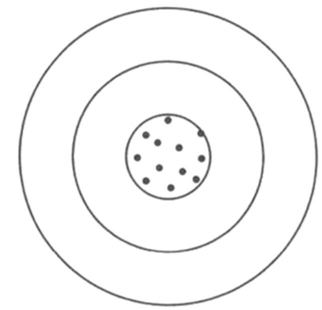
Lab 3a: Distribution functions

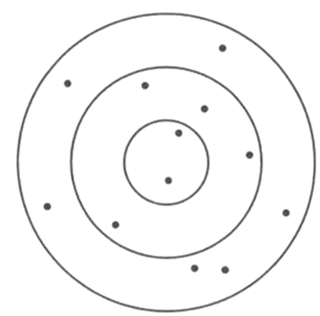
Outline

- Histogram
- Basic concepts
- Gaussian (normal) distribution
  - Limiting distribution
- Poisson distribution
  - Counting measurements



Random: small
Systematic: small

(a)

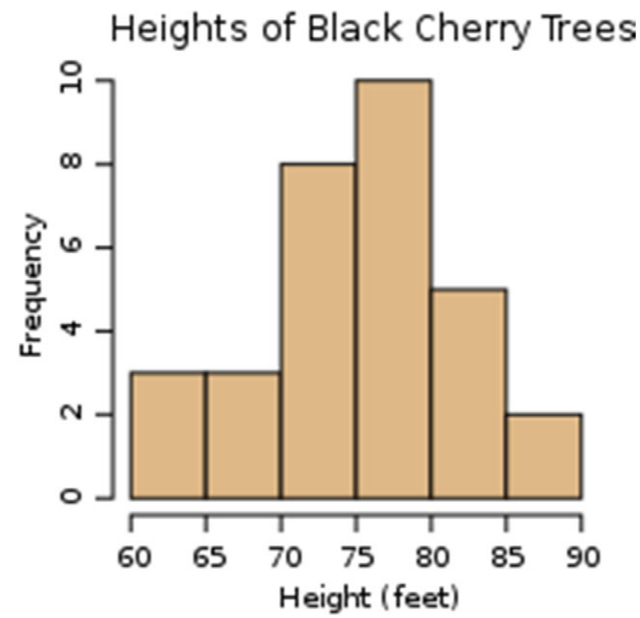


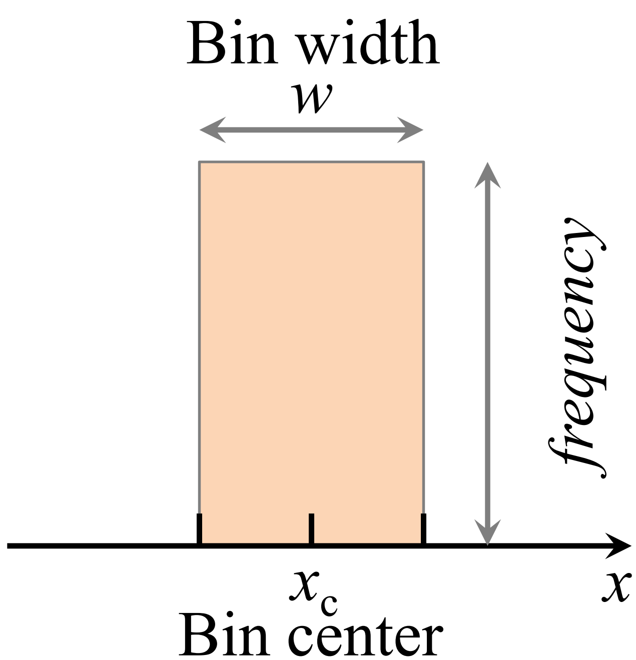
Random: large
Systematic: small

(c)

Histogram and bin

A **histogram** is a graphical representation of the distribution of data. It is a representation of tabulated frequencies, shown as adjacent rectangles (**bin**s), with an area equal to the frequency of the observations in the interval.



