第二章　随机变量及其分布

第二章　随机变量及其分布

- 随机变量
- 离散型随机变量的概率分布
- 随机变量的分布函数
- 连续型随机变量的概率密度
- 随机变量的函数的分布

§1 随机变量

一、随机变量概念的产生

在实际问题中，随机试验的结果可以用数量来表示：

1、有些试验结果本身与数值有关（本身就是一个数）

例：中国好声音盲选时，有几位导师为同一个选手转身

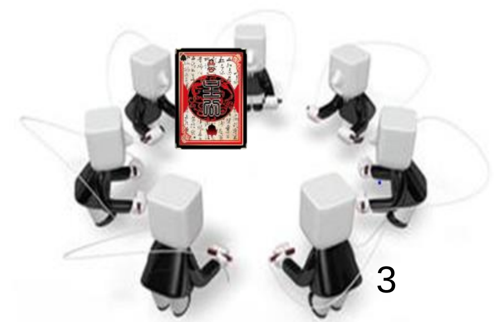
$S=\{0, 1, 2, 3, 4\}$

掷一颗骰子面上出现的点数

$S=\{1, 2, 3, 4, 5, 6\}$

玩三国杀（只有一个主公），玩10局同一人抽中主公角色的次数

$S=\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$



§1 随机变量

一、随机变量概念的产生

2、在有些试验中，试验结果看来与数值无关，但我们可以引进一个变量来表示它的各种结果。也就是说，把试验结果数值化

例1：将一枚硬币抛掷三次，观察出现正面和反面的情况，则样本空间是$S=\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$。

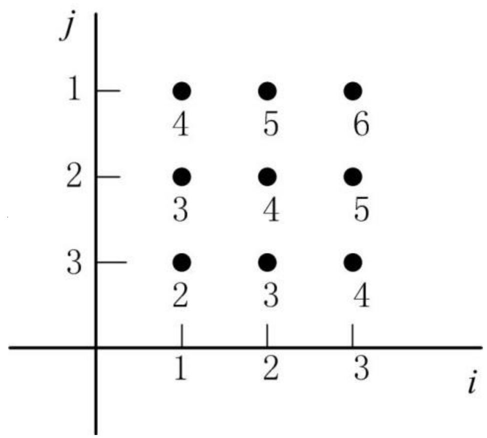
令X表示三次投掷得到正面H的总数，那么X是定义在S上的一个实单值函数。 定义域：样本空间S； 值　域：实数集合$\{0,1,2,3\}$

$$X=X(e)=\begin{cases}3 & e=HHH,\\ 2 & e=HHT,HTH,THH,\\ 1 & e=HTT,THT,TTH\\ 0 & e=TTT\end{cases}$$

一、随机变量概念的产生

例2：在一袋中装有编号分别为1,2,3的3只球，在袋中任取一只球，放回，再任取一只球，记录它们的号码，试验的样本空间为$S=\{e\}=\{(i, j)|i, j=1,2,3\}$，$i$，$j$分别为第1，第2次取到的球的号码。

以X记两球号码之和，对于试验的每一个结果$e=(i, j)\in S$，X都有一个指定的值$i+j$与之对应



X是定义在样本空间S上的单值实值函数，

定义域：样本空间S；值 域：实数集合$\{2,3,4,5,6\}$

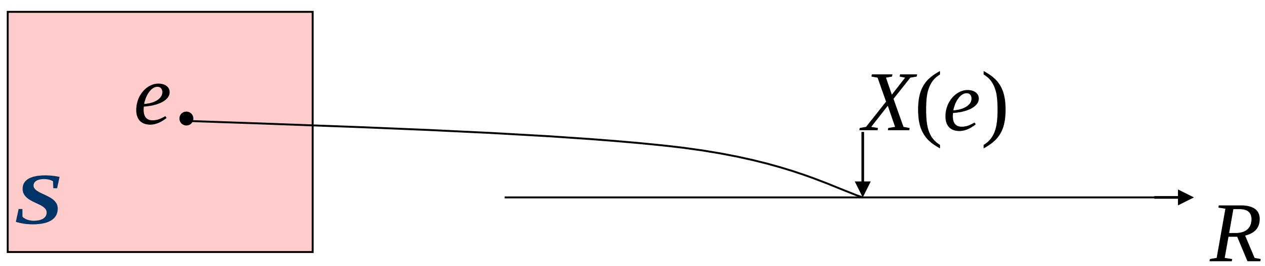
$$X=X(e)=X(i, j)=i+j,\ i, j=1,2,3$$

§1 随机变量

一、随机变量概念的产生

随机变量的定义

设E是一个随机试验，$S=\{e\}$是其样本空间．$X=X(e)$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**。



即对于每一个$e\in S$，有一实数$X=X(e)$与之对应，则称X为随机变量。

§1 随机变量

一、随机变量概念的产生

（1）随机变量通常用大写字母X, Y, Z或希腊字母ξ、η、ζ等表示，而表示随机变量所取的值时，一般采用小写字母x, y, z等。

（2）随机变量是定义在样本空间S上的实值单值函数，**S中的元素不一定是实数，而普通函数只是定义在实数轴上。**

（3）随机变量的取值具有随机性，它随试验结果的不同而取不同的值，试验之前仅知道它可能取值的范围，而不能预知它取什么值。它与普通函数的**区别在于随机变量取某一个值或在某个区间内取值均为随机事件**

（4）**随机变量的取值具有统计规律性。**由于试验结果的出现有一定的概率，因而随机变量取各个值也有一定的概率。

一、随机变量的分类

随机变量
- 离散型随机变量
  - 所有取值可以逐个一一列举
- 非离散型随机变量
  - 奇异型随机变量
  - 连续型随机变量
    - 全部可能取值不仅无穷多，而且还不能一一列举，而是充满一个区间

