

Review: Functions of Random Variables

$$Y = y(X) \quad \frac{dY}{dX} = y'(X)$$

conserve probability:
 $d(\text{Prob}) = f(Y) |dY| = f(X) |dX|$

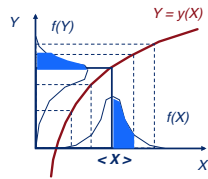
$$f(Y) = f(X) \left| \frac{dX}{dY} \right| = \frac{f(X)}{|y'(X)|}$$

mean value (biased)

$$\langle Y \rangle = y(\langle X \rangle) + \frac{1}{2} y''(\langle X \rangle) \sigma_X^2 + \dots$$

standard deviation (stretched)

$$\sigma_Y = \sigma_X \left| \frac{dy}{dx} \right|_{x=\langle X \rangle} + \dots$$



$$X = 100 \pm 10$$

$$Y = \sqrt{X}$$

$$\langle Y \rangle = ? \quad \sigma_Y = ?$$

The Central Limit Theorem

- (a.k.a. the **Law of Large Numbers**)
- Sum up a large number N of independent random variables X_i .
- The result resembles a Gaussian:

$$\sum_{i=1}^N X_i \rightarrow G(\mu, \sigma^2) \quad \text{as } N \rightarrow \infty$$

- The means and variances accumulate:

$$\mu = \sum_{i=1}^N \langle X_i \rangle \quad \sigma^2 = \sum_{i=1}^N \sigma^2(X_i)$$

- But **higher moments are forgotten**.
- The original distributions $f(X_i)$ don't matter -- **all shape information is lost**.
- This is why **Gaussians are special**.
- This is why measurements often give Gaussian error distributions.
- (Fast computers let us do more exact Monte Carlo analysis.)

Example: Coin Toss

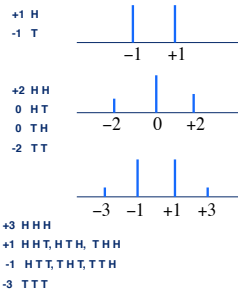
$C = +1$ if heads
 -1 if tails

$$\langle C \rangle = 0 \quad \sigma_C^2 = 1$$

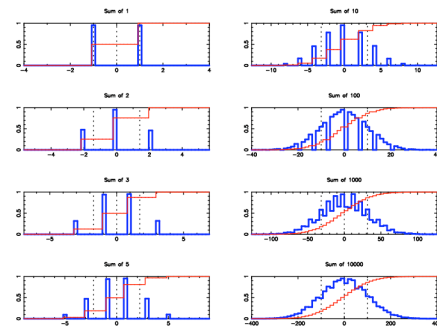
$$S_N = \sum_{i=1}^N C_i$$

$$\langle S_N \rangle = \sum_{i=1}^N \langle C_i \rangle = 0$$

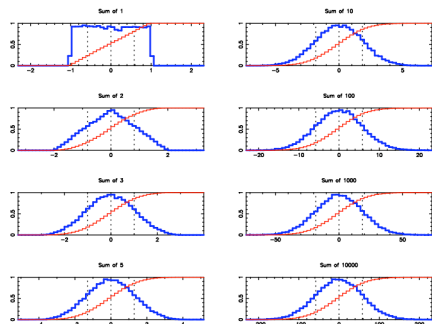
$$\sigma^2(S_N) = N \sigma_C^2 = N$$



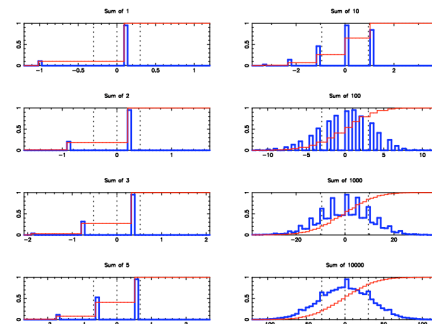
Coin Toss => Gaussian

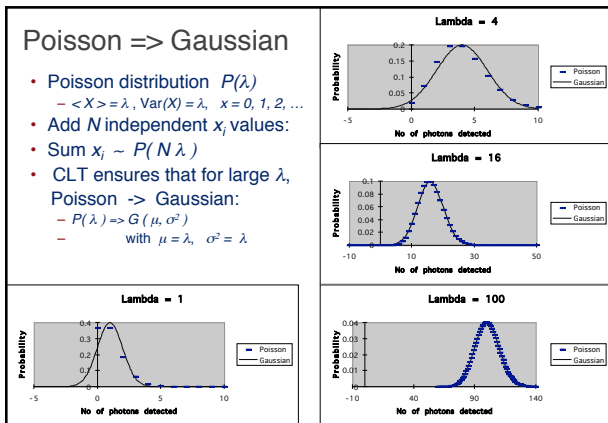


Uniform => Gaussian



Biased Coin => Gaussian





What is a Statistic?

