

---

## 不相等权重抽样 (G05NEF)

抽样 (Sampling) 这个名词有许多不同的含意，例如在讯号处理中是指从连续的讯号转换成离散的讯号，或是重复使用不同的声音重制成新的录音，或在受控制的实验中选取研究主题。在这里我们指的是在特定的资料中，随机的选取一或多个观测值。

在很多领域中都需要自资料中选取样本，它是许多算法的基础、个别配置实验设计的分配方法或是可以减少大量的数据。抽样的表现方式分成两种：

### 放回抽样 (with replacement):

当透过放回方式进行对原数据集合进行抽样时，样本值可能会重复出现。因此样本有可能会比原先的数据集更大。

### 不放回抽样 (without replacement):

当透过不放回方式对原数据集合进行抽样时，样本值最多只会出现一次。因此样本将不会比原始数据集要大。

这两个抽样方法可以进一步再分成两类：

### 相等权重抽样 (equal weights):

当采用相等权重取样，在原始资料中的每个资料都会与其他的资料一样有相同被选中的机率。

### 不相等权重抽样 (unequal weights):

若采用不相等权重取样，在原始资料中每个资料会依其所指定好的权重以成比例的方式被选中。

在 NAG 算法库 23 版之前，要进行其他三种抽样方法，可透过 [G05NEF](#) (使用不相等权重不放回抽样) 函数所附加的功能，它是构成任何重复抽样函数的基础。

上述所提的四种抽样的方法，可以透过以下的 NAG 函数调用：

- [G05TLF](#): 放回抽样、相等权重
- [G05TDF](#): 放回抽样、不相等权重
- [G05NDF](#): 不放回抽样、相等权重
- [G05NEF](#): 不放回抽样、不相等权重

这些函数的都有示例程序，可以帮助您了解如何利用它们来进行不同种类的抽样。

有些时候，我们会想要在数据集合中不同的阶层或子群体中取得样本，而不是在整个资料中进行抽样，此通常被称为分层抽样 (stratified sampling)。尽管 NAG 算法库并未包含分层抽样的函数，您仍然可以很直接的依原始数据集的分层，透过以上所提到函数进行分别抽样。