.....?

[github: https://github.com/hfl15]

博客园 首页 新随笔 联系 订阅 管理

数学基础（1）~ 概率论基础知识

概率论基础

出处：http://www.cnblogs.com/fanling999/p/6702297.html
参考：盛骤, 谢式千, 潘承毅. 概率论与数理统计, 第四版[M]. 高等教育出版社, 2008.

目录

0. 前言

本文主要旨在对概率统计的基础概念与知识进行概要的总结，一来是对自身学习过程的一个回顾，二来是便于使用到时可以参考，具体细节可以参考给出的书目1~5章。
概率论是数理统计的基础，也是很多机器学习模型的支撑，对相关的概念有所掌握和理解对后面的学习有很大帮助。

（1）给出了概率论中基本概念的定义，即随机试验、样本空间、随机事件、频率和概率、等可能概型、条件概率和独立性等概念。其中较为重要的概念有（a）概率是频率是客观存在，当试验次数充分大时频率将收敛于概率。（b）几个公式：乘法公式、全概率公式、贝叶斯公式，清楚理解其中划分的概念。

（2）对随机变量及其分布进行了总结。随机变量可以分为：离散型和非离散型，其中非离散型主要以连续型随机变量为主。离散型研究其分布律，而连续型则研究其概率密度（某一个点的改变不影响总体，因此求积分时开区间和闭区间没有太大区别）。常用的离散型随机分布：（0-1）分布、二项分布（伯努利试验）、泊松分布。常用的连续型随机分布有：均匀分布、指数分布、正态分布（高斯分布）。其中正态分布最为常用。

（3）多维随机变量的分布，可以基于乘法公式和条件概率公式对相应的概率进行计算，计算的是多个试验发生的联合概率。相互独立的概念较为重要，很多模型都基于独立性前提。

（4）随机变量的数字特征反应了分布的一些统计特性，数学期望反应了分布的均值、方差反应了分布的集中程度、协方差和相关系数反应了随机变量之间的相互作用关系。矩是数理统计中参数估计的重要参考。当随机变量很多时，协方差矩阵是很好的分析工具。

（5）大数定律和中心极限定理是概率论的核心基本理论。

1. 概率论的基本概念

1.1 基本概念

随机试验 （E） Random Experiment

（1）可以在相同的条件下重复地进行
（2）每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果
（3）进行一次试验之前不确定哪一个结果会出现
例子：抛一枚硬币，观察正面，反面出现的情况

样本空间 (S) Sample Space

随机试验所有可以能的结果组成的集合

样本点 Sample point

样本空间的元素，即每个可能的结果

随机事件 Random Event

随机试验E的样本空间S的子集称为随机事件

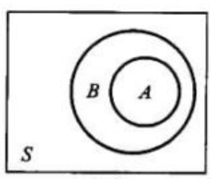
基本事件 fundmental Event

样本空间的单个元素，一个可能结果构成的集合

必然事件（全集）、不可能事件（空集） certain event / impossible event

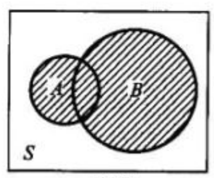
事件的关系与事件的运算 （类似于集合运算）

相等、和事件、积事件、差事件、互不相容（互斥）、逆事件（对立事件）



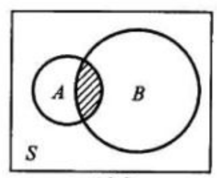
$A\subset B$

图 1—1



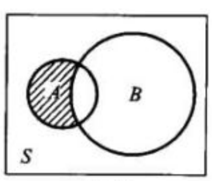
$A\cup B$

图 1—2

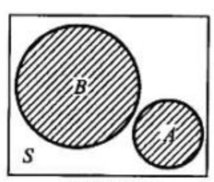


$A\cap B$

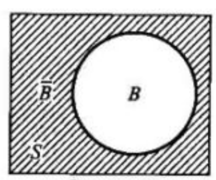
图 1—3



$A-B$



$A\cap B=\varnothing$



$B\cup\bar{B}=S, B\cap\bar{B}=\varnothing$

1.2 频率与概率

频率

定义：在相同条件下，进行n次试验，在这n次试验中，事件A发生的次数，称为事件A发生的**频数**，比值：*f=频数 / 试验次数*，称为事件A发生的**频率**。
基本性质：（1）0 <= f <= 1；(2) f(S) = 1；(3) 两两互不相融事件的可列可加性。
稳定性：当试验重复次数很大时，频率趋于稳定，可以用来表征事件A发生可能性的大小。

