a carrier and subsequently demodulating it with a mixer (multiplier/modulator) and low-pass filter.



$$R = \sqrt{X^2 + Y^2}$$
$$\theta = \text{atan2}(Y, X)$$

Input 1: Your noisy signal buried in noise, modulated and carried at ω_r

$$V_{input} = V_s sin(\omega_r t + \phi) + \text{noise}$$

Input 2: A reference signal which has the same frequency ω_r or its harmonic and a fixed stable phase relationship to the signal.

Operations: Multiply noisy signal and reference signal together, then time averaging or low pass filtering.

Output: A resultant signal encapsulates the amplitude of the signal V_s and the phase difference ϕ between signal and reference.

$$\begin{aligned} \langle V_{out} \rangle &= \frac{G}{T} \left[\int_0^{\pi/\omega_r} V_s sin(\omega_r t + \phi) dt + \int_{\pi/\omega_r}^{2\pi/\omega_r} -V_s sin(\omega_r t + \phi) dt \right] \\ &= \frac{2G}{\pi} \boxed{V_s cos(\phi)} \end{aligned}$$

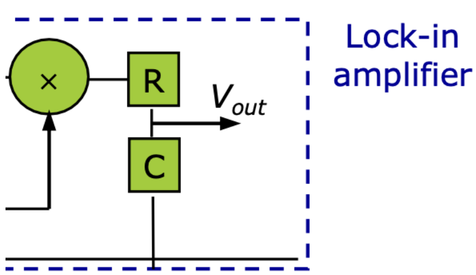
Or

$$\begin{aligned} \langle V_{out} \rangle &= V_s \; sin(\omega_r t + \phi) \times V_{ref} \; sin(\omega_r t) \\ &= \frac{1}{2} V_s V_{ref} [cos(\phi) - cos(2\omega_r + \phi)] \\ &\xrightarrow{low-pass} k \; \boxed{V_s cos(\phi)} \end{aligned}$$

**Reference signal is a square wave in this case for simplicity. Can be sinusoidal too.*

Key points:

- The portion of the signal that has the same frequency as the reference is effectively "squared" and thus has a time average value that can be all positive, negative or zero depending on ϕ.
- Any unwanted frequency, in which $\Delta\omega \neq 0$ has a time average value of zero, provided $\Delta\omega >> \frac{1}{RC}$. So, if RC is big enough, even noise close in frequency to ω_r will be removed.

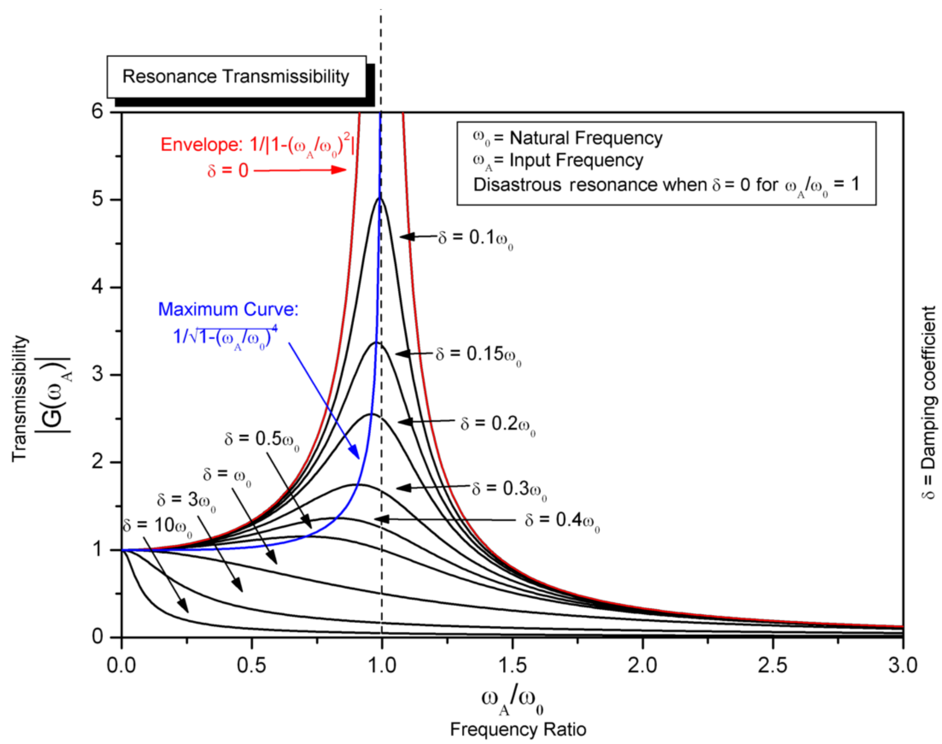


- If the averaging times are long enough, the frequency width can get very narrow, thereby reducing the noise contribution to the measurement. However, if we average too long may become insensitive to changes in the signal we are tying to detect.

