MATH 5490 - Principles of Stochastic Modeling Notes

Libao Jin

January 17, 2020

Contents

1 Overall Plan

(a) (1D) Analysis of static situation: given data points and find the probability density function.
(b) (1D) Discrete evolution in time (Ch 4, 6, 8, 10): Monte Carlo numerics/Brownian motion.
(c) (1D) Continuous motion in time: in theory (Ito Calculus, Wiener Process).
(d) (Multi-dimensional) Derivation of Navier-Stokes.

Advantages:

- Math
  - Theory of PDF structures.
  - Theory of PDF equation structures (= general diffusion-type PDE).
  - Relationship between stochastic and PDE's.
- Applications (general modeling concept).
  - Molecular flow (Navier Stokes).
  - Turbulent flow.
  - Population dynamics.
  - General diffusion problems.

2 Stochastic States

2.1 Probability Density Functions (PDF's)

2.1.1 Probability Density Functions (PDF's)

(a) Mean Value. Consider a random variable X, there are N measurement values $X_i, i = 1, \dots, N$. A basic characterization of the random variable X is given by the mean $\langle X \rangle$, which tells us which value we may expect for X. Therefore, the mean is often called an expectation value, where $\langle X \rangle$ is given by

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Here, we assume that N is sufficiently large.

(b) Moments. In general, we define a moment of n^{th} order $(n = 1, 2, \dots)$ by

$$\langle X^n \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i^n.$$

(c) Fluctuation. We define the fluctuation $\widetilde{X}_i = X_i - \langle X \rangle$ which is the deviations from the mean.

(d) Central Moments. Fluctuation can be characterized by the central moment (or the moment about the mean) of n^{th} order,

$$\langle \widetilde{X}^n \rangle = \frac{1}{N} \sum_{i=1}^{N} \widetilde{X}_i^n = \frac{1}{N} \sum_{i=1}^{N} (X_i - \langle X \rangle)^n.$$

(e) Distribution Function. Probabilities can be defined by means of theta functions, which are also called step functions or Heaviside functions. The theta function $\theta(z)$ of any variable z

can be defined by

$$\theta(z) = \begin{cases} 0 & \text{if } z < 0, \\ 1 & \text{if } z \geq 0. \end{cases}$$

We may replace z with $z = x - X$. Here, x represent any parameter, and X_i is one measured value of a random variable. Then we obtain

$$\theta(x - X_i) = \begin{cases} 0 & \text{if } x - X_i < 0, \text{ i.e., if } x < X_i, \\ 1 & \text{if } x - X_i \geq 0, \text{ i.e., if } x \geq X_i. \end{cases}$$

The probability for finding a value $X_i \leq x$ is one if $X_i \leq x$, and zero otherwise. Then the probability to find $X \leq x$ is

$$P(X \leq x) = \frac{1}{N} \sum_{i=1}^{N} \theta(x - X_i) = \langle \theta(x - X) \rangle.$$

Then we define the distribution function (cumulative distribution function)

$$F(x) = P(X \leq x) = \langle \theta(x - X) \rangle.$$

Then the probability $P(x \leq X \leq x + \Delta x)$ to find an X value between x and $x + \Delta x$ is given by

$$\begin{aligned} P(x \leq X \leq x + \Delta x) &= \langle \theta(x + \Delta - X) - \theta(x - X) \rangle \\ &= P(X \leq x + \Delta x) - P(X \leq x) \\ &= F(x + \Delta x) - F(x). \end{aligned}$$

(f) Distribution Function Properties.

- $P(x \leq X \leq x + \Delta x) \geq 0 \implies F(x + \Delta x) \geq F(x)$.
- $F(-\infty) = 0$.
- $F(\infty) = 1$.
- $0 \leq F(x) \leq 1$.

(g) Probability Density Function. A probability density function (PDF) is defined as a derivative of the distribution function $F(x)$,

$$f(x) = \frac{dF(x)}{dx} = \frac{d\langle \theta(x - X) \rangle}{dx} = \left\langle \frac{d\theta(x - X)}{dx} \right\rangle.$$

The meaning of the PDF is that $f(x)\, dx$ determines the probability for finding X in an infinitesimal interval between x and $x + dx$,

$$P(x \leq X \leq x + dx) = F(x + dx) - F(x) = f(x)\, dx.$$

Then

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx = \int_a^b \frac{dF(x)}{dx}\, dx = F(b) - F(a).$$

(h) PDF Properties. The PDF $f(x)$ has the following relevant properties,

- $f(x) \geq 0$.
- $f(-\infty) = f(\infty) = 0$.

- $\int f(x)\,dx = 1$.
- $\int g(x)f(x)\,dx = \langle g(X)\rangle$.

  Proof.

$$
\begin{aligned}
\int_{-\infty}^{\infty} g(x)f(x)\,dx &= \int_{-\infty}^{\infty} g(x)\frac{dF(x)}{dx}\,dx \\
&= \int_{-\infty}^{\infty} g(x)\frac{d\langle\theta(x-X)}{dx}\,dx \\
&= \int_{-\infty}^{\infty} g(x)\frac{d}{dx}\frac{1}{N}\sum_{i=1}^{N}\theta(x-X_i)\,dx \\
&= \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty} g(x)\frac{d\theta(x-X_i)}{dx}\,dx \\
&= \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty} g(x)\frac{d\theta(x-X_i)}{d(x-X_i)}\frac{d(x-X_i)}{dx}\,dx \\
&= -\frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty} g(x)\frac{d\theta(x-X_i)}{dX_i}\,dx \\
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{d}{dX_i}\int_{-\infty}^{\infty} g(x)\theta(x-X_i)\,dx \\
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{d}{dX_i}\lim_{L\to\infty}\int_{X_i}^{L} g(x)\cdot 1\,dx \\
&= -\frac{1}{N}\sum_{i=1}^{N}\lim_{L\to\infty}\frac{d}{dX_i}\int_{X_i}^{L} g(x)\,dx \\
&= \frac{1}{N}\sum_{i=1}^{N} g(X_i) \\
&= \langle g(X)\rangle.
\end{aligned}
$$

□

2.1.2 Delta Functions

The theta function was used for the representation of the distribution function $F(x)$. The PDF $f(x) = \langle d\theta(x-X)/dx\rangle$ was introduced as mean of the derivative of a theta function.

(a) Theta and Delta Functions as Limits. Consider the function $\theta_N(x)$ defined as follows,

$$\theta_N(x) = \frac{1+\tanh(Nx)}{2},$$

where $\tanh(x) = \sinh(x)/\cosh(x) = (e^x - e^{-x})/(e^x + e^{-x})$. Then taking the limit gives the theta function

$$\theta(x) = \lim_{N\to\infty}\theta_N(x) = \lim_{N\to\infty}\frac{1}{2} + \frac{1}{2}\frac{e^{Nx}-e^{-Nx}}{e^{Nx}+e^{-Nx}} = \begin{cases} 0 & \text{if } x<0, \\ \frac{1}{2} & \text{if } x=0, \\ 1 & \text{if } x>0. \end{cases}$$

To obtain $d\theta(x-X)/dx$, we consider the derivative of $\theta_N(x)$ with respect to x, and note that

$$\begin{aligned}
\frac{d\tanh(Nx)}{dx} &= \frac{d}{dx}\frac{e^{Nx}-e^{Nx}}{e^{Nx}+e^{Nx}} \\
&= \frac{d}{dx}\frac{e^{2Nx}-1}{e^{2Nx}+1} \\
&= \frac{2Ne^{2Nx}[(e^{2Nx}+1)-(e^{2Nx}-1)]}{(e^{2Nx}+1)^2} \\
&= \frac{4Ne^{2Nx}}{(e^{2Nx}+1)^2} \\
&= N\left(\frac{2e^{Nx}}{e^{2Nx}+1}\right)^2 \\
&= N\left(\frac{2}{e^{Nx}+e^{-Nx}}\right)^2 \\
&= \frac{N}{\cosh^2(Nx)}.
\end{aligned}$$

Then

$$\delta_N(x) = \frac{d\theta_N(x)}{dx} = \frac{d}{dx}\frac{1+\tanh(Nx)}{2} = \frac{1}{2}\frac{d\tanh(Nx)}{dx} = \frac{N}{2\cosh^2(Nx)}.$$

Then we obtain the delta function (or Dirac delta function or Dirac function)

$$\delta(x) = \lim_{N\to\infty}\delta_N(x) = \lim_{N\to\infty}\frac{d\theta_N(x)}{dx} = \frac{d}{dx}\lim_{N\to\infty}\theta_N(x) = \frac{d\theta(x)}{dx}.$$

To be more exact,

$$\delta(x) = \lim_{N\to\infty}\delta_N(x) = \lim_{N\to\infty}\frac{N}{2\cosh^2(Nx)} = \lim_{N\to\infty}\frac{2N}{(e^{Nx}+e^{-Nx})^2} = \begin{cases}\infty & x=0,\\ 0 & x\neq 0.\end{cases}$$

(b) Properties.

- $\int_{-\infty}^{\infty}\delta(x)\,dx = \int_{-\infty}^{\infty}\frac{d\theta(x)}{dx}\,dx = \theta(\infty)-\theta(-\infty) = 1.$
- Sifting property. $\int_{-\infty}^{\infty} g(x)\delta(x-a)\,dx = g(a)\int_{-\infty}^{\infty}\delta(x-a)\,dx = g(a).$

  Proof.

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x)\delta(x-a)\,dx &= \int_{-\infty}^{\infty} g(x)\frac{d\theta(x-a)}{dx}\,dx \\
&= \int_{-\infty}^{\infty} g(x)\cdot\left[-\frac{d\theta(x-a)}{da}\right]dx \\
&= -\frac{d}{da}\int_{-\infty}^{\infty} g(x)\theta(x-a)\,dx \\
&= -\frac{d}{da}\int_{a^-}^{\infty} g(x)\,dx \\
&= -\frac{d}{da}\int_{a}^{\infty} g(x)\,dx \\
&= g(a).
\end{aligned}$$

□

- Symmetry. $\int_{-\infty}^{\infty} g(x)\delta(-x)\,dx = \int_{-\infty}^{\infty} g(x)\delta(x)\,dx$.

(c) PDF Definition. By the PDF definition $f(x) = \langle d\theta(x-X)/dx \rangle$, then $f(x) = \langle \delta(x-X) \rangle$.

- $f(\pm\infty) = \langle \delta(\pm\infty - X) \rangle = 0$.
- $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \langle \delta(x-X) \rangle\,dx = \langle \int_{-\infty}^{\infty} \delta(x-X)\,dx \rangle = \langle 1 \rangle = 1$.
- $\int_{-\infty}^{\infty} g(x)f(x)\,dx = \int_{-\infty}^{\infty} g(x)\langle \delta(x-X) \rangle\,dx = \langle \int_{-\infty}^{\infty} g(x)\delta(x-X)\,dx \rangle = \langle g(X) \rangle$.

2.1.3 An Example: The Uniform Probability Density Function

(a) Uniform PDF. A uniform PDF is given as follows,

$$f(x) = \begin{cases} 1/(b-a) & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

This PDF satisfies the normalization condition to integrate to unity,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b dx = 1.$$

(b) Probability. The distribution function $F(x)$ is

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} \frac{\theta(b-y)-\theta(a-y)}{b-a}\,dy = \frac{\min(x,b)-\min(x,a)}{b-a}.$$

Then

$$P(c \le X \le d) = F(d) - F(c) = \frac{\min(d,b)-\min(d,a)-\min(c,b)+\min(c,a)}{b-a}.$$

(c) Moments. The mean value of a uniform PDF $f(x)$ is given by

$$\langle X \rangle = \int x f(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2},$$

which implies $\langle X \rangle$ is the mean position between a and b. The central moments is

$$\begin{aligned}
\langle \widetilde{X}^k \rangle &= \int_{-\infty}^{\infty} (x - \langle X \rangle)^k f(x)\,dx \\
&= \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^k dx \\
&= \frac{1}{b-a}\frac{1}{k+1}\left[\left(b - \frac{a+b}{2}\right)^{k+1} - \left(a - \frac{a+b}{2}\right)^{k+1}\right] \\
&= \frac{1}{b-a}\frac{1}{k+1}\left[\left(\frac{b-a}{2}\right)^{k+1} - \left(\frac{a-b}{2}\right)^{k+1}\right] \\
&= \frac{1}{b-a}\frac{1}{k+1}\left[\left(\frac{b-a}{2}\right)^{k+1} - (-1)^{k+1}\left(\frac{b-a}{2}\right)^{k+1}\right] \\
&= \begin{cases} \frac{1}{(k+1)}\left(\frac{b-a}{2}\right)^k & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}
\end{aligned}$$

In other words,

$$\langle \widetilde{X}^{2k} \rangle = \frac{1}{2k+1}\left(\frac{b-a}{2}\right)^{2k}, \langle \widetilde{X}^{2k+1} \rangle = 0, \forall k \in \mathbb{N}.$$

Specifically, the variance (the second order central moment) and the standard deviation is as follows, respectively,

$$\langle \widetilde{X}^2 \rangle = \frac{1}{3}\left(\frac{b-a}{2}\right)^2, \langle \widetilde{X}^2 \rangle^{1/2} = \frac{b-a}{2\sqrt{3}}.$$

(d) Parameters. For given values of the man and standard deviation we can calculate the model parameters a and b as follows,

$$\begin{aligned} \langle X \rangle &= \frac{a+b}{2} \implies a+b = 2\langle X \rangle, \\ \langle \widetilde{X}^2 \rangle^{1/2} &= \frac{b-a}{2\sqrt{3}} \implies b-a = 2\sqrt{3}\langle \widetilde{X}^2 \rangle^{1/2}. \end{aligned}$$

Then we can determine a and b,

$$b = \langle X \rangle + \sqrt{3}\langle \widetilde{X}^2 \rangle^{1/2}, a = \langle X \rangle - \sqrt{3}\langle \widetilde{X}^2 \rangle^{1/2}.$$

(e) Random Number Generation. To generate random numbers that have a specified distribution function $G(y)$. Then, it turns out that the random variable $Y = G^{-1}(X)$, which can be determined by solving the relation $G(Y) = X$ for Y, has the given distribution function $G(y)$ as follows,

$$F(y) = P(Y \le y) = P(G^{-1}(X) \le y) = P(X \le G(y)) = \begin{cases} 0 & \text{if } G(y) < 0, \\ G(y) & \text{if } 0 \le G(y) \le 1, \\ 1 & \text{if } G(y) > 1. \end{cases}$$

This method is called the inverse transformation method.