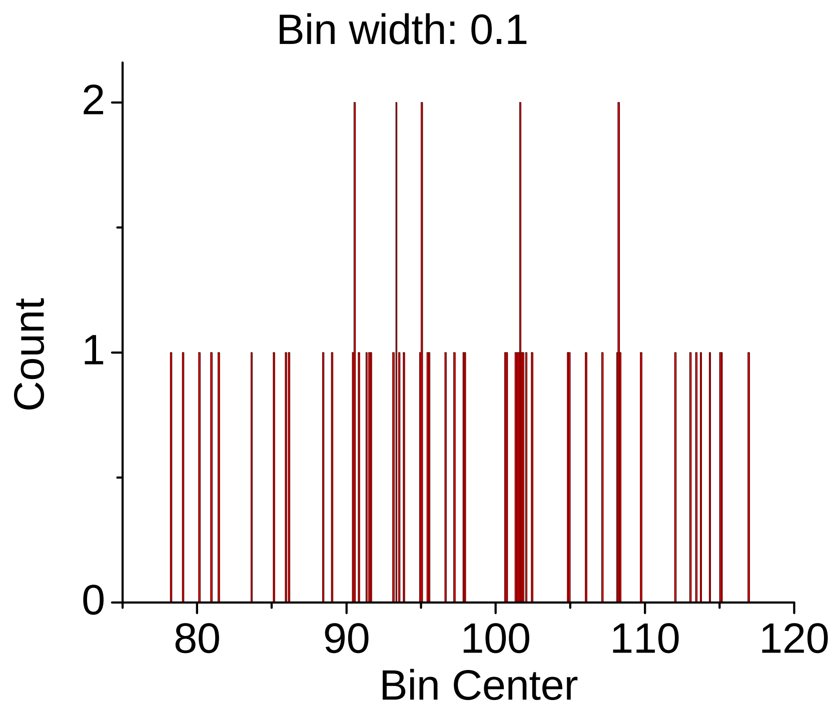


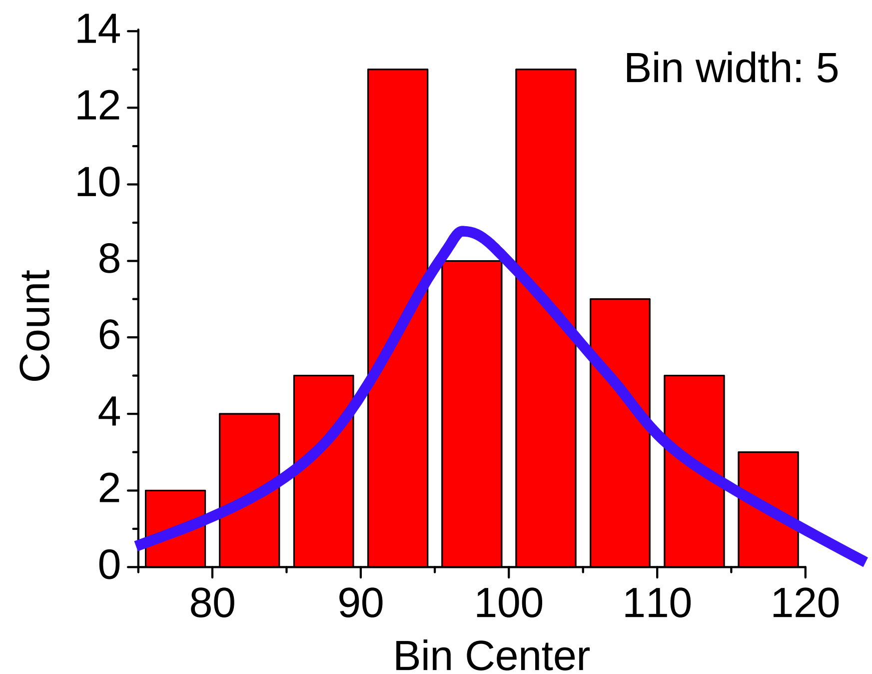
$$\text{Bin start} = x_c - w/2$$

$$\text{Bin end} = x_c + w/2$$

$$\text{Next Bin center} = x_c + w$$

Choose bin width properly





There is no "best" bin width. One commonly used "rule" (of # of bins) is:

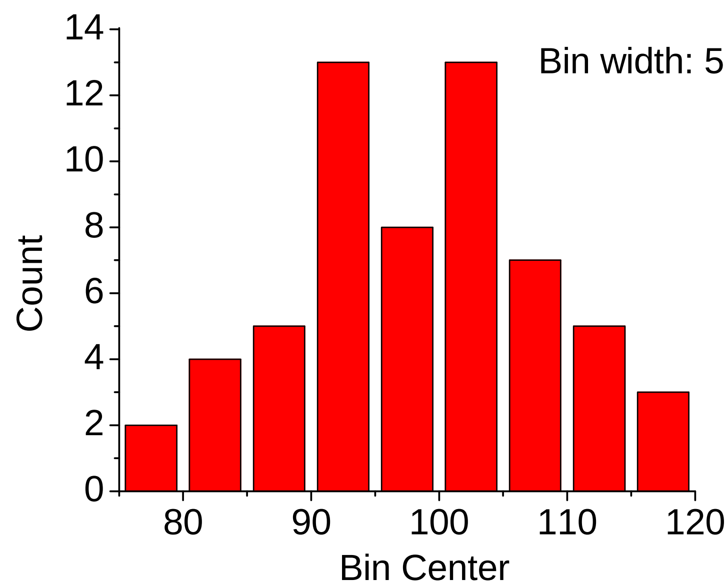
$$\# \text{ of bins} = \sqrt{\# \text{ of data points}}$$

Random variables

For a random variable x (e.g. measurements of a quantity), each measured value x_k occurs n_k times in N measurements.

the mean: $\bar{x} = \dfrac{1}{N}\sum_k x_k n_k$

$$N = \sum_k n_k$$



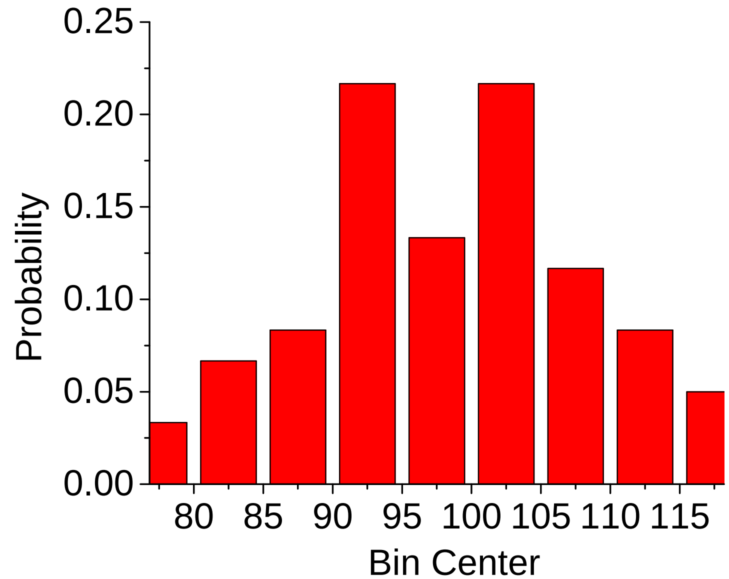
Probability distribution of N measurements

The fraction of N measurements that gave the result x_k (i.e. probability):

$$F_k = \frac{n_k}{N}$$

Normalization

$$\sum_k F_k = 1 \quad \left[N = \sum_k n_k \right]$$



F_k : the normalized probability of x.

The mean of x_k: $\quad \bar{x} = \sum_k x_k F_k$

True value and variance

For a random variable x (e.g. measurements of a quantity), the mean and variance of the limiting (a.k.a. parent) distribution ($N \to \infty$):

the true value: $$X = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_i$$

the true variance: $$\sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \left(x_i - X\right)^2$$

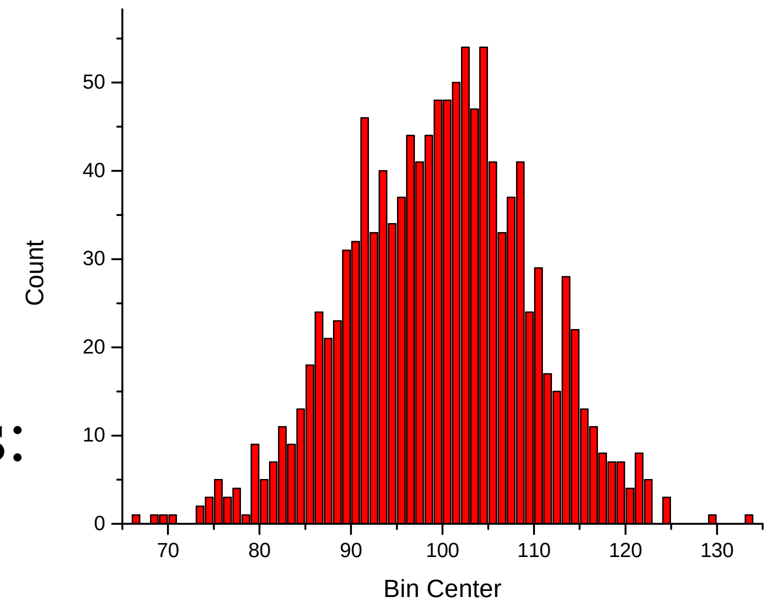
Limiting distributions

In the limit of $N \rightarrow \infty$, it is believed every measurement has a limiting distribution, which is often the Gaussian distribution. Also, discrete distributions approach a continuous function called probability distribution function (PDF):

