

# 食品卫生安全保障体系的数学模型及改进

## 摘 要

近年来，随着经济的迅速发展，我国的食品安全问题也日益突出，建立适合于我国国情的食品卫生安全保障体系已经成为关系到国计民生的迫切任务。本文在建立我国的人群食物摄入量模型、食品污染物分布模型的基础上建立食品安全评估模型，并对模型建立过程中对食物分类、样本选取、抽样数据不完整、不匹配时的处理方式等问题进行了探讨。

问题分析部分在对我国膳食结构的特点进行分析的基础上结合污染物因素提出了进行食物分类的三条原则，并对食物进行分类，同时，针对食品安全评估采用的 99.999% 右分位点进行了分析，发现由于个体对污染物承受能力的不同，该指标并不能真正反映我国的食品安全状况。

在人群食物摄入量模型的建立过程中，首先在对国家依据气候、地理、经济、民族风俗等因素进行地域划分的基础上提出了“层次化样本分配模型”，该模型利用我国人口普查数据中的性别、年龄结构属性作用于采样点的选取和样本数量的分配，用有限的样本按比例覆盖各种不同类型的采样点，并证明了按自然男女比例分配采样男性和女性的样本数量能很好的消除采样误差。通过对采样数据特征的分析，推测食物摄入量的概率密度函数应该是“偏正态”分布，在“最大熵”原理的指导下验证得出其概率密度函数为“对数正态分布”。接着讨论了食物摄入量模型中的时间因素，并针对抽样调查方案设计、食物分类等难点问题给出了专门的回答。

在污染物分布模型中，和食物摄入量模型类似的原理，在“最大熵”原理的指导下推测其概率密度函数应该是贝塔分布，并给出了概率密度的解析式形式。针对污染物含量测定中暴露出的采样率低，监测数据不完整的特点，提出了一种将污染物调查中的样本在时间轴上平移的方法，这种方法只改变样本标志值（即污染物在食物中的含量百分数）的均值而不改变分布的形状，较好的解决了数据不完整的问题。然后针对污染物分布模型中的难点问题给出了专门的回答。

风险评估模型基于食物摄入量和污染物分布模型，定义了“矩元法”对污染物摄入量进行评估，并给出了该模型的数值求解方法以及用于右分位点确定的相关算法。接着，对题目中提出的风险评估模型建立过程中的难点给出了专门的解答。

在模型求解过程中，使用肉类的相关数据对三个模型进行了求解，并对结果进行了分析。

最后，关于我国当前的食品卫生保障现状，提出了自己的一点建议。在文章的附录部分，根据食物摄入量模型制作了相应的调查问卷。

**关键词：**层次化样本分配 对数正态分布 贝塔分布 最大熵原理 矩元法

## 一、 问题重述（略）

## 二、 问题分析

目前，美国和欧盟的科学家主要在建立人群食物摄入量模型和污染物分布模型的基础上构造风险评估模型对居民的食品安全进行评估。其中，人群食物摄入量模型用于估计不同地区、不同性别、不同年龄、不同季节、不同劳动强度、不同经济收入的人群各类食品的一天摄入量；污染物分布模型用来估计各类食物中各种污染物的含量；风险评估模型用某些污染物每天摄入量的 99.999% 的右分位点来评估某一时刻的食品安全风险。

### 2.1. 食品污染及其分类

食品污染是指食物受到有害物质的侵袭，造成食品安全性、营养性或感官性状发生改变的过程<sup>[1]</sup>。根据污染物的性质，食品污染可分为三个方面：①，生物性污染：有微生物及其毒素，主要是细菌及细菌毒素，霉菌及霉菌毒素等；②，化学性污染：包括化学农药、有害金属、多环芳烃类、N-亚硝基化合物等；③，放射性污染：食品可以吸附或吸收外来的放射线核素，主要以半衰期较长的  $^{139}\text{Cs}$  和  $^{60}\text{Co}$  最具卫生学意义。其中，以有机磷、有机氯为代表的农药残留物和以汞、铅、镉、砷为代表的工业重金属对食品安全的威胁最大。生物性污染主要产生于食品的运输、加工环节，通过物流调配等人为手段可在短期内予以消除；化学性污染与工业发展带来的生态环境恶化有直接关系，并且由于现在人为的污染已经超过了生态环境的自愈能力，因此该类型的污染将长期存在；放射性污染只在特定环境中存在，不属于主流的食品污染源。综上，我国的食品卫生安全保障体系应该重点考虑农药、重金属造成的污染。

### 2.2. 衡量污染物摄入量的标准

按照国际通行的标准，如果把每个人每天某种污染物摄入量  $w$  看成是一个随即变量，食品卫生安全保障体系关心的不仅仅是  $w$  的均值，更关注它的 99.999% 的右分位点  $Z$ ，即保证绝大多数居民每人每天关于该种污染物摄入量少于  $w$ 。而实际上，对于不同类型的污染物，用每人每天的摄入量来衡量是否安全并不完全合适。例如，对于一个成人来说，每周铅的摄入量为  $0.025\text{mg/kgBW}$  ( $\text{mg/kgBW}$  表示每公斤体重所摄入受试试剂的毫克数) 时并不会发生铅中毒<sup>[2]</sup>，但是当儿童每周铅的摄入量达到这个水准时由于儿童肠胃对铅的吸收率比较高就会发生铅中毒。

从污染物代谢以及其毒害作用机理来看，用每人每天的污染物摄入量来衡量各种不同类型的污染物摄入情况也是不合理的。严格说来，由于危害描述一般是由毒理学试验获得的数据外推到人，医学上用人体的每日容许摄入量(ADI)值来衡量人体对农药、兽药和食品添加剂残留的承受能力，用暂定每周耐受摄入量(PTWI)来衡量对蓄积性污染物如铅、镉、汞的承受能力，用暂定每日耐受摄

入量(PTDI)来衡量对非蓄积性污染物如砷的承受能力<sup>[3]</sup>。为了便于污染物之间的横向比较，可以将周耐受摄入量折算成每天的耐受摄入量<sup>[2]</sup>，但是不同单位代表的物理含义并不相同。

因此，对于处于不同年龄段的个体来说，用每天每人对某类污染物的摄入量来衡量是否安全是不合理的，并且，即使是使用每公斤体重摄入污染物的含量作为衡量标准，对应到不同年龄段也有不同的参数值。*w* 的 99.999%右分位点只能保证绝大多数人的安全，但是不能反映高暴露人群污染物摄入量的特点，不能完全满足食品卫生保障体系的要求。故需要根据具体情况设定更精细的指标体系。

### 2.3. 食物分类方法分析

我国饮食文化源远流长，食物品种多而繁杂，对膳食摄入量进行统计首先需要进行必要的分类。食品安全保障体系的建立还包括对不同地区、不同季节里污染物的分布进行评估，因此食物分类不仅需要考虑到调查的易操作性和准确性，还应该考虑在污染物分布模型中由于该类型的食物受到诸如土壤、水等因素的影响而导致污染物呈现不同类型的分布的情况。譬如如若将肉禽类和鱼虾统一归结为肉类，由于肉禽类主要以谷物、植物为食，体内所累积的污染物主要来自于土壤、空气等，而鱼虾是水生动物，体内的污染物主要来自于水污染。

再则，污染物虽然在不同环境要素中存在的含量、形态以及危害程度不同，但是它们都可以通过食物链，在农产品中富集，进而造成食品污染，影响人类身体健康<sup>[4]</sup>。各类食物在食物链中所处的位置可能会影响到其中的污染物分布状况，因此，在对食物进行分类时为了增强横向的比较性，还要将食物分类到食物链的不同层次。

因此，对于食物保障评估体系中的食物分类制定了以下三条原则：

- 1， 污染物来源不同的食物必须分为不同的类别，如肉禽类、鱼虾类必须分为两个类别；
- 2， 在满足 1 的条件下，主成分比较类似的食物分为同一类，如大米和面粉，其中的主要成分均为淀粉；
- 3， 富集作用类似的食物放置于食物“金字塔”的同一层。

基于以上三原则，将食物分为水、谷物类、蔬菜类、水果类、豆类及其制品、肉禽蛋类、鱼虾水产类、奶类及奶制品、油脂类共 9 类，其在“食物金字塔”中的层次分布如图 1 所示。

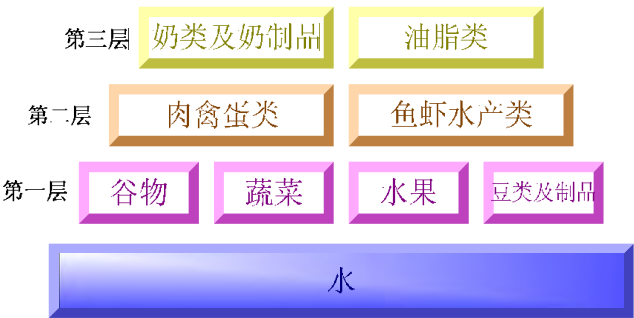


图 1 食物类型“金字塔”结构

### 三、 问题假设

- 1, 假设人群食物摄入量、污染物分布是独立的;
- 2, 在一段时间内, 一个地区的污染程度不会发生很大的变化;
- 3, 在一段时间内, 一个地区的食物结构不会发生很大的变化;

