ADA04 - 9am Tue 20 Sep 2022

Central Limit Theorem
(why Gaussians are special)

Statistics:
Sample Mean (unbiased)
vs
Optimal Average (unbiased, minimum variance)

Review: Functions of Random Variables

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

Conserve probability :

$$d(\text{Prob}) = f(Y)\ |\text{d}Y| = f(X)\ |\text{d}X|$$

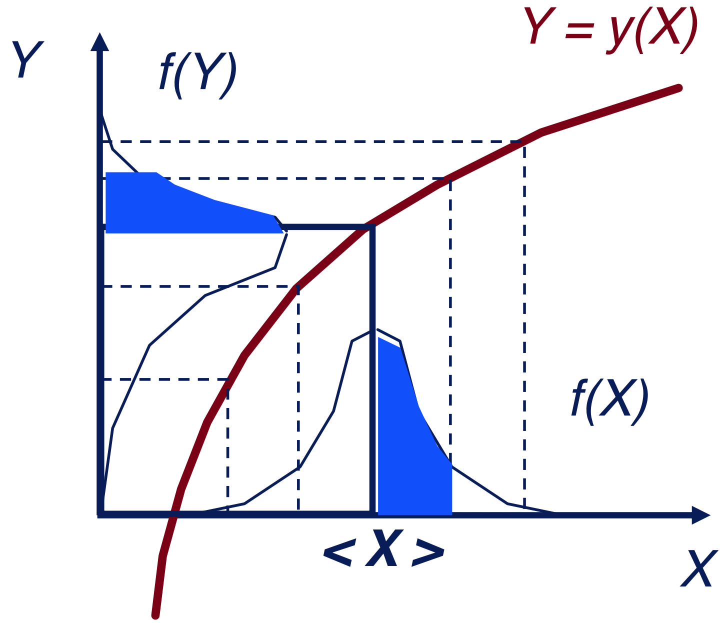
$$f(Y) = f(X)\ \left|\frac{dX}{dY}\right| = \frac{f(X)}{|y'(X)|}$$

mean value (biased)

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

standard deviation (stretched)

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



Power-law: $Y = X^p \quad \langle Y \rangle = \langle X \rangle^p + \text{bias}$

$$\frac{\sigma(Y)}{\langle Y \rangle} \approx p\, \frac{\sigma(X)}{\langle X \rangle} \qquad \frac{\text{bias}}{\langle Y \rangle} = \frac{p(p-1)}{2} \left(\frac{\sigma(X)}{\langle X \rangle} \right)^2$$

Example: $Y = \sqrt{X} \quad X = 100 \pm 10$

$$\langle Y \rangle \approx \sqrt{\langle X \rangle} = ? \quad \frac{\sigma(Y)}{\langle Y \rangle} = ? \quad \frac{\text{bias}}{\langle Y \rangle} = ?$$

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:

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

- The **means and variances accumulate** (algebra of random variables):