2.2 Models for Probability Density Function

How it is possible to find specific PDF shapes for certain observations?

(a) Design PDFs on the basis of any principle (e.g., the constraint considered below that the PDF has to maximize the uncertainty).
(b) Apply empirical PDF shapes that have desired properties.

2.2.1 Statistically Most-Likely Probability Density Functions

(a) Predictability. For the development of PDF models it is helpful to relate the shape of a PDF to a measure that characterizes the predictability of the state of a random variable. The consideration of the predictability of states of random variables can be used in the following way for the construction of PDFs. Apply information about the known moments combined with the constraint that the predictability (uncertainty) related to the PDF has to be minimal (maximal). First, we reduce our uncertainty by the given information (the known moments). Second, we are maximally uncommitted with respect to the missing information (the PDF shape).

(b) Measure of Uncertainty. The measure of uncertainty S, which is called entropy, is defined by

$$S = -\int_{-\infty}^{\infty} f(x)\ln(f(x))\,dx.$$

The combination of the definition of S with the uniform PDF shape shows that

$$S = -\frac{1}{b-a}\int_a^b \ln\left(\frac{1}{b-a}\right)\,dx = \frac{\ln(b-a)}{b-a}(b-a) = \ln(b-a) = \ln L,$$

where $L = b - a$. This expression illustrates the suitability of using S as a measure of uncertainty: the certainty S is minimal for $L \to 0$, and the uncertainty is maximal for $L \to \infty$.

(c) Statistically Most-Likely PDF. Assume that moments of n^{th}-order, $n = 1, 2, \ldots, s$ is given,

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n f(x)\,dx. \tag{2.1}$$

The goal is to construct a PDF that has s moments that agree with the given ones but maximizes the entropy S (i.e., the uncertainty). According to the calculus of variations, we extend the entropy S to a functional S^* by involving (2.1):

$$S^* = -\int_{-\infty}^{\infty}\left[f(x)\ln(f(x)) + \sum_{k=0}^{s}\mu_k x^k f(x) - f(x)\right]\,dx,$$

where μ_k are Lagrange multipliers which have to be chosen such that $\langle X^n \rangle$ is satisfied for all n. Note that the last term $f(x)$ modifies the multiplier μ_0. To find maximum of S^*, we can calculate the functional variation of S^* with regard to f,

$$\frac{\partial S^*}{\partial f} = -\int_{-\infty}^{\infty}\left[\ln(f(x)) + \sum_{k=0}^{s}\mu_k x^k\right]\,dx.$$

The functional variation is zero if the PDF $f(x)$ is given by

$$f(x) = \exp\left(-\sum_{k=0}^{s}\mu_k x^k\right),$$

which is called a statistically most-likely (SML) PDF. By introducing non-dimensional Lagrange multipliers λ_k we can also write the latter relation as

$$f(x) = \exp\left[\lambda_0 - \sum_{k=1}^{s}\frac{\lambda_k(x - \langle X \rangle)^k}{\langle \tilde{X}^2 \rangle^{k/2}}\right].$$

The $s + 1$ factors λ_k are uniquely determined by the normalization of $f(x)$ and s conditions (2.1). These conditions for the PDF can be written

$$\langle \tilde{X}^n \rangle = \int_{-\infty}^{\infty}(x - \langle X \rangle)^n f(x)\,dx,$$

where $n = 0, 1, \ldots, s$.

2.2.2 The Normal Probability Density Function

(a) Second-Order SML PDF. The normal PDF $f(x)$ of one random variable X is a second-order SML PDF. This PDF is given by

$$f(x) = \exp\left\{\lambda_0 - \lambda_1 \frac{x - \langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}} - \frac{\lambda_2}{2}\frac{(x-\langle X\rangle)^2}{\langle \widetilde{X}^2\rangle}\right\}.$$

The parameters λ_0, λ_1, and λ_2 have to be chosen such that the PDF satisfies for $n = 0, 1, 2$,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)\,dx, \\ 0 &= \int (x - \langle x\rangle) f(x)\,dx, \\ \langle \widetilde{X}^2\rangle &= \int_{-\infty}^{\infty} (x - \langle X\rangle)^2 f(x)\,dx. \end{aligned}$$

for any given mean $\langle X\rangle$ and variance $\langle \widetilde{X}^2\rangle$.

(b) Parameter calculation. A simple way to calculate the model parameters is to differentiate $f(x)$,

$$\frac{df}{dx} = \frac{f}{\langle \widetilde{X}^2\rangle^{1/2}}\left(-\lambda_1 - \lambda_2 \frac{x-\langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}}\right).$$

The integration of this relation leads to

$$\int_{-\infty}^{\infty} \frac{df}{dx}\,dx = -\frac{1}{\langle \widetilde{X}^2\rangle^{1/2}}\int_{-\infty}^{\infty}\left(\lambda_1 + \lambda_2 \frac{x-\langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}}\right) f\,dx = -\frac{\lambda_1}{\langle \widetilde{X}^2\rangle^{1/2}} \implies \lambda_1 = 0.$$

Integrate $(x - \langle X\rangle)f$, we find

$$\int_{-\infty}^{\infty} (x - \langle X\rangle)\frac{df}{dx}\,dx = -\frac{\lambda_2}{\langle \widetilde{X}^2\rangle}\int_{-\infty}^{\infty} (x-\langle X\rangle)^2 f\,dx = -\lambda_2.$$

Using integration by parts to the left-hand side, we have

$$\int_{-\infty}^{\infty} (x-\langle X\rangle)\frac{df}{dx}\,dx = \int \left[\frac{d(x-\langle X\rangle)f}{dx} - \frac{d(x-\langle X\rangle)}{dx} f\right] dx = (x-\langle X\rangle)f\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f\,dx = -1.$$

Then we can conclude that $\lambda_2 = 1$. The parameter λ_0 can be calculated by means of the condition,

$$1 = e^{\lambda_0}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\frac{(x-\langle X\rangle)^2}{\langle \widehat{X}^2\rangle}\right] dx.$$

Let us introduce the variable $y = (x - \langle X\rangle)/(2\langle \widehat{X}^2\rangle)^{1/2}$ to simplify the latter relation,

$$e^{-\lambda_0} = \sqrt{2\langle \widetilde{X}^2\rangle}\int_{-\infty}^{\infty} e^{-y^2}\,dy = \sqrt{2\pi\langle \widetilde{X}^2\rangle}.$$

Then the second-order SML PDF can be written as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Here, the model parameters μ and σ are given by

$$\mu = \langle X \rangle, \sigma = \langle \widetilde{X}^2 \rangle^{1/2}.$$

The PDF is a symmetric function about the mean. Thus, all the odd central moments disappear $(k = 1, 2, \ldots)$, $\langle \widetilde{X}^{2k-1} \rangle = 0$. The even central moments are determined by the formula $(k = 1, 2, \ldots)$, $\langle \widetilde{X}^{2k} \rangle = \frac{(2k)!}{2^k k!} \sigma^{2k}$.

(c) Skewness and Flatness. To see whether a random data set can be described by a normal PDF, it is helpful to use normalized moments of third-order (the skewness) and fourth-order (the flatness or kurtosis) as a reference.

$$m_3 = \frac{\langle \widetilde{X}^3 \rangle}{\langle \widetilde{X}^2 \rangle^{3/2}}, m_4 = \frac{\langle \widetilde{X}^4 \rangle}{\langle \widetilde{X}^2 \rangle^2}.$$

The skewness m_3 indicates deviations from the symmetry of fluctuations about the mean value.The flatness m_4 indicates the peak of a PDF. For the normal PDF, $m_3 = 0$ and $m_4 = 3$. A PDF with $m_4 \geq 3$ $(m_4 \leq 3)$ has a higher (lower) peak value than the normal PDF. Also, m_3 and m_4 have to satisfy

$$m_3^2 + 1 \leq m_4.$$

(d) Probability. The probability $P(a \leq X \leq b)$ for finding the random variable X between a and b is defined by

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] dx.$$

Let $A = (a - \mu)/(\sqrt{2}\sigma)$, $B = (b - \mu)/(\sqrt{2}\sigma)$, and $y = (x - \mu)^{\sqrt{2}\sigma}$. We can simplify the probability as follows,

$$\begin{aligned} P(a \leq X \leq b) &= \frac{1}{\sqrt{\pi}} \int_A^B e^{-y^2}\, dy \\ &= \frac{1}{\sqrt{\pi}} \left[\int_A^0 e^{-y^2}\, dy + \int_0^B e^{-y^2}\, dy \right] \\ &= \frac{1}{\sqrt{\pi}} \left[\int_0^B e^{-y^2}\, dy - \int_0^A e^{-y^2}\, dy \right] \\ &= \frac{1}{2} \left[\operatorname{erf}(B) - \operatorname{erf}(A) \right], \end{aligned}$$

where the error function $\operatorname{erf}(x)$ is defined as follows,

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2}\, ds.$$

(e) Error Function Approximation. The errof function $\operatorname{erf}(x)$ can be approximated below,

$$E(x) = \pm\sqrt{1 - e^{-Hx^2}}a, H = \frac{4/\pi + px^2}{1 + px^2}, p = -\frac{8}{3\pi}\frac{\pi - 3}{\pi - 4} = 0.1400.$$

(f) Example Probabilities.

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= \frac{\operatorname{erf}(1/2^{1/2}) - \operatorname{erf}(-1/2^{1/2})}{2} = \operatorname{erf}(2^{-1/2}) = 0.683, \\ P(\mu - \sigma \leq X \leq \mu + \sigma) &= \frac{\operatorname{erf}(2/2^{1/2}) - \operatorname{erf}(-2/2^{1/2})}{2} = \operatorname{erf}(2/2^{1/2}) = 0.955, \\ P(\mu - \sigma \leq X \leq \mu + \sigma) &= \frac{\operatorname{erf}(3/2^{1/2}) - \operatorname{erf}(-3/2^{1/2})}{2} = \operatorname{erf}(3/2^{1/2}) = 0.995. \end{aligned}$$

(g) Application of the Normal PDF. The normal distribution is used for a variety of applications, for example the distribution of

- intelligence quotients, where $\mu = 100, \sigma = 15$.
- heights of adult males in the U.S., where $\mu = 1.75\text{m}$, $\sigma = 0.07\text{m}$.
- test scores, where, e.g., $\mu = 80\%$, $\sigma = 10\%$.

