THE UNIVERSITY OF CHICAGO

THE STATISTICAL MECHANICS OF TRANSCRIPTIONAL CONTROL

A THESIS SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
CLAYTON W. SEITZ

CHICAGO, ILLINOIS
SPRING 20XX

TABLE OF CONTENTS

ABSTRACT

Eukaryotic transcription is episodic, consisting of a series of transcriptional bursts, Bursty transcriptional dynamics are well-exemplified by the transient expression of pro-inflammatory guanylate binding proteins (GBPs) - a group interferon-inducible GTPases that restrict the replication of intracellular pathogens [XXX]. Classical models of gene regulation explain transcriptional bursts by invoking stochastic binding and unbinding of transcription factors, RNA polymerase and mediator proteins at enhancer or promoter sequences. However, more recent studies have pointed towards a more cooperative picture of transcriptional control where phase-separated aggregates of DNA, RNA, and proteins form higher-order structures to control gene expression. For example, both chromatin immunoprecipitation and super resolution imaging have captured the phase separation of super-enhancer-binding proteins MED1 and BRD4 in transcriptional condensates at the *Essrb* genomic locus [XXX]. Furthermore, fluorescence microscopy techniques have colocalized MED1 and BRD4 with the GBP gene cluster alongside a reduction in the degree of disorder of 3D chromatin structure in murine macrophages after infection with *Mycobacterium tuberculosis*. Taken together, these results suggest that phase separation may play a role in the reorganization of chromatin structure during trancriptional control of innate immune response genes [XXX]. Here, we hypothesize that phase separation reduces the entropy of chromatin structure in order to induce bursty gene expression. Using single molecule localization microscopy (SMLM) to obtain super-resolution images of the H2B protein, we intend to demonstrate simultaneous (i) loss of disorder in chromatin structure (ii) formation of transcriptional condensates containing MED1 and BRD4 and (iii) non-Poissonian gene expression. The following sections discuss recent the biological evidence in more detail and summarize the single molecule microscopy techniques and biophysical models we employ to study the interactions between transcriptional condensates and the chromatin scaffold.

CHAPTER 1

SINGLE MOLECULE LOCALIZATION MICROSCOPY

1.1 Stochastic optical reconstruction microscopy

1.2 PSF engineering in single molecule localization microscopy

Born and Wolf's approximation, Lateral and axial point spread functions, Potentially double helix point spread functions for 3D storm

1.2.1 Fisher information theory for parameter estimation in single molecule microscopy

1.3 Bounding parameter uncertainty in single molecule localization

1.3.1 Point spread functions in widefield microscopy

Most detectors used for imaging have many elements (pixels) so that we can record an image projected onto the detector by a system of lenses. In fluorescence imaging, this is usually a relay consisting of an objective lens and a tube lens to focus the image onto the camera. Due to diffraction, any point emitter, such as a single fluorescent molecule, will be registered as a diffraction limited spot. The profile of that spot is often described as a Gaussian point spread function (Richardson and Wolf)

$$\mathrm{PSF}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right) \tag{1.1}$$

where (x_0, y_0) defines the coordinates of a single fluorescent molecule on a detector $\mathcal{D}$.

