Physics 566 Quantum Optics I

Lecture 2: Introduction to Stochastic Processes

Quantum mechanics is distinguished by two defining properties

(i) Intrinsic randomness.

(ii) Interfreme of indistinguishable processes.

However randomness and interference are features of classical physics and distinguishing those phenomena that have a uniquely quantum origin is a subtle business that has been at the heart of much of the activity in quantum optics since its inception. Certain nonclassical aspects arise from the corpuscular nature of the field, something not present in the classical theory. But this non-classicality can show up in a minimal way, as in Boltzmann's kinetic theory of gases, where the discretization of the fluid into discrete molecules leds to statistical fluctuations. Indeed, Einstein introduction of "light quanta," i.e. photons, in 1905 arose from his derivation of the Planck law (actually the Wien approximation at short wavelengths) based on the Boltzmann formula for the entropy of a gas of photons, and later for the fluctuation in the energy of blackbody which manefestly has a "wave component" and a "particle component" as we will see when we study the quantum theory of light. Unravelling these various stochastic and wave features and those that have a uniquely quantum description, unavailable in a classical dynamical description, will be one of the important tasks that we will persue in our studies.

To set the foundation let us begin with a review of some fundamental theory of probability and stochastic processes.

Probability

A "random variable" X can take on a value $x \in S$ for some set S of possible outcomes. S can be a discrete set or a continuous set. To each $x \in S$ we assign a probability $p(x) \geq 0$. If S is discrete (countable), let $S = \{x_i \mid i = 1, 2, \ldots, d\}$ (d can be ∞), then $p_i \equiv p(x_i)$ is the probability of outcome i. For the continuous case, $p(x)\,dx$ is the probability to have an outcome somewhere in the measurable interval $x \to x+dx$. Then $p(x)$ is a "probability density" with units $\frac{1}{[x]}$. General a probability distribution is normalized: $\sum_{i=1}^{d} p_i = 1$, $\int_{x\in S} dx\, p(x) = 1$.

- Probability and logic

Probability is the language of logic. The basic axiom of probability is that if two events A and B are mutually exclusive, the $P(A \text{ or } B) = P(A) + P(B)$. Quantum mechanics defies logic. Interference means that what appear to be multually exclusive possilities, e.g. a photon going one way or other at a beam splitter is not correct.
We define nonclassical phenomena to be those that can't be described by probability.

- Moments and Cumulants

The expected value $\langle x \rangle = \sum_i x_i p_i$, and the uncertainty is the root-mean-square deviation $\Delta x = \sqrt{\Delta x^2}$, $\Delta x^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x^2 \rangle$: Variance. Generally we define the moments of the distribut $\langle x^n \rangle = \sum_i (x_i)^n p_i$. A probability distribution is defined by all of its moments. We define the "moment generating function"

$$M(t) \equiv \langle e^{tX} \rangle = \sum_n \frac{t^n}{n!} \langle x^n \rangle$$

$$\Rightarrow \langle x^n \rangle = \left.\frac{d^n M}{dt^n}\right|_{t=0}$$

In many cases M does not converge, and it is more convenient to take $t \to ik$. Then the generating function is defined $\chi(k) = M(ik)$ and is called the "characteristic function"

$$\boxed{\text{Characteristic fnct:} \quad \chi(k) \equiv \langle e^{ikx} \rangle = \sum_i e^{ikx_i} p_i = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle}$$

The characteristic function is the Fourier transform of the probability distribution function and so is well defined since $\sum_i p_i^2$ is finite (Note, the moment generating function is sometimes defined $G(k) = \langle e^{kx} \rangle$, but this not not necessarily converge). The moments and the the coefficients in the Taylor series around $k=0$

$$\boxed{\text{Moments:} \qquad \langle x^n \rangle = (-i)^n \frac{d^n \chi(k)}{d k^n} \bigg|_{k=0}}$$