- The signal is often encoded at a high frequency where noise disturbances are small. 1/f noise is often the limiting factor.

4.2 Vibration isolation

Aim: limit the transmission of mechanical vibrations into experimental apparatus by decoupling the experiment from the movement of the surrounding.



1. Make the natural frequency, ω_0 of the decoupling system low. For example, low k value or air cushion.

$$\omega_0^2 = \frac{\gamma_{air} g}{z_t^{eq}} = \frac{k}{m}$$

2. Reduce the peak response at ω_0 by damping i.e. high δ

3. Make resonant frequency of experiment $>> \omega_0$

4.3 Thermal shielding

Aim: limit the transport of unwanted thermal energy into experimental apparatus

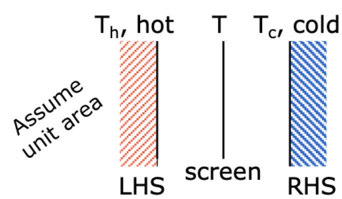
1. Reduce evaporation [lid], conduction [insulation/vacuum], and convection [vacuum]

2. Reduce radiation

   (a) Smooth and shiny surface has low ϵ. For example, matt black has $\epsilon \approx 0.95$ whereas polished Al foil has $\epsilon \approx 0.03$.

   (b) Insert n floating shields between the two surfaces. Power per unit area radiated by a body is given by Stefan Boltzmann law, $P = \sigma\epsilon T^4$ where ϵ is the emissivity (1 - reflectivity) and $\epsilon = 1$ for a black-body.

   With a single floating screen:



$$\text{net flow on LHS } = \sigma\epsilon T_h^4 - \sigma\epsilon T^4$$
$$\text{net flow on RHS } = \sigma\epsilon T^4 - \sigma\epsilon T_c^4$$

In thermal equilibrium, heat flow on LHS = heat flow on RHS:

$$\Rightarrow T_h^4 + T_c^4 = 2T^4$$

$$\text{So, overall flux to the right} = \sigma\epsilon\frac{1}{2}(T_h^4 - T_c^4)$$

Generalising to n number of heat shields:

$$\text{So, net radiation flow} = \frac{\sigma\epsilon}{n+1}(T_h^4 - T_c^4)$$

As n increases, net radiation flow decrease.

4.4 Differential measurements

Tricks:

1. Bridge: compare resistances
2. Strain gauge pair: one active and the other calibrate for the environment
3. Twisted wire pair: E & B induce the same current in each lead as they follow the same path through space
4. Differential amplifiers: ignore common mode signal

$$V_{out} = \frac{R_2}{R_1}(V_2 - V_1)$$

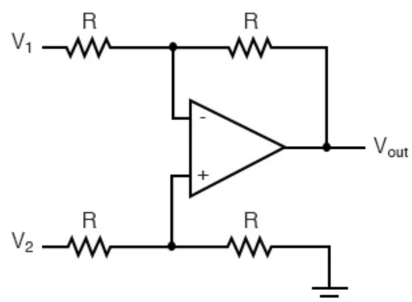
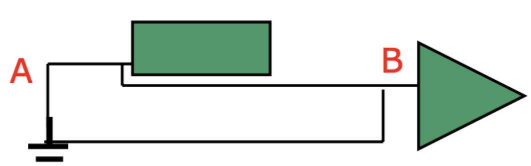


Figure 10: A differential amplifier circuit

4.5 Shielding of E + B fields

Aim: limit the transmission of unwanted electromagnetic disturbances into experimental apparatus

1. E field: Faraday cage. Conductor has zero internal field.
2. B field: high-permeability metal shield. Provides a low reluctance path for the B field lines.
3. Avoid earth loops, where the circuit is connected to earth ground at two or more points. The potential of the earth varies from point to point, two or more connections to the ground cause currents to flow. If the current flows through a signal carrying wire, the result is a noisy, offset signal.