2.2.3 The Gamma Probability Density Function

(a) Gamma PDF. The PDF of non-negative random variables is often modeled by the gamma PDF. This PDF is defined by

$$f(x) = \begin{cases} \dfrac{b}{\Gamma(a)\langle X\rangle}\left(\dfrac{bx}{\langle X\rangle}\right)^{a-1}\exp\left(-\dfrac{bx}{\langle X\rangle}\right) & \text{if } 0 \leq x, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Gamma(a)$ is the gamma function that is defined by the integral

$$\Gamma(a) = \int_0^\infty y^{a-1}e^{-y}\,dy,$$

which cannot be solved analytically, but it can be approximated by

$$\Gamma(a) = \sqrt{\frac{2\pi}{a}}\left(a\sinh\left(\frac{1}{a}\right) + \frac{1}{810a^6}\right)^a.$$

The gamma function has the following property,

$$\Gamma(a+1) = a\Gamma(a),$$

which can be used to calculate the gamma function at higher a — the accuracy of $\Gamma(a)$ increases with a. The above property can be proven by integration by parts,

$$\int_0^\infty \frac{dy^a e^{-y}}{dy}dy = 0 = a\int_0^\infty y^{a-1}e^{-y}\,dy - \int_0^\infty y^a e^{-y}\,dy.$$

The appearance of the gamma function can be seen by proving that the gamma PDF satisfies the normalization condition,

$$\int_0^\infty f(x)\,dx = \frac{b}{\Gamma(a)\langle X\rangle}\int_0^\infty \left(\frac{bx}{\langle X\rangle}\right)^{a-1}\exp\left(-\frac{bx}{\langle X\rangle}\right)dx = \frac{1}{\Gamma(a)}\int_0^\infty y^{a-1}e^{-y}\,dy = 1,$$

where the substitution $y = bx/\langle X\rangle$ was applied.

(b) Moments. The moments of $f(x)$ can be calculated in this way ($k = 1, 2, \ldots$),

$$\begin{aligned}\langle X^k\rangle &= \int_0^\infty x^k f(x)\,dx = \frac{b}{\Gamma(a)\langle X\rangle}\frac{\langle X\rangle^k}{b^k}\int_0^\infty \left(\frac{bx}{\langle X\rangle}\right)^k\left(\frac{bx}{\langle X\rangle}\right)^{a-1}\exp\left(-\frac{bx}{\langle X\rangle}\right)dx \\ &= \frac{1}{\Gamma(a)}\frac{\langle X\rangle^k}{b^k}\int_0^\infty y^{a+k-1}e^{-y}\,dy = \frac{\langle X\rangle^k}{b^k}\frac{\Gamma(a+k)}{\Gamma(a)},\end{aligned}$$

where $y = bx/\langle X\rangle$. The ratio of gamma functions, which is often denoted by Pochmammer's symbol $(a)_k = \Gamma(a+k)/\Gamma(a)$, can be rewritten by making use of $\Gamma(a+1) = a\Gamma(a)$,

$$\frac{\Gamma(a+k)}{\Gamma(a)} = \frac{(a+k-1)\Gamma(a+k-1)}{\Gamma(a)} = \frac{(a+k-1)(a+k-2)\Gamma(a+k-2)}{\Gamma(a)} = (a+k-1)(a+k-2)\cdots a.$$

Hence, the moments of $f(x)$ are given by

$$\langle X^k \rangle = \frac{(a+k-1)(a+k-2)\cdots a}{b^k} \langle X \rangle^k.$$

(c) Parameters. The mean is given by

$$\langle X \rangle = \frac{a}{b} \langle X \rangle,$$

and the second-order moment is

$$\langle X^2 \rangle = \frac{(a+1)a}{b^2} \langle X \rangle^2.$$

Then the variance is

$$\langle \widetilde{X}^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2 = \frac{a}{b^2} \langle X \rangle^2.$$

Then we can find a, b by the above relations,

$$a = b = \frac{\langle X \rangle^2}{\langle \widetilde{X}^2 \rangle}.$$

Therefore, we can write the gamma PDF for $x \geq 0$ as

$$f(x) = \frac{a}{\Gamma(a) \langle X \rangle} \left(\frac{ax}{\langle X \rangle} \right)^{a-1} \exp\left(-\frac{ax}{\langle X \rangle} \right).$$

(d) Non-Normality. The skewness m_3 and flatness m_4 implied by the gamma PDF are given by the expression

$$m_3 = -\frac{2}{\sqrt{a}}, m_4 = 3\left(1 + \frac{2}{a}\right).$$

For the finite values of the parameter a we find that both m_3 and m_4 are unequal and bigger than the corresponding values of a normal PDF. The values $m_3 = 0$ and $m_4 = 3$ for a normal PDF are recovered in the limit that $a \to \infty$.

(e) Probability Calculation. The probability cannot be performed analytically, so handle this by following the approach used for the calculation of the integral over the normal PDF. Then we have

$$\begin{aligned} P(c \leq X \leq d) &= \frac{a}{\Gamma(a)\langle X \rangle} \int_c^d \left(\frac{ax}{\langle X \rangle} \right)^{a-1} \exp\left(-\frac{ax}{\langle X \rangle} \right) dx \\ &= \frac{1}{\Gamma(a)} \int_C^D y^{a-1} e^{-y} \, dy \\ &= \frac{1}{\Gamma(a)} \left[\int_0^D y^{a-1} e^{-y} \, dy - \int_0^C y^{a-1} e^{-y} \, dy \right] \\ &= \frac{\Gamma_I(a, D) - \Gamma_I(a, C)}{\Gamma(a)}, \end{aligned}$$

where $y = ax/\langle X \rangle$ and the bounds $C = ac/\langle X \rangle$ and $D = ad/\langle X \rangle$, and the incomplete gamma function

$$\Gamma_I(a, x) = \int_0^x y^{a-1} e^{-y} \, dy,$$

which recovers the gamma function when $x \to \infty$. Note that the incomplete gamma can be approximated using the series expansion as follows:

$$\Gamma_I(a, x) \approx x^a \sum_{n=0}^{\infty} \frac{(-x)^n}{(a+n)n!}.$$

2.2.4 The Beta Probability Density Function

(a) Beta PDF. A beta PDF can apply to cases where the random variable is non-negative and has a finite range of variations, it is defined as follows,

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

where the beta function $B(a, b)$ is

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Note that

$$\int_0^1 f(x)\,dx = \frac{1}{B(a,b)} \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = 1.$$

(b) Probability Calculation. To calculate probabilities $P(c \le X \le d)$ we apply beta PDF in the probability definition,

$$\begin{aligned} P(c \le X \le d) &= \frac{1}{B(a,b)} \int_c^d x^{a-1}(1-x)^{b-1}\,dx \\ &= \frac{1}{B(a,b)} \left[\int_0^d x^{a-1}(1-x)^{b-1} - \int_0^1 x^{a-1}(1-x)^{b-1}\,dx \right] \\ &= \frac{B_I(a,b,d) - B_I(a,b,c)}{B(a,b)}, \end{aligned}$$

where $0 \le c, d \le 1$, and the incomplete beta function is

$$B_I(a,b,x) = \int_0^x y^{a-1}(1-y)^{b-1}\,dy,$$

which recovers the beta function $B(a, b)$ for $x = 1$. Similar to the incomplete gamma function, incomplete beta function can also be approximated using series expansion as below,

$$B_I(a,b,x) \approx x^a \sum_{n=0}^{\infty} \frac{\Gamma(1-b+n)}{\Gamma(1-b)} \frac{x^n}{(a+n)n!} = x^a \sum_{n=0}^{\infty} \frac{p_n x^n}{a+n},$$

where $p_n = \Gamma(1-b+n)/[\Gamma(1-b)n!\,]$.

2.2.5 SML-PDF: Beta and Gamma function

(a) Beta PDF.

$$\begin{aligned} \langle \ln x \rangle &= \int_0^n \ln x f(x)\,dx, \\ \langle \ln(1-x) \rangle &= \int_0^n \ln(1-x) f(x)\,dx. \end{aligned}$$

Then,

$$S^* = -\int \{ f \ln f - f - \mu_0 f + \mu_1 \ln x f + \mu_2 \ln(1-x) f \}\,dx.$$

then

$$\begin{aligned}
\frac{\partial S^*}{\partial f} &= -\int \{\ln f + 1 - 1 + \mu_0 + \mu_1 \ln x + \mu_2 \ln(1-x)\}\, dx \\
&= 0 \implies \\
\ln f &= -\mu_0 - \mu_1 \ln x - \mu_2 \ln(1-x), \\
f &= e^{-\mu_0 - \mu_1 \ln x - \mu_2 \ln(1-x)} = e^{-\mu_0} x^{-\mu_1}(1-x)^{\mu_2} = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}.
\end{aligned}$$

(b) Gamma PDF.

$$\begin{aligned}
\langle x \rangle &= \int_0^\infty x f(x)\, dx, \\
\langle \ln(x) \rangle &= \int_0^\infty \ln(x) f(x)\, dx.
\end{aligned}$$

Then,

$$S^* = -\int \{f \ln f - f + \mu_0 f + \mu_1 \ln x f + \mu_2 \ln(x) f\}\, dx.$$

then

$$\begin{aligned}
\frac{\partial S^*}{\partial f} &= -\int \{\ln f + 1 - 1 + \mu_0 + \mu_1 x + \mu_2 \ln(x)\}\, dx \\
&= 0 \\
\implies \ln f &= -\mu_0 - \mu_1 x - \mu_2 \ln(x), \\
f &= e^{-\mu_0 - \mu_1 x - \mu_2 \ln(x)} = e^{-\mu_0} e^{-\mu_1 x} x^{-\mu_2} \iff \frac{a}{\Gamma(a)\langle X \rangle} \left(\frac{ax}{\langle X \rangle}\right)^{a-1} e^{-ax/\langle X \rangle}.
\end{aligned}$$

2.3 Data Analysis

2.3.1 Calculation of Statistics

(a) Filtered PDFs. The definition $f(x) = \langle d\theta(x - X)/dx \rangle$ of a PDF involves a derivative. In order to calculate a PDF from measurements or simulation results, we have to represent the derivative in a discrete way. This can be done by using for the PDF the expression

$$\begin{aligned}
f_\Delta(x) &= \frac{1}{\Delta x} \left\langle \theta\left(x + \frac{\Delta x}{2} - X\right) - \theta\left(x - \frac{\Delta x}{2} - X\right) \right\rangle \\
&= \frac{1}{\Delta x} \int_{x - \Delta x/2}^{x + \Delta x/2} \left\langle \frac{d\theta(y - X)}{dy} \right\rangle dy \\
&= \frac{1}{\Delta x} \int_{x - \Delta x/2}^{x + \Delta x/2} f(y)\, dy \\
&= \frac{1}{\Delta x} \frac{\Delta N}{N},
\end{aligned}$$

where ΔN is defined by

$$\Delta N = \sum_{i=1}^N \left[\theta\left(x + \frac{\Delta x}{2} - X_i\right) - \theta\left(x - \frac{\Delta x}{2} - X_i\right)\right] = \sum_{i=1}^N \begin{cases} 1 - 1 = 0 & \text{if } X_i < x - \Delta x/2, \\ 1 - 0 = 1 & \text{if } x - \Delta x/2 \le X_i \le x + \Delta x/2, \\ 0 - 0 = 0 & \text{if } x + \Delta x/2 < X_i. \end{cases}$$

Consequently, ΔN measures the number of samples that are found in the interval $x - \Delta x/2 \leq X_i \leq x + \Delta x/2$. The PDF $f_\Delta(x) = \Delta N/(N\Delta x)$ represents, therefore, the relative number of samples around x normalized by the filter interval Δx.

(b) Properties of Filtered PDFs. For any functions $g(x)$, the filtered PDF $f_\Delta(x)$ has the property

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x) f_\Delta \, dx &= \frac{1}{\Delta x} \left\langle \int_{-\infty}^{\infty} g(x) \left[\theta\left(x - \left[X - \frac{\Delta x}{2}\right]\right) - \theta\left(x - \left[X + \frac{\Delta x}{2}\right]\right)\right] dx \right\rangle \\
&= \frac{1}{\Delta x} \left\langle \int_{X-\Delta x/2}^{\infty} g(x)\, dx - \int_{X+\Delta x/2}^{\infty} g(x)\, dx \right\rangle \\
&= \frac{1}{\Delta x} \left\langle \int_{X-\Delta x/2}^{\infty} g(x)\, dx + \int_{\infty}^{X+\Delta x/2} g(x)\, dx \right\rangle \\
&= \frac{1}{\Delta x} \left\langle \int_{X-\Delta x/2}^{X+\Delta x/2} g(x)\, dx \right\rangle \\
&= \langle g_\Delta(X) \rangle.
\end{aligned}$$

- By setting $g = 1$, we find that $f_\Delta(x)$ represents indeed a PDF because it integrates to one,
$$\int_{-\infty}^{\infty} f_\Delta(x)\, dx = \frac{1}{\Delta x} \left\langle X + \frac{\Delta x}{2} - \left(X - \frac{\Delta x}{2}\right) \right\rangle = 1.$$
- By setting $g = x$ we find
$$\int_{-\infty}^{\infty} x f_\Delta(x)\, dx = \frac{1}{\Delta x} \left\langle \frac{1}{2}\left(X + \frac{\Delta x}{2}\right)^2 - \frac{1}{2}\left(X - \frac{\Delta x}{2}\right)^2 \right\rangle = \frac{1}{2\Delta x} \langle 2X\Delta x \rangle = \langle X \rangle.$$
- By setting $g = (x - \langle X \rangle)^2$, we obtain
$$\begin{aligned}
\int_{-\infty}^{\infty} (x - \langle X \rangle)^2 f_\Delta(x)\, dx &= \frac{1}{3\Delta x} \left\langle \left(X - \langle X \rangle + \frac{\Delta x}{2}\right)^3 - \left(X - \langle X \rangle - \frac{\Delta x}{2}\right)^3 \right\rangle \\
&= \frac{\langle 6\widetilde{X}^2 \Delta x/2 + 2(\Delta x/2)^3 \rangle}{3\Delta x} \\
&= \langle \widetilde{X}^2 \rangle + \frac{(\Delta x)^2}{12}.
\end{aligned}$$

(c) Sample Number Effect. An increasing number of samples results in a much smoother PDF.

