第四章 仿真数据的统计分析

一. 仿真输入数据的分布拟合
二. 系统的性能测度及其估计
三. 终态仿真的置信区间
四. 稳态仿真的置信区间
五. 仿真输出估计精度比较与控制

一. 仿真输入数据的分布拟合

常用的方法有：

- 利用观察数据建立经验分布函数
- 通过对数据分布形式假定、参数估计和分布拟合优度检验等过程，确定输入随机变量的分布
- 直接用观察数据得到总体分布的估计

1. 经验分布函数

设$\{x_i\}$为仿真输入随机变量的观察值，将其从小到大排序后记为$\{x_{(i)}\}$，经验分布函数为

$$F_n(x)=\begin{cases}0, & x<x_{(1)}\\ \dfrac{i}{n}, & x_{(i)}\le x<x_{(i+1)} \quad i=1,2,\cdots,n-1\\ 1, & x\ge x_{(n)}\end{cases}$$

用经验分布的随机数做为仿真输入数据。

2. 分布拟合

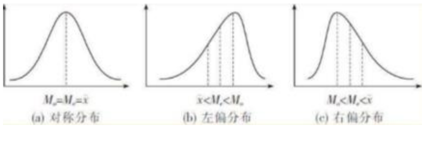
分布假设：依据样本数据提供的特征分析结果，确定其最可能服从的分布；

参数估计：使用样本数据，通过极大似然估计、最小二乘估计法、无偏估计法和矩量估计等方法估计参数

拟合优度检验：评价分布是否很好地实现了对观测数据的拟合

（1）统计分布的特征分析

位置参数：均值$\overline{x}$、众数M_0、中位数M_e

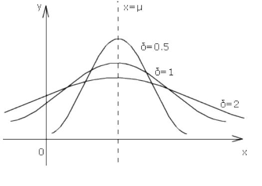


从分布角度看，众数始终是一组数据分布的最高峰值，中位数是位于一组数据中间位置上的值，均值则是全部数据的算术平均。

（1）统计分布的特征分析

尺度参数：刻画数据分布图形的聚集程度

正态分布 $N(\mu,\sigma^2)$

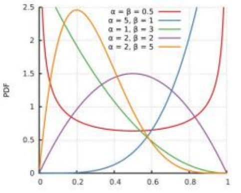


（1）统计分布的特征分析

形状参数:偏态SK、偏态K

偏态$SK=\frac{n\sum(x_i-\bar{x})^3}{(n-1)(n-2)s^3}$

偏态$K=\frac{n(n+1)\sum(x_i-\bar{x})^4-3\left[\sum(x_i-\bar{x})^2\right]^2(n-1)}{(n-1)(n-2)(n-3)s^4}$



（2）参数估计

假设总体的概率密度f(x|θ)是已知的，其中θ是未知参数，若$x_1,x_2,\dots x_n$为来自总体的一个样本，则它的联合概率密度函数可以写成

$$L(\theta;x_1,x_2,\dots x_n)=f(x_1,x_2,\dots x_n|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$$

对于离散分布

$$L(\theta;x_1,x_2,\dots x_n)=\sum_{i=1}^{n}P(X=x_i|\theta)$$

（2）参数估计

θ的极大似然估计值，可以使得似然函数L取最大值，为便于计算，对L两端取对数计算

$$\ln L(\theta;x_1,x_2,\dots x_n)=\sum_{i=1}^{n}\ln f(x_i|\theta)$$

以θ为自变量，求FOC

$$\frac{d\ln L(\theta;x_1,x_2,\dots x_n)}{d\theta}=\frac{d\sum_{i=1}^{n}\ln f(x_i|\theta)}{d\theta}=0$$

（3）分布拟合检验

设$\{x_i\}$为仿真输入随机变量X的观察值，零假设：

H_0：X的分布函数为$F(x)=F_0(x)$

其中$F_0(x)$为已知的理论分布。我们需要利用样本提供的信息判断这一假设是否合理，若合理，则接受假设，否则便拒绝假设。

注意：犯两类错误的概率。

假设检验大致可分为以下步骤：

① 提出假设H_0；

② 给定检验水平α, 一般α取0.05, 0.01, 或0.10；

③ 根据H_0的内容，选取适当的统计量T，并能确定出相应统计量的分布；

④ 利用相应分布的分位数，建立检验水平α下的拒绝域W；

⑤ 利用观察数据算出统计量的具体值；

⑥ 若统计量的具体值落入拒绝域W中，则在检验水平α下拒绝假设H_0，否则接受H_0。

χ^2检验　　**具体步骤:**

① 把实数轴分为k个不相交的区间$[a_0,\ a_1)$，$[a_1,\ a_2)$，…，$[a_{k-1},a_k)$，其中a_0,a_k可分别取$-\infty,+\infty$；

② 计算概率$p_i=P\{a_{i-1}<X\le a_i\}=F_0(a_i)-F_0(a_{i-1})$，$i=1,2,\dots,k$，并算出$np_i$，称之为理论频数；

③ 计算样本观察值$x_1,x_2,\dots,x_n$落在区间$[a_{i-1},a_i)$中的频数n_i，称之为经验频数；

④ 对给定水平α，查χ^2分布表得临界值，

$$\chi_\alpha(k-p-1):P\{\chi^2>\chi_\alpha(k-p-1)\}=\alpha$$

其中k是区间个数，p是分布中待估参数的个数；

⑤ 计算χ^2的具体值：　$\chi^2=\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}$

⑥ 若$\chi^2>\ \chi_\alpha(k-p-1)$ 则拒绝假设H_0，否则接受H_0。