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

- 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

$$C = +1 \quad \text{if heads}$$

$$\quad -1 \quad \text{if tails}$$

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

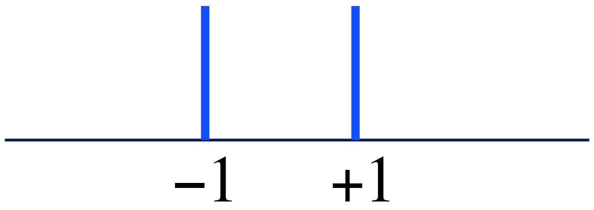
$$S_N \equiv \sum_{i=1}^{N} C_i$$

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

$$\sigma^2(S_N) = \sum_{i=1}^{N} \sigma^2(C_i) = N\,\sigma_C^2 = N$$

+1 H

-1 T

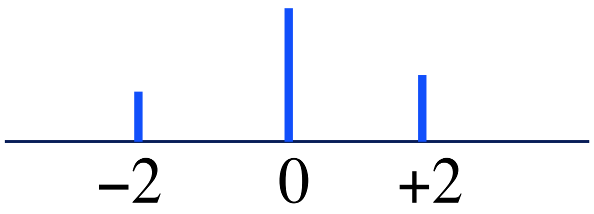


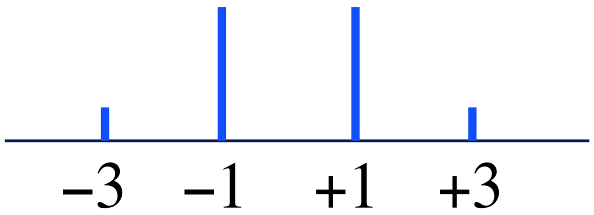
+2 H H

0 H T

0 T H

-2 T T





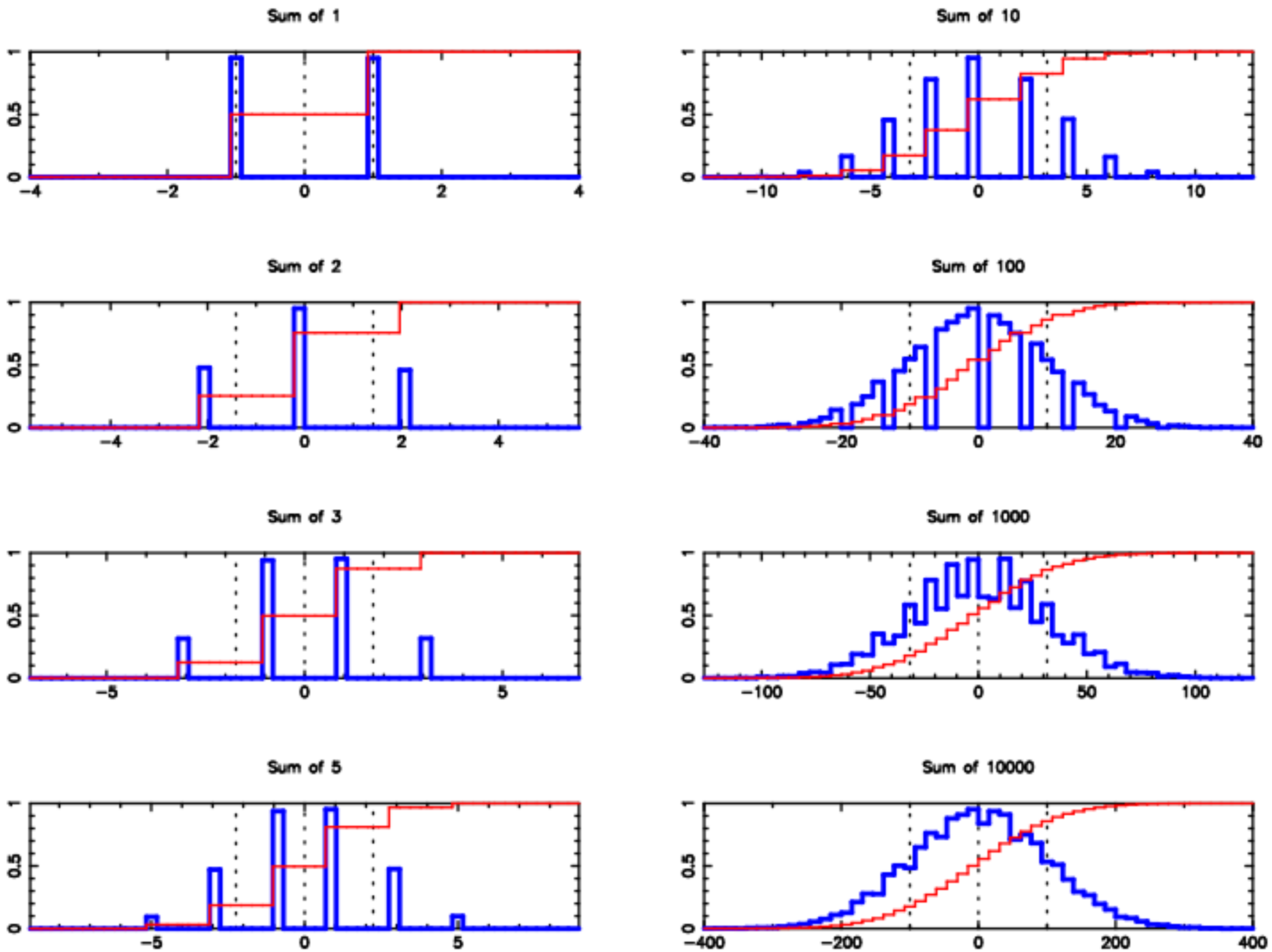
+3 H H H

+1 H H T, H T H, T H H

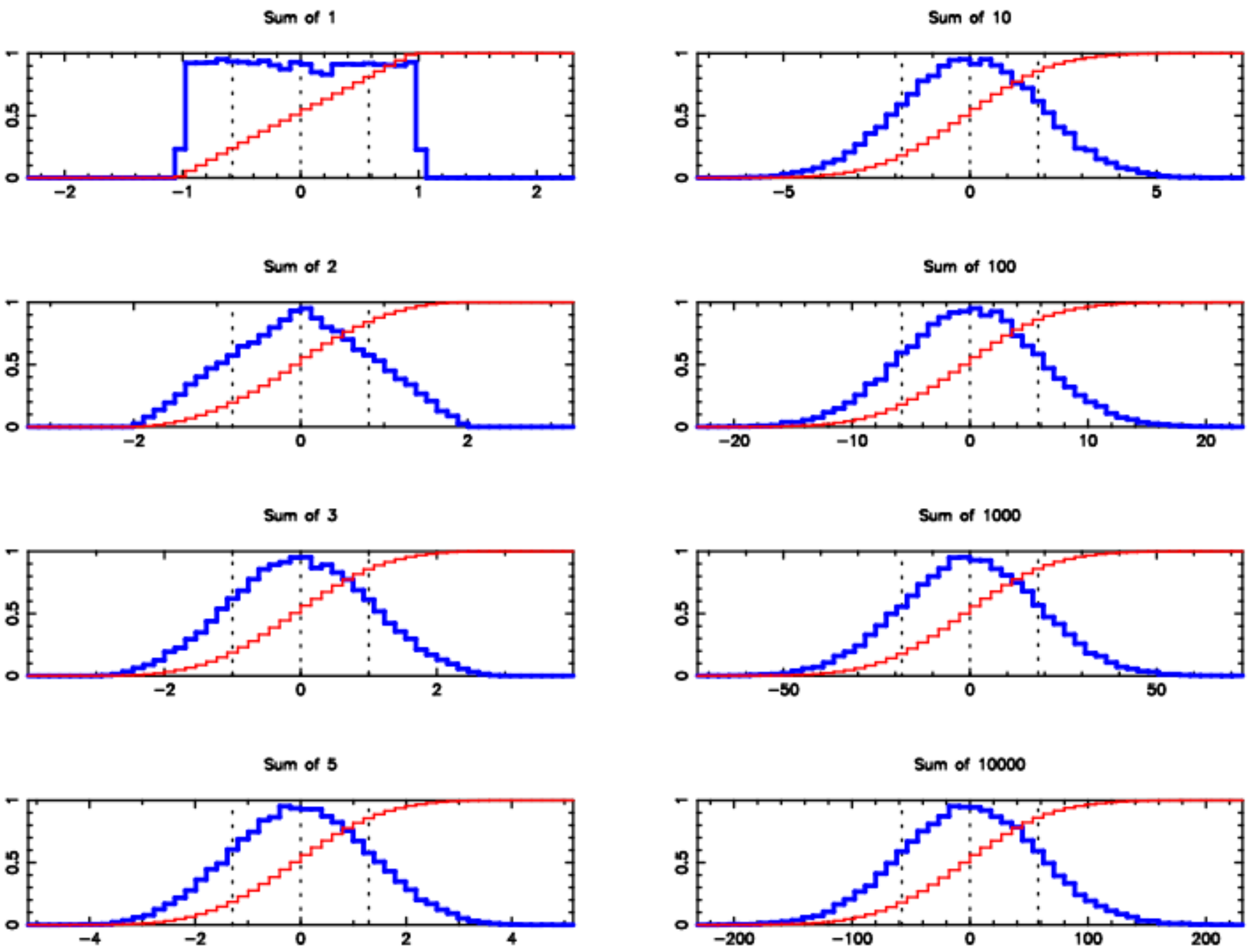
