《概率论与数理统计》随机变量及其分布

1 重点难点总结

1.1 随机变量

随机变量可以理解为一个法则/函数/映射，将随机试验的每一个结果和实数对应起来

定义 设随机试验的样本空间$S=e.\ X=X(e)$是定义在样本空间S上的实值单值函数，称$X=X(e)$为随机变量

随机变量的取值随试验的结果而定，而试验的各个结果出现有一定的概率，因而随机变量的取值有一定的概率

1.2 离散型随机变量及其分布律

离散型随机变量：取值是**有限个**或者**无限可列**多个的随机变量

离散型随机变量的例子：

- 丢骰子丢到的点数：只可能是0，1，2，...，6（有限个）
- 120急救电话台一昼夜收到的呼唤次数0，1，2，...（无限可列个）

1.2.1 离散型随机变量X的概率

设离散型随机变量X所有可能取值为$x_k(k=1,2,\cdots)$，X取各个可能值的概率，即事件$X=x_k$的概率为

$$P\{X=x_k\}=p_k,k=1,2,\cdots$$

上式也称为离散型随机变量X的**分布律**

p_k满足两个条件

- $p_k\geq 0,k=1,2,\cdots$
- $\sum_{k=1}^{\infty}p_k=1$

1.2.2 0-1分布

随机变量只能取0和1两个值，分布律为

$$P\{X=k\}=p^k(1-p)^{1-k},\quad k=0,1\quad (0<p<1)$$

则称X服从以p为参数的（0-1）分布或两点分布

例子：

- 新生儿的性别
- 商品是否合格

1.2.3 伯努利试验、二项分布

伯努利试验E的可能结果只有两个A和$\overline{A}$，将E独立重复的进行n次，则称这一串重复的独立试验为n重伯努利试验

1.2.4 泊松分布

泊松分布随机变量X的概率分布函数为

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!} \quad k=0,1,2,\cdots$$

例子：

- 深圳一段时间内发生交通事故的次数

以np为参数的二项分布概率值可以由参数为$\lambda=np$的泊松分布近似

1.3 随机变量的分布函数

定义：设X是一个随机变量，x是任意实数，函数

$$F(x)=P\{X\leq x\}, \quad -\infty<x<\infty$$

称为X的分布函数

若已知任意实数x_1,x_2就可以和分布函数，就可以知道X落在他们这个区间的概率

1.4 连续型随机变量及其概率密度

如果对于随机变量X的分布函数$f(x)$存在非负可积函数，使对于任意实数x有

$$F(x)=\int_{-\infty}^{x}f(t)dt$$

则称X为**连续型随机变量**，$f(x)$称为X的**概率密度函数**，简称**概率密度**

重点：若$f(x)$在x处连续，则$F'(x)=f(x)$

1.3.1 均匀分布

概率密度函数

$$f(x)=\begin{cases}\frac{1}{b-a} & a<x<b\\ 0 & \text{其他},\end{cases}$$

在区间a~b上服从均匀分布的随机变量X，落在其中任意等长度的子区间内的可能性是相同的

1.3.2 指数分布

概率密度函数

$$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & \text{其他},\end{cases}$$

其中$\theta>0$为常数，可以得到随机变量X的分布函数为

$$F(x)=\begin{cases}1-e^{\frac{-x}{\theta}} & x>0\\ 0 & \text{其他},\end{cases}$$

1.3.3 正态分布/高斯分布

概率密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty<x<\infty$$

性质：

- 曲线关于$x = \mu$对称
- 当$x = \mu$时取到最大值

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$$

z-score

1.4 随机变量的函数的分布

在实际中，我们常对某些随机变量的函数更感兴趣，例如测量圆轴截面的直径d，关心的却是截面面积$A = \frac{1}{4}\pi d^2$

2 学习过程中的思考、心得

随机变量和普通函数的差异在哪里?

随机变量的取值随试验的结果而定，在试验开始前不能预知它取什么值，且它的取值有一定的概率，但是函数输入确定的情况下输出的结果就是确定的

分布律记法：概率1以一定的规律分布在各个可能的随机变量取值上

分布函数是一个普通的函数，但是正是通过它，我们将能用数学分析的方法来研究随机变量

指数分布的分布函数是无记忆的，例如原件对他已使用过s个小时没有记忆

随机变量函数分布的解法思路：在计算"Y<=y"的时候，转换成"g(X)<=y"，再转换成一个和他等价的关于X的不等式，然后带入到X的方程中

3 学科交叉应用

机器学习中，有一个算法称为随机森林，其原理大概可以解释为，利用许许多多小的决策树（小的分类器），通过少数服从多数的形式来投票决定，那么如何说明随机森林的效果一定会优于单颗决策树呢?

随机森林的本质是一种装袋集成算法（bagging），对基评估器的预测结果进行平均或用多数表决的原则来决定集成评估器的结果。比如我们建立了25棵树，对任何一个样本而言，平均或多数表决的原则下，当且仅当有13棵以上的树判断错误的时候，随机森林才会判断错误。单独一颗决策树对红酒数据集的准确率在0.85上下浮动，假设一棵树判断错误的可能性位0.2ϵ，那13棵树以上都判断错误的可能性是

$$e_{random_forest} = \sum_{i=13}^{25} C_{25}^{i}(1-\epsilon)^{25-i} = 0.00369$$