Gaussian Process

Yiyuan Chen

16 December 2023

§1 Motivations: Why Gaussian?

§1.1 Random Diffusion

A particle experiences a random diffusion. Let τ be the time duration and $\rho(y)$ be the probability density at time τ where y is the distance between original place. We can observe $\rho(y)$ satisfies the following property:

$$\rho(y) \geqslant 0, \quad \int_{-\infty}^{+\infty} \rho(y) dy = 1, \quad \rho(y) = \rho(-y)$$

As a result, we can compute $\int_{-\infty}^{+\infty} y\rho(y)dy = 0$ and let $D = \int_{-\infty}^{+\infty} y^2\rho(y)dy$. Now, let $f(x,t)$ be the probability density at place x and time t with the initial value $f(0,0) = M$. By Conservation Law, we have

$$f(x, t+\tau) = \int_{-\infty}^{+\infty} f(x-y, t)p(y)dy$$

By Taylor expansion,

$$\begin{cases} f(x, t+\tau) = f(x,t) + \frac{\partial f}{\partial t} \cdot \tau + o(\tau) \\ f(x-y, \tau) = f(x,y) - \frac{\partial f}{\partial x} \cdot y + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} y^2 + o\left(y^2\right) \end{cases}$$

plunge into conservation equation, we get

$$\frac{\partial f}{\partial t} \cdot \tau = \frac{1}{2}\frac{\partial^2 f}{\partial x^2} \cdot D$$

Let $\frac{D}{2\tau} = C$,

$$\frac{\partial f}{\partial t} = C \cdot \frac{\partial^2 f}{\partial x^2}$$

Solve this partial differential equation, we get

$$f(x,t) = \frac{1}{\sqrt{2\pi Ct}} \exp\left(-\frac{x^2}{2Ct}\right)$$

§1.2 Maximum Entropy

Definition 1.1. The Entropy of a random variable is the following integral:

$$H(x) = -\int_{-\infty}^{+\infty} f_X(x) \log f_X(x) dx$$

Now, we want to find the random variable $X \in (-\infty, +\infty)$ such that the Entropy be the maximum under the following constraints:

$$E(X) = \mu \quad E\left(X^2\right) = \sigma^2$$

By Lagrange multiplier, we need optimize the following factorial:

$$\begin{aligned} G(f) = & -\int_{\mathbb{R}} f_X(x) \log f_X(x) dx + \lambda_1 \left(\int_{\mathbb{R}} f_X(x) dx - 1 \right) + \lambda_2 \left(\int_{\mathbb{R}} x f_X(x) dx - \mu \right) \\ & + \lambda_3 \left(\int_{\mathbb{R}} x^2 f_X(x) dx - \sigma^2 \right) \end{aligned}$$

We use Variational Method to find optimal. Denote f_0 is optimal solution. Let $H(t) = G\left(f_0 + tg\right)$ where $t \in \mathbb{R}$ and g is an arbitrary function. Since we have $H(0) = G\left(f_0\right) \geqslant G\left(f_0 + tg\right) = H(t)$, we deduce

$$\frac{\partial H(t)}{\partial t}|_{t=0} = 0$$

By computation, we have

$$\begin{aligned} \frac{\partial H(t)}{\partial t}|_{t=0} & = -\int g \log f + \int g + \lambda_1 \int g + \lambda_2 \int xg + \lambda_3 \int x^3 g \\ & = \int g(\log f + (\lambda_1 + 1) + \lambda_2 x + \lambda_3 x^2) \end{aligned}$$

Since g is an arbitrary function we have,

$$\log f + (\lambda_1 + 1) + \lambda_2 x + \lambda_3 x^2 = 0$$

Thus, we know

$$f(x) = \exp\left(\lambda_3 x^2 + \lambda_2 x + \lambda_1 + 1\right)$$

which is Gaussian!

> **Remark 1.2.** Random Walk

> **Remark 1.3.** If we change the condition of Random Variabe to $X \in [0, \infty)$ and $\int x f_X(x) dx = \mu$. Then, by similar discussion we use Maximum Entropy induce exponential distributing (No memory)
>
> $$\lambda \exp(-\lambda x) I_{(0,\infty)}(x)$$
>
> If the $X \in [a, b]$ with no moment constraint, then Maximum Entropy induce uniform distribution.

§1.3 Central Limit Theorem

We recall Central Limit Theorem

Theorem 1.4

Let $X_1, X_2, \cdot, X_n$ be i.d.d random variable and $E\left(X_k\right) = 0, \operatorname{Var}\left(X_k\right) = 1$. Then,

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \stackrel{n \to \infty}{\longrightarrow} \mathcal{N}(0, 1)$$

> **Remark 1.5.** From above theorem, we can find Gaussian distribution is universal.

Question 1.6. Recall Large Number law, compare it with Central Limit Theorem, and think about why they are different?

Before we prove Central Limit Theorem, we recall characteristic function.

