21
世纪
保险
精算
系列
教材

Actuarial
Science

21世纪保险精算系列教材
精算师考试用书
中国人民大学风险管理与精算中心主编

Actuarial Models

精算模型

肖争艳　编著

中国人民大学出版社

Actuarial Science

21世纪保险精算系列教材
精算师考试用书
中国人民大学风险管理与精算中心主编

Actuarial Models

精算模型

肖争艳　编著

中国人民大学出版社
·北京·

图书在版编目（CIP）数据

精算模型/肖争艳编著. —北京：中国人民大学出版社，2012.12
21世纪保险精算系列教材
ISBN 978-7-300-16695-7

Ⅰ.①精… Ⅱ.①肖… Ⅲ.①保险-计算方法-教材 Ⅳ.①F840.4

中国版本图书馆CIP数据核字（2012）第274454号

21世纪保险精算系列教材
精算模型
肖争艳 编著
Jingsuan Moxing

出版发行	中国人民大学出版社		
社　　址	北京中关村大街31号	邮政编码	100080
电　　话	010－62511242（总编室）		010－62511398（质管部）
	010－82501766（邮购部）		010－62514148（门市部）
	010－62515195（发行公司）		010－62515275（盗版举报）
网　　址	http://www.crup.com.cn		
	http://www.ttrnet.com（人大教研网）		
经　　销	新华书店		
印　　刷	北京鑫丰华彩印有限公司		
规　　格	185 mm×260 mm　16开本	版　　次	2013年1月第1版
印　　张	21.25 插页1	印　　次	2013年1月第1次印刷
字　　数	445 000	定　　价	39.00元

目　录

第 1 章

Chapter 1 随机变量的基本知识

一个精算模型总是试图描述未来可能发生的随机性损失。损失的随机性有三层含义：损失发生的次数是随机的；损失发生的时间是随机的；损失的大小是随机的。对一份特定的保单而言，这三种随机性至少满足一种。而在概率论里，这三种随机性都能用相应的随机变量来刻画。为了后面叙述的方便，本章将对本书经常用到的随机变量的基本知识进行简要的概述。

1.1 概率空间、随机变量及分布函数

随机试验是概率论中的基本概念。事件可以看作随机试验的一种结果。试验的结果事先不能准确地预言，但具有三种特征：

（1）可以在相同的条件下重复进行。

（2）每次试验的结果不止一个，但预先知道试验所有可能的结果。

（3）每次试验前不能确定哪个结果会出现。

记随机试验的基本结果为 ω，称为样本点。随机试验所有可能的结果组成的集合称为试验的样本空间，记为 Ω。Ω 中的样本点也称为基本事件，样本空间 Ω 称为必然事件，空集 $\varnothing$ 为不可能事件。Ω 的子集 A 由基本事件组成，通常称为事件。在实际问题中，人们一般不会对样本空间的所有子集感兴趣，而是关注某些事件及其发生的可能性大小。我们可用下面的概念来描述上述问题。

定义 1.1 设 Ω 是一个样本空间（或任意一集合），F 是 Ω 中某些子集组成的集合。如果满足：

（1）$\Omega \in F$；

（2）若 $A \in F$，则 $A^c \in F$；

（3）若 $A_n \in F(n=1, 2, \cdots)$，则

$$\bigcup_{n=1}^{\infty} A_n \in F$$

则称 F 为 σ 代数，(Ω, F) 为可测空间，F 中的元素称为事件。

如果 F 是事件的 σ 代数，则 $\varnothing \in F$；当 $A_n \in F(n \geqslant 1)$，$\bigcap_{n=1}^{\infty} A_n \in F$。

定义 1.2 设 (Ω, F) 是可测空间，$P(\cdot)$ 是定义在 F 上的实值函数。如果：

(1) 任意 $A \in F$，$P(A) \geqslant 0$；

(2) $P(\Omega)=1$；

(3) 对两两不相容的事件 A_1，A_2，…，A_n，…（即 $A_i \cap A_j=\varnothing$，$i \neq j$），有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)$$