-1 H T T, T H T, T T H

-3 T T T

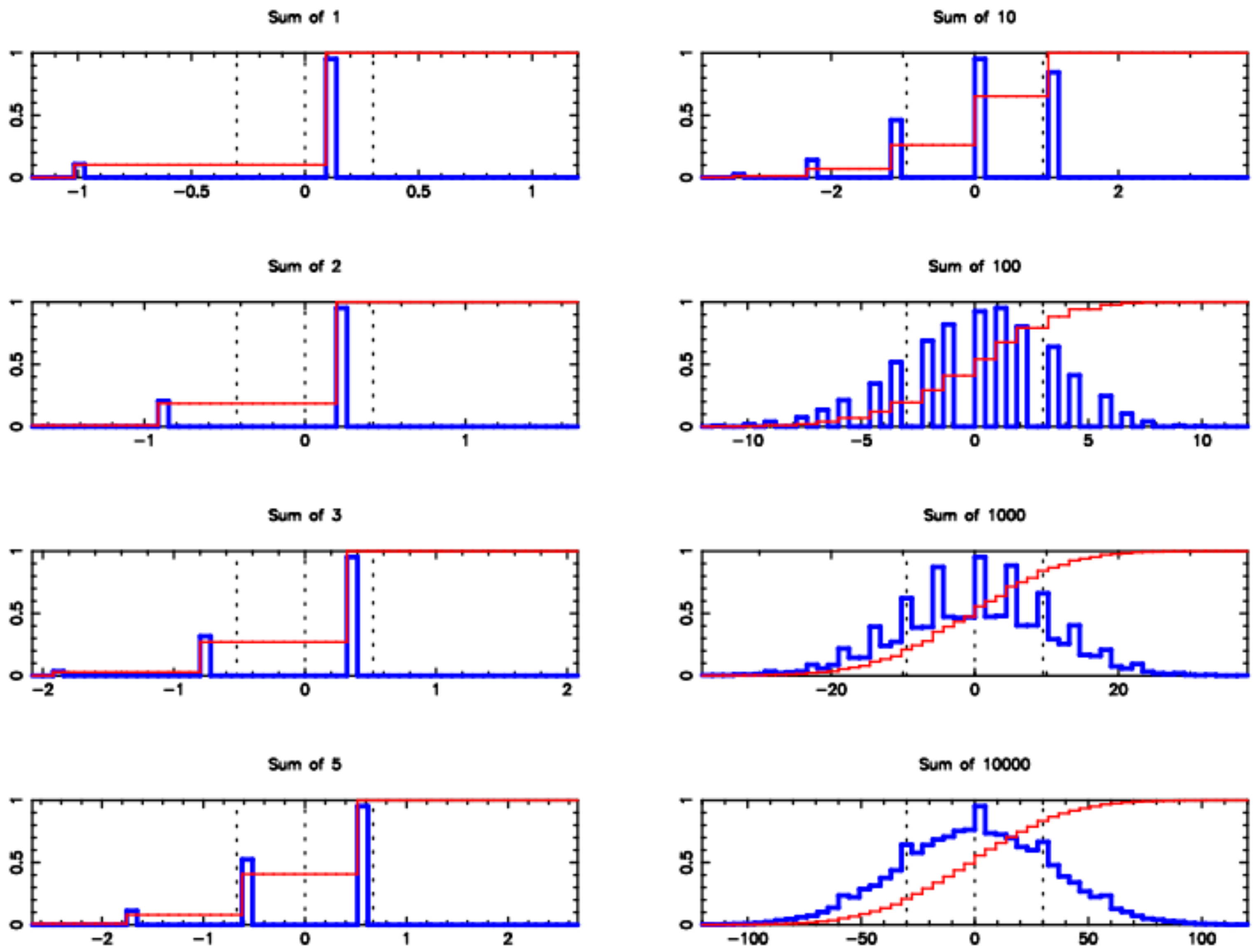
Coin Toss => Gaussian



Uniform => Gaussian

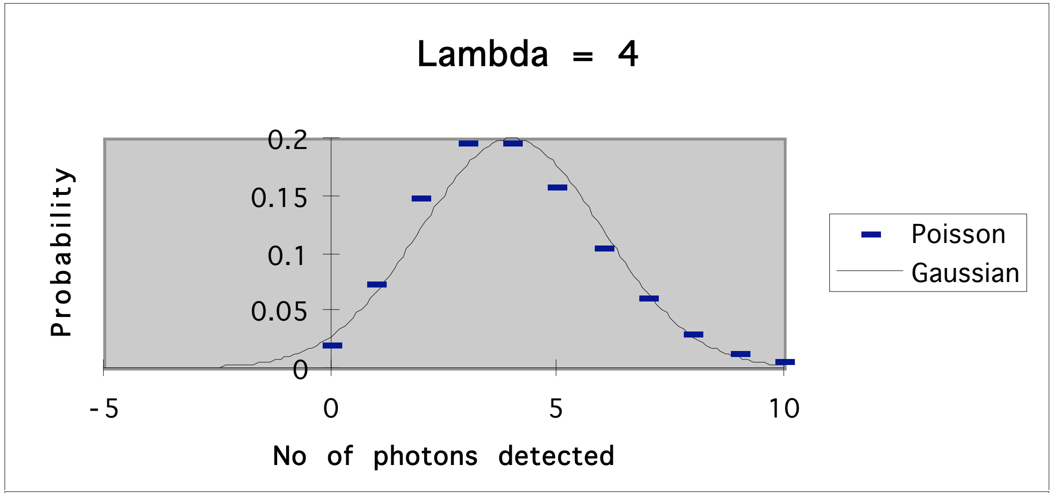


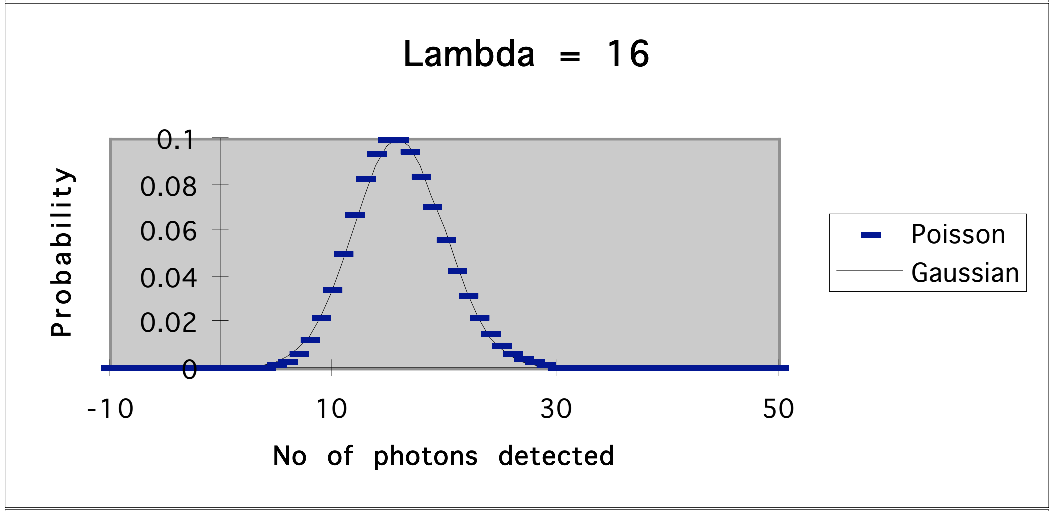
Biased Coin => Gaussian

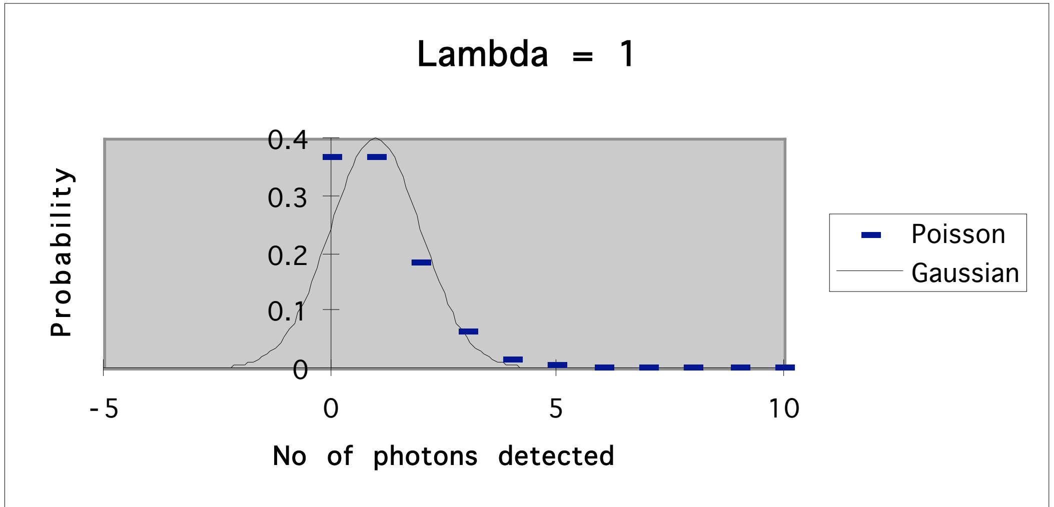


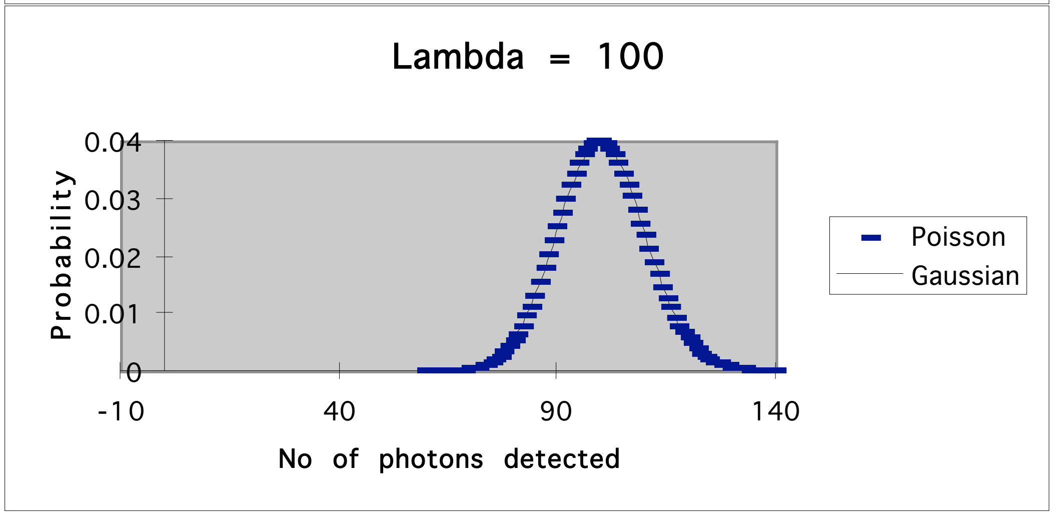
Poisson => Gaussian

- Poisson distribution $P(\lambda)$
  - $< X > = \lambda$, $\mathrm{Var}(X) = \lambda$, $x = 0, 1, 2, \ldots$
- Add N independent x_i values:
- Sum $x_i \sim P(N \lambda)$
- CLT ensures that for large λ, Poisson -> Gaussian:
  - $P(\lambda) \Rightarrow G (\mu, \sigma^2)$
  - with $\mu = \lambda$, $\sigma^2 = \lambda$



