第一节

贝叶斯滤波思想卡尔曼滤波和粒子滤波是它的多种实现，实际工程中使用的卡尔曼滤波变种，粒子滤波使用也比较多。

基础：概率论

目的：熟悉卡尔曼滤波和粒子滤波原理并学习应用它们。

第二节

贝叶斯滤波：用贝叶斯公式，对随机信号进行处理，减小它的不确定度。这里的随机信号实际是随机过程。

随机过程定义：x1,x2,...,xn都是随机变量，彼此之间不独立。（无法做随机试验）

随机试验：在相同条件下，试验可重复进行（随机试验之间相互独立，才能用大数定律，将频率赋值给概率）；一次试验，结果不确定，所有可能的结果已知；试验之前，实验结果预先知道。

如何研究随机过程：x1,...,xn不独立

1.找到xk=f(xk-1)的关系，就可以找到p(xk)=f(p(xk-1))的关系，因为前后时刻的变量之间有关系，这也体现了不独立性。

2.初值的选取：p(x1)=?

有的初值可以做随机试验，如随机游走xk = xk-1 + D,其中D的值有两种可能，向前走1米的概率是0.5，向后走1米的概率是0.5。初值p(x0)=0。

有的初值不可以做随机试验，只能使用主观概率。

假设随机过程变量之间的关系已经找到，如何设置初值呢？

p(x1) 使用主观概率时，不同的主观概率导致不同的结果。但我们希望尽量削弱不同主观概率带来的差异，这可以通过引入外部观测（证据，信息）即主观概率（先验概率，先于实验的概率）通过外部观测的修正（贝叶斯公式）可以得到相对客观的概率（后验概率，实验之后的概率）。

第三节

先验概率：实验之前的概率

后验概率：实验之后的概率

书《概率机器人》PDF与CDF混用，独立、无关、没有影响三个概念模糊

随机变量用大写字母表示，小写字母表示随机变量的取值，代表随机实验一个可能的结果。

例：抛硬币正面朝上 X=1 做一次随机实验，结果为正面朝上。

离散随机变量P(X = x) = Px

连续随机变量P(X <x) = 积分

条件概率：离散型随机变量，连续型随机变量

例：测温度，今天多少度？

答：首先给出先验概率分布，P(T=10) = 0.8,P(T=11) = 0.2

其次用温度计测量，Tm 为10.3摄氏度

最后求后验概率分布，P(T=10|Tm =10.3) = P(Tm = 10.3|T=10)P(T=10)/P(Tm=10.3)

P(T=11|Tm =10.3) = P(Tm = 10.3|T=11)P(T=11)/P(Tm=10.3)

标注的部分是似然概率：代表观测的准确度，(Tm = 10.3|T=10)当真实温度T=10时，温度计测得温度为10.3的概率。

第四个概率P(Tm=10.3)：温度计测量值为10.3的概率，利用全概率公式计算有，

P(Tm=10.3) = P(Tm = 10.3|T=10)\*P(T=10) + P(Tm=10.3|T=11)\*P(T=11)

T=10,T=11代表随机试验一个结果，结果不会影响分布律

而P(Tm=10.3)与T的取值无关，之和T的分布律有关，先验概率和似然概率已经确定，所以这个概率是一个常数。

改写贝叶斯公式P(T=10|Tm =10.3) = P(Tm = 10.3|T=10)P(T=10)const

即后验概率 = 似然概率\*先验概率\*常数

所有后验概率之和为1，常数 = 1/求和(似然概率\*先验概率)