则称 P 是 (Ω, F) 上的概率，(Ω, F, P) 是概率空间，$P(A)$ 称为事件 A 的概率。

事件 A 的概率可理解为对事件 A 发生的可能性的度量。由定义可知，事件的概率具有如下性质：

(1) $P(\varnothing)=0$；

(2) $0 \leqslant P(A) \leqslant 1$；

(3) 若 $A \in F$，$B \in F$，且 $A \subset B$，则 $P(A) \leqslant P(B)$；

(4) 若 $A \in F$，$B \in F$，则 $P(A \cup B)=P(A)+P(B)-P(A \cap B)$；

(5) 若 $A_n \in F$ $(n \geqslant 1)$，则 $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} P(A_i)$；

(6) 若 $A_n \in F$ $(n \geqslant 1)$，且 $A_n \subset A_{n+1}$ $(n \geqslant 1)$，则 $P\left(\bigcup_{n=1}^{\infty} A_n\right)=\lim\limits_{n \to \infty} P(A_n)$；

(7) 若 $A_n \in F$ $(n \geqslant 1)$，且 $A_{n+1} \subset A_n$ $(n \geqslant 1)$，则 $P\left(\bigcap_{n=1}^{\infty} A_n\right)=\lim\limits_{n \to \infty} P(A_n)$。

定义 1.3 设 (Ω, F, P) 是概率空间，X 是定义在 Ω 上的取值于实数域 R 的函数。如果对任意实数 $x \in R$，$(\omega | X(\omega) \leqslant x) \in F$，则称 $X(\omega)$ 是 F 上的随机变量，并称 $F_X(x)=P(X \leqslant x)$ 为 X 的累积分布函数，有时简称为分布函数。

分布函数 $F(x)$ 满足下列性质：

(1) $0 \leqslant F(x) \leqslant 1$；

(2) $F(x)$ 递增；

(3) $F(x)$ 右连续，即 $\lim\limits_{x \to x_0^+} F(x)=F(x_0)$；

(4) $\lim\limits_{x \to -\infty} F(x)=0$，$\lim\limits_{x \to +\infty} F(x)=1$。

随机变量可以用来描述现实中我们感兴趣但又无法预先知道的事件，如被保险人的损失次数、索赔额和保险公司的盈余额。在概率论中，常用的随机变量有两种类型：离散型随机变量和连续型随机变量。离散型随机变量常用来描述被保险人的理赔次数的分布。离散型随机变量的概率特性除了用分布函数表示外，还可以用分布律来描述（$p_k=P(N=k)$，$k=0$，1，2，…，n）。例如，某投保汽

车险的驾驶员在一年之内发生索赔次数 N 的分布如下：

$$p_0=P(N=0)=0.8,\quad p_1=P(N=1)=0.1,\quad p_2=P(N=2)=0.075$$
$$p_3=P(N=3)=0.025,\quad p_k=P(N=k)=0,\quad k\geqslant 4$$

常见的离散分布有泊松分布、负二项分布、几何分布、二项分布等。我们将在本书附录 1 中列出这些分布的性质。

连续型随机变量常用来描述损失额的大小，连续型分布的概率特征还可以用分布密度函数来刻画。

定义 1.4　设 (Ω, F, P) 是概率空间，X 是随机变量，$F_X(x)$ 是其分布函数，如果存在函数 $f(x)$，使得

$$F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

则 $f(x)$ 称为连续型随机变量 X 的分布密度函数。

用来描述损失大小的随机变量 X 的分布密度函数通常具有下列特征：

(1) 非负性。损失额应该都是非负的，因此 $P(x\geqslant 0)=1$ 且 $x\leqslant 0$ 时，$f(x)=0$。

(2) 损失额应该是连续变化的，因而 $f(x)$ 是连续的。

(3) 损失额小的保险事故发生的可能性较大，损失额大的保险事故发生的可能性较小，但不可以忽略。从直观上说，损失额分布密度函数的尾部较厚（见图 1—1）。

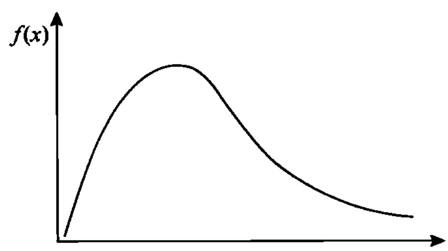

图 1—1　典型的损失额分布密度图

常用于描述损失的分布有指数分布、伽玛分布、对数正态分布、韦伯分布、帕累托分布等。本书附录 2 将列出一些常见分布的性质。

在实际问题中，对于某些随机试验的结果有时需要用多个随机变量来描述。例如，为了研究保险公司在一段时期内总理赔额的性质，不仅需要知道每次理赔额的大小，而且要知道这段时期内的理赔次数。随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 的分布不仅与随机变量 $X_1, X_2, \cdots, X_n$ 的分布有关，还依赖于 $X_1, X_2, \cdots, X_n$ 之间的相互关系。

