随机变量及其分布

by zijeff

2.1 随机变量

- 随机变量的引入

  我们高中就接触过**古典概型**中的**样本空间**S，我们需要找出样本空间所有的**样本点**e，进而计算一系列问题的概率。在实际问题中，样本点e既有可能是简单的数字，也有可能是文字上的描述。为了更好的进行数学分析，同时在**具体概率分析**的时候我们也并不关心结果的**具体文字描述**，所以基于以上两点原因，我们希望找到一个**对应关系**。这个**对应关系**使得随机试验的每个结果都被赋予一个数值， 在样本空间S和实数值建立一种对应关系（类似于函数中从定义域到值域的一种**映射**）。综上所述，我们需要引入某个东西来帮助我们分析，这个引入的东西就是**随机变量**了。

- 随机变量的概念

  定义：设随机试验的样本空间$S=\{e\}$，$X=X(e)$是定义在样本空间S上的单值实值函数。**称** $X=X(e)$为随机变量。

  > **Tip**
  >
  > - 随机变量通常用**大写字母**表示或**希腊字母**等表示
  > - 随机变量所取到的值通常用**小写字母**表示

  如定义所言，随机变量是**单值实值函数**，那它与我们之前学的函数又存在什么**区别**呢?

  > **Important**
  >
  > - 普通函数是定义在实数轴上的，而随机变量是定义在样本空间上的（样本空间的元素不一定是实数）。
  > - 随机变量随着试验的结果不同而取不同的值，由于试验的各个结果的出现具有一定的概率，因此随机变量的取值也有一定的概率规律。

2.2 离散型随机变量及其分布律

- 离散型随机变量的分布律

  设离散型随机变量X所有的可能取值为:

  $$x_k\ \ (k=1,2,\cdots)$$

  X取取各个可能值的概率，即事件$\{X=x_k\}$的概率，为:

  $$P(X=x_k)=p_k\ \ (k=1,2,\cdots)$$

  称此为离散型随机变量X的分布律。

  > **Note**
  >
  > 显然由**概率的定义**，易得如下性质:
  >
  > - $p_k\geq 0$，$k=1,2,\cdots$
  > - $\sum_{k=1}^{\infty}p_k=1$

  同时，分布律我们也可以用**表格**来表示。

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

表格直观地表示了随机变量取各个值的概率的规律。X取各个值各占一些概率，这些概率之和为1。

- **常见离散型随机变量的概率分布**

1. 常数随机变量（退化分布）：最简单的随机变量是常数，它在整个样本空间上仅取一个值。

$$X \equiv c$$
$$P(X \equiv c) = 1$$

2. 0-1分布（两点分布，伯努利分布）：

设随机变量X只能取到**0与1**这两个值，其分布为：

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

则称随机变量X服从以p为参数的0-1分布或者两点分布，**分布律表格**如下：

X	0	1
p_k	$1-p$	p

3. 二项分布（n重伯努利试验）：

设试验E只有两个可能的结果：A和$\overline{A}$。那我们称这种试验为**伯努利试验**，将E独立重复进行n次，则称这一系列的试验为**n重伯努利试验**。

我们设单次伯努利试验中$P(A)=p$，易得$P(\overline{A})=1-p$，其中$p \in (0,1)$，用随机变量X表示n重伯努利试验中事件A的发生次数，显然X的所有可能取值为$\{1,2,\cdots,n\}$。

接下来，我们推导**二项分布的分布律**。不失一般性地，也就是求当$X=k$，$k \in \{1,2,\cdots,n\}$时的概率。由高中学习的组合知识，显然我们只要在n次试验中选取k次事件A发生即可，从而得到如下分布律公式。

$$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}$$

符合上述分布律的分布称为**二项分布**，记作$X \sim b(n,p)$。易证，当$n=1$时，即为两点分布。

因此二项分布的分布律表格也就不难画出了，这里省略，读者可以自画。

4. 泊松分布：

设随机变量X的所有可能取值为$\{1,2,\cdots,n\}$，而取到各个值得概率为：

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

其中$\lambda > 0$是常数，则称X服从以λ为参数的泊松分布，记作$X \sim \pi(\lambda)$。

> **Important**
>
> 泊松分布的引入与其分布律的推导过程：

5. 几何分布：

考虑一个随机试验E，它只有**两个结果**，比如成功和失败，设两个结果的概率分别为p和$1-p$。现在我们将这个随机试验独立重复进行，直到其成功为止。用X表明所需要的**试验次数**，显然$X \in N_+$，同时易得X取每个值的时候的概率为：

$$P(X=k) = p(1-p)^{k-1}, k \in N_+$$

那么，我们称X服从于**几何分布**，其中p为参数，记作$X \sim G(p)$。

> **Note**
>
> **几何分布**具有一个比较重要的性质，即**无记忆性**。

> 若前m次试验均已失败，之后的试验次数X仍为**几何分布**，即如下式子成立：
>
> $$P(X=m+k|X>m)=p(1-p)^{k-1}$$

6. 负二项分布：

   给定正整数r，用X_r表示首次取得r次成功的试验次数。显然，$X_r\in\{r,r+1,r+2,\cdots\}$，并且：

   $$P(X_r=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}=C_{k-1}^{r-1}p^r(1-p)^{k-r}$$

