概率统计2——随机变量

随机变量的概念

定义1： 设Ω是某随机试验的样本空间.函数$X = h(\omega)$将样本空间中的样本点映射为固定的数字，则称函数X为该随机试验的随机变量.可以发现，随机变量的本质就是将随机试验的样本点映射为一个实数的过程。

根据随机变量的取值的数量状态的不同，可以将随机变量分为离散型和连续性两种。离散型随机变量的取值是有限可列或无限可列个的；而连续型随机变量的取值是无限多且绵密而不可列的。比如扔色子，摸球，观察某一段时间内某个路口的人流量等属于典型的离散型，而电灯泡的寿命,某地区某天的的降水量等则属于连续型。

离散型随机变量及其分布

定义2： 设X为某离散型随机变量，其中X的取值为$x_k, k \in \{1, 2, 3 \cdots\}$,则称与$x_k$对应的概率集合$p_k = P(x_k), k \in \{1, 2, 3 \cdots\}$为该随机变量的概率分布或者分布律.

例题1： 设有5个黑球和3个白球，每次以不放回的方式抽取1个，直到抽取到黑球时停止。现设X为在该试验中抽取到白球的个数，问$P(1 < X \leq 3)$

连续型随机变量及其分布

由于连续性随机变量的取值分布是无限且连续的，求出其中单个可能取值的概率是极限为0的无穷小量,因此分布律的方法对连续型随机变量不再适用，而应该使用**概率密度函数**加以描述。

定义3： 若存在一个可积函数$f(x) \geq 0$，对于任意实数$a \leq b$,存在$P\{a \leq X \leq b\} = \int_a^b f(x)dx$,则称$f(x)$为随机变量$X$的概率密度函数.

概率密度函数在某一点的函数值$f(x)$并不表明x点发生的概率大小$P(x)$(实际上在连续型随机变量中，单点发生的概率为0)，而是表明随机变量的取值分布在x点附近的密度大小。

概率密度函数的本质特征：当$a \to -\infty, b \to +\infty$时，有

$$P(-\infty \leq X \leq +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1$$

这表明，连续型随机变量的概率密度函数在整个实数域上的定积分必定为1

例题2：设

$$f(x) = \begin{cases} kx + 1 & 0 \leq x \leq 2 \\ 0 & \text{other} \end{cases}$$

求k的值.

分布函数的概念

定义4：随机变量X的取值小于等于某个自变量x时候的概率值函数即称为该随即变量的分布函数。即$F(x) = P(X \leq x)$,其中$x \in (-\infty, +\infty), F(x) \in [0, 1]$

分布函数的性质：

1. $0 \leq F(x) \leq 1$

2. $F(x)$是个非递减函数，这是因为随着x的增大，样本点增多，则概率必然也随之增大或保持不变
3. $\lim_{x\to+\infty} F(x) = F(+\infty) = 1$
4. $\lim_{x\to-\infty} F(x) = F(-\infty) = 0$
5. 对于离散型随机变量，其分布函数$F(x)$是右连续的，对于连续型随机变量，其分布函数$F(x)$是连续的

例题3: 设

$$F(x) = \begin{cases} a - e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases} \quad \lambda > 0$$

求a的值.

离散型随机变量的分布函数

对于离散型随机变量，设其所有可能取值由小到大排序为$x_1, x_2, \cdots, x_i, \cdots$,则

$$F(x) = P(X \le x) = \sum_{i=1}^{k} P(x_i)$$

其中$x_k \le x < k_{k+1}$

例题4: 设现有某离散型随机变量X的分布律如下：

$$\begin{array}{c|ccc} X & -1 & 2 & 3 \\ \hline P & \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{array}$$

求其分布函数$F(x)$的表达式.

连续型随机变量的分布函数

对于连续型随机变量而言，有$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$.