(d) Filter Interval Effect. For a relatively low number N of samples, the need to work with relatively smooth PDFs requires the use of a relatively large filter width Δx to have a sufficient number of samples in the intervals.

2.3.2 The First Fundamental Theorem of Probability

(a) The Problem with Randomness. The calculation of a mean value based on a finite number of samples will provide different results depending on the number N of sample values applied. We would like to known under which conditions we will have exact results that can be reproduced.

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

(b) The Law of Large Numbers. Bernoulli's theorem states that an infinite sequence of independent and identically distributed random numbers $X_1, X_2, \dots$ (the random numbers are independent and each variable has the same PDF) with finite mean converges to a mean $\langle X \rangle$,

$$\lim_{N\to\infty} \mu_N = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i = \langle X \rangle.$$

(c) Illustration. Consider the mean and standard deviation of normally distributed random numbers with $\langle X \rangle = 1$ and standard deviation $\langle \widetilde{X}^2 \rangle^{1/2} = 1$ for a varying number N of samples. The deviations from the exact values $\langle X \rangle = 1$ and $\langle \widehat{X}^2 \rangle^{1/2} = 1$ will be assessed in terms of the relative errors.

$$\Delta_m = \frac{\mu_N - \langle X \rangle}{\langle X \rangle}, \Delta_{sd} = \frac{\sigma_N - \langle \widetilde{X}^2 \rangle^{1/2}}{\langle \widetilde{X}^2 \rangle^{1/2}}.$$

Here $\sigma_N = \langle \widetilde{X}^2 \rangle_N^{1/2}$, where the fluctuation $\widetilde{X} = X - \mu_N$ refers to deviations from the mean μ_N obtained for a finite number of samples. The relative error multiplied with $N^{1/2}$ is independent of N. Then

$$\Delta_m = \Delta_{sd} = \frac{\varepsilon}{N^{1/2}},$$

where ε is a bounded random variable suggests that $|\varepsilon| < 2$. Hence, μ_N and σ_N converge to their exact values $\langle X \rangle$ and $\langle \widetilde{X}^2 \rangle^{1/2}$ proportional to $N^{-1/2}$.

(d) Unbiased Estimates. The unbiased estimate of mean is

$$\langle \mu_N \rangle = \frac{1}{N}\sum_{i=1}^{N} \langle X_i \rangle = \frac{N\langle X \rangle}{N} = \langle X \rangle.$$

And the unbiased estimate for the variance is

$$\sigma_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_N)^2.$$

2.3.3 The Second Fundamental Theorem of Probability

(a) The Generalized Problem with Randomness. What is the limiting behavior of the PDF of a sum of N independent and identically distributed random numbers

$$N\mu_N = X_1 + \cdots + X_N$$

as N approaches infinity

(b) The Central Limit Theorem. Let $X_1, X_2, \dots, X_N$ be a sequence of independent and identically distributed random numbers each having a finite mean μ and a finite variance $\sigma^2 > 0$. Further, let D_N be the PDF of the sum of N values X_i. Then, the Central Limit Theorem says that D_N converges independent of the original PDF to a normal distribution with mean $N\mu$ and variance $N\sigma^2$,

$$\sum_{i=1}^{N} X_i \sim D_N \implies \lim_{N\to\infty} D_N = \mathcal{N}(N\mu, N\sigma^2).$$

The Central Limit Theorem states that

$$\frac{1}{N}\sum_{i=1}^{N} X_i \sim d_N \implies \lim_{N\to\infty} d_N = \mathcal{N}(\mu, \sigma^2/N).$$

2.4 Real Distributions

2.4.1 Statistically Most-Likely Probability Density Functions

(a) Four-Order SML PDF. This PDF has a maximal entropy S (uncertainty) among all PDFs for which the first four moments agree with the first four moments of any data set. This PDF is given by

$$f(x) = \exp\left\{\lambda_0 - \lambda_1 \frac{x - \langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}} - \frac{\lambda_2}{2}\frac{(x - \langle X\rangle)^2}{\langle \widetilde{X}^2\rangle} - \frac{\lambda_3}{3}\frac{(x - \langle X\rangle)^3}{\langle \widetilde{X}^2\rangle^{3/2}} - \frac{\lambda_4}{4}\frac{(x - \langle X\rangle)^4}{\langle \widetilde{X}^2\rangle^2}\right\}.$$

This PDF must satisfy the constraint that $f(x)$ integrates to one, and the first four moments of $f(x)$ agree with the first four moments of any data set.

(b) Model Parameter Calculation. Differentiating $f(x)$ gives

$$\frac{df(x)}{dx} = \frac{f(x)}{\langle \widetilde{X}^2\rangle^{1/2}}\left\{-\lambda_1 - \lambda_2\frac{x - \langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}} - \lambda_3\frac{(x - \langle X\rangle)^2}{\langle \widetilde{X}^2\rangle} - \lambda_4\frac{(x - \langle X\rangle)^3}{\langle \widetilde{X}^2\rangle^{3/2}}\right\}.$$

Rearranging the above equation gives

$$\begin{aligned}
\frac{\langle \widetilde{X}^2\rangle^{1/2} df(x)}{\lambda_2 dx} &= -\frac{f(x)}{\lambda_2}\left\{\lambda_1 + \lambda_2\frac{x - \langle X\rangle}{\langle \widetilde{X}^2\rangle^{1/2}} + \lambda_3\frac{(x - \langle X\rangle)^2}{\langle \widetilde{X}^2\rangle} + \lambda_4\frac{(x - \langle X\rangle)^3}{\langle \widetilde{X}^2\rangle^{3/2}}\right\} \\
\implies \frac{(x - \langle X\rangle)f(x)}{\langle \widetilde{X}^2\rangle^{1/2}} &= -\frac{f(x)}{\lambda_2}\left\{\lambda_1 + \lambda_3\frac{(x - \langle X\rangle)^2}{\langle \widetilde{X}^2\rangle} + \lambda_4\frac{(x - \langle X\rangle)^3}{\langle \widetilde{X}^2\rangle^{3/2}}\right\} - \frac{\langle \widetilde{X}^2\rangle^{1/2} df(x)}{\lambda_2 dx}.
\end{aligned}$$

The multiplication of this relation with appropriate powers of $(x - \langle X\rangle)/\langle \widetilde{X}^2\rangle^{1/2}$ and integration then provides the conditions,

$$\begin{aligned}
0 &= \frac{1}{\langle \widetilde{X}^2\rangle^{1/2}}\int_{-\infty}^{\infty}(x - \langle X\rangle)f(x)\,dx = -\frac{\lambda_1 + \lambda_3 + \lambda_4 m_3}{\lambda_2}, \\
1 &= \frac{1}{\langle \widetilde{X}^2\rangle}\int_{-\infty}^{\infty}(x - \langle X\rangle)^2 f(x)\,dx = -\frac{\lambda_3 m_3 + \lambda_4 m_4 - 1}{\lambda_2}, \\
m_3 &= \frac{1}{\langle \widetilde{X}^2\rangle^{3/2}}\int_{-\infty}^{\infty}(x - \langle X\rangle)^3 f(x)\,dx = -\frac{\lambda_1 + \lambda_3 m_4 + \lambda_4 m_5}{\lambda_2}, \\
m_4 &= \frac{1}{\langle \widetilde{X}^2\rangle^2}\int_{-\infty}^{\infty}(x - \langle X\rangle)^4 f(x)\,dx = -\frac{\lambda_1 m_3 + \lambda_3 m_5 + \lambda_4 m_6 - 3}{\lambda_2},
\end{aligned}$$

Simplifying the above conditions gives

$$\begin{aligned}
\lambda_1 + \lambda_3 + \lambda_4 m_3 &= 0, \\
\lambda_2 + \lambda_3 m_3 + \lambda_4 m_4 &= 1, \\
\lambda_1 + \lambda_2 m_3 + \lambda_3 m_4 + \lambda_4 m_5 &= 0, \\
\lambda_1 m_3 + \lambda_2 m_4 + \lambda_3 m_5 + \lambda_4 m_6 &= 3.
\end{aligned}$$

3 Stochastic Changes

3.1 Motivation

(a) Diffusion. Diffusion processes are the processes of which the quantity is distributed until an equilibrium state is established (i.e., until the differences that drive the process are minimized). Examples of diffusion processes:

- Diffusion is responsible for the distribution of sugar throughout a cup of coffee.
- Diffusion is the mechanism by which oxygen moves into our cells.
- Sintering process (powder metallurgy, production of ceramics), the chemical reactor design, catalyst design in the chemical industry, doping during the production of semiconductors, and the transport of necessary materials such as amino acids within biological cells.

(b) Diffusion Model. Consider the model

$$y_n = y_{n-1} + r\varepsilon_{n-1}, n = 1, 2, \ldots,$$

where y_n refers to the position of any particle (for example, the height of any tracer above ground). The initial position y_0 is assumed to be given. Assume that

- y_0 is a deterministic parameter such that all the particles start at the same position;
- ε_k accounts for the effect of randomness and it is assumed to be normally distributed with mean $\langle \varepsilon_k \rangle = 0$ and variance $\langle \widehat{\varepsilon}_k^2 \rangle = 1$, and at each step, the noise process is considered independent, i.e., ε_k and ε_m are independent random variables, $\langle \varepsilon_k \varepsilon_m \rangle = 0$ for $k \neq m$, in other words, $\langle \varepsilon_k \varepsilon_m \rangle = \delta_{km}$, where δ_{km} refers to the Kronecker delta that is defined by

$$\delta_{km} = \begin{cases} 1 & \text{if } k = m, \\ 0 & \text{if } k \neq m. \end{cases}$$

- r is a deterministic parameter that modifies the intensity of randomness.

We can conclude that the solution y_n is given by $(n = 0, 1, \ldots)$

$$y_n = y_{n-1} + r\varepsilon_{n-1} = y_{n-2} + r\varepsilon_{n-2} + r\varepsilon_{n-1} = \cdots = y_0 + r(\varepsilon_0 + \cdots + \varepsilon_{n-1}).$$

3.2 Linear Stochastic Changes

Consider the first-order linear difference equation $(n = 1, 2, \ldots)$

$$y_n = ay_{n-1} + b + r\varepsilon_{n-1},$$

where a, b, and r are any deterministic model parameters. Assume that y_n has a normally distributed initial value y_0 at $n = 0$. The noise ε_k is assumed to be normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$. Also, we have $\langle \varepsilon_k \rangle = 0, \langle \varepsilon_k \varepsilon_m \rangle = \delta_{km}$.

3.2.1 One-Point Statistics

(a) Solution.

$$\begin{aligned}
y_1 &= ay_0 + b + r\varepsilon_0, \\
y_2 &= ay_1 + b + r\varepsilon_1 = a^2 y_0 + a(b + r\varepsilon_0) + b + r\varepsilon_1, \\
&\vdots \\
y_n &= a^n y_0 + \sum_{i=0}^{n-1} a^i (b + r\varepsilon_{n-i-1}),
\end{aligned}$$

(b) Moments.

$$\begin{aligned}
\langle y_n \rangle &= \left\langle a^n y_0 + \sum_{i=0}^{n-1} a^i (b + r\varepsilon_{n-i-1}) \right\rangle \\
&= a^n \langle y_0 \rangle + \sum_{i=0}^{n-1} \langle a^i (b + r\varepsilon_{n-i-1}) \rangle \\
&= a^n \langle y_0 \rangle + b \sum_{i=0}^{n-1} a^i \\
&= \begin{cases} a^n \langle y_0 \rangle + b\dfrac{1-a^n}{1-a} & \text{if } a \neq 1, \\ a^n \langle y_0 \rangle + nb & \text{otherwise.} \end{cases}
\end{aligned}$$

Note that

$$\widetilde{y}_n = y_n - \langle y_n \rangle = a^n \widetilde{y}_0 + \sum_{i=0}^{n-1} a^i r\varepsilon_{n-i-1}.$$

Then

$$\begin{aligned}
\langle \widetilde{y}_n^2 \rangle &= \left\langle \left(a^n \widetilde{y}_0 + \sum_{i=0}^{n-1} a^i r\varepsilon_{n-i-1} \rangle \right)^2 \right\rangle \\
&= a^{2n} \langle \widetilde{y}_0^2 \rangle + r^2 \sum_{i=0}^{n-1} a^{2i} \\
&= \begin{cases} a^{2n} \langle \widetilde{y}_0^2 \rangle + r^2 \dfrac{1-a^{2n}}{1-a^2} & \text{if } a \neq 1, \\ a^{2n} \langle \widetilde{y}_0^2 \rangle + nr^2 & \text{otherwise.} \end{cases}
\end{aligned}$$

(c) PDF. Consider the random numbers $X_i, i = 1, 2, \ldots, N$ that are normally distributed with mean μ_i and variance σ_i^2. Then the sum of X_i is also normally distributed, whose mean is the sum of all μ_i, and the variance is the sum of all variances σ_i^2. Therefore, the PDF of y_n is all normally distributed,

$$f_n(y) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{y - \mu_n}{2\sigma_n^2} \right) = \frac{1}{\sqrt{2\pi \langle \widetilde{y}_n^2 \rangle}} \exp\left(-\frac{y - \langle y_n \rangle}{2\langle \widetilde{y}_n^2 \rangle} \right).$$

3.2.2 Correlations

(a) Correlation Relevance. The consideration of the evolution of a stochastic process leads to the additional question about the typical lifetime of fluctuations. Consider the normalized correlation function

$$C_n(m) = \frac{\langle \widetilde{y}_n, \widetilde{y}_{n+m} \rangle}{\langle \widetilde{y}_n \widetilde{y}_n \rangle},$$

which is equal to the correlation coefficient between y_n and y_m if the variance is stationary (if $\langle \widetilde{y}_n \widetilde{y}_n \rangle = \langle \widetilde{y}_{n+m} \widetilde{y}_{n+m} \rangle$).

