

# 0 前言

**贝叶斯滤波**基于贝叶斯公式，通过上一时刻的状态及当前时刻的输入，对当前时刻的状态作出预测，并通过当前时刻的观测对预测作出更新（也可称为纠正），最终实现对当前时刻状态的估计。贝叶斯滤波思想是卡尔曼滤波、粒子滤波等算法的基础。本文从概率基础概念出发，逐步进行贝叶斯滤波的推导。

## 1 概率基础概念拾遗

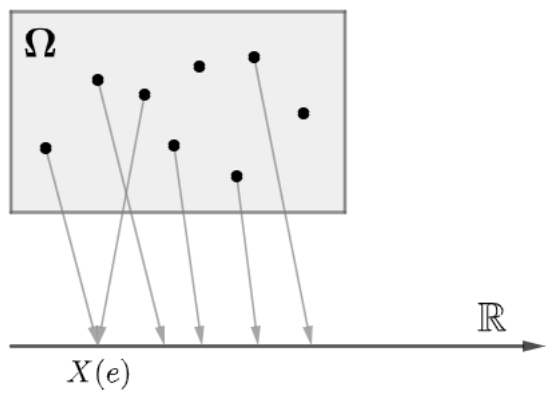
### 1.1 随机变量

#### 1.1.1 定义

定义在样本空间  $\Omega$  上的实值函数：

$$X = X(\omega), \omega \in \Omega$$

被称为**随机变量** (random variable)，如下图所示。 $\omega$  表示样本空间  $\Omega$  里的随机事件，可能与数量有关（如传感器随机测得的前方障碍物的距离值，随机掷骰子得到的点数），也可能与数量无关（如随机掷硬币的结果可能为正面或反面）。



随机变量表示随机试验各种结果的实值单值函数。随机事件不论与数量是否直接有关，都可以数量化，即都能用数量化的方式表达。

“随机变量  $X$  的取值为  $x$ ”就是满足等式  $X(\omega) = x$  的一切  $\omega$  组成的集合，简记为“ $X = x$ ”，这是  $\Omega$  的一个子集，即：

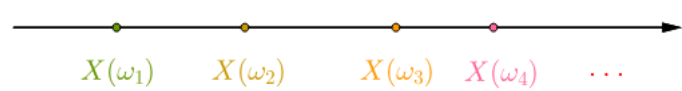
$$“X = x” = \{\omega : X(\omega) = x\} \subset \Omega$$

类似地，有：

$$“X \leq x” = \{\omega : X(\omega) \leq x\} \subset \Omega$$

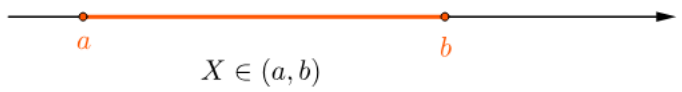
#### 1.1.2 离散型随机变量

如果随机变量的函数值是实数轴上孤立的点（有限个或者无限个），则称为**离散型随机变量**：



#### 1.1.3 连续型随机变量

如果随机变量的函数值是实数轴上某个区间上所有的值（也可以是  $(-\infty, +\infty)$  区间），则称为**连续型随机变量**：



## 1.2 累积分布函数

设  $X$  为一个随机变量，对于任意实数  $x$ ，称：

$$F(x) = P(X \leq x)$$

为随机变量  $X$  的**累积分布函数**（Cumulative Distribution Function，CDF），又称概率分布函数，简称**分布函数**。

若  $X$  为离散型随机变量，存在函数  $f_X(x)$ ，使其分布函数满足：

$$F(x) = P(X \leq x) = \sum_{u \leq x} f_X(u)$$

若  $X$  为连续型随机变量，存在函数  $f_X(x)$ ，使其分布函数满足：

$$F(x) = P(X \leq x) = \int_{-\infty}^x f_X(u)du$$

分布函数满足单调性、有界性和右连续性：

- 单调性  
对任意  $x_1 \leq x_2$ ，有  $F(x_1) \leq F(x_2)$
- 有界性  
对任意  $x$ ，有：

$$0 \leq F(x) \leq 1$$

且有：

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

- 右连续性  
对任意  $x_0$ ，有：

$$\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$$

### 1.3 概率质量函数

假设  $X$  为离散型随机变量，其全部可能取值为  $x_1, x_2, x_3, \dots$ ，其**概率分布列**可表述为：

$X$	$x_1$	$x_2$	$x_3$	$\dots$
$P$	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\dots$

则称：

$$f_X(x) = \begin{cases} p(x) = P(X = x), & x \in \{x_i\}, i = 1, 2, 3, \dots \\ 0, & \text{其它} \end{cases}$$

