普通高等教育“十一五”国家级规划教材

非寿险精算原理

FEISHOUXIAN JINGSUAN YUANLI

高洪忠 编著

中国财政经济出版社

普通高等教育“十一五”国家级规划教材

非寿险精算原理

高洪忠　编著

中国财政经济出版社

图书在版编目（CIP）数据

非寿险精算原理/高洪忠编著.—北京：中国财政经济出版社，2008.10
普通高等教育“十一五”国家级规划教材
ISBN 978-7-5095-0995-1

Ⅰ. 非…　Ⅱ. 高…　Ⅲ. 保险-精算学-高等学校-教材　Ⅳ. F840.4

中国版本图书馆 CIP 数据核字(2008)第 152682 号

中国财政经济出版社出版
URL：http：//www.cfeph.cn
E-mail：cfeph @ cfeph.cn

社址：北京市海淀区阜成路甲 28 号　邮政编码：100142
发行处电话：88190406　财经书店电话：64033436
北京财经印刷厂印刷　　各地新华书店经销
787×1092 毫米　16 开　22.75 印张　555 000 字
2008 年10月第 1 版　2008 年10月北京第 1 次印刷
印数：1—3 060　定价：38.00 元
ISBN 978-7-5095-0995-1/F·0825
（图书出现印装问题，本社负责调换）
本社质量投诉电话：010-88190744

前　　言

保险公司是经营风险的企业，其经营活动以风险管理为核心，具体包括：发现风险、分离风险、度量风险、承接风险、分散风险、评估负债等。公司从经营风险中获取利润，在这一过程中，保险精算扮演着重要角色。通常将保险业务分为寿险业务和非寿险业务。非寿险业务绝大多数是短期业务，保险事故发生的频率、每次事故造成的损失额都具有较大的不确定性；寿险业务大多数是长期业务，保险业务受利率变动影响很大，其保险给付通常事先确定。非寿险包括除人寿保险之外的其他所有保险业务，如财产保险、责任保险、短期健康保险和意外伤害保险等。

相应地，保险精算也分为两大类：寿险精算和非寿险精算。非寿险精算学是应用统计学和数学等工具，研究非寿险经营过程中的费率厘定和准备金评估等数量关系的一门交叉性学科。非寿险精算虽然起步较晚，但发展很快，目前已成为一个独立的精算学分支。

本书以李晓林教授编著的《精算学原理（第二卷）——风险统计》为蓝本，借鉴了国内外多本非寿险精算学教材，并结合作者的有关研究成果编写而成。从总体上看，非寿险精算的基本内容应该包括费率厘定和准备金评估两大部分，但考虑到再保险的定价和准备金评估具有自己的特点，因此，本书把再保险作为一个独立的部分进行处理，这就使得本书由费率厘定、准备金评估和再保险三大部分组成。本书具体分为十章，其中第一章为非寿险精算学的基础数学知识，主要是概率统计的有关背景知识，包括随机变量、概率、分布函数、各类母函数、常见的连续分布和离散分布等；第二章对风险模型进行介绍，包括集合风险模型的定义、性质、计算公式，个体风险模型以及参数变动/不确定性；第三章是信度理论，介绍了信度理论的基本思想、贝叶斯信度、两个经验贝叶斯模型——Bühlmann 模型和 Bühlmann – Straub 模型；第四章是无赔款优待系统，包括该系统的定义、稳定状态分析、无赔款优待系统对索赔倾向的影响等；第五至六章是费率厘定，主要包括纯保费法和损失率法，其中第六章综合了国内外的有关教科书，对多项分析法进行了较系统的介绍；第七至八章是关于非寿险业务责任准备金的评估，具体包括未到期责任准备金评估和未决赔款准备金评估；第九章对再保险精算的基本方法进行了介绍，重点是超赔分保合约的定价和再保险未决赔款准备金评估；第十章内容相对独立，对破产理论进行了简单介绍。

在本书的撰写过程中，作者根据自己在非寿险精算领域的有关研究工作，在部分章节

加入了一些个人研究成果。为了帮助读者掌握本书各章节的有关内容，本书第二至十章每章的最后都给出了部分练习题，供读者在学习过程中参考。这些题目大多数来自历年中国精算师考试的真题。本书既可以作为保险、金融、经济等专业非寿险精算课程的教材，也可以作为考生参加精算师考试的参考用书。