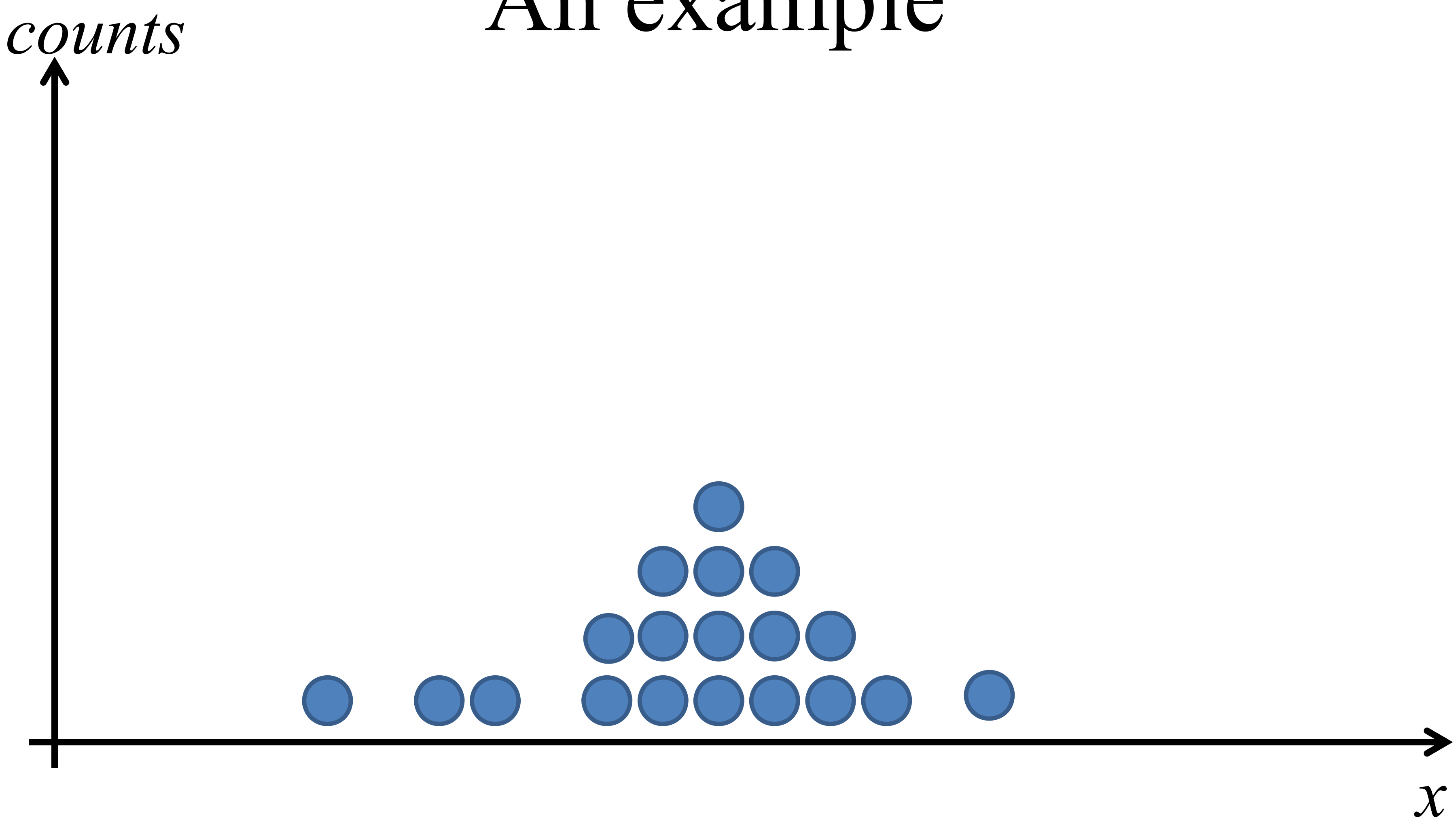
$$F_k \rightarrow f(x)\,dx$$

The probability that any one measurement will give an answer between x and $x+dx$ is:

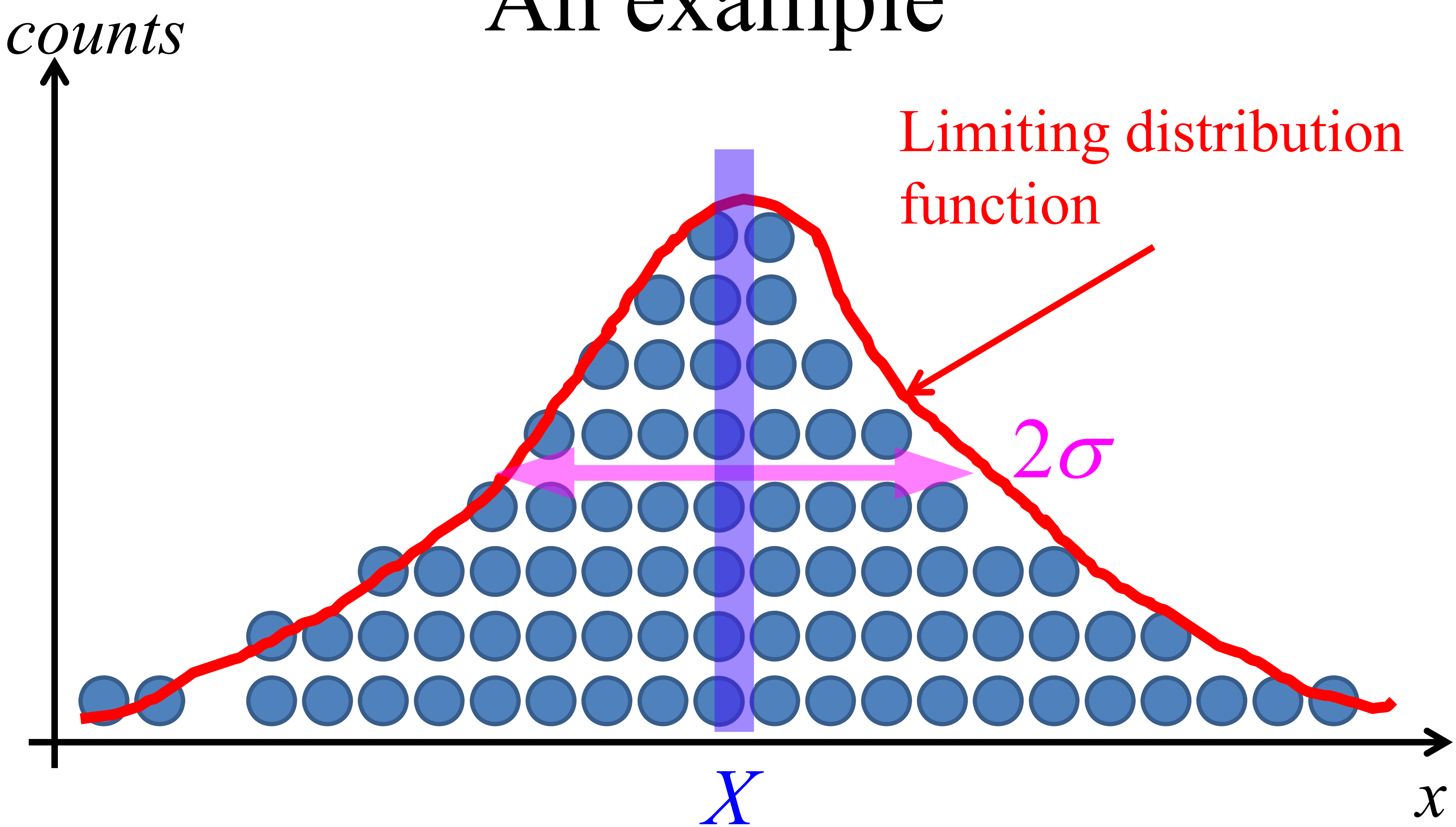
$$p(x) = f(x)\,dx$$



An example



An example



$$\left(\text{measured } x\right) = X \pm \sigma$$

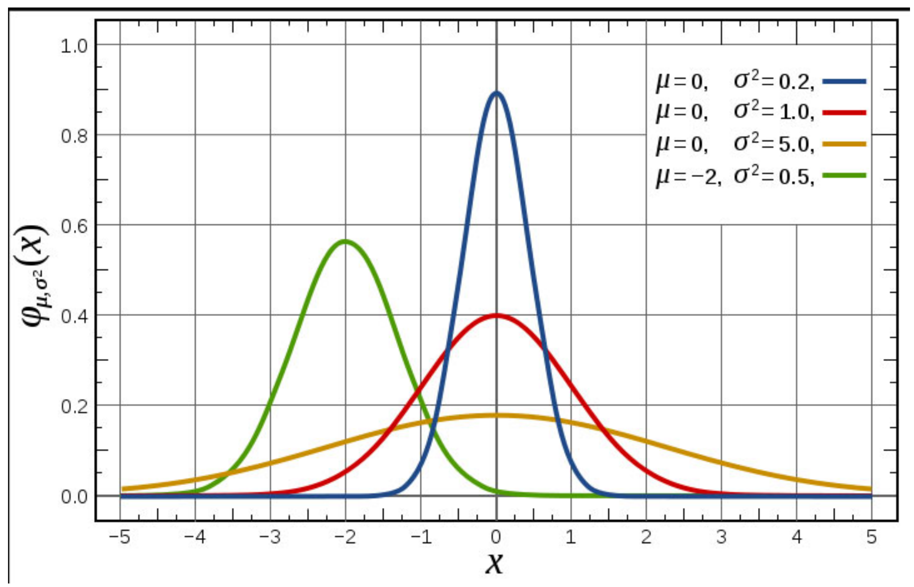
Probability distribution function (PDF)

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

The mean of x: $\bar{x} = \int_{-\infty}^{\infty} x f(x)\,dx$

The variance:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - X)^2 f(x)\,dx$$



http://en.wikipedia.org/wiki/Normal_distribution

