Math 227C: Introduction to Stochastic Differential Equations

Lecturer: Xiaohui Xie
Scribe: Xinwen Zhang

Lecture #3
4/8/2014

---

1 Gaussian Distribution

1.1 one-dimension Gaussian Distribution

This distribution can be represented by $X \sim N(\mu, \sigma^2)$. μ represents Expectation and δ^2 represent variance of the distribution. It has several properties shown below.

- Density function: $\rho(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Distribution function: $F(X \leq a) = \int_{-\infty}^{a} \rho(x)dx$
- Expectation: $E[X] = \int_{-\infty}^{\infty} \rho(x)dx = \mu$
- Variance: $Var[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} \rho(x)dx = \mu$
- The nth moment of X : $E[x^P] = \int_{-\infty}^{\infty} x^P \rho(x)dx$
- If we assume $\mu = 0$, the pth moment function of X can be simplified as:

$$
\begin{aligned}
&E[x^P] = \int_{-\infty}^{\infty} x^P \rho(x)dx \\
&= \int_{-\infty}^{\infty} x^P \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} -\sigma^2 x^{P-1} d e^{-\frac{x^2}{2\sigma^2}} \\
&= \frac{\sigma^2}{\sqrt{2\pi\sigma^2}} [(x^{P-1} e^{-\frac{x^2}{2\sigma^2}}) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} (P-1) x^{P-2} dx] \\
&= \sigma^2 (p-1) E(X^{P-2})
\end{aligned}
$$

And, If P is odd, $E[X^P] = 0$, In this way, we can calculate the pth moment function of X, if P is even.

1.2 multi-dimensional Gaussian Distriution

- In multi-dimensional Gaussian Distribution, Random Variable X is a vector: $X \subset R^n$

So, we have the probability density function: $\rho(x) = \frac{1}{\sqrt{(2\pi)^n detQ}} e^{-\frac{1}{2}(x-\mu)^T Q^{-1}(x-\mu)}$, in which, μ is the mean vector , $\mu \subset R^n$, Q is covariance matrix, $Q \subset R^{n*n}$.

- Expectation: $E[X] = \mu$

• Since X is a vector, which contains several random variables. We need to know variance of each random variable, and the correlation between each random veritable to describe this distribution. Matrix Q in the probability density function is the covariance matrix, in which, we can find variance of each random variable and their correlations:

$$\Sigma = Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{12} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{1n} & \cdots & \cdots & q_{nn} \end{bmatrix}$$

$Cov[X] = Q,$

$q_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$ is the covariance between X_i, X_j

$q_{ii} = \sigma^2(X_i) = E[(X_i - \mu_i)]$

• Transform multi-dimensional into independent scaler normal distributions.

It's hard to calculate with Covariance matrix in the power position, $\rho(x) = \frac{1}{\sqrt{(2\pi)^n detQ}} e^{-\frac{1}{2}x^T Q^{-1} x}$ Luckily there is a way to decomposit the probability density function, in that way there will be no matrix in the power of e.

Since Q is positive definite and symmetric matrix : Q can be decomposed like this:

$$Q = U \Lambda U^T$$

The columns of U are the eigenvectors of matrix Q and the diagonal elements of Λ are the eigenvalues.

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

$Q^{-1} = U\Lambda^{-1}U^T$, make $y = Ux$

$x^TQ^{-1}x = x^TU^T\Lambda^{-1}Ux = y^T\Lambda^{-1}y = \sum_{i=1}^{n}\frac{y_i^2}{\lambda_i} \quad (1)$

$$\Lambda = \begin{bmatrix} \frac{1}{\lambda_1} & ... & 0 \\ \vdots & \ddots & \vdots \\ 0 & ... & \frac{1}{\lambda_n} \end{bmatrix}$$

Since U consists of eigenvectors of Q, We can choose any U to make:

$det(U^T) = det(U) = \pm 1$

So that $det(Q) = det(U\Lambda U^T) = det(U)^2 det(\Lambda) = \lambda_1\lambda_2...\lambda_n \quad (2)$

From (1) (2)we can transform :

$$\rho(x) = \frac{1}{\sqrt{(2\pi)^n detQ}}e^{-\frac{1}{2}x^TQ^{-1}x} = \frac{1}{\sqrt{(2\pi)^n detQ}}e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{y_i^2}{\lambda_1}} = \prod_{i=1}^{n}\frac{1}{\sqrt{(2\pi\lambda_i)}}e^{-\frac{y_i^2}{2\lambda_i}}$$