§2 离散随机变量及其分布律

一、离散型随机变量的定义

离散型随机变量的定义

随机变量所有可能取的值是有限个或可列无限多个，这种随机变量称为**离散型随机变量。**

如X：取到次品的个数，Y：收到的呼叫数等都是离散型随机变量，但Z：电视机的寿命 不是离散型随机变量。

为了掌握随机变量*X*的统计规律性，不仅需要知道随机变量*X*的取值，而且还应知道*X*取每个值的概率

二、离散型随机变量概率分布的定义

设离散型随机变量X的**所有可能取值**为x_k（k=1, 2, …），称**X取各个可能值的概率**，即事件{$X=x_k$}的概率，

$$P\{X=x_k\}=p_k,\quad (k=1, 2, \ldots)$$

为X的**分布律**或**概率分布**(Probability distribution)。也可以表示为

X	x_1	x_2	…	x_k	…
p_k	p_1	p_2	…	p_k	…

二、离散型随机变量概率分布的性质

(1) $p_k \geq 0, k = 1, 2, \ldots$;

(2) $\sum_k p_k = 1$

证明：由于$\{X = x_1\} \cup \{X = x_2\} \cup \ldots$是必然事件，且$\{X = x_j\} \cap \{X = x_k\} = \Phi$, $k \neq j$

$$故\ 1 = P[\bigcup_{i=1}^{\infty}\{X = x_i\}] = \sum_{i=1}^{\infty} P\{X = x_i\},\ 即\ \sum_{k=1}^{\infty} \mathrm{P}_k = 1$$

三、求解离散型随机变量分布律

求解技巧

- 根据题意，求出X的所有可能取的值x_k
- 利用古典概率，条件概率，加法或乘法公式等方法求出随机变量X取这些值的概率p_k，**注意所求出的概率全体之和为1**
- 根据规则写出分布律。

例：掷两颗骰子，定义三个随机变量X,Y和Z:

X为点数和；Y为6点的骰子个数；Z为最大点数

求三个随机变量的分布律

例：某射手有5发子弹，射一次命中的概率为0.9，如果命中了就停止射击，否则一直射到子弹用尽。

(1) 求耗用子弹数X的分布律；(2)求最多用3发子弹的概率。

三、3种离散型随机变量

1. （0-1）分布

若随机变量X只取0和1，其分布律为

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

则称X服从参数为p的**（0-1）分布(两点分布)** Two-point distribution

其分布律也可以写成

X	1	0
p_k	p	$1-p$

应用场合: 凡是随机试验只有两个可能的结果

三、3种离散型随机变量

2. 伯努利试验、二项分布

设将试验E只有两个可能结果：A和$\overline{A}$，则称E为**伯努利（Bernoulli）试验**。将E独立重复进行n次，则称这n次独立试验为**n重伯努利试验。**

若以X表示n重伯努利试验事件A发生的次数，则称X服从参数为n, p的二项分布（binomial distribution）。记作$X \sim b(n, p)$，其分布律为：

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k}, \qquad (k=0,1,\ldots,n)$$

三、3种离散型随机变量

2. 伯努利试验、二项分布

伯努利概型对试验结果**没有等可能的要求**，但有下述要求：

- 每次试验条件相同；
- 每次试验只考虑两个互逆结果A或$\overline{A}$，且$P(A)=p$，$P(\overline{A})=1-p$；
- 各次试验相互独立。

注 意

二项分布描述的是n重伯努利试验中出现“成功”次数X的概率分布。

三、3种离散型随机变量

2. 伯努利试验、二项分布

二项分布 $b(n,p)$ 和0-1分布之间的关系

设试验E只有两个结果：A和 $\bar{A}$。记$p=P(A),0<p<1$

若X服从0-1分布，则$X \sim b(1, p)$；

把试验E在相同条件下，相互独立地进行n次，记X为n次独立试验中结果A出现的次数，X_i为第i次试验中结果A出现的次数，则$X_i \sim b(1, p)$，且$X= X_1+X_2+\ldots+X_n \sim b(n, p)$。

三、3种离散型随机变量

2. 伯努利试验、二项分布

维修问题

设有80台同类型设备，各台工作是相互独立的，发生故的概率都是0.01，且一台设备的故障能由一个人处理。考虑两种配备维修工人的方法，其一是由4人维护，每人负责20台；其二是由3个人共同维护80台。试比较这两种方法在设备发生故障时不能及时维修的概率的大小。

三、3种离散型随机变量

2. 伯努利试验、二项分布

维修问题

解：

按第一种方案

X：第1人维护的20台中同一时刻发生故障的台数

$A_i(i=1,2,3,4)$：第i人维护的20台中发生故障不能及时维修

则知80台中发生故障而不能及时维修的概率为

$$P\{A_1 \cup A_2 \cup A_3 \cup A_4\} \geq P(A_1) = 1 - \sum_{k=0}^{1} P\{X = k\}$$

$$= 1 - \sum_{k=0}^{1} \binom{20}{k} (0.01)^k (0.99)^{20-k} = 0.0169$$

$$P\{A_1 \cup A_2 \cup A_3 \cup A_4\} \geq 0.0169$$

按第二种方案

以Y记80台中同一时刻发生故障的台数。此时，$Y \sim b(80, 0.01)$，故80台中发生故障而不能及时维修的概率为

$$P\{Y \geq 4\} = 1 - \sum_{k=0}^{3} \binom{80}{k} (0.01)^k (0.99)^{80-k}$$

$$0.0087$$

第二种方案任务重了(平均每人维护约27台)，但是工作效率不仅没有降低，反而提高了

心灵感应实验

§2 离散随机变量及其分布律

三、3种离散型随机变量

3. 泊松(Poisson)分布

设随机变量X所有可能取的值为0,1,2,… ,而取各个值的概率为：

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

其中λ>0是常数，则称**X服从参数为λ的泊松分布**，记为$X\sim\pi(\lambda)$。

易知：

$$\begin{cases} P\{X=k\}\ge 0 \quad k=0,1,2,\dots \\ \sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=1 \end{cases}$$

