

第四章随机变量的概念

右武卫大将军

2019 年 7 月 3 日

1 引言

随机变量是赋予实验的每一个结果 ξ 的一个数, 记为 $x(\xi)$ 。这个数可以是游戏中的收益, 随机电源中的电压, 一个随机零件的价格或者随机实验中任何我们感兴趣的量。一个随机变量实质上是定义在实验结果所构成的集合 S 上的一个函数。

1.1 随机变量

给定一个实验, 实验的空间为 S , S 的子集构成的域称作事件, 并且赋予这些事件以概率。对实验中的每一个结果 ξ , 指定一个数 $x(\xi)$ 。于是建立了一个定义在集合 S 上的函数 x , 它的值域是一个数集。若函数 x 满足某些不太苛刻的条件, 那么就称这个函数为随机变量。

定义 1.1 随机变量 x 对实验中每一个结果 ξ 指定一个数 $x(\xi)$ 的函数。这一函数必须满足下面两个条件:

- 对于每一个 x , 集合 $\{x \leq x\}$ 是一个事件
- 事件 $\{x = \infty\}$ 和事件 $\{x = -\infty\}$ 的概率等于零

第二个条件表明, 虽然对一些结果允许 x 取 $+\infty$ 或者 $-\infty$, 但要求这些结果所构成集合具有零概率。

2 分布函数和密度函数

在集合 S 中, 组成事件 $\{X \leq x\}$ 的元素随 x 取值不同而变化。因此, 事件 $\{X \leq x\}$ 的概率 $P\{X \leq x\}$ 是依赖于 x 的一个数。这个数表示为 $F_X(x)$ 并称为它是随机变量 X 的累积分布函数。

定义 2.1 随机变量 X 的分布函数

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

是定义在从 $-\infty$ 到 ∞ 上的函数。

定义 2.2 分位点 一个随机变量 X 的 u 分位点是满足

$$u = P(X \leq x_u) = F(x_u)$$

的最小实数 x_u 。

因此 x_u 可以看作是函数 $u = F_X(x)$ 的逆变换。该函数的定义域是 $[0, 1]$ ，而值域是 R 。

2.0.1 分布函数的性质

在下面的定义中，表示式 $F(x^+)$ 和 $F(x^-)$ 分别表示函数 $F(x)$ 在 x 点的右极限和左极限，即

$$F(x^+) = \lim_{\eta > 0, \eta \rightarrow 0} F(x + \eta), \quad F(x^-) = \lim_{\eta > 0, \eta \rightarrow 0} F(x - \eta)$$

分布函数具有以下性质

- $F(+\infty) = 1 \quad F(-\infty) = 0$
- 它是 x 的非降函数
- $P(X > x) = 1 - F_X(x)$
- 函数 $F(x)$ 是右连续的，即 $F(x^+) = F(x)$
- $P\{x_1 < x \leq x_2\} = F(x_2) - F(x_1)$
- $P(X = x) = F_X(x) - F_X(x^-)$

2.1 连续型，离散型和混合型随机变量

如果随机变量 X 的分布函数 $F_X(x)$ 是连续的，则称 X 为连续型随机变量。在这种情况下，分布函数是左连续的，这意味着 $P(X = x) = 0$ 。

如果分布函数是仅有有限多个跳跃型间断点的阶梯型函数， X 被称作离散型随机变量。

2.2 概率密度函数

一个随机变量的分布函数的导数称为该随机变量的概率密度函数，记为 $f_X(x)$ ，即

$$f_X(x) := \frac{dF_X(x)}{dx}$$

从分布函数的单调非减性，概率密度函数满足：

$$f_X(x) \geq 0, \quad \forall x \in R$$

3 常用随机变量

定义 3.1 存在性定理

如果某一个函数 $f(x)$ 满足以下条件:

- 该函数在定义域 \mathbb{R} 上是非负的
- 该函数在整个定义域上的积分为 1
- 该函数的积分上限函数 $\int_{-\infty}^x f(t)dt$ 是右连续的

那么就一定可以构造一个实验和一个随机变量 X , 使得它的概率密度函数为 $f(x)$ 。

下面简要讨论一些常用的概率密度函数。

3.1 连续型随机变量

正态分布

正态分布或者高斯分布是最常用的分布之一。如果一个随机变量的概率密度函数为:

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

该函数对应的是一个钟形曲线, 该曲线是关于均值参数 μ 是对称的, 而方差参数 σ 控制钟形曲线宽度。高斯分布的分布函数的形式为:

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy := G\left(\frac{x-\mu}{\sigma}\right) \quad (2)$$

这里特意引入一个字母来表示标准正态分布的分布函数, 即

$$G(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \quad (3)$$

因为 $f_X(x)$ 仅依赖于两个参数 μ, σ , 故可以将某一个服从正态分布的随机变量 X 记为 $X \sim N(\mu, \sigma^2)$ 。特殊的, 当 $X \sim N(0, 1)$ 时, 该随机变量称为标准正态分布。

正态分布是概率论和统计学中最重要的分布之一。多种自然现象服从正态分布, 最重要的现象是如果总体误差是由无数个微小误差叠加在一起造成的, 那么总体误差服从正态分布。

指数分布

如果一个随机变量 X 的概率密度函数为:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{其他} \end{cases} \quad (4)$$

我们称这个随机变量是服从具有参数 λ 的指数分布。指数分布通常用来刻画某一个事件等待其发生的时间间隔的随机长度, 记随机变量 X 表示第一次事件发生的等待时间, 而 $q(t)$ 代表在 t 时间长度内事件没有出现的概率, 那么按照定义, 即 $P(X > t) = q(t)$ 。如果 t_1 和 t_2 表示两个相邻的互不重叠的区间, 那么根据常识, 这两个时间内是否发生事件是独立的, 即要求 $q(t_1)q(t_2) = q(t_1 + t_2)$ 。该方程实质上只有一个非平凡解:

$$q(t) = e^{-\lambda t}$$

所以该类随机变量的分布函数为：

$$F_X(t) = P(x \leq t) = 1 - q(t) = 1 - e^{-\lambda t} \quad (5)$$

进入就可以推导出上面的概率密度函数了。该随机变量有一个特别的性质称为**无记忆性**：设 $s, t \geq 0$ ，考虑事件 $\{X > t + s\}$ 和 $\{X > s\}$ 。由于事件 $\{X > t + s\} \subset \{X > s\}$ ，所以有

$$P(X > t + s | X > s) = \frac{P(X > t + s)}{P(X > s)} = \frac{e^{-(t+s)}}{e^{-s}} = e^{-t} = P(X > t) \quad (6)$$

当然这种无记忆性显然隐藏在两个时间段独立的前置假设当中。无记忆性简化了很多计算，这也是很多应用采用指数模型的原因。在这种模型下，只要没有坏的设备，就是和新的设备一样。