一维随机变量及其分布

随机变量的概念 I

随机变量是根据出现的试验结果取实值的函数。

请指出下面随机现象结果的区别

1. 掷一次骰子，出现的点数
2. 家用电器的使用寿命
3. 掷一枚硬币，观察向上的面
4. 抽取一件产品，检查其是否合格

随机变量的简单认识

1. 掷一次骰子，出现的点数 $X \in \{1, 2, 3, 4, 5, 6\}$
2. 家用电器的使用寿命 $X \in [0, \infty)$
3. 掷一枚硬币，观察向上的面 $X \in \{0(\text{正面}), 1(\text{反面})\}$
4. 抽取一件产品，检查其是否合格 $X \in \{0(\text{合格}), 1(\text{不合格})\}$

请指出上面的X是什么样的函数

随机变量的定义

定义2.1.1

设$(\Omega, \mathcal{F}, P)$为概率空间，称映射$X:\Omega \to R$为**随机变量**，如果对于任意$x \in \mathbb{R}$，有

$$A_x = \{\omega \mid X(\omega) \leq x\} \in \mathcal{F} \tag{1}$$

这个定义包含了下面几个意思

1. 随机变量是从样本空间到实轴的**映射**
2. 对于**任意**的$x \in \mathbb{R}$，满足(1)（反例见例题2.1.1）

随机变量和事件

假设，给定$a < b \in \mathbb{R}$

$$A_a = \{\omega : X(\omega) < a\}$$

$$A_b = \{\omega : X(\omega) < b\}$$

请定义下面的事件

$$\overline{A_a}, \quad \overline{A_b}, \quad A_a \cap A_b, \quad A_a - A_b$$

随机变量的分布函数 I

定义2.1.2

设$(\Omega, \mathcal{F}, P)$为概率空间，X为随机变量，X的**分布函数**F_X定义为

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \ \forall x \in \mathbb{R}$$

以后将$\{\omega \in \Omega : X(\omega) \leq x\}$简写为$X \leq x$。

分布函数和事件

根据分布函数的定义，显然的

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad \forall a < b \in \mathbb{R}$$

根据概率的上下连续性，对于$\forall a < b \in \mathbb{R}$，有下面的事实

$$
\begin{aligned}
&P(X = a) = F_X(a) - F_X(a-0),\\
&P(a \le X \le b) = F_X(b) - F_X(a-0),\\
&P(a \le X < b) = F_X(b-0) - F_X(a-0),\\
&P(a < X < b) = F_X(b-0) - F_X(a)
\end{aligned}
$$

当$F_X(x)$在a连续时，$F_X(a-0) = F_X(a+0) = F_X(a)$

分布函数的性质 I

1. **有界性** $0 \le F_X(x) \le 1$， $\forall x \in \mathbb{R}$

2. **单调性** 对于任意的$x_1 < x_2$，有$F_X(x_1) \le F_X(x_2)$ 并且

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to +\infty} F_X(x) = 1.$$

3. **右连续性** $\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$

$F(x)$是某个随机变量的分布函数$\Leftrightarrow$ $F(x)$满足上面的三条性质

课堂练习

1. 向半径为r的圆内随机抛一点，求此点到圆心距离X的分布函数，并且求$P(X > \frac{2r}{3})$
2. 口袋中有5个球，编号为1,2,3,4,5，从中任取3个，以X表示3个球中最大的号码，求X的分布函数，并作图。

当随机变量只能取有限个或者可数个函数值的时候，称为**离散型随机变量**。

假设一个定义在概率空间$\{\Omega, \mathcal{F}, P\}$上的随机变量$X$只有可数个取值，记作$a_1$，$a_2$，$\cdots$，$a_n$，$\cdots$，且

$$P(X = a_i) = p_i,\ i = 1, 2, \cdots, n, \cdots$$

通常称

$$X \sim \begin{pmatrix} a_1 & a_2 & \cdots & a_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

的右端为X的**分布列**，称$(p_1, p_2, \cdots, p_n, \cdots)$为**概率分布**。

分布列的性质

1. $p_i \geq 0, i = 1, 2, \cdots$
2. $\sum_{i=1}^{\infty} p_i = 1$
3. X的分布函数为

$$F_X(x) = P(X \leq x) = \sum_{a_i \leq x} p_i$$

$F_X(x)$的图像为右连续的阶梯函数。

$$P(b < X \leq c) = \sum_{b<a_i\leq c} p_i$$

课堂练习

一汽车沿着街道行驶，需要经过三个红绿灯，若每个信号灯显示红绿两种信号的时间相等，且各个信号灯工作相互独立。以X表示该汽车首次遇到红灯前已经通过的路口数，求X的分布列

