随机分析基础

邓燕飞

https://idengyf.github.io/

2023 年 5 月

内容提要

1. 随机变量 vs. 随机过程

随机现象

静态模型的随机现象 vs. 动态模型中随机现象

我们常通过随机试验来观察随机事件的统计规律性。

一般地，设 E 为一个试验，若不能事先准确地预测其结果，且在相同条件下可重复进行，则称为**随机试验。**

以 ω 表示随机实验一个可能结果，则称 ω 为随机实验 E 的一个**基本事件。**

全体基本事件的集合 $\Omega = \{\omega_i\}$ 称为基本事件空间或**样本空间。**

【例】E: 一个箱子里有 10 个球，分别标记为 $1, 2, \ldots, 10$。若从箱子里随机取一个球，令基本事件 ω_i 表示球上的数字是 i，则样本空间 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{10}\}$。$\omega_i$ 也可简记为 i，则 $\Omega = \{1, 2, \ldots, 10\}$。

随机变量

在上例中，令 $X=$ 随机抽取的球上的号码，则 X 的取值是 $1,2,3,\ldots,10$ 中的任一个，并且基本事实 ω_i 可以通过 X 来表示，即

$$\omega_i=\{X=i\},\ i=1,2,3,\ldots,10.$$

这是一个古典概率模型，有:

$$P(\omega_i)=P(X=i)\equiv p_i=\frac{1}{10},\ i=1,2,3,\ldots,10.$$

由于 X 这个变量的取值事先无法确定，并且是以一定概率取值，故称之为**随机变量。**

定义：一个**随机变量** X 是由**样本空间**映射至**数字集合**的一个**函数。**此数字集合或这一组数字被称为该变量的取值范围。

随机变量的离散型 vs. 连续型

定义：若随机变量 X 只能取有限个或可数个值，并以各种确定的概率取这些不同的值，则称 X 为**离散型随机变量。**

设 X 的取值为 $x_1, x_2, x_3, \ldots$，相应的概率为 $p_i = P(X = x_i),\ i = 1, 2, \ldots$，显然 $\{p_i\}$ 满足：

(i) $p_i \geq 0,\ i = 1, 2, \ldots;$

(ii) $\sum_{i=1}^{\infty} p_i = 1.$

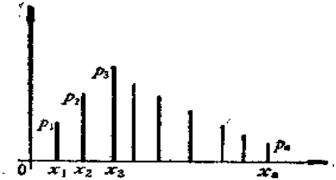

图: 离散型分布的随机变量

【例】E：扔两次同质硬币。基本事件空间（所有可能的结果）或**样本空间**为 { 正面正面，正面反面，反面正面，反面反面 }。定义**随机变量**为“正面的次数”（相当于定义域)，则该变量的取值区间为 $\{0,1,2\}$（相当于值域)。同质硬币意味着假设扔硬币后以相同概率为正面或反面，则该随机变量（“正面的次数”）的**概率质量函数**（或简称**概率函数**）为 $f(x)$（随机变量 X 分别取值 x 的可能性)：

$$f(x) = \begin{cases} 0.25 \\ 0.5 \\ 0.25 \end{cases} \quad \text{对于} \quad x = \begin{cases} 0 & \{\text{反面反面}\} \\ 1 & \{\text{正面反面，反面正面}\} \\ 2 & \{\text{正面正面}\} \end{cases}$$

语义：如果随机变量 X 取值充满某一个区间，并且 X 的值落在任何一个子区间的概率都是确定的，则称之为**连续型随机变量**。

定义：一个随机变量称为连续型，如果存在一个非负可积函数 $f(x)$，使得

$$P(a < X \leq b) = \int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

其表示为 X 落在区间 $(a, b]$ 的概率，F 是稍后要介绍的分布函数。

上式对于 $-\infty < a < b < +\infty$ 都成立，此时 $f(x)$ 称为随机变量 X 的分布密度函数（或简称分布密度、密度函数或密度）。

由于事件 $\{-\infty < X < +\infty\}$ 的概率为 1，所以

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1.$$

【例】张三要去餐厅会见同学。去之前，不知道会在餐厅呆多久。朋友多的话，可能呆得久一些。因此，去餐厅是一个有各种可能结果的概率实验。**随机变量** T 是在餐厅呆得时间长度，各种可能的结果是**样本空间**。随机变量 T 从样本空间映射至设定的一个时间范围比如 $[0,4]$（餐厅最早晚上 6 点开门营业最晚至 10 点）。由于时间是连续的，因此 $T\in[0,4]$ 是一个连续随机变量。**分布函数** $F(t)$ 指呆在餐厅的时长为 t 或短于 t 的概率。比如花 1.5 个小时至 2 个小时在餐厅的概率为 $\int_{1.5}^{2} f(t)\mathrm{d}t$，其中**概率密度函数**：