本书的第一章、第十章由张宁撰写，其他各章由高洪忠撰写，高洪忠对全书进行统稿。中央财经大学中国精算研究院的硕士研究生张亚男、李建奎、山泉、曾小琛、王卉等参加了本书的校对工作，并提出了许多建议。

作　者

2008 年 10 月

目　录

第一章

基础知识

厘定费率、准备金评估、再保险安排和偿付能力分析等构成了非寿险精算的核心内容，这些内容都与一个因素密切相关，那就是损失额，即保险公司所承保的标的发生的实际损失。保险公司根据损失额核定理赔金额（或称赔付额）。损失额和理赔额之间关系密切，但是并不完全相同，它们都具有一个本质属性——“不确定性”。实际中，这种不确定性可能是“是否有损失”，或者“什么时候发生损失”，还有可能是“损失额是多大”。

概率模型是描述不确定性的常用手段。这里把未来损失额或者理赔额看做一个随机变量。从整个公司层面上看，每张保单可能提出的索赔额是一个随机变量，而这一组保单所提出的索赔额之和就是该组保单的索赔总额，这仍然是一个随机变量。本章后续内容将讨论这些随机变量的性质。

本章主要介绍非寿险精算的基础数理知识。为了方便，这里省略了很多结论的具体证明过程，这些证明可以在有关概率统计教科书中找到；同时，为了帮助读者深入了解这一章的部分内容，也给出了一些关键性证明与例题。

第一节　基本概念

一、概率论基础

概率论有严格的数学公理化体系[①]：一个随机试验的可能结果称为基本事件；所有基本事件的全体组成一个基本空间 Ω；而随机变量 X 是定义在基本空间 Ω 上的取值为实数的函数，即基本空间 Ω 中每一个点，或者说每个基本事件都对应实轴上的一个点。

① 早期的古典概率没有坚实的数学基础。1933 年苏联数学家柯尔莫哥洛夫在其《概率论基础》中第一次给出了概率的测度论定义和一套严密的公理体系。他的公理化方法成为现代概率论的基础，使概率论成为严谨的数学分支，对概率论的迅速发展起了积极的作用。

（一）分布函数（distributed function）

分布函数通常记为 $F(x)$，它表示一个随机变量 X 小于或者等于某一个固定常数的概率，即 $F(x)=Pr(X\leqslant x)$，或 $F(x)=Pr(X\leqslant x)$。如果存在多个随机变量，分布函数常常加上下标以示区分，如随机变量 Y 的分布函数记为 $F_Y(x)$。

分布函数具有如下一些性质①：

1. $0\leqslant F(x)\leqslant 1$；
2. 分布函数是单调递增的；
3. $\lim\limits_{x\to+\infty}F(x)=1$，$\lim\limits_{x\to-\infty}F(x)=0$；
4. 分布函数是右连续的，即 $F(x+0)=F(x)$。

注意：在某些文献中，把概率分布函数定义为 $G(x)=Pr(X<x)$，这样定义的分布函数仍然具有前三条性质，但不再是右连续而是左连续。

（二）密度函数（density function）

对于随机变量 X，如果存在非负可积函数 $f(x)(-\infty<x<+\infty)$，使得对任意实数 a，b（$a<b$），都有：$Pr(a<X\leqslant b)\triangleq\int_a^b f(x)dx$，则称 $f(x)$ 为 X 的概率密度函数，并称 X 为连续型随机变量。

密度函数 $f(x)$ 是分布函数的一阶导数，但注意密度函数仅仅定义在"分布函数一阶导数存在的点"上②。

（三）随机变量的支撑集（support）与众数（mode）

随机变量所有可能取值的集合称为**支撑集**，它可以是离散的，也可以是连续的。

随机变量最可能的取值称为**众数**：对于离散型随机变量来说，以最大概率出现的值就是众数；对于连续型随机变量来说，使密度函数取到最大值时的自变量取值就是众数。

（四）生存函数（survival function）

生存函数一般记为 $S(x)$，它表示随机变量 X 大于某一给定值的概率：

$$S(x)=Pr(X>x)=1-F(x)$$

生存函数具有如下性质：（1）生存函数是单调递减的；（2）$\lim\limits_{x\to+\infty}S(x)=0$；（3）生存函数右连续。

（五）风险强度（hazard rate）

风险强度一般记为 $h(x)$，它是概率密度函数与生存函数的比值：$h(x)=f(x)/S(x)$。

① 满足这些性质的函数称为分布函数，测度论中证明了"任意分布函数一定是概率空间上某随机变量的概率分布函数"。

② 从数学角度，"$F(x)$ 是绝对连续的"当且仅当"随机变量是连续型的"，也就是说 $F(x)$ 的连续性并不能保证随机变量的连续性。概率密度函数完全有可能是"阶梯状"连续的。从实用角度可以不用深究或者认为是 L－S（Lebesgue－Stieltjes）积分。