概率

定义： 设E是随机试验，样本空间为S，对于E的每一个事件A赋予一个实数，记为P(A)，称为A的**概率**
性质：（1）非负性 P(A) >= 0；(2)规范性，P=1表示必然事件，等于P(S)；(3) 可列可加性（互不相容事件）。
（由频率的观察引申而来，事情发生的可能性是客观存在的）

1.3 等可能概型（古典概型）

满足两个性质：（1）试验的样本只包含**有限**个元素；（2）试验的基本事件，即每个可能的结果发生的**可能性相等**。
典型例子：抛硬币
长期实践的发现：**"概率很小的事件在一次试验中几乎是不发生"**（称之为**实际推理原理**）

算法学习(26)

综合(1)

最新评论

1. Re:HOOK API（四）——
止

您好，能发一份源码吗？103
qq.com

--

2. Re:HOOK API （一）——
础+一个鼠标钩子实例

认真学习中， 顺便说一下代码
UTTONDOWN写错了写成双

3. Re:使用R进行相关性分析

你好，请问如果数据集中既有
类别变量（factor或characte
做相关性矩阵？直接将类别变
值型变量后用cor函数意义岂
变了？比如：area这个变量，
3......

4. Re:使用R进行相关性分析

@我是酱油妹呀我建议你看官
一个tutorial专门讲如何画图，
细。...

5. Re:使用R进行相关性分析

你好，我用PerformanceAna
了相关图，但是字体、大小、
合杂志社发文章的要求，不知
改？可以教教我吗，或有什么
推荐吗？几乎逛遍各种贴吧论
相应......

--我

推荐排行榜

1.4 条件概率 conditional probability

> 假设A和B是试验E的事件，考虑A已经发生的情况下B发生的概率：**P(B|A) = P(AB) / P(A)**；满足概率的三个基本性质。
> **乘法公式**：P(AB) = P(A)P(B|A)
> **事件S的划分**：B_1, ... ,B_i, ... B_n
> **全概率公式**：P(A) = P(A|B_1) + ... + P(A|B_i) + ...+ P(A|B_n)
> **贝叶斯公式**：P(B_i|A) = P(B_iA)/P(A) = P(A|B_i)P(B_i) / (P(A|B_1) + ... + P(A|B_i) + ...+ P(A|B_n))
>
> Bayes' theorem
> $$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$
>
> **注意**：P(A) > 0, P(B_i) > 0

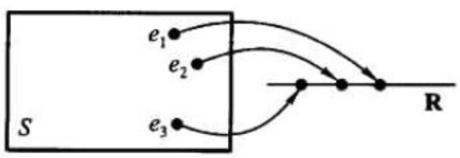
1.5 独立性 Independent

独立性是概率论和数理统计中很重要的概念，很多情况需要满足独立性才适用，一般根据实践来确定事件之间是否相互独立。

> **定义**：若 P(AB) = P(A)P(B)， 则称AB事件相互独立，即A和B两个事件的发生互不影响。
> **定理1**：若P(A) > 0， 且 P(B|A) = P(B) 等价于 AB相互独立
> **定理2**：若AB相互独立，则其对立事件也相互独立
> 可以很自然的推广到n个事件的情况

2. 随机变量及其分布

2.1 随机变量 random variable

> **定义**：设随机试验的样本空间为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**。
> 这样一来，样本空间可以很好的映射到一系列的实值上，方便了接下来各种性质的讨论。
>
> 
>
>
> 随机变量可以分为：**离散型随机变量**和**非离散型随机变量**，其中非离散型随机变量主要以**连续型随机变量为主**。
> **离散型随机变量**：随机变量可能取到的值时有限个数或可列无限多个
> **连续型随机变量**：随机变量可能取到的值时无限个数

2.2 随机变量的分布函数 distribution function.