$$f(t)=\frac{\mathrm{d}F(t)}{\mathrm{d}t}.$$

更一般地，若 X 是连续型随机变量，则其分布函数的一阶导即为概率密度函数 f

$$f(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x}.$$

离散随机变量的分布函数

设 X 为连续随机变量，令

$$F(x) = P(X \le x), \quad -\infty < x < \infty.$$

则称 $F(x)$ 是随机变量 X 的**分布函数**。

若 X 是离散型随机变量，其分布函数为：

$$F(x) = \sum_{i: x_i \le x} p_i.$$

换言之，随机变量 X 的分布函数 F 是从该随机变量的取值范围映射至概率域 $[0,1]$ 的函数。而随机变量 X 为 x 或小于 x 实现值的概率即表示为 $F(x)$。

假设一个离散型随机变量的概率分布为：

$$\begin{pmatrix} x_1 & x_2 & x_3 & \dots \\ p_1 & p_2 & p_3 & \dots \end{pmatrix} \overset{\text{上例}}{=} \begin{pmatrix} 1 & 2 & 3 & \dots & 10 \\ 0.1 & 0.1 & 0.1 & \dots & 0.1 \end{pmatrix}$$

分布函数的几何图示

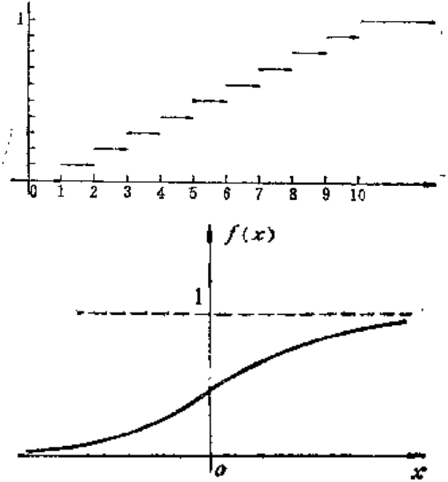

图: 离散- vs. 连续- 随机变量的分布函数

离散随机变量的一些类型

1）离散均匀分布

变量取值范围：$x \in \{a, a+1, \ldots, b-1, b\}$，

概率质量函数：$f(x) = \frac{1}{b-a+1}$。

比如扔骰子，扔出的可能点数为 $\{1, 2, \ldots, 6\}$，扔到任何点数的概率为 $\frac{1}{6}$。

2）离散泊松分布

变量取值范围：$x \in \{0, 1, 2, \ldots\}$，

概率质量函数：$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$。其中 λ 是某个为正值的参数（而在随机过程中，它又被称为到达率)。

比如，晚上看到流星的数量。

连续随机变量的一些类型

1）正态分布和标准正态分布

变量取值范围：$x \in [-\infty, +\infty]$，

概率密度函数：$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$，其中 μ 和 σ 分别是其均值和标准差。对于标准正态分布，$\mu = 0$ 和 $\sigma = 1$，因此概率密度函数为：$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$。

2）指数分布

变量取值范围：$x \in [0, \infty]$，

概率密度函数：$f(x) = \lambda e^{-\lambda x}$。其中 λ 是某个为正值的参数。比如一位刚失业者的失业时长。

高维随机变量

考虑两个随机变量 X_1 和 X_2，假设其概率密度函数分

$$
\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-(\frac{\sigma_{12}}{\sigma_1\sigma_2})^2}} e^{-\frac{1}{2[1-(\frac{\sigma_{12}}{\sigma_1\sigma_2})^2]}[(\frac{x_1-\mu_1}{\sigma_1})^2 - 2\frac{\sigma_{12}}{\sigma_1\sigma_2}\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + (\frac{x_2-\mu_2}{\sigma_2})^2]}, \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(\tilde{x}_1^2 - 2\rho\tilde{x}_1\tilde{x}_2 + \tilde{x}_2^2)}.
\end{aligned}
$$

其中 ρ 是随机变量 X_1 和 X_2 的相关系数，定义为两变量间的协方差 σ_{12} 除以各自的标准差 σ_1 和 σ_2。

每个随机变量的概率密度函数（即边际概率密度）独立于相关系数，为

$$
f(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\tilde{x}_i^2}, \quad i = 1, 2.
$$