为离散型随机变量  $X$  的**概率质量函数**（Probability Mass Function，PMF）。

也可采用如下的表示方法：

$$X \sim p(x)$$

读作  $X$  服从  $p(x)$  的概率分布。

概率质量函数表征了离散型随机变量在各特定取值上的概率，其满足**非负性**、**规范性**和**可加性**：

- 非负性

$$f_X(x_i) \geq 0$$

- 规范性和可加性

$$\sum_{i=1}^{\infty} f_X(x_i) = 1$$

离散型随机变量概率质量函数的叠加得到了离散型随机变量的累积分布函数。

### 1.4 概率密度函数

假设  $X$  为连续型随机变量，存在函数  $f_X(x)$ ，且满足：

- 非负性

$$f_X(x) \geq 0$$

- 规范性和可加性

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1$$

则称函数  $f_X(x)$  为连续型随机变量  $X$  的**概率密度函数** (Probability Density Function, PDF) , **连续型随机变量概率密度函数的积分得到了连续型随机变量的累积分布函数**。

概率密度函数表征了连续型随机变量的输出值在某个确定的取值点**附近**的可能性。连续型随机变量的输出值在某个确定的取值点上的概率为0, 但并不代表该事件不会发生。

## 1.5 联合概率

### 1.5.1 二维离散型随机变量的联合概率质量函数

若二维离散型随机变量  $(X, Y)$  所有可能的取值为

$$(x_i, y_j), i, j = 1, 2, 3, \dots$$

且存在函数  $f_{X,Y}(x, y)$ :

$$f_{X,Y}(x, y) = p(x, y) = P(X = x, Y = y), \quad (x, y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots$$

函数  $f_{X,Y}(x, y)$  满足如下性质 (即概率的三大公理) :

- 非负性

$$f_{X,Y}(x_i, y_i) \geq 0$$

- 规范性和可加性

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_i) = 1$$

则称函数  $f_{X,Y}(x, y)$  为  $(X, Y)$  的**联合概率质量函数** (Joint Probability Mass Function) , 又称联合分布列, 此定义可以推广到多维离散型随机变量。

**二维离散型随机变量联合概率质量函数的叠加得到了二维离散型随机变量的联合累积分布函数 (Joint Cumulative Distribution Function) :**

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f_{X,Y}(u, v)$$

### 1.5.2 二维连续型随机变量的联合概率密度函数

若对于二维连续型随机变量  $(X, Y)$ , 存在二元函数  $f_{X,Y}(x, y)$  满足:

- 非负性

$$f_{X,Y}(x, y) \geq 0$$

- 规范性和可加性

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$$

则称函数  $f_{X,Y}(x, y)$  为  $(X, Y)$  的**联合概率密度函数** (Joint Probability Density Function) , 此定义可以推广到多维连续型随机变量。

**二维连续型随机变量联合概率密度函数的积分得到了二维连续型随机变量的联合累积分布函数:**

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du$$

## 1.6 边缘概率

### 1.6.1 二维离散型随机变量的边缘概率质量函数

若二维离散型随机变量  $(X, Y)$  所有可能的取值为

$$(x_i, y_j), i, j = 1, 2, 3, \dots$$

且联合概率质量函数  $f_{X,Y}(x, y)$  为:

$$f_{X,Y}(x, y) = p(x, y) = P(X = x, Y = y), \quad (x, y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots$$

对  $j$  求和所得的函数:

$$f_X(x) = \sum_{j=1}^{\infty} f_{X,Y}(x, y_j) = \sum_{j=1}^{\infty} P(X = x, Y = y_j) = P(X = x), \quad x \in \{x_i\}, i = 1, 2, 3, \dots$$

称为  $X$  的**边缘概率质量函数** (Marginal Probability Mass Function) , 又称边缘分布列。

类似的对  $i$  求和所得的函数:

$$f_Y(y) = \sum_{i=1}^{\infty} f_{X,Y}(x_i, y) = \sum_{i=1}^{\infty} P(X = x_i, Y = y) = P(Y = y), \quad y \in \{y_j\}, j = 1, 2, 3, \dots$$

称为  $Y$  的边缘概率质量函数。

**二维离散型随机变量边缘概率质量函数的叠加得到了二维离散型随机变量的边缘累积分布函数 (Marginal Cumulative Distribution Function) :**

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_{u \leq x} f_X(u) = \sum_{u \leq x} \sum_{j=1}^{\infty} P(X = u, Y = y_j) = \sum_{u \leq x} P(X = u) \\ F_Y(y) &= P(Y \leq y) = \sum_{v \leq y} f_Y(v) = \sum_{v \leq y} \sum_{i=1}^{\infty} P(Y = v, X = x_i) = \sum_{v \leq y} P(Y = v) \end{aligned}$$

## 1.6.2 二维连续型随机变量的边缘概率密度函数

若二维连续型随机变量  $(X, Y)$  的联合概率密度函数为  $f_{X,Y}(x, y)$  , 则:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy$$