The "cumulants" $\langle \kappa^n \rangle$ are defined through the generating function

$$\boxed{\text{Cumulant generating function:} \quad K(t) \equiv \log(M(t)) = \log \langle e^{tX} \rangle \equiv \sum_{n=1}^{\infty} \frac{t^n}{n!} \langle \kappa_n \rangle}$$

$$\Rightarrow \langle \kappa_n \rangle = \frac{d^n K}{d t^n} \bigg|_{t=0}$$

Alternatively, we can define cumulants via the characteristic function

$$H(k) = \log(H(k)) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle \kappa_n \rangle \qquad \langle \kappa_n \rangle = \frac{1}{(i)^n} \frac{d^n H}{d k^n} \bigg|_{k=0}$$

Thus $\langle \kappa_1 \rangle = \langle x \rangle$, $\langle \kappa_2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \Delta x^2$, $\langle \kappa_3 \rangle = \langle (x - \langle x \rangle)^3 \rangle$

Generally $\langle \kappa_n \rangle \neq \langle (x - \langle x \rangle)^n \rangle$ for $n \geq 4$. The relationship between cumulants and moments is not simple; it's a combinatoric problem

- Multivariate Probability

Given two random variables, X Y, we can define the joint probability $p(x,y) = p(x \cup y)$.

- If X and Y are "statiscally independent" then $p(x,y) = p(x) p(y)$
- Conditional probability $p(x|y) = p(x)$ Given $p(y) = 1$ $= \frac{p(x,y)}{p(y)}$; also $p(y|x) = \frac{p(x,y)}{p(x)}$

$$\Rightarrow \text{Bayes rule} \quad p(y|x) = \frac{p(x|y)\, p(y)}{p(x)}$$

– "Marginal probability" $p(x_i) = \sum_j p(x_i, y_j)$; $p(y_j) = \sum_i p(x_i, y_j)$

Generalize to N-random variables $\vec{X} = (X_1, X_2, \ldots, X_N)$; $p(\vec{x}) = p(x_1 \cup x_2 \cup \cdots x_n)$.

Statistical independence $\Rightarrow$ $p(\vec{x}) = \prod_{i=1}^{N} p(x_i)$.

"Marginalize" over and M variables, e.g. $p(x_1, x_2) = \sum_{\{x_3, x_4 \ldots x_N\}} p(\vec{x})$

Aside Let $Y = \sum_{i=1}^{N} X_i$ (random variable) $\Rightarrow$

$\Rightarrow$ $\langle y \rangle = \sum_i \langle x_i \rangle$, $\Delta y^2 = \sum_i \Delta x_i^2$, if $\{x_i\}$ are statistically indep.

Multivariate characteristic function $\chi(\vec{k}) \equiv \langle e^{i\vec{k}\cdot\vec{x}} \rangle = \sum_{\{\vec{x}\}}' p(\vec{x})\, e^{i\vec{k}\cdot\vec{x}}$

$\Rightarrow$ Multivariate moments: E.g. $\langle x_i^n x_j^m \rangle = (-i)^{n+m} \dfrac{\partial^{n+m}}{\partial k_i^n \partial k_j^m} \chi(\vec{k}) \Big|_{\vec{k}=0}$

Covariance $\langle \Delta x_i \Delta x_j \rangle = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$

$= \langle \Delta x_i^2 \rangle \delta_{ij}$ if x_i + x_j statistically independent.

More generally, for statistically independent random variables, cummulants of a sum of random variables is the sum of the cummulants, e.g. two random variables $p(x_1, x_2) = p(x_1)\, p(x_2)$, Let $Y = X_1 + X_2$

$$K_Y(t) = \log \langle e^{t(x_1+x_2)} \rangle = \log \left(\langle e^{tx_1} \rangle \langle e^{tx_2} \rangle \right) = \log \langle e^{tx_1} \rangle + \log \langle e^{tx_2} \rangle$$

$$= K_{x_1}(t) + K_{x_2}(t)$$

Multivariate cumulants: Generating function $K(\vec{t}) = \log \langle e^{\vec{t}\cdot\vec{x}} \rangle$

