AS5001 (= SUPAAAA)
ADA= "Advanced" (Astronomical) Data Analysis

Keith Horne PandA 315A kdh1@st-and.ac.uk

ADA web page: http://star-www.st-and.ac.uk/~kdh1/ada/ada.html

All lecture pdfs, homework, projects, videos on Moodle.

Supplementary Texts:

Press et al. (CUP) ***Numerical Recipes : The Art of Scientific Computing***

(on the web at Numerical.Recipes)

Wall & Jenkins (CUP) ***Practical Statistics for Astronomers***

Gregory (CUP) ***Bayesian Logical Data Analysis for the Physical Sciences***

Opinionated Lessons in Statistics*, by Bill Press. OpinionatedLessons.org*

ADA= “Advanced” (Astronomical) Data Analysis

Goal: Build concepts and skills for analysing quantitative data.

~15 Lectures: develop basic principles, illustrate with examples, extend step-by-step to build expertise for advanced analysis of datasets.

50% 2 Homework sets: test understanding, build skills

50% 2 Projects: analyse real datasets (Keck, HST)

NO EXAM :)

Work steadily, ask questions, get help when you don’t understand, and you will succeed.

ADA 01 - 10am Mon 12 Sep 2022

Astronomical Data + Noise
Statistical vs Systematic errors
Probability distributions (pdf, cdf)
Mode, Mean, Median
Variance, standard deviation, MAD
Skewness, Kurtosis
Parameterised distributions
(Uniform, Gaussian, Lorentzian,
Poisson, Exponential, Chi^2)

ADA Lecture 1 Outline

- ***Astronomical Data Sets***
- ***Noise :***
  - ***statistical vs systematic errors***
- ***Probability distributions :***
  - ***Mean vs Median***
  - ***Variance (standard deviation) vs MAD***
  - ***Central moments (skewness, kurtosis)***
- ***Survey of parameterised distributions***
  - ***Uniform, Gaussian, Lorentzian, Poisson, Exponential, Chi-squared***

Astronomical Datasets

- (Almost) all our information about the Universe arrives as photons. (neutrinos, gravitational waves*)*
- ***Photon properties:***

position : $\vec{x}$

time : t

direction : α, δ

energy : $E = h\nu = hc / \lambda$

polarisation : (Stokes parameters, $\vec{\mathrm{p}} = \mathrm{I,Q,U,V}$)

- Astronomical datasets are (usually) photon distributions confined by a detector to (some subset of) these properties:

$$D_i = \int P_i(\vec{x},t,\alpha,\delta,\lambda,\vec{p})\ f(\vec{x},t,\alpha,\delta,\lambda,\vec{p})\ \mathrm{d}(\vec{x},t,\alpha,\delta,\lambda,\vec{p}) \quad + \quad Noise_i$$

Photon detection probability for data point *i*

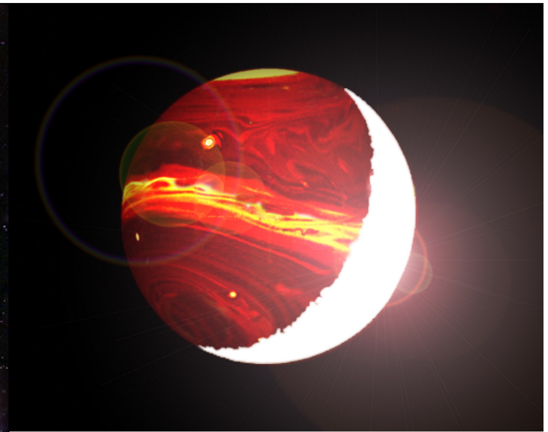
Photon distribution

Astronomical Datasets

- ***Direct imaging:***
  - ***size***
  - ***structure***

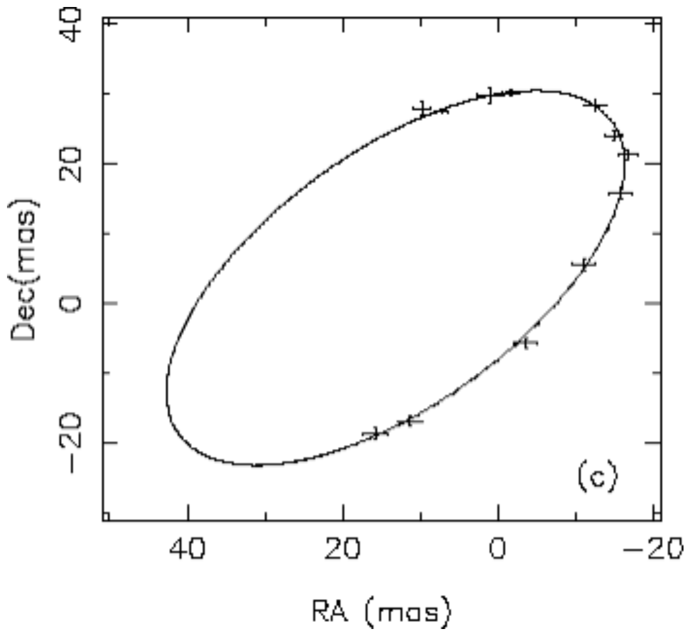
$$D(\alpha,\delta)$$



- ***Astrometry:***
  - ***distance***
  - ***parallax***
  - ***motion***
  - ***proper motion***
  - ***visual binary orbits***