随机变量的特征之基础篇

期望：$\mathbb{E}X = \int xf(x)\mathrm{d}x$，

方差：$\mathbb{V}X = \int (x - \mathbb{E}X)^2 f(x)\mathrm{d}x$，

协方差：$\mathbb{C}(X, Y) = \int\int (x - \mathbb{E}X)(y - \mathbb{E}Y)f(x, y)\mathrm{d}x\mathrm{d}y$，

相关系数：$\rho_{XY} = \frac{\mathbb{C}(X,Y)}{\sqrt{\mathbb{V}X\mathbb{V}Y}}$，

独立：$p(X \in A, Y \in B) = P(X \in A)P(Y \in B)$。

基本属性：

$$
\begin{aligned}
\mathbb{E}[a + bX] &= a + b\mathbb{E}X, \\
\mathbb{E}[bX + cY] &= b\mathbb{E}X + c\mathbb{E}Y, \\
\mathbb{E}(XY) &= \mathbb{E}X\mathbb{E}Y + \mathbb{C}(X, Y).
\end{aligned}
$$

$$\mathbb{V}X = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}X^2 - (\mathbb{E}X)^2,$$
$$\mathbb{V}(a + bX) = b^2\mathbb{V}X,$$
$$\mathbb{V}(X + Y) = \mathbb{V}X + \mathbb{V}Y + 2\mathbb{C}(X, Y).$$

$$\mathbb{C}(X, X) = \mathbb{V}X,$$
$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y,$$
$$\mathbb{C}(a + bX, c + dY) = bd\mathbb{C}(X, Y).$$

定理：令 X 为一个随机变量，$f(X)$ 和 $g(X)$ 是两个函数，则

$$f'(X)g'(X) \gtrless 0 \ \forall x \in X \quad \Rightarrow \quad \mathbb{C}(f(X), g(X)) \gtrless 0.$$

证明：略。

【应用】TFP 是一个随机变量（即上述 X），GDP 和 R&D 支出都关于 TFP 正相关，则 GDP 和 R&D 都顺周期。

随机变量的变换

1）正态分布随机变量的线性变换

考虑一个正态分布的随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$，那么，$Y = a + bX$ 的分布为?

期望：$\mathbb{E}(a + bX) = a + b\mathbb{E}X = a + b\mu$，

方差：$\mathbb{V}(a + bX) = b^2\mathbb{V}X = b^2\sigma^2$，

因此，$Y \sim \mathcal{N}(a + b\mu,\ b^2\sigma^2)$。

2）正态分布随机变量的指数变换

考虑一个正态分布的随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$，那么，$Y = e^X$ 的分布为?

因为 $X = \ln Y \sim \mathcal{N}(\mu, \sigma^2)$，故此随机变量 Y 是对数正态分布。

该对数正态分布的均值：$\mu_Y = e^{\mu + \frac{1}{2}\sigma^2}$，

该对数正态分布的方差：$\sigma_Y^2 = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$。

Cont'd

3）对数分布的变换

假设 Y 和 Z 为对数正态分布。则 Y^α（α 为常数）或 YZ 都是对数正态分布。令 $Y = e^{X_1}$，$Z = e^{X_2}$，其中 $X_i, i = 1, 2$ 是（联合）正态分布。因此，

$$Y^\alpha = e^{\alpha X_1};$$
$$YZ = e^{X_1 + X_2}.$$

(a) 由于 αX_1 是正态分布，因此 Y^α 是对数正态分布；

(b）由于 X_i 是（联合）正态分布，其和亦为（联合）正态分布，故此 YZ 是对数正态分布。

4）一般情形

假设有一个密度函数为 $f(x)$ 的随机变量 X，考虑 $y = y(x)$ 的一般变换，则 Y 的密度函数为？

定理：X 是一个随机变量，密度为 $f(x)$，取值范围为 $[a, b]$。让随机变量 Y 的实现值是 x 是单调增函数，即 $y = y(x)$，$y' > 0$，则 Y 在区间 $[y(a), y(b)]$ 上的密度函数 $g(y) = f(x(y))\frac{\mathrm{d}x}{\mathrm{d}y}$。

证明：略。

随机过程

1）随机游走（非平稳）

2）随机差分方程

$$x_t = ax_{t-1} + \epsilon_t \quad \overset{\text{等价于}}{\Longleftrightarrow} \quad x_t = ax_{t-1} + \mu + \upsilon_t.$$

其中：a 是一个常数；随机变量 $\begin{cases} \epsilon_t \overset{\text{i.i.d.}}{\sim} (\mu, \sigma^2) \\ \upsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2) \end{cases}$ 是独立同分布的正态分布，

即另有 $\begin{cases} \mathbb{C}(\epsilon_t, \epsilon_s) = 0; \\ \mathbb{C}(\upsilon_t, \upsilon_s) = 0 \end{cases} \quad \forall t \neq s$。