5 Chapter 5: 3 Probability distributions

What do our measurements tell us about the universe?

1. Principle of indifference: assign equal probabilities to events unless we have information distinguishing them.

2. Frequentist approach: assign probabilities on the basis of the relative frequencies of events in the limit of a large number of trials

3. Bayesian approach: examine what information we have and use that

5.1 Binomial

Definition: identical trials with only two possible outcomes.
Examples: 1d random walk, Ising model of a system of interacting spins.

n independent trials, with probability p of success and r number of successes:

$$p(r|p,n) = \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

$$X \sim B(np, np(1-p))$$

Key points:

- Peak around the mean = np
- As N increases, the peak width slowly narrows with respect to the range of N.
- The peak width depends on p, widest when p=0.5.
- When $p \neq 0.5$, skew shaped.

5.2 Poisson

Definition: for when there is a constant mean rate of occurrence λ over space or time, and each occurrences are independent of the previous.

Examples: radioactivity, shot-noise, mutation-rates, internet hits, detection of neutrino counts from a distant supernova

From Binomial distribution, let $n \to \infty$ and $p \to 0$ while $\lambda = np$ remains finite, we get Poisson distribution:

$$p(r|\lambda) = \frac{\lambda^r}{r!} \frac{1}{e^\lambda}$$

$$X \sim Po(\lambda, \lambda)$$

Key points: $\boxed{\text{variance} = \text{mean}}$

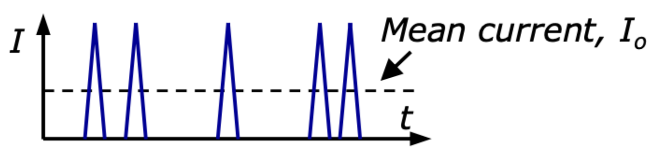
Comparison with Binomial:

1. Broader
2. Characteristic long upper tail
3. Number of events can greatly exceed λ but for Binomial $r \leq n$

5.2.1 Shot noise (Poisson noise)

Key idea: Fluctuation in current due to granularity of discrete charges/photons arriving at the detector.

Important trait: has a white temporal spectrum i.e. spread equally over all temporal frequencies.



Model:
Current pulse at some time t_n within time interval T:

$$I_n = e\delta(t - t_n) \xrightarrow{\text{Fourier series}} \frac{e}{T} + \frac{2e}{T} \sum_m cos \frac{2\pi m(t - t_n)}{T}$$

For a given normal mode m:

$$\Delta \langle I_n^2 \rangle_m = \left(\frac{2e}{T}\right)^2 \langle cos^2 \left(\frac{2\pi m(t - t_n)}{T}\right) \rangle = \frac{2e^2}{T^2}$$

Summing over all modes m and all electrons N arriving in time T:

$$\Delta \langle I^2 \rangle = \frac{2e^2 Nm}{T^2} = 2e \frac{Ne}{T} \frac{m}{T} = 2eI_{avg} \Delta f$$

$$\boxed{I_{rms} = \sqrt{2eI\Delta f} = \sqrt{\frac{2eI}{\Delta t}}}$$

where Δf is the frequency bandwidth or $\Delta f = \frac{1}{\Delta t}$. Valid for AC only (factor of 2).

Example: photon noise in light-capturing devices that result in grainy images.

Reducing shot noise:

- Take more measurements (increase N) as mean signal $N\lambda$ grows faster than noise $\sqrt{N\lambda}$ by a factor of $\sqrt{N}$.

$$\boxed{SNR = \sqrt{N}}$$

- Limit the detection bandwidth. For example, encoding the quantity you wish to measure with a monochromatic carrier using phase sensitive methods.

> Example: Uncertainty in current (alternate proof for derivation in exam)
> Number of electrons follows Poisson distribution.
>
> $$N = \frac{It}{e}$$
>
> Rearranging,
>
> $$\Delta I = \frac{e}{t}\Delta N = \frac{e}{t}\sqrt{\frac{It}{e}} = \sqrt{\frac{Ie}{t}}$$
>
> The actual ΔI has an extra factor of $\sqrt{2}$.

5.3 Gaussian

Example: Johnson noise, many experimental errors.