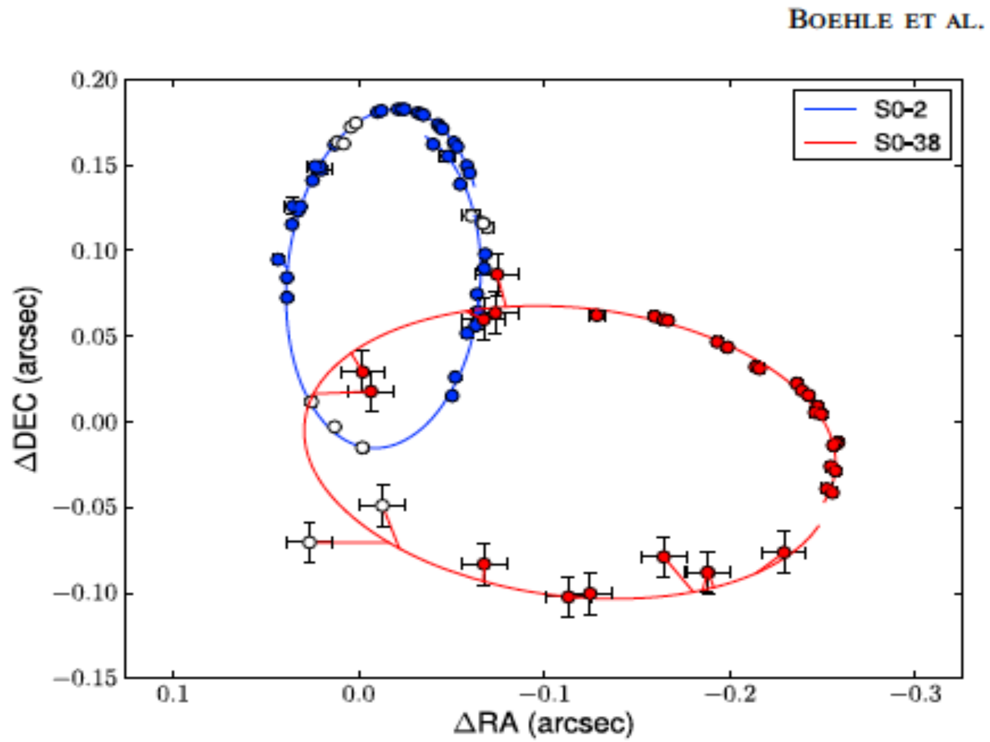
$$D(\alpha,\delta,t)$$

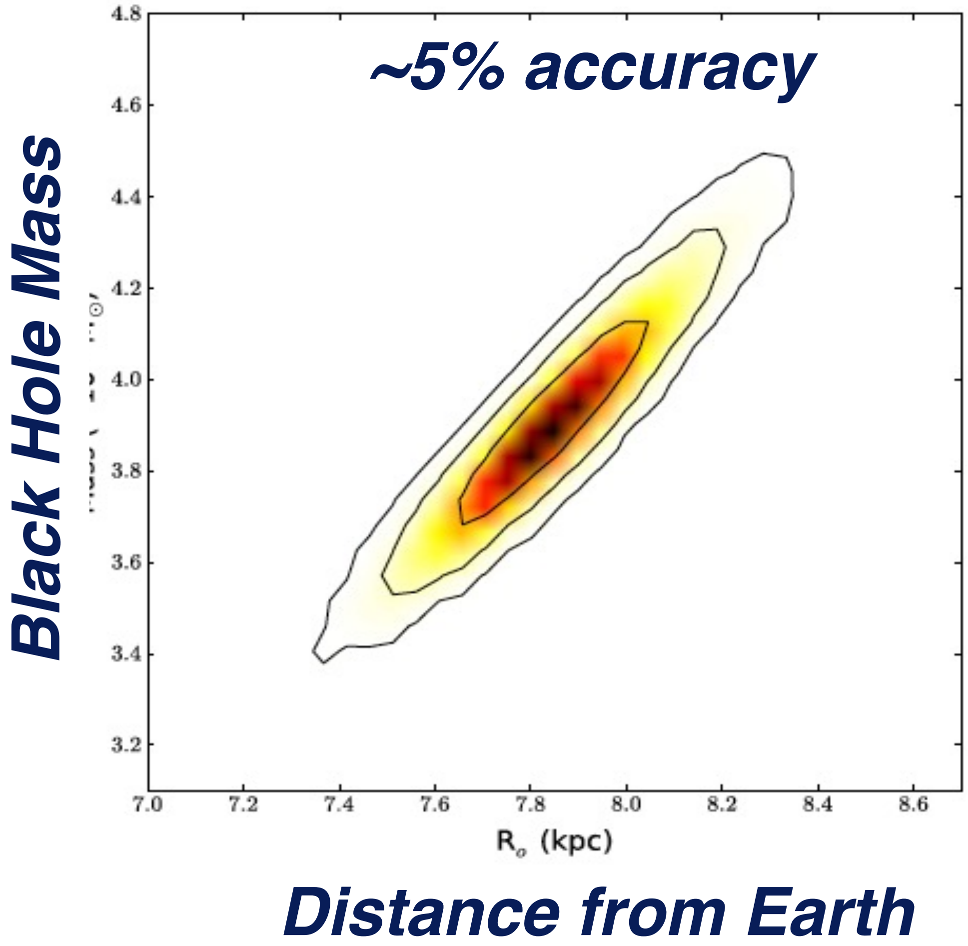


Black Hole Mass from Stellar Orbits

- Black Hole **in the Galactic Centre**
- Star orbits traced to find

$M_{BH} = (4.0 +/- 0.2\ ) \times 10^6\ \mathbf{M_\odot}$

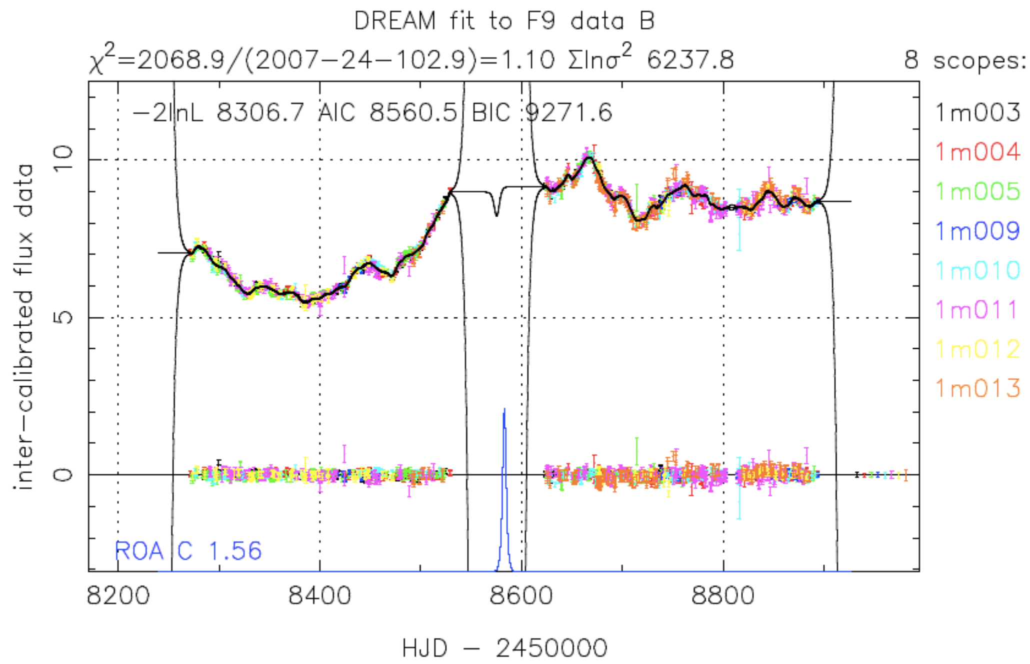




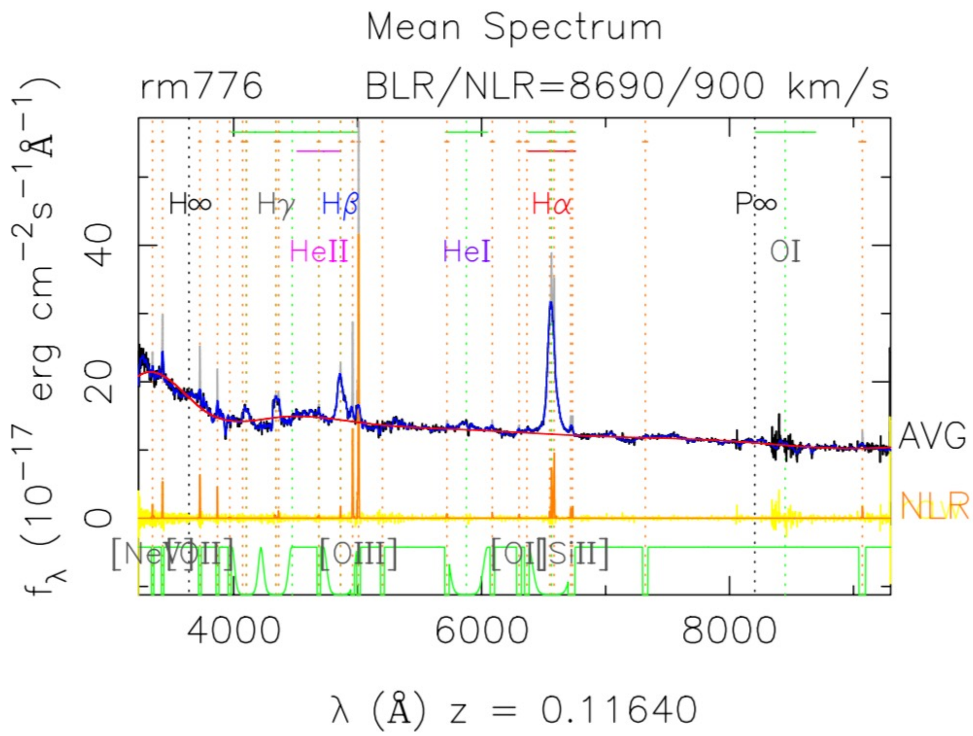
Boehle, Ghez et al (2016) ApJ

Astronomical Datasets

- ***Light curves:*** $D(t)$
  - ***Time variations***
  - ***Orbital periods***



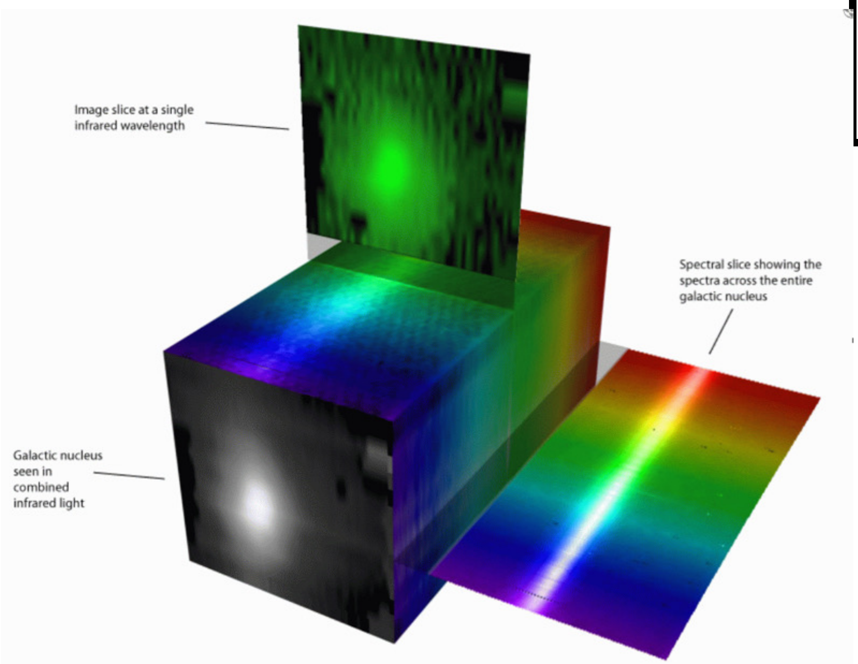
- ***Spectra:*** $D(\lambda)$
  - ***Physical conditions***
  - ***Temperature, density***
  - ***Velocities => masses***

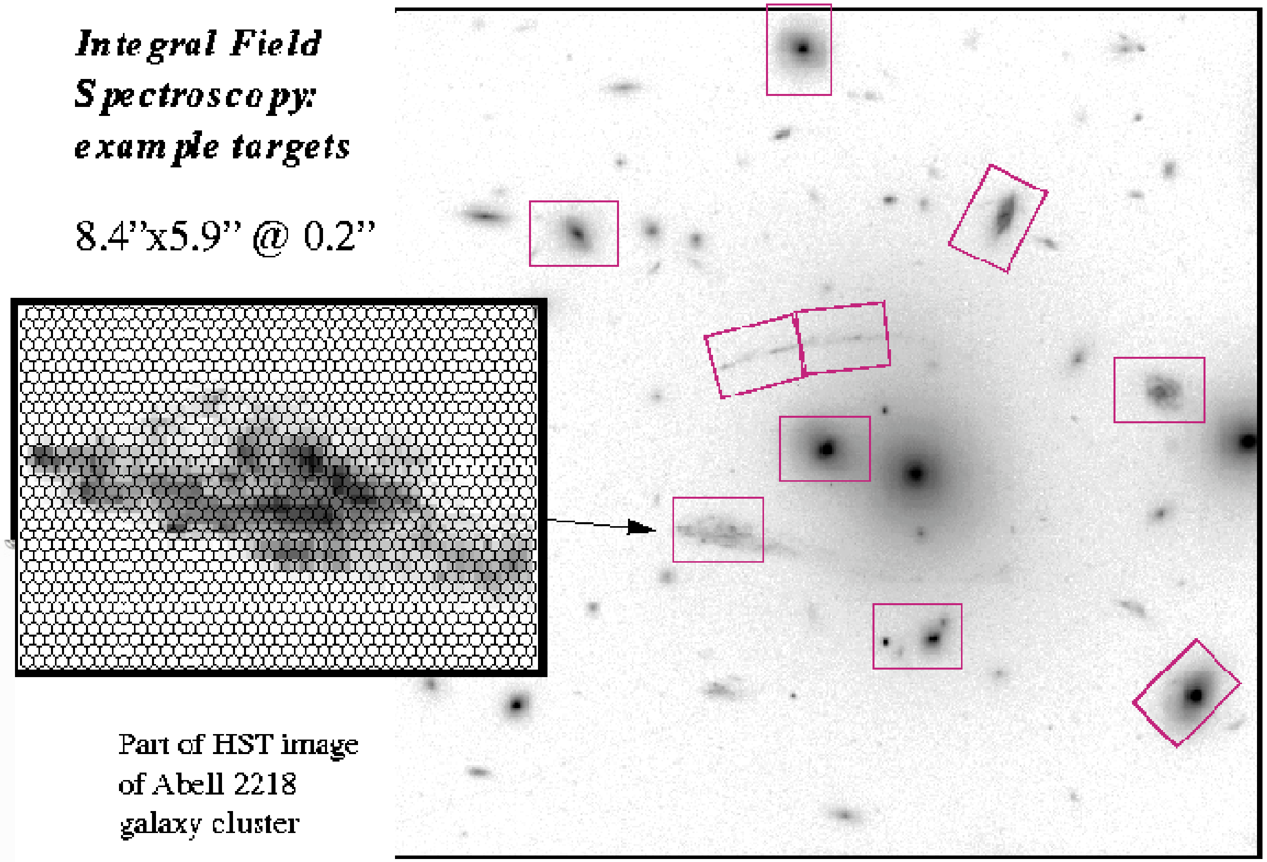


Integral-field Spectroscopy: $D(\alpha,\delta,\lambda)$

- *Close-packed array of fibres (or lenslets) giving spectra over a grid of positions on the sky.*

- *Probes spatial and spectral structure simultaneously.*

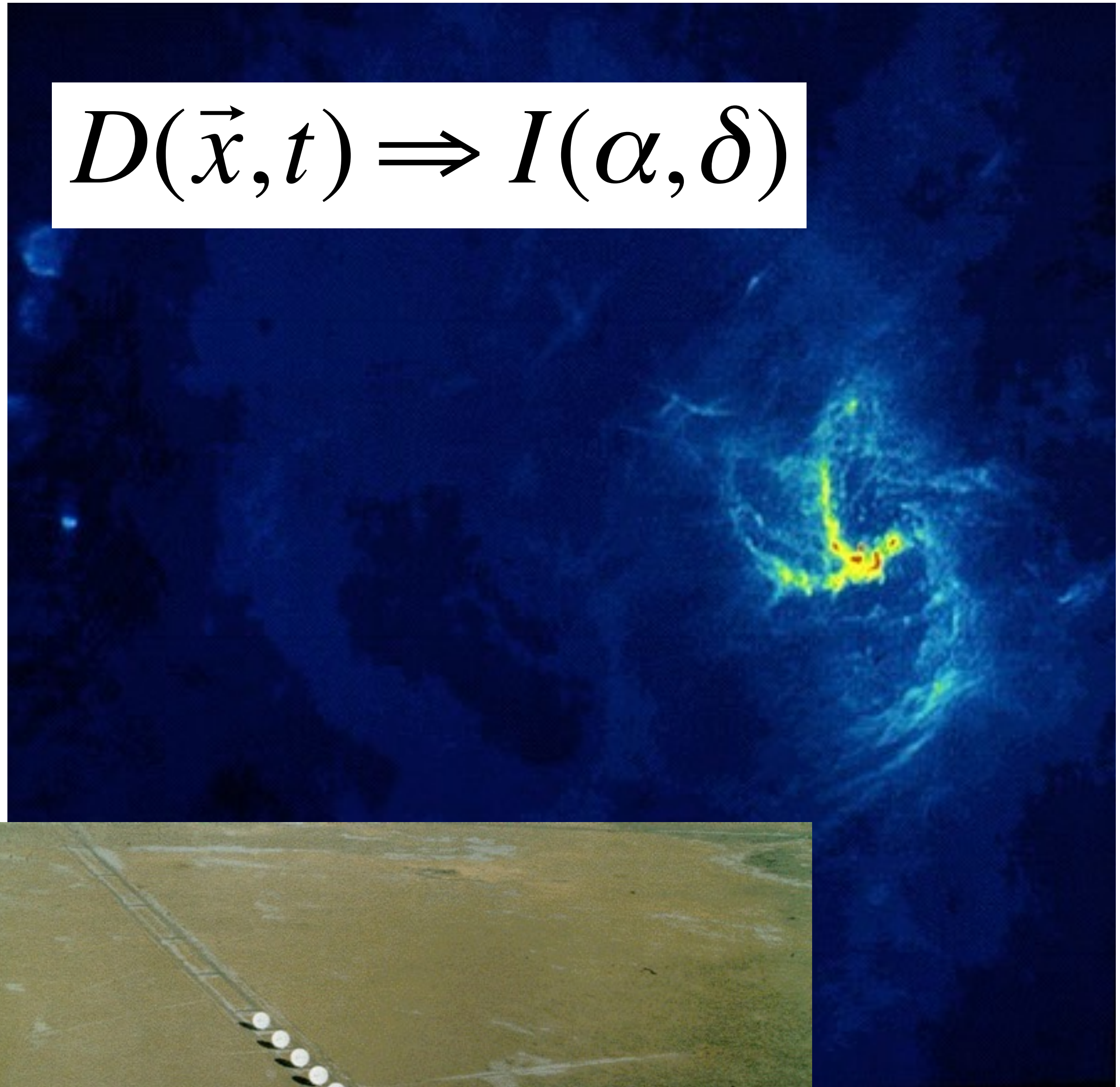




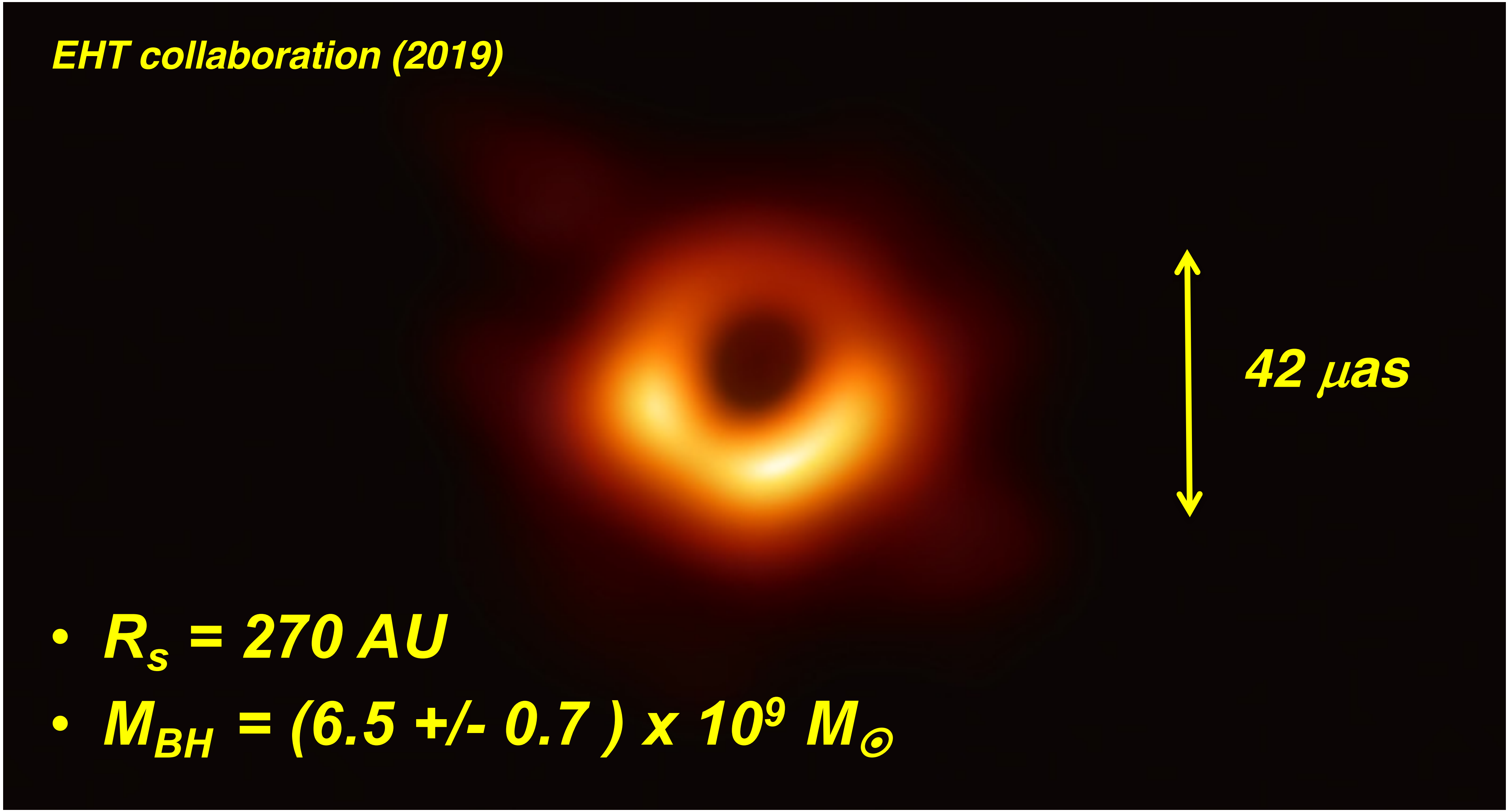
Interferometry:

- An example of ***Indirect Imaging***
- Use information about ***arrival time at different locations*** to infer angular structure of source.
- Picture: 6 cm radio map of “mini-spiral” of gas around Sgr A* (=black hole at the centre of our Milky Way galaxy).

$$D(\vec{x},t) \Rightarrow I(\alpha,\delta)$$



Black Hole in M87
imaged by the Event Horizon Telescope



Data are Data

There are many different types of data.