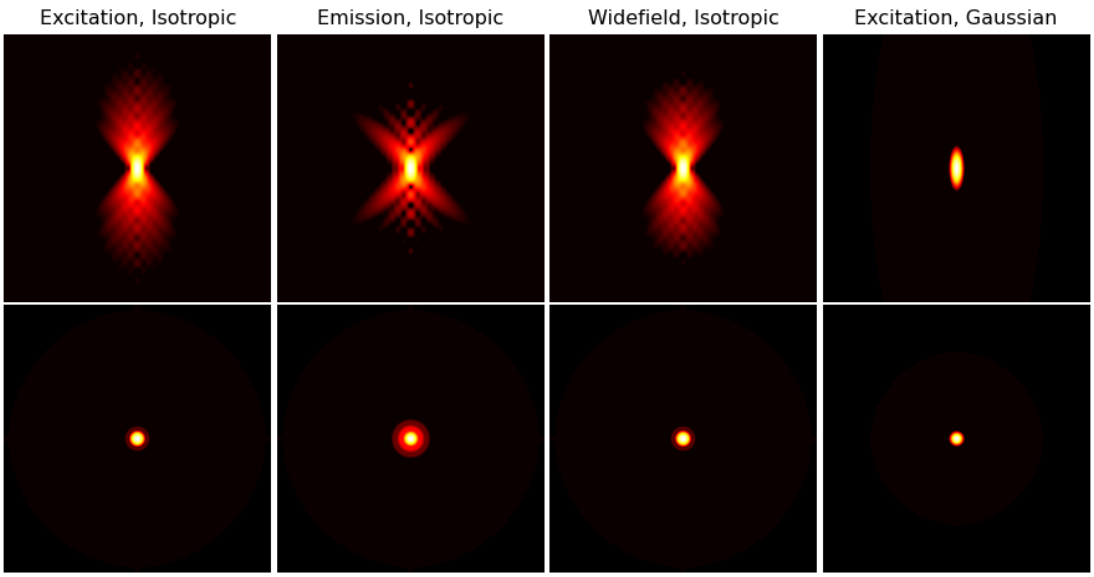

Figure 1.1: Cartoon of a excitatory-inhibitory neural network receving a time-dependent feedforward current $F(t)$, feedback current $R(t)$, and synaptic noise $\xi(t)$. Excitatory neurons are shown in red and inhibitory neurons shown in blue. Synaptic current is drawn to be instantaneous and syanptic weights J_{ij} are not functions of time. Feedforward inputs could be weighted Poisson processes where F defines the rates, for example.

1.3.2 Maximum a posteriori estimation of PSF parameters

Maximum a posteriori estimation of parameters

$$\begin{aligned}\theta^*_{\text{MAP}} &= \underset{\theta}{\text{argmax}}\, P(\theta|\mathcal{X}) \\ &= \underset{\theta}{\text{argmax}}\, \frac{\mathcal{L}_\theta \pi(\theta)}{\int_\theta \mathcal{L}_\theta \pi(\theta)}\end{aligned}$$

where $\mathcal{L}_\theta = \mathcal{L}(\theta|\mathcal{X})$ is the likelihood function and $\pi(\theta)$ a prior on the parameters. Furthermore, we will assume that $\pi(\theta)$ is a uniform distribution. An image captured by a camera can be loosely thought of as histogram of photon arrivals and a discretized form of the density $\text{PSF}(x, y)$ over an integration time τ. If $\text{PSF}(x, y)$ can be approximated as constant in time ($\tau << 1$), the value at a pixel approaches an integral of this density over

