北京师范大学
BEIJING NORMAL UNIVERSITY

2.3

随机变量及特征刻画

案例

在现实生活中，常关心一个事件 A 是否发生．如果用 1 代表事件 A 发生，0 代表事件 A 不发生，则无论事件 A 是什么，都可以把所关心的问题的样本空间和事件类统一写成

$$\Omega=\{0,1\},\quad \mathscr{F}=\{\varnothing,\{0\},\{1\},\Omega\}.$$

可用伯努利概率空间描述此类随机现象的概率分布问题.

启示

可通过映射把各类样本空间转化为实数空间；

可通过实数域上的概率空间来研究随机现象.

随机现象可纳入概率空间的框架研究， 各随机现象的样本空间千变万化，加大了随机现象的研究难度.

问题：是否可以统一

本节学习把抽象样本空间中的概率问题转化为实数空间中的概率问题的方法和途径.

随机变量及其分布函数

定义

设$(\Omega, \mathcal{F}, \mathbb{P})$为概率空间，$\xi$为定义在$\Omega$上的实函数．若对于任何实数 x 有

$$\{\xi \leqslant x\} \triangleq \{\omega \in \Omega: \ \xi(\omega) \leqslant x\} \in \mathscr{F},$$

就称 ξ 为**随机变量(random variable).**

$\{\xi \leq x\}$的含义：

- 是由满足条件$\xi(\boldsymbol{\omega}) \leq x$的样本点构成的**事件**；
- 是使得随机变量ξ落入区间$(-\infty, \boldsymbol{x}]$的样本点所构成的**集合**.

伯努利概率空间上定义随机变量

在投掷一枚硬币的实验中，样本空间$\Omega = \{$正，反$\}$，事件类$\mathscr{F} = \{\emptyset, \{正\}, \{反\}, \Omega\}$.

概率 $\mathbb{P}(\{正\}) = \mathbb{P}(\{反\}) = \frac{1}{2}$， $\mathbb{P}(\emptyset)=0$， $\mathbb{P}(\Omega) = 1$.

定义

$$\xi(\omega) = \begin{cases} 1, & \omega = 正, \\ 0, & \omega = 反, \end{cases}$$

试证明ξ为随机变量.

证明

对于任意 $x \in R$，有

$$\{\xi \le x\} = \begin{cases} \emptyset, & \text{如果 } x < 0 \\ \{\text{反}\}, & \text{如果 } 0 \le x < 1 \\ \Omega, & \text{如果 } x \ge 1 \end{cases}$$

所以$\{\xi \le x\} \in \mathcal{F}$，即$\xi$为随机变量.

- 是否为随机变量与概率的定义无关;
- 可用不同的随机变量来描述掷硬币实验.

可以证明

$$\tilde{\xi}(\omega) = \begin{cases} 3, & \omega = \text{正} \\ -2, & \omega = \text{反} \end{cases}$$

也是随机变量

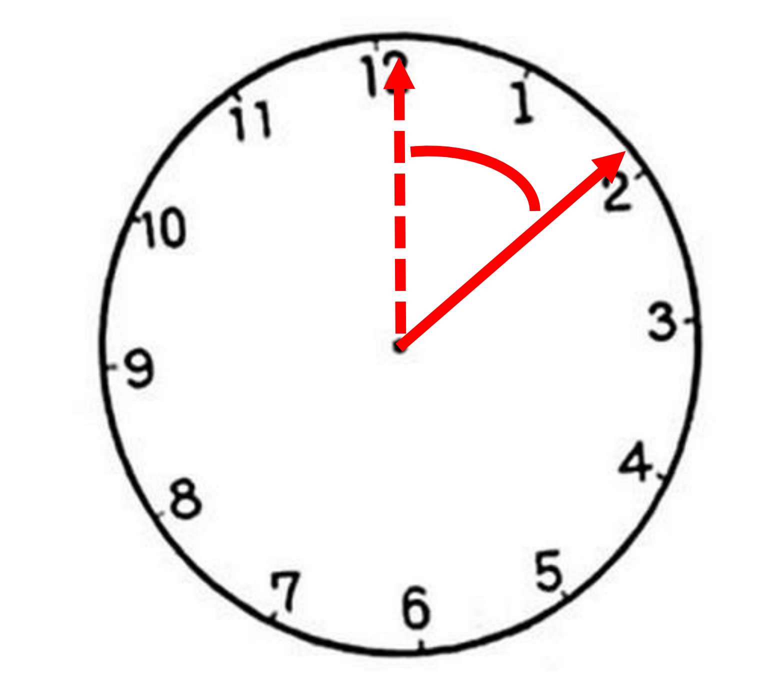
注：常数也可以看成随机变量.

$$\eta(\omega) \equiv c$$

但是它仅能表示必然事件和不可能事件.