三、3种离散型随机变量

3. 泊松(Poisson)分布

泊松定理

设λ>0是一个常数，n是任意正整数，设$np_n=\lambda$，则对于任一固定的非负整数，有

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

定理的条件$np_n=\lambda$(常数)意味着当n很大时p_n必定很小，因此，上述定理表明当n很大，p很小时有以下近似式

$$\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k e^{-\lambda}}{k!}$$

即以n，p为参数的二项分布的概率值可以由参数为$\lambda=np_n$的泊松分布的概率值近似。

机器维修问题

Story

三、3种离散型随机变量

3. 泊松(Poisson)分布

在保险公司里有2500个同一年龄和同社会阶层的人参加了人寿保险，在一年里每个人死亡的概率为0.002，每个参加保险的人在1月1日付1200元保险费，而在当年死亡时家属可由保险公司领赔偿金200000元，问 (1)保险公司亏本的概率是多少？ (2)保险公司获利不少于100万元，200万元的概率。

三、3种离散型随机变量

3. 泊松(Poisson)分布

例 已知某商场一天来的顾客数X服从参数为λ的泊松分布，而每个来到商店的顾客购物的概率为p，证明：此商场一天内购物顾客数服从参数为λp的泊松分布

§3 随机变量的分布函数

一、随机变量分布函数概念的产生

对非离散型随机变量，其取值不是离散的，有时可以充满整个区间，对于这种更一般的随机变量， 我们感兴趣的就不是它取到某个具体的数的概率，而是它的取值落在某一个区间上的概率，

比如：$P\{x_1<X\leq x_2\}$，$P\{X>a\}$

$$P\{x_1<X\leq x_2\}=P\{X\leq x_2\}-P\{X\leq x_1\},$$

$$P\{X>a\}=1-P\{X\leq a\}$$

为此我们引入随机变量**分布函数的概念**。

离散型 r.v. — 分布律

连续型 r.v. — 分布函数

§3 随机变量的分布函数

一、随机变量分布函数概念的产生

设X是随机变量，对任意实数x，事件$\{X\le x\}$的概率$P\{X\le x\}$称为随机变量X的**分布函数**（Distribution function），记为$F(x)$，即 $F(x)=P\{X\le x\}$。

易知，对任意实数$a, b\ (a<b)$，

$P\{a<X\le b\}=P\{X\le b\}-P\{X\le a\}=F(b)-F(a)$。

分布函数完整的描述了随机变量的统计规律性

§3 随机变量的分布函数

一、随机变量分布函数概念的产生

分布函数F(x)的性质

单调不减性： 若$x_1<x_2$, 则$F(x_1)\le F(x_2)$;

归一 性： 对任意实数x，$0\le F(x)\le 1$， 且

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

右连续性：对任意实数x，

$$F(x_0+0)=\lim_{x\to x_0^+}F(x)=F(x_0)$$

这三个性质是分布函数的充分必要性质

§3 随机变量的分布函数

一、随机变量分布函数概念的产生

已知r.v. X的概率分布或分布列求分布函数$F(x)$

例： 设随机变量X的分布律为

X	-1	2	3
p_k	0.25	0.5	0.25

求X的概率分布，并求$P\{X\le 0.5\}$，$P\{1.5<X\le 2.5\}$，$P\{2\le X\le 3\}$。

例： 设随机变量X的分布函数为

$$F(x)=\begin{cases}0 & x<-1\\ 0.4 & -1\le x<1\\ 0.8 & 1\le x<3\\ 1 & x\ge 3\end{cases}$$

求X的分布列

- $F(x)$的各间断点x_k的取值，就是X的所有可能取值
- X取这些值的概率为，

$$p_k=P\{X=x_k\}=F(x_k)-F(x_k-0)$$

- 根据规则写出分布列

§4 连续型随机变量及其概率密度

一、概率密度

概率密度的定义

对于随机变量X，若存在非负函数$f(x)$ $(-\infty<x<+\infty)$，使对任意实数x，都有

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

则称X为连续型随机变量，$f(x)$为X的**概率密度函数**，简称**概率密度**或**密度函数**（probability density function or probability density）。

注：连续型随机变量的分布函数是连续函数。

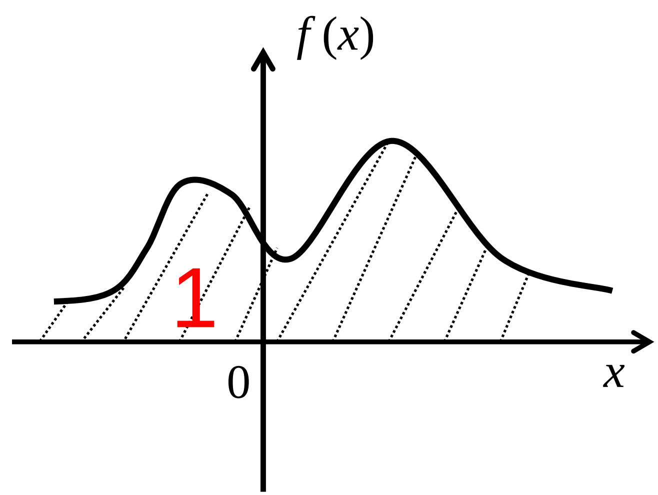
§4 连续型随机变量及其概率密度

一、概率密度

概率密度$f(x)$的性质

非负性：$f(x)\geq 0$，$(-\infty<x<\infty)$；

归一 性：$\int_{-\infty}^{+\infty} f(x)dx=1$



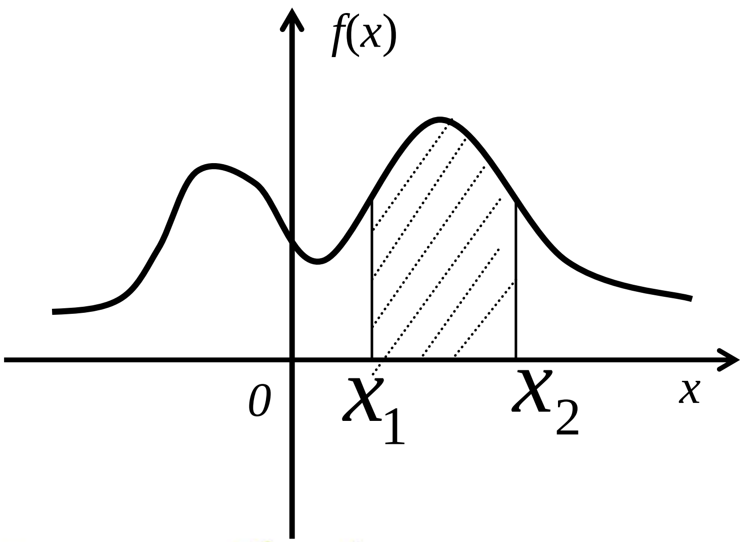
§4 连续型随机变量及其概率密度

一、概率密度

概率密度$f(x)$的性质

对于任意实数$x_1, x_2(x_1 \leq x_2)$,

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_2) = \int_{x_1}^{x_2} f(x)dx$$