> **定义**：设X是随机变量，x是任意实数，则分布函数为：
> $$F(x) = P(X \leqslant x), x \in (-\infty, \infty)$$
>
> 则对于任意实数x_1, x_2，有 P{x_1 < X <= x_2} = F(x_2) - F(x_1)
> **性质**：（1）F(x)是不减函数；（2）0<=F(x)<=1, 且 F(-inf) = 0, F(inf) = 1；（3）F(x+0) = F(x)，即F(x) 右连续

2.3 离散型随机变量及其分布律 $\boxed{\text{Discrete random variable}}$

> **分布律**：对于离散型随机变量X，可以取到的值有 x_1,...,x_i, ..., x_n， 对应的概率为 P(x_1),...,P(x_i), ..., P(x_n)。

常用离散型随机分布

（1）0-1分布 0-1 分布：two point distribution

> 事件只有发生和不发生两种可能，发生的概率为p，则不发生的概率为（1-p），
> 那么P{X=k} = $p^k(1-p)^{1-k}$, k = 0,1

（2）伯努利试验、二项分布 Bernoulli distribution

> **伯努利试验**：一次试验只有两种可能结果，发生A，或不发生A'，并且P(A) = p, P(A') = 1-p
> **n次独立重复的伯努利试验服从二项分布**：设X表示事件A发生的次数，则P{X=k} = C(n,k)$p^k(1-p)^{n-k}$, k = 0,1,...,n ，记为X~(n,p)，即X服从参数为n，p的二项分布。
> 注意：**重复**是指每次试验p不变；**独立**是指各次结果互不影响。

（3）泊松分布 Poisson Distribution 柏松分布

定义：记为X~PI(lambda)

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\ldots.$$

泊松定理：当n很大时，泊松分布近似等于二项分布，并且 lambda = np，即 $C(n,k)p^k(1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}(\lambda=np)$，定理表明当n很大时，p很小，上式常用来作二项分布概率的近似计算。
实际中很多事件服从泊松分布：一本书一页中的印刷错误数，某地区在一天内邮递遗失的信件数、某一医院在一天内的急诊病人数、某一地区一个时间间隔内发生交通事故的次数，在一个时间间隔内某种放射性物质发出的、经过计算机的粒子数等。
（可以发现这些例子中，都是**小概率事件**，从实际中与泊松定理联系起来。）

2.4 连续型随机变量及其概率密度 probability density function

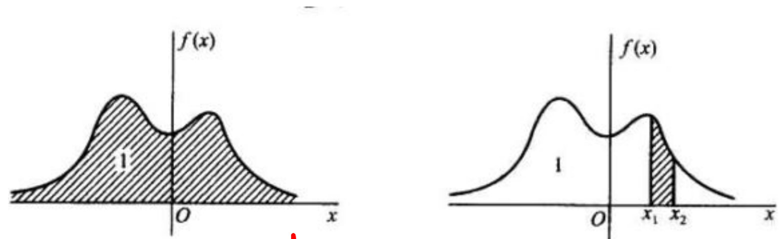
对于连续型随机变量X，$f(x)$称为X的**概率密度函数**，简称**概率密度**。分布函数定义如下：

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

概率密度函数的积分，即围成的面积，为随机变量落入某一区间的概率,如图所示：

$$P(x_1 \leqslant X \leqslant x_2)=\int_{x_1}^{x_2} f(t)dt$$

连续型：continuous random variable.