以上假设仅仅是基于模型之间的关系, 在模型的建立过程中会引入很多局部性的假设的。

### 四、 参数分类和符号说明

首先, 对食品安全保障体系中涉及的人群、食物类型、经济区域类型等进行必要划分。

将年龄划分为 5 个年龄段 (童年 0—5 岁、少年 6—18 岁、青年 19—35 岁、中年 36—50 岁、老年指 50 以上);

按照食物分类方法将食物划分为水、谷物类、蔬菜类、水果类、豆类及其制品、肉禽蛋类、鱼虾水产类、奶类及奶制品、油脂类 9 种类型, 其中, 水是第一类, 谷物是第二类, 依次类推;

选取典型性的污染物, 如果铅、镉、有机磷、有机氯, 设共有  $N$  类, 其中铅是第 1 类, 镉是第二类, 依次类推;

气候和地理差异在很大程度上决定一个地区的饮食结构, 根据这两个因素将我国分为十二个大区。值得注意的是, 描述区域的名称与传统的说明方式并不相同, 主要是在区域、气候差异的基础上依据饮食结构进行划分, 由于少数民族地区的饮食结构与汉族存在较大的差异, 因此少数民族聚居的地方都单独划分。十二个大区划分如下: 东北 (黑龙江、辽宁、吉林)、华北 (北京、天津、河北、山西、山东)、内蒙古、西南 (四川、重庆、云南、贵州)、西藏、华东 (上海、江苏、浙江)、华中地区 (安徽、江西、河南、湖北、湖南)、华南地区 (福建、海南、广东、台湾、香港、澳门)、西北地区 (陕西、甘肃、青海)、宁夏、广西、新疆, 东北是第一类地区, 华北是第二类地区, 其余大区编号依次类推;

每个地区根据经济发展水平的不同划分为 6 个不同类型的经济区域: 大城市、中小城市、一类农村、二类农村、三类农村、四类农村, 其中大城市包括北京、上海、天津、重庆、哈尔滨、沈阳、大连、济南、青岛、宁波、南京、广州、深圳、郑州、成都、西安、武汉、厦门; 其余的城市为中小城市 (不含县级市); 农村四类地区的划分以国家统计局《中国农村分区域综合经济实力研究报告》为依据。其中, 大城市是第一类区域, 中小城市是第二类区域, 依次类推。

符号说明如下:

$Q$ : 全国范围内进行膳食摄入量、膳食结构的抽样调查的样本总数;

$Z_d (1 \leq d \leq M)$ : 将某个区域 (东北大区、华北大区之类的大区或者省、市、乡镇等) 划分成  $M$  ( $M$  是自然数) 个更小一级的区域,  $Z_d$  表示其中的第  $d$  个区域;

$Z_{d,q}(1 \leq q \leq 6)$ : 将  $Z_d$  区域按经济发展水平分成大城市、中小城市、第一类农村等 6 个类型, 其中的第  $q$  类地区;

$R_d(1 \leq d \leq 12)$ : 从大区级的划分, 第  $d$  类地区的人口占全国总人口的百分比;

$D_q(1 \leq q \leq 6)$ : 第  $q$  类经济区域;

$R_{d,q}(1 \leq q \leq 6)$ :  $Z_d$  区域中 (此时  $Z_d$  表示地理上的区域), 第  $q$  类经济区域人口占

该区域总人口的百分比, 显然有  $\sum_{q=1}^6 R_{d,q} = 1$ ;

$A_i(1 \leq i \leq 5)$ : 第  $i$  个年龄段;

$V_{d,i}(1 \leq i \leq 5)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人口数量占该地区人口总数的百分比;

$V_{d,q,i}(1 \leq i \leq 5)$ :  $Z_{d,q}$  区域中, 第  $i$  个年龄段人口数量占该地区人口总数的百分比;

$F_j(1 \leq j \leq 9)$ : 第  $j$  种类型的食物;

$L_{d,i,j}(1 \leq i \leq 5, 1 \leq j \leq 9)$ :  $Z_d$  区域中, 第  $i$  个年龄段人群的食物结构中第  $j$  类食物占该人群食物摄入量的百分比;

$L_{d,q,i,j}(1 \leq i \leq 5, 1 \leq j \leq 9)$ :  $Z_{d,q}$  区域中, 第  $i$  个年龄段人群的食物结构中第  $j$  类食物占该人群食物摄入量的百分比;

$C_k(1 \leq k \leq N)$ : 第  $k$  类污染物;

$Male_d$ :  $Z_d$  区域中, 男性所占总人口的百分比;

$Male_{d,q}$ :  $Z_{d,q}$  区域中, 男性所占总人口的百分比;

$Male_{d,i}(1 \leq i \leq 5)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人口中, 男性所占的百分比;

$Male_{d,q,i}(1 \leq i \leq 5)$ :  $Z_{d,q}$  区域中, 第  $i$  个年龄段的人口中, 男性所占的百分比;

$Female_d$ :  $Z_d$  区域中, 女性所占总人口的百分比;

$Female_{d,q}$ :  $Z_{d,q}$  区域中, 女性所占总人口的百分比;

$Female_{d,i}(1 \leq i \leq 5)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人口中, 女性所占的百分比;

$Female_{d,q,i}(1 \leq i \leq 5)$ :  $Z_{d,q}$  区域中, 第  $i$  个年龄段的人口中, 女性所占的百分比;

$f_{d,i,j}(x)(1 \leq i \leq 5, 1 \leq j \leq 9)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人群关于第  $j$  类食物摄入量  $x$  的概率密度函数;

$g_{d,j,k}(x)(1 \leq j \leq 9, 1 \leq k \leq N)$ :  $Z_d$  区域中, 第  $k$  类污染物在第  $j$  类食物中分布的含量百分比  $x$  的概率密度函数;

$h_{d,i,k}(x)(1 \leq i \leq 5, 1 \leq k \leq N)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人群摄入的第  $k$  类污染物的含量  $x$  的概率密度函数;

$s_{d,i,k}(x)(1 \leq i \leq 5, 1 \leq k \leq N)$ :  $Z_d$  区域中, 第  $i$  个年龄段的人群摄入的第  $k$  类污染物的含量  $x$  的概率密度函数;

$s_{d,k}(x)(1 \leq k \leq N)$ :  $Z_d$  区域中, 人群摄入的第  $k$  类污染物的含量  $x$  的概率密度函数;

需要特别说明的是, 当被抽样的区域  $Z_d$  确定以后, 符号中参数的下标  $d$  可以被省略,  $R_q$ 、 $V_i$ 、 $L_{i,j}$ 、 $Male_i$ 、 $Female_i$ 、 $f_{i,j}$ 、 $g_{j,k}$ 、 $h_{i,k}$ 、 $s_k$  分别代表区域  $Z_d$  下的对应参数。同样当被抽样的区域  $Z_{d,q}$  被确定以后, 符号中参数的下标  $d, q$  也可以被省略。

## 五、 模型建立

按照建立食品卫生安全保障体系的基本要求, 我们首先针对人群食物摄入量、污染物分布和风险评估建立三个数学模型, 并在建立模型的过程中对涉及数据采集点的选取、调查方案的设计、抽样数据的采集方式、抽样内容以及所采集数据的处理流程都进行了详尽的描述。

### 5.1. 人群食物摄入量模型

人群的食物摄入量受到了地区、性别、年龄、季节、劳动强度、经济收入水平、城乡差别等诸多因素的影响, 在数学模型中对如此之多的因素做出定量描述是比较困难的, 并且由于其中涉及的大部分因素不存在着显著的量化指标, 操作性不强。众所周知, 食物摄入量模型需要根据国家的总膳食数据来建立, 但是对食物消费的普查难度甚至超过了全国人口普查, 因此需要尽可能的在使调查结果准确的情况下减少调查的工作量。这就要求尽可能的利用从其它一切渠道获得的信息。

#### 5.1.1. 层次化样本分配模型

在关于抽样地点的选择和分配上建立了层次化的抽样点选取模型, 首先将地区差异、经济收入水平、城乡差别等因素用人口比例统一量化后用于样本数量在

各采样地点之间进行分配。

为了衡量地理、气候因素对饮食结构和食物摄入量的影响，将全国分为了十二个大区，将样本数量在大区之间进行分配。

为了同时体现城乡差别和人均收入水平对食物结构的影响，按照经济水平，地区可以进一步分为六个经济类型区域（大城市、中小城市、一类农村、二类农村、三类农村、四类农村）。