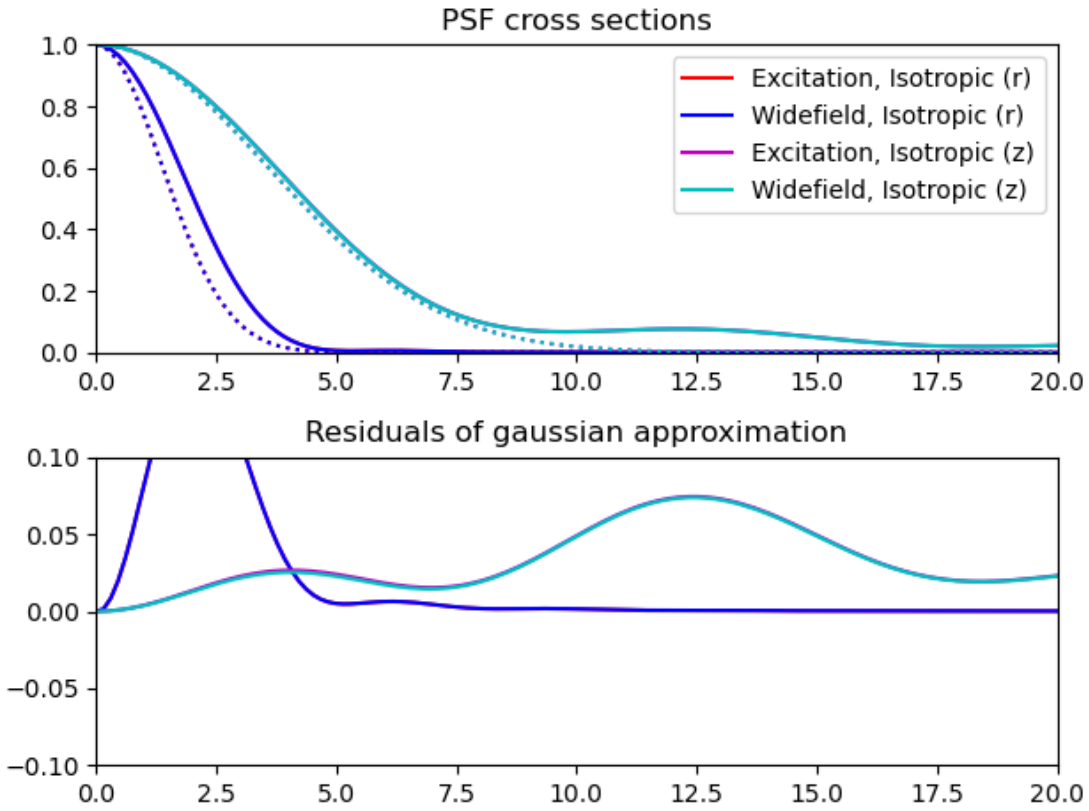

Figure 1.2: Cartoon of a excitatory-inhibitory neural network receving a time-dependent feedforward current $F(t)$, feedback current $R(t)$, and synaptic noise $\xi(t)$. Excitatory neurons are shown in red and inhibitory neurons shown in blue. Synaptic current is drawn to be instantaneous and syanptic weights J_{ij} are not functions of time. Feedforward inputs could be weighted Poisson processes where F defines the rates, for example.

the pixel:

$$\lambda_k = \gamma \cdot \int_{\mathcal{D}_k} q(x, y) dx dy \tag{1.2}$$

Put another way, the variable λ_k at a pixel k defines the probability of observing a photon per unit time and therefore we can define the number of photons arriving at each pixel as a random variable

$$P(S_k) = \text{Poisson}(\lambda_k)$$

It would be convenient to obtain an analytical expression for λ_k given the parameters of the point spread function $\theta = (x_0, y_0, \sigma)$. This can be done by making use of the error function

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dx$$

where $t = \frac{x}{\sqrt{2}\sigma}$. We just need to rescale this so it is appropriate for a standard normal distribution. We need something like

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-t^2) dx$$

so we rescale the error function by $\beta = 1/2\sqrt{2}$. Of course, this can also be used to construct the integral over an arbitrary interval $[a, b]$

$$\text{erf}(b) - \text{erf}(a) = \frac{2\beta}{\sqrt{\pi}} \int_{\frac{a}{\sqrt{2}\sigma}}^{\frac{b}{\sqrt{2}\sigma}} \exp(-t^2) dt$$

We use this result to compute the Poisson rate λ_k at pixel k. Now, the Poisso rate is in general given by

$$\lambda_k = \frac{1}{2\pi} \left(\int_{x_k}^{x_{k+1}} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2} \right) dx \right) \left(\int_{y_k}^{y_{k+1}} \exp\left(-\frac{(y - y_0)^2}{2\sigma^2} \right) dy \right)$$

Let $t = (x - x_0)/\sqrt{2}\sigma$ and $z = (y - y_0)/\sqrt{2}\sigma$ and write constant in terms of β

$$\lambda_k = \frac{4\beta^2}{\pi}\left(\int_{t_k}^{t_{k+1}} \exp\left(-t^2\right) dx\right)\left(\int_{z_k}^{z_{k+1}} \exp\left(-z^2\right) dy\right)$$
$$= \beta^2 \left[\mathrm{erf}(t_{k+1}) - \mathrm{erf}(t_k)\right]\left[(\mathrm{erf}(z_{k+1}) - \mathrm{erf}(z_k)\right]$$

where for example $t_k = (x_k - x_0)/\sqrt{2}\sigma$

Detectors often suffer from dark noise (thermal noise) and there may also be a background signal. Considering the former first, we define another r.v. W_k which represents Gaussian dark noise associated with pixel k. Unless otherwise specified we will always assume that $W_k \sim \mathcal{N}(m_k, \sigma^2_{w,k})$. We represent our corrupted signal as another random variable H_k. As a side note, we define S_k to have units of photons [p], while H_k has units of photoelectrons $[e^-]$. The conversion factor between the two is the gain of the detector element g_k $[e^- \mathrm{p}^{-1}]$.