例2.3.1

某人一觉醒来，打开收音机对时，请建立描述该人醒来时刻的概率空间，并在该空间上定义一个随机变量，以表示事件“他至少等50分钟”.



北京师范大学 BEIJING NORMAL UNIVERSITY

解： 取样本空间和事件类为

$$\Omega=(0,60],\ \mathscr{F}=\{A:\ A\subset\Omega, A\text{可求长度}\},$$

$\mathbb{P}$为几何概率．则可以用概率空间$(\Omega,\ \mathcal{F},\ \mathbb{P})$描述该人醒来的时刻．定义

$$\xi(\boldsymbol{\omega})=\boldsymbol{\omega},\forall\boldsymbol{\omega}\in\boldsymbol{\Omega}.$$

则ξ为随机变量.

可以将事件“他至少等50分钟”用ξ表示为：$\{\xi\leq 10\}$.

当然，我们也可定义

$$\eta=\begin{cases}1, & \text{如果他至少等50分钟}\\ 0, & \text{否则}\end{cases}$$

这时$\{\eta=1\}$就是事件“他至少等50分钟”．显然不能用随机变量η来描述事件“他等30分钟”（但可以用ξ表示该事件）．

定义

设ξ是定义在$(\Omega, \mathcal{F}, \mathbb{P})$上的随机变量. 对于任何$x \in R$，定义

$$F_\xi(x) \triangleq \mathbb{P}(\xi \leqslant x),$$

称F_ξ为随机变量ξ的分布函数，简称为分布函数.

在不至于引起混乱的情况下，常省略分布函数中的下标，即将$F_\xi(x)$简写为$F(x)$.

随机变量落入各个区间的概率完全由其分布函数所决定，即分布函数刻画了随机变量的随机变化规律.

如当$a<b$时，注意到

$$\{\xi \le a\} \subset \{\xi \le b\}, \quad \{a < \xi \le b\} = \{\xi \le b\} - \{\xi \le a\},$$

根据概率的可减性有

$$\mathbb{P}(a < \xi \leqslant b) = \mathbb{P}(\{\xi \leqslant b\} - \{\xi \leqslant a\}) = \mathbb{P}(\xi \leqslant b) - \mathbb{P}(\xi \leqslant a) = F(b) - F(a).$$

随机变量的分布函数就是我们所熟悉的一种函数，通过它可以刻画相关事件的概率，即它成为我们研究事件概率的工具.

例2.3.2

假设硬币的质地均匀，求随机变量

$$\xi(\boldsymbol{\omega}) = \begin{cases} 1, & \boldsymbol{\omega} = 正, \\ 0, & \boldsymbol{\omega} = 反, \end{cases}$$

的分布函数.

解： 由硬币质地均匀知$\mathbb{P}(\{正\}) = \mathbb{P}(\{反\}) = \frac{1}{2}$，

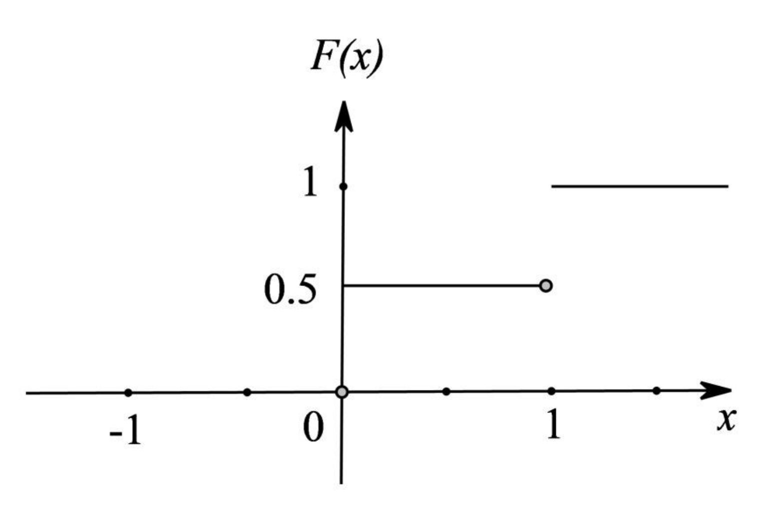
注意到当$x<0$时，$\{\xi \le \boldsymbol{x}\} = \varnothing$,