随机过程的分布

1）给定初始值求解。迭代，滞后算子法等。

$$x_t = a^t x_0 + a^{t-1}\epsilon_1 + a^{t-2}\epsilon_2 + \cdots + \epsilon_t = a^t x_0 + \sum_{s=1}^{t} a^{t-s}\epsilon_s.$$

2）分布特征

条件期望：$\mathbb{E}_0$ vs. $\mathbb{E}_t$

$\mathbb{E}_0$：经济主体知道 x_0 和 ϵ_0，但不知道 $\epsilon_1, \epsilon_2, \ldots$

$\mathbb{E}_t$：经济主体知道 $x_0, x_1, \ldots, x_t$ 和 $\epsilon_0, \epsilon_1, \ldots, \epsilon_t$，但不知道 $\epsilon_{t+1}, \epsilon_{t+2}, \ldots$

$$\begin{aligned} x_t &= a x_{t-1} + \epsilon_t, \\ &= a^t x_0 + \sum_{s=1}^{t} a^{t-s}\epsilon_s, \end{aligned}$$

$$\Rightarrow \quad \mathbb{E}_0 x_t = a^t \mathbb{E}_0 x_0 + \mathbb{E}_0 \sum_{s=1}^{t} a^{t-s}\epsilon_s = a^t x_0 + \mu \sum_{s=1}^{t} a^{t-s} = a^t x_0 + \mu \frac{1-a^t}{1-a}.$$

Cont'd

条件方差：$\mathbb{V}_0$ vs. $\mathbb{V}_t$

$$
\begin{aligned}
x_t &= a x_{t-1} + \epsilon_t, \\
&= a^t x_0 + \sum_{s=1}^{t} a^{t-s} \epsilon_s, \\
\Rightarrow \quad \mathbb{V}_0 x_t &= \sum_{s=1}^{t} a^{(t-s)^2} \mathbb{V}_0 \epsilon_s, \\
&= \sigma^2 \sum_{s=1}^{t} (a^2)^{t-s}, \\
&= \sigma^2 \sum_{i=0}^{t-1} (a^2)^i, \\
&= \sigma^2 \frac{1 - a^{2t}}{1 - a^2}.
\end{aligned}
$$

对于 $a \in (0,1)$，$t \uparrow \quad \Rightarrow \quad (\mathbb{V}_0 x_t) \uparrow$。

随机变量的长期特征

$$a \in (0,1) \quad \Rightarrow \quad \begin{cases} \lim\limits_{t\to\infty} \mathbb{E}_0 x_t = \frac{\mu}{1-a}, \\ \lim\limits_{t\to\infty} \mathbb{V}_0 x_t = \frac{\sigma^2}{1-a^2}. \end{cases}$$

因此，对于 $0 < a < 1$，长期视角下随机变量 x_t 是一个均值为 $\frac{\mu}{1-a}$ 和方差为 $\frac{\sigma}{1-a^2}$ 的正态分布。

通过计算得到 x_t 的显示解，易求同样 x_t 的期望和方差，因此时变中的 x_t 的密度函数为：

$$f(x_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{1}{2}(\frac{x_t-\mu_t}{\sigma_t})^2}, \quad \begin{cases} \mu_t = \mathbb{E}_0 x_t, \\ \sigma_t^2 = \mathbb{V}_0 x_t. \end{cases}$$

经济学应用：异质性部门

(1) 部门 i 的金融财富，假设其动态演化过程可由以下方程来描述（不同时间及不同个体间的冲击独立同分布)：

$$x_{it} = ax_{it-1} + \epsilon_{it}.$$

假设所有个体初始财富相同，$x_{i0} = x_0$。

大数定律在此处的应用：概率 vs. 份额。

社会总财富为

$$x_t = n \int_{-\infty}^{\infty} x_{it} f(x_{it}) \mathrm{d}x_{it} = n\mu_t.$$

其中，μ_t 是全部个体的平均财富或者说是任何个体 i 的预期财富。

以上假设个体是连续统，离散个体亦如是。

(2) RBC 模型中的对数技术变量是 AR(1) 过程。

更一般化的随机差分方程

$$x_t = \alpha_t x_{t-1} + \epsilon_t, \quad \alpha_t \overset{\text{i.i.d.}}{\sim} (a, \sigma_a^2), \ \ \epsilon_t \overset{\text{i.i.d.}}{\sim} (\mu, \sigma_\epsilon^2).$$

参考文献

方开泰、许建伦，1987，《统计分布》，科学出版社。

Klaus Wälde, 2012, Lecture Notes, Applied Intertemporal Optimization.