X 落在区间$(x_1, x_2]$的概率$P\{x_1 < X \leq x_2\}$等于区间$(x_1, x_2]$上的曲线$y=f(x)$之下的曲边梯形的面积

§4 连续型随机变量及其概率密度

一、概率密度

概率密度$f(x)$的性质

若$f(x)$在点x处连续，则有$F'(x)=f(x)$

在$f(x)$的连续点处有：

$$f(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{P(x<X\leqslant x+\Delta x)}{\Delta x}$$

则若不计高阶无穷小，有：$P(x<X\leqslant x+\Delta x)\approx f(x)\Delta x$

一、概率密度

概率密度$f(x)$的性质

注 意

- 概率密度不是概率
- $f(x)$ 的大小决定X落在区间$(x, x+\Delta x]$内的概率的大小。它反映了点x附近所分布的概率的“疏密”程度
- 对任意实数a，有：$P\{X=a\}=0$

  所以，由$P(A)=0$，不能推出$A=\Phi$，

一、概率密度

概率密度$f(x)$相关例题

例：设r.v.X的概率密度函数为：

$$f(x)=\begin{cases}\dfrac{1}{3} & 若x\in[0,1]\\[2ex] \dfrac{2}{9} & 若x\in[3,6]\\[2ex] 0 & 其它\end{cases}$$

若k使得 $P\{X\ge k\}=\dfrac{2}{3}$，则k的取值范围是？

例：设r.v.X的概率密度函数为：

$$f(x)=\begin{cases}a\sqrt{1-x^2}, & -1<x\le 1\\ 0, & 其他\end{cases}$$

(1)求常数a；(2)X的分布函数；
(3) $P(0<x\le 1/2)$

§4 连续型随机变量及其概率密度

一、概率密度

离散型r.v.与连续型r.v.的比较

	离散型r.v.	连续型r.v.
可能取值	$X_i(i=1,2,3)$ 数轴上一些离散的点	连续变动的x充满某一空间
概率微元及性质	$P_i=P\{X=x_i\}\quad(i=1,2,3,\ldots)$ $P_i\ge 0,\quad\sum_{i=1}^{\infty}P_i=1$	$f(x)dx$ $f(x)\ge 0,\quad\int_{-\infty}^{\infty}f(x)dx=1$
分布函数及性质	$F(x)=\sum_{x_i\le x}P_i$ 上升的右连续阶梯函数在x_i处有间断	$F(x)=\int_{-\infty}^{x}f(t)dt$ 上升的连续函数 $F'(x)=f(x)$
取值特点	$\forall a\in R$ $P\{X=a\}$不一定为零	$\forall a\in R$ $P\{X=a\}$一定为零

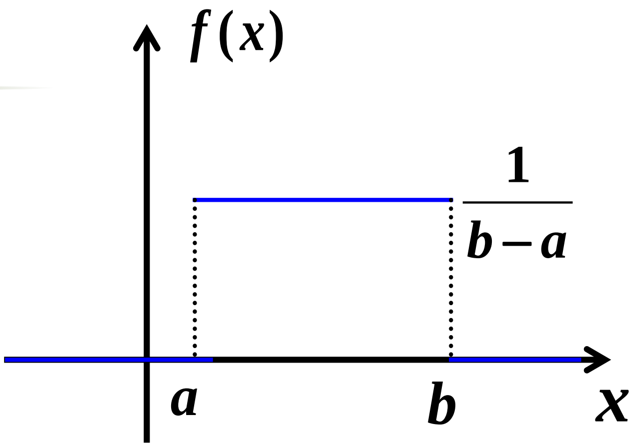
§4 连续型随机变量及其概率密度

二、常用的连续型随机变量

均匀分布

若连续型随机变量X具有概率密度

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



则称X在区间(a,b)上服从均匀分布，记为$X\sim U(a,b)$

$$f(x)\geq 0，且\int_{-\infty}^{+\infty} f(x)dx=1$$

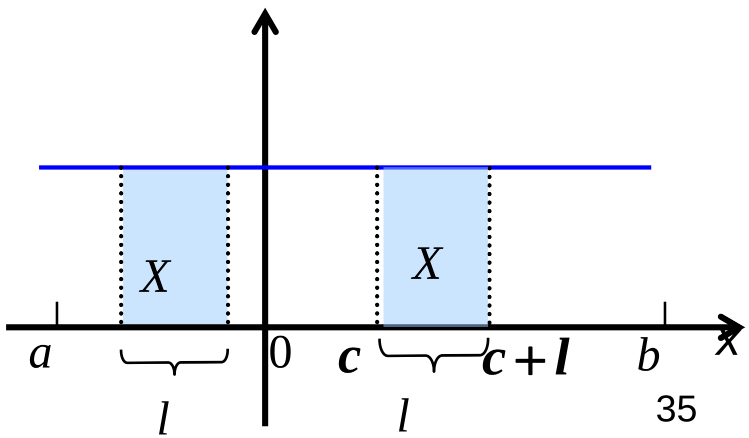
§4 连续型随机变量及其概率密度

二、常用的连续型随机变量

均匀分布的性质

均匀分布的**等可能性**：它落在区间(a, b)中任意等长度的子区间内的可能性是相同的；或者说它落在(a, b)的子区间内的概率只依赖于子区间的长度与子区间的位置无关

