ADA04 - 9am Tue 20 Sep 2022

Central Limit Theorem
(why Gaussians are special)

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

79

Review: Functions of Random Variables

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

Conserve probability :

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

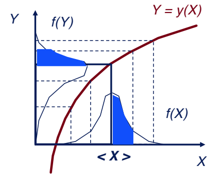
$f(Y) = f(X)\left|\dfrac{dX}{dY}\right| = \dfrac{f(X)}{|y'(X)|}$

mean value (biased)

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

standard deviation (stretched)

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



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

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

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

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

80

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:

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

- 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.)

81

Example: Coin Toss

$C = +1$ if heads

$\quad -1$ if tails

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

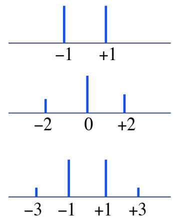
$$S_N = \sum_{i=1}^{N} C_i$$

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

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

+1 H
-1 T

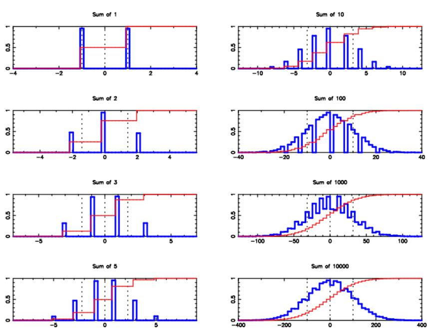
+2 HH
0 HT
0 TH
-2 TT



+3 HHH
+1 HHT, HTH, THH
-1 HTT, THT, TTH
-3 TTT

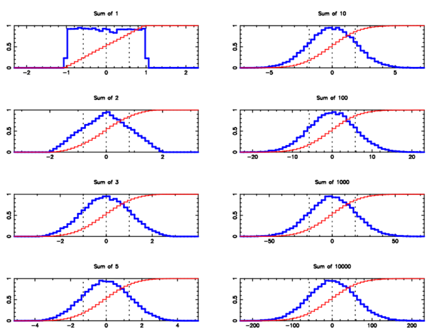
82

Coin Toss => Gaussian



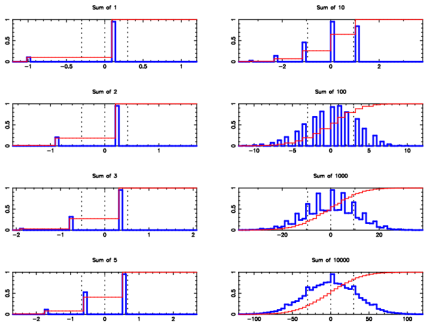
83

Uniform => Gaussian



84

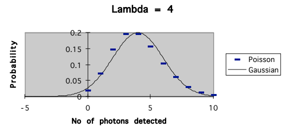
Biased Coin => Gaussian

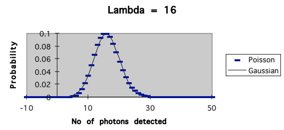


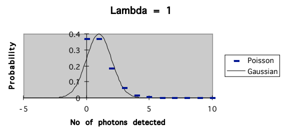
85

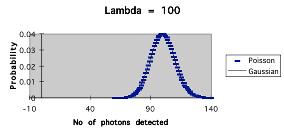
Poisson => Gaussian

- 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 $\mu = \lambda$, $\sigma^2 = \lambda$





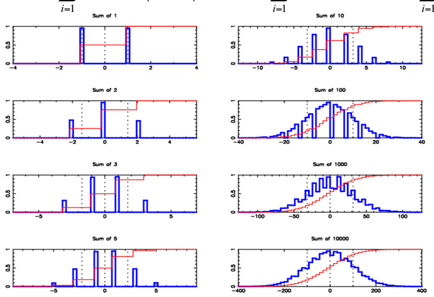




86

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

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



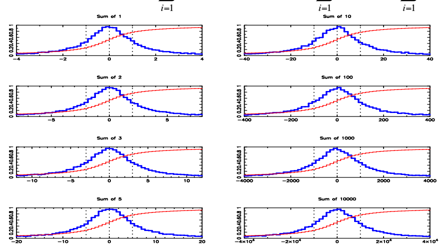
87

Convolution of Gaussians = Gaussian

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

Convolution of Lorentzians = Lorentzian

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



88

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

89

Sample Mean : Average of N data points

Sample Mean $\overline{X} \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$ 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 $\mathrm{Cov}[\,X_a, X_b\,] = \mathrm{Var}[\,X_a\,]\ \delta_{ab}$

90

Sample Mean: Unbiased and lower Variance

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

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

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

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

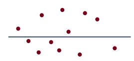
$$\mathrm{Var}\left(\bar{X}\right) = \frac{1}{N^2}\left(\sum_i \mathrm{Var}\left(X_i\right)\right) = \frac{N\,\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

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

91

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: $\dfrac{\sum_i w_i X_i}{\sum_i w_i}$ $\qquad \bar{X}$ uses weights $w_i = 1$
- **Which un-biased statistic is best ?**
  (best = minimum variance)

92

Inverse-variance weights are best!

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

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

- 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!

93

Optimising the weights

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

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

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

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

94

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

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

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

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

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

Memorise !

95

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 $N = 2$:

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

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

96

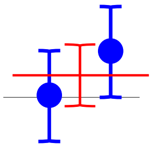
Averaging Data with Equal Error Bars

2 data points with equal error bars:

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

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

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



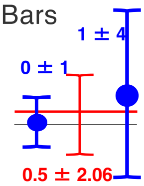
97

Averaging Data with Unequal Error Bars

2 data points with unequal error bars:

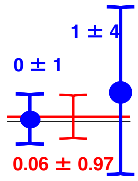
$$\overline{X} = \frac{0+1}{2} = \frac{1}{2}, \quad \sigma^2(\overline{X}) = \frac{1^2+4^2}{4} = \frac{17}{4}.$$

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



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

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



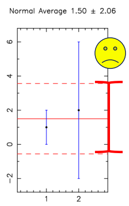
Optimal weights retain all the information.
Optimal Average always improves with new data.

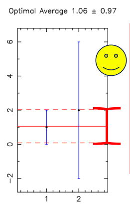
98

Compare: Sample Mean vs Optimal Average

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

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





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

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

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.

99

Fini -- ADA 04

100