2 Conditional Probability

2.1 independent

• Definition : P_1, P_2 are two events.They are independent with each other if the possibility of which, P_1, P_2 happens at the same time, is the product of possibility of P_1, P_2 happens separately.

$$\mu(P_1 \cap P_2) = \mu(P_1)\mu(P_2)$$

• This rule of independent also works for probability density function:

X,Y are independent random variables, if $\rho(x, y) = \rho(x)\rho(y)$

2.2 Conditional Probability

• If P_1, P_2 are not independent with each other, we can calculate conditional probability, that is the possibility of one thing happen, when given the fact that another thing have already happened.

For example: the following equation shows the possibility of P_1 happens when known that P_2 happened

$$\mu(P_1|P_2) = \frac{\mu(P_1 \cap P_2)}{\mu(P_2)}$$

• Bayes's rule tells us we can calculate the conditional probability of event P_1 conditioned on P_2 from the conditional probability of event P_2 conditioned on P_1:

$$\mu(P_2|P_1) = \frac{\mu(P_1|P_2)\mu(P_2)}{\mu(P_1)}$$

• Bayes's rule also works for the probability density function. We can get conditional probability density function of x conditioned on y, from probability density function of y conditioned on x:

$$\rho(x|y) = \frac{\rho(y|x)\rho(x)}{\rho(y)}, \rho(x) = \int_{-\infty}^{\infty} \rho(x,y)dy$$

3 Calculate density function: $X = \Phi(Y)$

Y is a random variable taking values in R^n. $\Phi : R^n \to R^m, X = \Phi(Y)$, suppose density function of y is known: $\rho(y)$. How to calculate density function of x ?

• The random variable Y has some range, so after the operation $\Phi(Y)$ the outcome also have a range, we call it range S

$$X = \Phi(Y), S = \{X|X = \Phi(Y) for\ some\ Y\} = range(Y)$$

• Random variable X has no possibility to go beyond that range :

$$X \not\subset S, \rho_x(x) = 0$$

• if $X \subset S$, We have Prob $\{X \subset (x, x + dx)\ \}$ = Prob $\{Y \subset (y, y + dy)\ \}$

So, $\rho_X(x)|dx| = \rho_Y(y)|dy|$,

Since $X = \Phi(Y)$, $dx = \Phi' dy$,

$$\rho_X(x) = \rho_Y(y)|\frac{dy}{dx}| = \rho_Y(y)\frac{1}{\Phi'(y)}\Big|_{y=\Phi^-1(x)}$$

• Generally : $X = \Phi(Y)$ do not need to be monotonic, this means several Y can map to same X, all we need to do is to add these parts together.

$m = n = 1$

$$\rho_X(x) = \sum_{Y:\Phi(Y)=X} \rho_Y(y) \frac{1}{\Phi'(y)}\Big|_{y=\Phi^-1(x)}$$

$m = n > 1$

$$\rho_X(x) = \sum_{Y:\Phi(Y)=X} \rho_Y(y) \frac{1}{|det(\frac{\partial \Phi}{\partial y})|}\Big|_{y=\Phi^-1(x)}, \quad \frac{\partial \Phi}{\partial y} : Jacobian\ matrix$$

4 Empirical estimation of probability density function

Many times we can't tell the arguments of a possibility density function, even if we have a model in hand, but the following method is a way to guess arguments:

• Suppose: $\rho_{\theta_1,\theta_2,...\theta_n(x)}$, $\theta_1, \theta_2, ...\theta_n$ are arguments of probability density function.

• We have observations from experiments :{ $X_1, X_2, ...X_n$}

• The likelihood if I choose that set of arguments is:

$$L = \prod_{i=1}^{n} \rho_{\theta_1,\theta_2,...\theta_n(x_i)}$$

Take log of the liklyhood:

$$LL = \sum_{i=1}^{n} log\rho_{\theta_1,\theta_2,...\theta_n(x)_i}$$

Then we can guess many sets of arguments $(\theta_1, \theta_2, ...\theta_n)$ and insert them into LL, and choose a set of arguments which can make LL to be Max