Dropping the pixel subscripts for the moment, the Fourier transform of a normal distribution $\mathcal{N}(\mu, \sigma^2_w)$, or its characteristic function reads

$$\tilde{P}(W) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{(x-\mu)^2}{2\sigma_w^2}} e^{-i\omega x} dx$$
$$= \frac{e^{-\frac{\sigma_w^2\omega^2}{2} + i\mu\omega}}{\sqrt{2\pi}}$$

On the other hand for $\lambda >> 1$ we can approximate the Poisson distribution as a normal distribution $\mathcal{N}(\lambda, \sqrt{\lambda})$ on countable support. We can then then write the likelihood function $\mathcal{L}_{\theta,k}$ as the inverse Fourier transform of the product of these two characteristic functions

$$\begin{aligned}\mathcal{L}_{\theta,k} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_k^2 \omega^2}{2} + i\mu_k \omega} e^{-\frac{\lambda_k \omega^2}{2} + i\lambda_k \omega} e^{-i\omega x} d\omega \\ &= \frac{1}{\sqrt{2\pi(\lambda_k + \mu_k^2)}} e^{-\frac{(-x+\mu_k+\lambda_k)^2}{2(\lambda_k+\mu_k^2)}}\end{aligned}$$

We can now return to the original optimization problem

$$\begin{aligned}\theta^*_{\text{MAP}} &= \underset{\theta}{\text{argmax}} \;\; \pi(\theta)\mathcal{L}_\theta \\ &= \underset{\theta}{\text{argmax}} \prod_k \mathcal{L}_{\theta,k} \\ &= \underset{\theta}{\text{argmax}} \sum_k \log \mathcal{L}_{\theta,k}\end{aligned}$$

$$\sum_k \log \mathcal{L}_{\theta,k} = \sum_k -\frac{(-x+\mu_k+\lambda_k)^2}{2(\lambda_k+\mu_k^2)} - \log\left(\sqrt{2\pi(\lambda_k+\mu_k^2)}\right)$$

Maximizing $\mathcal{L}_\theta$ is therefore equivalent to maximizing $\mathcal{L}_{\theta,k}$ at every pixel. Taking the derivative of the first term with respect to a parameter θ

$$\begin{aligned}&\frac{\partial}{\partial\theta}\left(\frac{(-x+\mu_k+\lambda_k)^2}{2(\lambda_k+\mu_k^2)}\right) \\ &= \left(\frac{4(\lambda_k+\mu_k^2)(-x+\mu_k+\lambda_k)\frac{\partial\lambda_k}{\partial\theta} - 2\frac{\partial\lambda_k}{\partial\theta}(-x+\mu_k+\lambda_k)^2}{4(\lambda_k+\mu_k^2)^2}\right)\end{aligned}$$

$$\frac{\partial}{\partial\theta}\log\left(\sqrt{2\pi(\lambda_k+\mu_k^2)}\right)=\frac{1}{2(\lambda_k+\mu_k^2)}\frac{\partial\lambda_k}{\partial\theta}$$

$$\begin{aligned}\frac{\partial\lambda_k}{\partial x_0}&=\beta^2\left(\int_{x_k}^{x_{k+1}}\frac{\partial}{\partial x_0}\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx\right)\left[(\mathrm{erf}(z_{k+1})-\mathrm{erf}(z_k)\right]\\&=\frac{\beta^2}{\sigma^2}\left(\int_{x_k}^{x_{k+1}}(x-x_0)\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx\right)\left[(\mathrm{erf}(z_{k+1})-\mathrm{erf}(z_k)\right]\end{aligned}$$

$$\int_{x_k}^{x_{k+1}}x\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx-x_0\int_{x_k}^{x_{k+1}}\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx$$

$$\begin{aligned}\int_{x_k}^{x_{k+1}}x\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx&=\sigma\sqrt{2\pi}\left|_{x_k}^{x_{k+1}}x-\int_{x_k}^{x_{k+1}}\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx\right.\\&=\sqrt{2\pi}\sigma\Delta-\mathrm{erf}(t_{k+1})-\mathrm{erf}(t_k)\end{aligned}$$