称为  $X$  的**边缘概率密度函数** (Marginal Probability Density Function) 。

类似的:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$$

称为  $Y$  的边缘概率密度函数。

**二维连续型随机变量边缘概率密度函数的积分得到了二维连续型随机变量的边缘累积分布函数:**

$$\begin{aligned} F_X(x) &= P(X \leq x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(u, y) dy du \\ F_Y(y) &= P(Y \leq y) = \int_{-\infty}^y f_Y(v) dv = \int_{-\infty}^y \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dx dv \end{aligned}$$

## 1.7 条件概率

### 1.7.1 二维离散型随机变量的条件概率质量函数

若  $(X, Y)$  是二维离散型随机变量, 对于固定的  $j$ , 若  $P(Y = y_j) > 0$  , 则称:

$$f_{X|Y}(x | y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad (x, y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots$$

为  $Y = y$  条件下离散型随机变量  $X$  的**条件概率质量函数** (Conditional Probability Mass Function) 。

同样的, 对于固定的  $i$ , 若  $P(X = x_i) > 0$  , 则称:

$$f_{Y|X}(y | x) = P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad (x, y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots$$

为  $X = x$  条件下离散型随机变量  $Y$  的条件概率质量函数。

简言之, 对于二维离散型随机变量:

$$\text{条件概率质量函数} = \frac{\text{联合概率质量函数}}{\text{边缘概率质量函数}}$$

二维离散型随机变量条件概率质量函数的叠加得到了二维离散型随机变量的条件累积分布函数（Conditional Cumulative Distribution Function）：

$$\begin{aligned} F_{X|Y}(x|y) &= P(X \leq x | Y = y) = \sum_{u \leq x} f_{X|Y}(u|y) = \sum_{u \leq x} \frac{f_{X,Y}(u,y)}{f_Y(y)} \\ F_{Y|X}(y|x) &= P(Y \leq y | X = x) = \sum_{v \leq y} f_{Y|X}(v|x) = \sum_{v \leq y} \frac{f_{X,Y}(x,v)}{f_X(x)} \end{aligned}$$

### 1.7.2 二维连续型随机变量的条件概率密度函数

假设二维连续型随机变量  $(X, Y)$  的联合概率密度函数为  $f_{X,Y}(x, y)$ ，若对于固定的  $Y = y$ ，有边缘概率密度函数  $f_Y(y) > 0$ ，则称函数：

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

为随机变量  $X$  在  $Y = y$  条件下的条件概率密度函数（Conditional Probability Density Function）。

若对于固定的  $X = x$ ，有边缘概率密度函数  $f_X(x) > 0$ ，则称函数：

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

为随机变量  $Y$  在  $X = x$  条件下的条件概率密度函数。

简言之，对于二维连续型随机变量：

$$\text{条件概率密度函数} = \frac{\text{联合概率密度函数}}{\text{边缘概率密度函数}}$$

二维连续型随机变量条件概率密度函数的积分得到了二维连续型随机变量的条件累积分布函数：

$$\begin{aligned} F_{X|Y}(x|y) &= P(X \leq x | Y = y) = \int_{-\infty}^x f_{X|Y}(u|y) du = \int_{-\infty}^x \frac{f_{X,Y}(u,y)}{f_Y(y)} du \\ F_{Y|X}(y|x) &= P(Y \leq y | X = x) = \int_{-\infty}^y f_{Y|X}(v|x) dv = \int_{-\infty}^y \frac{f_{X,Y}(x,v)}{f_X(x)} dv \end{aligned}$$

## 1.8 全概率

### 1.8.1 二维离散型随机变量的全概率公式

对于二维离散型随机变量  $(X, Y)$ ，已知条件概率质量函数：

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad (x,y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots \\ f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad (x,y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots \end{aligned}$$

容易得到二维离散型随机变量联合概率质量函数的等价形式：

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

由边缘概率质量函数与联合概率质量函数间的数量关系，得到二维离散型随机变量的全概率公式：

$$\begin{aligned} f_X(x) &= \sum_{j=1}^{\infty} f_{X,Y}(x, y_j) = \sum_{j=1}^{\infty} f_{X|Y}(x|y_j)f_Y(y_j) \\ f_Y(y) &= \sum_{i=1}^{\infty} f_{X,Y}(x_i, y) = \sum_{i=1}^{\infty} f_{Y|X}(y|x_i)f_X(x_i) \end{aligned}$$

### 1.8.2 二维连续型随机变量的全概率公式

对于二维连续型随机变量  $(X, Y)$ ，已知条件概率密度函数：

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \end{aligned}$$

容易得到二维连续型随机变量联合概率密度函数的等价形式：

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

由边缘概率密度函数与联合概率密度函数间的数量关系，得到二维连续型随机变量的全概率公式：

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y)f_Y(y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx$$