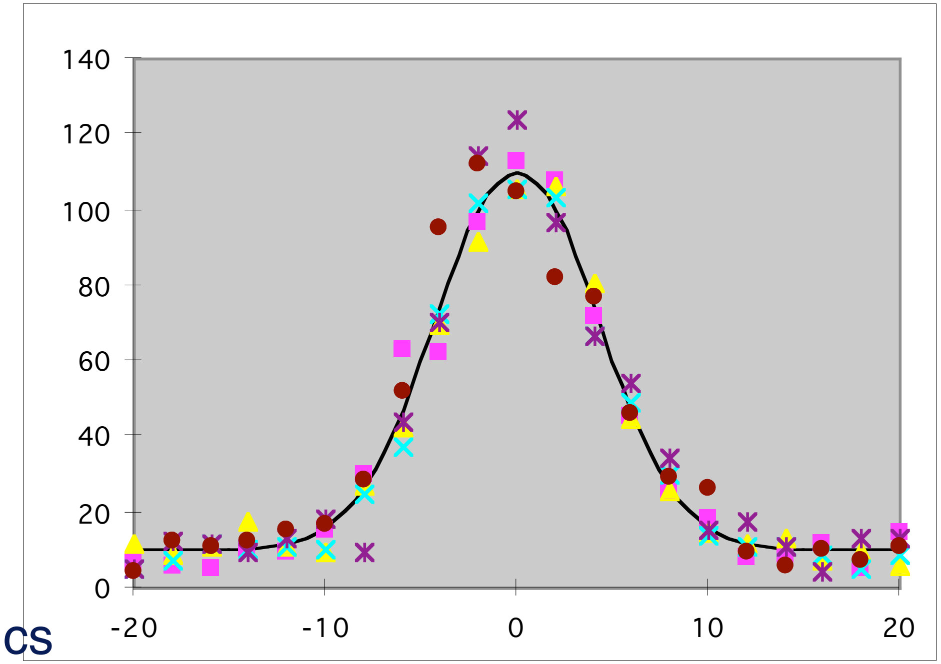
Photon properties define the dimensions of (most) astronomical datasets.

But: The same analysis techniques apply to all quantitative datasets.

(Astronomical or otherwise.)

Data are affected by Noise

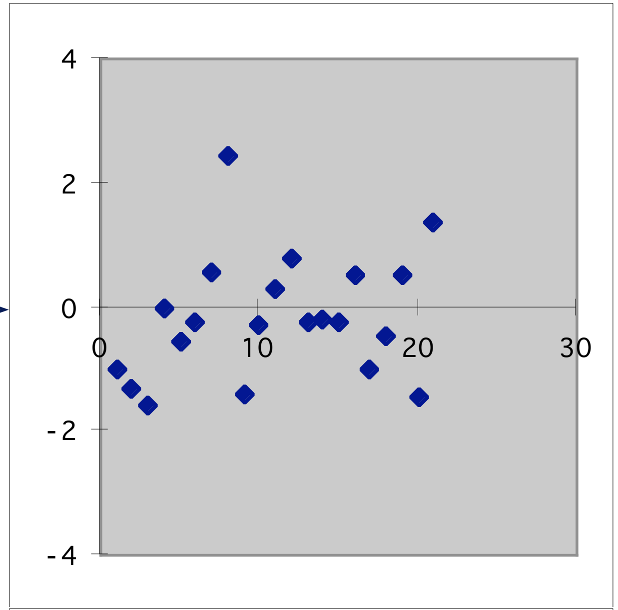
- Repetitions of the same experiment or observation give different results.
- e.g. spectral-line profile:



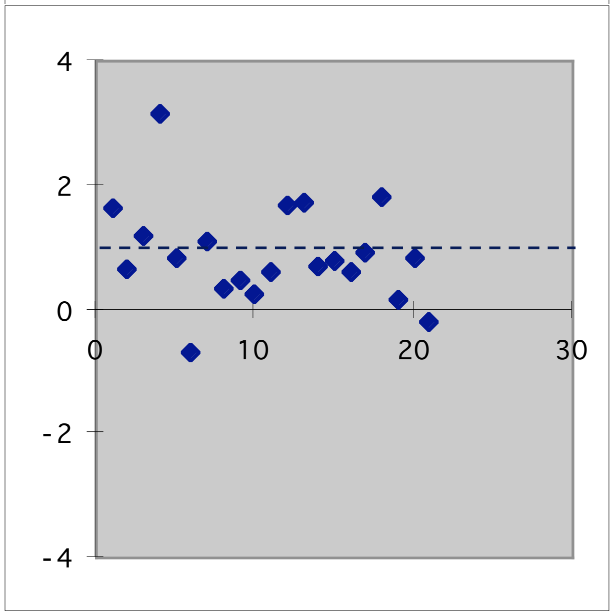
- Sources of noise:
- ***Quantum (Poisson) noise***
  - finite number of photons
- ***Thermal noise***
  - thermal fluctuations in the detector/electronics
- ***Rare events***
  - cosmic ray hits, instrument failures

Data Values as "Random Variables"

- Consider an ensemble of repeated measurements.
- Data values *"dance"* around.



- ***Statistical errors:***
  - From random nature of measurement process.
  - Can be reduced by averaging repeat measurements.
- ***Systematic errors (bias):***
  - Due to flawed measurement technique.
  - Bias remains after averaging repeat measurements.

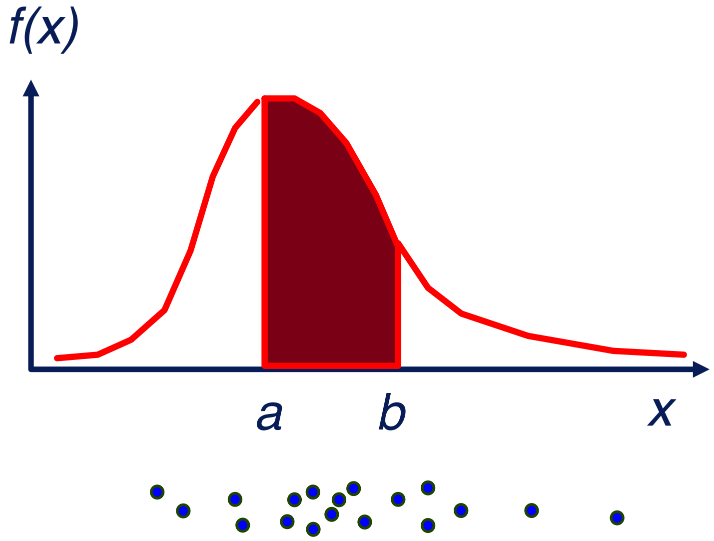


- ***Probability distributions*** describe this "dance" of the data values.

Probability Distributions (PDFs)

- ***Probability distribution*** *f(x)*
- aka: *probability density function* (pdf)
- defines the probability that *x* lies in some range:

$$P(a < x \le b) \equiv \int_a^b f(x)\,dx$$



- ***Probabilities add up to 1.***
- If *x* can take any value between -∞ and +∞ then

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Cumulative Probability Functions (CDFs)

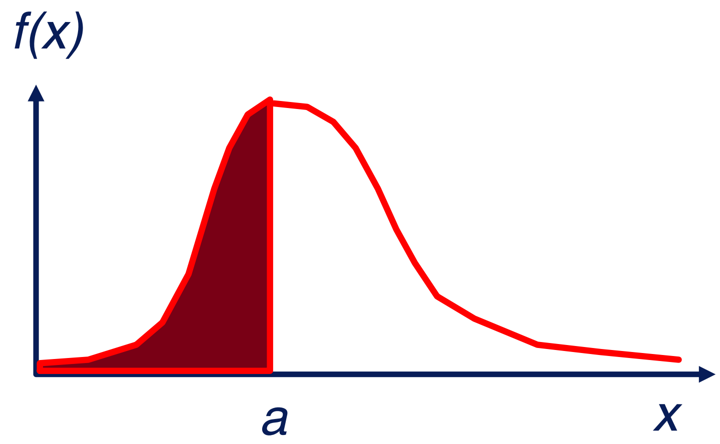
- Integrating *f(x)* gives the ***cumulative probability*** *F(a)* that $x \le a$:

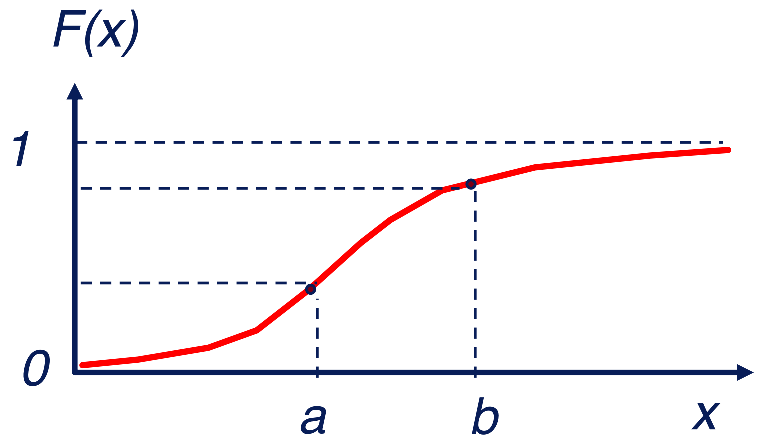
$$F(a) \equiv P(x \le a) \equiv \int_{-\infty}^{a} f(x)\,dx$$