当$0\le x <1$时，$\{\xi \le \boldsymbol{x}\} = \{\xi = \boldsymbol{0}\} = \{反\}$,

当$1\le x$时，$\{\xi \le \boldsymbol{x}\} = \{正，反\} = \Omega$,

由此可得 ξ 的分布函数为

$$F(x) = \mathbb{P}(\xi \leqslant x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1. \end{cases}$$

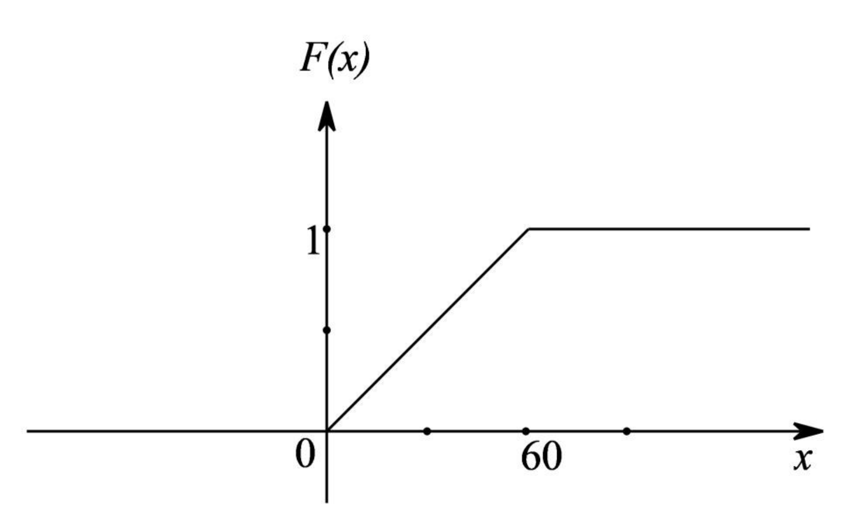


例2.3.3

在收音机对时例子中，用ξ表示这个人醒时的分钟数，求ξ的分布函数，并利用分布函数计算此人醒来时刻的分钟数落在区间(10, 20]上的概率.

解：这个人醒来的时刻是任意的，因此$\mathbb{P}$为几何概率．根据几何概率的计算公式，有

$$F(x)=\begin{cases}0, & x<0,\\ \frac{x}{60}, & 0\leqslant x<60,\\ 1, & x\geqslant 60.\end{cases}$$



所需求的概率为

$$\mathbb{P}(\xi\in(10,20])=F(20)-F(10)=\frac{1}{6}.$$

小结

随机变量

分布函数

离散型随机变量及其数学期望

离散型随机变量

定义

如果存在 $x_k \in R$，$1 \le k \le n$，使得 $\sum_{k=1}^{n} \mathbb{P}(\xi = x_k) = 1$,

就称ξ为离散型随机变量，称F_ξ为离散型分布函数．记

$$p_k = \mathbb{P}(\xi = x_k), \quad 1 \leqslant k \leqslant n,$$

称

$$\begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}$$

为ξ或F_ξ的密度矩阵，简称为密度.

例2.3.4

投掷一枚硬币，求随机变量ξ的密度矩阵.

$$\xi(\omega)=\begin{cases}1, & \omega=\text{正},\\ 0, & \omega=\text{反},\end{cases}$$

解： 显然，ξ的值域中只有两个数0和1，所以它是离散型随机变量，其密度矩阵为

$$\begin{pmatrix} 0 & 1 \\ \mathbb{P}(\xi=0) & \mathbb{P}(\xi=1) \end{pmatrix}.$$

一般地，如果ξ的密度矩阵为

$$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},$$

就称它所对应的分布为两点分布，称ξ服从两点分布.