With some work, one can show that the multivariate cummulants are

$$K(x_1, x_2, \ldots, x_n) = \sum_{\pi} (|\pi| - 1)!\, (-1)^{|\pi|-1} \prod_{B \in \pi} \langle \prod_{i \in B} x_i \rangle$$

where π lists all portions of $\{1, \ldots, n\}$, B lists all block of the portion, $|\pi|$ = number of parts in portion

e.g. $K(x,y,z) = \langle xyz \rangle - \langle xy \rangle \langle z \rangle - \langle xz \rangle \langle y \rangle - \langle yz \rangle \langle x \rangle + 2 \langle x \rangle \langle y \rangle \langle z \rangle$

– Some fundamental probability distributions

- Bernoulli trial $S = \{0, 1\}$ $p(x) = \begin{cases} p & x = 1 \\ q = 1-p & x = 0 \end{cases}$

$\langle x \rangle = p$ $\Delta x = \langle x^2 \rangle - \langle x \rangle^2 = p - p^2 = pq$

- Binomial distribution $k = \sum_{i=1}^{N} x_i$, X_i = Bernoulli trial $\quad S = \{0, 1, \ldots, N\}$

$$p(k) = \binom{N}{k} p^k q^{N-k} = \frac{N!}{k!\,(N-k)!}\, p^k (1-p)^{N-k}$$

: N statistically independent Bernoulli trial with k 1's, $N-k$ 0's

$$\langle k \rangle = Np \,, \qquad \Delta k^2 = Np(1-p)$$ (From N-statistically indepenlent x_i)

- Poisson distribution:

Limit of binomial distribution with $N \to \infty$, $p \to 0$, such that $Np \to \Lambda$

(Infinitely many Bernouli trails with infinestimal probability):

Random variable $k \in \{0, 1, \ldots, \infty\}$ (nonnegative integers)

$$p(n) = \lim_{\substack{N\to\infty,\, p\to 0 \\ Np \to \Lambda}} \frac{N!}{(N-n)!\, n!}\, p^n (1-p)^{N-n} = \lim \frac{\Lambda^n}{n!} \frac{N!}{(N-n)!\, N^n} \left(1 - \frac{\Lambda}{N}\right)^N \left(1 - \frac{\Lambda}{N}\right)^{-n} = \frac{\Lambda^n e^{-\Lambda}}{n!}$$

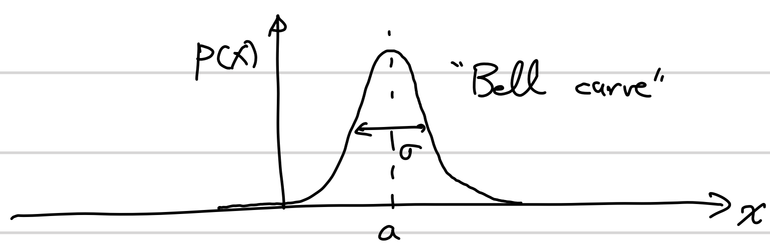
$$\langle n \rangle = \Lambda \,, \qquad \Delta n^2 = \langle n \rangle = \Lambda$$

- Gaussian (normal) distribution

Random variable $X \in \mathbb{R}$ (reals) : Continuous variable

$$p(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad \text{Normalized: } \int_{-\infty}^{\infty} dx\, p(x) = 1$$

$$\langle x \rangle = a \qquad \Delta x = \sigma$$



The normalized form of Gaussian is something every physicist should know by memory.
A Gaussian is fully specified by its first two moments $\langle x \rangle$ and $\langle x^2 \rangle$

Aside: A very useful Gaussian integral in complex plane.

$$\int_{-\infty}^{\infty} dx\, e^{-a(x-z)^2} = \sqrt{\frac{\pi}{a}} \,, \quad a, z \text{ complex}, \quad \mathrm{Re}(a) > 0$$