容易看出，$F'(x) = f(x)$，即连续性随机变量分布函数的导数是其概率密度函数。

例题5: 设

$$F(x) = \begin{cases} 0 & x < 0 \\ Ax^2 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

求：

1. A的值
2. 求其概率密度函数$f(x)$
3. 求$P(0.3 < X < 0.7)$

常见的重要分布

离散型随机变量的分布

0-1分布

0-1分布是一种离散型分布，且取值只能有两个，其分布律如下：

$$\begin{array}{c|cc} X & 1 & 0 \\ \hline P & p & 1-p \end{array}$$

其计算公式为：

$$P(X=k)=p^k(1-p)^{1-k}$$

其中$k=0,1$

几何分布

几何分布用以描述某事件发生的概率为p，对该事件进行独立重复，则该事件直至第k次发生，而前$k-1$次未发生的概率现象。

其计算公式公式为：

$$P(X=k)=(1-p)^{k-1}p$$

记作$X\sim G(p)$

二项分布

二项分布用以描述在进行了n次重复独立试验后事件发生k次的概率。

其计算公式公式是：

$$P(X=k)=C_n^kp^k(1-p)^{n-k},k=0,1,2,\ldots n$$

记作$X\sim B(n,p)$

$0-1$分布是二项分布在$n=1$时的特例

例题6: 设现有一套报警系统，每台报警器在发生危险时的报警概率为0.8，并且每台报警器独立工作。现要求在发生危险时该报警系统以大于0.99的可能报警，那么至少需要几台报警器？

课后作业1： 设有一批设备正在同时运行，每台设备发生故障的概率是0.01，那么问：

1. 1人运维20个设备，不能及时维修的概率是多大？
2. 3人运维80个设备，不能及时维修的概率是多大？

泊松分布

泊松分布主要用于描述**某段连续时间内**发生某个事件k次的概念。比如需要求晚上$6-7$的时间段内有50辆车经过某路口的概率值，便可以使用泊松分布。

其计算公式为：

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,3..$$

记作 $X\sim P(\lambda)$, 其中$\lambda>0$是一个给定的正数,其含义为在所需要求泊松分布概率的时间段内所预测事件的平均发生次数。

当二项分布的n很大而p很小时可以用$\lambda = np$的泊松分布进行近似，也就是说，泊松分布其本质上就是二项分布在上述条件下的极限形式，证明如下：

$$\lim_{n\to\infty} C_n^k(\frac{\lambda}{n})^k(1-(\frac{\lambda}{n}))^{n-k} = \lim_{n\to\infty} \frac{n\times(n-1)\times\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}(1-(\frac{\lambda}{n}))^{n-k} = \frac{\lambda^k}{k!}\lim_{n\to\infty}\frac{n\times(n-1)\times\cdots(n-k+1)}{n^k}(1-\frac{\lambda}{n})^{-k}(1-\frac{\lambda}{n})^n$$

其中

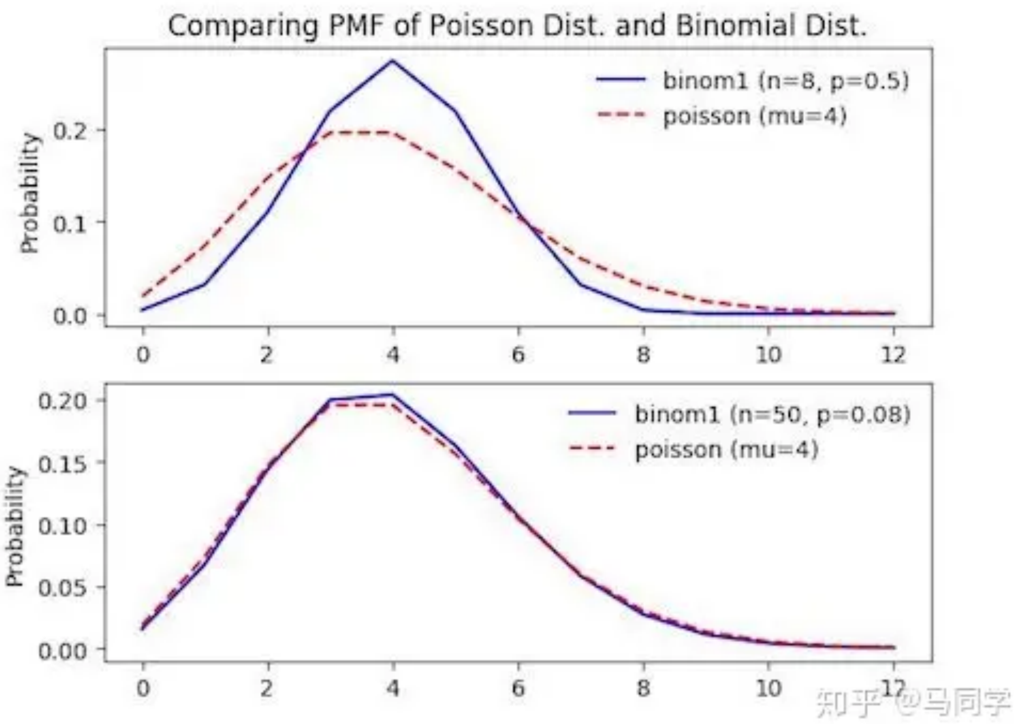
$$\lim_{n\to\infty}\frac{n\times(n-1)\times\cdots(n-k+1)}{n^k}(1-\frac{\lambda}{n})^{-k} = 1$$