如果ξ的密度矩阵为

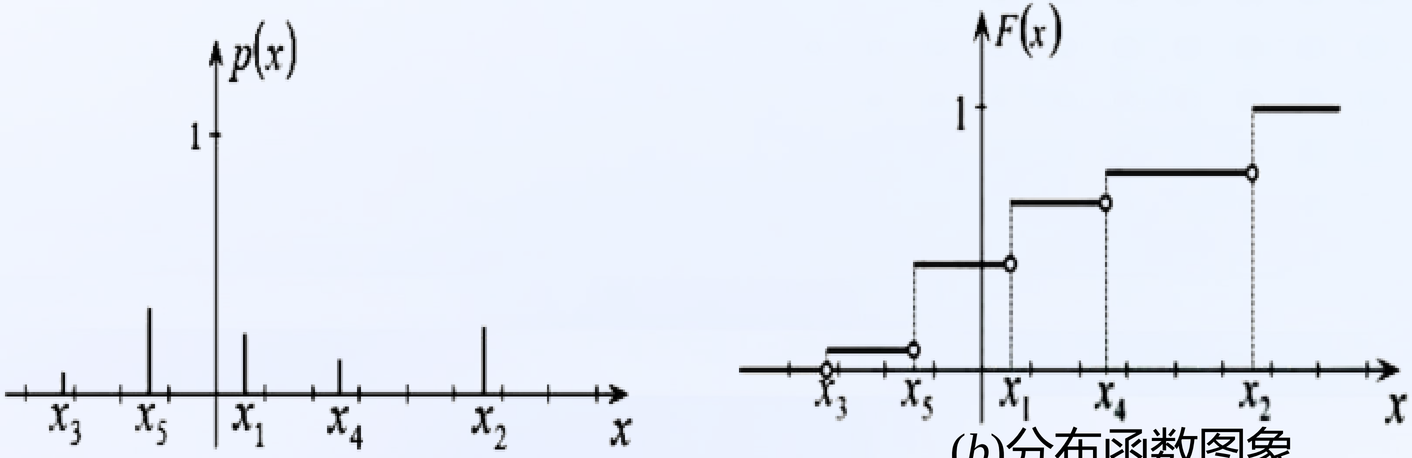
$$\begin{pmatrix} c \\ 1 \end{pmatrix}$$

就称它所对应的分布为单点分布或退化分布，称ξ服从单点分布.

对于离散型随机变量ξ，不难验证其分布函数为

$$F(x)=\sum_{k\in\{i:\,x_i\leqslant x\}}p_k,\quad \forall x\in\mathbb{R}.$$

离散型分布函数由密度唯一确定.



(a)密度图像 (b)分布函数图象

图2.7: 离散型随机变量的密度函数与分布函数

案例2.8

若ξ的密度矩阵为

$$\begin{pmatrix} 70 & 80 & 90 \\ 0.5 & 0.3 & 0.2 \end{pmatrix},$$

思考：其观测值的算术平均值随观测次数的增加有什么规律?

用k_{m1}，k_{m2}和k_{m3}分别表示观测中等于70，80和90的个数，则

$$\frac{1}{m}\sum_{k=1}^{m} x_k = 70\cdot\frac{k_{m1}}{m} + 80\cdot\frac{k_{m2}}{m} + 90\cdot\frac{k_{m3}}{m}$$

70*0.5+80*0.3+90*0.2=77

即观测值的算术平均会随着观测次数的增加稳定于

$$\mathbb{E}(\xi) \triangleq 70\times\mathbb{P}(\xi=70) + 80\times\mathbb{P}(\xi=80) + 90\times\mathbb{P}(\xi=90).$$

数学期望

定义

设ξ为离散型随机变量，其密度矩阵为

$$\begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}.$$

如果 $\sum\limits_{i=1}^{n} |x_i| p_i < \infty$，就称

$$\mathbb{E}(\xi) \triangleq \sum_{i=1}^{n} x_i p_i$$

为ξ（或其分布）的数学期望，又称为均值；否则称ξ的数学期望不存在.

- 随机变量的数学期望是一个实数，它是刻画随机变量的一种指标，在实际中有广泛的应用.
- $E(\xi)$是ξ的以概率为权重的加权平均，是ξ中心位置的一种度量.
- 如果两个随机变量的分布相同，则它们的数学期望相同，即数学期望完全由分布函数所决定，也由密度所决定.
- 数学期望的应用：比较两个随机变量的平均水平.（如两个班数学水平，两种投资方案的平均收益等）
- $E(\xi)$的估计方法：用ξ的重复观测值的算术平均值.

例2.3.5

求两点分布随机变量ξ的数学期望.

解： 由两点分布的密度矩阵和数学期望的定义知

$$\mathbb{E}(\xi) = 0 \times q + 1 \times p = p,$$

即两点分布的数学期望等于成功概率.

例2.3.6

投掷一枚质地均匀的硬币两次，对于：i=1, 2，定义

$$\xi_i(\omega)=\begin{cases}1, & \text{如果第}i\text{次出现正面},\\ 0, & \text{否则}.\end{cases}$$

求E(ξ_1)和E(ξ_2) .

解：ξ_1和ξ_2都服从两点分布，有共同的密度矩阵

$$\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix},$$

可知

$$\mathbb{E}(\xi_1)=\mathbb{E}(\xi_2)=\frac{1}{2}.$$

思考：上面的两个随机变量ξ_1和ξ_2是否相等？$\xi_1+\xi_2$的含义？

定理

（数学期望的线性性质）设随机变量ξ和η的数学期望都存在，$a, b \in R$，则

$$\mathbb{E}(a\xi + b\eta) = a\mathbb{E}(\xi) + b\mathbb{E}(\eta).$$

例2.3.7

投掷一枚质地均匀的硬币n次，求出现正面次数ξ的数学期望.