离散型随机变量

本章将要学习服从几种特殊分布的离散型随机变量

1. 二项分布
2. 泊松分布
3. 几何分布

二项分布

若一个随机变量的取值为$0,1,2,\cdots,n$，且

$$P(X=k)=b(k;n,p)=\left(\begin{array}{c} n \\ k \end{array}\right)p^k(1-p)^{n-k},\ k=0,1,\cdots,n,$$

这里我们称X服从**二项分布**，记为$X\sim B(n,p)$。

特别的，如果$n=1$，那么X只有0和1两个取值，我们称X服从**两点分布**。进一步，若$p=1$或者0，那么两点分布退化为**单点分布**。

二项分布的性质

若 $X \sim B(n,p)$，则

1. $P(X = k) \geq 0$，
2. 由二项式定理，

$$\sum_{k=0}^{n} P(X = k) = 1,$$

3.

$$P(X \leq k) = \sum_{i=0}^{k} b(i; n, p)$$

二项分布的例子

记n重伯努利试验中成功的次数为X，且一次成功的概率为p，那么

$$X \sim B(n,p)$$

那么n次试验中成功k次的概率为

$$P(X=k)=b(k;n,p)=\left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}$$

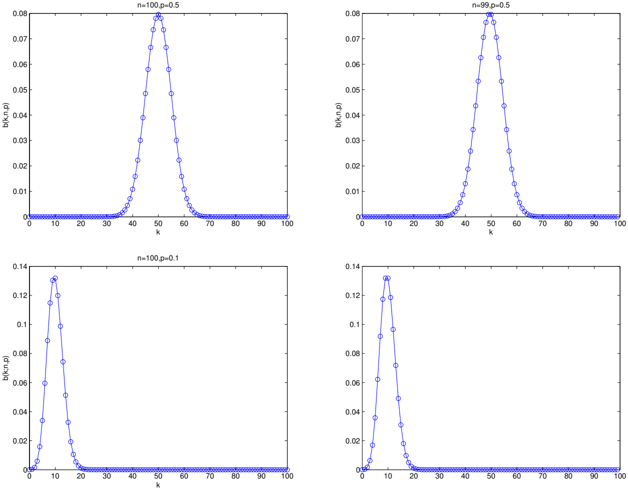
课堂练习

1. 设$X \sim B(10, 0.9)$，求$P(X = 8)$, $P(X \le 8)$, $P(3 \le X \le 9)$.
2. 设某种药物的治愈率是0.6，请问需要同时治疗多少位病人，才能使得至少有一位病人被治愈的概率为90%
3. 设随机变量$X \sim B(2, p)$，$Y \sim B(3, p)$，若$P(X \ge 1) = 5/9$，求$P(Y \ge 1)$。

二项分布的极值

定理2.2.1

设$X \sim B(n,p)$，当$(n+1)p$为整数m时，X取m和$m-1$的概率最大，且$b(m;n,p)=b(m-1;n,p)$；当$(n+1)p$不为整数时，X取$\lfloor (n+1)p \rfloor$的时候概率最大。


n=100,p=0.5
n=99,p=0.5
n=100,p=0.1
b(k,n,p)
b(k;n,p)
k

估计鱼塘中的鱼的总数 I

为了估计一个鱼塘中鱼的总数，我们做下面的试验

1. 网起一兜鱼，假设是100条，给这些鱼做上标记。
2. 把做过标记的鱼放回鱼塘中，让这些放回去的鱼欢快的游一段时间。
3. 再网起一兜，假设是80条，数出有标记的鱼的数目，假设有2条。

怎样根据试验的结果，估计鱼塘中鱼的总数呢?

