第 1 章

CHAPTER 1

绪　论

独立分量分析(independent component analysis,ICA)是信号处理领域在20世纪90年代后期发展起来的一项新处理方法。顾名思义,它的含义是把信号分解成若干个互相独立的成分。如果信号本来就是由若干独立信源混合而成的,我们自然希望能恰好把这些信源分解开来。从原理上说,只靠单一通道观察是不可能作这样的分解的,必需借助于一组把这些信源按不同混合比例组合起来的多通道同步观察。换句话说,ICA 是属于多导信号处理的一种方法。但是把一组观察信号分解成若干独立成分,分解结果肯定不是惟一的。因此分解总要施加一些约束条件,使答案接近于所期望的结果。

ICA 的发展是和盲信源分离(blind source separation,BSS)紧密联系的。BSS 的简单含义如图 1-1 所示。它的任务是只由多通道系统的输出数据 $\boldsymbol{X}$ 来判断其输入 $\boldsymbol{S}$ 和系统的传递函数 $\boldsymbol{H}$。所谓"盲"是指原理上它不要求对 $\boldsymbol{S}$ 和 $\boldsymbol{H}$ 具有先验知识。实际上任务的解答显然不是惟一的,因此免不了还是需要一些假设。一般至少需要假设多通道输入 $\boldsymbol{S}$ 中各分量互相独立、零均值且方差为 1。不难看出,BBS 问题的提法和 ICA 十分接近,只是前者的研究范畴更宽,处理手段也更多些。

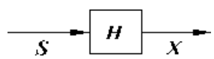

图 1-1　盲信源分解简释

在以往的信号处理技术中多导信号的分解有两项本质统一、并被熟知的技术,即主分量分析(principal component analysis,PCA)和奇异值分解(singular value decomposition,SVD),它们都建立在线性代数的基础上。其含义大致如下:

设有由 M 个通道观察值(每通道 N 点采样数据)组成的数据阵 $\boldsymbol{X}$。根据代数原理,在 $\boldsymbol{X}$ 满足一定条件的前提下,必可对它作如下分解:

$$\underset{M\times N}{\boldsymbol{X}} = \underset{M\times M}{\boldsymbol{U}}\underset{M\times N}{\boldsymbol{\Sigma}}\underset{N\times N}{\boldsymbol{V}^{\mathrm{T}}} \tag{1-1}$$

式中 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都是正交归一阵,即 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{I}_M$,$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=\boldsymbol{I}_N$($\boldsymbol{I}$ 表示单位

阵)。$\boldsymbol{\Sigma}$ 是准对角阵,在 $M<N$ 情况下

$$\boldsymbol{\Sigma}=\underbrace{\begin{bmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & \\ 0 & & \sigma_M & \end{bmatrix}}_{N} \Big\updownarrow M$$

不失一般性,通常设 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_M \geqslant 0$,称各 σ_i 为奇异值。

如果记 $\boldsymbol{U}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_M]$,$\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_N]$,则(1-1)式又可写成:

$$\boldsymbol{X}=\sum_{m=1}^{M}\sigma_m \boldsymbol{u}_m \boldsymbol{v}_m^{\mathrm{T}} \tag{1-2}$$

$\boldsymbol{u}_m$、$\boldsymbol{v}_m$ 分别称为左、右奇异矢量,其维数分别是 $M\times1$ 和 $N\times1$。

这样一来,$M\times N$ 维的原始数据就被分解成维数相同的 M 个子矩阵,这就是奇异值分解。联系生理测量,如果 M 是体表电极的数目,则 $\boldsymbol{u}_m$ 反映分解出的第 m 个子矩阵在各电极位置处的空间模式(spatial pattern),而 $\boldsymbol{v}_m$ 则反映该模式随时间变化的过程,即时间模式(temporal pattern)。

对 $\boldsymbol{X}$ 求协方差阵(假设各导记录的均值皆为 0),则有:

$$\boldsymbol{C}_X=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}} \tag{1-3}$$

式中

$$\boldsymbol{\Lambda}=\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}=\mathbf{Diag}[\sigma_1^2,\sigma_2^2,\cdots,\sigma_M^2]=\mathbf{Diag}[\lambda_1,\lambda_2,\cdots,\lambda_M]$$