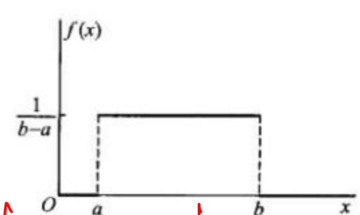


（1）均匀分布 Uniform Distribution

随机变量落入区间（a，b）中任意等长度的子区间内的可能性是相同的。或者说它落入（a，b）区间内的概率只依赖于子区间内的长度而与子区间的位置无关。 X~U（a,b）

```
if a<x<b :
    f(x) = 1/(b-a)
else :
    f(x) = 0
```

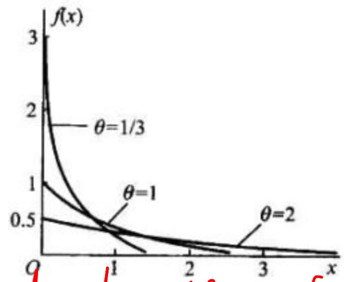
均匀分布图：



（2）指数分布 Exponential Distribution

```
if x>0 :
    f(x) = (1/theta)*exp(-x/theta)
else
    f(x) = 0
```

指数分布图：



（3）正态分布 （高斯(Gauss)分布） Standard Normal Distribution

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

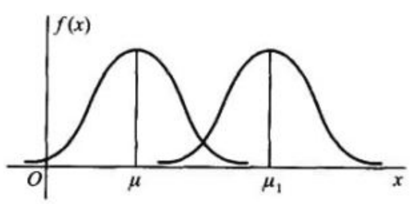
其中μ, σ为常数，分别表示均值和标准差，

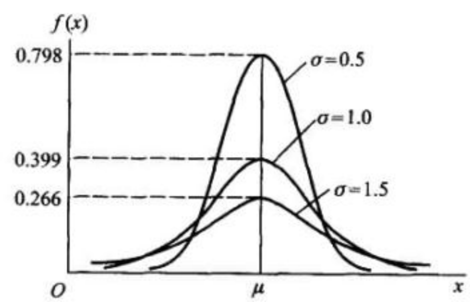
$$X \sim N(\mu, \sigma^2)$$

均值为0，方差为1的正态分布称为标准正态分布，记为X~N(0,1)。任何一个正态分布都可以通过标准化操作转换成标准正态分布，标准化操作等式如下：

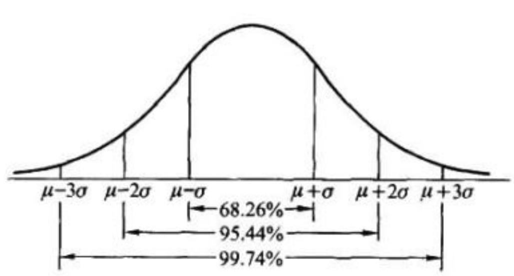
$$X^* = \frac{X - \mu}{\sigma}$$

高斯分布的不同参数的影响：



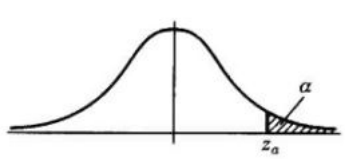


高斯分布的“3σ”法则： $\mu - 3\sigma$ 范围的覆盖率已经达到99.%以上。



高斯分布的上α 分位点

$$P(X > z_\alpha) = \alpha, 0 < \alpha < 1$$



2.5 随机变量的函数的分布

随机变量X的函数Y=g(X)也是一个随机变量，可以根据X的分布率或概率密度求出Y的分布率或概率密度。

3. 多维随机变量及其分布

多维随机变量是在一维上的扩展，以二维随机变量为例

（1）离散型随机变量

分布函数：

$$F(x, y) = P(X \leqslant x, Y \leqslant y), -\infty < x < \infty, -\infty < y < \infty$$

分布率：

$$P(X = x_i, Y = y_i) = p_{ij}, i = 1, 2, \ldots (p_{ij} \geqslant 0, \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1)$$