Characteristic Function $\chi(k) = e^{ika}\, e^{-\frac{\sigma^2 k^2}{2}} \Rightarrow -i \frac{\partial \chi}{\partial k}\Big|_k = \langle x \rangle = 0 \,, \quad -\frac{\partial^2 \chi}{\partial k^2}\Big|_0 = \langle x^2 \rangle = \sigma^2 + a^2 = \Delta x^2 + \langle x \rangle^2$

Cumulant generator $H(k) = \log(\chi(k)) = ika - \frac{\sigma^2}{2} k^2$

$$\Rightarrow \langle \kappa_1 \rangle = a = \langle x \rangle \,, \quad \langle \kappa_2 \rangle = \sigma^2 = \Delta x^2 \,, \quad \langle \kappa_3 \rangle = 0 \quad n \geq 3$$

$\Rightarrow$ All cumulants of a Gaussian vanish for $n \geq 3$. The higher order cumulants are thus a measure of how different the moments are from those of a Gaussian.

When $\langle x \rangle = 0$, $\chi(k) = e^{-\frac{\sigma^2 k^2}{2}}$. It the follows $\langle x^n \rangle = \begin{cases} 0 & n\text{-odd} \\ (n-1)!!\,\sigma^n & n\text{-even} \end{cases}$

<u>Multivariate Gaussian</u> : Joint probability of $\vec{x} = (x_1, x_2, \dots x_N) \in \mathbb{R}^N$

$$p(\vec{x}) \equiv \frac{1}{(2\pi \det \mathbb{C})^{1/2}} \exp\left\{-\frac{1}{2}\sum_{ij}^{N}(x_i - a_i)\,\mathbb{C}^{-1}_{ij}\,(x_j - a_j)\right\}$$

where $\mathbb{C}$ is a symmetric positive matrix known as the covariance matrix

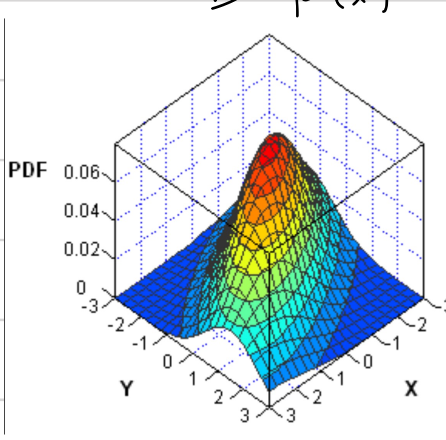
$\langle x_i \rangle = a_i$ $\quad\quad$ $\langle \Delta x_i \Delta x_j \rangle = C_{ij}$ (covariance)

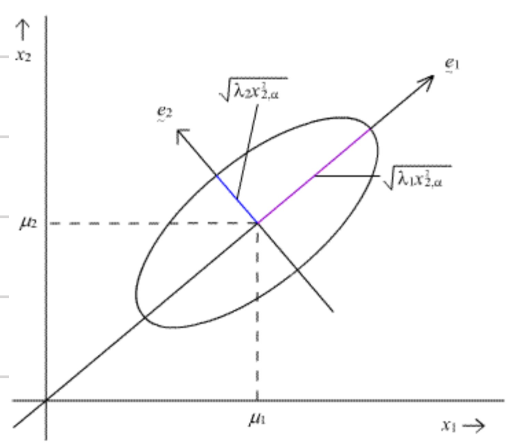
The multivariate Gaussian is fully specified by in mean and covariance matrix

the Gaussian factorizes in terms of the eigenvectors of $\mathbb{C}$: Let $\mathbb{C}\vec{u}^{(\alpha)} = \sigma_\alpha^2 \vec{u}^{(\alpha)}$

Let $\vec{x} = \sum x_\alpha \vec{u}^{(\alpha)}$ coordinate in eigenbasis $\quad\quad$ $\vec{a} = \sum a_\alpha \vec{u}^{(\alpha)}$

$$\Rightarrow p(\vec{x}) = \prod_{\alpha=1}^{N} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}}\, e^{-\frac{(x_\alpha - a_\alpha)^2}{2\sigma_\alpha^2}}$$





<u>Central Limit Theorem</u>:

Let $Y = \sum_{i=1}^{N} X_i$ be a random variable, where $\{X_i\}$ are statisticall independent but drawn from the identical probability distribution

We have already seen $\langle y \rangle = \sum \langle x_i \rangle = N\langle x \rangle$, $\quad \Delta y^2 = \sum \Delta x_i^2 = N\Delta x^2$

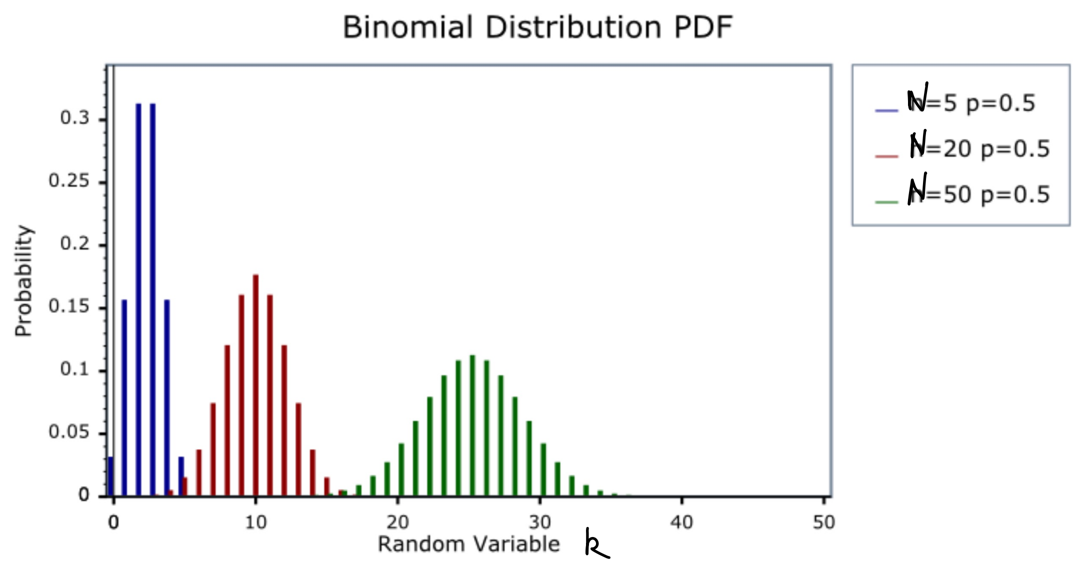
According to the Central Limit theorem, as $N \to \infty$

$$\boxed{p(y) = \frac{1}{\sqrt{2\pi N \Delta x^2}}\, e^{-\frac{(y - N\langle x \rangle)^2}{2N\Delta x^2}}}$$

That is as $N \to \infty$ Y is normally distributed, with mean $N\langle x \rangle$ and variance $N\Delta x$

Example: The binomial distribution is the sum of identical Bernoulli trail of mean $\langle x \rangle = p$ and variance $\Delta x^2 = p(1-p)$. Thus as the number N of trials $\to \infty$, the binomial distribution converges to a Gaussian