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

于是：

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

实际来说：当$n \geq 100, np \leq 10$时，用泊松分布近似近似二项分布效果是较为理想的。



泊松分布更详细推导过程，参考：**泊松分布的现实意义是什么，为什么现实生活多数服从于泊松分布?**

泊松分布的计算可以使用既定的表格进行查询，参考：**泊松分布函数表**

例题7: 某电话台一天之内用户呼叫的次数服从$X \sim P(3)$,问明天该电话台的呼叫次数不超过5次的概率是多少?

例题8: 设某银行证券部有1000个账户，每个账户中客户的投资额是10万元，每个客户在一年内提现2万元的概率是0.006，问：要保证以0.95可能性以上的资金安全周转，则每年存留在证券部的现金储备至少是多少?

解：由于客户之间不存关联性，因此可以认为每个账户的提现是一种独立重复行为。设随机变量X为每年在证券部提现的客户数，因此$X \sim B(1000, 0.006), \lambda = np = 6$。由于X为提现用户数，则2X为每年提现的总金额，设x为每年预留的储备金。故有$P(2X \leq x) \geq 0.95$.由于$\lambda = np = 6$,因此该二项分布可以近似使用泊松分布进行计算，则$P(X \leq \frac{x}{2}) = \sum_{k=0}^{\frac{x}{2}} \frac{6^k}{k!}e^{-6} \geq 0.95$, 查表可知，$x \geq 20$

超几何分布

定义5： 有N个元素，其中N_1个属于第一类，N_2属于第二类，从中取n个元素.设X为 n 个元素中的第一类的个数。其计算公式为：

$$P(X=k) = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}, k=0,1,2,\ldots min\{n, N_1\}$$

记作$X \sim H(n, N_1, N)$

> 超几何分布是一种不放回抽样，但是当$n \ll N$时，$N \approx N-n$,此时就相当于是做了n次概率为$\frac{N_1}{N}$的放回抽样试验，也就是独立重复试验，满足二项分布的条件，即：$n \ll N$时，有：

$$P(X=k)=\frac{C_{N_1}^k C_{N-N_1}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}$$

课后作业2： 设有10000粒种子，发芽率是0.99，现取200粒种子，求至多一粒种子不发芽的概率（结果请用合适的分布近似）

连续型随机变量的分布

均匀分布

定义6：若在某段特定区间上，随机变量的取值的分布是均匀的，则称该随机变量在该区间上服从均匀分布。其概率密度函数如下：

$$f(x)=\begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & other \end{cases}$$

记作$X \sim U[a, b]$

其分布函数如下：

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

例题9: 假设公共汽车从七点开始每隔 15 分钟发车一趟，某人从七点至七点半间到车站的时间服从均匀分布，问：

1. 等车不超过五分钟的概率
2. 等车超过十分钟的概率

正态分布

定义7：大量的自然和社会现象，诸如身高体重都满足“中间多，两头少”的特征，即**随机变量取值的分布大量地集中在其均值附近，越远离均值则分布越稀疏**。正态分布即是用来描述这种普遍存在的现象的，因此，正态分布也是最为重要的概率分布，其概率密度函数为：

$$\phi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-u)^2}{2\sigma^2}}, x \in (-\infty, +\infty)$$

记作$X \sim N(\mu, \sigma^2)$.