## 1.9 随机过程

**随机过程** (Stochastic Process) 是定义在  $\Omega \times T$  上的二元函数  $X(\omega, t)$ ，简单的定义为一组随机变量的集合，即指定一参数集  $T$ （又称指标集，通常为时间集），其中：

- 对于固定的时间  $t$ ， $X(\omega, t)$  为随机变量，简记为  $X(t)$  或  $X_t$ ；
- 对于固定的  $\omega$ （即固定每一时刻对应的随机变量的取值）， $X(\omega, t)$  为时间  $t$  的一般函数，称为**样本函数** (Sample Function) 或样本轨道 (Sample Path)，简记为  $x(t)$ 。随机过程也可定义为一组样本函数的集合。

注意：随机变量  $X(t)$  或  $X_t$  并不是时间  $t$  的函数，它只表示所有样本函数  $x(t)$  在  $t$  时刻的取值。

依据随机变量和参数集的连续离散情况，随机过程有以下四种类型：

- **连续型随机过程**：随机变量连续，参数连续；
- **离散型随机过程**：随机变量离散，参数连续；
- **连续型随机序列**：随机变量连续，参数离散；
- **离散型随机序列**：随机变量离散，参数离散。

不同于确定过程，**随机过程中的随机变量彼此间不独立**。应该理解，**贝叶斯滤波处理的是一个随机过程**，而且往往是一个连续型随机过程。

# 2 贝叶斯滤波

## 2.1 贝叶斯公式

### 2.1.1 二维离散型随机变量的贝叶斯公式

对于二维离散型随机变量  $(X, Y)$ ，由其条件概率质量函数与全概率公式，容易得到其贝叶斯公式：

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\sum_{i=1}^{\infty} f_{Y|X}(y|x_i)f_X(x_i)}, \quad (x,y) \in \{x_i, y_j\}, i, j = 1, 2, 3, \dots$$

二维离散型随机变量的贝叶斯公式可通过作图的方式轻松证得。

### 2.1.2 二维连续型随机变量的贝叶斯公式

#### (1) 结论

对于二维连续型随机变量  $(X, Y)$ ，由其条件概率密度函数与全概率公式，容易得到其贝叶斯公式：

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx}$$

#### (2) 推导

二维连续型随机变量的贝叶斯公式无法通过作图的方式推得，下面进行公式推导，首先计算二维连续型随机变量的条件累积分布函数：

$$\begin{aligned}
F_{X|Y}(x|y) &= P(X \leq x | Y = y) \\
&= \sum_{u=-\infty}^x P(X = u | Y = y) \quad \Rightarrow \text{化连续为离散无穷小的累加} \\
&= \sum_{u=-\infty}^x \frac{P(Y = y | X = u)P(X = u)}{P(Y = y)} \quad \Rightarrow \text{二维离散型随机变量的贝叶斯公式} \\
&= \lim_{\epsilon \rightarrow 0} \sum_{u=-\infty}^x \frac{P(y \leq Y \leq y + \epsilon | X = u)P(u \leq X \leq u + \epsilon)}{P(y \leq Y \leq y + \epsilon)} \quad \Rightarrow \text{化无穷小为极限形式} \\
&= \lim_{\epsilon \rightarrow 0} \sum_{u=-\infty}^x \frac{[F_{Y|X}(y + \epsilon | u) - F_{Y|X}(y | u)][F_X(u + \epsilon) - F_X(u)]}{F_Y(y + \epsilon) - F_Y(y)} \quad \Rightarrow \text{分布函数性质} \\
&= \lim_{\epsilon \rightarrow 0} \sum_{u=-\infty}^x \frac{[f_{Y|X}(\xi_1 | u) \cdot \epsilon][f_X(\xi_2) \cdot \epsilon]}{f_Y(\xi_3) \cdot \epsilon} \quad \Rightarrow \text{拉格朗日中值定理, } \xi_1, \xi_3 \in (y, y + \epsilon), \xi_2 \in (u, u + \epsilon) \\
&= \lim_{\epsilon \rightarrow 0} \sum_{u=-\infty}^x \frac{f_{Y|X}(y | u)f_X(u)}{f_Y(y)} \cdot \epsilon \quad \Rightarrow \epsilon \rightarrow 0 \text{ 时, } \xi_1 \rightarrow y, \xi_2 \rightarrow u, \xi_3 \rightarrow y \\
&= \int_{-\infty}^x \frac{f_{Y|X}(y | u)f_X(u)}{f_Y(y)} du \quad \Rightarrow \text{积分定义} \\
&= \int_{-\infty}^x \frac{f_{Y|X}(y | x)f_X(x)}{f_Y(y)} dx \quad \Rightarrow \text{替换自变量符号 } u \text{ 为 } x
\end{aligned}$$