定义 1.5　对于 n 维随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$，它的 n 维分布函数（或联合分布函数）定义为：

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

如果 $f(x_1, x_2, \cdots, x_n)=\dfrac{\partial^n F}{\partial x_1 \partial x_2 \cdots \partial x_n}$ 对所有 $(x_1, x_2, \cdots, x_n)\in R^n$ 都存在，则称 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 是连续型随机向量，函数 $f(x_1, x_2, \cdots, x_n)$ 是 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 的联合密度函数，并且

$$F(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(t_1,t_2,\cdots,t_n)\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n$$

如果 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 的所有可能取值是有限个 n 维向量或可列无穷个 n 维向量，则称 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 是离散型随机向量，称 $P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)$ 为 n 维随机向量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 的联合分布列。

定义 1.6 设 $F(x_1, x_2, \cdots, x_n)$ 是 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$ 的联合分布函数，则 $X_1, X_2, \cdots, X_n$ 的边际分布定义为：

$$\begin{aligned}&F_{k_1,k_2,\cdots,k_n}(x_{k_1},x_{k_2},\cdots,x_{k_n})\\&=F(\infty,\cdots,\infty,x_{k_1},\infty,\cdots,\infty,x_{k_2},\infty,\cdots,\infty,x_{k_n},\infty,\cdots,\infty)\end{aligned}$$

特别地，X_k 的边际分布为：

$$F_{X_k}(x_k)=F(\infty,\cdots,\infty,x_k,\infty,\cdots,\infty)$$

1.2 生存函数与危险率函数

在寿险精算中，通常用生存函数和危险率函数来描述个体的生存时间。设随机变量 X 表示个体的生存时间，即个体从初始时刻开始直至死亡、发生疾病、失效或者违约的时间。描述生存时间统计特征的基本函数是生存函数，它反映被观察个体在任意时刻 $x(x\geqslant 0)$ 仍然生存的概率，我们将其定义为：

$$S(x)=P(X>x) \tag{1—2—1}$$

当 X 为连续型随机变量时，生存函数与累积分布函数互补，即 $S(x)=1-F(x)$，这里 $F(x)=P(X\leqslant x)$，同时，生存函数也是概率密度函数 $f(x)$ 的积分，即 $S(x)=P(X>x)=\int_x^{\infty} f(y)\mathrm{d}y$，因此

$$f(x)=-\frac{\mathrm{d}S(x)}{\mathrm{d}x} \tag{1—2—2}$$

$S(x)$ 的图形叫做生存曲线，陡峭的生存曲线表示较低的生存概率或较短的生存时间，平缓的生存曲线表示较高的生存概率或较长的生存时间。

【例 1—1】

生存时间随机变量服从指数分布，其生存函数为 $S(t)=\mathrm{e}^{-\frac{1}{\theta}t}(t\geqslant 0, \theta>0)$，图 1—2 为指数生存曲线。

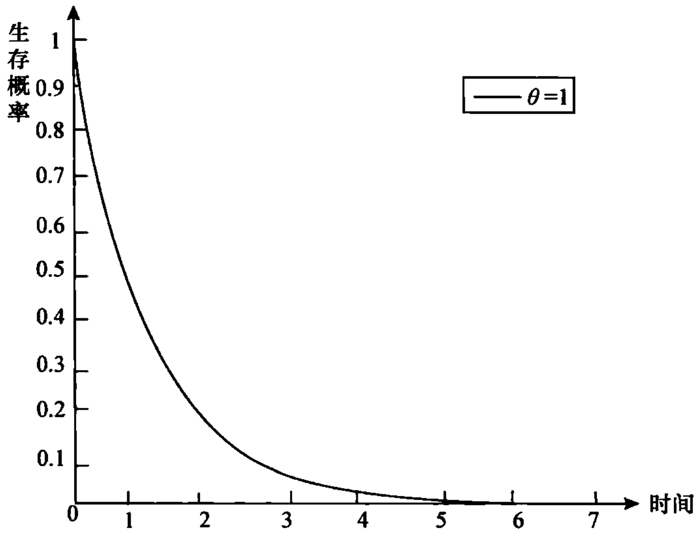

图 1—2　指数生存曲线

当 X 为离散型随机变量时，假设其概率分布函数 $p(x_j)=P(X=x_j)(j=1, 2, \cdots, n)$，其中 $x_1<x_2<\cdots<x_n$，那么 X 的生存函数为：