在采样区域限定到某个大区内的某个经济类型区域的基础后，按照各年龄段里人口数量对样本数量进行分配，在此基础上根据该年龄段男女比例确定男性样本和女性样本的数量。

表 1 中列出了对于在抽样调查中的一个样本，所需要得到的信息。抽样点的采集任务描述形如：连续 12 个月（采集日期定为每个月的 10 号），采集东北地区（东北三省）某个人均收入在 1000 元/月—1500 元/月的城市（中、小城市经济类型）里的年龄在 10-18 岁之间的 30 个男孩和 20 个女孩的食物摄入量信息。可以看出，抽样点自身的选取包含了采集时间、地区、人均收入、城市/乡村，年龄段、性别的信息，具体调查时只需要在满足这些约束条件的前提下，对各种基本食物类型的摄入量进行统计。

表 1 抽样调查信息基本内容

姓名	性别	地区	年龄	城/乡（收入水平）	职业类型	采样时间	9 类基本食物的日摄入量

层次化的抽样模型的特点主要体现在样本数量的分配和抽样点的选取上，其中抽样点的选取流程如图 2 的前两级所示，先选取典型大区，然后在分区中进一步选取典型经济水平区域。

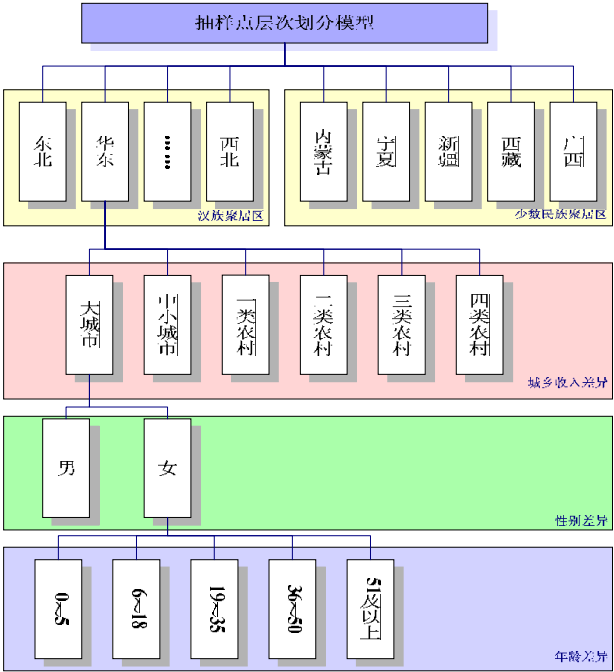


图 2 样本点的层次化分配模型

在符号说明部分已经定义在全国进行抽样的样本总数为  $Q$ ，各采集点的样本数量分配过程如算法 1 所表示：

算法 1 样本数量的分配过程

- 1, 首先在大区之间进行样本数量的分配, 第  $d(1 \leq d \leq 12)$  大区分配到的样本数为  $Q \times R_d$ ;
- 2, 然后在大区内部进行样本分配, 第  $d(1 \leq d \leq 12)$  大区第  $q(1 \leq q \leq 6)$  类经济区域分配到的样本数量为  $Q \times R_d \times R_{d,q}$ ;
- 3, 在  $Z_{d,q}$  区域内进行样本分配, 第  $d(1 \leq d \leq 12)$  大区第  $q(1 \leq q \leq 6)$  类经济区域中的第  $i$  个年龄段的样本数量应为  $Q \times R_d \times R_{d,q} \times V_{d,q,i}$ ;
- 4, 分别对  $Z_{d,q}$  区域内第  $i$  个年龄段的样本数量按照该年龄段内男女比例进行分配, 则该年龄段内的男性样本数量为  $Q \times R_d \times R_{d,q} \times V_{d,q,i} \times Male_{d,q,i}$ , 女性样本的数量为  $Q \times R_d \times R_{d,q} \times V_{d,q,i} \times Female_{d,q,i}$

### 5.1.2. 样本统计方法

在问题分析部分中已经提出, 不同年龄阶段的人群对污染物的承受能力存在差异, 因此, 为了后续的风险评估, 需要分不同的年龄阶段对食物摄入量进行统计。

对于  $Z_{d,q}$  区域内第  $i$  个年龄段的样本而言, 其中有  $Q \times R_d \times R_{d,q} \times V_{d,q,i} \times Male_{d,q,i}$  个男性样本, 有  $Q \times R_d \times R_{d,q} \times V_{d,q,i} \times Female_{d,q,i}$  个女性样本, 可以发现男性样本和女性样本数量的比值就是  $Z_{d,q}$  区域第  $i$  个年龄段的男女比例, 因此, 可认为把这两类样本看做一个统一的样本进行食物摄入量的统计和分别对男、女性进行统计, 在此基础上求该年龄段的人群的食物摄入量模型的结果是一致的。以下给出证明:

**证明:** 为了求解  $Z_{d,q}$  区域内第  $i$  个年龄段的人群的食物摄入量模型, 采集一个男女样本比例符合自然男女比例的样本进行统计求解, 和分别采集若干个男性、女性样本先求解男/女性食物摄入量模型再求解该年龄段人群的模型得到的结果是一致的。

以下模型和参数均在  $Z_{d,q}$  区域下的第  $i$  个年龄段下有效。设对男性第  $j$  类食物摄入量采样  $N_1$  个样本, 女性第  $j$  类食物摄入量采样  $N_2$  个样本, 令男性的  $N_1$  个样



本中出现摄入第  $j$  类食物数量为  $x$  的概率为  $N_1(x)/N_1$ ，其中  $N_1(x)$  为  $N_1$  个样本中摄入第  $j$  类食物数量为  $x$  的样本个数，女性的  $N_2$  个样本中出现摄入第  $j$  类食物数量为  $x$  的概率为  $N_2(x)/N_2$ ，其中  $N_2(x)$  为  $N_2$  个样本中摄入第  $j$  类食物数量为  $x$  的样本个数。定义一组混合样本  $N$ ，其中男性  $N_1$  名，女性  $N_2$  名，当样本无穷大时，男女比例自然会趋近于自然比例， $N(x)$  表示混合样本中摄入第  $j$  类食物数量为  $x$  的样本个数，则有

$$\frac{N(x)}{N} =$$

定义男性第  $j$  类食物摄入量的概率密度函数为  $f_1(y)$ ，女性第  $j$  类食物摄入量概率密度函数为  $f_2(y)$ ，第  $i$  个年龄段人群第  $j$  类食物摄入量概率密度函数为  $f(y)$ 。由离散抽样求概率密度函数的原理，有以下等式：

$$\begin{aligned} \int_0^y f(y)dy &= \lim_{N \rightarrow +\infty} \frac{N(y)}{N} = Male_{d,q,i} \times \int_0^y f_1(y)dy + Female_{d,q,i} \times \int_0^y f_2(y)dy \\ &= \int_0^y (Male_{d,q,i} \times f_1(y) + Female_{d,q,i} \times f_2(y))dy \end{aligned}$$

$$\text{即有 } f(y) = Male_{d,q,i} \times f_1(y) + Female_{d,q,i} \times f_2(y) \quad (1)$$

(1)式表明，先分别对单独的男/女性种群做统计，然后以男女所占比例为系数对得到的概率密度函数进行线性组合可以得到该年龄段的人群关于第  $j$  类食物摄入量的概率密度函数。注意到这里的样本空间都应该是无穷大，由于实际上进行采集时的样本空间不可能无限大，因此，人为的设定混合样本中男女比例就是自然比例，可以使抽样调查的结果更加准确。

根据这个证明可以得到以下几个很重要的结论：

1，在食物摄入量模型中，食物摄入量的概率密度函数可以进行线性组合型的运算，其系数就是该类型样本空间在混合后整个样本空间中所占的比例；

2，单独的男女种群中关于某类食物摄入量的概率密度函数是独立的，同理，一个地区 A 划分成的不同区域 B、C 之间关于某类食物摄入量的概率密度函数  $f_b$ 、

$f_c$  也是独立的，若已知 B、C 中人口数量之比  $b, c$ ，则  $f_a = \frac{b}{b+c} \times f_b + \frac{c}{b+c} \times f_c$

3，采样点对各个大区下属的不同类型的经济水平区域进行直接采样，对数据进行统计、处理直接反映的是该大区下各类型经济水平区域的食物摄入量分布的概率密度函数，但是，通过该大区下各种不同类型经济水平区域的概率密度函数依照之间的人口比例进行线性组合就可以衡量整个大区的食物摄入量分布的概率密度函数，进而可以衡量全国的状况，赋予了对食物摄入量分布进行评估时

极大的灵活性。

4, 同一个大区依照经济发展水平被划分成了若干区域, 按照层次化样本分配模型进行样本数量的分配后, 这些区域所采集的样本不做任何处理混合后就等效于对该大区做无差别抽样的效果。

5,

### 5.1.3. 各年龄段的食物摄入量模型

通过 5.1.1 中的层次化样本分配模型和 5.1.2 的样本统计, 可以得到  $Z_{d,q}$  区域下离散化的概率分布情况 (若用  $P(x)$  表示样本中出现食物摄入量为  $x$  的概率, 则

$P(x) = N_x / N$ , 其中,  $N$  为样本空间总数,  $N_x$  是样本空间中统计出的食物摄入量  $x$  出现的次数), 若是已知概率分布函数的解析式形式, 通过拟和就可以测定解析表达式中的参数, 进而得到准确的分布。