它在不同领域有不同名称，如在寿险精算中常称之为死亡效力。

因为 $h(x) = -S'(x)/S(x) = -d\ln S(x)/dx$，所以对任何 $h(x)$ 存在的区间 $(0,b)$，可以用 $h(x)$ 构造出生存函数：$S(x) = e^{-\int_0^b h(x)dx}$。

（六）随机向量

概率空间上的 n 个随机变量 $X_1, X_2, \cdots, X_n$，其整体 $X = (X_1, X_2, \cdots, X_n)$ 称为 n 维**随机向量**，或者 n 维随机变量。其中，二维随机变量可以看做平面上的随机点，而三维随机变量可以看做空间中的随机点。

对任意的 n 维随机变量 $X = (X_1, X_2, \cdots, X_n)$，称

$$F(x_1, x_2, \cdots, x_n) = Pr(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n)$$

为随机变量的联合分布函数。

$$F_{X_i}(x_i) = F(+\infty, \cdots, x_i, \cdots, +\infty) = Pr(X_1 < +\infty, \cdots, X_i < x_i, \cdots, X_n < +\infty)$$

称为 n 维随机变量关于 X_i 的边缘（或者边际）分布函数。

随机向量也分为连续型和离散型。

联合密度函数定义为联合分布函数的偏导数，边际密度函数定义为边际分布函数的偏导数，分别记为 $f(x_1, x_2, \cdots, x_n)$ 和 $f_{X_i}(x)$。联合分布函数是联合密度函数的 n 重积分，边际分布函数是边际密度函数的积分①。

如果 n 维随机向量 $(X_1, \cdots, X_n)$ 分别具有密度函数 $p_1(x_1), \cdots, p_n(x_n)$，则它们相互独立的充要条件是：$\prod_{i=1}^{n} p_i(x_i)$ 是 $(X_1, \cdots, X_n)$ 的联合密度函数。

【例 1－1】 二维随机变量 (X, Y) 服从二维正态分布，则它们的联合密度函数为：

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-u_1)^2}{\sigma_1^2} - \frac{2\rho(x-u_1)(y-u_2)}{\sigma_1\sigma_2} + \frac{(y-u_2)^2}{\sigma_2^2}\right]\right\}$$

它的边缘分布都服从正态分布：$X \sim N(u_1, \sigma_1^2)$，$Y \sim N(u_2, \sigma_2^2)$。

【例 1－2】 N 维随机变量 $X = (X_1, X_2, \cdots, X_n)^T$ 服从 n 维正态分布，它的密度函数可以写为：

$$f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}\left|\sum\right|^{1/2}}\exp\left[-\frac{1}{2}(x-U)^T \sum{}^{-1}(x-U)\right]$$

其中，$x = (x_1, x_2, \cdots, x_n)^T$；而 $U = (u_1, u_2, \cdots, u_n)^T$ 为参数向量；$\sum = (\sigma_{ij})_{n\times n}$ 是 n 阶正定矩阵，也是相关系数矩阵（参考“相关系数”）。特别的，当 $n=2$ 时，即为二维正态分布的密度函数。

【例 1－3】 多项分布。$(X_1, \cdots, X_r)$ 服从参数为 $(n; p_1, \cdots, p_r)$ 的多项分布，它的分布律为：

$$Pr[(X_1, \cdots, X_r) = (k_1, \cdots, k_r)] = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}$$

---

① 数学上通过 Borel 集定义联合密度函数，对于应用没有必要作如此抽象的刻画，并且通常情况下也不用考虑多重黎曼积分转化为累积积分的条件。事实上在 L－S 积分条件下，可以不用考虑如此复杂，只需要 $P(x,y)$ 非负可积即可。

其中，$k_1,\cdots,k_r$ 为非负整数且 $\sum_1^r k_i = n, \sum_1^r p_i = 1$。特别的，当 $r=2$ 时为二项分布（见“离散分布”）。