定义随机变量X表示80条鱼中有记号的鱼的个数，那么$X \sim B(80, \frac{100}{N})$，$X = 0, 1, 2, \cdots, 80$。

由于**小概率事件在一次试验中基本不发生，即一次随机试验出现的结果是发生可能性大的，或者是最大的**，那么根据定理2.2.1，有

$$2 = (80 + 1)\frac{100}{N}$$

即$N = 4050$

泊松分布 I

如果一个随机变量只能取非负整数值，并且

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

则称X服从泊松(Poisson)分布，记为$X\sim Pois(\lambda)$。

泊松定理

设随机变量$X \sim B(n, p_n)$, $0 < p_n < 1$，并且有$\lim_{n\to\infty} np_n = \lambda$，那么

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

泊松分布描述的是“稀有事件

Proof: 记$\lambda_n = np_n$，那么

$$b(k;n,p_n) = \left(\begin{array}{c} n \\ k \end{array} \right) p_n^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

$$= \frac{\lambda_n^k}{k!} 1 \cdot \left(1-\frac{1}{n}\right) \cdot \left(1-\frac{2}{n}\right) \cdots \left(1-\frac{k-1}{n}\right) \left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

固定k，有

$$\lim_{n\to\infty} \lambda_n^k = \lambda^k,$$

$$\lim_{n\to\infty} (1-\tfrac{\lambda_n}{n})^{n-k} = e^{-\lambda} (hint: \lim_{x\to\infty} (1+\tfrac{1}{x})^x = e)$$

以及

$$\lim_{n\to\infty} \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{k-1}{n}\right) = 1$$

故

$$\lim_{n\to\infty} b(k;n,p_n) = \frac{e^{-\lambda}\lambda^k}{k!}$$

泊松分布的应用 I

对于可以用二项分布$X \sim B(n, p_n)$描述的随机变量，如果同时满足

1. $n \to \infty$, $p_n \to 0$
2. $\lim_{n\to\infty} np_n = \lambda$

我们就可以用泊松分布代替二项分布。

孕妇生三胎的问题

假如一个孕妇生三胞胎的概率是$p = 10^{-4}$，求在100000个孕妇中，有2个人生三胞胎的概率。

解：由于每个孕妇生几胞胎这个试验是相互独立的，故定义随机变量X为生三胞胎的人数，$X \sim B(100000, 10^{-4})$。

$A_i = \{100000$个孕妇中有i个人生三胞胎$\}$

$$P(A_2) = P_{\text{binom}}(X = 2) = b(2, 10^5, 10^{-4})$$
$$= \binom{10^5}{2}(10^{-4})^2(1 - 10^{-4})^{10^5-2} = 0.002269293$$

注意$p \ll 1$，$n \gg 1$，$pn = 10$，我们用泊松分布来计算一下

$$P(A_2) = P_{\text{pois}}(X = 2) = 0.00226996$$

近似效果还是可以的，绝对误差在第5位有效数字上。

多人合作问题 I

同类设备有80台，各台设备是否正常工作是独立事件，每台设备的故障率是$p = 0.01$，一台设备发生故障时需要安排一人维修。求 由三个人共同负责维修80台设备的时候，设备发生故障无人维修的概率。

分析： 当三个人共同负责80台设备的时候，只要有四台以及以上的设备同时发生故障，就有设备没有人去维修。

令 X为同时发生故障的设备数，$A = \{$至少有四台设备同时发生故障$\}$。

$$P(A) = P_{\text{binom}}(X \geq 4) = 1 - P(X \leq 3) = 0.008659189$$

用泊松分布计算

$$P(A) = P_{\text{pois}}(X \geq 4) = 0.009079858$$

多人合作问题 II

当n不是很大的时候，泊松分布对二项分布的近似并不理想。

几何分布 I

定义随机试验，是一次次的做伯努利试验，直到第一次成功为止，记录所做的伯努利试验次数。

设每次独立试验成功的概率为p，记首次成功的时候做的伯努利试验的次数为X，那么$X=k$代表事件

$A=\{$第$1,2,\cdots,k-1$次试验都是失败的，第k次试验是成功的$\}$

则有

$$P(A)=P_{\text{geom}}(X=k)=(1-p)^{k-1}p$$

称X服从**几何分布**，记作$X\sim Geo(p)$。

矿井逃生问题 I

设一个地下采矿点有5个可以上升到地面的通道，如果事故发生，只有一个通道可以逃生，且由于没有照明，所以遇险者**每次只能随意**的在5个通道中选择一个，若发现该通道不通，则需要返回原点再随意的选择一个。求