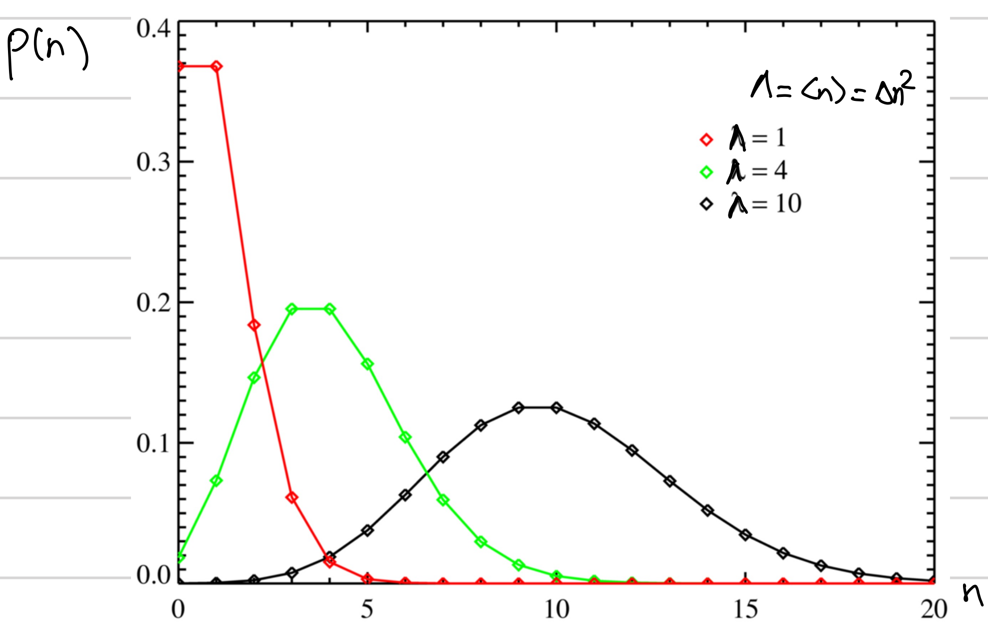
$$\lim_{N\to\infty} p(k) = \lim_{N\to\infty} \binom{N}{k} p^k (1-p)^k \to \frac{1}{\sqrt{2\pi N p(1-p)}}\, e^{-\frac{(k-Np)^2}{2Np(1-p)}}$$



Example: The Poisson distribution is limit of a binomial distribution, with $Np \to \Lambda$. As $\Lambda \to \infty$, the Poisson distribution should then converge to a Gaussian

$$\text{Poisson} \quad p(n) = \frac{\Lambda^n e^{-\Lambda}}{n!} = \frac{\langle n\rangle^n e^{-\langle n\rangle}}{n!}, \qquad \langle n\rangle = \Lambda \qquad \Delta n^2 = \langle n\rangle = \Lambda$$

$$\Rightarrow \text{ as } \langle n\rangle \to \infty \qquad \boxed{p(n) \to \frac{1}{\sqrt{2\pi\langle n\rangle}}\, e^{-\frac{(n-\langle n\rangle)^2}{2\pi\langle n\rangle}}}$$



Stochastic Process

A Stochastic Process is a random variable as a function of time $X(t)$, i.e., $X(t)$ is drawn from some probability distribution

$$p\,(X(t)=x) = p(x,t)$$

Given a time series, on random variable $X(t_1), X(t_2), \ldots X(t_k)$; we define the joint probability distribution $p(x_k t_k; \cdots x_2 t_2; x_1 t_1)$ typically ordered $t_k > \ldots > t_2 > t_1$.

Stationary Process: A stochastic process is said to b "stationary" if the "statistics" of $X(t)$ do not change with time. Formally, the joint probability distribution

$$p(x_k\, t_k; \ldots x_2 t_2, x_1 t_1) = p(x_k\, t_k+T; \ldots x_2\, t_2+T;\ x_1\, t_1+T)$$

Thus, the joint probabilities depend only on the time difference $t_i - t_j$, and not the origin of time.

Thus; the one-time probabilities are independent of time $p(x,t) = p(x)$

the two time probabilities $p(x_2 t_2; x_1 t_1) = p(x_2\, t_2 - t_1,\ x_1 0)$

Ergodic Process: Intuitively, a stochastic process is said to be ergodic if over time one uniformly samples the probability distribution. In that case, an ensemble average over the probability distribution is equal to a time average in the limit that the averaging time goes to infinity. A stochastic process can be ergodic only if it is also stationary, and typically we consider "wide-sense" stationary, meaning stationary for one & two-point correlations:

$$\Rightarrow \langle X(t)\rangle = \int dx\ x\ p(x,t) = \underset{\text{stationary}}{\int dx\ x\, p(x)} \simeq \lim_{T\to\infty} \int_{-T/2}^{T/2} x(t)\,dt = \overline{x}$$

$$\langle x(t+\tau)x(t)\rangle = \int dx_1 dx_2\ x_1 x_2\ p(x_1 t;\ x_2\ t+\tau) = \int dx_1 dx_2\ x_1 x_2\ p\,(x_1 0; x_2 \tau)$$

$$= \lim_{T\to\infty} \int_{-T/2}^{T/2} dt\ x(t)\,x(t+\tau) = \overline{x(t)x(t+\tau)}$$

Markov Process: A Markov Process has "no memory"! Thus the probability of X taking on a value x at time t depends only on the value immediately before, and not the whole history.