在第 i 次投掷硬币实验中，定义

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次掷出正面}\\ 0, & \text{第 } i \text{ 次掷出反面}\end{cases}$$

则$\xi=\sum_{i=1}^{n}\xi_i$,, 且 ξ_i 服从两点分布，$\mathbb{E}(\xi_i)$=1/2

再由数学期望的线性性质知

$$\mathbb{E}(\xi)=\sum_{i=1}^{n}\mathbb{E}(\xi_i)=\sum_{i=1}^{n}\frac{1}{2}=\frac{n}{2}.$$

小结

离散型随机变量

数学期望及其线性性质

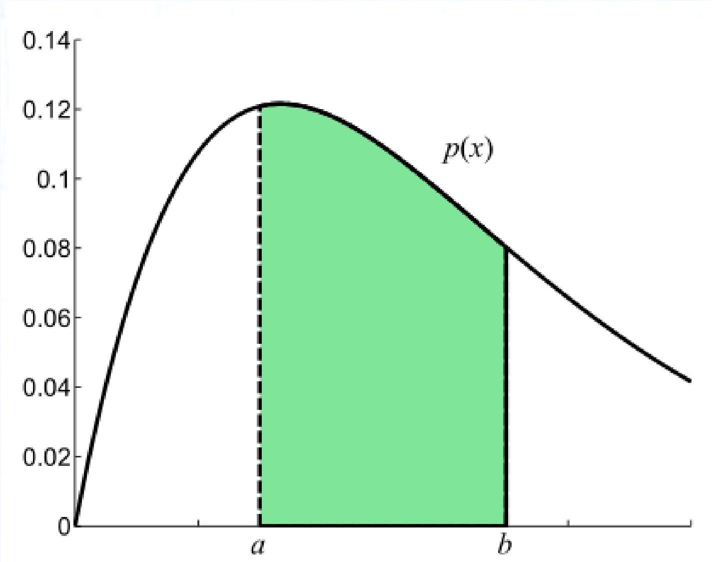
北京师范大学
BEIJING NORMAL UNIVERSITY

第四次作业

105页练习2.19、2.20、2.21、2.22

连续型随机变量及其数学期望

曲边梯形：



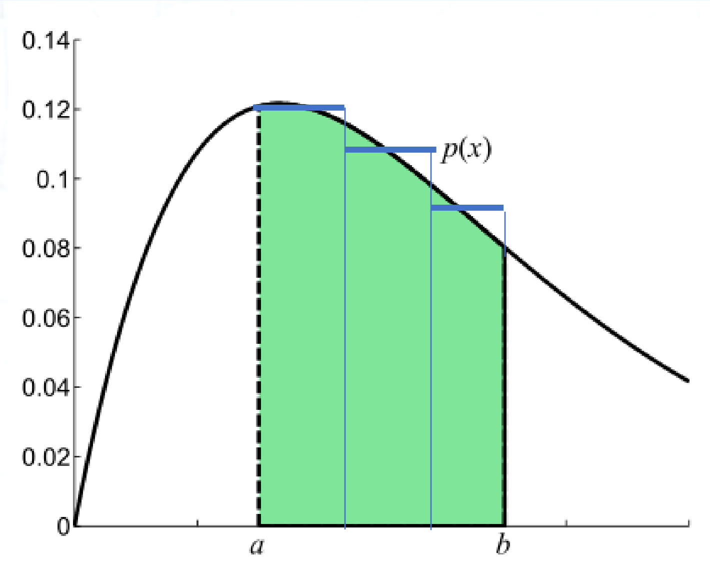


图2.8 定积分与曲边梯形的面积

称其面积为函数$p(x)$在区间$(a,b]$上的**定积分**，称$p(x)$为**被积函数**，如图2.8所示，为讨论方便，我们用

$$\int_a^b p(x)dx$$

表示函数$p(x)$在区间$[a,b]$上的定积分.