Definition 1.7. Let X be a random variable, $f_X(x)$ is its probability distribution. Then its characteristic function[1] is

$$\begin{aligned}\phi_X(\omega) &= E(\exp(j\omega x)) \\ &= \int_{-\infty}^{+\infty} \exp(j\omega x) f_X(x) dx\end{aligned}$$

Now, we compute characteristic function of $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Since $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$,

$$\begin{aligned}\phi_X(\omega) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + j\omega x\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu-j\sigma^2\omega\right)^2 + j\mu\omega - \frac{1}{2}\sigma^2\omega^2\right) dx \\ &= \exp\left(j\mu\omega - \frac{1}{2}\sigma^2\omega^2\right)\left(\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu-j\sigma^2\omega\right)^2\right) dx\right) \\ &= \exp\left(j\mu\omega - \frac{1}{2}\sigma^2\omega^2\right)\end{aligned}$$

Question 1.8. How to compute $\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu-j\sigma^2\omega\right)^2\right) dx$? Which theorem in complex analysis we use?

> **Remark 1.9.** Why we use characteristic function to research random variable? There are many reasons.
>
> - Firstly, Fourier Transformation is one to one. Thus Characteristic Function is 'Character' of random variable.
> - Secondly, it provide us a convenient way to deal with the sum of random variable. It's similar to Lagrange Transformation in ordinary differential equation.
> - Third, you will see in Gaussian Distribution characteristic function avoid computing inverse of covariance matrix. It's much convenient!!!

Proof. Since

$$\phi_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(\omega) = \prod_{k=1}^{n} \phi_{X_k}\left(\frac{\omega}{\sqrt{n}}\right) = \phi_{X_1}^n\left(\frac{\omega}{\sqrt{n}}\right)$$

and

$$\begin{aligned}\phi\left(\frac{\omega}{\sqrt{n}}\right) &= E\left(\exp\left(j\frac{\omega}{\sqrt{n}}X\right)\right) \\ &= E\left(1 + j\frac{\omega}{\sqrt{n}}X - \frac{1}{2}\frac{\omega^2}{n}X^2 + o\left(\frac{1}{n}\right)\right) \\ &= 1 - \frac{\omega}{2n} + o\left(\frac{1}{n}\right)\end{aligned}$$

[1] Since $f_X(x) \geq 0$, $\phi_X(w)$ is positive defined.

We conclude

$$\phi_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(\omega) = \left(1 - \frac{\omega^2}{2n} + 0\left(\frac{1}{n}\right)\right)^n \stackrel{n\to\infty}{\longrightarrow} \exp\left(-\frac{\omega^2}{2}\right)$$

Thus,

$$f_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(x) \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

□

Remark 1.10. Recall Law of Large Numbers: Let $X_1, X_2, \cdot, X_n$ be i.d.d random variable and $E(X_k) = m$. Then,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \stackrel{n\to\infty}{\longrightarrow} m$$

The proof is similar and we know when $\delta > 0$,

$$\frac{X_1 + \cdots + X_n}{n^{1+\delta}} \to 0 \quad (n \to \infty)$$

According to the power of n the average of random variables present different property: still random variable or fix number. So what is the critical point? The answer is $\sqrt{n \log n \log n}$.

Finally, we consider discrete model for random diffusion. Let $X(t)$ be a stochastic process and $t = n\Delta t$. In every time period Δt a particle will move Δx with 50% to right and 50% to left. Denote S_n be the step number to right.Then $S_n = X_1 + X_2 + \cdot + X_n$ where

$$X_k = \begin{cases} 1 & 0.5 \\ 0 & 0.5 \end{cases}$$

Then,

$$\begin{aligned} X(t) &= S_n \cdot \Delta x + (n - S_n)(-\Delta x) \\ &= (2S_n - n)\Delta x \end{aligned}$$

It's easy to find $E(X(t)) = 0$ and

$$\begin{aligned} \operatorname{Var}(X(t)) &= (\Delta x)^2 \operatorname{Var}(2S_n - n) \\ &= 4(\Delta x)^2 \operatorname{Var}(S_n) \\ &= (\Delta x)^2 n \end{aligned}$$

Now,

$$\begin{aligned} X(t) &= (2S_n - n)\Delta x \\ &= \frac{s_n - \frac{n}{2}}{\sqrt{\frac{1}{4}n}} \cdot \sqrt{n}\Delta x \end{aligned}$$

Where $\sqrt{n}\Delta x = \Delta x \cdot \sqrt{\frac{t}{\Delta t}} = \sqrt{\frac{\Delta x^2}{\Delta t} \cdot t}$. If we let $\frac{\Delta x^2}{\Delta t} \to D$ when $\Delta t \to 0$, then

$$X(t) \to \mathcal{N}(0,1)\sqrt{Dt} = \mathcal{N}(0, \sqrt{Dt})$$

when $\Delta t \to 0$. This implies $f_X(x) = \frac{1}{\sqrt{2\pi Dt}} \exp\left(-\frac{x^2}{2Dt}\right)$

§2 Gaussian Process

§2.1 Basic Properties of Gaussian Distribution and Gaussian Process

Definition 2.1. A stochastic process $X(t)$ is a Gaussian process if for arbitrary $n \in \mathbb{Z}$ and arbitrary $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$, $X = (X(t_1), \cdots, X(t_n))^\top \sim \mathcal{N}(\mu, \Sigma)$