Finite number of measurements

In reality, the number of measurements is finite, i.e. $N \neq \infty$. The average value is the best estimation of true mean value, while the standard deviation is the best estimation of true variance.

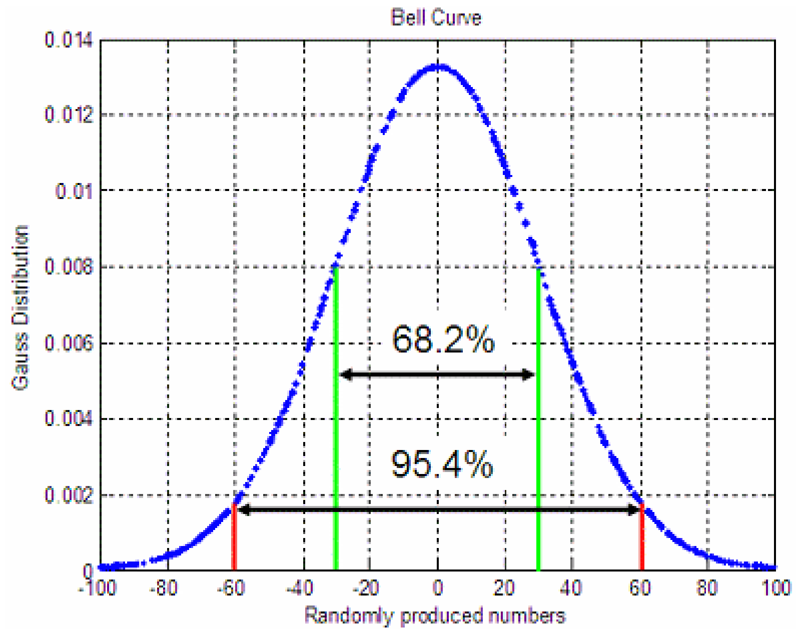
$$X \approx \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2 \approx \sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

These estimations will be “justified” in the next lecture.