$$F(-\infty) = 0 \qquad F(+\infty) = 1$$

$$P(a < x \le b) = \int_{a}^{b} f(x)\,dx$$

$$= F(b) - F(a)$$



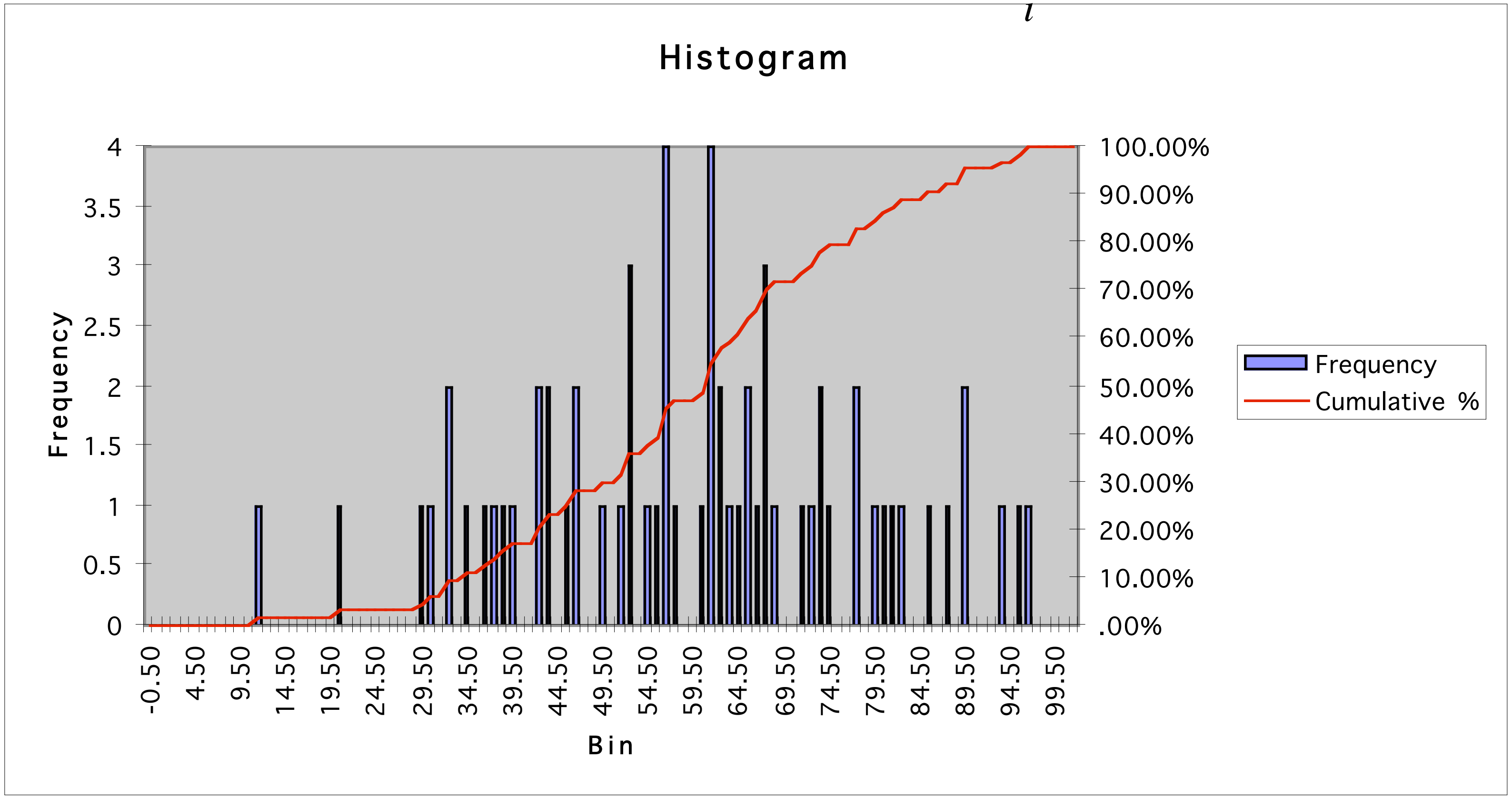


Discrete Probability Distributions

- Example:
  - Exam marks
  - Photons per pixel

$$f(x) \equiv \sum_i p_i \delta(x - x_i)$$

$$F(x) \equiv \sum_i p_i \text{ for all } x_i \leq x$$

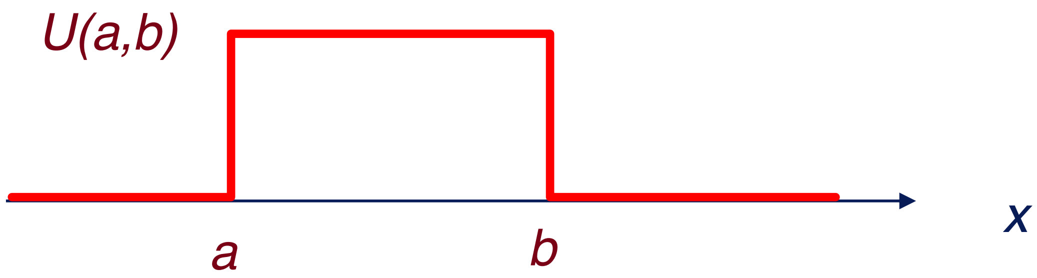


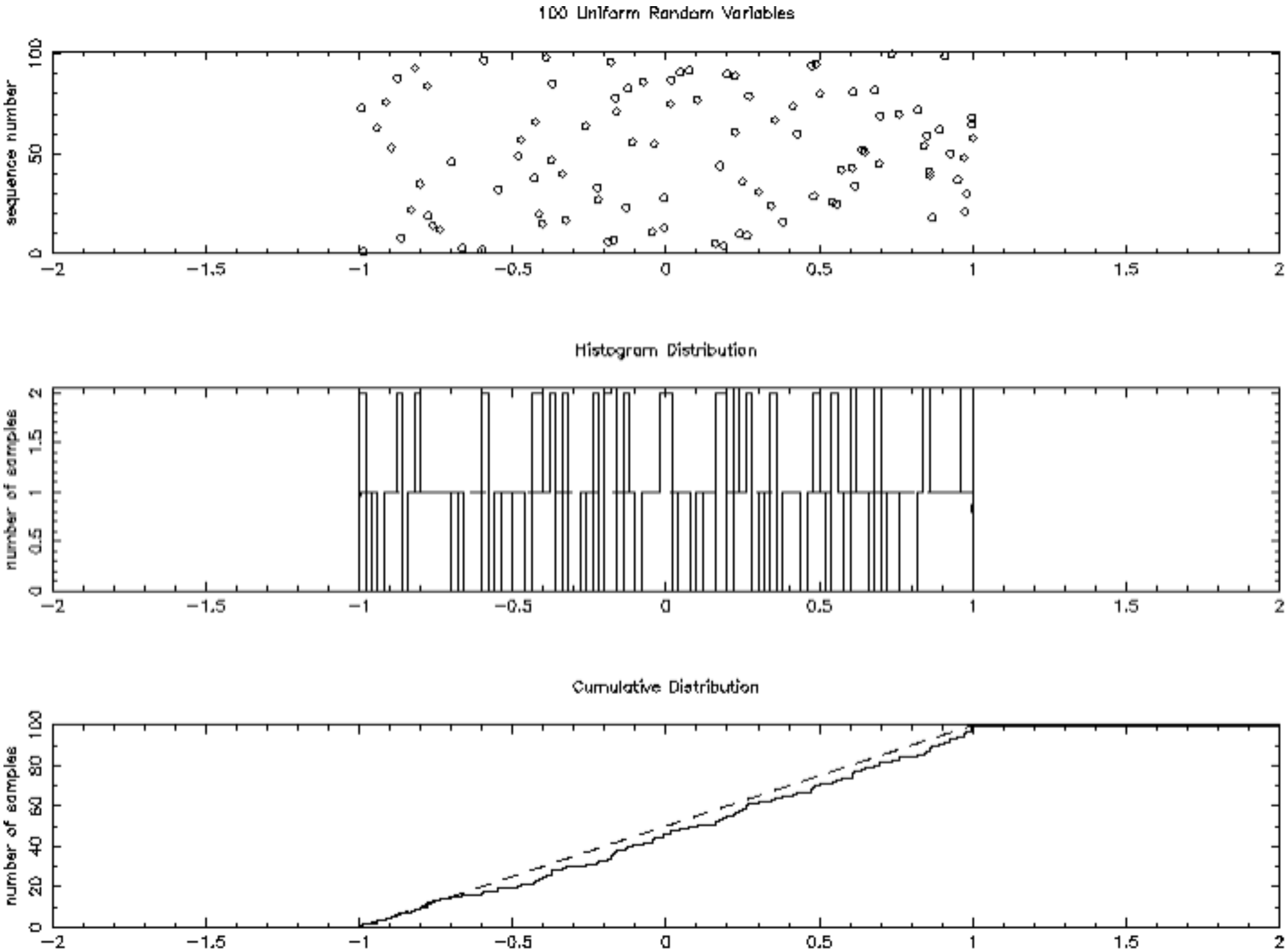
Uniform Distribution *U(a,b)*

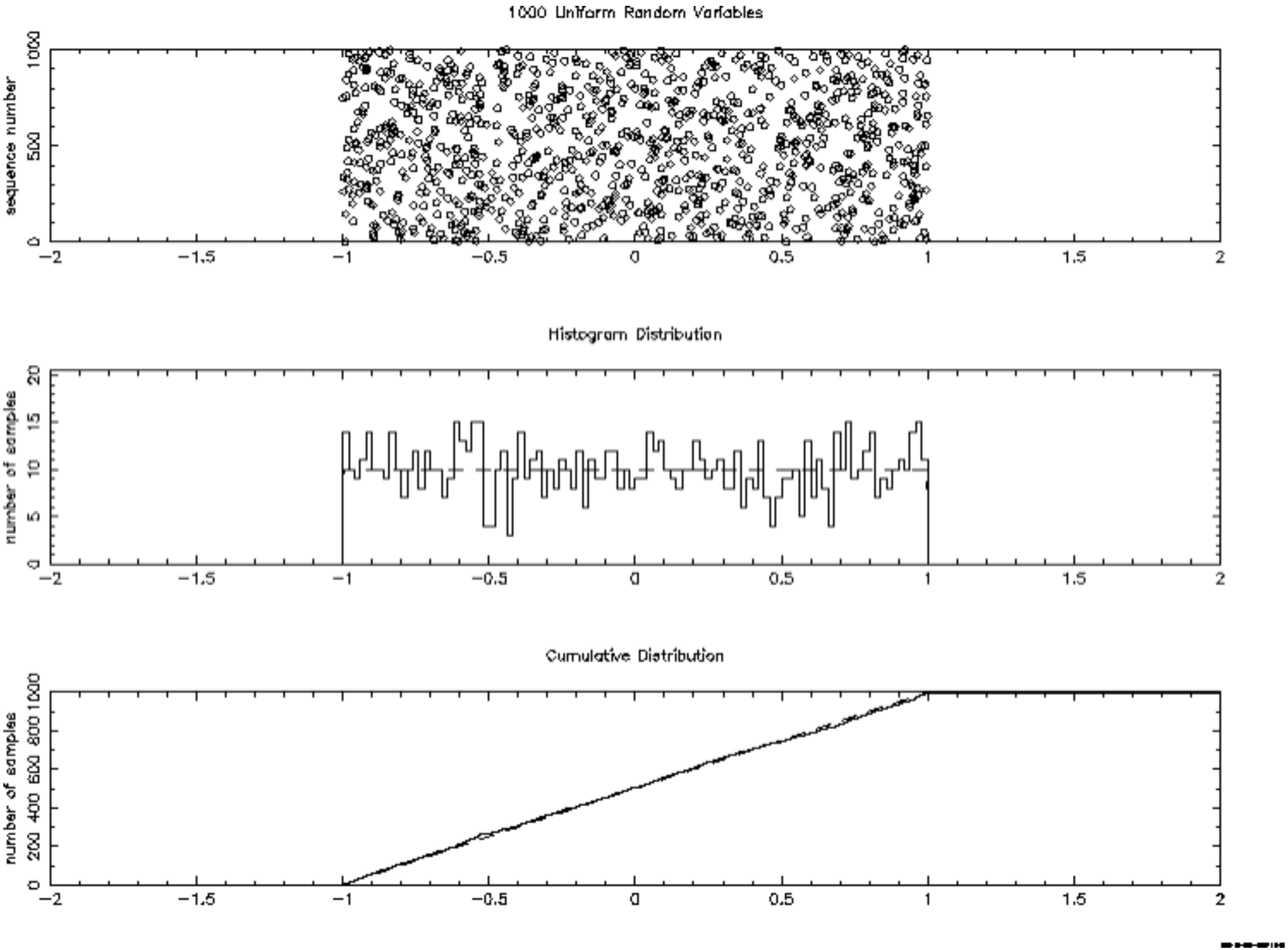
- Also known as a "**boxcar**" or "**tophat**" distribution:

$$f(x) = \frac{1}{|b-a|} \quad \text{for } a < x < b$$

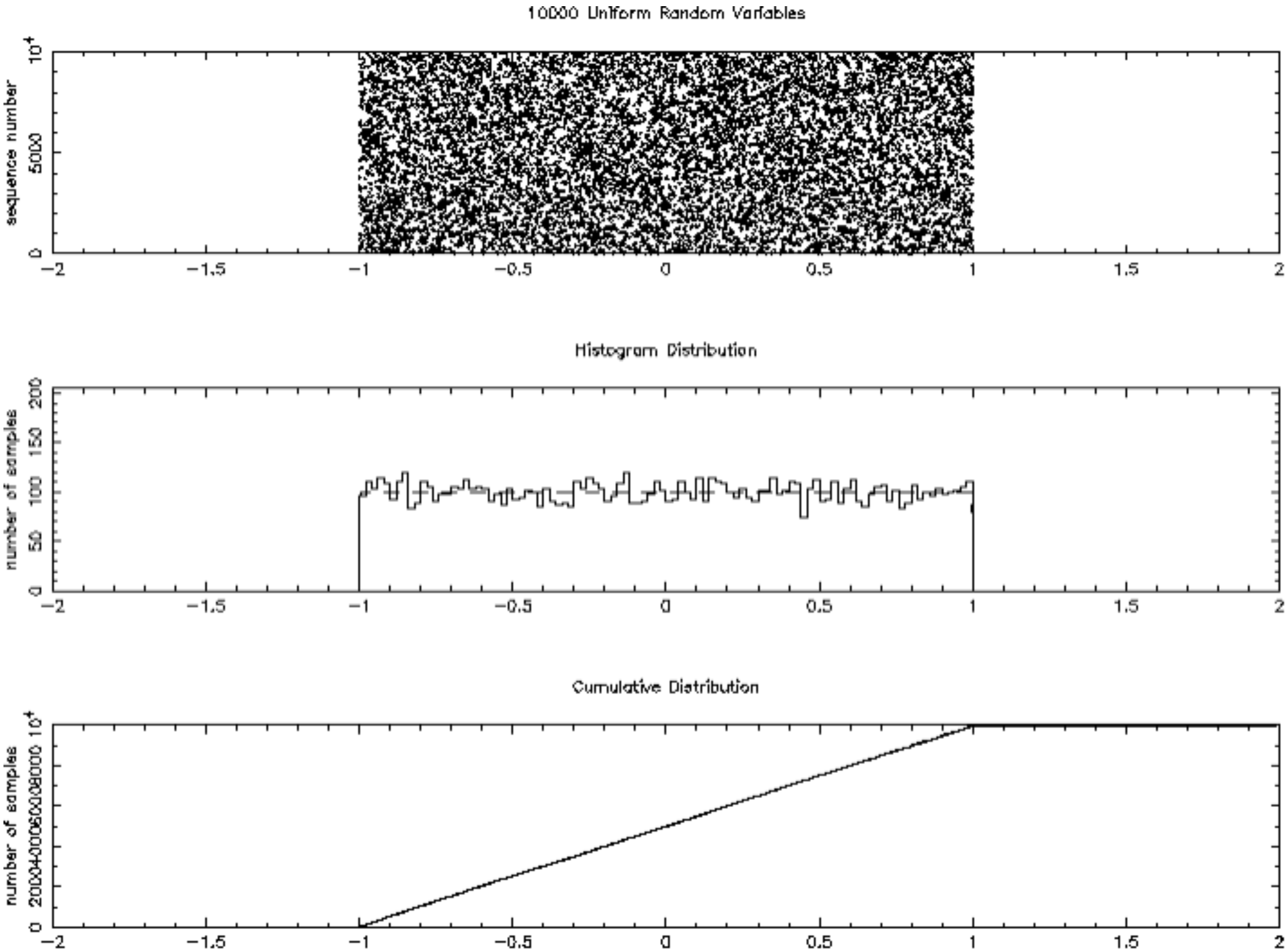
$$f(x) = 0 \qquad \text{otherwise.}$$



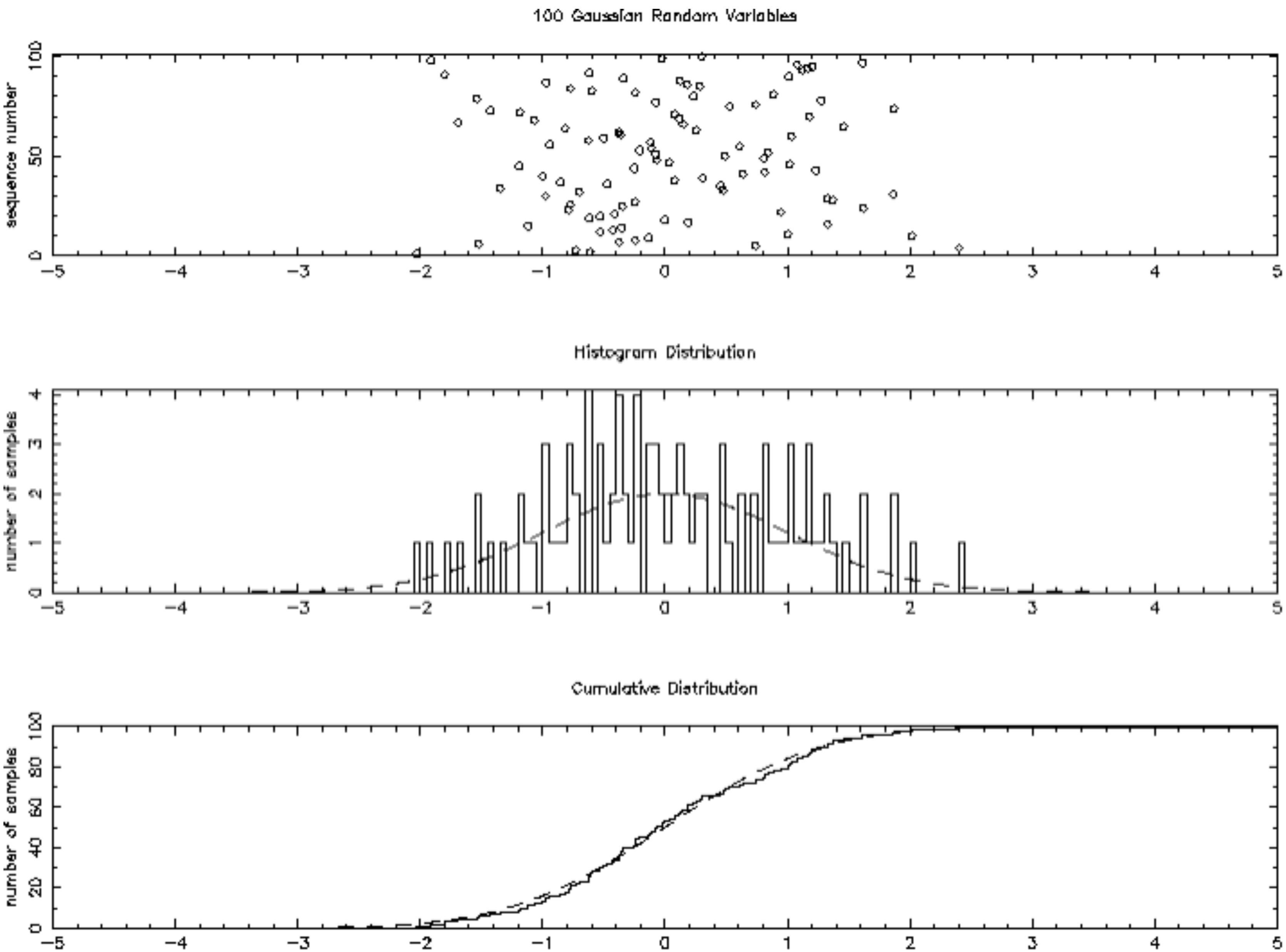

100 Uniform Random Variables
sequence number
Histogram Distribution
number of samples
Cumulative Distribution
number of samples

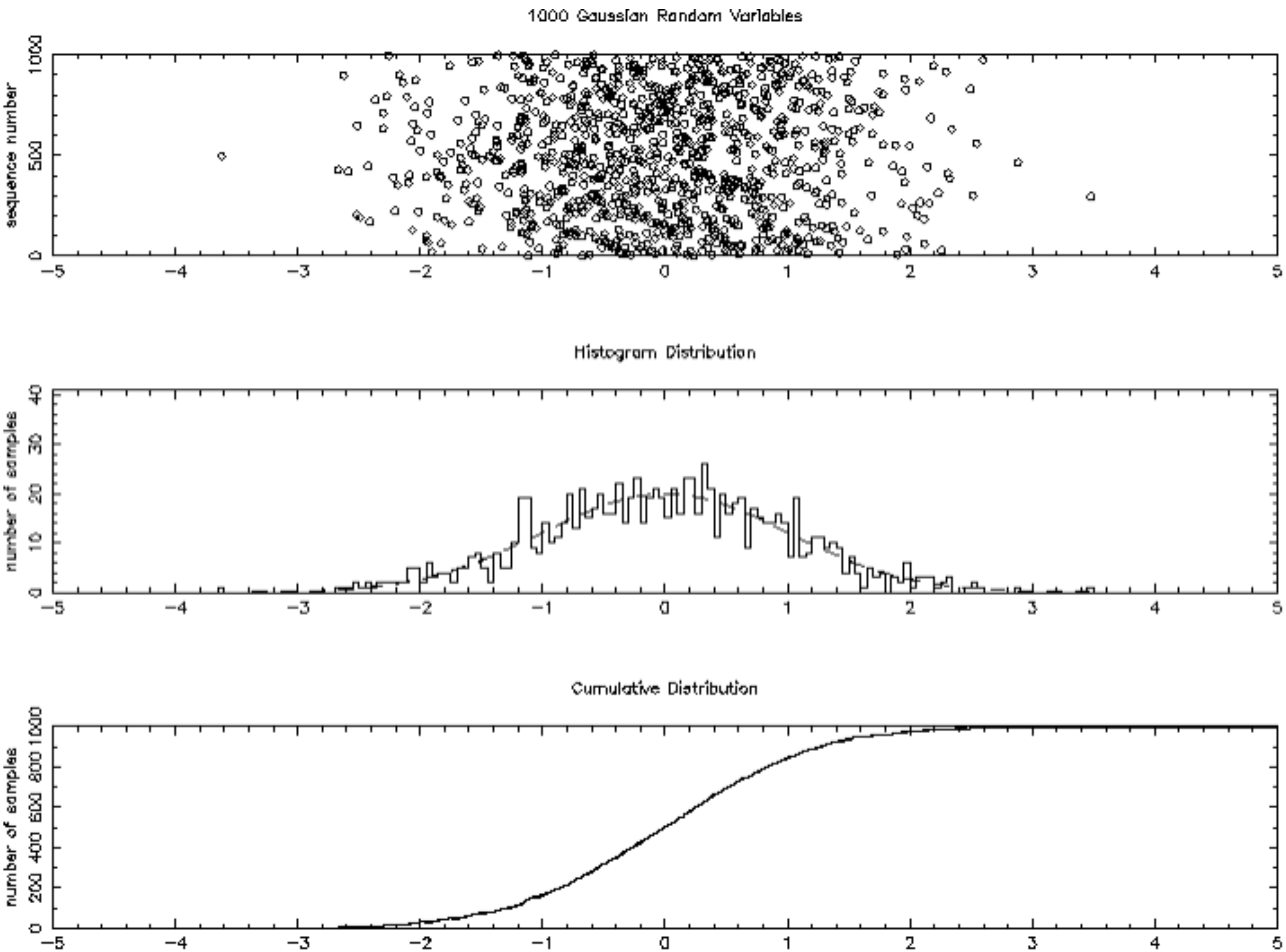


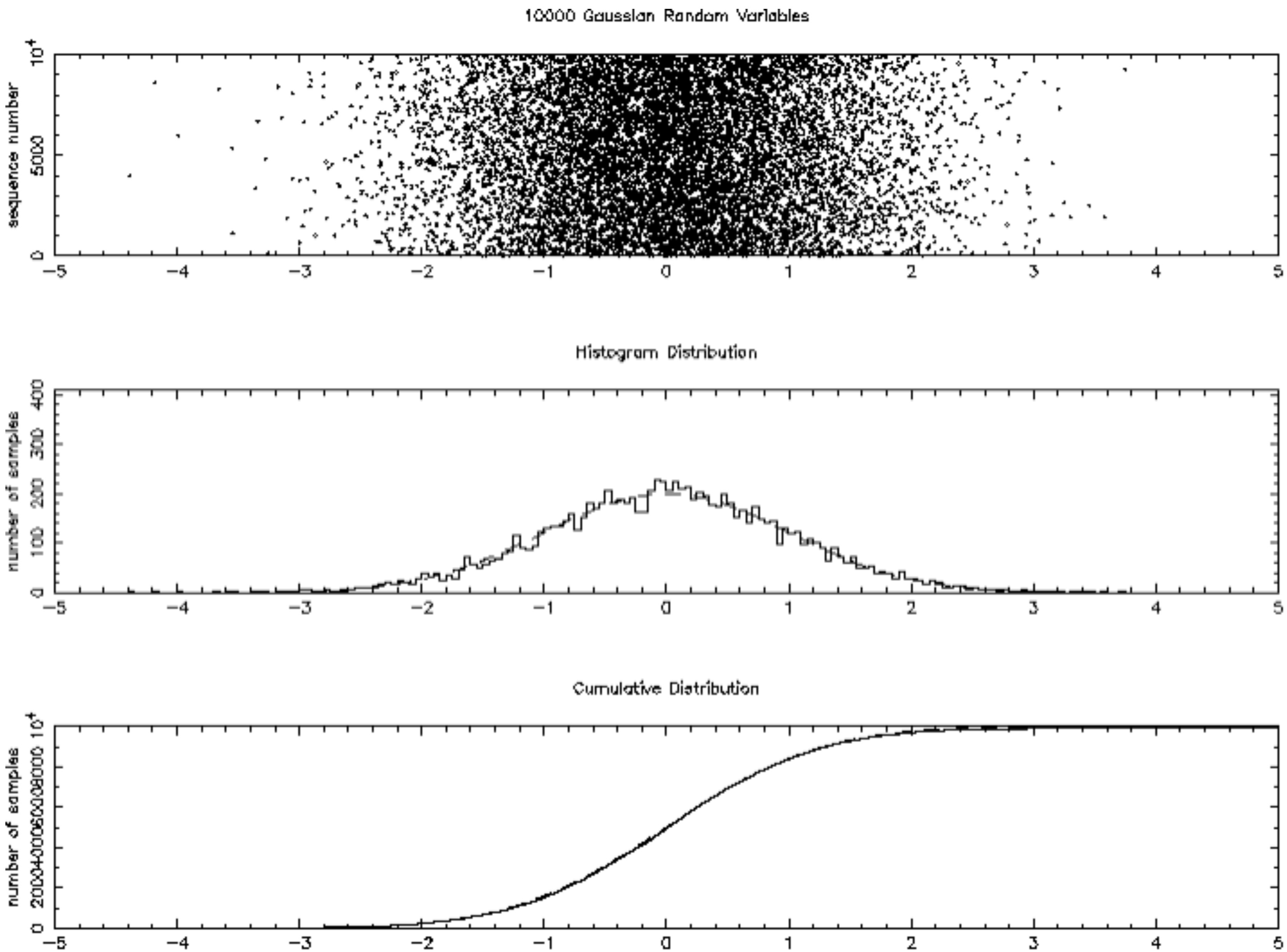
1000 Uniform Random Variables
sequence number
Histogram Distribution
number of samples
Cumulative Distribution
number of samples



Note that the histograms converge to *f(x)* and *F(x).*


100 Gaussian Random Variables
sequence number
Histogram Distribution
number of samples
Cumulative Distribution
number of samples
-5
-4
-3
-2
-1
0
1
2
3
4
5


1000 Gaussian Random Variables
sequence number
0
500
1000
-5
-4
-3
-2
-1
0
1
2
3
4
5
Histogram Distribution
number of samples
0
10
20
30
40
Cumulative Distribution
number of samples
0
200
400
600
800
1000



Note that the histograms converge to *f(x)* and *F(x).*

Moments of Distributions

- The **moments** of a distribution characterise its **location**, **width** and **shape**.

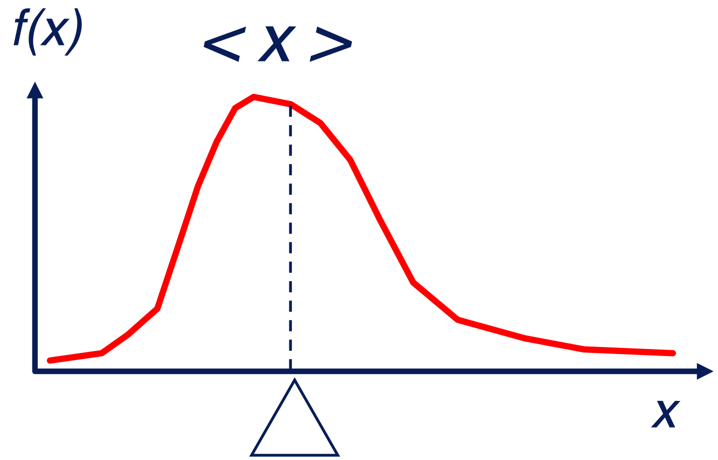
- Strong physical analogy with moments in mechanics of rigid bodies:
  - Centre of mass = first moment
  - Moment of inertia = second (central) moment
  - Higher moments => info on the shape of the distribution

Location measures: Mode, Mean and Median

- ***Mode*** (highest probability density)
- ***Mean*** (centre of mass)

= probability-weighted average of *x*

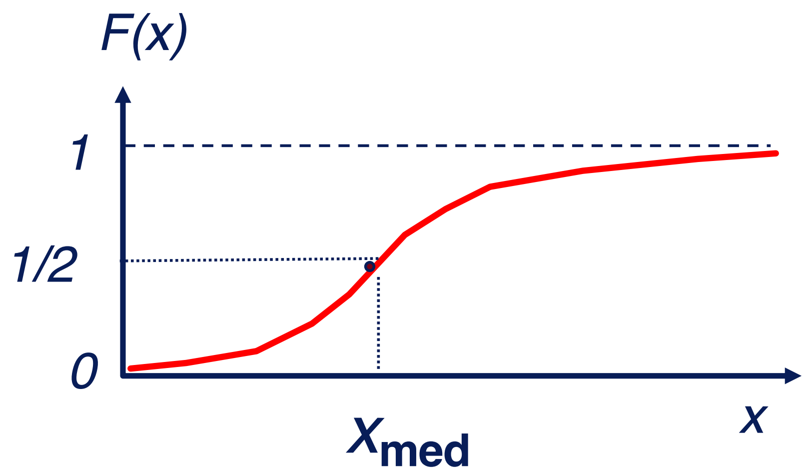
$$\langle x \rangle \equiv \int f(x)\; x\; dx$$



- ***Median*** (50th percentile)

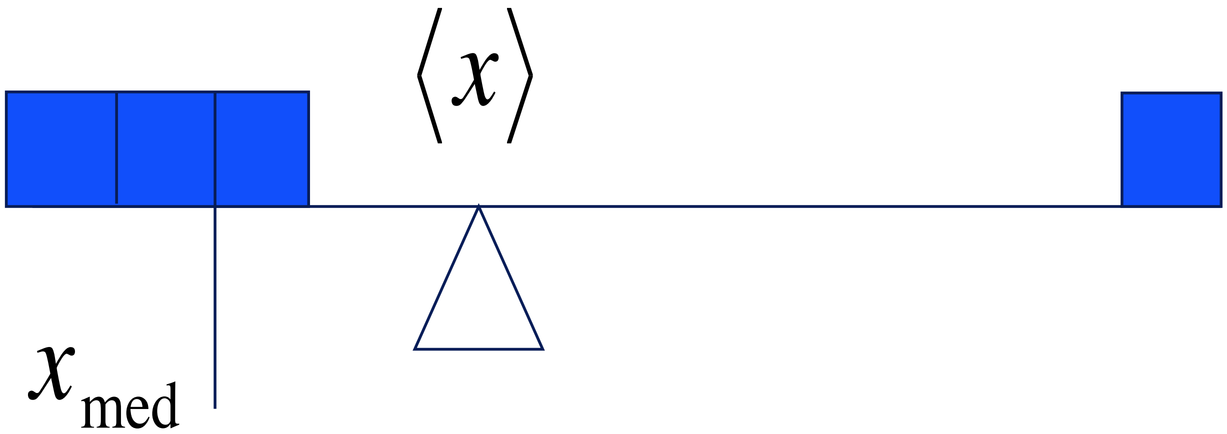
$$F(x_{\mathrm{med}}) \equiv \frac{1}{2}$$

$$P(x < x_{\mathrm{med}}) = P(x > x_{\mathrm{med}})$$



Mean vs Median

- **Median** is less sensitive to the long wings of a distribution -- the outliers.