第四节

连续随机变量下的贝叶斯滤波

《概率论与数理统计》茆诗松第一版（十章）推荐

传感器的准确度，似然：likelihood 可能性

独立未必没有函数关系 Y=f(X) ,Y与X可能独立，也可能不独立。

连续随机变量的贝叶斯公式：

离散随机变量贝叶斯公式，上一节已讲即后验概率=常数\*似然概率\*先验概率

那么连续 P(X<x|Y=y) 化积分为求和。

虽然形式相同，但是随机变量下的贝叶斯公式是通过推到得到的。

第五节

似然概率与狄拉克函数

X表示系统状态，Y表示观测，它们都是连续的随机变量，贝叶斯公式该如何使用。

例：测温度

首先给一个先验概率密度，然后得到观测值通过温度计测量，计算后验概率。

似然概率模型：等可能模型（概率密度为常数）、阶梯型（概率密度值分段）、直方图型（化连续为离散）、正态分布型

后验分布与观测值有关。

重要定理：若先验概率服从正态分布，似然概率服从正态分布，后验概率就会服从正态分布。

Mathmatica软件计算

先验概率（预测值），似然概率（观测值），后验概率的方差比前两者都要小。

狄拉克函数

似然概率的方差趋于0，就会变成狄拉克函数。它本质上为离散随机变量的必然事件的概率密度。

第六节

随机过程的贝叶斯滤波

1-5 X 先验概率 Y观测 求后验概率（一个随机变量）

现在一个随机过程有多个随机变量了，有一个初值，有k个观测值

第七节

卡尔曼滤波

首先，回顾贝叶斯滤波的三个公式。

卡尔曼滤波相对于贝叶斯滤波多了一些假设，即假设状态方程和观测方程都是线性方程，另外Q和R服从正态分布。

应用 more is different

问题：

1. 请用计算机生成一个含有正态噪声的信号，并用KF滤波；
2. 传感器融合：问题，已知x=t^2为信号，有两个不同传感器对x进行观测

产生了两组测量数据ya1-yak,yb1-ybk

已知两个观测方程，与观测的噪声

求传感器融合下的后验估计，用代码实现。

注意：使用矩阵形式卡尔曼滤波，f h可以不是方阵列，阶数也可以不相同，泰勒展开。

马尔可夫与观测独立

第九节

粒子滤波原理详解，应用最广泛，原理最复杂，术语最多，主要用于静态环境，或动态可预测环境，如汽车，电池电量等，视频跟踪，封闭环境导航等。

（理论、代码、重采样方法）

从贝叶斯滤波开始，无穷积分没有解析解。

由大数定律引发的遐想

大数定律：设x为随机变量，期望存在，对x采样，做n次随机实验，则有。。。

暗示了当n足够大时，期望和平均值相等，积分和求和近似相等。

大数定律暗示了可以用一堆粒子来近似概率密度，这就是粒子滤波。

粒子越多，精度越高。

缺点：需要大量粒子，如何让少量粒子表示PDF。

引入权重的概念，让少量粒子有较高的权重。

粒子的位置与权重完全决定了cdf，所有权重加起来等于1。

所谓粒子是每次随机实验中随机变量的取值。

粒子的位置和权重该如何设置呢？如何采样？

权重：原则PDF高的地方wi大，可以按PDF比例分配，也可以n很大时权重1/n，也可以两者综合。

贝叶斯滤波，已知Xn的pdf为fn(x),在fn(x)中采了n个样本（怎么采）

正态分布比较好采样，软件中有库。

那么概率密度就可以用粒子的形式表示出来。

将概率密度求事件，如何操作。（本质是对事件做随机实验，如果事件比较复杂可以拆分为多个小事件的和）

通用方法：利用傅里叶变换来做，思想是将事件拆分为子事件。尽量保证粒子数不改变。

粒子滤波算法

1. 设置初值
2. 生成X0的样本
3. 生成X0i的权重w0i
4. 生成X1-的样本，根据X0的事件
5. 计算f1-，改变粒子位置，但未改变粒子权重
6. 预测步
7. 观测到了一个数据
8. 后验概率计算，更新步骤并未改变粒子位置，改变了粒子权重。

重采样：为了解决粒子退化的问题，少量粒子有较高权重，大量的粒子权重极低。

为什么粒子会退化：因为粒子数不能太多，似然概率一般假设为正态分布，权重大的粒子是权重小的粒子的近1000倍。

粒子退化的坏处：权重将不会更新，失去了更新的作用。

为了解决粒子退化问题，重采样应运而生。

当前步的更新会导致粒子退化，粒子退化会导致下一步的更新失效，粒子退化本身没什么问题（代表传感器很精确）。

重采样，按概率进行复制与淘汰，权重高的更有可能被多次复制，从而保证整个粒子数不定。把所有粒子权重设置为1/n。可以缓解但是不能根除粒子退化问题，只能解决更新失效的问题。

重采样有一定减弱粒子退化的能力，更主要减弱更新失效问题，必然导致粒子多样性的丧失。减慢粒子滤波的速度。

第十二节粒子滤波拾遗

采样方法：如何在复杂PDF上CIA杨

预测方程：X=f(t), 怎么由X = f(t) 高精度改写 Xk = F(Xk-1)

一般均匀分布和正态分布是比较好采样的。

采样粒子的特点：PDF大的粒子多，PDF小的粒子少

思想：通过某种方式在均匀分布的粒子中去掉一些粒子，从均匀分布到其他分布（复杂分布、正态分布）。

具体怎么做：高PDF的地方有更大的概率保留，低PDF的地方有更大的概率去掉。

重采样的思想也是类似的。

1. 均匀分布生成粒子x1,x2,...xn
2. 取一个直线M，使得M>f(x)
3. 对每一个采样的粒子做审判

生成一个随机数a,a~U(0,M)

看a将在哪一个区间

若a在（0，f(xi)）xi保留，反之去掉（参考了重采样的思想实现了从均匀分布到任意分布的采样）

对高斯分布采样也可得到任意分布

（以上为接收拒绝采样法的基本思想）

接收拒绝采样法：

待采样f(x)，容易采样的g(x)又叫建议分布

1. 找到M，使得Mg(x)>f(x)
2. 在g(x)采样一个粒子x1
3. 生成一个a服从均匀分布U(0,Mg(x)),若a在(0,f(x))之间，则保留，反之则拒绝
4. 在g(x)采样一个粒子x2
5. ...