对于任意长度的子区间$(c, c+l)$, $a \le c < c+l \le b$，有

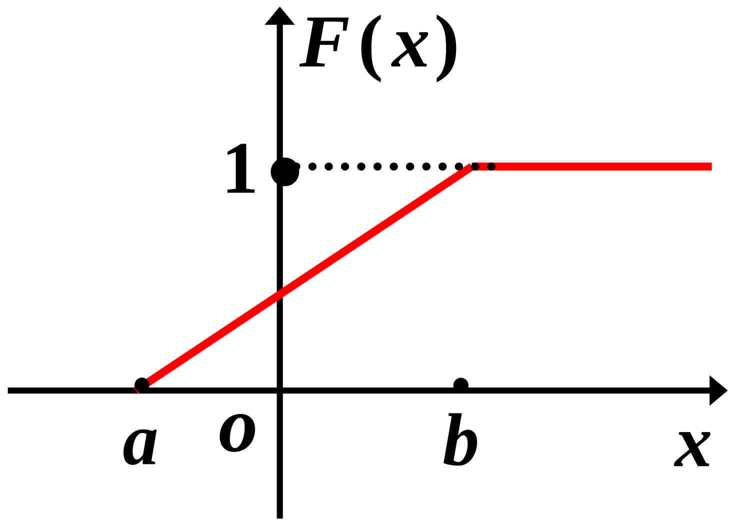


二、常用的连续型随机变量

均匀分布的分布函数

X的分布函数为

$$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$$



§4 连续型随机变量及其概率密度

二、常用的连续型随机变量

均匀分布相关例题

例 设随机变量X在区间[2, 5]上服从均匀分布，现对X进行3次独立观测，试求至少有两次观测值大于3的概率

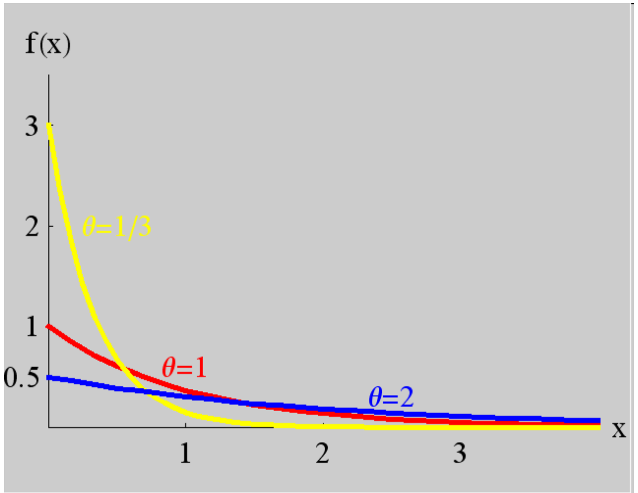
例 设随机变量X在区间[1, 6]上服从均匀分布，求方程$t^2 + Xt + 1 = 0$有实根的概率

二、常用的连续型随机变量

指数分布

若连续型随机变量X的概率密度为

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



其中$\theta>0$为常数，则称X服从参数为θ的指数分布

二、常用的连续型随机变量

指数分布的无记忆性

对于任意s，t>0，有

$$P\{X>s+t \mid X>s\}=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>s\}}$$

$$=\frac{P\{X>s+t\}}{P\{X>s\}}=\frac{1-F(s+t)}{1-F(s)}=\frac{e^{-(s+t)/\theta}}{e^{-(s)/\theta}}=e^{-(t)/\theta}=P\{X>t\}$$

“无记忆性”

二、常用的连续型随机变量

指数分布相关例题

例 设顾客在某银行的窗口等待服务的时间X(min)服从指数分布，其概率密度为

$$f_X(x)=\begin{cases}\dfrac{1}{5}e^{-x/5}, & x>0\\ 0 & \text{其他}\end{cases}$$

某顾客在窗口等待服务，若超过10min他就离开。他一个月要到银行5次。以Y表示一个月内他未等到服务而离开窗口的次数。写出Y的分布律，并求$P\{Y\geqslant 1\}$。

二、常用的连续型随机变量

正态分布

若连续型随机变量X的概率密度为

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

其中μ和σ都是常数, μ任意, $\sigma>0$, 则称X服从参数为μ和σ的正态分布。记为$X\sim N(\mu,\sigma^2)$

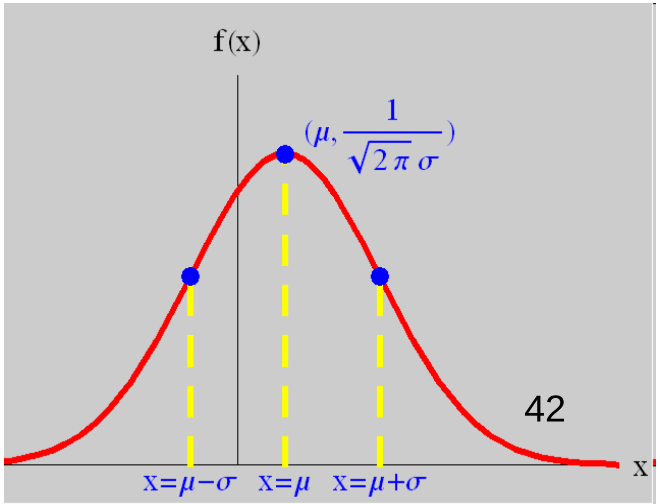
二、常用的连续型随机变量

正态分布的性质

- 曲线关于$x=\mu$对称。这表明对于任意$h>0$有

$$P\{\mu-h<X\le\mu\}=P\{\mu<X\le\mu+h\}$$

- 当$x=\mu$时取到最大值 $f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$
- X离μ越远，$f(x)$的值越小。$x\to\pm\infty$时，$f(x)\to 0$
- 曲线在$x=\mu\pm\sigma$处有拐点
- 曲线以x轴为渐近线