$$S(x)=P(X>x)=\sum_{x_j>x}p(x_j) \tag{1—2—3}$$

【例 1—2】

假设生存时间 X 服从离散均匀分布，概率分布函数为：

$$p(x_j)=P(X=j)=\frac{1}{2},\quad j=1,2$$

其生存函数为：

$$S(x)=P(X>x)=\sum_{x_j>x}p(x_j)=\begin{cases}1, & 0\leqslant x<1\\ 0.5, & 1\leqslant x<2\\ 0, & x\geqslant 2\end{cases}$$

危险率函数是生存分析中的另一个基本函数，它描述被观察个体在某时刻存活的条件下，在以后的单位时间内死亡的（条件）概率。危险率函数又称条件瞬时死亡率、死亡密度。在人口学中，它被称为死亡力，在可靠性研究中称为条件失效率。

危险率函数的定义为：

$$h(x)=\frac{f(x)}{S(x)} \tag{1—2—4}$$

显然，$h(x)$ 是在生存到时刻 x 的条件下的死亡密度。

我们注意到，式（1—2—2）和式（1—2—4）都是对个体在 x 时刻死亡密度

的瞬时度量，但 $f(x)$ 只需个体在初始时刻生存即可，而 $h(x)$ 需要个体在时刻 x 生存，这也是称 $f(x)$ 是时刻 x 死亡的无条件密度，而 $h(x)$ 是条件密度的原因。

将式（1—2—2）代入式（1—2—4），有

$$h(x)=-\frac{\mathrm{d}S(x)/S(x)}{\mathrm{d}x}=-\frac{\mathrm{d}\ln S(x)}{\mathrm{d}x} \tag{1—2—5}$$

式（1—2—5）两边从 0 到 t 积分，得

$$\int_0^x h(y)\mathrm{d}y=-\ln S(x) \tag{1—2—6}$$

所以

$$S(x)=\mathrm{e}^{-\int_0^x h(y)\mathrm{d}y} \tag{1—2—7}$$

在实际应用中，还经常用到累积危险率函数，记为 $H(x)$，其定义为：

$$H(x)=\int_0^x h(y)\mathrm{d}y=-\ln S(x) \tag{1—2—8}$$

因此

$$S(x)=\mathrm{e}^{-H(x)} \tag{1—2—9}$$

与生存函数相比，危险率函数通常可以反映死亡（失效）机制更为详细的信息。因此，在概括性地描述生存数据时，危险率函数通常占据主导地位。

1.3 随机变量的数字特征

随机变量的概率分布（分布函数或密度函数）包含了随机变量的全部信息。但是在很多情况下，人们并不需要获得一个随机变量的全部信息，只需要了解它的某些特征值，如均值、最大值、最小值、与均值的离散程度、分布的对称情况。这些信息通常可以由随机变量的数字特征来反映。

定义 1.7 （1）如果 $\sum\limits_{j=1}^{\infty}|x_j|^k p_j<\infty$，取值于 $\{x_j\}$ 的离散型随机变量 X 的 k 阶原点矩定义为：

$$\mu_k=\sum_j x_j^k P(X=x_j)$$

（2）连续型随机变量 X 的 k 阶原点矩定义为：

$$\mu_k=E(X^k)=\int_{-\infty}^{\infty}x^k\mathrm{d}F(x)=\int_{-\infty}^{\infty}x^k f(x)\mathrm{d}x$$

如果$\int_{-\infty}^{\infty} |x|^k \mathrm{d}F(x) < \infty$，则 $F(x)$ 为分布函数。

特别地，$\mu_1 = E(X)$ 称为 X 的数学期望或均值，用来描述平均水平。在精算模型中，设每张保单实际赔付额为 X，则均值 $E(X)$ 通常称为纯保费，它是保费定价的基础。