（七）数学期望（expectation）

数学期望定义为 $\int_{-\infty}^{+\infty} xdF(x)$，记为 $E(x)$，是 $L-S$（Lebesgue - Stieltjes）积分：如果随机变量是离散形式的，则期望可以写为求和符号形式；如果连续型随机变量的密度函数存在并连续，则它的数学期望可以写为：$\int_{-\infty}^{+\infty} xf(x)dx$ ①。

数学期望还可以这样计算：

$$\int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} S(x)dx = \int_{-\infty}^{+\infty} [1-F(x)]dx$$

证明：设 $F(x) = \int_{-\infty}^{x} f(t)dt$，则有

$$\int_{-\infty}^{+\infty} [1-F(x)]dx = \int_{-\infty}^{+\infty} \left[\int_{x}^{+\infty} f(t)dt\right]dx = \int_{-\infty}^{+\infty} \left[\int_0^t f(t)dx\right]dt$$

$$= \int_{-\infty}^{+\infty} tf(t)dt = E(x)$$

在证明过程中进行了积分顺序交换。证毕。

设 $H(x)$ 是定义在实数域上的实值函数，一般如果自变量是随机变量，则 $H(x)$ 也是随机变量，定义它的数学期望 $E(H(x))$：$E[H(x)] = \int_{-\infty}^{+\infty} H(x)dF(x)$。积分是 $L-S$（Lebesgue - Stieltjes）积分，当分布函数为可导函数时，定义变为普通的黎曼积分。这称为随机变量函数的数学期望，矩和母函数都可以通过这种方式定义。

数学期望具有线性性质，即期望作为一个数学算子是线性的：

$$E(a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_0 + a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

如果 $X_1,X_2,\cdots,X_n$ 相互独立，则有：

$$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$$

（八）原点矩与中心矩

设 X 是随机变量，k 是一正整数，若 $a_k = E(X^k)$ 存在，则称它是 X 的 k 阶**原点矩**；若 $b_k = E[X-E(X)]^k$ 存在，则称它是 X 的 k 阶**中心矩**。原点矩、中心矩都是随机变量的数字特征。

当 $k=1$ 时，X 的一阶原点矩 $\mu = E(X)$ 就是它的**数学期望**（expectation value）。

当 $k=2$ 时，X 的二阶中心矩 $\sigma^2 = E[X-E(X)]^2$ 就是它的**方差**（variance），而称 σ 为**标准差**（standard deviation）。

① 定义还要求该无穷积分绝对收敛，即 $\int_{-\infty}^{+\infty} |xf(x)| dx < +\infty$。

当 $k=3$ 时对应着 X 的三阶中心矩 b_3，它与标准差三次方的比值 b_3/σ^3 称为**斜度**(skewness)。

当 $k=4$ 时对应着 X 的四阶中心矩 b_4，它与标准差四次方的比值 b_4/σ^4 称为**峰度系数**(kurtosis)。

直接计算原点矩和中心矩非常复杂，采用第二节将介绍的母函数进行计算较为方便。

(九) 方差

方差是二阶中心矩：$Var(X) = \int_{-\infty}^{+\infty}(x-E(x))^2dF(x)$。经简单计算后，可得

$$Var(X)=E(X^2)-[E(X)]^2$$

方差具有如下性质：

1. a,b 为任意常数，则

$$Var(a+bX)=b^2Var(X)$$

2. 设 $a_1,a_2,\cdots,a_n$ 为任意常数，则

$$Var(\sum_{i=1}^{n}a_iX_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jE\{[X_i-E(X_i)][X_j-E(X_j)]\}$$

3. 如果随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，从上式可以直接推出

$$Var(\sum_{i=1}^{n}a_iX_i) = \sum_{i=1}^{n}a_i^2Var(X_i)$$

4. Chebyshev 不等式：如果随机变量 X 的期望、方差都存在，则对任意的 $\varepsilon>0$，有

$$Pr(|X-E(X)|\geqslant\varepsilon)\leqslant\frac{1}{\varepsilon^2}Var(X)$$

实际中，由 Chebyshev 不等式得到的上界过于宽松，没有太大应用价值。

(十) 协方差与相关系数，协方差阵与相关系数矩阵

对于两个随机变量 X,Y 而言，其数学期望 $E(X)$、$E(Y)$ 与方差 $Var(X)$、$Var(Y)$ 从不同方面描述了各自的特征，但却不能反映它们之间的关系。**协方差**以及相关系数正是能够描述两者关系的数字特征。