(b) Correlation Calculation.

$$\widetilde{y}_n = y_n - \langle y_n \rangle = a y_{n-1} + b + r\varepsilon_{n-1} - (a\langle y_{n-1} + b) = a\widetilde{y}_{n-1} + r\varepsilon_{n-1}.$$

Then

$$\langle \widetilde{y}_n \widetilde{y}_{n+1} \rangle = \langle \widetilde{y}_n (a\widetilde{y}_n + r\varepsilon_n) \rangle = a\langle \widetilde{y}_n \widetilde{y}_n \rangle + r\langle \widetilde{y}_n \varepsilon_n \rangle = a\langle \widetilde{y}_n \widetilde{y}_n \rangle.$$

That implies

$$\langle \widetilde{y}_n \widetilde{y}_{n+m} \rangle = a^m \langle \widetilde{y}_n \widetilde{y}_n \rangle \implies C_n(m) = a^m.$$

3.2.3 Solution Features

(a) Means and Variances.

(b) Correlations.

3.2.4 Monte Carlo Simulation

(a) Example. Consider the equation

$$y_n = a y_{n-1} + b + r\varepsilon,$$

where $a = 0.5$, $b = 1$, $r = 0.8$, and $y_0 = 0$.

(b) One-Point Statistics.

3.3 Diffusion

Consider the random walk models (drunkard's walk models) to illustrate the application of the linear stochastic first-order difference equation. We will focus on the Wiener process,

$$y_n = y_{n-1} + r\varepsilon_{n-1},$$

where ε_{n-1} is normally distribution and characterized by $\langle \varepsilon_k \rangle = 0$ and $\langle \varepsilon_k \varepsilon_m \rangle = \delta_{km}$. We will also assume that ε_k is independent of the random initial position y_0.

3.3.1 Random Walk Model

(a) One-Point Statistics. Let $a = 1$ and $b = 0$. The mean of y_n equals the mean initial value for the case $a = 1$ and $b = 0$ considered,

$$\langle\langle y_n \rangle = \langle y_0 \rangle,$$

and the variance of y_n is given by

$$\langle \widetilde{y}_n^2 \rangle = \langle \widetilde{y}_0^2 \rangle + nr^2. \tag{3.1}$$

Hence, y_n is normally distributed according to

$$y_n \sim \mathcal{N}(\langle y_0 \rangle, \langle \widetilde{y}_0^2 + nr^2).$$

(b) Correlations. The correlations of y_n are determined by

$$\langle \widetilde{y}_n, \widetilde{y}_{n+m} \rangle = \langle \widetilde{y}_n, \widetilde{y}_n \rangle, \tag{3.2}$$

where $m = 0, 1, 2, \ldots$. The meaning of this result can be rewritten as

$$\langle \widetilde{y}_n (\widetilde{y}_{n+m} - \widetilde{y}_n) \rangle = 0.$$

(c) Time Dependence. We may use $y_n = y_{n-1} + r\varepsilon$ as a model for a continuous diffusion in time $n\Delta t$. For this case, the variance (3.1) should be a function of $n\Delta t$. The latter is the case if we parameterize the noise coefficient r by

$$r = \sqrt{D\Delta t},$$

where D is the diffusion coefficient. We can write the variance

$$\langle \widetilde{y}_n^2 \rangle = \langle \widetilde{y}_0^2 \rangle + Dn\Delta t.$$

Therefore, D is the derivative of the variance by time $n\Delta t$. D determines the increase of the position variance (which describes the spreading of a plume). By using $r = (D\Delta t)^{1/2}$, we can write the diffusion model as

$$y_n = y_{n-1} + \sqrt{D\Delta t}\varepsilon_{n-1} \sim \mathcal{N}(\langle y_0 \rangle, \langle \widetilde{y}_0^2 + Dn\Delta t).$$

Equation (3.2) implies for the normalized correlation function $C_n(m)$ that

$$C_n(m) = 1.$$

3.3.2 The Wiener Process

(a) Model Reformulation.

$$W_n = W_{n-1} + \sqrt{\Delta t}\varepsilon_{n-1} \implies \Delta W_{n-1} = \sqrt{\Delta t}\varepsilon_{n-1} \implies y_n = y_{n-1} + \sqrt{D}\Delta W_{n-1}.$$

(b) Wiener Process. The Wiener process W_{n-1} is normally distributed according to

$$W_n \sim \mathcal{N}(\langle W_0, \langle \widetilde{W}_0^2 + n\Delta t).$$

And the correlations of W_n are determined by

$$\langle \widetilde{W}_n, \widetilde{W}_{n+m} \rangle = \langle \widetilde{W}_n, \widetilde{W}_n \rangle.$$

(c) Wiener Process Change. The change $\Delta W_n = \sqrt{\Delta t}\varepsilon_n$ of a Wiener process is normally distributed,

$$\Delta W_n \sim \mathcal{N}(0, \Delta t).$$

The correlation properties of ΔW_n also follow from $\Delta W_n = \sqrt{\Delta t}\varepsilon_n$,

$$\langle \Delta W_n \Delta W_n \rangle = \begin{cases} \Delta t & \text{if } n = m, \\ 0 & \text{if } n \neq m, \end{cases}$$

where we made use of $\langle \varepsilon_k \varepsilon_m \rangle = \delta_{km}$. We applied $\Delta \widetilde{W}_n = \Delta W_n$, which is the same because $\langle \Delta W_n \rangle = 0$. By dividing both sides by $(\Delta t)^2$ we can write

$$\left\langle \frac{\Delta W_n}{\Delta t} \frac{\Delta W_m}{\Delta t} \right\rangle = \begin{cases} \frac{1}{\Delta t} & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

That implies that the variance of the derivative $\Delta W_n / \Delta t$ of W_n does not exist for $\Delta t \to 0$. Consequently, W_n is not differentiable.

3.3.3 Diffusion Model

(a) Definition of Concentration. Consider an instantaneous emission from a point source, i.e., the emission of a mass M at time zero at a fixed position y_0. For this case, the mean concentration C is given by M times the PDF $f_n(y)$ for finding a parcel at a step n at a position y,

$$C_n = M f_n(y) = \frac{M}{\sqrt{2\pi}\sigma_n} \exp\{-\frac{(y - y_0)^2}{2\sigma_n^2},$$

where $\mu = y_0$, $\sigma_n^2 = Dn\Delta t$. The above equation describes the temporal evolution of the mean concentration in one dimension: the y-axis. Note that

$$M = \int_{-\infty}^{\infty} C_n \, dy = \int_{-\infty}^{\infty} \frac{M}{\sqrt{2\pi}\sigma_n} \exp\{-\frac{(y - y_0)^2}{2\sigma_n^2} \, dy.$$

(b) Initial Condition. The best way to calculate the initial concentration is to consider the limit $\sigma_n^2 \to 0$. Then

$$C_0 = M\delta(y - y_0) = \lim_{\sigma_n \to 0} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{(y - y_0)^2}{2\sigma_n^2}\right\}.$$

(c) Boundary Effects. Assume that there is a total reflection of material at $y = 0$, the presence of such a totally reflecting boundary can be taken into account by assuming that there is a hypothetical source at $y = -y_0$. The contributions of the sources at $y = y_0$ and $y = -y_0$, then result in the concentration

$$C_n = \frac{M}{\sqrt{2\pi}\sigma_n} \left[\exp\left\{-\frac{(y - y_0)^2}{2\sigma_n^2}\right\} + \exp\left\{-\frac{(y + y_0)^2}{2\sigma_n^2}\right\}\right].$$

Integrate C_n over the range $0 \leq y < \infty$,

$$\int_0^{\infty} C_n \, dy = \frac{M}{\sqrt{2\pi}\sigma_n} \left[\int_0^{\infty} \exp\left\{-\frac{(y - y_0)^2}{2\sigma_n^2}\right\} dy + \int_0^{\infty} \exp\left\{-\frac{(y + y_0)^2}{2\sigma_n^2}\right\} dy\right] = M.$$

Next, consider the case of a totally absorbing boundary at $y = 0$. The presence of such a boundary can be accounted for by assuming that there is a hypothetical source at $y = y_0$. We have to consider the difference of both distributions to ensure that $C_n = 0$ at the boundary $y = 0$,

$$C_n = \frac{M}{\sqrt{2\pi}\sigma_n} \left[\exp\left\{-\frac{(y - y_0)^2}{2\sigma_n^2}\right\} - \exp\left\{-\frac{(y + y_0)^2}{2\sigma_n^2}\right\}\right].$$

(d) Ground Concentrations. The mean concentration development in time $t = n\Delta t$ can be used to find a corresponding two-dimensional concentration in a x-y plane, where x and y refer to the horizontal and vertical coordinates. Assume that the substance is transported along the x-direction with a constant velocity $U = x_n/(n\Delta t)$. By using the relation $n\Delta t = x_n/U$, the variance $\sigma_n^2 = Dn\Delta t = Dx_n/U$ becomes a function of x_n. Let us calculate the ground concentration at $y = 0$ for the case without boundary to illustrate the use of this approach. The ground concentration for the case without boundary is given by

$$C_n = M\sqrt{\frac{U}{2\pi D x_n}} \exp\left\{-\frac{1}{2}\frac{y_0^2}{Dx_n}\right\}.$$

It is convenient to introduce the non-dimensional positions $x_{*n} = Dx_n/(Uy_0^2)$ and concentration $C_{*n} = C_n y_0/M$. Then, the equation becomes

$$C_{*n} = \frac{1}{\sqrt{2\pi} x_{*n}} \exp\left\{-\frac{1}{2x_{*n}}\right\}.$$

We can find the maximum position of the ground concentration C_{*n} by calculating the first and second derivatives of C_{*n} with respect to x_{*n},

$$\frac{dC_{*n}}{dx_{*n}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2x_{*n}}\right\}\left(\frac{1}{2x_{*n}^{1/2} x_{*n}^2} - \frac{1}{2x_{*n}^{3/2}}\right) = \frac{1-x_{*n}}{2x_{*n}^2} C_{*n}.$$
$$\frac{d^2C_{*n}}{dx_{*n}^2} = \frac{(1-x_{*n})^2}{4x_{*n}^2} C_{*n} + C_{*n}\left(-\frac{1}{x_{*n}^3} + \frac{1}{2x_{*n}^2}\right) = C_{*n}\frac{3x_{*n}^2 - 6x_{*n} + 1}{4x_{*n}^4}.$$

Therefore, the maximal ground concentration can be obtained at $x_{*n} = 1$,

$$C_{*n} = \frac{e^{-1/2}}{\sqrt{2\pi}} \approx 0.242.$$

3.4 Brownian Motion

3.4.1 Brownian Motion Model

(a) Brownian Motion Model. Follow Lagnevin's approach by considering the following stochastic difference equation system,

$$\frac{x_n - x_{n-1}}{\Delta t} = v_{n-1}, \tag{3.3}$$
$$\frac{v_n - v_{n-1}}{\Delta t} = -\frac{1}{\tau}\left(v_{n-1} - \sqrt{D}\frac{\Delta W_{n-1}}{\Delta t}\right), \tag{3.4}$$

where x_n and v_n refer to the position and velocity of a Brownian particle, respectively, D is the diffusion coefficient, τ represents a characteristic time scale, and the change of the Wiener process is defined by $\Delta W_{n-1} = (\Delta t)^{1/2}\varepsilon_{n-1}$. Note that $\frac{\sqrt{D}}{\tau}\frac{\Delta W_{n-1}}{\Delta t}$ provides a random input (as a model for the random impacts of water molecules on a pollen grain), and $-\frac{v_{n-1}}{\tau}$ models the relaxation of the pollen velocity due to the damping influence of surrounding water molecules.