The Gaussian (normal) distribution

$$G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$$



2 key parameters of a Gaussian PDF.: $\boxed{X, \sigma}$

Normalization: $$\int_{-\infty}^{\infty} G_{X,\sigma}(x)\,dx = 1$$

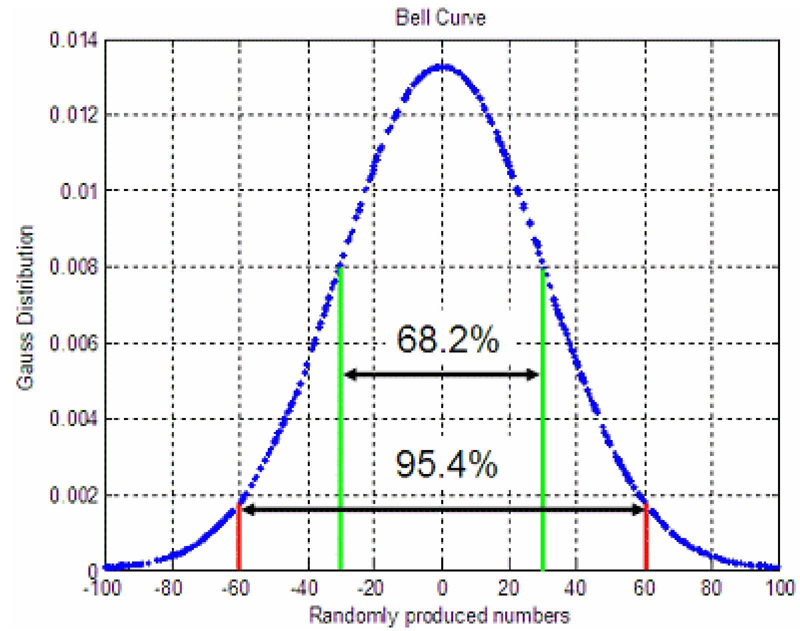
The mean: $$\bar{x} = X$$

The variance: $$\sigma_x^2 = \sigma^2$$

The Gaussian (normal) distribution

$$G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$$

The standard deviation: σ



$$\text{Prob}(\text{within} \pm \sigma) = \int_{X-\sigma}^{X+\sigma} G_{X,\sigma}(x)\,dx \approx 0.682$$

$$\text{Prob}(\text{within} \pm 2\sigma) = \int_{X-2\sigma}^{X+2\sigma} G_{X,\sigma}(x)\,dx \approx 0.954$$

$$\text{Prob}(\text{within} \pm 3\sigma) = \int_{X-3\sigma}^{X+3\sigma} G_{X,\sigma}(x)\,dx \approx 0.997$$

The Poisson distribution (I)

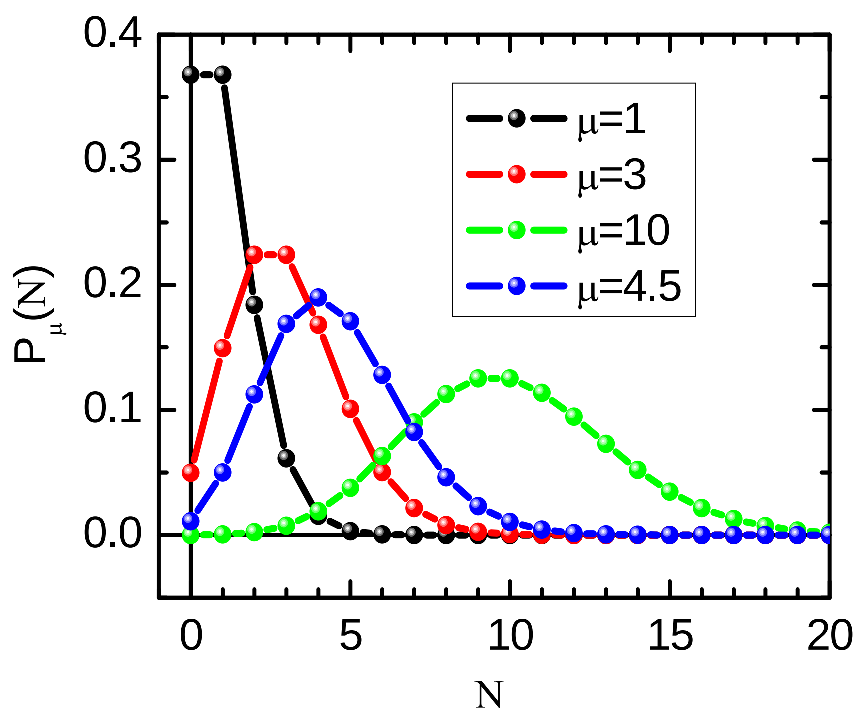
- Events with constant rate / probability.

e.g.: radioactive decay, birth count, etc.