为了讨论食物摄入量这个随机变量的分布, 先介绍**熵最大原理**。任何物质系统除了都受到或多或少的外部约束外, 其内部总是具有一定的自由度, 这种自由度导致系统内的各元素处于不同的状态。而状态的多样性, 状态的丰富程度 (混乱程度、复杂程度) 的定量计量标尺就是熵, 熵最大就是事物状态的丰富程度自动达到最大值。简而言之, 事物总是在约束下争取 (或呈现) 最大的自由权, 这是自然界的根本原则。

根据信息熵理论, 对于食物摄入量这个随机变量,  $N$  次抽样调查 (即  $N$  个样本) 组成的广义集合中, 每个样本出现不同的标志值 (食物摄入量) 的概率不相同, 假设标志值为  $x_1$  的个体有  $n_1$  个, 标志值为  $x_2$  的个体有  $n_2$  个, 标志值为  $x_3$  的个体有  $n_3$  个.....标志值为  $x_k$  的个体有  $n_k$  个, 则该广义集合出现的概率  $P$  依照概率的多项式分布公式如下:

$$P = \left[ \frac{N!}{n_1! n_2! \dots n_k!} \right] \times P_1^{n_1} \times P_2^{n_2} \times P_3^{n_3} \times \dots P_k^{n_k} \quad (2)$$

其中  $p_1$ 、 $p_2$  等分别是  $x_1$ 、 $x_2$  等等对应的出现概率。

由于这个广义集合是已经确定的, 因此  $p_1$ 、 $p_2$  等是这个广义集合的属性, 是一些常数, 而  $n_1$ 、 $n_2$  则是自变量, 即各个  $n_i$  值的不同对应着不同的广义集合。

把(2)两边取对数, 则可以得到(3), 其中  $C$  定义为这个集合的复杂程度。

$$\ln P = C + \sum_{i=1}^k n_i \ln p_i \quad (3)$$

将(3)推广到连续变量的应用场合, 则演化为(4)

$$\ln p = C + \int f(x) \ln p(x) dx \quad (4)$$

其中，积分遍及自变量的一切可能取值， $p(x)$ 是各个 $x$ 值的出现概率。

对于积分 $\int f(x) \ln p(x) dx$ ， $p(x)$ 为已知函数，而在复杂度最大时， $f(x)$ 也是明确唯一的函数，所以其积分为一个固定数值。此时，满足使用拉格朗日方法中的公式的条件，利用文献[6]的结论，可以把(4)改写成

$$\ln P = -\int N \times f(x) \times \ln \frac{f(x)}{p(x)} dx \quad (5)$$

即在 $p(x)$ 已知(可以作为一个约束条件)的情况下，抽样得到的分布函数 $f(x)$

的出现概率 $P$ 应该满足这个公式。 $-\int N \times f(x) \times \ln \frac{f(x)}{p(x)} dx$ 被称为是**条件熵**，随机

变量（某个年龄段食物摄入量分布）的分布应该使条件熵尽可能的大。

下面对 $Z_{d,q}$ 区域中第 $i$ 个年龄段下的某类食物的摄入量分布的形状做部分定性的分析：

1，根据经验，同一个年龄段的关于同一类食物的摄入量相差一般不大，大部分样本的摄入量集中在一个值的附近，在概率密度曲线上会形成一个“峰”，类似于正态分布；

2，由于不存在食物摄入量为0的情况，因此该概率密度曲线会过0点，而在该类食物摄入量为无穷大时，对应的概率应是一个无穷小量而不是0，即该分布的定义域为 $[0, +\infty]$ ，不可能和正态分布一样关于某个峰值对称，因此该分布必定不是一个正态分布，结合1，推测其为“偏正态分布”。

3，正态分布的代数平均数为常数，表示该分布所衡量的参数变化稳定在一定的水平，而偏正态分布在峰值左右两侧并不对称，为了使参数能在峰值两侧的一段区间内的概率稳定，故其几何平均数应为一个常数。

综合以上三点考虑，满足这三点要求的比较常见的分布是对数正态分布，并且该分布在生物学中得到了广泛应用<sup>[7]</sup>，该分布的解析式形式如下

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \quad (6)$$

下面，需要验证对数正态分布是否满足信息熵最大的条件从而是否适用于衡量随机变量（食物摄入量）的概率密度分布。验证过程如下：

验证过程 对数正态分布是否使随机变量分布的熵最大

1，验证 $\int_0^{+\infty} f(x) dx (x > 0)$ 是否等于1，即其是否是一个概率密度函数，很显然成立；

2，该分布变量的对数的几何分布是常数的，表示为 $\int_0^{+\infty} (\ln x) f(x) = \mu$ ；

3,  $\sigma^2 = \int_0^{+\infty} (\ln x - \mu)^2 f(x) dx$ , 即该分布对数的标准差为固定值

4, 根据文献[6], 定义一个新函数  $F$  (其中  $f$  就是指概率密度函数  $f(x)$  ),

$$F = -\int f \ln f dx + C_1 (\int f dx - 1) + C_2 [\int (\ln x) f dx - a] + C_3 [(\ln x - a)^2 f dx - \sigma^2]$$

5, 根据熵最大理论,  $F$  对  $f$  的偏微商为 0, 根据 1, 2, 3 中的等式关系, 可以验证其是否是对数正态分布 (理论上可行)。

因此,  $Z_{d,q}$  区域第  $i$  个年龄段对摄入第  $j$  类的食物量的概率分布函数具有式(6)

$$\text{的形式 } f_{i,j}(x) = \frac{1}{\sqrt{2\pi}\sigma_{i,j}x} \exp\left[-\frac{(\ln x - \mu_{i,j})^2}{2\sigma_{i,j}^2}\right] (1 \leq i \leq 5, 1 \leq j \leq 9)。 \quad (7)$$

#### 5.1.4. 食物摄入量模型随时间的变化

人的食物摄入量和食物结构会随着时间的变化, 比如冬季肉类所占膳食的比重比夏季大, 夏季里水果在膳食中所占比重比冬季大。在设计抽样方案时, 必须考虑到时间的因素, 因此设计了连续十二个月抽样模型, 以反映时间对膳食结构和食物摄入量模型的变化。

通过对相关资料历史数据的查阅发现, 一定区域内的饮食结构年与年之间比较稳定, 随着季节的变化会有较大的波动。并且波动的曲线并不平滑, 不适合于用连续函数进行拟和。因此, 人群食物摄入量模型并没有强制的加入到影响模型变化的因素中去。

具体的做法是对月份之间的数据做分析, 统计出参数。比如, 12 个月内食物摄入量随时间的变化, 食物结构 (即各类食物所占的比重) 随时间的变化, 在进行某个月的评估时调用对应月份的参数。

#### 5.1.5. 食物摄入量模型中的难点问题分析及处理

1, 如何设计抽样调查方案使调查结果能尽量反映全国的实际情况

根据统计学原理, 抽样点的选取总会存在抽样误差, 只有当样本空间趋近于无穷大时, 才能真实的反映客观事实, 而这种理想情况在操作上是不能实现的。在设计抽样模型时, 我们希望用一次普查调查的结果就能够反映各乡镇、各县市、各省区和全国的情况, 这就要求样本需要覆盖各种类型的膳食结构和变化迥异的膳食摄入量, 因此在抽样点的选取过程中, 模型考虑了气候、地理、城乡差别、经济发展水平、不同年龄阶段的人群以及民族生活习惯, 以尽可能的反映实际情况。

为了使同时期采集的样本可重用同时减少工作量, 抽样点的样本分配模型中充分考虑了全国性的人口普查所提供的人口结构、数量的特点, 按照一定的比例在各种类型的区域内分配样本数量。在模型的数据统计部分证明了按照自然男女

比例构造的样本空间可以直接混合以代替混合人群样本空间进行统计、分析而不会改变原有的概率密度函数。因此，若希望统计某个地理区域（大区）的食物摄入量模型时，只需将处于本大区内所有的样本混合，即可用于对本大区内的情况进行统计和分析。

在模型的数据统计部分的证明中还提供了一种方法，不需要按照比例分配样本数量，在得到每个基本元单位（大区下的经济类型区域）的食物摄入量概率密度函数后，把元单位的概率密度函数按照本元单位人口占其上一级单位（在这里是大区）比例进行线性组合也可以代表其上一层次（在这里是大区）的概率密度函数。

2，中国居民消费的食品种类比其他国家居民消费的食品种类复杂得多，细分将达数千种以上，在实际调查过程中进行如此详细的分类，其调查工作量太大，而如果随意粗糙进行分类，则将影响调查的精度，因此需要根据污染物分布模型的数据合理设计抽样调查中食物的分类方法。

对于这个问题，在问题分析的“食物分类方法分析”中已有了详细说明，根据食物摄入量模型和污染物分布模型的特点，以及食物链对污染物（特别是重金属）的富集作用，提出了进行食物分类的三条原则。