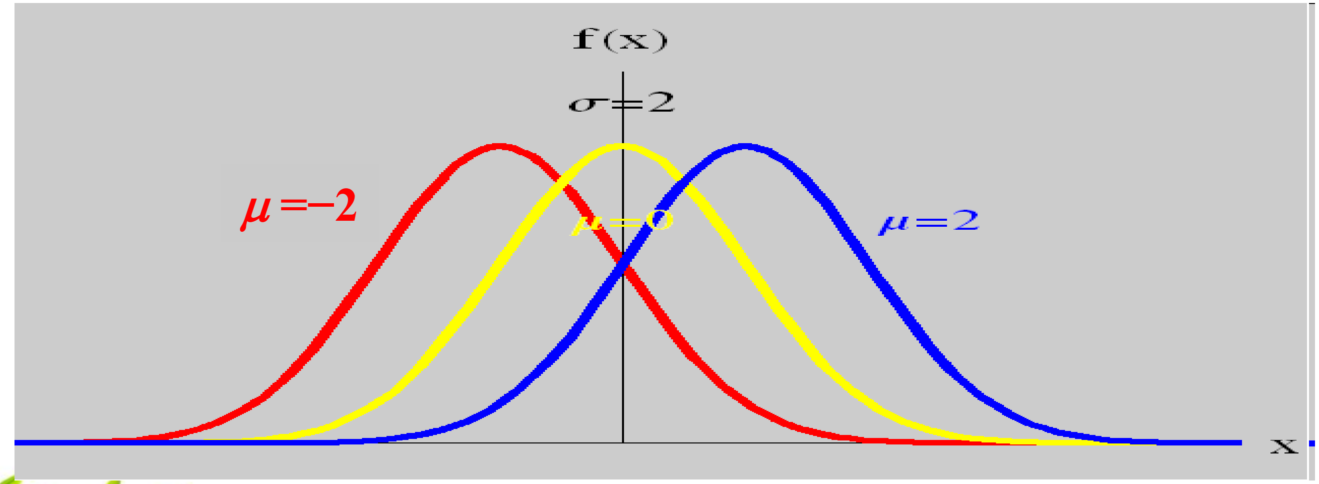
§4 连续型随机变量及其概率密度

二、常用的连续型随机变量

正态分布的两个参数

μ— 位置参数

固定 σ，对于不同的 μ，对应的 $f(x)$的形状不变化，只是沿着x轴平移。



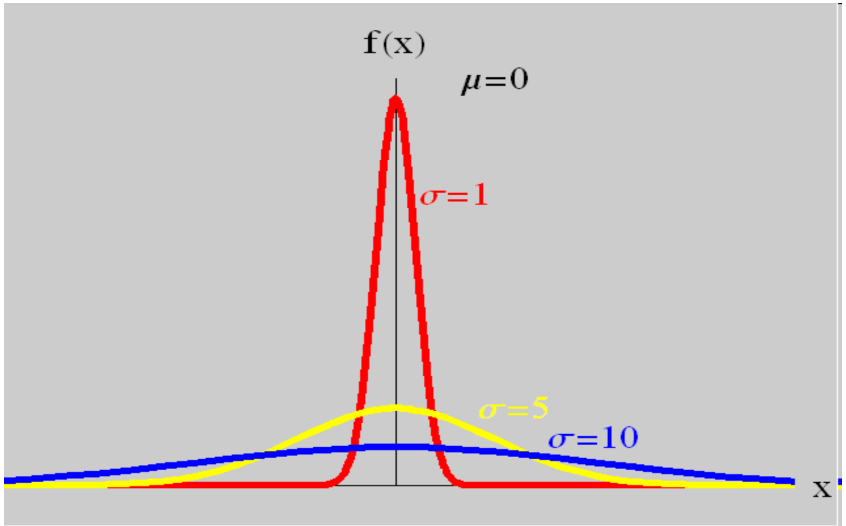
二、常用的连续型随机变量

正态分布的两个参数

σ— 形状参数

固定 μ，对于不同的σ，$f(x)$ 的形状不同。由于

$f(\mu)=\dfrac{1}{\sqrt{2\pi}\,\sigma}$，所以$\sigma$越小，$f(x)$变得越尖，而$X$落在$\mu$附近的概率越大