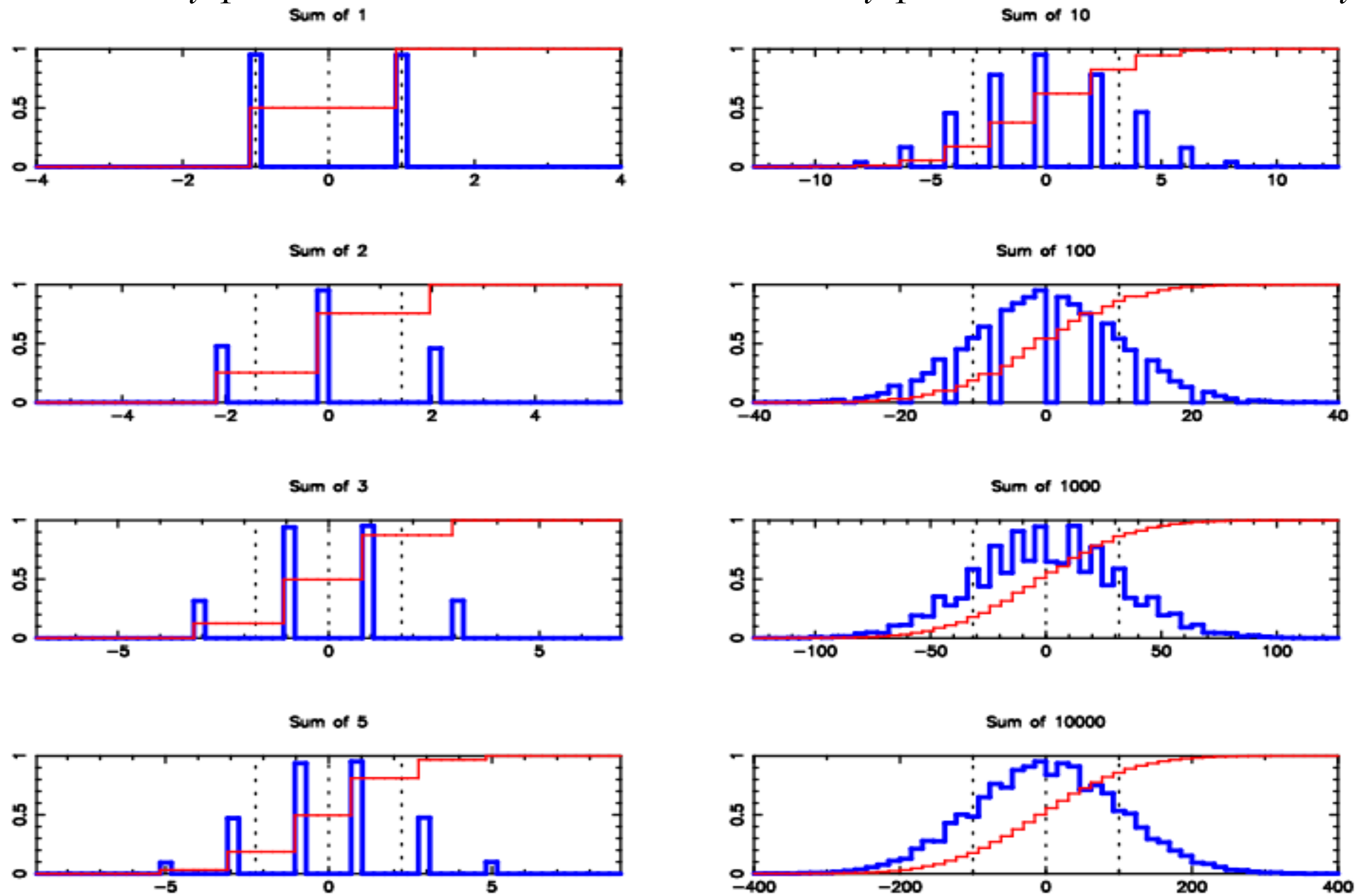






Recap : Central Limit Theorem : (e.g. Coin Toss => Gaussian)

$$X_i = \mu_i \pm \sigma_i \qquad \sum_{i=1}^{N} X_i \Rightarrow G\left(\mu, \sigma^2\right) \qquad \mu = \sum_{i=1}^{N} \mu_i = 0 \qquad \sigma^2 = \sum_{i=1}^{N} \sigma_i^2$$