3，要通过万分之一（甚至更小）的抽样率得到的数据建立起全国比较准确的人群食品摄入量模型，因而要确定合理的技术路线，充分利用从其他一切渠道可以获得的信息。

要得到全国比较准确的人群食品摄入量模型，样本空间的大小是一个很重要的因素。在限定了抽样率的情况下，进行随机抽样是不合理的，会放大抽样误差。

通过利用气候、地理因素对全国进行分区，通过经济水平、城乡差别因素将抵御划分为6个不同类型的经济水平区域，进而根据人口普查可获得的男女比例和年龄结构，控制样本的选取，以尽可能的达到按比例均匀抽样的目的，用较少的样本数量来反映全国的平均水平。并且证明了采样的样本可以直接混合，独立抽样调查形成的概率密度函数可以进行线性运算。

## 5.2. 污染物分布模型

污染物分布模型主要是根据食品卫生检测部门日常对市场上食物的检测数据和市场上各类食品的流通量以及进出口口岸的检测数据来估计市场上各种食品的污染物含量。由于实际情况的限制，这些数据往往是抽样率很低的随机抽样数据，因此首先应当对总体的分布情况进行合理的猜测，然后基于现有的有限数据对该猜测进行检验。

### 5.2.1. 污染物分布模型的建立

假设当前分析的目标为第 $Z_d$ 区域中，第 $k$ 类污染物在第 $j$ 类食物中分布的含量百分比 $x$ 的概率密度函数。污染物的含量百分比仅能出现在区间 $[0, 1]$ 上，并且在一段时间内，一个地区的平均污染程度基本保持不变，因此可以认为该变量的算术平均数和几何平均数都固定于一个定值上。此外，食品中污染物的含量百分比应该是一个随机的量，因此，结合复杂性原理，我们可以认为污染物分布的含量百分比符合Beta分布。

基于以上的分析，我们可以得出 $Z_d$ 区域中，第 $k$ 类污染物在第 $j$ 类食物中分

布的含量百分比  $x$  的概率密度函数如下:

$$g_{j,k}(x) = \frac{1}{B_{j,k}(p_{j,k}, q_{j,k})} x^{p_{j,k}-1} (1-x)^{q_{j,k}-1} (1 \leq j \leq 9, 1 \leq k \leq N)$$

$$\text{其中 } B_{j,k}(p_{j,k}, q_{j,k}) = \int_0^1 x^{p_{j,k}-1} (1-x)^{q_{j,k}-1} dx = \frac{\Gamma(p_{j,k})\Gamma(q_{j,k})}{\Gamma(p_{j,k} + q_{j,k})} = \frac{(p_{j,k} + 1)(q_{j,k} + 1)}{(p_{j,k} + q_{j,k} + 1)}$$

令  $m = p_{j,k}, n = q_{j,k}$ , 完整的污染物含量百分比模型如下:

$$g_{j,k}(x) = \frac{(m+n+1)}{(m+1)(n+1)} x^{m-1} (1-x)^{n-1}, \text{其中 } (1 \leq j \leq 9, 1 \leq k \leq N)$$

### 5.2.2. 污染物模型随时间的变化

研究表明, 植物体内的污染物会随着季节的变化而变化, 如图 3 所示[8], 气候、土壤环境的变化都会影响到植物的新陈代谢, 进而影响到对各种物质的吸收。

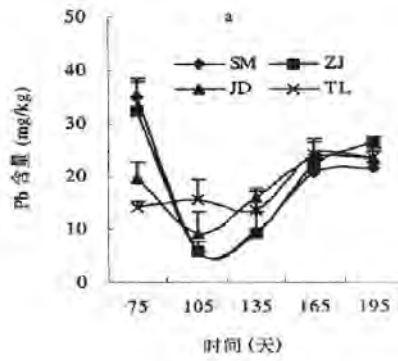


图 3 香草类植物体内铅含量随时间变化的曲线[8]

生物学上有对植物体内的铅、镉、有机磷等污染物含量随季节的变化(mg/kg)有专门的研究, 其中的污染物含量主要指均值, 如文献[8], 从各个地区的环境保护机构也比较容易获得相应的数据。如果能够采集每个希望的时间点污染物变化固然很好, 但是这个条件很难得到满足, 污染物分布模型中需要解决的问题主要是样本不匹配的问题, 这将在下一节中进行讨论。

### 5.2.3. 污染物调查中的样本在时间轴上移动的方法

由于采样的复杂性, 采样数据是抽样率很低的随机抽样数据, 并且由于食品的季节性、区域性和多样性特点, 日常检测无法获得详细的、完整的分类数据。举一种典型的应用情况, 现在需要对某地区全年谷物中的污染物分布进行评估, 但是缺少几个月份的数据, 在上一小节中已经提到, 污染物在植物体内的含量会随着季节发生很明显的变化, 这样会对全年的评估造成不良影响。

提出一种用于根据真实样本构造具有分布形状类似的样本集合的方法以解决数据不完整的问题, 其基本思想类似于连续函数中的插值。在采样这个环节中, 所采集的样本数据全部是离散化的, 和连续函数不同, 构造出的伪样本数据也应该是离散的, 并且统计上服从相似形状分布。

伪样本空间应该满足以下原则:

1, 伪样本空间中样本标志值（被测量值）的概率密度函数和真实样本空间中样本标志值具有相似的分布, 即概率密度函数的形式不发生变化;

2, 由于样本空间仅仅是在时间上发生变化, 没有地区性、人口数量、结构、生态环境发生剧烈变化的可能性, 因此, 对于污染物调查中样本分布的参数均值和标准差来说, 参数均值肯定会有变化, 但是由于整个区域内的所有包含该类污染物的食物中污染物的含量也都会随着时间一起变化, 因此, 可近似认为其标准差变化较小, 甚至不变, 特别是在相邻月份, 可近似认为其标准差不变。

根据信号处理中的时域和频域理论, 衡量一个分布的形状（在信号处理中指信号的包络线）是否发生变化, 在频域中的反映就是验证在时域中与之卷积[9]（一种物理运算）的函数在频域中分布, 值得注意的是, 正态分布函数就是频域中的抽样函数在时域中的反映, 其在频域中的响应形式就是一个常数, 因此, 可以考虑用标准差为 1 的正态分布在时域中与样本空间做卷积, 只使得样本空间的均值发生了变化, 而包络线的形状不变。

在上一小节中已经提到, 关于某种污染物在某类食物中的含量随着时间变化的均值是比较容易得到的, 在此基础上, 我们进行伪样本空间的构造。已知污染物分布的概率密度函数的包络线与正态分布比较类似, 其间也涉及到了均值、方差两个参数, 令其概率密度函数为  $f_{\mu,\sigma}(x)$ 。下面用一个均值为  $a$ , 方差为 1 的正态分布去与之做时域上的卷积, 很显然  $f_{\mu,\sigma}(x)$  和这个正态分布是独立的, 因此合成后的函数的均值即

$$\int_{-\infty}^{+\infty} x \times (f_{\mu,\sigma}(x) * \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x-a)^2}{2})) dx$$

由于  $\frac{1}{\sqrt{2\pi}} \exp(-\frac{(x-a)^2}{2})$  只是将一个标准正态分布在时间轴上平移了  $a$  个单

位, 且与  $f_{\mu,\sigma}(x)$  是独立分布的, 因此  $\int_{-\infty}^{+\infty} x \times (f_{\mu,\sigma}(x) * \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x-a)^2}{2})) dx$  最终

化简为  $a \int_{-\infty}^{+\infty} x \times f_{\mu,\sigma}(x) dx$ , 恰为原均值的  $a$  倍。

#### 5.2.4. 污染物测量中涉及的难点问题

1, 污染物测量中的数据也是抽样率很低的随机抽样数据, 应该怎么样充分利用这些数据去建立模型?

和人群食物摄入量模型面临的情况类似, 在样本比较少时, 根据样本的概率统计进行概率密度函数的推导是不切实际的, 应该先通过理论推导出分布的具体形式, 然后用少量样本去检验分布的正确性以及拟和参数。

2, 食品的季节性、区域性、多样性特点, 日常检测无法获得详细的、完整的分类数据, 如何利用这批数据去提高精度?