$\lambda_m=\sigma_m^2$ 称为特征值。$\boldsymbol{U}$ 的每一列称为一个特征矢量。(1-3)式便是对 $\boldsymbol{C}_X$ 的主分量分解。

不论 SVD 或 PCA,分解出的分量都是按能量大小排序的($\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_M \geqslant 0$,因此 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_M \geqslant 0$)。而且,如果原始数据的秩小于 M(这意味着 M 导数据不是线性无关的),则某些奇异值和特征值将等于 0。因此,SVD 和 PCA 作为数据压缩和去除噪声的工具是很有用的;它们也是提取信号特征的一种途径。

但是按 SVD 或 PCA 原理做出的分解只能保证分解出来的各分量不相关,却不能保证这些分量互相独立(除非它们都是高斯型过程,因为对高斯型信号不相关便意味着独立)。这就使得这样的分解缺少实际(生理)意义,因而降低了所提取特征的典型性。在许多电生理测量中,观察值实际上是由若干相对独立的信源的加权和组成的,例如,胎儿心电记录中的胎儿心电(FECG)、母亲心电(MECG)和肌电;脑电记录中的自发脑电、诱发脑电及其他干扰成分(如快速眼动(REM),头皮肌电、工频干扰);心脏阻抗信号中的呼吸干扰等。在这类情况下,采用 ICA 来分解独立分量,再从各独立分量中提取有关特征,就可能会更有生理意义,有助于进一步的模式识别。因此,ICA 技术的发展引起生物医学工程界较大注意。它也在一些其他科学技术领域引起重视。实际上 ICA 理论最早就是针对所谓"鸡尾酒会问题"(从酒会的嘈杂人声中提取所关心对象的语音)这一声学问题发展起来的。在通讯领域它的应用也广为人知。例如针对码分多址(CDMA)通讯建立的模型符合含噪 ICA 的基本假设,因此可以采用 ICA 技术把各用户码分解开来。又如,在经济数据的处理中可以利用 ICA 作经济动向的预测。此外,旋转电机监测、地震预报与地质勘探、图像处理、遥测遥感等领域都有 ICA 成功应用的报导。仅从 Web of Science 上近年发表的有关 ICA 论文的检索,涉及的学科包括信息、通讯、生命、材料、电力、机械、

化学等。正是由于 ICA 的重要性，一些权威学术刊物，如 Proc. IEEE、IEEE Trans. on Biomedical Engineering、IEEE Trans. on Signal Processing、Neurocomputing 等，相继出版了有关专辑，还出版了多本有关专著(见本章附录 A)。就研究内容看，早期工作集中于信源简单线性组合的情况。近期除进一步改进这类情况的有关算法并推广其应用外，更注意探讨一些更深入难题的解决。如：各信号传递过程中有延迟或与通道发生卷积、系数时变、稀疏组合以及非线性、非平稳等情况。

图 1-2 是 ICA 最简单的框图说明。多导观察 $\boldsymbol{X}$ 是多个信源 $\boldsymbol{S}$ 经混合矩阵 $\boldsymbol{A}$ 组合而成($\boldsymbol{X}=\boldsymbol{AS}$)。现在的任务是：在 $\boldsymbol{S}$ 与 $\boldsymbol{A}$ 均为未知的条件下，求取一个解混矩阵 $\boldsymbol{B}$，使得 $\boldsymbol{X}$ 通过它后所得输出 $\boldsymbol{Y}(\boldsymbol{Y}=\boldsymbol{BX})$ 是 $\boldsymbol{S}$ 的最优逼迫。从处理技术上看，依据独立性作分解势必涉及概率密度函数或高阶统计量；而且处理过程中常要引入非线性环节。从这一意义上看，ICA 技术高于常用的只建立在二阶统计量上的线性处理技术。

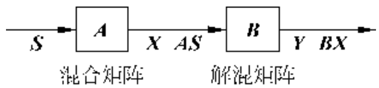

图 1-2　ICA 的简单框图说明

进行 ICA 分解的基本原则可以粗略地概括为两条：

(1) 原则一：非线性去相关　求解混阵 $\boldsymbol{B}$ 使其任意两输出 $y_i, y_j(i\neq j)$ 不但本身不相关，而且经非线性变换后的分量 $g(y_i), h(y_j)$ 也不相关。函数 g, h 自然要适当选择。

