PHYS 500: Distributions

Lecture 3: Basic properties of known distributions

Probability Distributions-a

- Of the many probability distributions that are involved in the analysis of experimental data, three play a fundamental role: the **binomial distribution**, the **Poisson distribution**, the **Gaussian distribution** and the **Lorentzian distribution** .
- The **binomial** distribution is generally applied to experiments in which the result is one of a small number of possible final states, such as the number of “heads” or “tails” in a series of coin tosses, or the number of particles scattered forward or backward relative to the direction of the incident particle in a particle physics experiment.
- The **Poisson** distribution is generally appropriate for counting experiments where the data represent the number of items or events observed per unit interval. It is important in the study of random processes such as those associated with the radioactive decay of elementary particles or nuclear states, and is also applied to data that have been sorted into ranges to form a frequency table or histogram.

Binomial (or Bernoulli) Distribution

- Let an experiment (like tossing a coin) which has two possible outcomes: Outcome A with probability *p* and outcome B with probability *q* (with *p*+*q*=1). Let's assume that we perform *n* trials of this experiment. The probability of the outcomes does not change in each trial. Then the probability of having *x* times the outcome A is given by:

$$P_A(x;\ n,p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

- This probability is characterized by a mean value μ and and a variance σ^2 given by:

$$\mu = np, \quad \sigma^2 = np(1-p)$$

Poisson Distribution-a

- The Poisson distribution represents an approximation to the binomial distribution for the special case where the average number of successes is much smaller than the possible number; that is, when $\mu << n$ because $p << 1$.
- For such events the large number of n makes the use of binomial distribution impractical.
- Furthermore, neither the number n of possible events nor the probability p for each is usually known. What may be known instead is the average number of events μ expected in each time interval or its average value.
- The Poisson distribution provides an analytical form appropriate to such investigations that describes the probability distribution in terms of just the variable x and the parameter μ.

In practice Poisson approximates binomial when at least we have $n \geq 50$ trials and $np \leq 5$.

Poisson Distribution-b

- The form of the Poisson distribution is:

$$\lim_{p \to 0} P_A\left(x;\ n,\ p\right) = P_p\left(x;\ \mu\right) \equiv \frac{\mu^x}{x!} e^{-\mu}$$

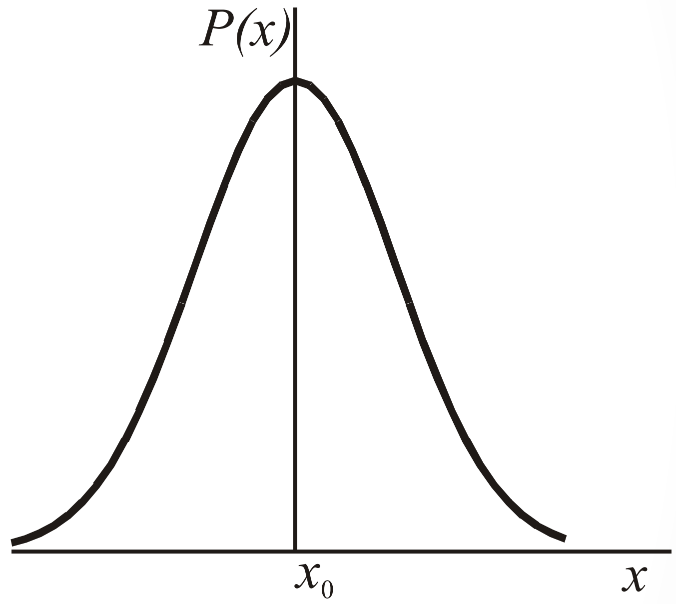
- Poisson distribution has the following mean and variance:

$$\left\langle x \right\rangle = \mu, \quad \sigma^2 = \mu$$

We must not forget that Poisson distribution is a discrete distribution with x=0, 1, 2, …

Probability Density-a

- Let's consider that we measure a physical quantity which has a real value x_0. A measurement gives us a value *x.* We make *N* measurements and we plot the curve $P(x)=n(x)/N$.
- It can be proved that as *N* gets very large, **and provided that the distribution results from random errors in measurement**, the curve takes the smooth shown in the figure. This curve is symmetrical with respect to the real value x_0 and tends to zero for very large values of *x.*