在上一小节中已经提出了一种方法, 即用方差为 1 的正态分布和样本空间做卷积, 可以部分解决数据的完整性问题。

### 5.3. 风险评估模型

根据人群食物摄入量模型中提出的方法，利用全国人口普查所得的年龄结构、人口比例数据可以得到抽样点以上任何层次地域的各种年龄段各类食物的摄入量模型，摄入量的概率密度函数呈对数正态分布。

污染物分布模型所评估的区域和人群食物摄入量模型是一致的，即若需要评估第  $d$  大区第  $q$  类经济水平区域的食品安全卫生状况，则需要有对应地区关于各类食物、各个年龄段的食物摄入量模型和污染物分布模型。

#### 5.3.1. 矩元法污染物摄入量评估模型

在限定在对某区域  $Z_{d,q}$  进行评估的条件下，采用“矩元法”对食品安全进行评估，矩元法的思想是建立不同年龄结构人群关于各类食物的食物摄入量矩阵  $H$ ，其中的元素  $H_{i,j}$  ( $1 \leq i \leq 5, 1 \leq j \leq 9$ ) 表示第  $i$  个年龄段的人摄入第  $j$  类食物的重量分布的概率密度函数，即  $f_{i,j}(x)$ ；建立记录各种污染物在各种类别的食物中分布的概率密度矩阵  $C$ ，其中的元素  $C_{j,k}$  ( $1 \leq j \leq 9, 1 \leq k \leq N$ ) 表示第  $k$  类污染物在第  $j$  类食物中百分比含量分布的概率密度函数，即  $g_{j,k}(x)$ 。

定义这矩阵  $H$  和  $C$  之间的叉乘运算  $\times$ ，令  $H \times C = S$ ，则  $S$  是一个  $5 \times N$  的矩阵，其中  $S_{i,k}$  是关于一个新的变量第  $i$  个年龄阶段的人摄入的第  $k$  类污染物的数量分布的概率密度函数，函数形式如下：

$$s_{i,k}(z) = \sum_{j=1}^9 L_{i,j} \times \left( \int_{x=0}^{+\infty} \int_{y=0}^1 f_{i,j}(x) \times g_{j,k}(y) \times dx dy \right) \quad (7)$$
$$z = xy$$

代入  $z = xy$  的条件，消去变量  $y$ ，然后利用积分消去  $x$ ，就得到了关于  $z$  的分布函数。 $s_{i,k}(z)$  表示的物理含义是第  $i$  个年龄阶段的人摄入的第  $k$  类污染物的分布。其中涉及到了  $L_{i,j}$ ，即第  $i$  个年龄阶段的人群的食物结构，用百分数表示，这个食物结构的参数是在为建立膳食摄入量模型而进行数据采样调查时一并统计生成。

#### 5.3.2. 模型求解

由(7)可以看出，直接求解析式解是相当困难的， $f_{i,j}(x)$  已经是对数正态分布， $g_{j,k}(y)$  是贝塔分布，直接变换成这样一个解析式是比较困难的，由于时间有限，



在尝试多次未果后，我们转用更容易操作的数值解法。由于已经了解析式的表示形式，用该等式来对生成的样本数据进行鉴别。

采用蒙特卡罗模拟求解算法对二重积分进行近似求解，在进行求解之前已经完成了两个工作：①，通过对食物摄入量抽样数据进行统计，用数据拟和已经得到了  $Z_{d,q}$  区域内食物摄入量模型的参数；②，通过对测量污染物分布的抽样数据进行统计，用数据拟和已经得到了  $Z_{d,q}$  区域内各种污染物分布模型的参数。

蒙特卡罗模拟求解的基本流程是：随机的在食品摄入模型的对数正态分布函数曲线以及污染物分布模型的 Beta 分布函数曲线上取点，将这些点的横坐标相乘并将结果保存下来，在经过大量的取值与运算之后，将这些结果按值划分到不同的区域中去，并统计各个区域中的点的个数，由于在假设两个模型中的数据是没有相关性的（即完全独立）前提下，食物摄入量与污染物含量百分比的乘积即为一个污染物摄入量值，因此只要结合总的采样点个数，就可以估算出污染物摄入量落在不同区间的概率，即可估算出污染物摄入量的函数值并描绘出函数曲线。

采用蒙特卡罗模拟求解有两个明显的优点：

1，对食物摄入量模型和污染物分布模型进行原始的数据统计已经相当困难，风险评估模型中指标基本上没有直接进行测量的可能，因为污染物不是摆放在物体表面以供取样调查；

2，进行模拟求解可以生成远远超过真实采样数据的大量的数据点，充分利用了食物摄入量模型和污染物分布模型中提供的解析式函数进行数据生成，能够在一定程度上排除异常点所产生的负作用。

算法求解的流程图如图 4 所示。

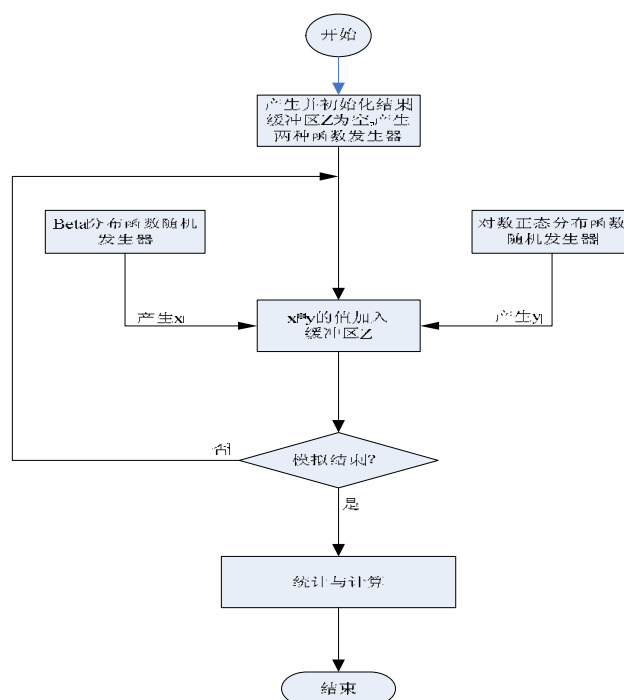


图 4 算法求解流程

下面对算法的具体工作过程予以描述。

---

**Step 1:** 初始化结果缓冲区  $Z$ ，产生两种分布函数发生器；  
**Step 2:** 从对数正态分布函数发生器中随机产生一个值  $X$  (该值的物理含义即为一个可能的食物摄入量值)，同时从 **Beta** 分布函数发生器中随机产生一个值  $Y$  (该值的物理含义为一个可能的污染物含量百分比)，将两者的乘积  $X \times Y$  (该值的含义即作为一种可能的污染物摄入量值)加入结果缓冲区  $Z$  中；  
**Step 3:** 判断模拟是否结束，若结束，记录该模拟过程产生的结果数  $N$ ，转 **Step 4** 进行统计和计算，否则，转 **Step 2**，继续模拟过程；  
**Step 4:** 该步骤根据前三个步骤得到的结果缓冲区  $Z$  中的数据以及模拟重复总次数  $N$ ，即可以估计出食物摄入量的概率分布函数。

---

统计计算方法为：假设结果缓冲区  $Z$  中的数据均分布在区间  $[a, b]$  上，则将区间  $[a, b]$  划分成  $n$  个子区间  $[a_1, b_1], [a_2, b_2] \cdots [a_k, b_k] \cdots [a_n, b_n]$ ，各区间的长度分别为  $L_1, L_2 \cdots L_k \cdots L_n$ ，将缓冲区  $Z$  中的数据按其值分别放入各个区间，并记录每个区间包含的数据个数，对于区间  $[a_k, b_k]$ ，其长度为  $L_k$ ，假设  $Z$  中有  $W_k$  个数据落在该区间上，则可以近似的认为污染物摄入量落在区间  $[a_k, b_k]$  上的概率可近似为  $\frac{W_k}{N}$ ，同理可以得出其它区间的情况，进而得到整体分布。

### 5.3.3. 右分位点的确定

由于建立污染的摄入量模型的目标是保证绝大多数居民的食品安全，因此我们最关心的问题并不仅仅是由该模型反映出来的每人每天污染物摄入量的均值，而是它的 99.999% 的右分位点。因为如果这个分位点的数值明显小于由食品卫生安全主管部门制定的，经过大量试验被证明是安全的标准，那么我们就完全有足够的理由认为当前的食品卫生状况是安全的。

因此，如何精确的找到污染物摄入模型的这个右分位点，对于食品卫生安全状况的评测具有非常重要的意义。基于前面的分析，我们知道对于污染物摄入量分布函数，直接求解析式解是相当困难的，并且即使是求出了污染物摄入量分布模型的解析函数，求其分位点有时也会非常困难，因此我们的求解方法依然基于蒙特卡罗模拟求解算法，下面是对该求解算法的详细描述。

#### 算法 3 右分位点的确定算法

---

**Step 1:** 根据蒙特卡罗模拟求解算法的模拟结果将区间划分成  $n$  个子区间  $[a_1, b_1], [a_2, b_2] \cdots [a_k, b_k] \cdots [a_n, b_n]$ ，根据前面描述的统计计算方法，计算出各个区

间上的概率值  $P_k = \frac{W_k}{N}$ 。

Step 2: 初始化  $CurSpan = 1$ ，表示从第一个区间开始累加；初始化阈值  $\theta$ ，该阈值表示分位点精度的可接受范围，即：当以点  $h$  作为有作为分布函数 99.999% 的右分位点，如果  $(P(x \leq h) - 99.999\%) < \theta$ ，则认为该点可接受，即以该点作为 99.999% 的右分位点满足精度要求；初始化  $TotalProbability = 0$ ，该变量表示当前累积的区间概率之和；初始化分位点的初始坐标为  $z = 0$ 。

Step 3: 累加当前区间的概率值即： $TotalProbability = TotalProbability + P_{CurSpan}$ （其中

$P_{CurSpan} = \frac{W_{CurSpan}}{N}$ ），令  $z = b_{CurSpan}$ （将分位点赋予该区间的右端点值），判断该分位