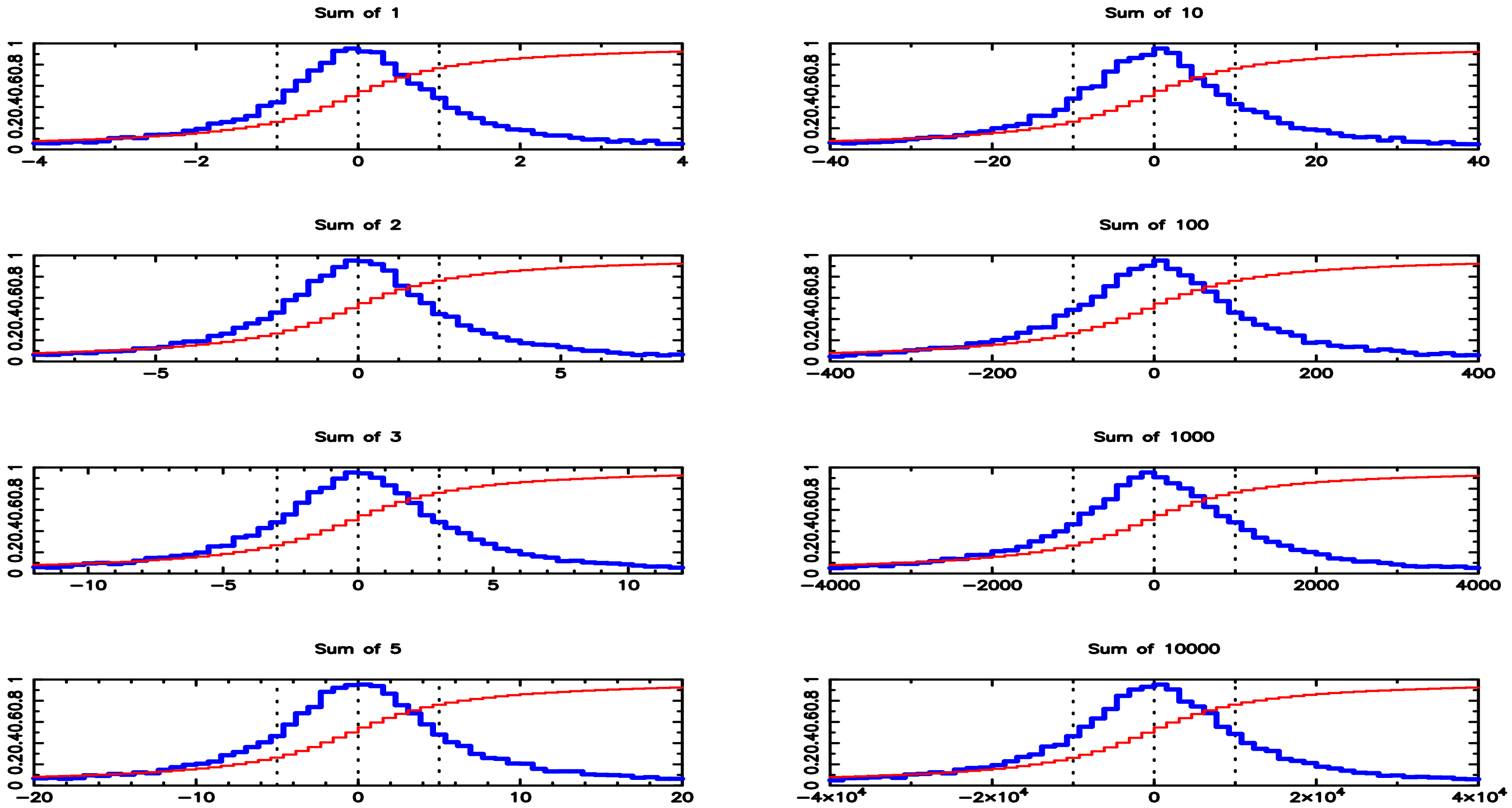


Convolution of Gaussians = Gaussian

$$X_i \sim G\left(\mu_i,\sigma_i^2\right) \qquad \sum_{i=1}^{N} X_i \sim G(\mu,\sigma^2) \quad \mu = \sum_{i=1}^{N} \mu_i \quad \sigma^2 = \sum_{i=1}^{N} \sigma_i^2$$

Convolution of Lorentzians = Lorentzian

$$X_i \sim L\left(\mu_i,\sigma_i\right) \qquad \sum_{i=1}^{N} X_i \sim L(\mu,\sigma) \quad \mu = \sum_{i=1}^{N} \mu_i \quad \sigma = \sum_{i=1}^{N} \sigma_i$$



Definition : 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:

$$\overline{X} \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$$

- 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 $\overline{X} \neq \langle X \rangle$ Why?
- If $\langle X_i \rangle = \langle X \rangle$, then $\langle \overline{X} \rangle = \langle X \rangle$.

Sample Mean : Average of N data points

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

It has a probability distribution,

with a mean value:

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

and a variance:

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

---

assuming Cov[X_a , X_b] = Var[X_a] δ_{ab}

Sample Mean: Unbiased and lower Variance

If X_i have the same mean, $< X_i > \; = \; < X >$, then:

$$\left\langle \overline{X} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \left\langle X_i \right\rangle = \frac{N \left\langle X \right\rangle}{N} = \left\langle X \right\rangle$$

$\therefore \quad \overline{X}$ is an <u>unbiased estimator</u> of $\left\langle X \right\rangle$.

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

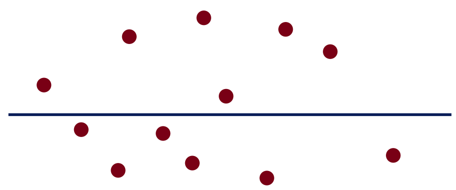
$$\mathrm{Var}\left(\overline{X}\right) = \frac{1}{N^2} \left(\sum_i \mathrm{Var}\left(X_i\right) \right) = \frac{N \, \sigma^2}{N^2} = \frac{\sigma^2}{N}$$

$\therefore \quad \sigma\left(\overline{X}\right) = \dfrac{\sigma}{\sqrt{N}}$, i.e. $\overline{X}$ "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 Cov[X_i, X_j] = $\sigma_i^2\,\delta_{ij}$) :

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

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

$$w_i = \frac{1}{\mathrm{Var}(X_i)} \equiv \frac{1}{\sigma_i^2}\,.$$

- Let's verify this -- it's important!

Optimising the weights

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

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

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

$$(\text{Note: } \sum w_i^2 \sigma_i^2 = \sum w_i \text{ for } w_i = 1/\sigma_i^2)$$

The Optimal Average

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

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

N datapoints: $X_i = \langle X \rangle \pm \sigma_i$

$$\langle X_i \rangle = \langle X \rangle \quad \mathrm{Cov}\left[X_i, X_j\right] = \sigma_i^2 \delta_{ij}$$

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

- And minimum variance:

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

Memorise !

Compare: Equal vs Optimal Weights

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

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

$$\mathrm{Var}\left[\overline{X}\right] = \frac{\sigma_1^2 + \sigma_2^2}{4} \qquad \mathrm{Var}\left[\hat{X}\right] = \frac{1}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}$$

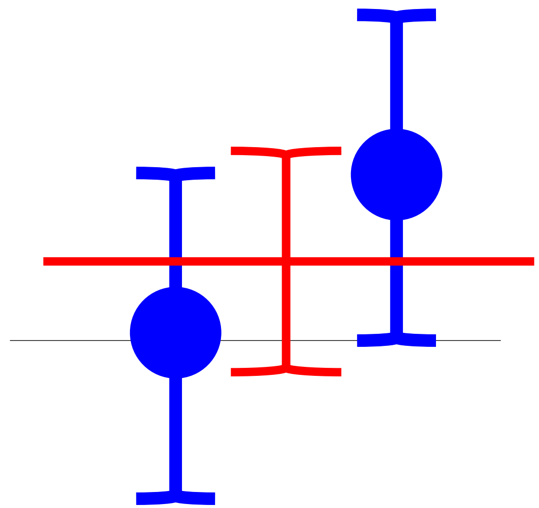
Averaging Data with Equal Error Bars

2 data points with equal error bars:

$$\overline{X} = \frac{0+1}{2} = \frac{1}{2}, \qquad \sigma^2(\overline{X}) = \frac{1^2+1^2}{4} = \frac{1}{2}.$$