R软件可以用函数*integrate*计算定积分，例如计算

$$\int_0^1 x^2 \boldsymbol{dx}$$

的R程序代码如下:

```
integrate(function(x) x^2,0,1)
```

运行该代码后，在控制台窗口中显示的输出结果为:

```
0.3333333  with  absolute  error  <  3.7e-15
```

可利用R语言的在线帮助了解该函数的使用方法:

```
?integrate
```

连续型随机变量

定义

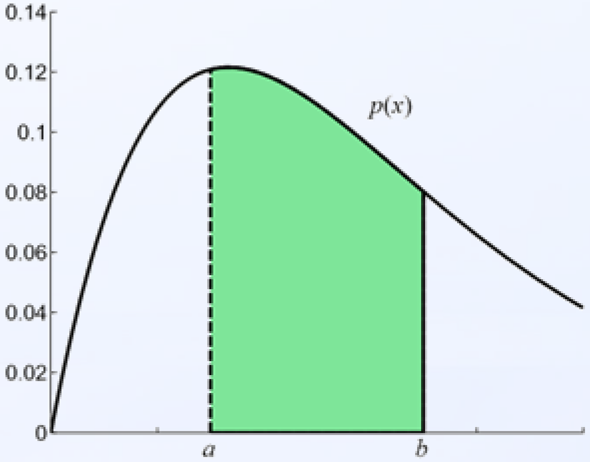
如果存在非负函数 $p(x)$，使ξ的分布函数

$$F(x)=\int_{-\infty}^{x} p(t)\mathrm{d}t, \quad \forall x \in \mathbb{R}$$

就称ξ为连续型随机变量；称F为连续型分布；称$p(x)$为ξ或 F 的密度函数，简称为密度.

基本性质与解释

概率与密度的关系： 随机变量ξ落入区间$(a, b]$内的概率等于以密度函数为顶、以横轴$(a, b]$为底的曲边梯形的面积.



$$\mathbb{P}(a < \xi \leqslant b) = \int_a^b p(t)\mathrm{d}t$$

等于曲边梯形的面积

密度函数的概率含义：

$$\mathbb{P}\left(\xi\in\left[x-\frac{\Delta}{2},x+\frac{\Delta}{2}\right]\right)=\int_{x-\frac{\Delta}{2}}^{x+\frac{\Delta}{2}}p(t)\mathrm{d}t\approx p(x)\Delta.$$

因此，密度函数$p(x)$刻画了ξ落在x邻域内的概率大小.

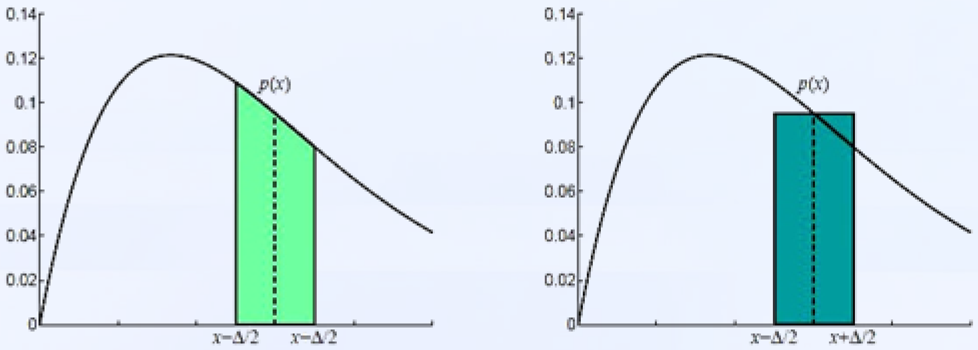


图　密度函数$p(x)$的内涵

例2.3.8

证明在收音机对时例子中，这个人醒时的分钟数ξ是连续型随机变量.

定义

$$p(x)=\begin{cases}1/60, & 0<x<60,\\ 0, & \text{其他}.\end{cases}$$

称为(0,60)的均匀分布，记为 $\xi \sim U(0,60)$.

由例2.3.3中ξ的分布函数$\boldsymbol{F}(\boldsymbol{x})$的表达式(2.21)知：当$\boldsymbol{x}<\mathbf{0}$时有

$$\boldsymbol{F}(\boldsymbol{x})=\mathbf{0}=\int_{-\infty}^{\boldsymbol{x}}\boldsymbol{p}(\boldsymbol{t})\boldsymbol{d}\boldsymbol{t}\ ;$$