1. 第三次才选择正确的通道的概率
2. 选择成功时已经选择其他错误的通道次数不大于6次的概率

解：定义随机变量X为选择正确时总共选择的次数，那么$X \sim Geo(\frac{1}{5})$

1. 第三次才选择正确的概率

$$P_{\text{Geo}}(X=3) = \left(\frac{4}{5}\right)^2 \frac{1}{5}$$

矿井逃生问题 II

2 选择正确前选择其他通道的次数不大于6

$$P_{\text{Geo}}(X \le 7) = \sum_{k=1}^{7} \left(\frac{4}{5}\right)^{k-1} \frac{1}{5}$$

几何分布具有“无记忆性”，也就是说当前k次试验都不成功时，第$k+n$次试验首次成功与k的大小无关，即

$$P(X = k + n \mid X > k) = P(X = n)$$

矿井逃生问题 III

证明

$$
\begin{aligned}
P(X = k+n \mid X > k) &= \frac{P((X = k+n) \cap (X > k))}{P(X > k)} \\
&= \frac{P(X = k+n)}{1 - P(X \le k)} \\
&= \frac{(1-p)^{n+k-1} p}{(1-p)^k} \\
&= (1-p)^{n-1} p = P(X = n)
\end{aligned}
$$

作为练习，请大家验证

$$
P(X > k+n \mid X > k) = P(X > n)
$$

课堂练习

袋中有m个白球，n个黑球，现有放回的摸球，直到摸到白球停止，请问已知摸球的次数大于3，那么在第5次后停止的概率。

一维连续型随机变量 I

若随机变量的可能取值充满$\mathbb{R}$上的一个区间(a,b)，则称其为**连续随机变量**。

请思考下面的问题：在$[a,b]$区间随机的投点，求落点位置X的分布函数

定义2.3.1

设$(\Omega, \mathcal{F}, P)$为概率空间，X为其上的随机变量，F_X为X的分布函数。如果存在**非负可积**的函数f_X，使得

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t, \quad \forall x \in (-\infty, \infty)$$

那么称X为**连续型随机变量**，称$f_X(x)$为X的**分布密度函数**。

由Newton-Leibniz公式，在$f_X(x)$的连续点上，有

$$f_X(x) = F'_X(x)$$

1. 若$f_X(x)$是X的分布密度函数，那么
   1. $f_X(x) \geq 0, \ \forall x \in (-\infty, \infty)$
   2. $\int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = 1$
2. 若定义在$(-\infty, \infty)$上的函数f满足上面的两点，那么令

$$F(x) = \int_{-\infty}^{x} f(t) \, \mathrm{d}t, x \in (-\infty, \infty)$$

那么$F(x)$一定是某随机变量的分布函数。

显然的

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,\mathrm{d}x,\ \forall a < b \in \mathbb{R}$$

而

$$P(X = a) = \lim_{h \to 0+} P(a - h < X \leq a) = \lim_{h \to 0+} \int_{a-h}^{a} f_X(x)\,\mathrm{d}x = 0$$

上面的式子表明连续型随机变量取任意的单点值的概率为零，换句话说，连续型随机变量的分布特性不能通过列举它取单点值的概率表示。

根据Riemann积分的性质，存在$\eta \in (m, M)$，其中$m = \min_{[x-\Delta x, x+\Delta x]} f_X(x)$，$M = \max_{[x-\Delta x, x+\Delta x]} f_X(x)$ 使得

$$\begin{aligned}\eta \Delta x &= \int_{x-\Delta x}^{x+\Delta x} f_X(t)\,\mathrm{d}t = F_X(x+\Delta x) - F_X(x-\Delta x) \\ &= P(x-\Delta x < X \leq x+\Delta x)\end{aligned}$$

说明连续型随机变量在密度函数取值大的点附近取值的概率也较大。

例子

设随机变量X的分布密度函数为

$$f_X(x) = \frac{a}{1+x^2}, x \in (-\infty, \infty)$$

1. 确定a的值
2. 求X的分布函数
3. 求$P(X^2 \leq 1)$

学习重点

本章将要学习服从几种特殊分布的连续型随机变量

1. 均匀分布
2. 指数分布
3. 正态分布

均匀分布 I

如果连续型随机变量X的分布密度函数为

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{当} x \in [a,b] \\ 0, & \text{当} x \notin [a,b] \end{cases}$$

则称X服从$[a,b]$上的**均匀分布**，记为$X \sim U[a,b]$，其分布函数为

$$F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0, & \text{当} x < a \\ \dfrac{x-a}{b-a}, & \text{当} a \le x \le b \\ 1, & \text{当} x > b \end{cases}$$

