

## STAT243 Lecture 9.5 Generating Random Variables

### Logic ▾

在仿真研究中，我们需要从各种常见分布（如正态、Gamma、Beta、Poisson、t 等）中生成随机变量。

这些分布的采样方法通常已内置于 **Python** 和 **R** 中，且实现使用了高效且可靠的算法，因此这里仅介绍其核心原理。

许多统计计算与 Monte Carlo 方法的教材（如 Gentle、Robert & Casella）中对这些方法有详细论述。

其中多数方法利用了**分布间的关系**——例如，Beta 随机变量可由两个 Gamma 随机变量构造而得。

## 1 Multivariate Distributions

对于多元分布（如多元正态、多元 t），

`mvtnorm`（R 包）或 `scipy.stats.multivariate_normal`（Python）提供了概率密度与 CDF 计算工具。

### 1.1 多元正态的生成方法

若协方差矩阵为  $\Sigma$ ，则可通过 **Cholesky 分解** 生成：

```
Python
1 L = np.linalg.cholesky(covMat) # L 为下三角矩阵
2 x = L @ np.random.normal(size = covMat.shape[0])
```

此时生成的  $x$  满足  $x \sim \mathcal{N}(0, \Sigma)$ 。

#### ⚠ Remark ▾

若协方差矩阵奇异，可使用带 pivoting 的 Cholesky 分解，将秩亏部分对应行置零，从而生成满足约束的随机向量。但需注意输出向量的重新排序。

## 2 Inverse CDF Method (反CDF法)

要生成  $X \sim F$ ，其中  $F$  是累积分布函数 (CDF)，若  $F$  可逆，则可按以下步骤：

1. 生成  $Z \sim \mathcal{U}(0, 1)$ ；
2. 设  $X = F^{-1}(Z)$ 。

即：

$$X = F^{-1}(Z), \quad Z \sim \mathcal{U}(0, 1)$$

对于离散分布，可使用离散化的 CDF。

对于多元分布，可按条件分布顺序采样：

$$f(x_1)f(x_2|x_1)f(x_3|x_1, x_2) \cdots f(x_k|x_1, \dots, x_{k-1})$$

该方法在存在解析条件分布时尤为有效。

## | 3 Rejection Sampling (拒绝采样)

### | 3.1 基本思想

若目标密度为  $f(x)$ ，我们希望从中采样。  
引入辅助变量  $u$ ，则可写作：

$$f(x) = \int_0^{f(x)} du$$

即  $(X, U)$  的联合分布为：

$$(X, U) \sim \mathcal{U}(x, u) : 0 < u < f(x)$$

但我们无法直接从  $f(x)$  采样，因此使用“包络函数”  $cg(x)$  来近似  $f(x)$ ，  
其中  $g(x)$  易于采样，且存在常数  $c > 0$  满足：

$$cg(x) \geq f(x), \quad \forall x$$

### | 3.2 算法步骤

1. 生成  $x \sim g(x)$ ；
2. 生成  $u \sim \mathcal{U}(0, 1)$ ；
3. 若  $u \leq \frac{f(x)}{cg(x)}$ ，则接受  $x$ ；否则拒绝并重采样。