其中μ是正态分布的均值或期望，σ是 正态分布的标准差，σ^2是正态分布的方差。

其分布函数为：

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

正态分布的性质：

1. $y=\phi(x)$是以$x=\mu$为对称轴的钟形曲线
2. $x=\mu$时取得最大值$\frac{1}{\sqrt{2\pi}\sigma}$
3. $y=\phi(x)$以 x 轴为水平渐近线

4. σ固定时，μ变化则曲线左右平移
5. μ固定时，σ 变大，最高点下移，曲线趋于平缓;σ 变小，最高点上移，曲线趋于陡峭。

标准正态分布
称$\mu = 0, \sigma = 1$的正态分布为标准正态分布，其概率密度函数为：

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty$$

其分布函数为：

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad -\infty < x < +\infty$$

标准正态分布的性质：

1. $\phi_0(x) = \phi_0(-x)$
2. $\Phi_0(-x) = 1 - \Phi_0(x)$

一般正态分布标准化

一般正态分布需要先转化为标准正态分布进行计算，其推导过程如下：

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-u)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} d\frac{t-\mu}{\sigma} = \Phi_0(\frac{x-\mu}{\sigma})$$

即：

$$\Phi(x) = P(X \le x) = P(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}) \xlongequal{Y = \frac{X-\mu}{\sigma}} P(Y \le \frac{x-\mu}{\sigma}) = \Phi_0(\frac{x-\mu}{\sigma})$$

所以可知：

$X \sim N(\mu, \sigma^2)$时，则$Y = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$

标准正态分布的计算可以通过如下表格进行查询：

标准正态分布的函数分布表

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

例题10：如果$X \sim N(3,3^2)$, 求$P(|X-3|>6)$

课后作业3：在大型考试中，学生的成绩往往是满足正态分布的。教师通常根据学生的分数估计μ与 σ ,然后将分数在$\mu-\sigma$到 μ之间的学生评为C等，问：被评为C等的学生的比例是多大?

课后作业4： 如果$X \sim N(0,\sigma^2)$，计算$Y=|X|$ 的密度.(如果你学过定积分的话就做，否则不用做)

课后作业5： $X \sim N(\mu, \sigma^2)$,计算满足 $P(|X-\mu| \le c)$ 的 c 值.

课后作业6： 如果$X \sim N(\mu, \sigma^2)$，证明$P(|X-\mu| \le 0.675\sigma) = 0.5$

随机变量函数的分布

定义8：已知某随机变量X及其分布$F_X(x)$，现有另一以X为自变量的随机变量函数$Y = h(X)$，则称Y的分布函数$F_Y(x)$为随机变量X满足函数 h 的分布。

离散型随机变量函数的分布

设有离散型随机变量 X，其取值依次为 $x_1, x_2, \cdots$, 另有离散型随机变量$Y = h(X)$,其取值依次为$y_1, y_2, c \ldots$，则显有$P(Y = y_i) = \sum_j P(X = x_j)$,其中$x_j$满足$h(x_j) = y_i$.

例题11: 设X为某离散型随机变量，其分布如下：

X	-1	0	1	2
P	0.2	0.3	0.4	0.1

求$Y = X^2$的分布

连续型随机变量函数的分布

设有连续型随机变量X，设其概率密度函数为$f_X(x)$，其分布函数为$F_X(x)$. 另有连续型随机变量$Y = g(X)$,其分布函数为$F_Y(x)$. 则

$$F_Y(x) = P(Y \le x) = P(g(X) \le x) \xlongequal{g\text{严格单调递增}} P(X \le g^{-1}(x)) = F_X(g^{-1}(x))$$

$$F_Y(x) = P(Y \le x) = P(g(X) \le x) \xlongequal{g\text{严格单调递减}} P(X \ge g^{-1}(x)) = 1 - F_X(g^{-1}(x))$$