(b) Linear Second-Order Difference Equation. Consider

$$\begin{aligned} v_n &= \left(1 - \frac{\Delta t}{\tau}\right) v_{n-1} + \frac{1}{\tau}\sqrt{D}\Delta W_{n-1} \\ &= \left(1 - \frac{\Delta t}{\tau}\right) \frac{x_n - x_{n-1}}{\Delta t} + \frac{1}{\tau}\sqrt{D}\Delta W_{n-1} \\ &= \left(1 - \frac{\Delta t}{\tau}\right)\left[\left(1 - \frac{\Delta t}{\tau}\right)\frac{x_{n-1} - x_{n-2}}{\Delta t} + \frac{1}{\tau}\sqrt{D}\Delta W_{n-2}\right] + \frac{1}{\tau}\sqrt{D}\Delta W_{n-1}, \end{aligned}$$

where

$$\frac{x_n - x_{n-1}}{\Delta t} = \left(1 - \frac{\Delta t}{\tau}\right)\frac{x_{n-1} - x_{n-2}}{\Delta t} + \frac{1}{\tau}\sqrt{D}\Delta W_{n-2}.$$

Then

$$\begin{aligned}x_n &= x_{n-1} + \left(1 - \frac{\Delta t}{\tau}\right)(x_{n-1} - x_{n-2}) + \frac{1}{\tau}\sqrt{D}\Delta t \Delta W_{n-2}. \\ &= x_{n-1} + \left(1 - \frac{\Delta t}{\tau}\right)(x_{n-1} - x_{n-2}) + \frac{1}{\tau}\sqrt{D\Delta t}\varepsilon_{n-2}.\end{aligned}$$

(c) Comparison with Diffusion Model. The Diffusion Model is given by

$$x_n = x_{n-1} + \sqrt{D\Delta t}\varepsilon,$$

where $r = \sqrt{D\Delta t}$. It is worth noting that $\Delta t = \tau$, the Brownian motion model reduces to the diffusion model.

3.4.2 Discrete Brownian Motion Statistics

(a) Joint PDF. The joint process (x_n, v_n) is normally distributed.
(b) Solution. To find the means and variances of the joint normal PDF of x_n and v_n, we need to compute the solution of (3.3) and (3.4). Let $a = 1 - \Delta t/\tau$ and $r_B = (D\Delta t/\tau^2)^{1/2}$, where r_B is the noise intensity in the Brownian motion velocity equation. Then (3.3) and (3.4) become

$$\begin{aligned}x_n &= x_{n-1} + \Delta t v_{n-1} = \cdots = x_0 + \Delta t(v_0 + v_1 + \cdots + v_{n-1}), \\ v_n &= a v_{n-1} + r_B\varepsilon_{n-1} = a^n v_0 + r_B(a^{n-1}\varepsilon_0 + \cdots + a^1\varepsilon_{n-2} + a^0\varepsilon_{n-1}).\end{aligned}$$

Then we can plug v_n into x_n, then

$$\begin{aligned}x_n &= x_0 + \Delta t\{v_0\frac{1-a^n}{1-a} + r_B\Delta t\left[\varepsilon_0\frac{1-a^{n-1}}{1-a} + \varepsilon_1\frac{1-a^{n-2}}{1-a} + \cdots + \varepsilon_{n-3}\frac{1-a^2}{1-a} + \varepsilon_{n-2}\frac{1-a}{1-a}\right] \\ &=\end{aligned}$$

(c) Model 1:

$$y_n = a y_{n-1} + b + r\varepsilon_{n-1}$$

(d) Model 2:

$$y_n = a y_{n-1} + b + r\varepsilon_{n-1}^2$$

3.5 Population Dynamics

3.5.1 A Stochastic Logitic Model

(a) Logistic Model.

$$P_n = P_{n-1} + aP_{n-1}\left(1 - \frac{P_{n-1}}{K}\right),$$

where K is the carrying capacity, and the model parameter $a = \Delta t/T$ determines the transition rate to the equilibrium state. Here Δt denotes a time interval, and T is a characteristic time scale. There are two equilibrium states $P_n = 0$ and $P_n = K$ that can be realized depending on the initial population P_0: these equilibrium states imply $P_n - P_{n-1} = 0$. Rewriting the equation gives

$$\frac{P_n - P_{n-1}}{\Delta t} = \frac{P_{n-1}}{T}\left(1 - \frac{P_{n-1}}{K}\right).$$

(b) Stochastic Logistic Model. Randomization of K is questionable because the poppulation is negative. The consideratoin of negative and positive values of the growth time T represents an appropriate mean to reflect varying conditionos for a population development. We assume that T^{-1} is normally distributed,

$$T^{-1} = \mu + \sigma \frac{\Delta W_{n-1}}{\Delta t}.$$

We apply $\Delta W_{n-1} = (\Delta t)^{1/2} \varepsilon_{n-1}$ as before, then we have the model

$$\frac{P_n - P_{n-1}}{\Delta t} = P_{n-1}(1 - P_{n-1})\left(\mu + \sigma \frac{\Delta W_{n-1}}{\Delta t}\right).$$

3.5.2 One-Point Statistics and Correlations

4 Stochastic Evolution

4.1 PDF Evolution Equations

4.1.1 The Kramers-Moyal Equation

(a) PDF Definition. The PDF of a random variable X is defined by the expression $f(x) = \langle \delta(x - X) \rangle$ refers to a delta function. The expression $f(x) = \langle \delta(x - X) \rangle$ also can be used for a stochastic process that changes in time. The PDF $f(x,t)$ at the time t is then defined by

$$f(x,t) = \langle \delta(x - X(t)) \rangle.$$

At the later time $t + \Delta t$, the PDF is given by

$$f(x, t + \Delta t) = \langle \delta(x - X(t + \Delta t)) \rangle. \tag{4.1}$$

(b) Kramers-Moyal Equation. Consider the instantaneous PDF involved in (4.1),

$$\delta(x - X(t + \Delta t)) = \delta(z), z = x - X(t + \Delta t).$$

Consider the Taylor expansion

$$\begin{aligned} f(x, t + \Delta t) &= \delta(z) \\ &= \sum_{n=0}^{\infty} \frac{\delta^{(n)}(z_0)}{n!}(z - z_0)^n \\ &= \delta(z_0) + \sum_{n=1}^{\infty} \left(\frac{d}{dx}\right)^n \left[\frac{(-1)^n}{n!}(z_0 - z)^n \delta(z_0)\right] \\ &= f(x,t) + \sum_{n=1}^{\infty} \left(-\frac{d}{dx}\right)^n \frac{(z_0 - z)^n \delta(z_0)}{n!}, \end{aligned}$$

where $\delta^{(n)}$ refers the n^{th}-order derivative of $\delta(z_0)$, $z_0 = x - X(t)$. Then reorder terms we have

$$\lim_{\Delta t \to 0} \frac{f(x, t+\Delta t) - f(x,t)}{\Delta t} = \sum_{n=1}^{\infty} \left(-\frac{d}{dx}\right)^n \lim_{\Delta t \to 0} \frac{(z_0 - z)^n \delta(z_0)}{n!\,\Delta t}.$$

Then, we have

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t) f(x,t),$$

where

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle (z_0 - z)^n \delta(z_0) \rangle}{n!\,\Delta t f(x,t)}.$$

(c) Kramers-Moyal Coefficients. The Kramers-Moyal coefficients $D^{(n)}(x,t)$ can be rewritten by using the definitions $z_0 - z = X(t+\Delta t) - X(t)$ and $z_0 = x - X(t)$,

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle [X(t+\Delta t) - X(t)]^n \delta(x - X(t)) \rangle}{n!\,\Delta t f(x,t)}.$$

The conditional mean is defined for any function $g(X(t))$ by

$$\frac{1}{f(x,t)} \langle g(X(t)) \delta(x - X(t)) \rangle = \langle g(X(t)) | X(t) = x \rangle = \langle g(X(t)) | x, t \rangle.$$

Then the Kramers-Moyal coefficients can be written

$$D^{(n)}(x,t) = \lim_{\Delta t \to \infty} \frac{\langle (X(t+\Delta t) - X(t))^n | x, t \rangle}{n!\,\Delta t}.$$

(d) Markov Process. Stochastic processes for which ΔX does only depend on the previous state $X(t)$ are referred to as Markov processes.

4.1.2 The Pawula Theorem

(a) Pawula's Theorem. Consider a negative function $H(p) \geq 0$ as

$$H(p) = \langle (\Delta X^k + p \Delta X^{k+m})^2 | x, t \rangle = \langle \Delta X^{2k} | x, t \rangle + 2p \langle \Delta X^{2k+m} | x, t \rangle + p^2 \langle \Delta X^{2k+2m} | x, t \rangle,$$

where $\Delta X = X(t+\Delta) - X(t)$, and we assume that $k \geq 1$ and $m \geq 0$. The first two derivatives of $H(p)$ by p are given by

$$\frac{dH}{dp} = 2\langle \Delta X^{2k+m} | x, t \rangle + 2p \langle \Delta X^{2k+2m} | x, t \rangle, \frac{d^2 H}{dp^2} = 2 \langle \Delta X^{2k+2m} | x, t \rangle.$$

These two derivatives show that $H(p)$ has a minimum at

$$p_c = -\frac{\langle \Delta X^{2k+m} | x, t \rangle}{\langle \Delta X^{2k+2m} | x, t \rangle}.$$

Then the minimum $H_{\min}$ of $H(p)$ is given by

$$H_{\min} = \langle \Delta X^{2k} | x, t \rangle - \frac{\langle \Delta X^{2k+m} | x, t \rangle^2}{\langle \Delta X^{2k+2m} | x, t \rangle}.$$

The function $H(p) \geq 0$ for all p, then we have $H_{\min} \geq 0$, which implies

$$\langle \Delta X^{2k} | x, t\rangle \langle \Delta X^{2k+2m} | x, t\rangle \geq \langle \Delta X^{2k+m} | x, t\rangle^2,$$

which is trivial for $m = 0$. Consider $m \geq 1$, we find that

$$(2k)!\, D^{(2k)} (2k+2m)!\, D^{(2k+2m)} \geq [(2k+m)!\, D^{(2k+m)}]^2.$$

(b) Consequences of Pawula's Theorem. The consequences of Pawula's theorem can be seen by considering the cases that $D^{(2k)} = 0$ and $D^{(2k+2m)} = 0$, respectively.

4.1.3 The Fokker-Planck Equation

(a) Fokker-Planck Equation. Neglecting coefficients $D^{(n)}$ with $n \geq 3$. Hence we consider the following Fokker-Planck Equation:

$$\frac{\partial f(x,t)}{\partial t} = -\frac{-\partial D^{(1)}(x,t) f(x,t)}{\partial x} + \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2},$$

where

$$D^{(1)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle \Delta X | x, t\rangle}{\Delta t},$$
$$D^{(2)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle \Delta X^2 | x, t\rangle}{2\Delta t}.$$

(b) Mean Equation.

$$\int_{-\infty}^{\infty} x \frac{\partial f(x,t)}{\partial t}\, dx = -\int_{-\infty}^{\infty} x \frac{-\partial D^{(1)}(x,t) f(x,t)}{\partial x}\, dx + \int_{-\infty}^{\infty} x \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2}\, dx,$$
$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} x f(x,t)\, dx = \int_{-\infty}^{\infty} D^{(1)}(x,t) f(x,t)\, dx,$$
$$\frac{d\langle X\rangle}{dt} = \langle D^{(1)}\rangle.$$

(c) Variance Equation.

$$\int_{-\infty}^{\infty} x^2 \frac{\partial f(x,t)}{\partial t}\, dx = -\int_{-\infty}^{\infty} x^2 \frac{-\partial D^{(1)}(x,t) f(x,t)}{\partial x}\, dx + \int_{-\infty}^{\infty} x^2 \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2}\, dx,$$
$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} x^2 f(x,t)\, dx = 2\int_{-\infty}^{\infty} x D^{(1)}(x,t) f(x,t)\, dx - 2\int_{-\infty}^{\infty} x \frac{\partial D^{(2)}(x,t) f(x,t)}{\partial x}\, dx,$$
$$\frac{d\langle X^2\rangle}{dt} = 2\langle X D^{(1)}\rangle + 2\langle D^{(2)}\rangle.$$

Then, we have

$$\frac{d\langle \widetilde{X}^2\rangle}{dt} = 2\langle X D^{(1)}\rangle + 2\langle D^{(2)}\rangle - 2\langle X\rangle\langle D^{(1)}\rangle = 2\langle \widetilde{X}\widetilde{D}^{(1)}\rangle + 2\langle D^{(2)}\rangle.$$

4.2 Solution to the Fokker-Planck Equation

The equation considered is given by

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[F(t) + G(t)(x - \langle X\rangle)] f(x,t) + \frac{\partial^2 D(t) f(x,t)}{\partial x^2}.$$

4.2.1 The Solution Approach

(a) The Solution Approach. The solutoins $f(x,t)$ to the Fokker-Planck equation involve

- information about the intial PDF $f(x_0, t_0)$, and
- information about the trainsition from the intial PDF to any asymptotic PDF, which is determined by the PDF evolution equation.