Probability Density-b

- The function $P(x)$ is a probability density. This means:
- A) That the probability of finding a measurement x between two values a and b ($a \le x \le b$) is given by

$$P\left(a,b\right) = \int_a^b P(x)\,dx$$

- The function is normalized to 1:

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

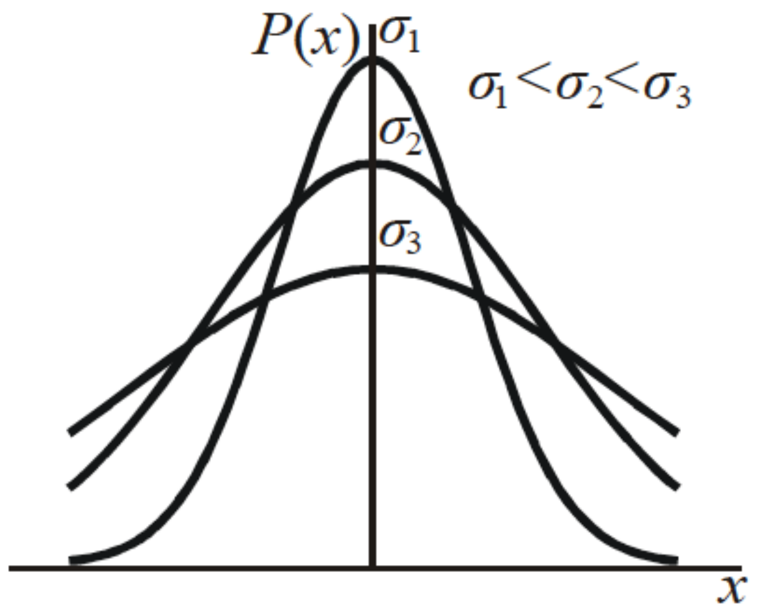
- This function is called **Gaussian distribution**

The Gaussian Distribution-a

- The actual form of the Gaussian distribution is:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-x_0)^2/2\sigma^2}$$

- The number σ is called **standard deviation** and is very important because it determines both the width and the height of the curve. As σ increases the width of the curve increases so the height decreases (can you explain why?).



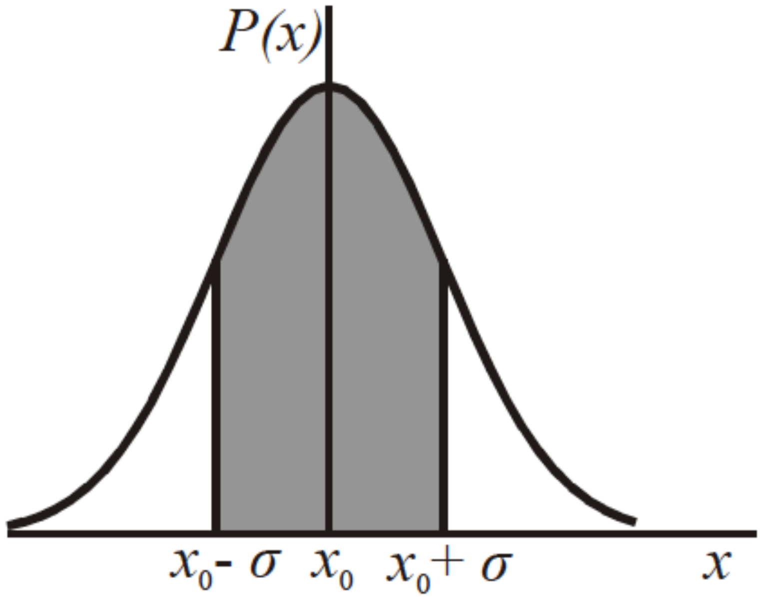
The Gaussian Distribution-b

- Another interesting characteristic of the Gaussian distribution is that the integral

$$P(-\sigma,\ \sigma) = \int_{x_0-\sigma}^{x_0+\sigma} P(x)\,dx$$

which is the area of the shaded region is always constant to a value 0.68

- Of course in an experiment we do not have an infinite number of measurements.