定义 1.8　(1) 如果 $\sum_{j=1}^{\infty} |x_j - \mu_1|^k p_j < \infty$，取值于 $\{x_j\}$ 的离散型随机变量 X 的 k 阶中心矩定义为：

$$\mu'_k = \sum_j (x_j - \mu_1)^k P(X = x_j)$$

(2) 连续型随机变量 X 的 k 阶中心矩定义为：

$$\mu'_k = E((X - \mu_1)^k) = \int_{-\infty}^{\infty} (x - \mu_1)^k \mathrm{d}F(x) = \int_{-\infty}^{\infty} (x - \mu_1)^k f(x) \mathrm{d}x$$

特别地，当 $k=2$ 时，μ'_2 称为 X 的方差，记为 $Var(X)$，通常用 σ^2 来表示。计算方差还有一个常用的公式：

$$\mu'_2 = E(X - \mu_1)^2 = E(X^2) - \mu_1^2$$

方差的平方根称为 X 的标准差，记为 $Std(X)$，通常用 σ 来表示。它与随机变量 X 具有相同的量纲。

除了均值和方差外，还经常使用以下指标来描述随机变量的分布特征：

(1) 变异系数 (coefficient of variation)：$r = \sigma / \mu_1$。

(2) 偏度：$r_3 = \mu'_3 / \sigma^3$。

(3) 峰度：$k = \mu'_4 / \sigma^4$。

方差和标准差、变异系数都可以用来描述分布的离散程度。方差和标准差以均值为中心计算分布的离散程度，它们的优点是直观、容易理解。但是，一方面，其数值的大小取决于随机变量 X 本身水平高低的影响，也就是说与 X 的均值大小有关。X 的绝对水平高，离散程度的测度值自然也大；绝对水平低，离散程度的测度值自然也小。另一方面，它们与随机变量 X 的计量单位相同，采用不同计量单位计量的随机变量，其离散程度的测度值也不同。因此，用方差和标准差无法比较平均水平不同或计量单位不同的不同组别的随机变量与均值的离散程度。

变异系数则从相对的角度来观察差异和离散程度。由于变异系数等于标准差除以均值，分子和分母具有相同的量纲，因此消除了量纲的影响。在比较两个不同分布的差异程度时，变异系数比标准差更好。变异系数越大，说明分布的离散程度越大；变异系数越小，说明分布的离散程度越小。

偏度和峰度这两个指标用来描述分布形状。偏度是对分布偏斜方向和程度的测度。当分布对称时，离差三次方后正负离差可以相互抵消，因此偏度 $r_3 = 0$。当分布不对称时，正负离差不能抵消，就形成了正的或负的偏态。若 $r_3 > 0$，说

明正向偏差值较大，可以判断分布函数正偏或者右偏。若 $r_3<0$，说明负离差数值较大，可判断为负偏或者左偏。r_3 的绝对值越大，说明偏斜程度越大。

峰度是对分布密度为平峰或尖峰程度的测度。峰度的值通常是与标准正态分布相比较而言的。可以证明，如果随机变量服从标准正态分布，则峰度值为 3。若 $k>3$，说明分布比正态分布更尖，为尖峰分布；若 $k<3$，说明分布比正态分布更平，为平峰分布。

对于二维随机变量 (X, Y)，除了讨论 X 和 Y 的数学期望和方差，还需要讨论 X 和 Y 之间的相互关系的数字特征。

定义 1.9 称

$$E(X-E(X))E(Y-E(Y))$$

为随机变量 X 和 Y 的协方差，记为 $Cov(X, Y)$。称

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

为随机变量 X 和 Y 的相关系数。

从定义 1.9 可以推出协方差和相关系数具有如下简单性质：

(1) $Var(X+Y)=Var(X)+Var(Y)-2Cov(X,Y)$；

(2) $Cov(X,Y)=E(XY)-E(X)E(Y)$；

(3) $|\rho_{XY}|\leqslant 1$；

(4) $|\rho_{XY}|=1$ 的充要条件是，存在常数 a，b 使得 $P(Y=a+bY)=1$。

从性质 (4) 可以看出，ρ_{XY}是一个可以用来表示 X，Y 之间线性关系紧密程度的量，当 $|\rho_{XY}|$ 较大时，我们通常说 X，Y 线性相关程度较高；当 $|\rho_{XY}|$ 较小时，我们通常说 X，Y 线性相关程度较低；当 $\rho_{XY}=0$ 时，则称 X 和 Y 不相关。