课堂练习

1. 设随机变量$X \sim U[0,10]$，对X进行4次独立观测，求至少有3次观测值大于5的概率。
2. 在$[0,1]$上任取一点记为X，求$P(X^2 - \frac{3}{4}X + \frac{1}{8} \geq 0)$

指数分布 I

定义

若连续型随机变量X的分布密度函数为

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, \text{当} x > 0, \\ \quad 0, \text{当} x \le 0 \end{cases}$$

其中$\lambda > 0$为参数，记为$X \sim \text{Exp}(\lambda)$, 其分布函数为

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, \text{当} x > 0 \\ 0, \quad \text{当} x \le 0 \end{cases}$$

指数分布 II

某窗口接待一位顾客的服务时间T服从参数为$\frac{1}{10}$的指数分布，即

$$f_T(t) = \begin{cases} \frac{1}{10}e^{-\frac{1}{10}t}, \text{当} t > 0, \\ 0, \text{当} t \leq 0 \end{cases}$$

假设一次服务时间超过15分钟，顾客即评价为“不满意”，求

1. 某位顾客不满意的概率
2. 十位顾客中恰有两位顾客评价为不满意的概率
3. 十位顾客中最多有两位评价为不满意的概率
4. 十位顾客中至少有两位评价为不满意的概率

指数分布 III

设事件$A = \{$某位顾客不满意$\}$

$$P(A) = \int_{15}^{\infty} \frac{1}{10} e^{-\frac{1}{10}x} \, \mathrm{d}x$$

由于每位顾客接受服务的试验相互独立，所以某一窗口服务多位顾客是n重伯努利试验

记不满意的顾客数为Y，$Y \sim B(10, P(A))$

那么题目中的第$2,3,4$小问分别对应下面三个概率

$$P(Y = 2), \; P(Y \leq 2), \; P(Y \geq 2)$$

根据二项分布公式

$$P(Y = 2) = \left(\begin{array}{c} 10 \\ 2 \end{array} \right) P(A)^2 P(\bar{A})^8$$

指数分布 IV

$$P(Y \leq 2) = \sum_{k=0}^{2} \left(\begin{array}{c} 10 \\ k \end{array} \right) P(A)^k P(\bar{A})^{10-k}$$

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \sum_{k=0}^{1} \left(\begin{array}{c} 10 \\ k \end{array} \right) P(A)^k P(\bar{A})^{10-k}$$

指数分布也具有无记忆性，即

$$P(X > t + s \,|\, X > s) = P(X > t)$$

正态分布 I

连续型随机变量服从的最常见的分布是**正态分布**

正态分布的密度函数

若X服从正态分布，那么其分布密度函数为

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

其中$\mu \in \mathbb{R}$和$\sigma > 0$为正态分布的参数，记$X \sim N(\mu, \sigma^2)$。

特别的，当$\mu = 0, \sigma = 1$时，称X服从**标准正态分布**。

正态分布的密度函数图像

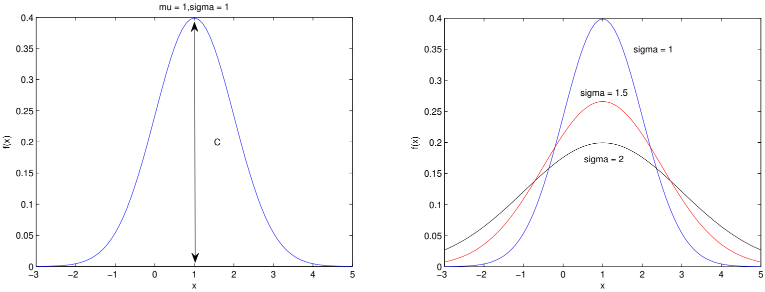


Figure : Left: $C = \frac{1}{\sqrt{2\pi}\sigma}$, right: $\mu = 1$

标准正态分布 I

当$X \sim N(0,1)$，其密度函数为

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

其分布函数为$\Phi(x)$，

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

显然

$$\begin{aligned}
\Phi(-x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{\frac{-t^2}{2}} \, \mathrm{d}t = -\frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{\frac{-\xi^2}{2}} \, \mathrm{d}\xi \\
&= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{\frac{-\xi^2}{2}} \, \mathrm{d}\xi = 1 - \Phi(x)
\end{aligned}$$

标准正态分布 II

正态分布的积分不是初等函数，故对于服从一般正态分布的随机变量X

$$
\begin{aligned}
P(a < X < b) &= \frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{\frac{-(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}\left(\frac{x-\mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{t^2}{2}}\,\mathrm{d}t \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}
$$