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Let $z = \frac{x-\mu}{\sigma}$, obtain the standard normal:

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$X \sim N(\mu, \sigma^2)$$

The limit of Poisson and Binomial distribution is Gaussian:
As λ is large, Poisson $\rightarrow$ Gaussian with $X \sim N(\lambda, \lambda)$
As N is large, Binomial $\rightarrow$ Gaussian with $X \sim N(Np, Np(1-p))$

Central Limit Theorem:
A quantity that depends on the sum of many independent variables will tend to have a Gaussian probability distribution. For $N > 30$.

Application: **multiple causes of error**, tend to lead to a measured quantity having a Gaussian error distribution. So the fact that each error may have been drawn from an unknown or even non-Normal distribution is irrelevant, as long as the number of different contributions is large.

5.3.1 Johnson noise

Key idea: A fluctuation in current caused by thermal fluctuations of stationary charge carriers.

Important trait: has a white temporal spectrum

Why it is Gaussian: summation of a large number of random fluctuations.

Approach: determining the amplitude of the noise is equivalent to summing the energy in the normal modes of electrical oscillation along a shorted transmission line connected to two resistors of resistance R.

Model:
By the equipartition law of thermodynamics, every mode of oscillation contributes approximately kT average energy to the oscillation.

Power contribution from frequency band Δf:

$$dP = k_B T \Delta f$$

$$\Delta\langle V^2\rangle = (2R)^2 \Delta\langle I^2\rangle = 4R^2 \frac{dP}{R} = 4Rk_B T\Delta f$$

$$\Delta V_{rms} = \sqrt{4Rk_B T\Delta f}$$

$$\boxed{P_{rms} = 4k_B T\Delta f}$$

$$\boxed{\Delta V_{rms} = \frac{P_{rms}}{I}}$$

Reducing the error:

- Cool the apparatus since P_{rms} reduces as T decreases. The noise fluctuations are independent of the particular signal you are measuring, and so the signal-to-noise ratio scales linearly with the mean signal level.
- Limiting the detection bandwidth. For example, encoding the quantity you wish to measure with a monochromatic carrier using phase sensitive methods.

6 Chapter 6: Inference

How do we estimate parameters from our data?

6.1 Bayesian parameter estimation

Bayes' theorem, in essence, is an expression for conditional probability. When used in hypothesis testing, it allow us to update our belief of a model given new data. The initial belief in our model is the prior probability.

$$p(\underset{\sim}{a}|data) = \frac{p(data|\underset{\sim}{a}) \times p(\underset{\sim}{a})}{p(data)}$$

$p(\underset{\sim}{a})$ is the prior, $p(data|\underset{\sim}{a})$ is the likelihood function and $p(\underset{\sim}{a}|data)$ is the posterior probability for $\underset{\sim}{a}$.

Stage 1: before the experiment

3 cases of prior knowledge:

1. Do not know the order of magnitude of our parameter → log-uniform/reciprocal distribution i.e. $p(a)\ da \propto 1/a\ da$
2. We know its within some range → uniform distribution within that range
3. Some preference for the values of the model parameters, either based on someone else's data or some other information.

Warning: For the third case, if there is a strong prior for some particular parameter values, it means even a likelihood function strongly peaked elsewhere may not be enough to alter that prior probability significantly.

Stage 2: after the experiment

The extent to which certain model parameters are to be preferred will depend on whether the data measured could reasonably have been explained on the basis of the prior parameter values. This is captured by the likelihood function.

For uniform prior in the model parameters, Bayesian parameter estimation is equivalent to the maximum likelihood method.

6.2 Maximum likelihood method

$$\text{Likelihood} = L(y_1 ... y_N \,|\, \underset{\sim}{a}) = \prod_i p\,(y_i \,|\, \underset{\sim}{a})\, dy_i$$

where y_i are the data points and $\underset{\sim}{a}$ is the vector of the model parameters.

Idea: find the value of $\underset{\sim}{a}$ that best maximizes L.