- A limiting case of binomial distribution

The probability of N counts in a definite interval:

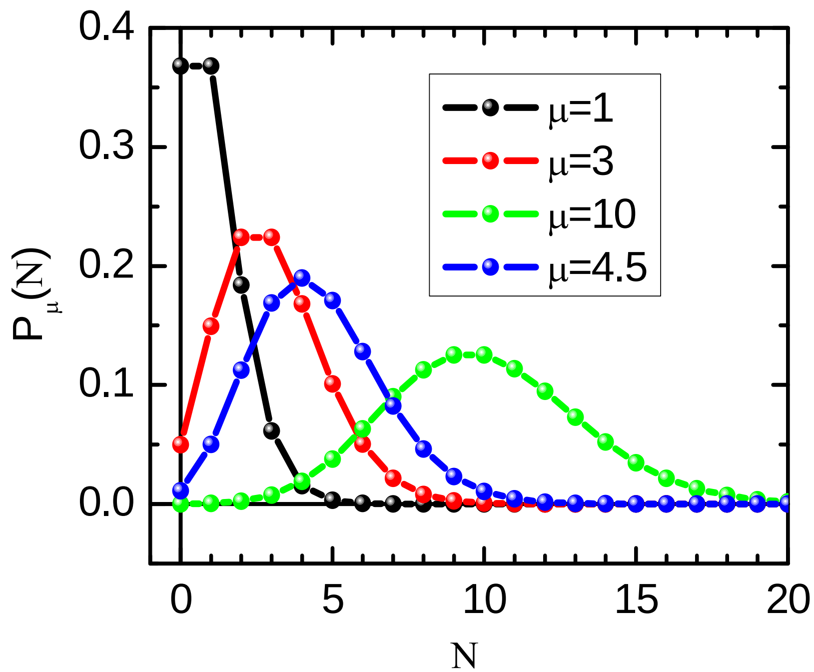
$$P_\mu(N) = e^{-\mu} \frac{\mu^N}{N!}$$



The Poisson distribution (II)

Normalization:

$$\sum_{\nu=0}^{\infty} P_\mu(N) = 1$$



the mean: $\bar{N} = \mu$

the average "rate" = N/interval

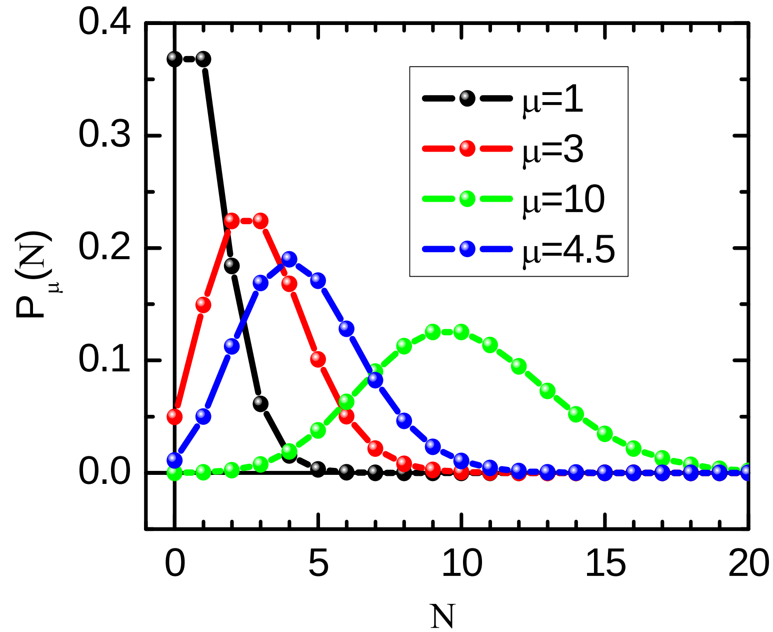
the standard deviation: $\sigma_N = \sqrt{\mu}$

* The origin of "square-root rule" for counting experiments.

The Poisson distribution (III)

At the limit $\mu \gg 1$

The Poisson distribution quickly approach the Gaussian distribution.



the mean: $\bar{N} = \mu$

the standard deviation: $\sigma_N = \sqrt{\mu}$

$$P_\mu(N) = e^{-\mu} \frac{\mu^N}{N!} \xrightarrow{\mu \gg 1} \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(N-\mu)^2}{2\mu}}$$

Lab 3a: Radioactive decay

Decay rate of ^{137}Cs is approximately a constant. The measured rate (R) depends on the distance between source and detector. (Why?)

Each run of experiments has n consecutive intervals (Δt). The decay count N_i ($i = 1, \dots, n$) at each interval follow Poisson distribution.

The average decay rate: $R = \dfrac{\bar{N}}{\Delta t}$

Standard deviation: $$\sigma_R = \frac{\sigma_{\bar{N}}}{\Delta t} = \frac{\sigma_N / \sqrt{n}}{\Delta t}$$