故，二维连续型随机变量的条件概率密度函数为：

$$f_{X|Y}(x|y) = \frac{dF_{X|Y}(x|y)}{dx} = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

代入全概率公式：

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx}$$

上式即为二维连续型随机变量的贝叶斯公式，推导完毕。

## 2.2 先验概率、似然概率与后验概率

在二维连续型随机变量的贝叶斯公式中，有如下定义：

- $f_X(x)$  被称为**先验概率密度** (Prior Probability Density)，表示根据以往的经验和分析，在本次试验或采样前便可获得的随机变量  $X$  的概率密度；
- $f_{Y|X}(y|x)$  被称为**似然概率密度** (Likelihood Probability Density)，表示在状态随机变量  $X$  取值为  $x$  的条件下，观测随机变量  $Y$  取值为  $y$  的概率密度，状态为因，观测为果，即**由因推果**；
- $f_{X|Y}(x|y)$  被称为**后验概率密度** (Posterior Probability Density)，表示在观测随机变量  $Y$  取值为  $y$  的条件下，状态随机变量  $X$  取值为  $x$  的概率密度，状态为因，观测为果，即**由果推因**。

此外，当  $y$  为定值时， $\eta = \left[ \int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx \right]^{-1}$  为一常数，常被称为贝叶斯公式的**归一化常数**。

因此，二维连续型随机变量的贝叶斯公式可表示为：

$$\text{后验概率密度} = \eta \times \text{似然概率密度} \times \text{先验概率密度}$$

## 2.3 再谈似然概率

上文中提到，似然概率密度函数  $f_{Y|X}(y|x)$  表示在状态随机变量  $X$  取值为  $x$  的条件下，观测随机变量  $Y$  取值为  $y$  的概率密度。**似然概率密度函数表征了传感器检测精度**，对于给定的状态条件  $X = x$ ，观测结果  $Y = y$  的概率分布通常有三种模型：

### (1) 等可能型

观测值在状态量真值附近呈均匀分布，此时的似然概率密度函数为常数。

### (2) 阶梯型

观测值在状态量真值附近呈阶梯分布，此时的似然概率密度函数为分段常数。

### (3) 正态分布型

观测值在状态量真值附近呈高斯分布，此时的似然概率密度函数为高斯函数：

$$f_{Y|X}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2\sigma^2}}$$

若假定似然概率密度函数为高斯函数，此时，似然概率密度函数的均值  $x$  代表状态量真值， $\sigma$  代表传感器检测精度范围。若同时假定先验概率密度函数为高斯函数，即：

$$f_X(x) \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad f_{Y|X}(y|x) \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

则

$$f_{X|Y}(x|y) \sim \mathcal{N}\left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\mu_2, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)$$

由于

$$\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^2}{1 + \frac{\sigma_1^2}{\sigma_2^2}} < \sigma_1^2$$

且

$$\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_2^2}{1 + \frac{\sigma_2^2}{\sigma_1^2}} < \sigma_2^2$$

故，**后验概率密度函数方差既小于先验概率密度函数方差，也小于似然概率密度函数方差，系统不确定度降低。**

若  $\sigma_1^2 \gg \sigma_2^2$ ，则近似有：

$$f_{X|Y}(x|y) \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

此时，后验倾向于观测。

若  $\sigma_1^2 \ll \sigma_2^2$ ，则近似有：

$$f_{X|Y}(x|y) \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

此时，后验倾向于先验。

## 2.4 贝叶斯滤波推导

### 2.4.1 问题建模

#### (1) 问题描述

对于某状态量随机变量  $X$ ，从初始时刻 0 开始，对其进行观测，得到 0 ~ k 时刻的观测值：

$$y_0, y_1, y_2, \dots, y_k$$

求解 k 时刻状态量随机变量  $X_k$  的最优估计  $\hat{x}_k$ 。

#### (2) 求解思路

以贝叶斯公式为求解方向，将问题转化为求解状态量随机变量  $X_k$  后验概率密度函数的期望：

$$\hat{x}_k = E[f_{X_k}^+(x)]$$

进而要求解状态量随机变量  $X_k$  的先验概率密度函数与似然概率密度函数。我们认为，**k 时刻的状态量随机变量  $X_k$  与且仅与上一时刻的状态量随机变量  $X_{k-1}$  有关，k 时刻的观测量随机变量  $Y_k$  与且仅与 k 时刻的状态量随机变量  $X_k$  有关**，其中的数量关系我们分别称之为**状态方程**与**观测方程**：

$$\begin{cases} X_k = f(X_{k-1}) + Q_k & \Rightarrow \text{状态方程} \\ Y_k = h(X_k) + R_k & \Rightarrow \text{观测方程} \end{cases}$$