Straight-line fitting

$$\underline{\text{independent}}\ y_i \text{ with } \underline{\text{no systematic error}} = \begin{cases} \underline{\text{Gaussian random error}} :\ \sigma_i \\ \text{independent variable with } \underline{\text{no error}} :\ x_i \\ \text{model} : f(x_i | \underset{\sim}{a}) \end{cases}$$

For the ith measurement:

$$p(y_i | \underset{\sim}{a}) = \frac{1}{\sigma_i \sqrt{2\pi}}\, e^{-\frac{[y_i - f(x_i)]^2}{2\sigma_i^2}}$$

$$\begin{aligned} \therefore\ ln(L) &= -\frac{1}{2} \sum_i \left[\frac{y_i - f(x_i)}{\sigma_i} \right]^2 - \sum ln(\sigma\sqrt{2\pi}) \\ &= -\frac{1}{2}\chi^2 - \sum ln(\sigma\sqrt{2\pi}) \end{aligned}$$

To maximise L, we minimise χ^2.

$$\boxed{\chi^2 = \sum_i \left[\frac{y_i - f(x_i)}{\sigma_i} \right]^2} \tag{1}$$

Case 1: assume σ_i is **the same** for all y_i
By differentiating equation (1) with respect to c and m, we get:

$$\overline{y} = \hat{m}\overline{c} + \hat{c} \tag{2}$$

$$\overline{xy} = \hat{m}\overline{x^2} + \hat{c}\,\overline{x} \tag{3}$$

Ingredients: $\overline{xy}$, $\overline{y}$, $\overline{x}$, $\overline{x^2}$. Rearranging, we get:

$$\boxed{\hat{m} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}} = \frac{cov(x,y)}{var(x)}$$

$$\boxed{\hat{c} = \overline{y} - \hat{m}\overline{x}} = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2}$$

Using $\sigma_m^2 = \sum_i \left(\frac{\partial \hat{m}}{\partial y_i}\right)^2 \hat{\sigma}^2$, where $\hat{\sigma}^2$ is the deviation of data from best-fit model given by $\hat{\sigma}^2 = \frac{1}{N-2}\sum_i [y_i - f(x_i)]^2 = \frac{\sigma_i^2 \chi^2}{N-2}$. We obtain:

$$\boxed{\sigma_m^2 = \frac{\sigma^2}{N \times \text{var}(x)} = \boldsymbol{\frac{\chi^2 \sigma_i^2}{N-2} \cdot \frac{1}{N \cdot \text{var}(x)}}}$$

where $(N-2)$ is like the degrees of freedom.

$$\boxed{\sigma_c^2 = \frac{\hat{\sigma}^2\, \overline{x^2}}{N \times [\overline{x^2} - \overline{x}^2]} = \boldsymbol{\sigma_m^2 \cdot \overline{x^2}}}$$

Case 2: σ_i is **not** the same for all y_i.

1. Weighted by $1/\sigma_i^2$
2. Normalised by $\sum_i 1/\sigma_i^2$ instead of N

$$\boxed{\text{Weighted mean: } \overline{y} = \frac{\sum_i y_i/\sigma_i^2}{\sum_i 1/\sigma_i^2},\ \sigma_y^2 = \frac{1}{\sum_i 1/\sigma_i^2}}$$

More importantly,

$$\hat{\sigma}^2 = \frac{1}{N-2}\sum_i \frac{1}{\sigma_i^2} \times \frac{[y_i - (\hat{m}x_i + \hat{c})]^2}{\sum_i 1/\sigma_i^2}$$

6.3 Hypothesis testing - χ^2 as measure of goodness of fit

Question to ask: what is the significance of mismatch between model and data?
Answer: using χ^2 test as a measure of goodness of fit

Determining fit of model

Step 1: determine degrees of freedom, $df = N_{data} - N_{param}$
Step 2: compute χ^2 using formula:

$$\chi^2 = \sum_i \left[\frac{y_i - f(x_i)}{\sigma_i}\right]^2$$

Step 3: compare df with χ^2 value.

1. if $\boxed{\chi^2 \approx df}$, model f **fits** the data as we $|y_i - f(x_i)| \approx \sigma_i$ for a good model.
2. if $\chi^2 >> df$, model is likely wrong
3. $\chi^2 << df$, very strange, the estimates of σ_i might be too large

Optimising model parameters

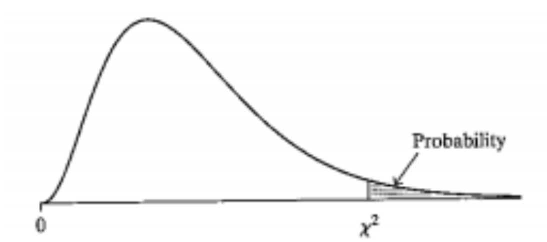
Step 1: find expression for χ^2:

$$\chi^2 = \sum_i \left[\frac{y_i - f(x_i)}{\sigma_i}\right]^2$$

Step 2: minimise χ^2 with respect to the model parameters via differentiation and equating to zero.