- **二项分布的泊松逼近**

  当试验次数n很大的时候，计算二项分布的精确概率变得十分麻烦。因此，我们需要寻找**近似计算**二项分布的方法。这里我们先介绍**泊松近似**，后面我们将介绍二项分布的正态近似。

  泊松定理：设$\lambda>0$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k，有如下式子成立

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

  该定理的**证明如下**：由已知条件可知，$p_n=\frac{\lambda}{n}$，将其带回等式左边得到

  $$\binom{n}{k}(\frac{\lambda}{n})^k[1-(\frac{\lambda}{n})]^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$$

  继续变形后得到

  $$\frac{\lambda^k}{k!}[1\cdot(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})](1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}$$

  值得注意的是，我们始终关心的是上述式子在两个特定整数λ,k下随着$n\to\infty$时的**极限值**，又因为上述式子的项数是**有限可列**的，所以我们可以大胆地使用**极限的运算法则**。

  直接将上面这一长串式子，拆分成几个部分：

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

  $$\lim_{n\to\infty}[1\cdot(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})]=1$$

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

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

  综上所诉，将各个极限值带回**待证等式**，即可得到原式成立，证毕。

2.3 随机变量的分布函数

- **分布函数的概念**

  1. 概念的引入：对于随机变量X，我们不仅要知道X取哪些值，还要知道X取这些值的概率；而且在实际解决问题中，更重要的是想知道X在**任意有限区间**(a,b)内取值的概率。
  2. 分布函数的定义：设X是一个随机变量，x是任意实数，函数

     $$F(x)=P(X\le x)\ \ (-\infty<x<+\infty)$$

     函数$F(x)$称为X的分布函数。

> **Tip**

> 现在我们可以思考一个问题，为什么我们要定义分布函数$F(x) = P(X \le x)$，这背后有什么说法吗?
>
> 首先，我们知道分布函数的提出与定义是为了求解X在**任意有限区间**(a, b)内取值的概率。所以，现在我们不妨考虑求解随机变量X落在区间$(x_1, x_2]$内的概率。显然，求解过程如下:
>
> $$P(x_1 < X \le x_2) = P(X \le x_2) - P(X \le x_1)$$
>
> 所以我们就可以令$P(X \le x_2) = F(x_2), P(X \le x_1) = F(x_1)$，因此也就给出了分布函数的定义。

- 分布函数的性质
  1. $F(x)$是一个**不减函数**。
  2. $0 \le F(x) \le 1$，且$F(x)$的极限满足如下等式:

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

  3. $F(x + 0) = F(x)$，即$F(x)$是右连续的。

> **Important**
>
> 由分布函数得一些计算上的重要公式:
>
> 1. $P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)$
> 2. $P(X > x) = P(X < \infty) - P(X \le x) = 1 - F(x)$
> 3. $P(X = x) = F(x) - F(x - 0)$
> 4. $P(X < x) = F(x - 0)$
> 5. $P(a \le X \le b) = F(b) - F(a - 0)$
> 6. $P(a \le X < b) = F(b - 0) - F(a - 0)$
> 7. $P(a < X < b) = F(b - 0) - F(a)$

一般地，设离散型随机变量X的分布律:

$$P(X = x_k) = p_k \ \ (k = 1, 2, \cdots)$$

由概率的**可列可加性**可得X的分布函数为:

$$F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_k)$$

即，化简后**分布函数**为:

$$F(x) = \sum_{x_i \le x} p_k$$

这里和式是对所有满足$x_k \le x$的k求和的。所以**分布函数**$F(x)$在$x = x_k \ (k = 1, 2, \cdots)$处有跳跃，其跳跃值为$p_k$。

> **Note**
>
> **不同**的随机变量,它们的分布函数可能**相同**。

2.4 连续型随机变量及其概率密度

连续型随机变量X所有可能取值充满一个区间，对这种类型的随机变量，不能像离散型随机变量那样，以指定它取每个值概率的方式，去给出其概率分布， 而是通过给出概率密度函数的方式进行计算。

- 概率密度的概念与性质

1. 概率密度函数的定义：如果对于随机变量X的分布函数$F(x)$，存在**非负可积函数**$f(x)$，使对于任意实数x有下式成立：

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

   则称X为连续型随机变量，其中函数$f(x)$称为X的**概率密度函数**，简称**概率密度**。

   > ⓘ Note
   >
   > 连续型随机变量的分布函数是**连续函数**。

2. 概率密度函数的性质：
   - $f(x) \geq 0$
   - $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$，也就意味着**概率密度函数**与坐标横轴围成的**面积为定值1**
   - 对于任意的实数$x_1, x_2, (x_1 < x_2)$，$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)\mathrm{d}x$
   - 若$f(x)$在点x处连续，则有$F'(x) = f(x)$成立