点是否可接收（即判断  $|TotalProbability - 99.999\%| < \theta$  是否成立），若成立，说明，

该分位点可以作为 99.999% 的右分位点，转 Step 4，否则，若  $CurSpan = n$ ，说明在当前的划分方式下，无法找到满足精度要求的右分位点，转 Step 1，对区间进行重新划分。若  $CurSpan < n$ ，则令  $CurSpan = CurSpan + 1$ ，重复 Step 3，继续寻找。

Step 4: 判断  $|TotalProbability - 99.999\%| < |TotalProbability + P_{CurSpan+1} - 99.999\%|$  是

否成立，若成立，说明  $b_{CurSpan+1}$  比  $b_{CurSpan}$  具有更高的精度，令  $z = b_{CurSpan+1}$ ，

$TotalProbability = TotalProbability + b_{CurSpan+1}$ ， $CurSpan = CurSpan + 1$ ，重复 Step 4 继

续寻找可能的更精确点。否则，说明已找到当前最精确的右分位点，转 Step 5，结束查找过程。

Step 5: 过程结束， $z$  即为满足精度要求的污染物分布函数的 99.999% 右分位点。

---

由以上的算法描述可知，该算法具有下面几个特点：

1. 如果需要该算法会经过反复迭代最后总能找到一个满足精度要求的点作为 99.999% 的右分位点；

2. 运用该算法找出的右分位点一定是当前划分条件下所能找到的满足条件的最佳分位点，这就最大限度的保证了精度要求；

3. 因为该算法从根据区间值的大小顺序从左到右依次进行搜索，而污染物分布函数的 99.999% 右分位点往往落在最右端的几个区间内，因此为了最大限度的提高右分位点的精度，可以在划分区间时，将靠右的区间划分成尽量小的区间，而靠左边的可以适当放大以减小计算的复杂性；

4. 该算法是在蒙特卡罗模拟求解算法结果的基础上建立的，因此蒙特卡罗算法模拟的精确性是该算法精确性的基础，要尽量提高蒙特卡罗算法对污染物分布曲线模拟的精确度，比较好的方法是加大模拟次数。

### 5.3.4. 风险评估模型建立过程中的难点问题

#### 1, 样本数据不匹配问题

在我们的模型中,目的是为了评估某个地区的安全食品状况,而不是评价特定的一群人的饮食安全,在进行样本选取时,只需要人群食品摄入量模型中的调查对象是人群结构中的典型集合,污染物分配模型中也是相同的情况,这种数据不匹配并不会对结果造成很大的影响。

#### 2, 模型中数据分配不匹配问题

若是食品分类方法比现在更细,那么现在统计可以基于历史数据的组合;

若是食品分类方法比现在更粗糙,可以对地区的饮食习惯进行调查,譬如调查谷物类中大米和面粉的比例,假定这一饮食结构在短期内均没有发生变化,那么也可以从历史数据中分离出相关的统计信息,只是可能会降低数据的准确度。

#### 3, 如何提高右分位点的精度问题

在 5.3.3 的右分位点的确定算法中已经对右分位点的精度做出了考虑,通过调整步长在不大大增加运算量的前提下提高精度。从模型角度考虑,为了提高右分位点的精度,若是在小区域中就把概率密度函数计算出来,然后做线性组合求大区等上一层次区域的概率密度函数,会损失精度,应该把样本混合后进行统计。

## 六、 模型求解

在模型的描述中都已经对求解方法做了详细的描述,以下求解的数据来自于参考文献上统计数的平均值,通过时间上的变换处理后,求解出两个参数,做出的曲线。由于数据太少,且不太准确,所展示的是曲线的形状和各种指标的数量级,经过和一些文献中的统计数据进行比较,基本符合实际情况。

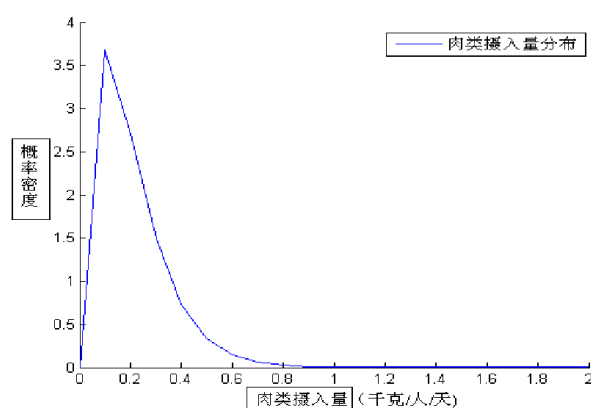


图 5 每个人每天摄入的肉类数量的概率密度函数

由上图可以看出,一个人每天摄入的肉类在 100g 左右。

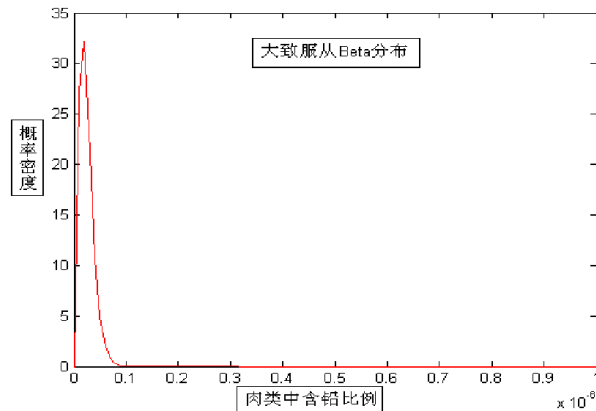


图 6 铅在肉类中的分布模型

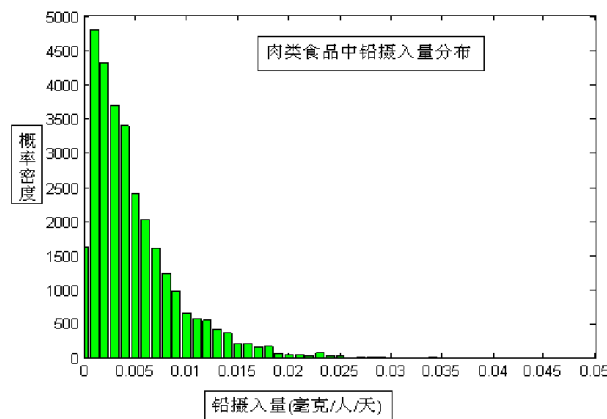


图 7 基于前两个模型的风险评估

对于铅而言，根据模拟结果，其 99.999% 的右分位点大约为 0.027mg，PWTI 的值是 0.025mg/周.千克，上面的结果是基于平均体重为 60kg 得到的数据，在该体重下的耐受能力为 0.214mg/人.天，因此，可以认为该地区的肉类食品的铅污染程度比较轻，满足食品要求。

## 七、 对当前食品安全保障的一点建议

“民以食为天，食以安为先”，如何保证食品安全是关系到国计民生的重大课题。作为一个拥有 13 亿人口的发展中国家，由于种种原因，当前中国的食品安全质量状况不容乐观：农药残留超标时有发生；各种工业环境污染物的存在及治理方面的困难；有害元素，包装材料，微生物和各种病原体的污染；有害生物和疫病多次发现；国内外人口流动的增加，新的疾病出现；食品携带疾病并传播的紧急情况时有发生。为解决以上的问题，参考国际社会和发达国家在食品安全质量控制方面积累的经验，建立完善的食品安全保障体系无疑是最为稳妥的解决方案。

### (1) 改进食品卫生安全监督管理体制和运作机制

长期以来，我国实行分散的多部门分段式食品卫生安全监管体制，一个监管环节由一个部门负责，分段监管为主，品种监管为辅。由农业部门负责初级农产

品生产环节的监管，质检部门负责食品生产加工环节的监管，卫生部门负责餐饮业和食堂等消费环节的监管，工商部门负责流通环节的监管，而食品药品监管部门负责综合监督、组织协调和依法查处重大事故。在一种食品从生产到消费的链条中，五个部门承担不同环节的监督责任。为避免出现各个部门互相推诿责任的情况，考虑建立全国统一的食品卫生安全监管模式，明确各个监管环节的责任主体，普及公众食品卫生安全知识，提高监管部门执法的信息透明度，实现“行政执法、行业自律、舆论监督、群众参与”的覆盖全社会各个层面的监管机制，

## (2) 完善食品卫生安全法律法规以及相应的安全标准体系

健全的法律体系是保障食品质量和完善监管体系的基础。目前我国还没有一个完整的食品安全监管法律体系，只有《中华人民共和国食品卫生法》，且仅仅针对食品的生产领域和流通领域。相比较而言，欧美发达国家大多建立了涵盖所有食品类别和食品链各个环节的法律法规体系，为制定监管政策、检测标准以及质量认证工作提供了有效依据。因此应加强与食品卫生安全有关的立法工作，同时对现有法律法规进行调整和协调，要从国家标准、行业标准和企业规范三个层面上来健全食品卫生安全标准体系，与国际先进模式接轨。