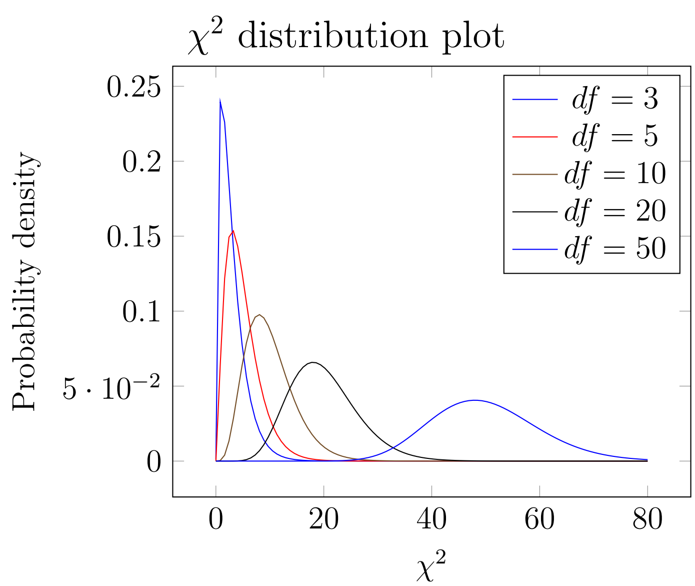
Step 3: solve simultaneous equation.

Using the table of critical p-values:



$$p = P(X \geq \chi^2)$$

The p-value, which is the area under graph in the upper tail, tells you the probability that the actual value of χ^2 is as large as or larger than your value. If $p < 0.05$ i.e. χ^2 lies in the shaded region, the deviations are statistically significant at 5% significance level and the model being tested have to be rejected.



As $N \to \infty, \chi^2 \sim N(df, 2\,df)$

Warning:

1. Make sure the assumptions of straight-line fitting above hold
2. Do not use χ^2 as a means of comparing different models. If two competing theories fit our data comparably well, then we will likely prefer the one with the fewer model parameters, even if its value of chi-squared is slightly higher.

6.4 Non-parametric statistics

If the underlying distribution is not known, need to use non-parametric statistics.

- Runs test: detecting non-randomness
- Sign test: is the distribution of x the same as the distribution of y?
- Mann-Whitney test: do 2 samples come from the same dist.?
- Kolmogorov-Smirnov test: do 2 probability distributions differ?

7 Tripos Q&A

1. **Q: When digitising an analogue signal, explain why adding noise can be beneficial when combined with oversampling in time. (2010 P2 Q5)**

   A: On timescale much shorter than the shortest period present in the signal, we can model the signal as constant. This constant signal is unlikely to exactly match one of the digitization levels, and hence will be mis-measured to the closest level. Adding symmetric noise with standard deviation comparable with the level spacing can remove this error. The noise will trigger adjacent digitization levels and then averaging the signal will recover the true value.

   One must not average over timescales longer than the shortest period present in the signal, as then one would be averaging the signal, hence the signal must be oversampled relative to the Nyquist criterion of twice the maximum frequency present in the signal.

---

2. **Q: Explain how the χ^2 test can be used to compare theoretical models with data, including a discussion of 'degrees of freedom'. (2009 P2 Q13)**

   A: Given data with some associated uncertainty and a theoretical model that predicts what the observed distribution should be, we can use χ^2 to determine goodness of fit. We have to assume that the independent variables, X, has no errors and the measurements, Y, is independent of each other and has Gaussian errors. The χ^2 statistic is:

$$\chi^2 = \sum_i \left(\frac{\text{measured} - \text{predicted}}{\text{std dev}} \right)^2$$

   For a good fit, the value of χ^2 should be the number of data points minus the number of known parameters. This difference, N, is known as the degrees of freedom of the dataset. If χ^2 is significantly higher than N, then the model is certainly wrong. It if it too low, it is indicative that we might have overestimated the degrees of freedom in the model or the errors are too too large.