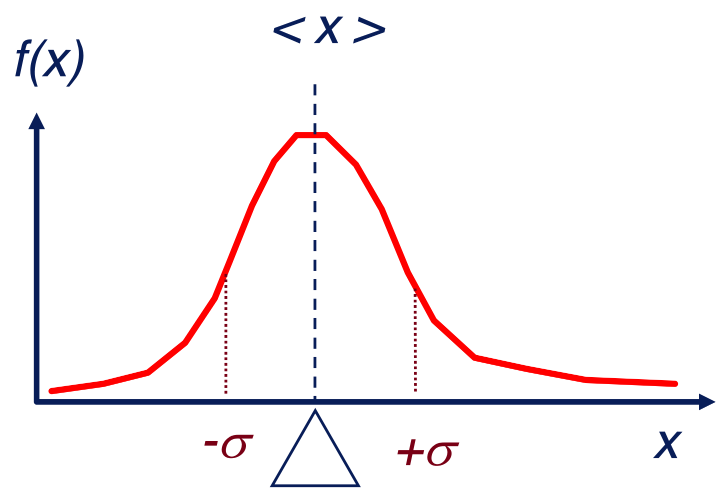
Width Measures: Standard Deviation, MAD

- **Standard deviation** σ measures **width** of distribution.

- **Variance** σ^2 (moment of inertia)

$$\sigma^2(x) = \sigma_x^2 = \mathrm{Var}(x) \equiv \left\langle [x - \langle x \rangle]^2 \right\rangle$$

$$= \int f(x)[x - \langle x \rangle]^2 \, dx$$



Mean Absolute Deviation (MAD):

$$\mathrm{MAD} \equiv \left\langle \left| x - x_{\mathrm{med}} \right| \right\rangle$$

Shape : Higher-order (Central) Moments

- General form: $m_n \equiv \left\langle \left[\frac{x - \langle x \rangle}{\sigma} \right]^n \right\rangle$

(n th central moment in units of σ^n)

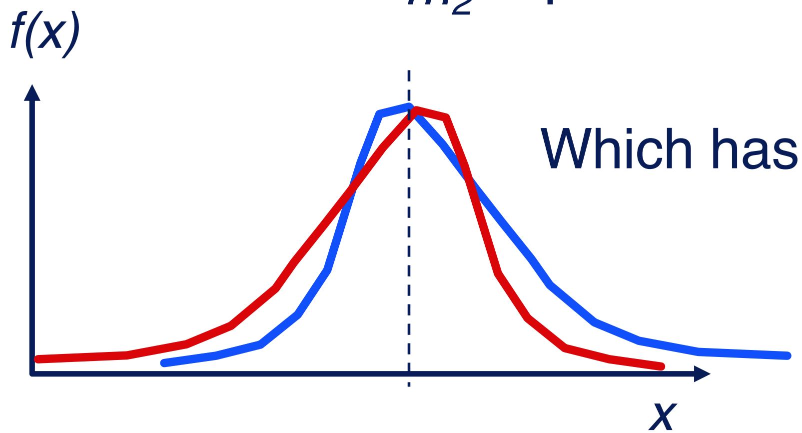
$m_1 = ?$

$m_2 = ?$

Higher central moments $n = 3, 4, \ldots$ define the **shape** of the distribution.

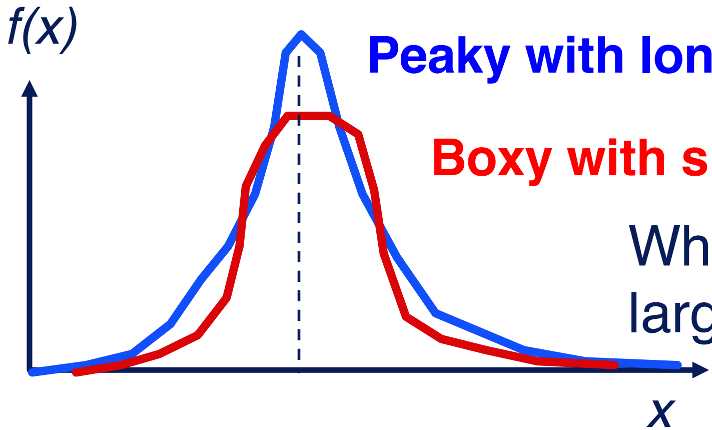
- **Skewness (m_3) :**

  (asymmetric tails)



Which has $m_3 > 0$?

- **Kurtosis (m_4) :**



Peaky with long tails

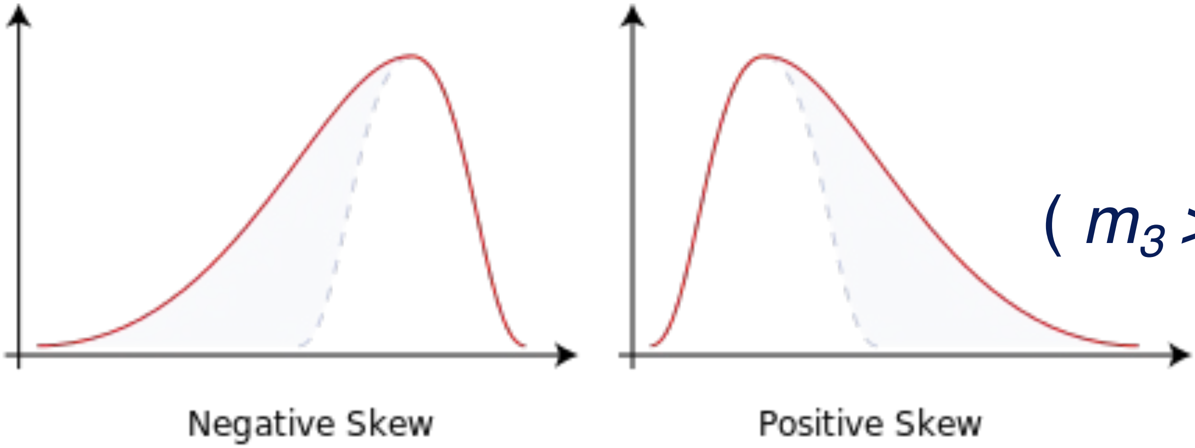
Boxy with short tails

Which has larger m_4 ?

If you know ***all*** the moments, you know the full shape.

Skew and Kurtosis

($m_3 < 0$)

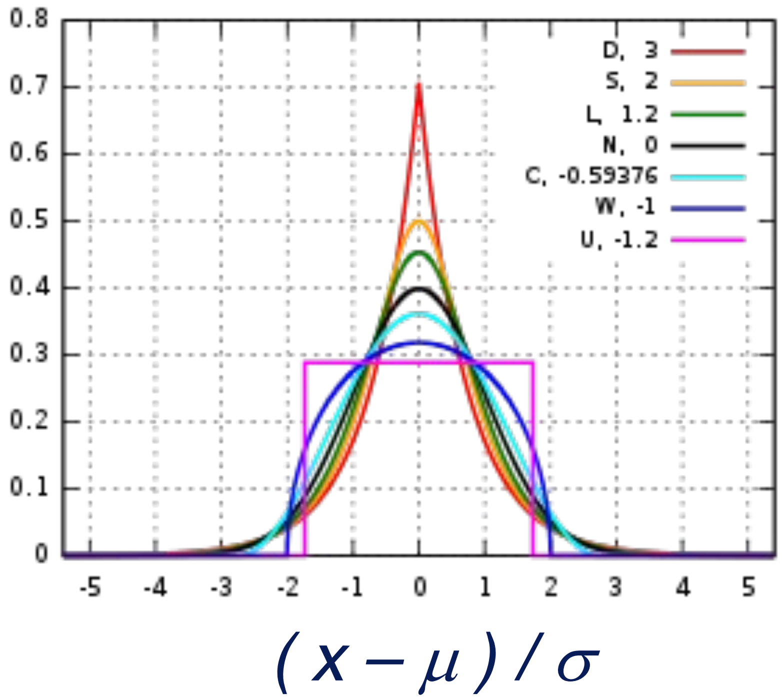


($m_3 > 0$)

"Leptokurtic"
($m_4 > 3$) with longer tails, like a kangaroo (leaps)

"Mesokurtic"
($m_4 = 3$) like a Gaussian.

"Platykurtic"
($m_4 < 3$) with shorter tails, like a platypus.



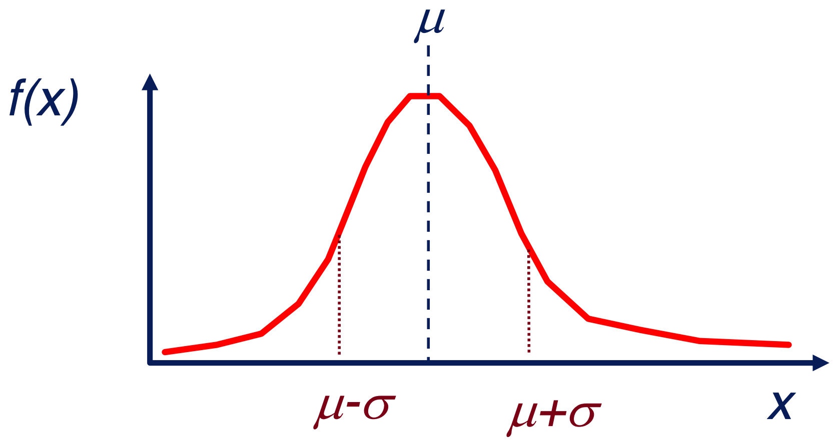
$m_4 > 3$ increases peak and wings relative to a Gaussian

Excess Kurtosis ($m_4 - 3$) defined relative to the kurtosis of a Gaussian.

Gaussian Distribution $G(\mu, \sigma^2)$

- Also known as a **Normal** distribution. $N(\mu, \sigma^2)$
- Physical example: thermal Doppler broadening

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



- 2 parameters:
- Mean (expected) value:
- $E(x) = \langle x \rangle = \mu$
- Variance: $Var(x) = \sigma^2(x) = \sigma^2$
- Standard deviation (dispersion) σ
- Full width at half maximum (FWHM)

$$\text{FWHM} = \sqrt{8\ln 2}\,\sigma \approx 2.3\sigma$$

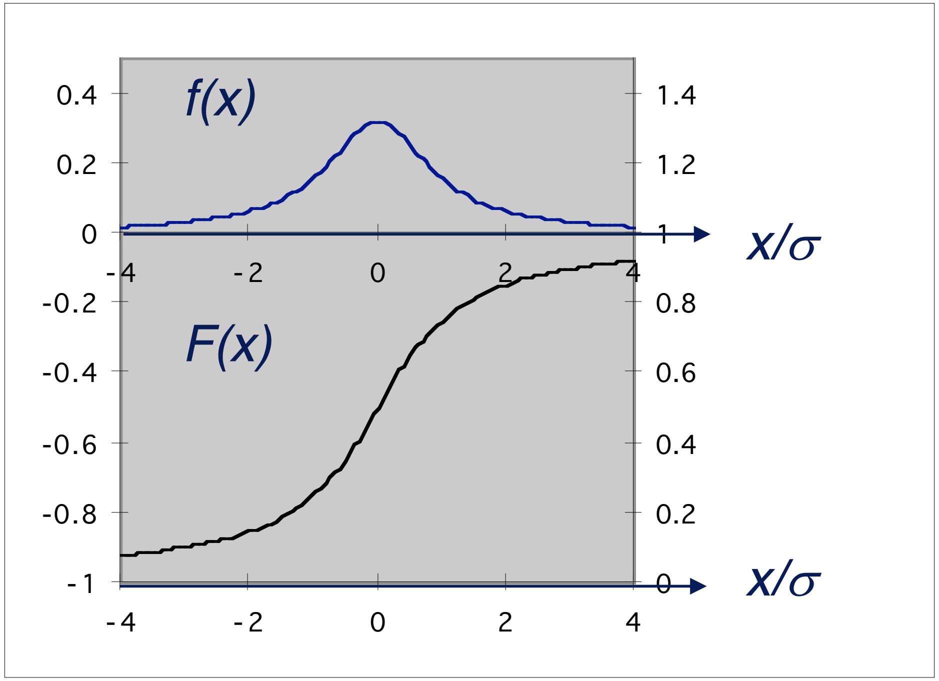
- 32% probability that x is outside $\mu \pm \sigma$
- 4.5% for x outside $\mu \pm 2\sigma$
- 0.3% for x outside $\mu \pm 3\sigma$

Lorentzian (Cauchy) Distribution $L(\mu,\sigma)$

- Peak at $x = \mu$, HWHM $= \sigma$.
- Physical example: damping wings of spectral lines.