1.4 随机变量的矩母函数和母函数

定义 1.10 设随机变量 X 的分布函数为 $F(x)$，其特征函数、矩母函数和母函数分别定义为：

(1) 特征函数：$\varphi_X(t)=E(e^{itX})=\int_{-\infty}^{+\infty}e^{itx}\mathrm{d}F(x)=\int_{-\infty}^{\infty}e^{itx}f(x)\mathrm{d}x$（或者 $=\sum_k e^{itx_k}P(X=x_k)$）。

(2) 矩母函数：$M_X(t)=E(e^{tX})=\int_{-\infty}^{+\infty}e^{tx}\mathrm{d}F(x)=\int_{-\infty}^{\infty}e^{tx}f(x)\mathrm{d}x$（或者 $=\sum_k e^{tx_k}P(X=x_k)$）。

(3) 母函数：$P_X(t)=E(t^X)=\int_{-\infty}^{+\infty}t^x\mathrm{d}F(x)=\int_{-\infty}^{\infty}t^x f(x)\mathrm{d}x$（或者 $=$

$\sum_k t^{x_k} P(X=x_k)$)。

从定义可以看出，若随机变量 X 的特征函数、矩母函数和母函数都存在，它们之间存在如下关系：

$$\begin{aligned} &P_X(t)=\varphi(-i\ln t)\\ &\varphi_X(t)=P_X(e^{it})\\ &M_X(t)=P_X(e^t)\\ &P_X(t)=M_X(\ln t) \end{aligned} \tag{1—4—1}$$

特征函数是一个实变量的复值函数，因为 $|e^{itx}|=1$，因此它对一切实数 t 都有定义。随机变量的特征函数通过傅立叶积分变换与分布函数建立了唯一确定的一一对应关系。因此，求随机变量的概率分布可以等价地求该随机变量的特征函数。由于特征函数只与分布函数有关，因此称为分布的特征函数。

矩母函数和母函数的被积函数都是一个实值函数，由于积分未必存在，两者对一切实数 t 未必有定义。因此，不是所有的分布函数都对应有矩母函数和母函数。但由于两者的定义中避免使用复数被积函数，因此使用起来较为方便。一般地，矩母函数常用于研究连续型随机变量的分布性质，母函数常用于研究离散型随机变量的分布性质。从定义式可以看出，两者在原点总是有定义的，即 $M_X(0)=1$，$P_X(0)=1$。假定 X 的矩母函数（或母函数）在原点的某邻域 $|t|<r$ 内存在，则在该邻域内，$M_X(t)$，$P_X(t)$ 具有如下与特征函数 $\varphi_X(t)$ 类似的性质：

（1）与分布函数具有一一对应的特征。对于具有两个不同分布的随机变量，不可能存在相同的矩母函数（母函数、特征函数）。

（2）矩母函数、特征函数和母函数还可以用来获得 X 的数字特征。它们之间存在如下关系：

$$E(X)=M'_X(0)=-i\varphi'_X(0)=P'_X(1) \tag{1—4—2}$$

$$E(X^2)=M''_X(0)=-\varphi''_X(0) \tag{1—4—3}$$

$$E(X^k)=M^k_X(0)=(-i)^k\varphi^{(k)}_X(0) \tag{1—4—4}$$

（3）若 X_1，X_2，…，X_n 为相互独立的随机变量，则独立随机变量和 $S=X_1+X_2+\cdots+X_n$ 的矩母函数为各随机变量的矩母函数（母函数、特征函数）的乘积，即

$$M_S(t)=M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t) \tag{1—4—5}$$

$$P_S(t)=P_{X_1}(t)P_{X_2}(t)\cdots P_{X_n}(t) \tag{1—4—6}$$

$$\varphi_S(t)=\varphi_{X_1}(t)\varphi_{X_2}(t)\cdots\varphi_{X_n}(t) \tag{1—4—7}$$