1.3.3 Fisher information and the Cramer-Rao bound

Suppose we are given samples from a normal distribution with unknown parameters. We then decide to build a model normal distribution. Perhaps we define a likelihood function over the $\mu-\sigma$ plane based on a set of samples, does the likelihood of the data vary? If the likelihood surface is flat, all parameter sets would be equally likely and the data does not appear to carry much information about the parameters. If the surface has a number of bumps or inflection points, then we expect our data does carry information about the parameters. This "bumpiness" of the likelihood surface is measured by computing the variance

the second derivative of the likelihood over the parameter space

The log-likelihood of an image under parameters $\theta = (x_0, y_0, \sigma^2, \sigma_w^2, g_k)$, assuming each pixel is an independent but not identical random variable, reads

$$\ell(H|\theta) = \log \prod_k^{M^2} P(H_k|\theta) = \sum_{k=1}^{M^2} \ell(H_k|\theta)$$

When there are many parameters, the Fisher Information (second moment of the score) is a matrix

$$I_{ij}(\theta) = \mathop{\mathbb{E}}_{\theta} \left[\frac{\partial}{\partial \theta_i} \left(\ell(H|\theta) \right) \frac{\partial}{\partial \theta_j} \left(\ell(H|\theta) \right) \right]$$

The only elements of the matrix of interest are on the diagonal, which allows us to write

$$I(\theta_i) = \mathop{\mathbb{E}}_{\theta} \left[\sum_k \frac{\partial^2}{\partial \theta_i^2} \ell(H_k|\theta) \right]$$

At this point, we need to evaluate the second derivative w.r.t each of the parameters in $\theta = (x_0, y_0, \sigma^2)$. We have shown that the model for the number of photoelectrons at a pixel is

CHAPTER 2

STATISTICAL MECHANICS OF GENE REGULATION

2.1 Introduction

2.2 Transcriptional bursts and chromatin architecture

2.3 Transcriptional condensates: a phase separation model for transcriptional control

Heavily influenced by Young's group at MIT

2.4 Brief introduction to polymer dynamics

2.5 Statistical mechanics of transcriptional condensates

2.6 The chemical master equation and finite state projection

The central assumption underlying a Markov process, is the memoryless property

$$P(X_t|X_{t-1}, X_{t-2}, ..., X_{t-N}) = P(X_t|X_{t-1})$$

A single Markov chain is the set of states $\boldsymbol{X} = \{X_1, X_2, ..., X_N\}$. Such a set can be generated provided that $P(X_t|X_{t-1})$ is known. To capture $P(X_t|X_{t-1})$ for all possible pairs X_t and X_{t-1}, we define a square transition matrix $T \in \mathcal{R}^{N\times N}$ where $N = |\Omega|$. As such, the elements of T represent the probability of a transition from a state ω_j to ω_i in a unit time

$$T_{ij} = \Pr\left(X_t = \omega_i, \mid X_{t-1} = \omega_j\right)$$

Under these definitions, the row T_i represents the present time, and is a conditional probability distribution $P(\omega|X_{t-1} = \omega_j)$ which requires that

$$\sum_j T_{ij} = \sum_j P(X_t = \omega_j | X_{t-1} = \omega_i) = 1$$

The matrix T is not necessarily symmetric $T_{ij} \neq T_{ji}$. One should note that the columns T_j *do not* define a probability distribution $P(X_t = \omega_i|X_{t-1} = \omega_j)$ and therefore do not necessarily sum to unity. The probability $P(X_t = \omega_i|X_{t-1} = \omega_j)$ has no meaning in this context, since we have defined the rows to represent a probability of the future given the present. We simply sample $X_t \sim P(X_t = \omega_j|X_{t-1} = \omega_i)$, assign $i = j$, and repeat. It directly follows from the fundamental rules of probability, the first order dynamics for a **particular** state ω_i: $P(\omega_i, t)$ is given by

$$P(\omega_i, t + dt) = P(\omega_i, t) + \mathcal{J}_i dt \tag{2.1}$$

The net probability current $\mathcal{J}_i$ must be

$$\mathcal{J}_i = \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t)$$

The first is a sum on a column and the second a sum on a row. This can be simplified further by noticing that the normalization condition implies