$$f(x) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (x-\mu)^2}$$

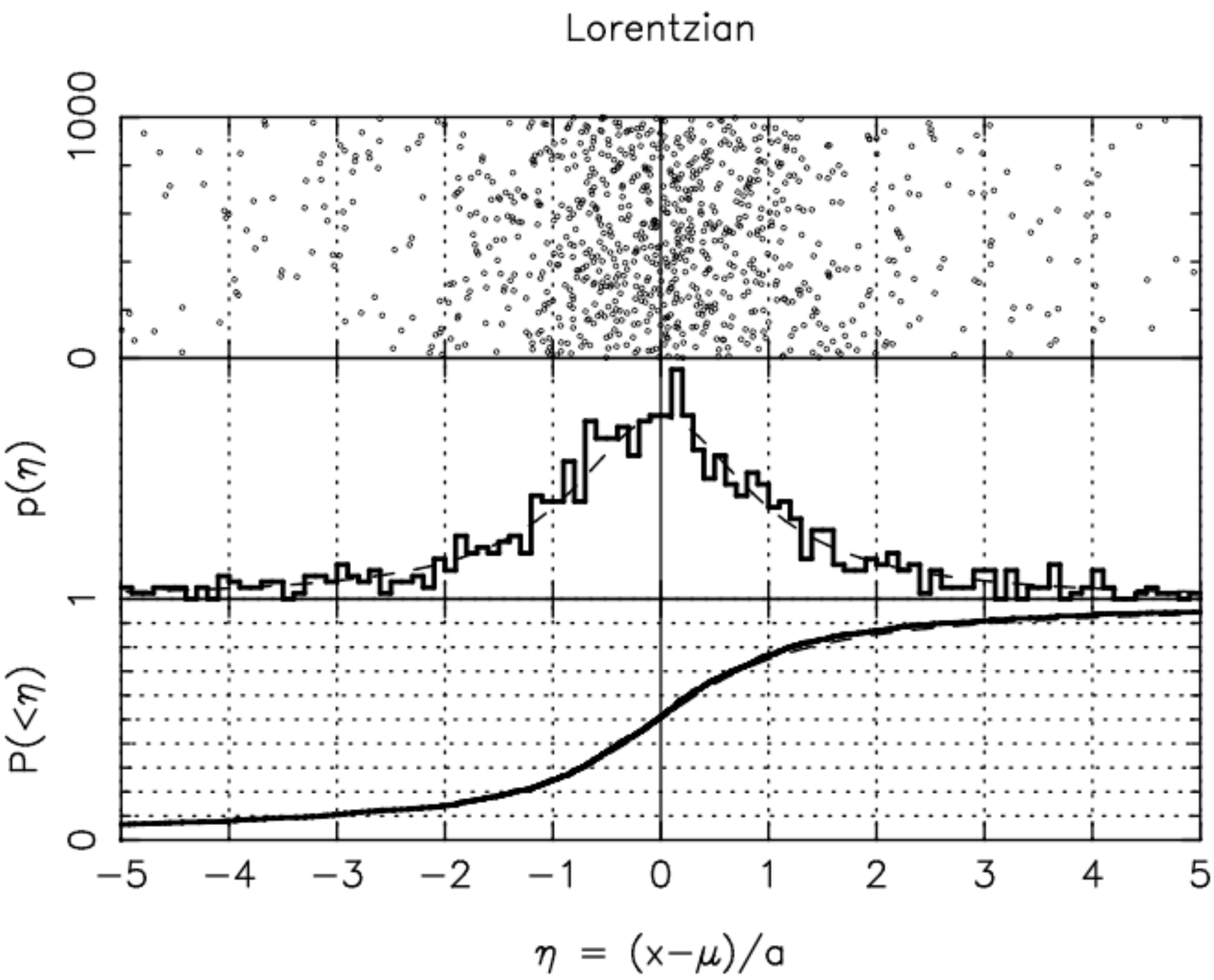
$$F(x) = \frac{1}{\pi} \tan^{-1}\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}$$



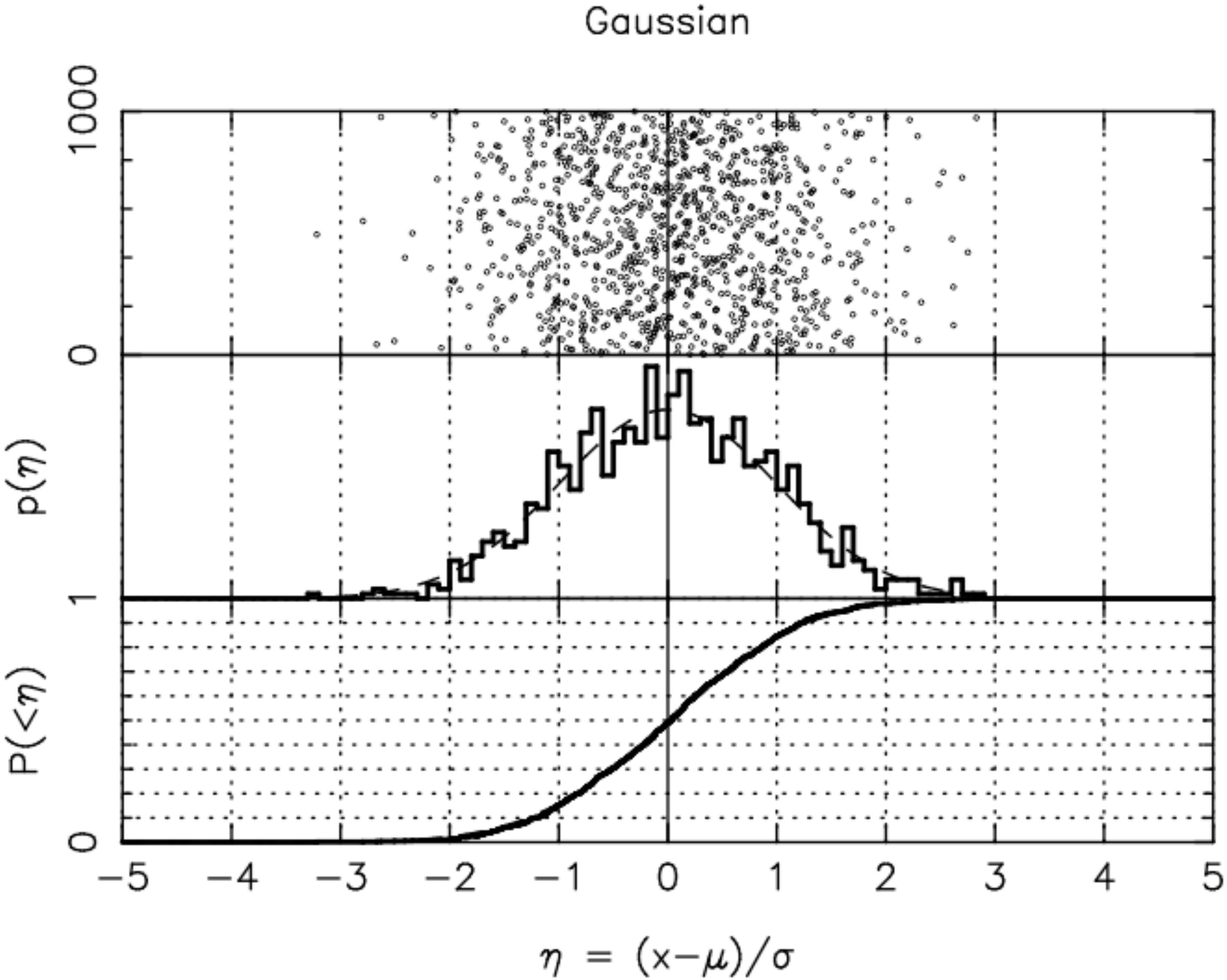
- Pathological: wings so broad that all moments diverge! ☹

$$\langle x \rangle \equiv \frac{\sigma}{\pi} \int_{-\infty}^{\infty} \frac{x\,\mathrm{d}x}{\sigma^2 + (x-\mu)^2} \propto \ln\left(\left|1 + x^2\right|\right)\Big|_{-\infty}^{\infty} = \infty - \infty$$

$$\left\langle (x-\mu)^2 \right\rangle = \infty$$



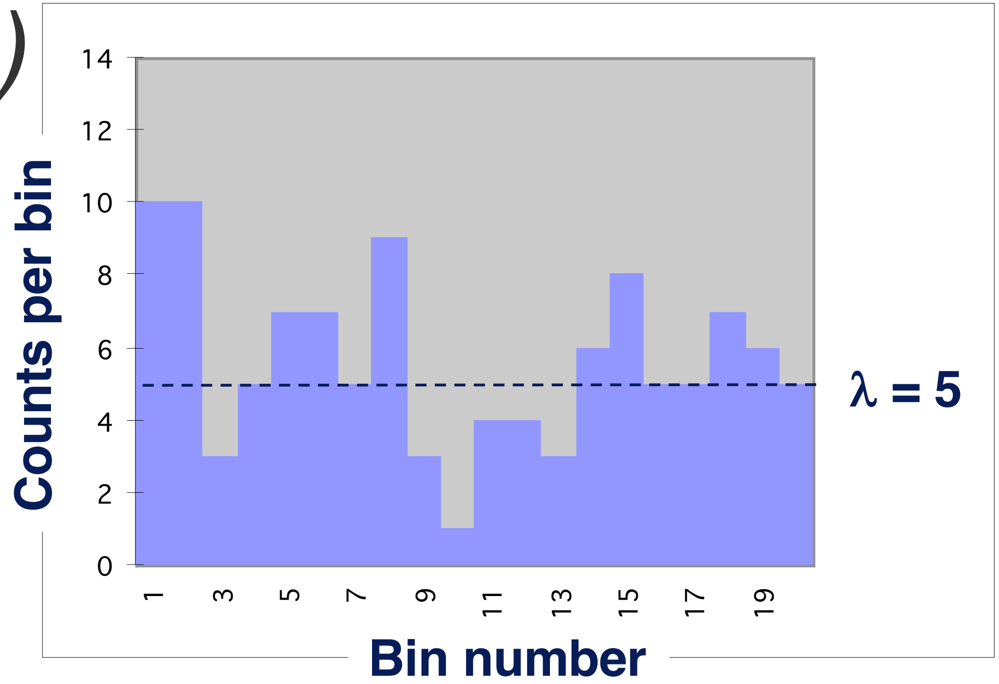
Lorentzian
1000
0
p(η)
1
P(<η)
0
−5
−4
−3
−2
−1
0
1
2
3
4
5
η = (x−μ)/a

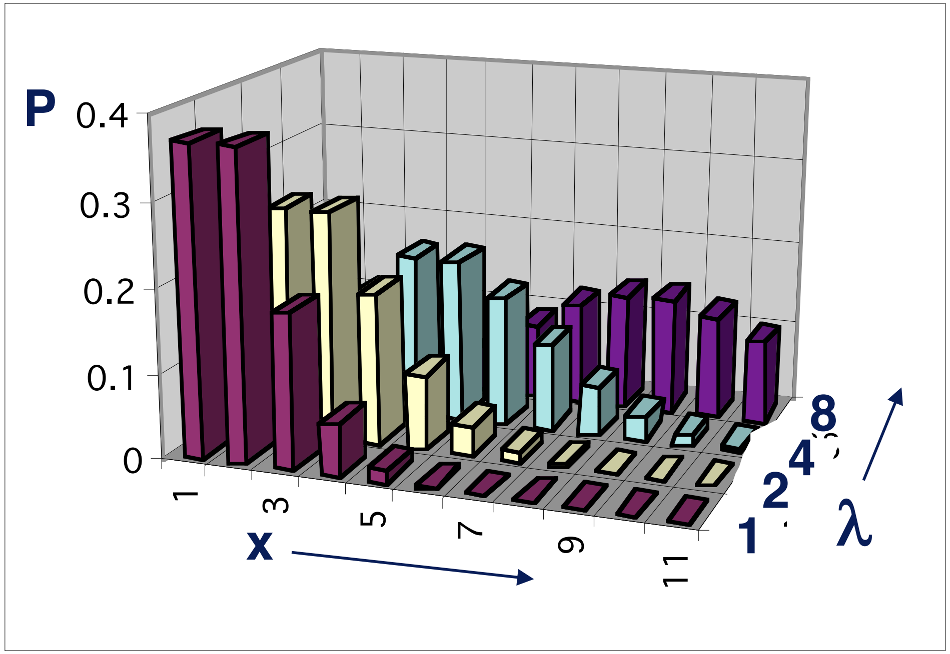


Gaussian
1000
0
1
0
p(η)
P(<η)
−5
−4
−3
−2
−1
0
1
2
3
4
5
η = (x−μ)/σ

Poisson Distribution $P(\lambda)$

- A discrete distribution
- Describes counting statistics:
  - Raindrops in bucket per time interval
  - Photons per pixel during exposure
- **λ = mean count rate**
  - **Not necessarily an integer !**





$$f(x) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \delta(x-n)$$

$$P(x=n) = e^{-\lambda} \frac{\lambda^n}{n!} \qquad n = 0,1,2,\ldots$$

$$\langle x \rangle = \lambda$$

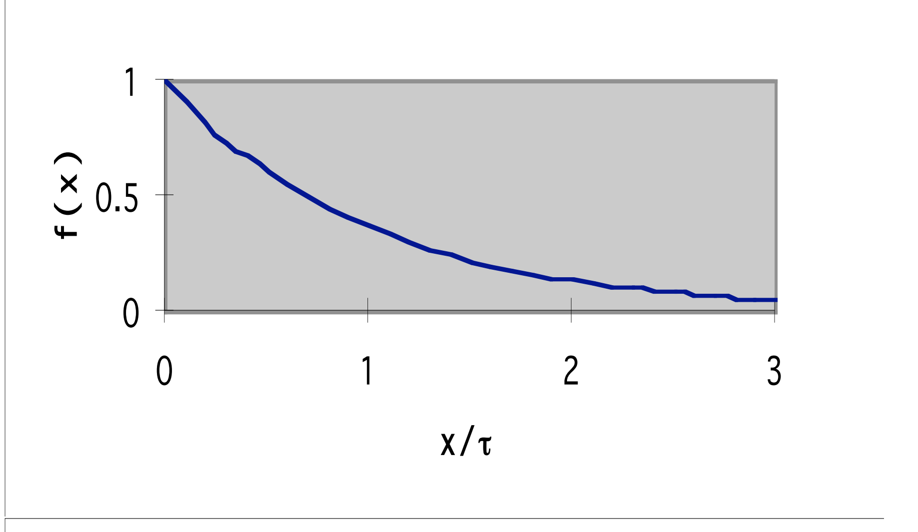
$$\sigma^2(x) = \lambda \Rightarrow \boxed{\sigma(x) = \sqrt{\langle x \rangle}}$$

Exponential Distribution *E(τ)*

- Distribution of time intervals between random events
  - Raindrops, photons, radioactive decays, lightbulbs burning out, etc.