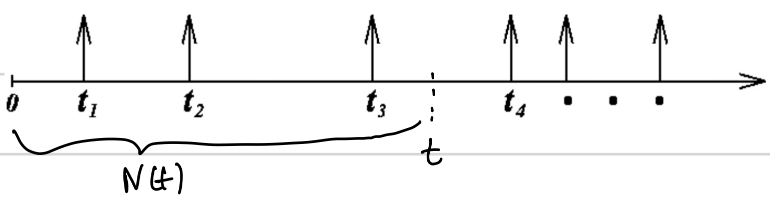
Formally the conditional probability

$$p(x_k t_k |\ x_{k-1} t_{k-1};\ x_{k-2}\, t_{k-2}; \ldots x_2 t_2;\ x_1 t_1) = p(x_k t_k |\ x_{k-1}\, t_{k-1}) \quad \text{where } t_k > t_{k-1} > t_{k-2} \cdots t_2 > t_1$$

A "true" Markov process has no memory, even for a differential time interval. There is no such thing physically, but, as we will see, many quantum optical systems are well approximated as Markov processes.

Example: Poisson Process

A Poisson process describes the counting of random events that arrive at the counter at a constant rate, but event event is random and completely uncorrelated from any other event



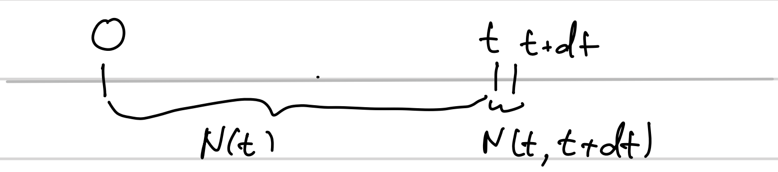
(Most randomly spaced events)

This is a stationary, ergodic, Markov process.

We define the random varible $N(t) = \#$ of counts $[0,t)$ and $N(t_a, t_b) =$ $\#$ between t_a and t_b $= N(t_a) - N(t_b)$. $N(t)$ is a process with rate λ in two equivalent definitions:

(1) The Probability distribution of $N(t_a, t_b)$ is a Poisson distribution with mean $\lambda(t_b - t_a)$, and every nonoverlapping interval is statistically independent.

(2) In an infinitesimal time interval dt there can be no more than one count with probality $P(N(t, t+dt) = 1) = \lambda dt$, statically independent of the arrivals outside of this interval.



Let us show that definition (2) leads to the Poisson distibution.

Let $P(n,t) =$ Probability $(N(t) = n)$.

$$P(1, dt) = \lambda dt \qquad P(0, dt) = 1 - \lambda dt$$

$$\Rightarrow P(0, t+dt) = P(0,t)\, P(0,dt) = P(0,t)(1 - \lambda dt) \Rightarrow \frac{dP(0,t)}{dt} = -\lambda P(0,t)$$

$$\Rightarrow P(0,t) = Ce^{-\lambda t}, \text{ and since } P(0, t=0) = 1, \; C = 1$$

By similar reasoning, $P(n, t+dt) = P(n,t)(1 - \lambda dt) + P(n-1, t)\lambda dt$

$$\Rightarrow \frac{dP(n,t)}{dt} + \lambda P(n,t) = \lambda P(n-1,t) \Rightarrow \frac{d}{dt}\left[e^{\lambda t} P(n,t)\right] = e^{\lambda t} \lambda P(n-1,t)$$

Iterating one finds $\boxed{P(n,t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}}$ the Poisson distribution