$f(x)$  被称为状态转移函数， $h(x)$  被称为观测函数。

对于 0 时刻的初始状态量随机变量  $X_0$ ，认为观测值  $y_0$  即为其真值，其后验概率密度函数即为其先验概率密度函数。我们可以根据经验知识（建模精度和传感器精度）写出 0 时刻的初始状态量随机变量  $X_0$  的后验概率密度函数  $f_{X_0}^+(x)$ 、k 时刻过程噪声随机变量  $Q_k$  的概率密度函数  $f_{Q_k}(x)$  和 k 时刻观测噪声随机变量  $R_k$  的概率密度函数  $f_{R_k}(x)$ 。

#### (3) 符号定义

- 各时刻的状态量随机变量



$$X_0, X_1, X_2, \dots, X_k$$

- 各时刻的观测量随机变量

$$Y_0, Y_1, Y_2, \dots, Y_k$$

- 各时刻的观测值

$$y_0, y_1, y_2, \dots, y_k$$

- 各时刻的过程噪声随机变量

$$Q_1, Q_2, \dots, Q_k$$

- 各时刻的观测噪声随机变量

$$R_1, R_2, \dots, R_k$$

- 各时刻的过程噪声随机变量概率密度函数

$$f_{Q_1}(x), f_{Q_2}(x), \dots, f_{Q_k}(x)$$

- 各时刻的观测噪声随机变量概率密度函数

$$f_{R_1}(x), f_{R_2}(x), \dots, f_{R_k}(x)$$

- 各时刻的状态量随机变量先验概率密度函数

$$f_{X_0}^-(x), f_{X_1}^-(x), f_{X_2}^-(x), \dots, f_{X_k}^-(x)$$

- 各时刻的状态量随机变量后验概率密度函数

$$f_{X_0}^+(x), f_{X_1}^+(x), f_{X_2}^+(x), \dots, f_{X_k}^+(x)$$

- 各时刻状态量随机变量与观测量随机变量的似然概率密度函数

$$f_{Y_1|X_1}(y_1 | x), f_{Y_2|X_2}(y_2 | x), \dots, f_{Y_k|X_k}(y_k | x)$$

#### (4) 重要假设

- $X_0$  分别与  $Q_1, Q_2, \dots, Q_k$  相互独立;
- $X_0$  分别与  $R_1, R_2, \dots, R_k$  相互独立;
- $X_{k-1}$  与  $Q_k$  相互独立;
- $X_k$  与  $R_k$  相互独立。

#### (5) 重要定理

★ 条件概率里的条件可以作逻辑推导。例如：

$$P(X = 1 | Y = 2, Z = 3) = P(X + Y = 3 | Y = 2, Z = 3) = P(X + Y = 3 | Y = 2, Z - Y = 1)$$

### 2.4.2 预测步推导

已知 0 时刻状态量随机变量  $X_0$  的后验概率密度函数  $f_{X_0}^+(x)$ ，状态转移函数  $f(x)$ ，1 时刻过程噪声随机变量  $Q_1$  的概率密度函数  $f_{Q_1}(x)$ ，求解 1 时刻状态量随机变量  $X_1$  的先验概率密度函数  $f_{X_1}^-(x)$ 。

类似二维连续型随机变量贝叶斯公式的推导过程，我们从求解  $X_1$  的先验累积分布函数  $F_{X_1}^-$  入手。

$$\begin{aligned}
F_{X_1}^-(x) &= P(X_1 \leq x) \\
&= \sum_{u=-\infty}^x P(X_1 = u) \Rightarrow \text{化连续为离散无穷小的累加} \\
&= \sum_{u=-\infty}^x \sum_{v=-\infty}^{+\infty} P(X_1 = u | X_0 = v) P(X_0 = v) \Rightarrow \text{全概率公式} \\
&= \sum_{u=-\infty}^x \sum_{v=-\infty}^{+\infty} P[X_1 - f(X_0) = u - f(v) | X_0 = v] P(X_0 = v) \Rightarrow \text{条件概率里的条件可以作逻辑推导} \\
&= \sum_{u=-\infty}^x \sum_{v=-\infty}^{+\infty} P[Q_1 = u - f(v) | X_0 = v] P(X_0 = v) \Rightarrow \text{状态方程} \\
&= \sum_{u=-\infty}^x \sum_{v=-\infty}^{+\infty} P[Q_1 = u - f(v)] P(X_0 = v) \Rightarrow \mathbf{X}_{k-1} \text{与} \mathbf{Q}_k \text{相互独立} \\
&= \sum_{u=-\infty}^x \left\{ \lim_{\epsilon \rightarrow 0} \sum_{v=-\infty}^{+\infty} f_{Q_1}[u - f(v)] \cdot \epsilon \cdot f_{X_0}^+(v) \cdot \epsilon \right\} \Rightarrow \text{类似二维连续型随机变量贝叶斯公式的推导过程，将点概率化为概率} \\
&= \sum_{u=-\infty}^x \left\{ \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{+\infty} f_{Q_1}[u - f(v)] f_{X_0}^+(v) dv \cdot \epsilon \right\} \Rightarrow \text{积分定义} \\
&= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{Q_1}[u - f(v)] f_{X_0}^+(v) dv du \Rightarrow \text{积分定义} \\
&= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{Q_1}[x - f(v)] f_{X_0}^+(v) dv dx \Rightarrow \text{替换自变量符号} u \text{为} x
\end{aligned}$$