上面的$\Phi\left(\frac{b-\mu}{\sigma}\right)$和$\Phi\left(\frac{a-\mu}{\sigma}\right)$都可以查表计算。

标准正态分布 III

对于$X \sim N(\mu, \sigma^2)$，查表计算有

$$P(|X-\mu| \le 3\sigma) = 0.9973$$

命题2.3.1

设 $X \sim N(\mu, \sigma^2)$，令$Y = \frac{X-\mu}{\sigma}$，那么$Y \sim N(0,1)$。

证明稍后给出。

正态分布的应用一 I

设$X \sim N(-1, 2^2)$，求$P(-5 \le X < 1)$, $P(-2 \le X \le 2)$, $P(|X| \ge 1)$。

解：

$$P(-5 \le X < 1) = \Phi\left(\frac{1-(-1)}{2}\right) - \Phi\left(\frac{-5-(-1)}{2}\right)$$
$$= \Phi(1) - \Phi(-2)$$

$$P(-2 \le X \le 2) = \Phi\left(\frac{2-(-1)}{2}\right) - \Phi\left(\frac{-2-(-1)}{2}\right)$$
$$= \Phi(\frac{3}{2}) - \Phi(-\frac{1}{2})$$

$$P(|X| \ge 1) = 1 - P(|X| \le 1)$$
$$= 1 - \left(\Phi\left(\frac{1-(-1)}{2}\right) - \Phi\left(\frac{-1-(-1)}{2}\right)\right)$$

正态分布的应用二 I

假设一次测量误差$X \sim N(0, 10^2)$,现在独立重复的进行100次测量，求误差的绝对值超过19.6的次数不小于3的概率。

解：先求出一次测量误差大于19.6的概率。

$$\begin{aligned} P(|X| > 19.6) &= 1 - P(|X| \le 19.6) \\ &= 1 - P(\left|\frac{X-0}{10}\right| \le 1.96) \\ &= 1 - [\Phi(1.96) - \Phi(-1.96)] \\ &= 2 - 2\Phi(1.96) \\ &= 0.05 \end{aligned}$$

正态分布的应用二 II

由于每次测量都是独立的，所以设Y为100次测量中误差大于19.6的次数，那么$Y \sim B(100, 0.05)$，

$$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - \sum_{k=0}^{2} b(k; 100, 0.05) = 0.881737$$

大家也可用泊松分布计算下看看。

正态分布的应用二 I

公共汽车车门的高度是按照男子与车门顶碰头的机会0.01以下来设计的。设男子的身高$X \sim N(170, 6^2)$。确定车门高度。

解：设车门高度为h，按照题意，

$$P(X > h) < 0.01,$$

所以

$$\Phi(\frac{h-170}{6}) = P(X \leq h) = 1 - P(X > h) > 0.99$$

查一下正态分布表，近似的

$$\frac{h-170}{6} \geq 2.33$$

正态分布的应用二 II

即

$$h \geq 183.98$$

如果取整数的话，$h = 184\text{cm}$.

正态分布的应用三 I

设某只股票的初始价格为$S_0 = 40$元，预期收益率μ为每年16%，波动率为每年20%.在Black-Scholes模型下，股票在每个时刻t的价格S_t为随机变量

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right)$$

其中$B_t \sim N(0,t)$，估计六个月后这只股票的价格范围（允许出错的概率为5%）。

解：对于每个给定的时间t,随机变量S_t为另一个随机变量B_t的函数，易知的是在$t = 0.5$（半年）的时候

$$\begin{aligned}\ln(S_{0.5}) &= \ln(40) + [0.5(0.16 - \frac{0.2^2}{2}) + 0.2B_{0.5}]\\ &= 3.758879 + 0.2B_{0.5}\end{aligned}$$