当$0 \le x < 60$时,

$$F(x) = \frac{x}{60} = \int_0^x \frac{1}{60} dt = \int_{-\infty}^x p(t) dt;$$

当60≤ x时,

$$F(x) = 1 = \int_0^{60} \frac{1}{60} dt = \int_{-\infty}^x p(t) dt.$$

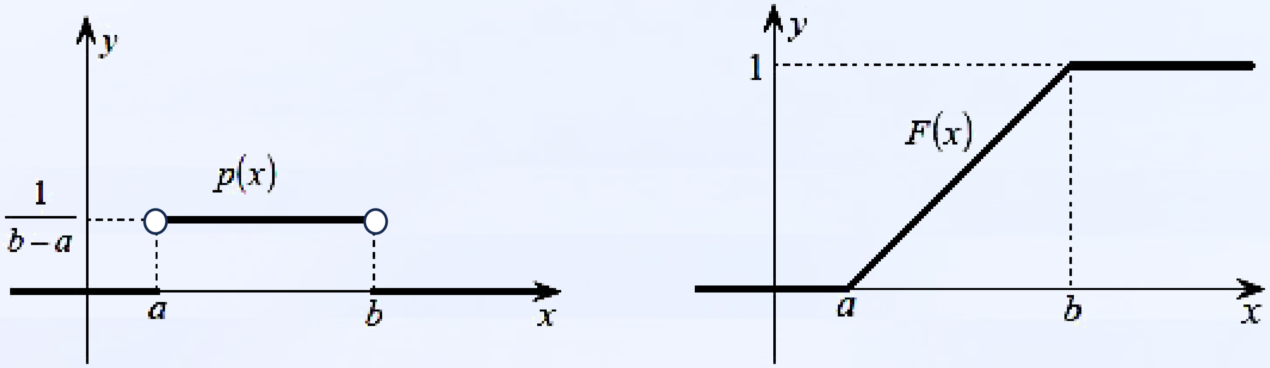
因此ξ为连续型随机变量.

定义2.3.6 （均匀分布）

给定实数$a<b$,称以

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

为密度函数的分布为(a,b)区间上的均匀分布，简记为$U(a,b)$. 若ξ的分布为$U(a,b)$，就称它服从$U(a,b)$，简记为$\xi\sim U(a,b)$.



$U(a,b)$的密度函数和分布函数的图像

定义

设ξ为连续型随机变量，密度函数为$p(x)$. 如果

$$\int_{-\infty}^{\infty} |x|p(x)\mathrm{d}x < \infty,$$

称

$$\mathbb{E}(\xi) \triangleq \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$$

为ξ (或其分布)的数学期望，或均值；否则称ξ的数学期望不存在.

北京师范大学 BEIJING NORMAL UNIVERSITY

- 随机变量的数学期望是一个实数，它是刻画随机变量的一种指标，在实际中有广泛的应用.
- $E(\xi)$是ξ的以概率为权重的加权平均，是ξ中心位置的一种度量.
- 如果两个随机变量的分布相同，则它们的数学期望相同，即数学期望完全由分布函数所决定，也由密度所决定.
- 数学期望的应用：比较两个随机变量的平均水平.（如两个班数学水平，两种投资方案的平均收益等）
- $E(\xi)$的估计方法：用ξ的重复观测值的算术平均值.

例

设$\xi \sim U(0,60)$，写出计算$E(\xi)$的程序代码，并给出计算结果.

在R语言中，用dunif 计算均匀分布的密度函数值，因而可以用如下的程序代码计算$U(0,60)$的数学期望：

```
integrate(function(x)  x*dunif(x, 0, 60), -Inf, Inf)
```

其中 -Inf 和 Inf 分别代表 $-\infty$ 和 ∞ . 上述程序代码的计算结果为29.99986 .