（2）连续型分布函数（概率密度对应三维空间积分）

分布函数：

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) dxdy, f(x, y) > 0$$

概率：

$$P((X, Y) \in G) = \iint_G f(x, y) dxdy$$

（3）其他概念

边缘分

$$F_X(x) = P(X \leqslant x) = P(X \leqslant x, Y < \infty) = F(x, \infty)$$

$$F_Y(y) = P(Y \leqslant y) = P(X < \infty, Y \leqslant y) = F(\infty, y)$$

条件分布
离散型随机变量：

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}, j = 1, 2, \ldots$$

连续型随机变量：

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

相互独立

$$F(x, y) = F_X(x) F_Y(y)$$

$$f(x, y) = f_X(x) f_Y(y)$$

$$P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j)$$

两个随机变量的分布函数
Z = X+Y
Z = Y/X
Z = XY

M = max{X, Y}
N = min{X, Y}

4. 随机变量的数字特征

4.1 数学期望 Expection

数学期望简称**期望**，又称为**均值**。 数学期望完全由随机变量的分布所确定，若X服从某一分布，也称E(X)是这一分布的数学期望。
离散型随机变量:

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

连续型随机变量:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

数学期望几个重要性质

1. 设C是常数，则有E(C) = C.
2. 设X是一个随机变量，C是常数，则有：E(CX) = CE(X).
3. 设X，Y是两个随机变量，则有：E(X+Y) = E(X)+E(Y)
4. 设X, Y是相互独立的随机变量，则有：E(XY) = E(X)E(Y)

4.2 方差 Variance

定义
方差表达了随机变量X的取值与其数学期望的偏离程度。

$$D(X) = Var(X) = E([X - E(X)]^2)$$

$$\sigma(X) = \sqrt{D(X)}$$

由定义可知，方差实际上就是随机变量X的函数

$$g(X) = (X - E(X))^2$$

的数学期望，因此
离散型随机变量的方差：

$$D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k$$

连续型随机变量的方差：

$$D(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx$$

随机变量X的方差可按下列公式计算（常用）：

$$D(X) = E(X^2) - [E(X)]^2$$

标准化变量:
期望为0，方差为1

$$X^* = \frac{X - \mu}{\sigma}$$

方差的几个重要性质:

1. 设C为常数，则D(C) = 0
2. 设X是随机变量，C是常数，则有

$$D(CX) = C^2 D(X), D(X + C) = D(X)$$

3. 设X，Y是两个随机变量，则有

$$D(X+Y) = D(X) + D(Y) + 2E((X-E(X))(Y-E(Y)))$$

,若X,Y相互独立，则

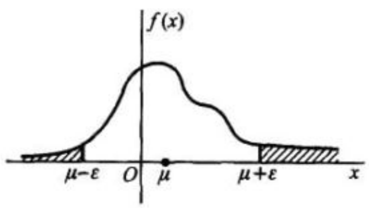
$$D(X+Y) = D(X) + D(Y)$$

4. D(X)=0的充要条件是X以概率1取常数E(X)，即 P{X=E(X)} = 1

切比雪夫（Chebyshev）不等式：
切比雪夫不等式给出了在随机变量的分布未知，而只知道E(X)和D(X)的情况下估计概率 P{|X-E(X)|<epsilon}的界限，epsilon是任意正数。

$$P(|X-\mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

$$P(|X-\mu| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{\varepsilon^2}$$



4.3 协方差及相关系数

协方差： Covarivance 讨论两个方差变化趋势

$$Cov(X,Y) = E(X-E(X))E(Y-E(Y))$$

相关系数：

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

由定义可知： Cov(X,Y) = Cov(Y,X), Cov(X,X) = D(X)

协方差性质：

1.

$$Cov(aX,bY) = abCov(X,Y)$$

ab是常数

2.

$$Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y)$$

相关系数性质：

1.

$$|\rho_{XY}| \leqslant 1$$

2.