故，1 时刻状态量随机变量  $X_1$  的先验概率密度函数为：

$$f_{X_1}^-(x) = \frac{dF_{X_1}^-(x)}{dx} = \int_{-\infty}^{+\infty} f_{Q_1}[x - f(v)] f_{X_0}^+(v) dv$$

推导完毕。可以发现，**先验概率密度函数本质来源于状态方程。**

### 2.4.3 更新步推导

已知 1 时刻观测量随机变量  $Y_1$  的取值  $y_1$ ，求解 1 时刻状态量随机变量与观测量随机变量的似然概率密度函数  $f_{Y_1|X_1}(y_1 | x)$ ，并联合预测步得到的 1 时刻状态量随机变量  $X_1$  的先验概率密度函数  $f_{X_1}^-(x)$ ，求解 1 时刻状态量随机变量  $X_1$  的后验概率密度函数  $f_{X_1}^+(x)$ 。

首先，求解似然概率密度函数  $f_{Y_1|X_1}(y_1 | x)$ ：

$$\begin{aligned}
f_{Y_1|X_1}(y_1 | x) &= \lim_{\epsilon \rightarrow 0} \frac{F_{Y_1|X_1}(y_1 + \epsilon | x) - F_{Y_1|X_1}(y_1 | x)}{\epsilon} \Rightarrow \text{导数的定义} \\
&= \lim_{\epsilon \rightarrow 0} \frac{P(y_1 \leq Y_1 \leq y_1 + \epsilon | X_1 = x)}{\epsilon} \Rightarrow \text{累积分布函数的性质} \\
&= \lim_{\epsilon \rightarrow 0} \frac{P[y_1 - h(x) \leq Y_1 - h(X_1) \leq y_1 - h(x) + \epsilon | X_1 = x]}{\epsilon} \Rightarrow \text{条件概率里的条件可以作逻辑推导} \\
&= \lim_{\epsilon \rightarrow 0} \frac{P[y_1 - h(x) \leq R_1 \leq y_1 - h(x) + \epsilon | X_1 = x]}{\epsilon} \Rightarrow \text{观测方程} \\
&= \lim_{\epsilon \rightarrow 0} \frac{P[y_1 - h(x) \leq R_1 \leq y_1 - h(x) + \epsilon]}{\epsilon} \Rightarrow \mathbf{X}_k \text{与} \mathbf{R}_k \text{相互独立} \\
&= \lim_{\epsilon \rightarrow 0} \frac{F_{R_1}[y_1 - h(x) + \epsilon] - F_{R_1}[y_1 - h(x)]}{\epsilon} \Rightarrow \text{累积分布函数的性质} \\
&= f_{R_1}[y_1 - h(x)] \Rightarrow \text{导数的定义}
\end{aligned}$$

可以发现，**似然概率密度函数本质来源于观测方程。**

然后，联合预测步得到的 1 时刻状态量随机变量  $X_1$  的先验概率密度函数  $f_{X_1}^-(x)$ ，求解 1 时刻状态量随机变量  $X_1$  的后验概率密度函数  $f_{X_1}^+(x)$ ：

$$f_{X_1}^+(x) = \eta_1 \cdot f_{Y_1|X_1}(y_1 | x) \cdot f_{X_1}^-(x) = \eta_1 \cdot f_{R_1}[y_1 - h(x)] \cdot f_{X_1}^-(x)$$

其中，归一化常数  $\eta_1$  为：

$$\eta_1 = \left[ \int_{-\infty}^{+\infty} f_{Y_1|X_1}(y_1 | x) f_{X_1}^-(x) dx \right]^{-1} = \left\{ \int_{-\infty}^{+\infty} f_{R_1}[y_1 - h(x)] f_{X_1}^-(x) dx \right\}^{-1}$$