- Anything you measure or compute from the data.
- Any function of the data.
- Because the data "jiggle", every statistic also "jiggles".
- Example: the average of N data points is a statistic:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$
- It has a definite value for a particular dataset.
- It has a probability distribution describing how it "jiggles" with the ensemble of repeated datasets.
- Note that $\bar{X} \neq \langle X \rangle$ Why? If $\langle X_i \rangle = \langle X \rangle$ then $\langle \bar{X} \rangle = \langle X \rangle$

Average of N data points

Sample Mean $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ is a statistic.

It has a pdf, with a mean value:

$$\langle \bar{X} \rangle = \left\langle \frac{1}{N} \sum_i X_i \right\rangle = \frac{1}{N} \left\langle \sum_i X_i \right\rangle = \frac{1}{N} \sum_i \langle X_i \rangle$$

...and a variance:

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{N} \sum_i X_i\right) = \frac{1}{N^2} \text{Var}\left(\sum_i X_i\right) = \frac{1}{N^2} \sum_i \text{Var}(X_i)$$

assuming $\text{Cov}[X_a, X_b] = \text{Var}[X_a] \delta_{ab}$

Sample Mean: Unbiased and lower Variance

If X_i have the same mean, $\langle X_i \rangle = \langle X \rangle$, then:

$$\langle \bar{X} \rangle = \frac{1}{N} \sum_{i=1}^N \langle X_i \rangle = \frac{N \langle X \rangle}{N} = \langle X \rangle$$

$\therefore \bar{X}$ is an unbiased estimator of $\langle X \rangle$. 😊

If X_i all have the same variance, $\text{Var}[X_i] = \sigma^2$, and are uncorrelated, $\text{Cov}[X_i, X_j] = \sigma^2 \delta_{ij}$, then:

$$\text{Var}(\bar{X}) = \frac{1}{N^2} \left(\sum_i \text{Var}(X_i) \right) = \frac{N \sigma^2}{N^2} = \frac{\sigma^2}{N}$$

$\therefore \sigma(\bar{X}) = \frac{\sigma}{\sqrt{N}}$, i.e. \bar{X} "jiggles" much less 😊 than a single data value X_i does.

Many other Unbiased Statistics

- Sample median (half points above, half below)

- $(X_{\max} + X_{\min}) / 2$
- Any single point X_i chosen at random from sequence
- Weighted average: $\frac{\sum_i w_i X_i}{\sum_i w_i}$ \bar{X} uses weights $w_i = 1$
- Which un-biased statistic is best? (best = minimum variance)

Inverse-variance weights are best!

- Variance of the weighted mean (assume $\text{Cov}[X_p, X_j] = \sigma_j^2 \delta_{ij}$):

$$\text{Var}\left[\frac{\sum_i w_i X_i}{\sum_i w_i}\right] = \frac{\text{Var}\left[\sum_i w_i X_i\right]}{\left(\sum_i w_i\right)^2} = \frac{\sum_i w_i^2 \text{Var}[X_i]}{\left(\sum_i w_i\right)^2} = \frac{\sum_i w_i^2 \sigma_i^2}{\left(\sum_i w_i\right)^2}$$

- What are the optimal weights?
- The **variance** of the weighted average is **minimised** when:

$$w_i = \frac{1}{\text{Var}(X_i)} \equiv \frac{1}{\sigma_i^2}$$
- Let's verify this -- it's important!