## (3) 推行食品卫生安全质量管理体系（生产供应链管理，信息数据库）

对食品实施“从田间到餐桌”的一整套完整连贯的立法、检测、监控、执法体系，对全国的食品从生产、加工、包装、贮运、销售和进出口各个环节严格监管。首先从生产源头抓起，通过加强环境保护和规范化肥、农药、饲料的使用，建立和完善食品原产地的环境监测网络；其次强化生产过程控制，对于食品生产企业实行危害分析与关键控制点（HACCP）生产管理和良好操作规范（GMP）管理模式；再次建立和完善市场准入制，加强对上市农产品的监测，防止不安全的产品进入市场流通；最后加大市场农产品监测力度，严格检测程序，建立食品安全信息系统，定期发布相关信息，防止不合格产品进入普通消费者家庭，达到减少食品引起的危害风险，保护消费者健康和促进食品产业以及中国外经济贸易的发展的目标。

## (4) 建立食品卫生安全风险分析体系

建立完整的风险分析体系，确定和实施适当的风险控制措施，为相关部门的决策提供技术数据，是完善整个食品卫生保障体系必不可少的关键步骤。我国目前虽已对食品生产企业设立了 A/B/C/D 四个等级标准，但这种分类方法十分粗略并且实施时由于从事安全监督和风险等级的人员缺乏相关技术背景而导致此种评估方法效果并不明显。目前国际对风险分析体系框架已达成共识，该框架包括三大部分：风险评估、风险管理和风险沟通，然后在此框架内制定食品标准。其中风险评估包括危害识别、危害的特性描述和危害显性评估三个环节；在风险评估的基础上由管理机构实行风险管理，而风险沟通贯穿于风险分析整个过程，一方面通过有效的信息发布使公众健康免于受到不安全食品的危害，另一方面将管理机构的风险分析程序向大众公开，接受舆论监督。在我国应用风险分析的好处在于，可以更好地改善食品安全的管理，改进食品生产、制作和贸易的操作规范。

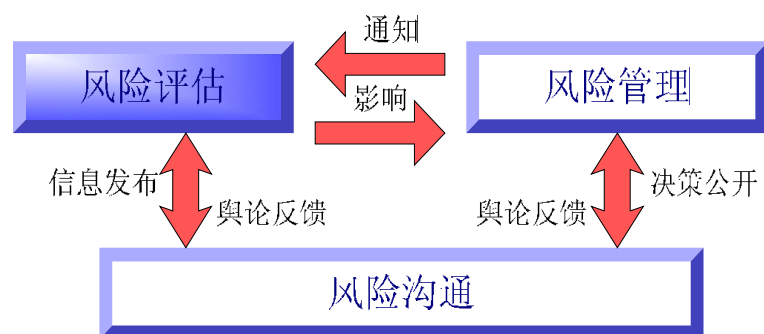


图 8 食品卫生风险分析框架

## 八、 参考文献

- [1] 曹小红等 食品安全与卫生 科学出版社 2006 年 7 月
- [2] 高俊全 李筱薇 赵京玲 2000 年中国总膳食研究-膳食铅、镉摄入量 卫生研究 第 35 卷第 6 期
- [3] <http://www.cctnet.com.cn/Html/djxx/14010981.html>
- [4] 何家骏 李奇 农业生态环境与食品安全 江苏农业科学 2007 年第 1 期
- [5] 杨晓光 孔灵芝等 中国居民营养与健康状况调查的总体方案 中华流行病学杂志 2005 年 7 月第 26 卷第 7 期
- [6] 张学文 概率分布的统一 <http://xiqxs.idm.cn/zhangxw%20web/ZCL/part3/C18b.htm>
- [7] Koch A L J Theoret Bio, 1969;23;251
- [8] 翁高艺, 孙小峰, 吴龙华 铜锌铅复合污染土壤上香蕈植物的生长和重金属吸收动态 土壤 2006, 38 (5): 602-608
- [9] <http://source.eol.cn/gjpxw/thujsj/004/text/chap1/section4/right03.htm>

## 九、 附录

附录 1:

### 消费者食品卫生安全调查问卷

填报说明:请在符合您情况的选项后的“□”内打“√”。

1. 您的性别:

先生 ☐ 女士 ☐

2. 您的年龄:

0-5 岁 ☐ 6-18 岁 ☐ 19-35 岁 ☐ 35-50 岁 ☐ 50 岁以上 ☐

**3. 您的学历:**

中学或以下 ☐ 大学 ☐ 研究生或以上 ☐

**4. 您的月收入:**

800 元以下 ☐ 800-1500 元 ☐ 1500-3000 元 ☐

3000-5000 元 ☐ 5000-10000 元 ☐ 10000 元以上 ☐

**5. 您的家庭成员共有几人?**

一人 ☐ 两人 ☐ 三人 ☐ 四人 ☐

四人以上 ☐ (请注明人数\_\_\_\_\_)

**6. 您的家庭月食品消费支出:**

100 元以下 ☐ 100-300 元 ☐ 300-500 元 ☐

500-1000 元 ☐ 1000 元以上 ☐

**7. 您所在地区:**

省 \_\_\_\_\_ 市 \_\_\_\_\_ 区/县 \_\_\_\_\_

**8. 您对食品安全关注吗?**

关注 ☐ 一般 ☐ 不关注 ☐

**9. 您做出购买决定的主要影响因素有:**

价格优先 ☐ 质量优先 ☐ 价格质量并重 ☐

**10. 您经常购买食品的场所有:**

批发市场 ☐ 农贸市场 ☐ 超市 ☐ 便利店 ☐ 马路市场 ☐ 其他 \_\_\_\_\_

**11. 您认为最安全的食品购物场所是:**

批发市场 ☐ 农贸市场 ☐ 超市 ☐ 便利店 ☐ 马路市场 ☐ 其他 \_\_\_\_\_

**12. 您家庭经常购买的食物和每日大概消耗数量 (单位 kg) 是: (多选题)**

蔬菜 ☐ \_\_\_\_\_ 水果 ☐ \_\_\_\_\_

肉禽蛋 ☐ \_\_\_\_\_ 水产 ☐ \_\_\_\_\_

粮油 ☐ \_\_\_\_\_ 调味品 ☐ \_\_\_\_\_

糖酒茶 ☐ \_\_\_\_\_ 其他 \_\_\_\_\_

**13. 您认为不安全的食品有: (多选题)**

蔬菜 ☐ 水果 ☐ 肉禽蛋 ☐ 水产 ☐

粮油 ☐ 调味品 ☐ 糖酒茶 ☐ 其他 \_\_\_\_\_

**14. 您最担心的食品安全问题是:**

蔬菜中农药高残留 ☐ 食品中违规使用添加剂 ☐ 注水肉 ☐

病死牲畜肉问题 ☐ 非食用油流入餐桌 ☐

食物中毒问题 ☐ 其他 \_\_\_\_\_

**15. 您知道负责食品安全的主要职能部门有哪些: (多选题)**

农业部门 ☐ 质量技术监督部门 ☐ 食品和药品监督部门 ☐

卫生部门 ☐ 工商部门 ☐ 公安部门 ☐ 城管部门 ☐ 其他 \_\_\_\_\_



**16. 如果买到劣质食品，您会采用以下何种方式维权？**

与经营者交涉 ☐ 向消协投诉 ☐

向有关行政部门申诉 ☐ 算了，自认倒霉 ☐ 其他 \_\_\_\_\_

**17. 您的投诉是否得到解决：**

是 ☐ 否 ☐

**18. 您对投诉解决是否满意：**

是 ☐ 否 ☐

**19. 您认为造成目前食品安全问题众多的原因是：（多选题）**

不法食品生产加工企业和个人利欲熏心 ☐

对失信企业 and 个人的惩罚力度不够 ☐

各主管部门职责不明 ☐

各执法部门沟通不畅、缺乏配合 ☐ 其他 \_\_\_\_\_

**20. 您认为提高食品安全水平最有效的做法有：（多选题）**

政府部门加强检验检疫 ☐ 曝光典型案例 ☐

对出现不安全食品的生产加工、经营企业加大处罚力度 ☐

行业自律 ☐ 优化食品安全风险评估、管理执法制度 ☐

普及科学消费知识，提高购买者的感观鉴别能力 ☐ 其他 \_\_\_\_\_

**21. 您对目前政府对食品安全监管工作力度的判断是：**

到位 ☐ 一般 ☐ 不到位 ☐

**22. 您是否了解食品卫生安全方面的立法或相关国家标准？**

十分了解 ☐ 一般了解 ☐ 不了解 ☐

**23. 您是否觉得我国应加强食品卫生安全方面的法律法规和标准制定？**

有必要 ☐ 不必要 ☐ 无所谓 ☐

**24. 按满分 100 分计，您对当前食品安全状况打多少分？**

100 分 ☐ 90 ☐ 80 ☐ 70 ☐ 60 ☐ 60 以下 ☐

**25. 您对提高食品安全监管还有哪些意见和建议：**

谢谢参与！