### | 3.3 接受概率

接受率为：

$$P(\text{accept}) = \frac{1}{c} = \frac{\int f(x)dx}{\int cg(x)dx}$$

因此，理想的  $g(x)$  应：

- 具有比  $f(x)$  更“胖的尾部”；
- 使  $c$  尽量小，以减少拒绝次数。

### | 3.4 几何直观

我们在  $cg(x)$  曲线下均匀采样  $(x, u)$ ，  
仅保留位于  $f(x)$  下方的点，即  $u < f(x)/cg(x)$ 。

---

### | 3.5 拒绝采样中的 Squeezing 技巧

若  $f(x)$  计算代价高，可使用其下界  $f_{\text{low}}(x)$ ：

$$u \leq \frac{f_{\text{low}}(x)}{cg(x)}$$

满足条件时可直接接受，避免计算完整  $f(x)$ ，从而减少计算负担。

---

### | 3.6 例：截断正态分布

若标准正态分布截断点在  $a > 0$ ，  
直接采样并拒绝效率极低。

Robert (1995) 提出用**平移指数分布**作为包络函数：

$$g(x) = \lambda e^{-\lambda(x-a)}, \quad x > a$$

其中  $\lambda$  根据截断点选择, 使  $cg(x)$  能包络正态尾部。  
当  $a < 0$  时, 可通过取负对称实现。

---

## | 4 Adaptive Rejection Sampling (自适应拒绝采样) (optional)

RS 的难点在于选取良好的包络函数。

**自适应拒绝采样 (ARS)** 适用于连续、可微且对数凹的密度。

### | 4.1 基本思想

1. 在  $\log f(x)$  空间上, 用切线与割线构造上下包络;
2. 上包络对应分段指数函数;
3. 从上包络采样, 再通过“squeezing”判定是否接受;
4. 每次接受需要计算  $f(x)$  的新点后, 更新包络。

ARS 能动态改进效率, 但要求  $f(x)$  对数凹。

---

## | 5 Importance Sampling

### | 5.1 目标

估计期望:

$$\phi = E_f[h(Y)] = \int h(y)f(y)dy$$

若直接从  $f(y)$  采样困难, 可引入辅助分布  $g(y)$ , 则:

$$\phi = \int h(y) \frac{f(y)}{g(y)} g(y) dy$$

Monte Carlo 估计量为:

$$\hat{\phi} = \frac{1}{m} \sum_{i=1}^m h(y_i) \frac{f(y_i)}{g(y_i)}, \quad y_i \sim g(y)$$

定义权重:

$$w_i = \frac{f(y_i)}{g(y_i)}$$

若  $f$  仅已知至比例常数 (如贝叶斯情境), 使用归一化权重:

$$w_i^* = \frac{w_i}{\sum_j w_j}$$

---

### | 5.2 性质与建议

- 只需  $f, g$  有**相同 support**, 不必要求  $cg(x) \geq f(x)$ ;
- 若  $g$  尾部比  $f$  轻, 方差可能爆炸;
- 为降低方差, 应确保  $w_i$  大时  $h(y_i)$  小, 即避免过度权重。

估计量方差为:

$$\text{Var}(\hat{\phi}) = \frac{1}{m} \text{Var} \left( h(Y) \frac{f(Y)}{g(Y)} \right)$$

启发：

若  $h(y)$  大的区域贡献主要， $g(y)$  应更多地采样这些区域。  
在稀有事件估计中，这意味着过采样稀有事件，再用权重修正。

---

## 5.3 Sampling Importance Resampling (SIR)

若目标是从  $f$  中生成样本而非仅估计期望：

1. 从  $g(y)$  生成样本  $y_i$ ；
2. 根据权重  $w_i$  进行有放回重采样。

此过程称为 **SIR (Sampling Importance Resampling)**。

---

## 6 Ratio of Uniforms (比值法) (optional)

该方法基于如下构造：

若  $(U, V)$  在集合

$$C = (u, v) : 0 \leq u \leq \sqrt{f(v/u)}$$

上均匀分布，则随机变量：

$$X = \frac{V}{U}$$

的密度与  $f(x)$  成比例。

### 6.1 算法

1. 选取矩形区域包含  $C$ ：

$$0 \leq u \leq \sup_x \sqrt{f(x)}, \quad \inf_x x \sqrt{f(x)} \leq v \leq \sup_x x \sqrt{f(x)}$$

2. 在矩形中采样  $(u, v)$ ；
3. 若  $u \leq \sqrt{f(v/u)}$ ，接受并令  $x = v/u$ 。

可根据  $f(x)$  形状裁剪矩形以提升效率。

**推荐阅读：** Monahan (2001)，在其第 323 页提出适用于离散分布的改进版比值法。