对于两个随机变量 X,Y，若 $Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$ 存在，则称它为 X,Y 的协方差。

协方差可以看做两个随机变量函数的数学期望。特别的，$Var(X)=Cov(X,X)$。协方差可以这样计算：

$$Cov(X,Y)=E(XY)-E(X)\cdot E(Y)$$

协方差具有如下性质：

1. $Cov(X,Y)=Cov(Y,X)$；
2. 对任意实数 a,b，$Cov(aX,bY)=abCov(X,Y)$；
3. $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$。

相关系数 $\rho(X,Y)$ 定义为：

$$\rho(X,Y)=\frac{E[(X-E(X))(Y-E(Y))]}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{Cov(X,Y)}{Cov(X,X)Cov(Y,Y)}$$

它具有如下性质：

1. 对任意的 X,Y，有 $|\rho(X,Y)|\leqslant 1$；

2. 如果 X,Y 相互独立，则 $\rho=0$；

3. $|\rho(X,Y)|=1$ 的充要条件是：存在常数 a,b（其中 $b\neq 0$），使得 $Pr(Y=a+bX)=1$，即 X,Y 线性相关。

注意：相关系数实际是线性相关系数，它只描述了 X,Y 之间的线性相关程度，而不能反映出其他相关关系。

衡量多个随机变量的关系需要用到协方差阵和相关系数矩阵。记 X 为 n 维随机向量 $(X_1,X_2,\cdots,X_n)$，若 $\sigma_{ij}=Cov(X_i,X_j)$ 都存在，其中 $i,j=1,2,\cdots,n$，则称下列矩阵为 X 的协方差阵：

$$\sum=\begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}\end{bmatrix}$$

协方差阵 $\sum$ 具有两个重要性质：（1）$\sum$ 是对称矩阵；（2）$\sum$ 是非负定矩阵。

设 $X=(X_1,X_2,\cdots,X_n)$，若 $\rho_{ij}=\rho(X_i,X_j)$ 都存在，其中 $i,j=1,2,\cdots,n$，则称下列矩阵为 X 的相关系数阵：

$$\begin{bmatrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{bmatrix}$$

（十一）条件分布、条件期望与条件方差

如果 X,Y 是离散型随机变量，对一切满足 $Pr(Y=y_k)>0$ 的 y_k，当 $Y=y_k$ 时，定义 $X=x_j$ 的条件概率为：

$$Pr(X=x_j|Y=y_k)=\frac{Pr(X=x_j,Y=y_k)}{Pr(Y=y_k)}$$

当 $Y=y_k$ 时，定义 X 的条件分布函数为：

$$F(x|y_k)=Pr(X\leqslant x|Y=y_k)=\sum_{x_j\leqslant x}Pr(X=x_j|Y=y_k)$$

当 $Y=y_k$ 时，定义 X 的条件期望为：

$$E(X|Y=y_k)=\int_{-\infty}^{+\infty}xdF(x|y_k)=\sum_j x_jPr(X=x_j|Y=y_k)$$

如果 X,Y 是连续型随机变量并且具有联合密度函数 $f(x,y)$，则对所有的 $f_Y(y)>0$，当 $Y=y$ 时，定义 X 的条件密度函数为：

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}$$

当 $Y=y$ 时，定义 X 的条件分布函数为：

$$F(x\mid y)=Pr(X\leqslant x\mid Y=y)=\int_{-\infty}^{x}f(t\mid y)\,dt$$

当 $Y=y$ 时，定义 X 的条件期望为：

$$E(X\mid Y=y)=\int_{-\infty}^{+\infty}x\,dF(x\mid y_k)=\int_{-\infty}^{+\infty}xf(x\mid y)\,dx$$