The Poisson distribution tends to the when λ -> ∞.

The Gaussian distribution is an approximation to the binomial distribution for the special limiting case where the number of possible different observations n becomes infinitely large and the probability of success for each is finitely large $np >> 1$.

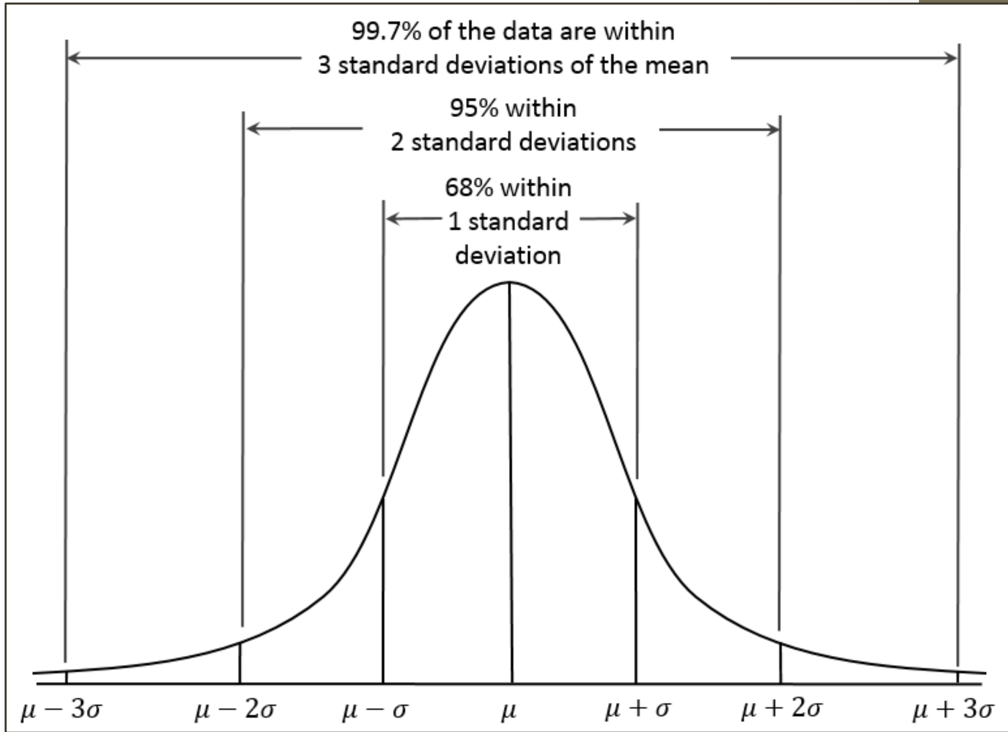
The Standard Gaussian Distribution

- It is generally convenient to use a standard form of the Gaussian equation obtained by defining the dimensionless variable $z=(x-\mu)/\sigma$ where we get:

$$P(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- The standard Gaussian distribution is related to the **error function**.

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2/2}\, du$$



The Lorentzian Distribution-a

- The Lorentzian distribution is an appropriate distribution for describing data corresponding to resonant behavior, such as the variation with energy of the cross section of a nuclear or particle reaction or absorption of radiation in the Mössbauer effect.
- The probability density is defined as:

$$p_L(x;\ \mu, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(x-\mu)^2 + (\Gamma/2)^2}$$

- The distribution is symmetric about its mean μ with a width characterized by its half-width Γ. The most striking difference between the Lorentzian and the Gaussian distribution is that it does not diminish to 0 as rapidly;

The Lorentzian Distribution-b

- It is obvious that from the symmetry of the curve, the mean, the median and the most probable value are the same.
- The standard deviation is not defined for the Lorentzian distribution as a consequence of its slowly decreasing behavior for large deviations.
- The width of Lorentzian distribution is instead characterized by the **half-width Γ** which is the width of the curve measured between the levels of half maximum probability.