$$\begin{aligned} T_{ij} &= 1 - \sum_j T_{ij}(1 - \delta_{ij}) \\ &= 1 - \sum_j T_{ij} + \sum_j T_{ij}\delta_{ij} \end{aligned}$$

$$
\begin{aligned}
\mathcal{J}_i &= \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t) \\
&= \sum_i \left(1 - \sum_j T_{ij} + \sum_j T_{ij} \delta_{ij} \right) P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t) \\
&= |\Omega| - |\Omega| + \sum_i \sum_j T_{ij} P(\omega_j, t) \delta_{ij} - \sum_j T_{ij} P(\omega_i, t) \\
&= \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)
\end{aligned}
$$

Notice that the Kronecker delta effectively just swaps the index. Taking the limit of (1.1), we arrive at the **master equation**

$$
\frac{\partial P(\omega_i)}{\partial t} = \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)
$$

It is common to then define an operator $\boldsymbol{W}$ s.t. $W_{ij} = T_{ij}$ and $W_{ii} = -\sum_j T_{ij}$

$$
\frac{dP(\omega_i)}{dt} = \sum_j W_{ij} P(\omega_j) \rightarrow \frac{dP(\boldsymbol{\omega})}{dt} = \mathcal{J}(\boldsymbol{\omega}) = \mathbf{W} P(\boldsymbol{\omega})
$$

This operator form has a solution in terms of a matrix exponential

$$
P(\boldsymbol{\omega}, t) = \exp(\mathcal{J}(\boldsymbol{\omega}))
$$

This matrix exponential is intractable for large $|\Omega|$. However, in the Finite State Projection algorithm, it is possible to truncate the state space $\Omega \rightarrow \tilde{\Omega}$ and obtain good estimates $\tilde{P}(\boldsymbol{\omega}, t)$ with some certificate of accuracy.

2.6.1 Sequential tempered Markov Chain Monte Carlo

Appendices

APPENDIX A

DERIVATION OF THE FOKKER PLANCK EQUATION

A.0.1 Kramers-Moyal Expansion

Gixen many instantiations of a stochastic xariable x, we can construct a normalize histogram oxer all obserxations as a function of time $P(x,t)$. Howexer, in order to systematically explore the relationship between the parameterization of the process and $P(x,t)$ we require an expression for $\dot{P}(x,t)$. If we make a fundamental assumption that the exolution of $P(x,t)$ follows a Markox process i.e. its exolution has the memoryless property, then we can write

$$P(x',t) = \int T(x',t|x,t-\tau)P(x,t-\tau)dx \tag{A.1}$$

which is known at the Chapman-Kolmogorox equation. The factor $T(x',t|x,t-\tau)$ is known as the *transition operator* in a Markox process and determines the exolution of $P(x,t)$ in time. We proceed by writing $T(x',t|x,t-\tau)$ in a form referred to as the Kramers-Moyal expansion

$$\begin{aligned} T(x',t|x,t-\tau) &= \int \delta(u-x')T(u,t|x,t-\tau)du \\ &= \int \delta(x+u-x'-x)T(u,t|x,t-\tau)du \end{aligned}$$

If we use the Taylor expansion of the δ-function

$$\delta(x+u-x'-x) = \sum_{n=0}^{\infty} \frac{(u-x)^n}{n!} \left(-\frac{\partial}{\partial x}\right)^n \delta(x-x')$$

Inserting this into the result from aboxe, pulling out terms independent of u and swapping

the order of the sum and integration gixes

$$T(x', t|x, t-\tau) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n \delta(x-x') \int (u-x)^n T(u,t|x,t-\tau) du \tag{A.2}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n \delta(x-x') M_n(x,t) \tag{A.3}$$

noticing that $M_n(x,t) = \int (u-x)^n T(u,t|x,t-\tau) du$ is just the nth moment of the transition operator T. Plugging (2.6) back in to (2.4) gixes

$$P(x,t) = \int \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n M_n(x,t)\right) \delta(x-x') P(x,t-\tau) dx \tag{A.4}$$

$$= P(x', t-\tau) + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n [M_n(x,t) P(x,t)] \tag{A.5}$$