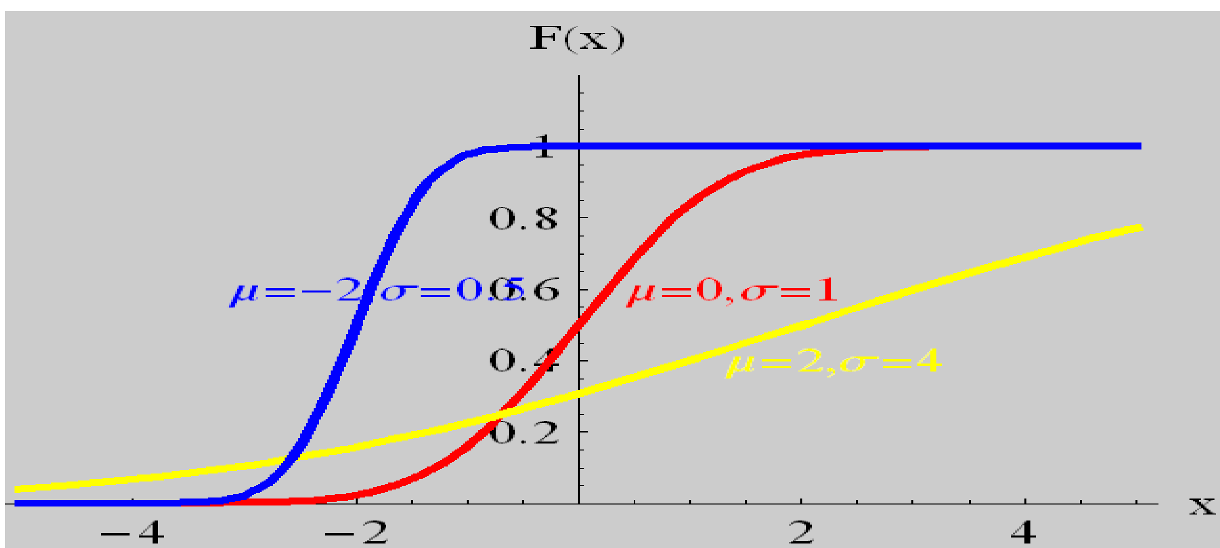


§4 连续型随机变量及其概率密度

二、常用的连续型随机变量

正态分布的分布函数

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$



Story

二、常用的连续型随机变量

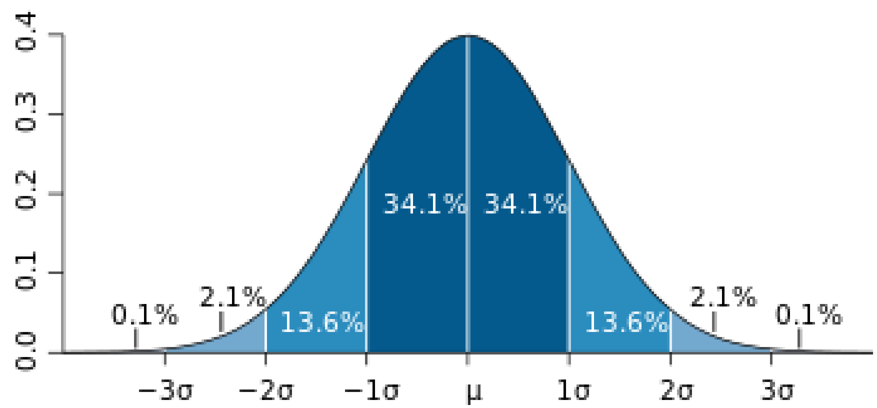
标准正态分布

当正态分布$X\sim N(\mu,\sigma^2)$中$\mu=0$，$\sigma=1$时，称X服从标准正态分布。其概率密度和分布函数分别用$\varphi(x)$、$\Phi(x)$表示，既有：

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad -\infty<x<\infty$$

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt,\quad -\infty<x<\infty$$

$$\Phi(-x)=1-\Phi(x)$$



二、常用的连续型随机变量

标准正态分布

引 理

若$X\sim N(\mu,\sigma^2)$，则$z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$

证明： $z=\dfrac{X-\mu}{\sigma}$的分布函数为

$$P\{Z\le x\}=P\left\{\frac{X-\mu}{\sigma}\le x\right\}=P\{X\le\mu+\sigma x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\mu+\sigma x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

令$\dfrac{t-\mu}{\sigma}=u$，得 $P\{Z\le x\}=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du=\Phi(x)$

由此知$z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$

二、常用的连续型随机变量

标准正态分布

于是，若$X\sim N(\mu, \sigma^2)$，则它的分布函数$F(x)$可写成

$$F(x)=P\{X\le x\}=P\left\{\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

对于任意区间$(x_1, x_2]$，有

$$P\{x_1\le X\le x_2\}=P\left\{\frac{x_1-\mu}{\sigma}\le\frac{X-\mu}{\sigma}\le\frac{x_2-\mu}{\sigma}\right\}$$

$$=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

Story

二、常用的连续型随机变量

正态分布相关例题

例 设$\ln(X)\sim N(1,4)$，求$P\{0.5<X<2\}$

解： 由题意得，$\dfrac{\ln X-1}{2}\sim N(0,1)$

$$P\{0.5<X<2\}=\Phi\left(\frac{\ln 2-1}{2}\right)-\Phi\left(\frac{\ln 0.5-1}{2}\right)$$

$$=\Phi(-0.153)-\Phi(-0.847)$$

$$=0.8023-0.5596$$

$$=0.2497$$

若$\ln X\sim N(\mu,\sigma^2)$，则称X服从对数正态分布，求X的密度函数

计算$\int_{-\infty}^{\infty}e^{-at^2}dt\ \ (a>0)$

二、常用的连续型随机变量

正态分布相关例题

例 设某工程队完成某项工程所需时间X(天)近似服从$N(100,5^2)$，工程队上级规定：若工程在100天内完成，可以得到奖金10万元；在100~115天内完成可得奖金3万元；若超过115天完成，罚款5万元。求该工程队在完成该工程时，获奖金额的分布律。

解： 设Y为事件：获奖金额； $X\sim N(100,5^2)$

$$Y=\begin{cases}10 & 0<X<100\\ 3 & 100\le X\le 115\\ -5 & 115\le X<+\infty\end{cases}$$

由$X\sim N(100,5^2)$，可以得到

$$P\{Y=10\}=P\{X<100\}=\Phi\left(\frac{100-100}{5}\right)=0.5$$

$$P\{Y=3\}=P\{100\le X\le 115\}$$

$$=\Phi\left(\frac{115-100}{5}\right)-\Phi\left(\frac{100-100}{5}\right)=\Phi(3)-\Phi(0)=0.4987$$

$$P\{Y=-5\}=P\{X>115\}=1-\Phi\left(\frac{115-100}{5}\right)$$

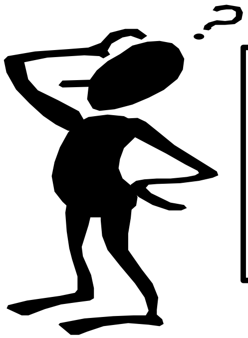
$$=1-\Phi(3)=0.0013$$

Y的分布律：

Y	-5	3	10
P_k	0.0013	0.4987	0.5000

§5 随机变量的函数的分布

设$g(x)$是定义在随机变量X的一切可能值x的集合上的函数，若随机变量Y随着X取值x的值而取$y=g(x)$的值，则称随机变量Y为随机变量X的函数，记作$Y=g(X)$



如何根据已知的随机变量X的分布求得随机变量$Y=g(X)$的分布？

一、随机变量$Y=f(X)$的分布的求解

离散型情况

例 设r.v.X的分布律为：

X	-2	-1	0	1	2
P	0.3	0.2	0.1	0.3	0.1

试求：1) $Y=3X+2$的分布律；　2) $Z=X^2$的分布律

解： 1) 列表 ⇒

X	-2	-1	0	1	2
$Y=3X+2$	-4	-1	2	5	8
P	0.3	0.2	0.1	0.3	0.1

由上表可得$Y=3X+2$的分布律：

2) 列表 ⇒

X	-2	-1	0	1	2
$Z=X^2$	4	1	0	1	4
P	0.3	0.2	0.1	0.3	0.1

⇒ $Z=X^2$的分布律：

Z	0	1	4
P	0.1	0.4	0.4

§5 随机变量的函数的分布

一、随机变量$Y=f(X)$的分布的求解

离散型情况

方法总结

一般，若X的分布律为：

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	P_1	P_2	$\cdots$	P_i	$\cdots$