$$|\rho_{XY}| = 1$$

的充要条件是，存在常数a,b，使P(Y=a+bX) = 1

3. 当

$$|\rho_{XY}| = 0$$

X和Y 不相关

> _注意：_相关系数也称为线性相关系数，它是一个可以用来描述随机变量（X，Y）的两个分量X，Y之间的线性关系紧密程度的数字特征。当相关系数较小时，X，Y的线性相关程度较差；当相关系数=0时称X，Y不相关。**值得注意的是，不相关是指X，Y之间不存在线性关系，它们还可能存在除线性关系之外的关系。**
> **X，Y相互独立是对X，Y的一般关系而言。X，Y相互独立则X，Y一定不相关；反之，若X，Y不相关则X,Y不一定相互独立。**
> **特别的，**对于二维正态随机变量（X，Y），X和Y不相关与X和Y相互独立是等价的。

4.4 矩、协方差矩阵

k阶原点矩（k阶矩）

$$E(X^k), k = 1, 2, \ldots$$

k阶中心矩

$$E([X - E(X)]^k), k = 2, 3, \ldots$$

k+l阶混合矩

$$E(X^k Y^l), k, l = 1, 2, \ldots.$$

k+l阶混合中心矩

$$E([X - E(X)^k [Y - E(Y)]^l), k, l = 1, 2, \ldots$$

协方差矩阵

$$c_{ij} = Cov(X_i, X_j) = E([X_i - E(X_i)][X_j - E(X_j)]), i, j = 1, 2, \ldots, n$$

一般n维随机变量的分布是不知道的，或者是太复杂，以至于在数学上不易处理，因此在实际引用中协方差矩阵显得尤为重要。

5. 大数定律及中心极限定理

大数定律：随机变量序列的前一些项的算数平均在某种条件下收敛到这些项的均值的算术平均值；
中心极限定理：在相当一般的条件下，当独立随机变量的个数不断增加时，其和的分布趋于正态分布。

5.1 大数定律

Law of Large Numbers.

弱大数定理（辛钦大数定理）：对于相互独立且同分布的序列而言

$$\lim_{n \to \infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| < \varepsilon) = 1$$

伯努利大数定理：

$$\lim_{n \to \infty} P(|\frac{f_A}{n} - p| < \varepsilon) = 1$$

当试验次数很大时，便可以用事件的频率来代替事件的概率。

5.2 中心极限定理

Central limit theorem.

定理一（独立同分布的中心极限定理）：

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

$$Y_n \sim N(0, 1)$$

定理二（李雅普诺夫（Lyapunov）定理），前提：各随机变量相互独立。
无论各个随机变量X_k (k=1,2,..)服从什么分布，只要满足定理的条件，那么它们的和，当n很大时就近似服从正态分布。

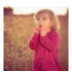
定理三（棣莫弗一拉普拉斯（ De Moivre－ Laplace）定理）正态分布是二项分布的极限分布。当n充分大时可以使用正态分布作为二项分布的近似。二项分布的标准化变量服从标准正太分布。

https://github.com/hfl15

分类： 数学基础

标签： 机器学习**,** 数据挖掘**,** 概率论**,** 数学

好文要顶 关注我 收藏该文



.....?
关注 - 3
粉丝 - 51
+加关注

0 0

« 上一篇：git入门基础
» 下一篇：数学基础（2）~ 数理统计基础知识

posted @ 2017-04-13 15:36 .....? 阅读(5680) 评论(3) 编辑 收藏

评论列表

#1楼 2017-04-13 19:11 桂。

我也打算这一两周内开始着手入门机器学习呢，祝你不断进步~

支持(0) 反对(0)

#2楼[楼主] 2017-04-13 19:37 .....?

@ 桂。
好的，谢谢，一起加油(*^_^*)

支持(0) 反对(0)

#3楼 2017-04-13 23:40 牛腩

支持支持

支持(0) 反对(0)

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 登录 或 注册， 访问 网站首页。

相关博文：
· 概率论基础
· 概率论——数学——大学课程学习
· 概率论01
· 数学基础-概率论
· 2011考研数学概率论基础复习必备知识点