提议分布是不唯一的，M取得越小越好，粒子接受概率会更高，采样效率会更高。

预测方程的写法（贝叶斯滤波核心）

观测方程一般更好写一些

预测方程难写

一般X是随时间变化的函数，将t采样后就是离散的函数

我们希望由Xk与tx的关系找到Xk和Xk-1、t的关系

常微分方程的数值解法：欧拉法、改进欧拉法、龙格库塔法等在不改变维数的情况下提高精度。

第十三节EKF

由状态方程和观测方程得到贝叶斯滤波

预测步和更新步

若状态方程和观测方程是线性的且预测误差和观测误差服从正态分布

贝叶斯滤波的无穷积分有解析解

贝叶斯滤波变成卡尔曼滤波

若状态方程和观测方程为非线性函数，那么衍生出粒子滤波

但是需要重采样比较慢，粒子滤波的精度是随着粒子的增多而提高的

更快的但精度第一点的方法就是扩展卡尔曼滤波

对状态方程和观测方程线性化，设状态变量服从正态分布（但实际不是，存在误差）

泰勒展开保留一阶项就是线性关系的了。

公式是根据贝叶斯公式推到而来的。

预测是在前一时刻估计展开，观测是在当前时刻预测展开。

预测步（期望方差），卡尔曼增益，更新步（期望方差）

EKF算法：在估计方差和卡尔曼增益得时候用线性化后的雅克比

1. UKF

贝叶斯滤波中最复杂的算法

EKF速度快，精度差，较稳定

PF 速度慢，精度高，较稳定

UKF 速度较快，精度较高，很不稳定，易崩溃（矩阵维度越高越容易崩溃）

UKF 学术界较多，工业界较少

假设X~N(0,1) Y = f(x)

EKF 对函数近似 f(x) = f(0) + f’(0)(X - 0) (线性化)

UKF 对pdf近似， Y= f(x),E(Y)和E(Y^2)

若能通过某种方式直接求出Y的期望与方差，用一个期望方差与其相同的正态分布取近似它的Y概率密度，将非线性问题转化为卡尔曼滤波问题

怎么做？

UT变换（无迹变换，无损变换）

X~N(0,1) Y = g(x) 求E(Y) D(Y)

精确解存在，但是积分不好算

UT变换，取3个粒子近似代替N（0,1）

Y = f(x)的方差怎么算呢，非线性变换后的y，求均值和方差

用UT变换所计算出来的期望和方差是对真实的值的近似

EKF具有一阶精度，UKF具有二阶精度

UKF中由于正态分布近似，所有随机变量都是正态分布，可以用卡尔曼滤波来算

讲者表示不会

以上是根据贝叶斯估计得来的

矩阵形式的卡尔曼滤波（需要协方差矩阵正定，才能做LLT分解）

需要对矩阵做2次分解，矩阵一旦不正定算法就崩溃了。

第十五节 UKF代码

第十六节（理论完结篇）

1. 总结
2. 拾遗（怎么和市面上的教程相兼容，独立无关马尔可夫，UT变换的精度问题）
3. 直方图滤波（粒子滤波，多个点，直方图滤波是一个一个柱形，中心点是粒子，面积是权重）

贝叶斯滤波的本质是信息的融合

信号部分反应本质，并且往往带有噪声

滤波准原则：传感器所传递的信息是否独立，Q与R的设置

重要的几节：第三讲、第四讲、第六讲、第七讲、第十讲、第十二讲

独立、无关、马尔可夫

数学上的无关与统计上的无关

数学上的独立与统计上的独立

数学 XY之间不存在一个映射，XY独立或无关

概率 没有无关的概念，只有不相关 Cov(X,Y) =0 XY不相关

独立 P(A|B) = P(A) 则A与B相互独立，B的发生与否不会影响A的分布律（可能性）

马尔可夫链：未来只与现在有关，与过去无关

更贴切的说明：未来的可能性只与现在的可能性有关，与过去的可能性无关（这里的只与限定现在，不是过去）

若xk = f(xk-1) + Qk 是一个马尔可夫链，充分条件

贝叶斯滤波只能处理马尔可夫链的过程

UT变换的二阶精度可以保证二阶期望没有误差，但是更高阶的方差必然存在误差。