关于条件概率有全概率公式，分别考虑离散连续两种情形：

1. 如果 X,Y 是离散型随机变量，概率分布为 $X=\{x_i\}$ 和 $Y=\{y_i\}$，则对于任意事件 B，有

$$Pr(B)=\sum_i Pr(B\mid X=x_i)Pr(X=x_i)$$

其中，$Pr(B\mid X=x_i)=\sum_{y_j\in B}Pr(Y=y_i\mid X=x_i)$。

2. 如果 X,Y 是连续型随机变量，则对于任意事件 B，有

$$Pr(B)=\int_{-\infty}^{+\infty}Pr(B\mid\xi=x)f_i(x)\,dx$$

其中，$Pr\{Y\mid X=x\}=\int_B f_Y(y\mid\xi=x)\,dy$。条件期望具有如下性质：

1. 如果随机变量 X,Y 相互独立，则条件期望等价于无条件期望，即 $E(X\mid Y)=E(X)$。

2. 条件期望和普通数学期望一样，也具有线性，即对于常数 a,b，有

$$E[(aX+bY)\mid Z]=aE(X\mid Z)+bE(Y\mid Z)$$

3. 设 X,Y 为两个随机变量，$g(Y)$ 是 Y 的实值函数，则有下式成立

$$E[X\cdot g(Y)\mid Y]=g(Y)\cdot E(X\mid Y)$$

4. 如果 X,Y 相互独立，$g(x)$ 为实值函数，则有：

$$E[g(X)\mid Y]=E[g(X)]$$

5. 设 X,Y 为任意两个随机变量，$g(x)$ 为实值函数，则有：

$$E[E(g(X)\mid Y)]=g(X)$$

此公式称为**全期望公式**。条件方差的定义类似于普通方差。在随机变量 Y 取定的条件下，随机变量 X 的条件方差定义为：

$$Var(X\mid Y)=E[(X-E(X\mid Y))^2\mid Y]$$

或者

$$Var(X\mid Y)=E(X^2\mid Y)-[E(X\mid Y)]^2$$

条件方差满足下面恒等式：

$$Var(X)=E[Var(X\mid Y)]+Var[E(X\mid Y)]$$

证明：
$$\begin{aligned}Var(X)&=E(X^2)-[E(X)]^2\\&=E[E(X^2\mid Y)]-\{E[E(X\mid Y)]\}^2\\&=E[E(X^2\mid Y)]-\{E[E(X\mid Y)]\}^2\\&=E\{Var(X\mid Y)+[E(X\mid Y)]^2\}-\{E[E(X\mid Y)]\}^2\\&=E[Var(X\mid Y)]+Var[E(X\mid Y)]\end{aligned}$$

【例 1-4】 条件数学期望的应用——最佳预测问题。本书将在后面章节中介绍线性模

型，这是最佳预测的一个特例，即在线性函数类中寻找目标。如果把寻找目标的范围不限制在 X 的线性函数类中，这就是普通意义下的最佳预测，也称为一随机变量对另外一随机变量的回归。

假设 X,Y 是相联系的连续型随机变量，数学期望和方差有限；其次，假设可以通过试验对随机变量 X 进行观测。要求根据对 X 的观测值来估计随机变量 Y 的值，并且希望误差最小。

解：为解决所提出的问题，通常选择一个函数 $y=g(x)$，并以 $\hat{Y}=g(X)$ 的值作 Y 的估计（预测）值；其次，需要选一种度量来衡量 $\hat{Y}$ 的近似程度，一般用均方差表示①：

$$E(Y-\hat{Y})^2=E[Y-g(X)]^2$$

这样问题就变为寻找适当的函数 $y=g(x)$，使均方差最小，也就说明两者最接近。一般如果回归函数 $\mu(x)$ 有最小均方差，则称为最优的，即对于任意回归函数 $g(x)$，有

$$E[Y-\mu(X)]^2\leqslant E[Y-g(X)]^2$$

下面证明：$y=\mu(x)=E(Y|X)$ 是随机变量 X 对 Y 的最优回归函数。

证明：
$$\begin{aligned}E[Y-g(X)]^2&=E[Y-\mu(X)+\mu(X)-g(X)]^2\\&=E[Y-\mu(X)]^2+E[\mu(X)-g(X)]^2+2E\{[Y-\mu(X)][\mu(X)\\&\quad-g(X)]\}\end{aligned}$$

而
$$\begin{aligned}E\{[Y-\mu(X)][\mu(X)-g(Y)]\}&=E\{E\{[Y-\mu(X)][\mu(X)-g(X)]|X\}\}\\&=E\{[\mu(X)-g(X)]\{E[Y-\mu(X)]|X\}\}\end{aligned}$$