Approximating the derixatixe as a finite difference and taking the limit $\tau \to 0$ gixes

$$\dot{P}(x,t) = \lim_{\tau \to 0} \left(\frac{P(x,t) - P(x,t-\tau)}{\tau}\right) \tag{A.6}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n [M_n(x,t) P(x,t)] \tag{A.7}$$

which is formally known as the Kramers-Moyal (KM) expansion. The Fokker-Planck equation is a special case of (2.10) where we neglect terms $n > 2$ in the *diffusion approximation.*

Consider the following Ito stochastic differential equation

$$d\vec{x} = F(\vec{x}, t) + G(\vec{x}, t)dW$$

The SDE gixen aboxe corresponds to the Kramers-Moyal expansion (KME) of a transition density $T(x', t'|x, t)$ see (Risken 1989) for a full derixation.

$$\frac{\partial P(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n [M_n(x,t)P(x,t)] \tag{A.8}$$

where M_n is the nth moment of the transition density. In the diffusion approximation, the KME becomes the Fokker-Planck equation (FPE) (Risken 1989). For the sake of demonstration, consider the univariate case with random variable x and the form of $T(x', t'|x, t)$ is a Gaussian with mean $\mu(t)$ and variance $\sigma^2(t)$. In this scenario, the FPE applies because $M_n = 0$ for all $n > 2$. Given that the drift $M_1(x, t) = \mu(t)$ and the diffusion $M_2(x, t) = \sigma^2(t)$, the FPE reads

$$\frac{\partial P(x,t)}{\partial t} = \left(-\frac{\partial}{\partial x}M^{(1)}(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}M^{(2)}(t)\right) P(x,t) \tag{A.9}$$

We can additionally define the term in parentheses as a differential operator acting on $P(x, t)$

$$\hat{\mathcal{L}}_{FP} = \left(-\frac{\partial}{\partial x}M^{(1)}(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}M^{(2)}(t)\right) \tag{A.10}$$

It is common to additionally define the probability current $J(x, t)$ as

$$J(x,t) = \left(M^{(1)}(t) - \frac{1}{2}\frac{\partial}{\partial x} M^{(2)}(t) \right) P(x,t) \tag{A.11}$$

This definition provides some useful intuition. The value of $J(x,t)$ is the net probability flux into the interval between x and $x + dx$ at at time t. This also allows us to write the FPE as a continuity equation

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x} \tag{A.12}$$

2.1 Poisson processes

The Poisson process can be derived very quickly by noticing that it is simply the continuous-time limit of the Binomial distribution.

$$B(m;t) = \binom{n}{m} \lambda^m (1-\lambda)^{n-m}$$

Since λ is the fraction of successes, the expected number of successes is $\mu = n\lambda$

$$\begin{aligned} B(m;t) &= \binom{n}{m} \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{n-m} \\ &= \binom{n}{m} \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{n} \left(1 - \frac{\mu}{n}\right)^{-m} \end{aligned}$$

$$B(m;t) = \frac{n!}{m!(n-m)!}\left(\frac{\mu}{n}\right)^m\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-m}$$
$$= \frac{n!}{(n-m)!}\left(\frac{1}{n}\right)^m\left(1-\frac{\mu}{n}\right)^{-m}\frac{\mu^m\left(1-\frac{\mu}{n}\right)^n}{m!}$$

In the first term, we can take the first m subterms of the numerator $n! = n(n-1)...(n-m)$ and, since $n >> m$, each term will cancel with one factor of n from the term $1/n^m$. This leaves

$$B(m;t) = \frac{(n-m)!}{(n-m)!}\left(1-\frac{\mu}{n}\right)^{-m}\frac{\mu^m\exp(-\mu)}{m!}$$

We now take the continuous time limit i.e. $n \to \infty$ and, again, since $m << n$ we are left with

$$\lim_{n\to\infty} B(m;t) = \frac{\mu^m\exp(-\mu)}{m!}$$

If an event can be detected will probability γ, the rate of the Poisson process will be reduced by that factor i.e., $\lambda' = \gamma\lambda$. Therefore, the mean and variance of the process becomes $\mu = \gamma\lambda\Delta t$