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

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

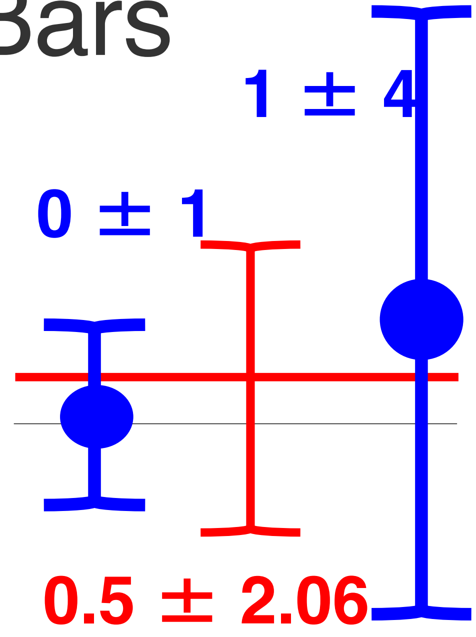


Averaging Data with Unequal Error Bars

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 $\sigma(\bar{X}) > \sigma(X_1)$.

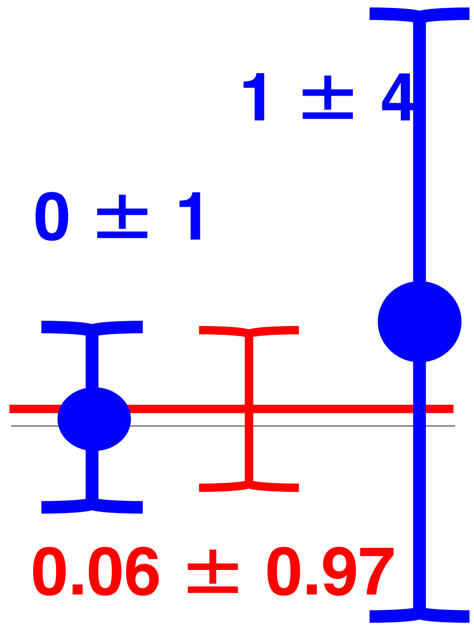


$$\hat{X} = \frac{\frac{0}{1^2} + \frac{1}{4^2}}{\frac{1}{1^2} + \frac{1}{4^2}} = \frac{1}{17}, \quad \sigma^2(\hat{X}) = \frac{1}{\frac{1}{1^2} + \frac{1}{4^2}} = \frac{1}{17/16} = \frac{16}{17}.$$

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

Optimal weights retain all the information.

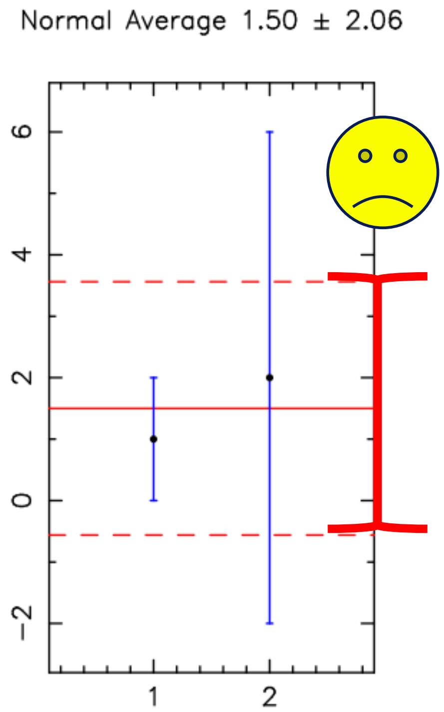
Optimal Average always improves with new data.

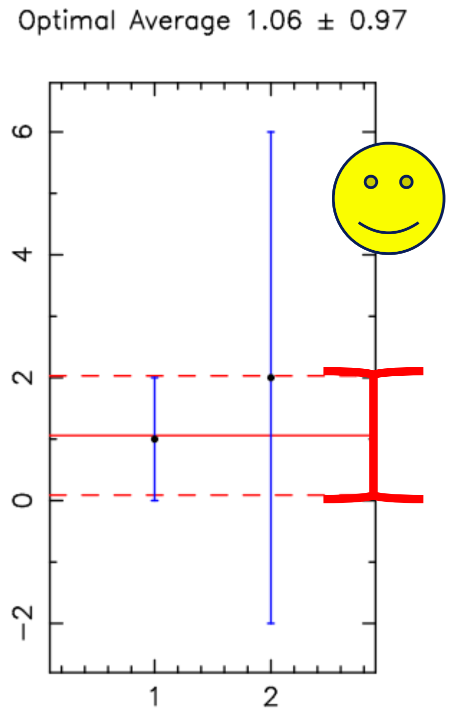


Compare: Sample Mean vs Optimal Average

$$\overline{X} \equiv \frac{1}{N}\sum_i X_i$$

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





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

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

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.

Fini -- ADA 04