When $n = 1$, $X \sim \mathcal{N}(\mu, \sigma)$ we know $E(X) = \mu, \mathrm{Var}(X) = \sigma^2$ and

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

When $n = 2$, $X \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ we know $E(X_i) = \mu_i, \mathrm{Var}(X_i) = \sigma_i^2$ and $E((X_1 - \mu_1)(X_2 - \mu_2)) = \rho$ where ρ is the linear relationship of two random variables.

$$\begin{aligned} f_{X_1 X_2}(x_1, x_2) = & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ & \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1 \cdot \sigma_2}\right)\right) \end{aligned}$$

For higher dimensional Gaussian Distribution, we have

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{\Sigma}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)$$

where $E(\mathbf{X}) = \boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{\Sigma} = E((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top)$, $\mathbf{\Sigma}_{i,j} = E((X_i - \mu_i)(X_j - \mu_j))$.

Question 2.2. How to compute the integral $\int_{\mathbb{R}^n} f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$ [2]? Hint: Since $\mathbf{\Sigma}$ is symmetric and positive defined, we have $\mathbf{\Sigma} = A^\top A$.

Proposition 2.3

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, then the characteristic function of $\mathbf{X}$ is

$$\phi_{\mathbf{X}}(\boldsymbol{\omega}) = \exp(j\boldsymbol{\omega}^\top \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^\top \mathbf{\Sigma} \boldsymbol{\omega})$$

Remark 2.4. Notice we can avoid computing inverse of $\mathbf{\Sigma}$ in characteristic field!

§2.2 Linearity of Gaussian Process

Theorem 2.5

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, $\mathbf{Y} = A\mathbf{X}$, then $\mathbf{Y} \sim \mathcal{N}(A\boldsymbol{\mu}, A^\top \mathbf{\Sigma} A)$,

Proof.

$$\begin{aligned} \phi_{\mathbf{Y}}(\boldsymbol{\omega}) &= E(\exp(j\boldsymbol{\omega}^\top \mathbf{Y})) \\ &= E(\exp(j\boldsymbol{\omega}^\top A\mathbf{X})) \\ &= E(\exp(j(A\boldsymbol{\omega})^\top \mathbf{X})) \\ &= \phi_{\mathbf{X}}(A^\top \boldsymbol{\omega}) \\ &= \exp(j\boldsymbol{\omega}^\top A\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^\top A\mathbf{\Sigma} A^\top \boldsymbol{\omega}) \end{aligned}$$

[2] Notice classical result: $\int_{-\infty}^{+\infty} \exp(-\frac{1}{2}y^2)dy = \sqrt{2\pi}$

□

Corollary 2.6

If joint distribution is Gaussian, then marginal distribution is also Gaussian.

However, the opposite of the above claim is not true.

Example 2.7

We want to find a distribution which is not Gaussian but its marginal distribution is Gaussian. Firstly, we let

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right)+k(x_1,x_2)$$

We want find $k(x_1,x_2)$ satisfies

$$\int_{-\infty}^{+\infty}k(x_1,x_2)\,dx_1=\int_{-\infty}^{+\infty}k(x_1,k_2)\,dx_2=0$$

And we want $f_{X_1X_2}(x_1,x_2)$ is a probability density. Finally, we find

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi}\exp\left(-\frac{x_2^2+x_2^2}{2}\right)(1+\sin x_1\sin x_2)$$

is desired function.

Theorem 2.8

For a random vector $\mathbf{X}\in\mathbb{R}^n$, if $\forall\alpha\in\mathbb{R}^n$ we have $\alpha^\top\mathbf{X}\sim\mathcal{N}$, then $\mathbf{X}$ is Gaussian.

Proof.

$$\begin{aligned}\phi_{\mathbf{X}}(\boldsymbol{\omega})&=E\left(\exp\left(j\boldsymbol{\omega}^\top\mathbf{X}\right)\right)\\&=\phi_{\boldsymbol{\omega}^\top\mathbf{X}}(1)\\&=\exp(j\mu_{\boldsymbol{\omega}^\top\mathbf{X}}-\frac{1}{2}\sigma^2_{\boldsymbol{\omega}^\top\mathbf{X}})\end{aligned}$$

$$\mu_{\boldsymbol{\omega}^\top\mathbf{X}}=E(\boldsymbol{\omega}^\top\mathbf{X})=\boldsymbol{\omega}^\top\mathbf{X}$$

$$\begin{aligned}\sigma^2_{\boldsymbol{\omega}^\top\mathbf{X}}&=E((\boldsymbol{\omega}^\top\mathbf{X}-\boldsymbol{\omega}^\top\boldsymbol{\mu})^2)\\&=\boldsymbol{\omega}^\top E((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top)\boldsymbol{\omega}\\&=\boldsymbol{\omega}^\top\boldsymbol{\Sigma}\boldsymbol{\omega}\end{aligned}$$

□