但 $E[Y-\mu(X)|X]=E(Y|X)-E[\mu(X)|X]=E(Y|X)-\mu(X)=0$

即 $E[Y-g(X)]^2\geqslant E[Y-\mu(X)]^2$

上面过程实际上证明了最优回归函数的存在性。但是一般情况下，回归函数 $\mu(x)=E(Y|X=x)$ 很难具体求出，所以人们一般在某一具体函数类中求最优回归。例如，在线性函数类中求最优回归函数，即为我们所熟知的线性回归，或者称为线性预测。

（十二）超额损失随机变量（excess loss variable）

给定阈值 d，$Pr(X>d)>0$，定义超额损失随机变量为 $Y=(X-d|X>d)$。

$(X-d|X>d)$ 是左截尾平移随机变量：左截尾是指不考虑小于 d 的值，平移是指在余下的值中要除去 d。在寿险里面它可以解释为死亡年龄，非寿险里面就可以解释为超额损失（超过某固定损失的损失）。

它的数学期望为 $E(Y)=E(X-d|X>d)$，这是一个条件期望，称为"平均超额损失函数"（mean excess loss function）。它的 k 阶矩为：

$$E_X^k(d)=\frac{\int_d^{+\infty}(x-d)^k f(x)dx}{1-F(d)}\quad\text{（如果变量是连续型的）}$$

$$=\frac{\sum_{x_j>d}(x_j-d)^k p(x_j)}{1-F(d)}\quad\text{（如果变量是离散型的）}$$

$E(Y)$，$S(x)$ 之间具有如下关系：

① 这种度量是一种抽象距离，含义直观，一般求解比较方便，在数理统计中有广泛应用。

$$E(Y)=E_X(d)=\frac{\int_d^{+\infty}(x-d)f(x)dx}{1-F(d)}$$

$$=\frac{-(x-d)S(x)\Big|_d^{+\infty}+\int_d^{+\infty}S(x)dx}{S(x)}=\frac{\int_d^{+\infty}S(x)dx}{S(x)}$$

（十三）左置零平移随机变量（left - censored and shifted random variable）

左置零平移随机变量定义为：$Y=(X-d)_+=\begin{cases}0, & X<d\\ X-d & X\geqslant d\end{cases}$

根据该定义，比 d 小的部分虽然没有被忽略但是被置为0。这里需要注意区分它与超额损失随机变量的区别：在后者中，任何比阈值低的部分没有被记录。在风险模型中，如果比阈值低的部分没有被记录就是左截尾，如果这部分被记为某一个非常小的损失（可以被视为0），就是左置零。

它的矩为 $E[(X-d)_+^k]=\int_d^{+\infty}(x-d)^k dF(x)$；

如果随机变量是连续的，有 $E[(X-d)_+^k]=\int_d^{+\infty}(x-d)^k f(x)dx$；

如果随机变量是离散的，有 $E[(X-d)_+^k]=\sum_{x_j>d}(x_j-d)^k p(x_j)$。

（十四）止损随机变量（Limited Loss random variable）

止损随机变量定义为：$Y=X\wedge u=\begin{cases}X, & X<u\\ u, & X\geqslant u\end{cases}$，表达式中的运算符 $\wedge$ 为取小运算，即求两者之间的最小值。它也被称为右截尾随机变量（right - censored random variable）：因为所有超过定值 u 的部分都被设为 u。在风险模型中，这种情况很常见，例如，经常把超过一定数额的风险（这称为自留风险）进行再保险，这样超过 u 的风险就被再保险合约所规避了，公司所面临的风险最大值就是 u。

因为 $(X\wedge u)+(X-u)_+=X$，所以可以给出一个保险策略：购买一个最大限额为 u 的保单和一个免赔额为 u 的保单就可以覆盖整个风险。

止损随机变量的期望值为 $E(X\wedge u)$，这被称为有限额的数学期望（limited expected value）：

$$E(X\wedge u)=\int_0^u xf(x)dx+u[1-F(u)]$$

止损随机变量的 k 阶矩为：

$$E[(X\wedge u)^k]=\int_{-\infty}^u x^k f(x)dx+u^k[1-F(u)]\quad（如果随机变量是连续的）$$

$$E[(X\wedge u)^k]=\sum_{x_j\leqslant u}x_j^k p(x_j)+u^k[1-F(u)]\quad（如果随机变量是离散的）$$

连续型止损随机变量的 k 阶原点矩计算过程如下：

$$E[(X\wedge u)^k]=\int_{-\infty}^0 x^k f(x)dx+\int_0^u x^k f(x)dx+u^k[1-F(u)]$$