例题12：设随机变量的概率密度函数为$f_X(x), Y = 3X + 1$

1. 求Y的分布函数及其概率密度函数
2. 若X满足$0-4$区间上的均匀分布，求Y的概率密度函数以及分布函数

解1:

$$F_Y(x) = P(Y \le x) = P(3X + 1 \le x) = P(X \le \frac{x-1}{3}) = F_X(\frac{x-1}{3})$$

$$f_Y(x) = F_Y(x)' = \frac{1}{3} f_X(\frac{x-1}{3})$$

解2: 由于

$$f_X(x)=\begin{cases}\frac{1}{4} & 0\le x\le 4\\ 0 & other\end{cases}$$

则

$$f_Y(x)=\frac{1}{3}f_X(\frac{x-1}{3})=\begin{cases}\frac{1}{12} & 1\le x\le 13\\ 0 & other\end{cases}$$

由于

$$F_X(x)=\begin{cases}0, & x\le 0\\ \frac{x}{4}, & 0<x<4\\ 1, & 1\le x\end{cases}$$

则

$$F_Y(x)=F_X(\frac{x-1}{3})=\begin{cases}0 & x<1\\ \frac{x}{12} & 1\le x<13\\ 1 & 13\le x\end{cases}$$

我们可以得到如下推论：

若X服从$[a.b]$ 上的均匀分布，则 $Y=kx+c(\neq 0)$ 服从相应区间上的均匀分布,即：

$k>0$时:

$$f_Y(x)=\begin{cases}\frac{1}{kb-ka} & ka+c\le x\le kb+c\\ 0 & other\end{cases}$$

$k<0$时：

$$f_Y(x)=\begin{cases}\frac{1}{ka-kb} & kb+c\le x\le ka+c\\ 0 & other\end{cases}$$

例题13：设$X\sim N(\mu,\sigma^2),Y=aX+b,a\neq 0$,求$f_Y(x)$

解：
$a>0$时：

$$F_Y(x)=P(Y\le x)=P(aX+b\le x)=P(X\le\frac{x-b}{a})=F_X(\frac{x-b}{a})=\Phi(\frac{x-b}{a})$$

于是

$$f_Y(x)=F_Y(x)'=\frac{1}{a}\phi(\frac{x-b}{a})=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\frac{x-b}{a}-\mu)^2}{2\sigma^2}}\frac{1}{a}=\frac{1}{\sqrt{2\pi}a\sigma}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$$

$a<0$时

$$f_Y(x)=\frac{1}{\sqrt{2\pi}-a\sigma}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$$

因此综上：

$$f_Y(x)=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}},\text{即：}Y\sim N(a\mu+b,a^2\sigma^2)$$

根据上例，**当$Y=\frac{x-\mu}{\sigma}$时，显然有$Y\sim N(0,1)$,这正是正态分布转化为标准正态分布的数学原理**

课后作业7：证明若随机变量X的概率密度函数为$f_X(x),Y=kX+b(k\neq 0)$,则 $f_Y(x)=\frac{1}{|k|}f_X(\frac{x-b}{k})$

例题14： $X \sim N(0,1), Y = X^2$,求$f_Y(x)$

解：

$x < 0$时：

$$F_Y(x) = P(Y \leq x) = P(X^2 \leq x) = 0$$

此时不存在满足要求的随机变量值，因此分布函数值为0

$x \geq 0$ 时：

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= \Phi_0(\sqrt{x}) - \Phi_0(-\sqrt{x}) \\ &= \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

$$f_Y(x) = F_Y(x)' \xlongequal{\text{变限积分求导}} \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})} x^{-\frac{1}{2}}$$

因此,

$$f_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})} x^{-\frac{1}{2}} & 0 \leq x \\ 0 & x < 0 \end{cases}$$

课后作业8：设随机变量X,其概率密度函数为：

$$f_X(x) = \begin{cases} \frac{1}{x+1} & 0 < x < e-1 \\ 0 & other \end{cases}$$

$Y = \sqrt{X}$,求$f_Y(x)$