   > ⓘ Note
   >
   > 第一条和第二条性质是判定一个函数$f(x)$是否为某个**连续型随机变量**X的概率密度函数的**充要条件**。

3. 利用上述性质得以下计算公式：
   - $P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)\mathrm{d}x$
   - $P(X > a) = 1 - P(X \leq a) = 1 - F(a) = \int_{-\infty}^{\infty} f(x)\mathrm{d}x - \int_{-\infty}^{a} f(x)\mathrm{d}x = \int_{a}^{\infty} f(x)\mathrm{d}x$

> Important
>
> **连续型随机变量**取任一指定实数值a的概率均为0。即$P(X = a) = 0$
>
> 证明如下：
>
> 从随机变量本身性质出发，我们可以得到如下不等关系：
>
> $$0 \leq P(X = a) \leq P(a - \delta < X \leq a) = F(a) - F(a - \delta)$$
>
> 当$\delta \to 0^+$时，得到$P(X = a) = 0$，因此我们可以得到如下两条重要性质：
>
> - 由$P(A) = 0$，我们不能推出$A = \varnothing$
> - 对于连续型随机变量而言，$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$
>
> 也就是说：**连续型随机变量取值落在某区间的概率与端点无关**。

- **常见连续型随机变量及其概率分布**
  1. 均匀分布：若连续型随机变量X具有概率密度

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

     则称X在(a, b)上服从均匀分布，记为$X \sim U(a, b)$

     均匀分布的意义在于：在区间(a, b)上服从均匀分布的随机变量X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。

     对概率密度函数积分后可以得到**分布函数**，如下：

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

  2. 指数分布：若连续型随机变量X具有概率密度

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

其中，$\theta>0$是常数，则称X服从参数为θ的指数分布。易知$f(x)\geq 0$，且$\int_{-\infty}^{\infty}f(x)\mathrm{d}x=1$

对概率密度函数积分后可以得到**分布函数**，如下：

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

> **Important**
>
> 指数分布也具有**无记忆性**的重要性质。
>
> 也就是说，对于任意的$s,t>0$，有下式成立：
>
> $$P(X>s+t\mid X>s)=P(X>t)$$

3. 正态分布：若连续型随机变量X具有概率密度

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

其中$\mu,\sigma(\sigma>0)$为常数，则称X服从参数为μ,σ的正态分布或高斯分布，记为$X\sim N(\mu,\sigma^2)$。

正态分布具有如下重要性质：

- 曲线关于$\mu=x$对称，这表明对于任意的$h>0$，有 $P(\mu-h<X\leq\mu)=P(\mu<X\leq\mu+h)$
- 当$x=\mu$时取到**最（极）大值**，$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$
- 在$x=\mu\pm\sigma$处曲线有拐点
- 曲线以x轴为渐近线
- μ称为正态分布概率密度曲线的位置参数，意味着曲线的位置将完全由参数μ所决定。
- 改变σ时，曲线对称轴不改变，而形状在改变。σ越小，图形越高越瘦，σ越大，图形越矮越胖。

2.5 随机变量的函数分布

从实际出发，生活中存在各式各样的随机变量，而不同的随机变量之间往往不是独立存在的，之间存在着某种关系。确切的说，某些随机变量之间存在函数关系，我们希望通过已知的随机变量X和函数关系，从而快速得到随机变量Y的分布。

- **离散型随机变量的函数的分布**

  如果X是离散型随机变量，其函数$Y=g(X)$也是离散型随机变量，若X的分布律为：

  $$P(X=x_k)=p_k$$

  则其函数$Y=g(X)$的分布律为：

  $$P(Y=g(x_k))=p_k$$

  值得注意的是，若$g(x_k)$中有值相同的，应将相应的p_k合并。

- **连续型随机变量的函数的分布**

  分布函数微分法：这是求$Y=g(X)$的分布密度的一种方法。具体步骤为：先用定义求Y得分布函数$F_Y(y)$，再求导数得Y的**分布密度**$f_Y(y)$。

  $$F_Y(y)=P(Y\leq y)=P(g(X)\leq y)=\int_S f_X(x)\mathrm{d}x$$

其中积分区域S是满足$g(x) \leq y$的所有取值，即为该不等式的解集$S = \{x \mid g(x) \leq y\}$，最后就可以求微分得到分布密度了。

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y}$$

除开这种方法之外，还有一种方法可以求解，我们先引入一条定理。

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

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}| & \alpha < x < \beta \\ 0 & \text{其他} \end{cases}$$

其中$\alpha = \min\{g(-\infty), g(\infty)\}, \beta = \max\{g(-\infty), g(\infty)\}$

如果$f(x)$在有限区间$[a, b]$以外等于零则只需假设在上恒有$g'(x) > 0$（或者$g'(x) < 0$）。此时，边界条件变为：

$$\alpha = \min\{g(a), g(b)\}, \beta = \max\{g(a), g(b)\}$$

> **Tip**
>
> 这个定理通俗来讲就是：若$g(x)$为**单调函数**，就可以使用这个定理快速得到概率密度。