正态分布的应用三 II

那么

$$\frac{ln(S_{0.5}) - 3.758879}{0.2} = B_{0.5} \sim N(0, 0.5)$$

亦即

$$X = \frac{ln(S_{0.5}) - 3.758879}{0.2\sqrt{0.5}} \sim N(0, 1)$$

0.5年后去估计此股票的价格范围，若出错率为0.05，即

$$P(|X| \leq y) = 2\Phi(y) - 1 = 0.95,\ \Phi(y) = 0.975$$

查分布表知

$$y = 1.96$$

即

$$\left|\frac{ln(S_{0.5}) - 3.758879}{0.2\sqrt{0.5}}\right| \leq 1.96$$

正态分布的应用三 III

得

$$S_{0.5} \in [32.51, 56.60]$$

练习

1. 设$X \sim N(3,2^2)$，求$P(2 < X \leq 5)$，$P(|X| > 2)$，确定c使得$P(X < c) = P(x > c)$。
2. 某地区成年男子的体重X服从正态分布$N(\mu,\sigma^2)$，若已知$P(X \leq 70) = 0.5$，$P(X \leq 60) = 0.25$。(1)求μ，σ,(2)若在这个地区随机抽出5名成年男子，问其中至少有两个人的体重超过65的概率是多少？

一维随机变量函数的分布 I

问题：已知X的分布，求$f(X)$的分布

离散型随机变量

列举$f(X)$在离散点上的相应取值，然后将$f(X)$取相同值的概率相加。

例：设X的分布列为

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.15 & 0.2 & 0.2 & 0.2 & 0.25 \end{pmatrix}$$

求$Y = X^2$，$Z = 2X - 1$，$W = |X| + 1$的分布列。

解：直接求Y，Z，W的分布列

$$Y \sim \begin{pmatrix} 4 & 1 & 0 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$$

一维随机变量函数的分布 II

$$Z \sim \begin{pmatrix} -5 & -3 & -1 & 1 & 3 \\ 0.15 & 0.2 & 0.2 & 0.2 & 0.25 \end{pmatrix}$$

$$W \sim \begin{pmatrix} 3 & 2 & 1 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$$

连续型随机变量

先求$f(X)$的分布函数，再对分布函数求导数得到$f(X)$的分布密度函数

例：$X \sim N(0,1)$，求$Y = X^2$的分布密度函数。
第一步：先求Y的分布函数$F_Y(y)$。显然的，对$y < 0$有

$$P(Y \leq y) = 0$$

对$y \geq 0$，有

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X < \sqrt{y})$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

一维随机变量函数的分布 II

第二步：对$F_Y(y)$关于y求导数，得到

$$
\begin{aligned}
f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} &= \frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}\left(\frac{1}{2\sqrt{y}}\right) - \frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}\left(-\frac{1}{2\sqrt{y}}\right) \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}
\end{aligned}
$$

即

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

$aX+b$的分布 I

设X的分布函数为f_X，求$Y=aX+b$的分布密度函数，其中$a\neq 0$

记Y的分布函数为F_Y，分布密度函数为f_Y，那么

1. 当$a>0$时
   第一步：求Y的分布函数

$$F_Y(y)=P(Y\leq y)=P(aX+b\leq y)=P\left(X\leq\frac{y-b}{a}\right)$$
$$=\int_{-\infty}^{\frac{y-b}{a}} f_X(x)\,\mathrm{d}x$$

第二步：对$F_Y(y)$关于y求导，有

$$f_Y(y)=\frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

$aX+b$的分布 II

2 当$a<0$时，有

$$F_Y(y) = P(Y \le y) = P(aX+b \le y) = P(X \ge \frac{y-b}{a})$$
$$= \int_{\frac{y-b}{a}}^{+\infty} f_X(x)\,\mathrm{d}x$$

关于y求导，有

$$f_Y(y) = \left(-\frac{1}{a}\right) f_X\left(\frac{y-b}{a}\right)$$

$aX+b$的分布 III

综上所述

$$f_Y(y)=\frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$$

作为特列，若$X\sim N(\mu,\sigma^2)$，且$Y=\frac{X-\mu}{\sigma}$,那么由上式

$$f_Y(y)=\frac{1}{\frac{1}{\sigma}}f_X\left(\frac{y+\frac{\mu}{\sigma}}{\frac{1}{\sigma}}\right)=\sigma\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\sigma y+\mu-\mu)^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$$

即$Y\sim N(0,1)$，与命题2.3.1吻合。

例子

1. 请大家验证：$X \sim N(\mu, \sigma^2)$，那么若$a \neq 0$，则$aX + b \sim N(a\mu + b, a^2\sigma^2)$。并根据这个结论，说明命题2.3.1
2. 设随机变量X的密度函数是

$$f_X(x) = \begin{cases} \frac{2x}{\pi^2}, & 0 < x < \pi, \\ 0, & \text{其他} \end{cases}$$

求$Y = sin(X)$的密度函数。