推导完毕。

### 2.4.4 递推流程

由预测步和更新步的推导结果，可得到由 0 时刻状态量随机变量  $X_0$  的后验概率密度函数  $f_{X_0}^+(x)$  到 k 时刻状态量随机变量  $X_k$  的后验概率密度函数  $f_{X_k}^+(x)$  的递推流程：

$$\begin{aligned} f_{X_0}^+(x) &\xRightarrow{\text{预测}} f_{X_1}^-(x) = \int_{-\infty}^{+\infty} f_{Q_1}[x - f(v)]f_{X_0}^+(v)dv \xRightarrow{\text{观测更新}} f_{X_1}^+(x) = \eta_1 \cdot f_{R_1}[y_1 - h(x)] \cdot f_{X_1}^-(x) \\ &\xRightarrow{\text{预测}} f_{X_2}^-(x) = \int_{-\infty}^{+\infty} f_{Q_2}[x - f(v)]f_{X_1}^+(v)dv \xRightarrow{\text{观测更新}} f_{X_2}^+(x) = \eta_2 \cdot f_{R_2}[y_2 - h(x)] \cdot f_{X_2}^-(x) \\ &\dots \\ &\xRightarrow{\text{预测}} f_{X_k}^-(x) = \int_{-\infty}^{+\infty} f_{Q_k}[x - f(v)]f_{X_{k-1}}^+(v)dv \xRightarrow{\text{观测更新}} f_{X_k}^+(x) = \eta_k \cdot f_{R_k}[y_k - h(x)] \cdot f_{X_k}^-(x) \end{aligned}$$

其中，归一化常数  $\eta_k$  为：

$$\eta_k = \left\{ \int_{-\infty}^{+\infty} f_{R_k}[y_k - h(x)]f_{X_k}^-(x)dx \right\}^{-1}$$

最终，可得到 k 时刻状态量随机变量  $X_k$  的最优估计  $\hat{x}_k$ ：

$$\hat{x}_k = E[f_{X_k}^+(x)] = \int_{-\infty}^{+\infty} x f_{X_k}^+(x)dx$$

## 2.4.5 完整算法框架

### (1) 设初值

初始 0 时刻状态量随机变量  $X_0$  的后验概率密度函数：

$$f_{X_0}^+(x)$$

### (2) 预测步

k 时刻状态量随机变量  $X_k$  的先验概率密度函数：

$$f_{X_k}^-(x) = \int_{-\infty}^{+\infty} f_{Q_k}[x - f(v)]f_{X_{k-1}}^+(v)dv$$

### (3) 更新步

k 时刻状态量随机变量  $X_k$  的后验概率密度函数：

$$f_{X_k}^+(x) = \eta_k \cdot f_{R_k}[y_k - h(x)] \cdot f_{X_k}^-(x)$$

归一化常数  $\eta_k$ ：

$$\eta_k = \left\{ \int_{-\infty}^{+\infty} f_{R_k}[y_k - h(x)]f_{X_k}^-(x)dx \right\}^{-1}$$

### (4) 求解状态量后验估计

k 时刻状态量随机变量  $X_k$  的后验估计：

$$\hat{x}_k^+ = E[f_{X_k}^+(x)] = \int_{-\infty}^{+\infty} x f_{X_k}^+(x)dx$$

## 2.5 贝叶斯滤波的缺点及解决方法

### 2.5.1 缺点

从上文的推导及结论中可以发现，求解预测步中的先验概率密度函数  $f_{X_k}^-(x)$ 、更新步中的归一化常数  $\eta_k$ 、最终的最优估计  $\hat{x}_k$  时均涉及到**无穷积分**，而大多数情况无法得到解析解，使得贝叶斯滤波算法的直接应用十分困难。

### 2.5.2 解决办法

为了解决贝叶斯滤波中的无穷积分问题，通常从两个角度出发：

#### (1) 作理想假设

- 假设状态转移函数  $f(x)$  和观测函数  $h(x)$  均为线性函数，过程噪声随机变量  $Q_k$  和观测噪声随机变量  $R_k$  均服从均值为 0 的正态分布——卡尔曼滤波（Kalman Filter）

- 假设状态转移函数  $f(x)$  和（或）观测函数  $h(x)$  为非线性函数，过程噪声随机变量  $Q_k$  和 观测噪声随机变量  $R_k$  均服从均值为 0 的正态分布——扩展卡尔曼滤波（Extended Kalman Filter）和无迹卡尔曼滤波（Unscented Kalman Filter）

## (2) 化连续为离散

将无穷积分转化为数值积分，一般有以下方法：

- 高斯积分（不常用）
- 蒙特卡罗积分（粒子滤波，Particle Filter）
- 直方图滤波

针对本节内容中提到的卡尔曼滤波、扩展卡尔曼滤波、无迹卡尔曼滤波、粒子滤波、直方图滤波等常用滤波算法，将在后续文章中进行详细展开讨论。

# 3 参考

1. [马同学数学课程](#)
2. [百度百科 - 分布函数](#)
3. [百度百科 - 随机过程](#)
4. [如何从深刻地理解随机过程的含义？](#)
5. [b站忠实的王大头《贝叶斯滤波与卡尔曼滤波》系列教学视频](#)
6. [你真的搞懂贝叶斯滤波了吗？](#)