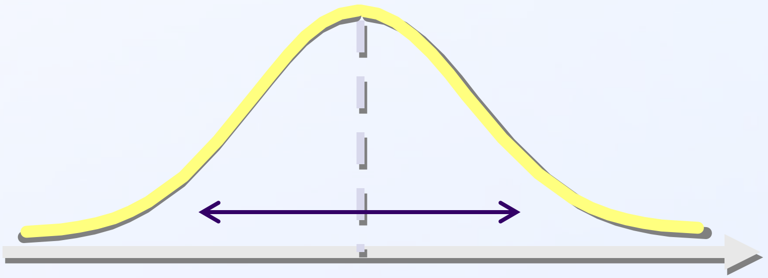
小结

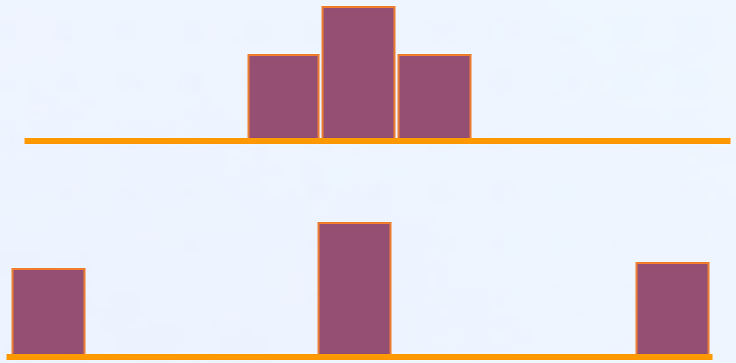
连续型随机变量及其期望

密度函数的概率含义

随机变量的方差

问题： 如何刻画随机变量ξ的集中（离散）程度？





要刻画随机变量的集中程度，首先要有一个代表中心位置的点，然后考察随机变量集中于这个点的程度．由于数学期望是代表随机变量的中心位置，因此考虑

$$\eta = \left(\xi - \mathbb{E}(\xi)\right)^2$$

来刻画随机变量的集中程度．

ξ不能事先预知. 所以η也不能预知, 用它表达集中程度不合适．

定义

设ξ^2的数学期望存在，称

$$D(\xi) \triangleq \mathbb{E}\left(\xi - \mathbb{E}(\xi)\right)^2$$

为ξ(或其分布)的方差，称$\sqrt{D(\xi)}$为其标准差.

方差的解释

方差表示随机变量（平均）集中于数学期望的程度. 它越小ξ越集中于均值. 特别，当$D(\xi)=0$时，$\mathbb{P}\big(\xi=\mathbb{E}(\xi)\big)=\mathbf{1}$.

- 离散型随机变量

若随机变量ξ的密度矩阵为$\begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}$，则方差

$$D(\xi)=\sum_{k=1}^{n}\left(x_k-\mathbb{E}(\xi)\right)^2 p_k.$$

- 连续型随机变量

若ξ的密度函数为$p(x)$，则方差

$$D(\xi)=\int_{-\infty}^{\infty}\left(x-\mathbb{E}(\xi)\right)^2 p(x)\mathrm{d}x.$$

定理

设ξ^2的数学期望存在，则

$$D(\xi)=\mathbb{E}(\xi^2)-(\mathbb{E}(\xi))^2 .$$

显然

$$(\xi-\mathbb{E}(\xi))^2=\xi^2-2\xi\mathbb{E}(\xi)+(\mathbb{E}(\xi))^2 .$$

注意到$E(\xi)$为实数，利用数学期望的线性性质得

$$D(\xi)=\mathbb{E}(\xi^2)-(2\mathbb{E}(\xi))\,\mathbb{E}(\xi)+(\mathbb{E}(\xi))^2=\mathbb{E}(\xi^2)-(\mathbb{E}(\xi))^2 ,$$

即结论成立.

定理

若ξ的方差存在，ξ_i表示随机变量ξ的第i次重复观测，a_i为实数，则

$$D\left(\sum_{i=1}^{n} a_i\xi_i\right)=\sum_{i=1}^{n} a_i^2 D\left(\xi_i\right),$$

例

掷一枚质地均匀的硬币n次，求出现正面次数ξ的方差.

沿用之前的记号，注意到$\xi^2{}_i=\xi_i$ 可得

$$D\left(\xi_i\right)=\mathbb{E}\left(\xi_i^2\right)-\left(\mathbb{E}\left(\xi_i\right)\right)^2=\mathbb{E}\left(\xi_i\right)-\left(\frac{1}{2}\right)^2=\frac{1}{4},$$

再由上述定理得

$$D(\xi)=D\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}D\left(\xi_i\right)=\frac{n}{4}.$$

定义

对于随机变量ξ，称 $Z=\dfrac{\xi-\mathbb{E}(\xi)}{\sqrt{D(\xi)}}$ 为ξ的**标准化**，或ξ的**标准得分**.

随机变量ξ的标准得分的期望为0，方差为1，即标准化消除了ξ数学期望和方差的特征，它是一个无量纲的量.

问题

有甲乙两名学生分别参加了A省和B省的高考，如何判断哪一名学生的成绩更好？你有什么办法吗？

再比如参加同一个省两次英语考试，成绩如何比较？

用标准得分可以解决一部分问题，比如比较在本省学生中的相对位置.

利用等值理论，但目前还没有完全解决.

在实际应用中，人们经常需要计算随机变量的函数$f(\xi)$的数学期望，这里 f 是一个函数．如ξ^2就是随机变量的函数，它是平方函数在ξ处的值．一般 $f(\xi)$ 也是随机变量. 可以借助于ξ的密度计算$f(\xi)$的数学期望，具体公式如下：

当ξ有密度矩阵时，

$$\mathbb{E}(f(\xi)) = \sum_{k=1}^{n} f(x_k) p_k;$$

当ξ的密度函数为$p(x)$时，

$$\mathbb{E}(f(\xi)) = \int_{-\infty}^{\infty} f(x) p(x) \, \mathrm{d}x.$$

小结

随机变量的方差

随机变量的标准化

第五次作业

106页练习2.24、2.25、2.26、2.27、2.28、2.29