（4）若 X_1，X_2，…，X_n 为独立同分布的随机变量，随机变量 N 取值于正整数，且与 X_1，X_2，…，X_n 独立，则独立随机变量和 $S=X_1+X_2+\cdots+X_N$ 的矩母函数（母函数、特征函数）分别为：

$$M_S(t)=P_N(M_X(t)) \tag{1—4—8}$$
$$P_S(t)=P_N(P_X(t)) \tag{1—4—9}$$
$$\varphi_S(t)=P_N(\varphi_X(t)) \tag{1—4—10}$$

1.5 条件概率和条件期望

定义 1.11 设 (Ω, F, P) 是概率空间，事件 B 满足 $P(B)>0$，给定事件 B 发生，事件 A 发生的条件概率定义为：

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

不难验证，条件概率 $P(\cdot \mid B)$ 满足概率测度定义中的三个条件，即

(1) 对于任意事件 A，有 $P(A\mid B)\geqslant 0$。

(2) $P(\Omega\mid B)=1$。

(3) 对两两不相容的事件 $A_1, A_2, \cdots, A_n$（即 $A_i\cap A_j=\varnothing$，$i\neq j$），有

$$P(\bigcup_{i=1}^{\infty}A_i\mid B)=\sum_{i=1}^{\infty}P(A_i\mid B)$$

既然条件概率符合上述三个条件，那么对概率所证明的一些重要结果都适用于条件概率。例如，若 $A_1\in F$，$A_2\in F$，则

$$P(A_1\cup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1\cap A_2|B)$$

下面列出两个与条件概率有关的公式。

全概率公式：设 (Ω, F, P) 是概率空间，A 为事件，$B_1, B_2, \cdots, B_n$ 是 Ω 的一个划分（即满足 $B_i\in F$ $(1\leqslant i\leqslant n)$，$B_i\cap B_j=\varnothing$ $(i\neq j)$，$\bigcup_{i=1}^{n}B_i=\Omega$），且 $P(B_i)>0(1\leqslant i\leqslant n)$，则

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n)$$

在很多实际问题中，$P(A)$ 有时不易直接求得，却很容易找到 Ω 的一个划分 $B_1, B_2, \cdots, B_n$，且 $P(B_i)$ 和 $P(A\mid B_i)$ 或为已知，或容易求出，那么可以使用全概率公式求得 $P(A)$。

贝叶斯公式：设 (Ω, F, P) 是概率空间，A 为事件，$B_1, B_2, \cdots, B_n$ 是 Ω 的一个划分，且 $P(A)>0$，$P(B_i)>0$，则

$$P(B_i\mid A)=\frac{P(A\mid B_i)P(B_i)}{\sum_{j=1}^{n}P(A\mid B_j)P(B_j)},\quad i=1,2,\cdots,n$$

由条件概率很自然地引出条件概率分布的概念。

设 (X, Y) 是二维离散型随机变量，其分布律为：

$$P(X=x_i,Y=y_j)=p_{ij}, \quad i,j=1,2,\cdots$$

（X，Y）关于 X 和 Y 的边缘分布律分别为：

$$P(X=x_i)=p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij}, \quad i=1,2,\cdots$$

$$P(Y=y_i)=p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij}, \quad j=1,2,\cdots$$

定义 1.12　设（X，Y）是二维离散型随机变量，对于固定 j，若 $P(Y=y_j)>0$，则称

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}}, \quad i=1,2,\cdots$$

为在 $Y=y_j$ 条件下随机变量 X 的条件分布律。

同样，对于固定 i，若 $P(X=x_i)>0$，则称

$$P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\cdot}}, \quad j=1,2,\cdots$$

为在 $X=x_i$ 条件下随机变量 Y 的条件分布律。

固定 x_i，由于 $P(Y=y_j|X=x_i)>0$ 且 $\sum_{j=1}^{\infty}P(Y=y_j \mid X=x_i)=1$，因此 $P(\cdot|X=x_i)$ 可以看作分布列。于是可定义随机变量 Y 在 $X=x_i$ 条件下的期望为：

$$E(Y \mid X=x_i)=\sum_{j=1}^{\infty}y_jP(Y=y_j \mid X=x_i)$$

注意：$E(Y \mid X=x_i)$ 的 x_i 只是 X 的一个特定值，给定 x_i，$E(Y \mid X=x_i)$ 是一个与 x_i 有关的常数。随着 x_i 的改变，条件期望 $E(Y \mid X=x_i)$ 也在改变，因此 $E(Y \mid X=x_i)$ 可看作 x_i 的函数。事实上可以证明，存在一个定义在随机变量 X 的值空间，取值于实数的可测函数 $f(\cdot)$ 使得