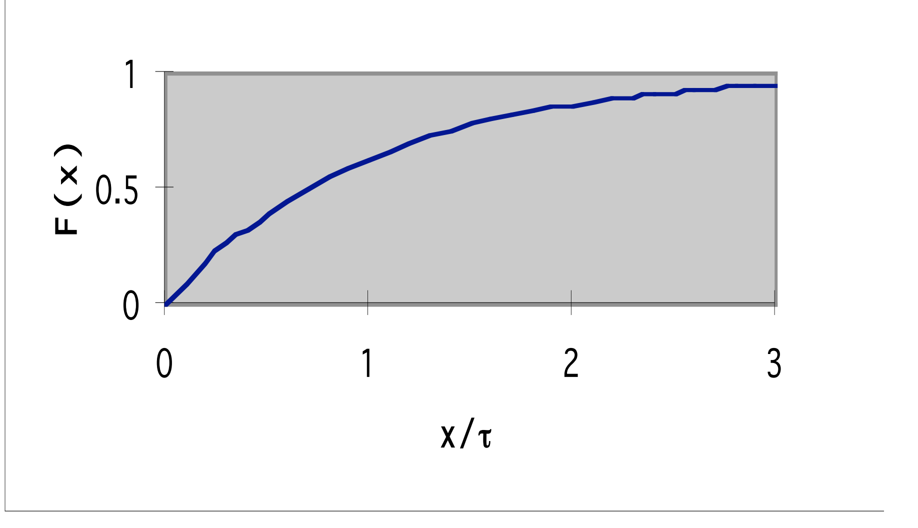
$$f(x) = \frac{1}{\tau} e^{-x/\tau}$$

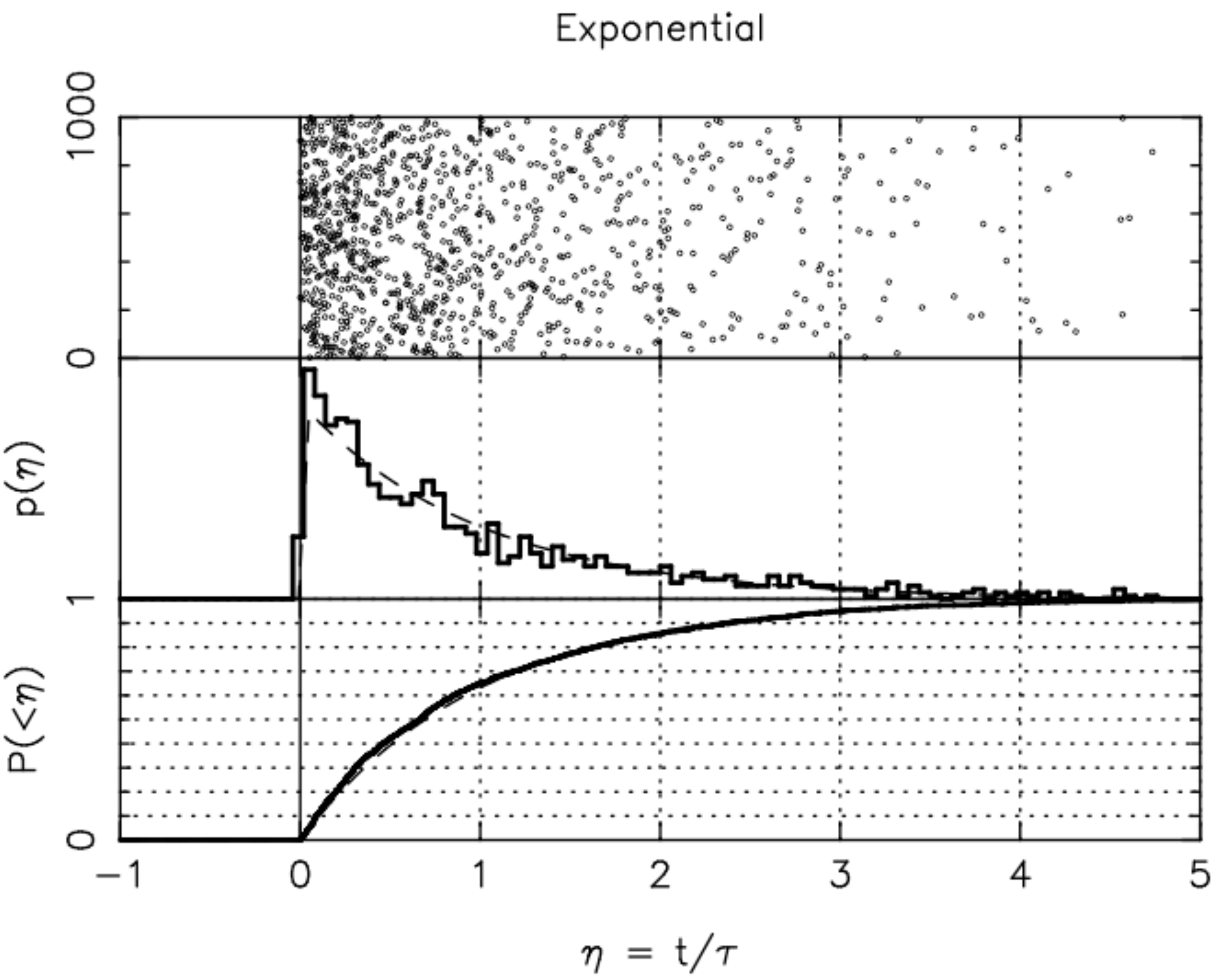
$$F(x) = 1 - e^{-x/\tau}$$



$$\langle x \rangle = \tau = \text{ mean time between events}$$

$$\text{Var}[x] = \langle (x-\tau)^2 \rangle = \tau^2$$





Exponential
1000
0
p(η)
1
P(<η)
0
−1
0
1
2
3
4
5
η = t/τ

Chi-Squared Distribution χ^2_N

- Sum of squares of *N* independent Gaussian random variables

$\chi^2_N \equiv$ Chi-Squared with N degrees of freedom

X and Y are independent Gaussian random variables.

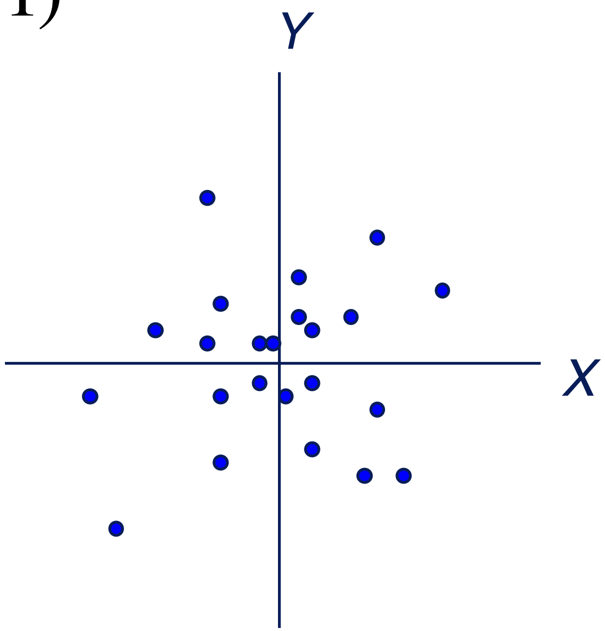
i.e. $X \sim G(0,1)$ $\quad Y \sim G(0,1)$

then $X^2 \sim \chi^2_1$ $\quad Y^2 \sim \chi^2_1$

$$X^2 + Y^2 \sim \chi^2_2$$

and so on for each new degree of freedom:

$$\chi^2_N + \chi^2_M \sim \chi^2_{N+M}$$



Chi-Squared = "Badness of Fit"

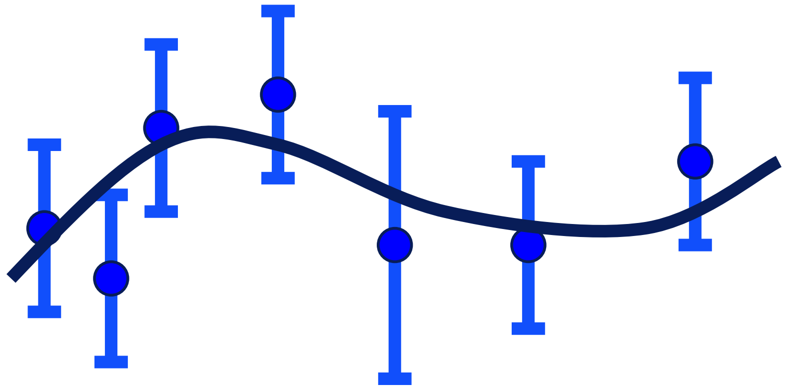
$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{D_i - \mu_i(\alpha)}{\sigma_i} \right)^2 \sim \chi^2_{N-P}$$

D_i = data value

σ_i = 1 - σ error bar

$\mu_i(\alpha)$ = model predicted data value

α = parameters of the model



N = number of data points

P = number of fitted parameters

$N - P$ = degrees of freedom

χ^2 distribution with *N* degrees of freedom

$$f(x) = \frac{1}{\Gamma(N/2)\, 2^{N/2}}\, x^{(N/2-1)}\, e^{-x/2}$$

$$\Gamma(1) = 1 \qquad\qquad \Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(n) = (n-1)! \qquad \Gamma(x+1) = x\,\Gamma(x)$$

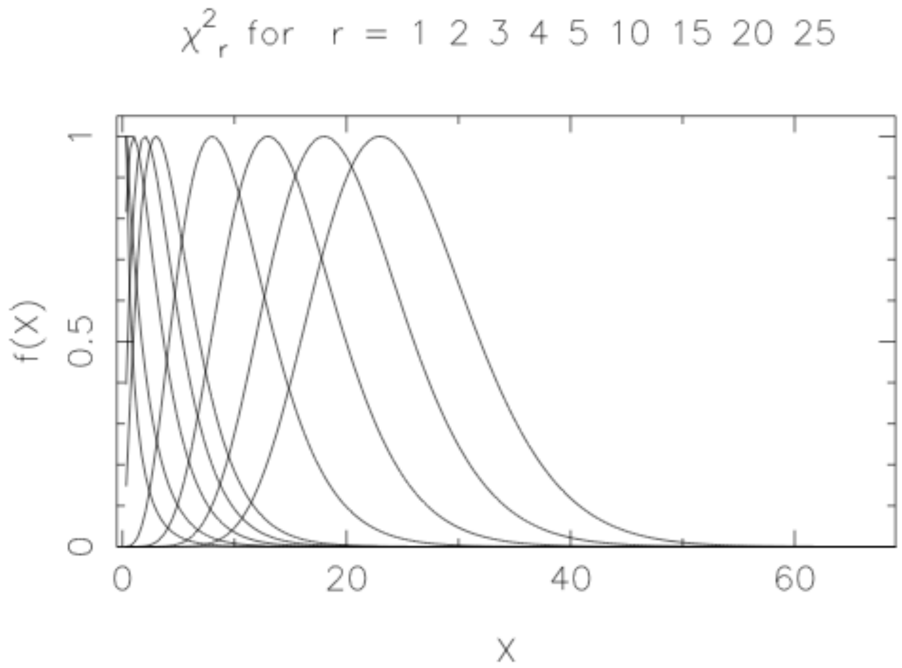
$$e.g. \quad \Gamma(3/2) = (1/2)\,\Gamma(1/2) = \sqrt{\pi}\,/\,2$$

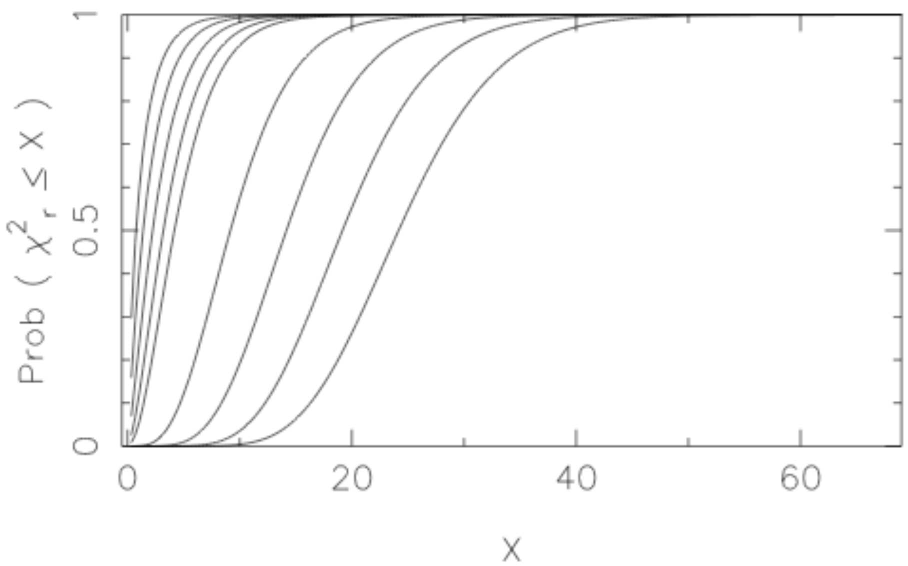
$$\chi_1^2 : \quad f(x) = \left(\frac{e^{-x}}{2\,\pi\,x} \right)^{1/2}$$

$$\chi_2^2 : \quad f(x) = \frac{1}{2} e^{-x/2}$$

$$\left\langle \chi_N^2 \right\rangle = N$$

$$\sigma^2\left(\chi_N^2\right) = 2N$$





χ^2_N and reduced χ^2_N distribution

- Sum of squares of *N* independent Gaussian random variables

$\chi^2_N = \text{chi} - \text{squared}$

with N degrees of freedom

$$\left\langle \chi^2_N \right\rangle = N$$

$$\sigma^2\left(\chi^2_N\right) = 2N$$

$$\sigma\left(\chi^2_N\right) = \sqrt{2N}$$

Reduced χ^2_N

$$\left\langle \frac{\chi^2_N}{N} \right\rangle = 1$$

$$\sigma^2\left(\frac{\chi^2_N}{N}\right) = \frac{2}{N}$$

$$\sigma\left(\frac{\chi^2_N}{N}\right) = \sqrt{\frac{2}{N}}$$