则$Y=g(X)$的分布律为：

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_i)$	$\cdots$
P	P_1	P_2	$\cdots$	P_i	$\cdots$

注意：若$g(x_1)$，$g(x_2)$，$\ldots g(x_i)$中有相同的值，则应合同值列并

§5 随机变量的函数的分布

一、随机变量$Y=f(X)$的分布的求解

连续型情况

若X是连续r.v.，则$Y=g(X)$一般也是r.v.；若已知r.v.X的概率密度$f_X(x)$，则可以用分布函数法求$Y=g(X)$的概率密度

方法总结

- 先求X的定义域$\boldsymbol{\Omega}_X$，确定$Y=g(X)$的值域Ω_Y；
- 对于任意的$Y\in\Omega_Y$，写出Y的分布函数；

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)=P(x\in G_Y)=\int_{G_Y} f_s(x)dx$$

$$\left(\text{其中}G_Y=\{x|g(x)\le y\}\right)$$

- 写出$F_Y(y)$在$(-\infty,\infty)$上的表达式
- 求导可得 $f_Y(y)=F_Y'(y)$

一、随机变量$Y=f(X)$的分布的求解

连续型情况

例 已知若X~N(0,1)，试求Y=e^{X}的概率密度$f_Y(y)$

解： 由X的定义域$\Omega_X=(-\infty,\infty)$，可确定$Y$=e^{X}的值域$\Omega_Y=(0,\infty)$；

对任意的$Y\in\Omega_Y=(0,\infty)$，Y的分布函数：

$$F_Y(y)=P(Y\le y)=P\left(e^X\le y\right)=P\{X\le\ln y\}=\int_{-\infty}^{\ln y}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

而当$y\le 0$时，$F_Y(y)=P(Y\le y)=P\left(e^X\le y\right)=P(\Phi)=0$

求导可以得到：
$$f_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\,y}e^{-\frac{(\ln y)^2}{2}} & y>0\\ 0 & y\le 0\end{cases}$$

§5 随机变量的函数的分布

一、随机变量$Y=f(X)$的分布的求解

连续型情况

定 理

设随机变量X具有概率密度$f_X(x)$, $-\infty<x<\infty$, 又设函数g(x)处处可导且恒有$g'(x)>0$（或恒有$g'(x)<0$），则$Y=f(X)$是连续型随机变量，其概率密度为

$$f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)|, & \alpha<y<\beta \\ 0 & else \end{cases}$$

其中$\alpha=\min\{g(-\infty), g(\infty)\}$, $\beta=\max\{g(-\infty), g(\infty)\}$, $h(y)$是g(x)的反函数

§5 随机变量的函数的分布

一、随机变量$Y=f(X)$的分布的求解

连续型情况

定理证明

1) 讨论$g'(x)>0$的情况，此时$g(x)$严格单调增加，它的反函数存在，且在(α,β)严格单调增加，可导。

当$y=g(x)$在(α,β)取值，

当$y\le\alpha$时，$F_Y(y)=P(Y\le y)=0$

当$y\ge\beta$时，$F_Y(y)=P(Y\le y)=1$

当$\alpha<y<\beta$时，$F_Y(y)=P(Y\le y)$

$$=P(g(X)\le y)=P(X\le h(y))=\int_{-\infty}^{h(y)}f(x)dx$$

则，概率密度函数

$$f_Y(y)=\begin{cases} f_X[h(y)]h'(y), & \alpha<y<\beta \\ 0 & else \end{cases}$$

2) 讨论$g'(x)<0$的情况可以同样地证明，此时有

$$f_Y(y)=\begin{cases} f_X[h(y)][-h'(y)], & \alpha<y<\beta \\ 0 & else \end{cases}$$

综上所述，可以得到：

$$f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)|, & \alpha<y<\beta \\ 0 & else \end{cases}$$

其中 $\alpha=\min\{g(-\infty),\ g(\infty)\}$, $\beta=\max\{g(-\infty),\ g(\infty)\}$, $h(y)$ 是 $g(x)$ 的反函数

§5 随机变量的函数的分布

一、随机变量$Y=f(X)$的分布的求解

连续型情况

定 理

若r.v. $X\sim N(\mu,\sigma^2)$， $Y=aX+b(k\neq 0)$, 则$Y\sim N(a\mu+b, a^2\sigma^2)$

即服从正态分布的随机变量的线性变换仍然服从正态分布

证明： $Y=aX+b$是单调函数，其反函数为： $x=\dfrac{Y-b}{a}$

$Y=aX+b$的值域为$(-\infty,\infty)$，X的概率密度$f(x)$为： $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

由前面的定理易得Y概率密度$f_Y(y)$为：

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{|a|}=\frac{1}{\sqrt{2\pi}\sigma|a|}e^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}}$$

特别，取$a=\dfrac{1}{\sigma},b=-\dfrac{\mu}{\sigma}$得：$Y=\dfrac{X-\mu}{\sigma}\sim N(0,1)$

一、随机变量$Y=f(X)$的分布的求解

连续型情况

例 若X服从区间$(0,\pi)$上的均匀分布，试求$Y=\sin X$的概率密度$f_Y(y)$

解： 由X的值域$\Omega_X=(0,\pi)$，可确定$Y=\sin X$的值域$\Omega_Y=(0,1)$；

当$y\in(0,1)$，

$$F_Y(y)=P(Y\le y)=P(\sin X\le y)=P(X\in(0,\arcsin y]\cup[\pi-\arcsin y,\pi))$$

$$=\int_0^{\arcsin y}\frac{1}{\pi}dx+\int_{\pi-\arcsin y}^{\pi}\frac{1}{\pi}dx=\frac{2}{\pi}\arcsin y$$

显然，当$y\le 0$时，$F_Y(y)=0$；当$y\geqslant 0$时，$F_Y(y)=1$；

从而，$Y=\sin X$的概率密度为：

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi}\dfrac{1}{\sqrt{1-y^2}}, & 1>y>0\\ 0 & else\end{cases}$$

谢　　谢!