$$f(x_i)=E(Y|X=x_i), \quad \forall x_i$$

于是，随机变量 Y 关于 X 的条件期望可定义为：

$$E(Y|X)=f(X)$$

显然，$E(Y \mid X)$ 也是一个随机变量，它是 X 的函数。

类似地，可定义随机变量 X 关于 Y 的条件期望。

对于连续型随机变量（X，Y），由于对任意 x，y，有 $P(X=x)=P(Y=y)=0$，因此不能直接用条件概率公式来引入条件分布函数。这时可以用极限的方法来定义。

定义 1.13　设（X，Y）是连续型随机变量。给定 y，设对任意固定的正数 ε，$P(y-\varepsilon<Y\leqslant y+\varepsilon)>0$，且若对任意实数 x，极限

$$\lim_{\varepsilon\to0+}P(X\leqslant x|y-\varepsilon<Y\leqslant y+\varepsilon)=\lim_{\varepsilon\to0+}\frac{P(X\leqslant x,y-\varepsilon<Y\leqslant y+\varepsilon)}{P(y-\varepsilon<Y\leqslant y+\varepsilon)}$$

存在，则称此极限是在 $Y=y$ 条件下 X 的条件分布函数，记为 $F_{X|Y}(x\mid y)$。

设（X，Y）的分布函数为 $F(x,y)$，分布密度函数为 $f(x,y)$，若在点（x，y）处 $f(x,y)$ 连续，边缘密度 $f_Y(y)$ 连续，且 $f_Y(y)>0$，则可以证明

$$F_{X|Y}(x\mid y)=\int_{-\infty}^{x}\frac{f(u,y)\mathrm{d}u}{f_Y(y)}$$

因此，若记 $f_{X|Y}(x\mid y)$ 为在 $Y=y$ 条件下 X 的条件分布密度函数，则

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

类似地，可以定义 $F_{Y|X}(y\mid x)$ 和 $f_{Y|X}(y\mid x)$。

给定 $Y=y$，X 的条件期望定义为：

$$E(X\mid Y=y)=\int_{-\infty}^{\infty}xf_{X|Y}(x\mid y)\mathrm{d}x$$

注意：$E(X\mid Y=y)$ 的 y 只是一个 Y 的特定值，因此给定 y，$E(X\mid Y=y)$ 是一个与 y 有关的常数。随着 y 的改变，条件期望 $E(X\mid Y=y)$ 也在改变，因此 $E(X\mid Y=y)$ 是 y 的函数。对于连续型随机变量（X，Y），也可以证明，存在一个定义在随机变量 Y 的值空间，取值于 R 的可测函数 g 使得

$$E(X|Y)=g(Y),\ \text{a. e}$$

$g(Y)$ 称为 X 关于 Y 的条件期望。由于 $E(X\mid Y)$ 是 Y 的函数，因此也是一个随机变量。

条件期望具有以下常用性质：

(1) 对任意函数 g，有 $E[g(Y)|Y]=g(Y)$；

(2) 对任意函数 f，g，有 $E[f(Y)X+g(Y)|Y]=f(Y)E(X|Y)+g(Y)$；

(3) 若 X 和 Y 相互独立，则 $E(X|Y)=E(X)$；

(4)（迭代期望法则）$E[E(X|Y)]=E(X)$；

(5) 若 $E(X\mid Y)=E(X)$，则 $Cov(X,Y)=0$。

在条件期望的基础上，给定 $Y=y$，X 的条件方差和 X 关于 Y 的条件方差分别定义为：

$$Var(X|Y=y)=E\{[X-E(X|Y=y)]^2|Y=y\}=E(X^2|Y=y)-[E(X|Y=y)]^2$$

$$Var(X|Y)=E\{[X-E(X|Y)]^2|Y\}=E(X^2|Y)-[E(X|Y)]^2$$

条件方差具有以下常用性质：

(1) 若 X 和 Y 相互独立，则 $Var(X|Y)=Var(X)$。

(2) 方差分解（decomposition of variance）：

$$Var(X)=Var[E(X|Y)]+E[Var(X|Y)] \tag{1—5—1}$$