Optimising the weights

- To minimise the variance of the weighted average, set:

$$0 = \frac{\partial}{\partial w_k} \left(\frac{\sum_i w_i^2 \sigma_i^2}{\left(\sum_i w_i\right)^2} \right) = \frac{2 w_k \sigma_k^2}{\left(\sum_i w_i\right)^2} - \frac{2 \sum_i w_i^2 \sigma_i^2}{\left(\sum_i w_i\right)^3} \left(\frac{\partial \left(\sum_i w_i\right)}{\partial w_k} \right)$$

$$= \frac{2}{\left(\sum_i w_i\right)^2} \left(w_k \sigma_k^2 - \frac{\sum_i w_i^2 \sigma_i^2}{\left(\sum_i w_i\right)} \right) \Rightarrow w_k = \frac{1}{\sigma_k^2}$$

(Note: $\sum_i w_i^2 \sigma_i^2 = \sum_i w_i$ for $w_i = 1/\sigma_i^2$)

The Optimal Average

- Good principles for constructing statistics:
 - **Unbiased** -> no systematic error
 - **Minimum variance** -> smallest possible statistical error
- Optimal (inverse-variance weighted) average:

N datapoints: $X_i = \langle X \rangle \pm \sigma_i$
 $\langle X_i \rangle = \langle X \rangle$ $\text{Cov}[X_i, X_j] = \sigma_i^2 \delta_{ij}$

$$\hat{X} = \frac{\sum_i X_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}$$

- Is unbiased, since: $\langle \hat{X} \rangle = \langle X \rangle$

$$\sigma^2(\hat{X}) = \frac{1}{\sum_i 1 / \sigma_i^2}$$

- And minimum variance:

Memorise !

Compare: Equal vs Optimal Weights

- Both are unbiased: $\langle \hat{X} \rangle = \langle \bar{X} \rangle = \langle X_i \rangle = \langle X \rangle$
- Bad data spoils the Sample Mean (information lost).
- Optimal average ALWAYS improves with more data.
- Consider $N = 2$:

$$\bar{X} = \frac{X_1 + X_2}{2} \qquad \hat{X} = \frac{\frac{X_1}{\sigma_1^2} + \frac{X_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

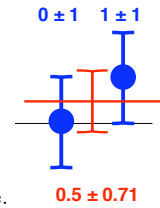
$$\text{Var}[\bar{X}] = \frac{\sigma_1^2 + \sigma_2^2}{4} \qquad \text{Var}[\hat{X}] = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

Averaging with Equal Error Bars

2 data points with equal error bars:

$$\bar{X} = \frac{0+1}{2} = \frac{1}{2}, \quad \sigma^2(\bar{X}) = \frac{1^2+1^2}{4} = \frac{1}{2}$$

$$\hat{X} = \frac{\frac{0}{1^2} + \frac{1}{1^2}}{\frac{1}{1^2} + \frac{1}{1^2}} = \frac{1}{2}, \quad \sigma^2(\hat{X}) = \frac{1}{\frac{1}{1^2} + \frac{1}{1^2}} = \frac{1}{2}$$



In this case $\hat{X} = \bar{X}$ since the σ_i are all the same.

Averaging with Unequal Error Bars

2 data points with unequal error bars:

$$\bar{X} = \frac{0+1}{2} = \frac{1}{2}, \quad \sigma^2(\bar{X}) = \frac{1^2+4^2}{4} = \frac{17}{4}$$

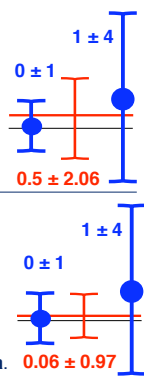
Information lost since $\sigma(\bar{X}) > \sigma(X_1)$. ☹️

$$\hat{X} = \frac{\frac{0}{1^2} + \frac{1}{4^2}}{\frac{1}{1^2} + \frac{1}{4^2}} = \frac{1}{17}, \quad \sigma^2(\hat{X}) = \frac{1}{\frac{1}{1^2} + \frac{1}{4^2}} = \frac{1}{17/16} = \frac{16}{17}$$

Now $\sigma(\hat{X}) < \sigma(X_1)$. ☺️

Optimal weights retain all the information.

Optimal Average always improves with new data. 0.06 ± 0.97



Compare: Equal vs Optimal Weights

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

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

Normal Average 1.50 ± 2.06

$$\hat{X} = \frac{\sum_i X_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}$$

$$\sigma^2(\hat{X}) = \frac{1}{\sum_i 1 / \sigma_i^2}$$

Optimal Average 1.06 ± 0.97

Equal weights:
 Poor data degrades the result.
 Better to ignore "bad" data.
 Information lost.

Optimal weights:
 New data always improves the result.
 Use ALL the data, but with appropriate 1 / Variance weights.
Must have good error bars.