某些特殊概率分布之间的相互变换

杨航锋

1 离散分布的情况下

已知 $rand\_n()$ 表示可以均匀产生$\{1, 2, \cdots, n\}$之间的离散整数发生器，可以记作 $rand\_n() \sim DU(1, n)$ (DU表示离散均匀分布)。现有 $rand\_7()$ 怎么构造 $rand\_10()$ 呢? 或者，可以先思考稍微简单点的问题怎么通过 $rand\_10()$ 构造 $rand\_7()$ 呢? 很容易想到既然 $rand\_10()$ 可以均匀产生$\{1, 2, \cdots, 10\}$，那么只需要把大于7的数字过滤掉即可，遵循这个算法可以得到如下代码:

```
def rand_7():
    rand = float("inf")
    while rand > 7:
        rand = rand_10()
    return rand
```

要证明这个算法产生 $rand\_7()$ 的正确性，即证明 $p(x = k) = \dfrac{1}{7}, k \in \{1, 2, \cdots, 7\}$ 。因为 $rand\_10()$ 可能在第一次就产生合格的 k ，也可能第二次才产生合格的 k ，可能在第m次才产生合格的 k ，在这里仅证明 $k = 1$ 的情形，其它同理即可。故由概率论和幂级数（等比数列）相关知识可得

$$
\begin{aligned}
p(x = 1) &= \frac{1}{10} + \frac{3}{10} \times \frac{1}{10} + \cdots + (\frac{3}{10})^{m-1} \times \frac{1}{10} \\
&= \frac{1}{10} \times (1 + \frac{3}{10} + \cdots + (\frac{3}{10})^{m-1}) \\
&= \frac{1}{10} \times \frac{1}{1 - \dfrac{3}{10}} \\
&= \frac{1}{7}
\end{aligned}
$$

至此证明完毕。根据上面的证明可以得到一个一般性的结论：当 $a > b$ 时， $rand\_a()$ 可以实现 $rand\_b()$ 。

```python
def rand_b():
    rand = float("inf")
    while rand > b:
        rand = rand_a()
    return rand
```

现在再来分析 $rand\_7()$ 怎么构造 $rand\_10()$，如果能够把$rand_7()$映射到 $rand\_t()$ 且 $t > 10$ 时就能够利用上述结论解决该问题。构造 $(rand\_7() - 1) \times 7 + rand\_7()$，首先分析 $rand\_7() - 1$ 的取值范围为 $\{0, 1, \cdots, 6\}$，那么 $rand\_7() - 1) \times 7$ 的取值范围为 $\{0, 7, \cdots, 42\}$，而且每个数都只有一种组合得到，所以可以构造映射函数 $rand\_49()$，即 $rand\_49() = (rand\_7() - 1) \times 7 + rand\_7()$。故可编写如下代码

```python
def rand_10():
    rand = float("inf")
    while rand > 10:
        rand = (rand_7() - 1) * 7 + rand_7() #rand_49()
    return rand
```

上述代码可能有些瑕疵，从概率学的角度来说 $rand$ 有很大的可能性会大于10，因此 $while$ 循环将需要执行多次才能产生符合要求的 $rand$，从而可以优化该代码，让 $rand$ 与最接近49且小于49的10的倍数做比较，于是判断条件可以修改为 $rand \leqslant 40$ 故 $rand \in \{1, 2, \cdots, 40\}$，然后通过模运算 $rand\%10 + 1$ 映射到$\{1, 2, \cdots, 10\}$。

```python
def rand_10():
    rand = float("inf")
    while rand > 40:
        rand = (rand_7() - 1) * 7 + rand_7()
    return rand % 10 + 1
```

从特殊到一般归纳假设，假设有离散整数发生器 $rand\_a()$ 和 $rand\_b()$ 且 $a \neq b$，利用 $rand\_a()$ 表示 $rand\_b()$：

1. 如果 $a > b$ 则进入步骤2；否则，构造 $rand\_a^2 = (rand\_a - 1) \times a + rand\_a$，如果 $a^2 < b$ 继续构造 $rand\_a^3 = (rand\_a^2 - 1) \times a^2 + rand\_a^2$ 直到$a^k > b$，此时得到 $rand\_a^k$ 记为 $rand\_A$；
2. 经过步骤1有 $a > b$ 或者 $a^k > b$，利用下面代码构造 $rand\_b$

```python
def rand_b():
    rand = float("inf")
    while rand > b * (A // b): #表示与最接近A且小于A的b的倍数作比较。
        rand = rand_A()
    return rand % b + 1
```

2 连续分布的情况下

不妨考虑一般化的情况，假设 $rand\_ab() \sim U(a,b)$ 、 $rand\_cd() \sim U(c,d)$ ，如何通过 $rand\_ab()$ 来构造 $rand\_cd()$ 呢？从几何的角度上看，可以把区间 $[a,b]$ 上的点 x 一一映射到区间 $[c,d]$ 上，只需要构造线性变换 $f : \dfrac{x-a}{b-a} \times (d-c) + c$ 即可。如果能够证明 $\vartheta = \dfrac{rand\_ab() - a}{b-a} \times (d-c) + c$ 且 $\vartheta \sim U(c,d)$ 那么也就是构造出了 $rand\_ab()$ 到 $rand\_cd()$ 之间的映射关系。接下来将证明该结论：

由 $rand\_ab() \sim U(a,b)$ 可知其概率密度函数为

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0 \end{cases}$$

又因为 $\vartheta = \dfrac{X-a}{b-a} \times (d-c) + c$ ，故 ϑ 的概率分布函数为 $F_\vartheta(\theta)$

$$\begin{aligned} F_\vartheta(\theta) &= p(\vartheta \leqslant \theta) = p\left(\frac{d-c}{b-a}X - \frac{d-c}{b-a}a + c \leqslant \theta\right) \\ &= p\left(X \leqslant \frac{b-a}{d-c}(\theta - c) + a\right) \\ &= F_X\left(\frac{b-a}{d-c}(\theta-c)+a\right) \\ &= \int_a^{\frac{b-a}{d-c}(\theta-c)+a} \frac{1}{b-a} dx \\ &= \frac{\theta}{d-c} - \frac{c}{d-c} \end{aligned}$$

其中 $a \leqslant \dfrac{b-a}{d-c}(\theta-c)+a \leqslant b$ ，化简即 $c \leqslant \theta \leqslant d$ 。对上式求导得 $f_\vartheta(\theta)$

$$f_\vartheta(\theta) = \begin{cases} \dfrac{1}{d-c}, & c \leqslant \theta \leqslant d \\ 0 \end{cases}$$

至此证明完毕。