First, we represet the PDF $f(x)$ as

$$f(x,t) = \int_{-\infty}^{\infty} f(x,t;x_0,t_0)\,dx_0,$$

where $f(x,t;x_0,t_0) = \langle \delta(x - X(t))\delta(x_0 - X(t_0))\rangle$ represents the two-point PDF.

4.3 Stochastic Differential Equations

4.3.1 Nonlinear Markovian Stochastic Equations

(a) Approach.

- Determine the general structure of stochastic difference equations.
- Represent the obtained stochastic difference equation as a stochastic differential equation.
- Define stochastic intergration for the calculation of solutions of stochastic differential equations.

(b) Stochastic Difference Equation.

$$\frac{X_n - X_{n-1}}{\Delta t} = a(X_{n-1}, t_{n-1}) + b(X_{n-1}, t_{n-1})\frac{\Delta W_{n-1}}{\Delta t},$$

where X_n represents the variable considered (the particle position, particle velocity or population density), and we have $\Delta W_{n-1} = (\Delta t)^{1/2}\varepsilon_{n-1}$.

(c) Stochastic Differential Equation. By considering an infinitesimal time interval $\Delta \to 0$ and defining time t by $t = n\Delta t$, the stochastic model can be written as

$$\frac{dX}{dt}(t) = a(X,t) + b(X,t)\frac{dW}{dt}(t). \tag{4.2}$$

(d) Stochastic Integration. Integrate equation (4.2) from t to $t + dt$,

$$X(t+dt) - X(t) = \int_t^{t+dt} a(X(s), s)\,ds + \int_t^{t+dt} b(X(s), s)\frac{dW}{ds}(s)\,ds,$$

where $dt \to 0$ is an infinitesimal time interval. The Itô definition is to take the function values of $a(X(s), s)$ and $b(X(s), s)$ at $X(t)$ and t, such that

$$X(t+dt) - X(t) = a(X(t), t)dt + b(X(t), t)dW(t),$$

where $dW(t) = W(t+dt) - W(t)$.The Stratonovich definition is to take $a(X(s), s)$ and $b(X(s), s)$ at the mean value $[X(t+dt) + X(t)]/2$ and t.

4.3.2 Relationship to the Fokker-Planck Equation

(a) PDF Equation. To find the PDF equation that is implied by the stochastic model (4.2) we consider thee Kramers-Moyal equation,

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t) f(x,t),$$

where

$$D^{(n)}(x,t) = \lim_{\Delta t \to 0} \frac{\langle (X(t+\Delta t) - X(t))^n | X(t) = x \rangle}{n!\,\Delta t}.$$

Then we can derive the following by the continuous case of $X(t+dt) - X(t)$ as follows,

$$X(t+\Delta t) - X(t) = a(X(t),t)\Delta t + b(X(t),t)\Delta W(t).$$

Then the Kramers-Moyal coefficients is

$$\begin{aligned} D^{(n)}(x,t) &= \lim_{\Delta t \to 0} \frac{\langle [a(X(t),t)\Delta t + b(X(t),t)\Delta W(t)]^n | X(t) = x \rangle}{n!\,\Delta t}, \\ &= \lim_{\Delta t \to 0} \frac{(\Delta t)^{n/2}}{n!\,\Delta t} \left\langle \left[a(x,t)\sqrt{\Delta t} + b(x,t)\frac{\Delta W(t)}{\sqrt{\Delta t}} \right]^n \right\rangle, \end{aligned}$$

Then we can calculate Kramers-Moyal coefficients as follows

$$\begin{aligned} D^{(1)}(x,t) &= \lim_{\Delta t \to 0} \frac{(\Delta t)^{1/2}}{1!\,\Delta t} \left\langle \left[a(x,t)\sqrt{\Delta t} + b(x,t)\frac{\Delta W(t)}{\sqrt{\Delta t}} \right]^1 \right\rangle = a(x,t), \\ D^{(2)}(x,t) &= \lim_{\Delta t \to 0} \frac{(\Delta t)}{2!\,\Delta t} \left\langle \left[a(x,t)\sqrt{\Delta t} + b(x,t)\frac{\Delta W(t)}{\sqrt{\Delta t}} \right]^2 \right\rangle \\ &= \lim_{\Delta t \to 0} \frac{(\Delta t)}{2!\,\Delta t} \left\langle a^2(x,t)\Delta t + 2a(x,t)b(x,t)\Delta W(t) + b^2(x,t)\frac{\Delta W^2(t)}{\Delta t} \right\rangle = \frac{1}{2}b^2(x,t), \\ D^{(3)}(x,t) &= \cdots = D^{(n)}(x,t) = \cdots = D^{(\infty)} = 0. \end{aligned}$$

(b) Correlations. Assume that $t \le t' = t + r$, where r is any non-negative time. Consider (4.2) at $t + r$ instead of t,

$$\frac{dX(t+r)}{d(t+r)} = a(X(t+r), t+r) + b(X(t+r), t+r)\frac{dW}{d(t+r)}(t+r).$$

The differentiation of X by $t + r$ can be replaced by a derivative by r. We multiply this equation with $\widetilde{X}(t)$ and average,

$$\left\langle \widetilde{X}(t)\frac{dX(t+r)}{dr} \right\rangle = \langle \widetilde{X}(t) a(X(t+r), t+r) \rangle + \left\langle \widetilde{X}(t) b(X(t+r), t+r)\frac{dW}{dt}(t+r) \right\rangle.$$

Note that dW/dt at $t + r$ is independent of $X(t)$ and $X(t + r)$, and dW/dt vanishes in the mean,

$$\left\langle \widetilde{X}(t) b(X(t+r), t+r)\frac{dW}{dt}(t+r) \right\rangle = \langle \widetilde{X}(t) b(X(t+r), t+r) \rangle \left\langle \frac{dW}{dt}(t+r) \right\rangle = 0.$$

Then we have

$$\frac{d\langle \widetilde{X}(t) X(t+r)}{dr} = \langle \widetilde{X}(t) a(X(t+r), t+r) \rangle.$$

We may replace X and a with $\widetilde{X}$ and $\widetilde{a}$ and obtain

$$\frac{d\langle \widetilde{X}(t) \widetilde{X}(t+r)}{dr} = \langle \widetilde{X}(t) \widetilde{a}(X(t+r), t+r) \rangle.$$

4.3.3 Linear Markovian Stochastic Equations

Consider the linear stochastic Markovian differential equation,

$$\frac{dX}{dt}(t) = -\frac{X - \langle X \rangle}{\tau} + \frac{\sqrt{D}}{\tau}\frac{dW}{dt}(t),$$

where D is a diffusion coefficient, and τ is the cahracteristic relaxation time scale of fluctuations. Averaging the above equation gives

$$\frac{d\langle X \rangle}{dt} = 0.$$

Hence, the mean $\langle X \rangle$ is a constant, i.e., $\langle X \rangle = \langle X_0 \rangle$.

(a) PDF. The evolution of $X(t)$ can be described by the Fokker-Planck equation

$$\frac{\partial f(x,t)}{\partial t} = \frac{1}{\tau}\frac{\partial (x - \langle X \rangle) f(x,t)}{\partial x} + \frac{D}{2\tau^2}\frac{\partial^2 f(x,t)}{\partial x^2}.$$

In particular, when $F = 0$, $G = -1/\tau$, the solution of the above equation is given by

$$f(x,t) = \int \frac{1}{\sqrt{2\pi\beta}} \exp\left\{-\frac{(x-\alpha)^2}{2\beta}\right\} f(x_0, t_0)\, dx_0,$$

where $f(x_0, t_0)$ refers to any intial PDF.

4.3.4 Summary

(a) PDF Evolution Equation. Given the Kramers-Moyal equation

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t) f(x,t).$$

The Kramers-Moyal equation implies Pawula's theorem that shows that there are two possibilities: we can either work with an equation that involves an infinite number of Kramers-Moyal coefficients $D^{(n)}(x,t)$, or we can work with Fokker-Planck equation,

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial D^{(1)}(x,t) f(x,t)}{\partial x} + \frac{\partial^2 D^{(2)}(x,t) f(x,t)}{\partial x^2},$$

which does only involve the first two Kramers-Moyal coefficients. The neglect of $D^{(n)}$ with $n \geq 3$ is justified if the stochastic process considered has a continuous sample path, this means if jump processes (i.e., processes involving instantaneous unbounded changes) are not considered.

(b) Solutions of the Fokker-Planck Equation. It was shown that this equation can be solved analytically if we consider the specific Fokker-Planck equation,

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[F(t) + G(t)(x - \langle X \rangle)] f(x,t) + \frac{\partial^2 D(t) f(x,t)}{\partial x^2}.$$

It turns out the sotlution to the above equation is given by a normal PDF integrated over the initial conditions,

$$f(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\beta}} \exp\left\{-\frac{(x-\alpha)^2}{2\beta}\right\} f(x_0, t_0)\, dx_0,$$

where α and β are functions of t, and α does not depend on x_0. Asymptotically (i.e., for $t \to \infty$), α and β relax to the mean $\langle X \rangle$ and variance $\langle \widetilde{X}^2 \rangle$ of the process considered. Then, the PDF $f(x,t)$ becomes independent of the initial PDF $f(x_0, t_0)$: $f(x,t)$ is then given by a normal PDF with mean $\langle X \rangle$ and variance $\langle \widetilde{X}^2 \rangle$.

(c) Stochastic Process Equations. Instead of asking how the PDF of a stochastic process evolves, we may ask how the underlying stochastic process evovles in time. In generalization of the stochastic difference equation, we considered the model for the evolution of the stochastic process $X(t)$,

$$\frac{dX}{dt}(t) = a(X,t) + b(X,t)\frac{dW}{dt}(t). \tag{4.3}$$

We caclucate the Kramers-Moyal coefficients that are implied by the above equation as follows,

$$\begin{aligned} D^{(1)}(x,t) &= a(x,t), \\ D^{(2)}(x,t) &= \frac{1}{2}b^2(x,t), \\ D^{(3)}(x,t) &= D^{(4)}(x,t) = \cdots = D^{(\infty)}(x,t) = 0. \end{aligned}$$

By using the coefficient relations we see that the Fokkeer-Planck equation corresponds to the stochastic model

$$\frac{dX}{dt}(t) = F(t) + G(t)[X(t) - \langle X \rangle] + \sqrt{2D(t)}\frac{dW}{dt}.$$

(d) Applications to Modeling. The stochastic differntial equation (4.3) can be used for the modeling of any nonlinear processes. On the other hand, (4.3) describes a Markovian stochastic process, and this assumption is often not rigorously satisfied. It was shown that the Markovian velocity model is not incorrect but only less complete than the non-Markovian model, which describes proceses that take place over the time scale τ_f (over which accelerations change) and over the time scale τ (over which velocities change).

$$\begin{aligned} \frac{dX}{dt}(t) &= b\frac{dW}{dt}(t), \\ \frac{dv}{dt}(t) &= -\frac{v(t) - \langle V \rangle}{\tau} + \sqrt{\frac{4e}{3\tau}}\frac{dW}{dt}. \end{aligned}$$

5 Stochastic Multivariate Evolution

5.1 Motivation

(a) Fluid Dynamics. consider the motion of fluids (e.g., atmospheric motions) in order to illustrate the need for methods for the calculation of the evolution of several random variables. The prediction of fluid flow requires the calculation of the mean velocity $U_i(\mathbf{x}, t)$ of molecules, which represents the ith component ($i = 1, \ldots, 3$) of the fluid velocity at the position $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ at time t. The fluid velocity $U_i(\mathbf{x}, t)$ and fluid mass density $\rho(\mathbf{x}, t)$ have to satisfy a coupled system of partial differential equations, which represent the conservation of mass and momentum,

$$\begin{aligned} \frac{D\rho}{Dt} + \rho\frac{\partial U_m}{\partial x_m} &= 0, \\ \frac{DU_i}{Dt} + \frac{1}{\rho}\frac{\partial \rho\sigma_{im}}{\partial x_m} &= 0, \end{aligned}$$

where $\sigma_{im}(\mathbf{x}, t)$ refers to the variance of molecular velocities, this means $\sigma_{im} = \overline{v_i v_m}$. We use the sum convention for repeated subscripts, this means we have for example

$$\frac{\partial U_m}{\partial x_m} = \frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial x_2} + \frac{\partial U_3}{\partial x_3}.$$

The total derivative (or substantial or material derviative) of any property $Q(\mathbf{x}, t)$ is defined by

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + U_m \frac{\partial Q_m}{\partial x_m}.$$

The meaning of DQ/Dt can be seen by considering the property Q at $\mathbf{x} = \mathbf{x}(t)$, where $\mathbf{x}(t)$ is a point that follows the fluid velocity U_i, i.e., $\mathbf{x}(t)$ is determined by

$$\frac{dx_i(t)}{dt} = U_i(\mathbf{x}(t), t).$$

The total derivative DQ/Dt at $\mathbf{x} = \mathbf{x}(t)$ reads

$$\begin{aligned}
\frac{DQ(\mathbf{x}(t), t)}{Dt} &= \frac{\partial Q(\mathbf{x}(t), t)}{\partial t} + U_m(\mathbf{x}(t), t)\frac{\partial Q(\mathbf{x}(t), t)}{\partial x_m} \\
&= \frac{\partial Q(\mathbf{x}(t), t)}{\partial t} + \frac{\partial Q(\mathbf{x}(t), t)}{\partial x_m}\frac{dx_m(t)}{dt} \\
&= \frac{dQ(\mathbf{x}(t), t)}{dt}.
\end{aligned}$$

(b) Closure Problem. The conservation of mass and momentum are unclosed because the variance σ_{im} of molecular velocities is unknown. The variance σ_{im} has to satisfy a convservation equation,

$$\frac{D\sigma_{ij}}{Dt} + \frac{1}{\rho}\frac{\partial \rho \overline{v_i v_j v_m}}{\partial x_m} + \frac{\partial U_i}{\partial x_m}\sigma_{mj} + \frac{\partial U_j}{\partial x_m}\sigma_{mi} = -\frac{2}{T}\left(\sigma_{ij} - \frac{\sigma_{kk}}{3}\delta_{ij}\right),$$

where $\overline{v_i v_j v_m}$ is the triple correlation of molecular velocities, T is a charactersitic relaxation time scale, and δ_{ij} refers to the Kronecker delta.

5.2 Data Analysis Concepts for Joint Random Variables

5.2.1 Joint Probability Density Functions

(a) Joint PDF. Define the joiont PDF of two variables X and Y by

$$f(x, y) = \langle \delta(x - X)\delta(y - Y)\rangle.$$

The joint PDF $f(x, y)$ has the properties

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x, y)\, dy &= \int_{-\infty}^{\infty} \langle \delta(x - X)\delta(y - Y)\rangle\, dy = \langle \delta(x - X)\rangle = f(x), \\
\int_{-\infty}^{\infty} f(x, y)\, dx &= \int_{-\infty}^{\infty} \langle \delta(x - X)\delta(y - Y)\rangle\, dx = \langle \delta(y - Y)\rangle = f(y),
\end{aligned}$$

where $f(x)$ and $f(y)$ are called marginal PDFs. Other properties of joint PDF $f(x,y)$ are

$$\begin{aligned}
&f(x,y) \geq 0, \\
&f(-\infty,y) = f(\infty,y) = f(x,-\infty) = f(x,\infty) = 0, \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dydx = 1, \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dydx = \langle g(X,Y)\rangle.
\end{aligned}$$

Then the probability for joint events $a \leq X \leq b$ and $c \leq Y \leq d$,

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\,dx\,dy.$$

The validity of this relation can be seen by using the definition of $f(x,y)$,

$$\int_c^d \int_a^b f(x,y)\,dx\,dy = \left\langle \int_c^d \int_a^b \frac{d\theta(x-X)}{dx}\frac{d\theta(y-Y)}{dy}\,dx\,dy \right\rangle = \langle [\theta(b-X)-\theta(a-X)][\theta(d-Y)-\theta(c-Y)]\rangle.$$

5.2.2 Application to Optimal Modeling

(a) Optimal Model. We consider a set of (X_i, Y_i) data, where $i = 1, 2, \ldots, N$. We want to find a model $y_M(x)$ that agrees as good as possible with the given data. The particular problem was to find a model $y_M(x)$ that minimizes the least-square error

$$\begin{aligned}
E^2 &= \frac{1}{N}\sum_{i=1}^{N}[Y_i - y_M(X_i)]^2 \\
&= \langle [Y - y_M(X)]^2 \rangle \\
&= \int_{-\infty}^{\infty} \langle [Y - y_M(X)]^2 | x \rangle f(x)\,dx \\
&= \int_{-\infty}^{\infty} \langle [Y - y_M(x)]^2 | x \rangle f(x)\,dx.
\end{aligned}$$

5.3 The Fokker-Planck Equation

5.3.1 Definition of Multivariate Probability Density Functions

The generalization of the Fokker-Planck equation to an equation for the joint PDF of a vectorial stochastic process $\mathbf{X}(t) = \{X_1(t), X_2(t), \ldots, X_N(t)\}$ requires a relevant step: the definition of a multivariate PDF $f(\mathbf{x}, t)$ using the theta and delta functions for several variables.

(a) Multivariate THeta and Delta Functions. For a vectorial process $\mathbf{X}(t) = \{X_1(t), X_2(t), \ldots, X_N(t)\}$, the corresponding theta and delta functions are given by

$$\begin{aligned}
&\theta[\mathbf{x} - \mathbf{X}(t)] = \theta(x_1 - X_1(t))\theta(x_2 - X_2(t))\cdots\theta(x_N - X_N(t)), \\
&\delta[\mathbf{x} - \mathbf{X}(t)] \\
&\qquad = \delta(x_1 - X_1(t))\delta(x_2 - X_2(t))\cdots\delta(x_N - X_N(t)).
\end{aligned}$$

Hence, muultivariate theta and delta functions are products of all the theta and delta functions of single variables.

(b) Multivariate PDFs. Averaging the above delta function gives the joint PDF $f(\mathbf{x}, t)$

$$f(\mathbf{x}, t) = \langle \delta[\mathbf{x} - \mathbf{X}(t)] \rangle.$$

In terms of the normalization property of delta functions we find that this definition staisfies the normalization condition for the joint PDF $f(\mathbf{x}, t)$,

$$\int f(\mathbf{x}, t)\, d\mathbf{x} = \int \langle \delta[\mathbf{x} - \mathbf{X}(t)] \rangle\, d\mathbf{x} = \langle 1 \rangle = 1,$$

where $d\mathbf{x} = dx_1 dx_2 \cdots dx_N$ is a multivariate differential given by the product of all differential involved. Two-point PDFs can be defined correspondingly. For example, the two-point PDF $f(\mathbf{x}, t; \mathbf{x}', t')$ for having joing events $(\mathbf{x}, t)$ and $(\mathbf{x}', t')$ is defined by

$$f(\mathbf{x}, t; \mathbf{x}', t') = \langle \delta[\mathbf{x} - \mathbf{X}(t)]\delta[\mathbf{x}' - \mathbf{X}(t')] \rangle.$$

The one-point PDF $f(\mathbf{x}, t)$ can be recoveered from this definition,

$$f(\mathbf{x}, t) = \int_{-\infty}^{\infty} f(\mathbf{x}, t; \mathbf{x}', t')\, d\mathbf{x}' = \int_{-\infty}^{\infty} \langle \delta[\mathbf{x} - \mathbf{X}(t)]\delta[\mathbf{x}' - \mathbf{X}(t')] \rangle\, d\mathbf{x}' = \langle \delta[\mathbf{x} - \mathbf{X}(t)] \rangle.$$

A PDF $f(\mathbf{x}, t|\mathbf{x}', t')$ conditioned on $\mathbf{X}(t') = \mathbf{x}'$ can be defined in corresponding to the definition for a single-variable PDF,

$$f(\mathbf{x}, t|\mathbf{x}', t') = \frac{f(\mathbf{x}, t; \mathbf{x}', t')}{f(\mathbf{x}', t')} = \frac{\langle \delta[\mathbf{x} - \mathbf{X}(t)]\delta[\mathbf{x}' - \mathbf{X}(t')] \rangle}{\langle \delta[\mathbf{x}' - \mathbf{X}(t')] \rangle} = \langle \delta[\mathbf{x} - \mathbf{X}(t)|\mathbf{X}(t') = \mathbf{x}'] = \langle \delta[\mathbf{x} - \mathbf{X}(t)|\mathbf{x}', t'].$$

In terms of this definition the one-point PDF $f(\mathbf{x}, t)$ can be written

$$f(\mathbf{x}, t) = \int_{-\infty}^{\infty} f(\mathbf{x}, t|\mathbf{x}', t') f(\mathbf{x}', t')\, d\mathbf{x}'.$$

5.3.2 The Fokker-Planck Equation

(a) Fokker-Planck Equation. Let us consider an N-dimensional stochastic vector process $\mathbf{X}(t) = \{X_1(t), X_2(t), \dots, X_N(t)\}$.

6 Final Exam Review

6.1 Equivlibrium Models (Analytic Models)

(a) When $-\infty < x < \infty$, normal PDF.

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

where $\mu = \langle X \rangle$, $\sigma^2 = \langle \widetilde{X}^2 \rangle$. Support:

- FPE.
- Central Limit Theorem.
- SML.

(b) When $0 < x < \infty$, gamma PDF.

$$f(x) = \frac{a}{\Gamma(a)\langle X\rangle}\left(\frac{ax}{\langle X\rangle}\right)^{a-1} e^{-ax/\langle X\rangle},$$

where $\langle X\rangle$ is the mean and $a = \frac{\langle X\rangle^2}{\langle \widetilde{X}^2\rangle}$. No support.

(c) When $0 < x < 1$. Beta PDF.

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}.$$

$$\begin{aligned} a &= \langle X\rangle\left[\frac{\langle X\rangle(1-\langle X\rangle)}{\langle \widetilde{X}^2\rangle} - 1\right], \\ b &= (1-\langle X\rangle)\left[\frac{\langle X\rangle(1-\langle X\rangle)}{\langle \widetilde{X}^2\rangle} - 1\right]. \end{aligned}$$

(d) Limitations.
- One-trend processes.
- Generalization to multi-trend processes.

6.2 General Evolutoin

- Kramers-Moyal Equation.

$$\frac{\partial f}{\partial t}(x,t) = \sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^u D^{(u)}(x,t)f(x,t).$$

  - No jumps: Fokker-Planck Equation.

$$\frac{\partial f}{\partial t}(\mathbf{x},t) = -\frac{\partial D_i(\mathbf{x},t)f(\mathbf{x},t)}{\partial x_i} + \frac{\partial^2 D_{ij}(\mathbf{x},t)f(\mathbf{x},t)}{\partial x_i \partial x_j}.$$

  We have the analytical conclusion, PDF, $\langle X_k\rangle$ and $\langle \widetilde{X}_k\widetilde{X}_n\rangle$.

  - Modeling, numerical solution.

$$\frac{dX_i}{dt} = a_i[\mathbf{X}(t),t] + b_{ij}[\mathbf{X}(t),t]\frac{dW_k}{dt}(t).$$

6.3 Linear Evolutoin

- Fokker-Planck Equation.

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x_m}[F_m(t) + G_{mk}(t)(x_k - \langle X_k\rangle)]f(\mathbf{x},t) + \frac{\partial^2 D_{nm}(t)f(\mathbf{x},t)}{\partial x_n x_m}.$$

Close to equilibrium.

$$\frac{dX_m}{dt} = F_m(t) + G_{mk}(t)(x_k - \langle X_k\rangle) + b_{nk}(t)\frac{dW_k}{dt}(t).$$

Solution.

$$f(\mathbf{x},t) = \int f(\mathbf{x},t|\mathbf{x}',t')\,d\mathbf{x}',$$

where

$$f(\mathbf{x},t|\mathbf{x}',t') = -\frac{1}{(2\pi)^{N/2}\sqrt{\det\beta}}\exp\{-\frac{1}{2}\beta_{ij}^{-1}(x_i-\alpha_i)(x_i-\alpha_j)\}.$$

And

$$\frac{\alpha_k}{\alpha_n} = F_k + G_{kn}.$$

6.4 Model Parameters

What variables are considered?

- Model 1: x.
- Model 2: (x,v).
- Model 3: (x,v,a).

$$\frac{dX}{dt} = -\frac{X-\langle X\rangle}{\tau} + \frac{\sqrt{D}}{\tau}\frac{dW}{dt}.$$

Then we have normal PDF:

$$\langle X\rangle = \text{constant}, \langle\widetilde{X}(t)\widetilde{X}(s)\rangle \xrightarrow{t\to\infty} \frac{D}{2\tau}e^{-(t-s)/\tau}.$$

Final:

- Normal PDF.
- Optimal Model.
- PDF e.g. MC solutiono.
- SDE e.g. Solutoin via PDF eg.