(2) 原则二：使输出尽可能非高斯化　在输出某分量 y 的方差恒定的条件下，将输入 $\boldsymbol{X}$ 各分量作线性组合 $y=\sum_i b_i x_i$ 。优化选择各权重 b_i，使 y 尽可能非高斯化。则 y 的非高斯性的每一个局部极大值给出一个独立分量(理由见 6.4 节)。

在信号处理领域内，盲信号处理(包括盲辨识、盲解卷、盲信源分解、盲通道均衡等)是 20 世纪 90 年代后期发展起来的。ICA 是它的一个组成部分。文献报导上一般把 ICA 的最早提出归功于 Jutten 与 Herault 在 1986 年和 1991 年发表在 Signal Processing 上的一组论文[1,2,3]。论文虽然比较实用，但理论上不够完善。其后，Comon 在 1994 年发表的论文[3,4]，从理论上作了比较严密的讨论。20 世纪末，随着技术的发展出现了各种不同的处理算法。其中目前被较广泛采用的有以下几个单位的研究成果：

(1) 美国加州大学生物系计算神经生物学实验室是以 Sejnowski 为学术带头人的研究室。他们提出的信息极大化(infomax，也就是 information maximization)及其扩展算法(extended infomax)已被广泛采用[5,6]。实验室的网址是：http://www.cnl.salk.edu。他们的工作成果集中在其网页的 ICA-CNL page 上(CNL 是计算神经实验室的缩写)。人们不但可以从此网址上取得该室 ICA 最新研究信息，而且可以下载有关软件。他们的工作与脑电、脑磁及磁共振图像结合得较紧。

(2) 日本 Riken 的数量神经科学实验室以 Shun-ichi Amari 为学术带头人的研究组。网址是：http://www.brain.riken.jp/lab/mns/amari。他们的工作多是结合互信息极小化(minimization of mutual information，MMI)进行的，而且主要是采用人工神经网络来达到优化目的。Amari 提出的“在学习过程中用自然梯度(natural gradient)代替常用的

最陡下降梯度”的算法可以在非正交坐标系下给出计算量小而且收敛性更好的系数递推公式[7,8]，目前也获得了广泛应用。

(3) 芬兰赫尔新基工业大学神经网络研究中心（以 Erkki Oja 为首）和赫尔辛基大学计算机科学系神经信息研究组（以 Aapo Hyvarinen 为首）的工作。他们的网址分别是：http://www.cis.hut.fi/～oja 和 http://www.cs.helsinki.fi/aapo.hyvarinen。他们的早期工作主要围绕非线性 PCA 进行，其特点是分解虽然仍是基本分量的线性组合，但组合权重中则引入非线性函数，因而实际上是把高阶统计量引入到 PCA 过程中，而不是只考虑二阶统计特征。Hyvarinen 工作中另一引人注意的亮点是：提出了一种立足于逐次提取独立分量的“固定点算法”(fixed point algorithm)[9,10]。文献中常称之为快速 ICA (fast ICA)。方法是建立在使提取信号非高斯性极大化概念上的，其计算量较小，收敛速度也较快。

除以上几个研究单位外当然还有许多其他研究组和人士在 ICA 领域进行工作，如法国学者 J. F. Cardoso（网址：http://tsi.enst.fr/～cardoso）；他提出的成批数据处理算法 JADE[11,12]，计算量虽然大些，但至今仍被人采用。又如英国格拉斯哥大学计算机科学系的 Mark Girolami（网址：http://www.dcs.gla.ac.uk/～girolami）在 ICA 发展前期也有较多贡献。2000 年编辑出版的专著：Advances in Independent Component Analysis，反映了 ICA 领域的新进展。读者还可以从网页 http://www.cnl.salk.edu/～tewon/ica_cnl.html 上链接到更多 ICA 研究人员的网页上。

ICA 涉及的基础理论和应用范围十分广泛，本书不可能完全覆盖。实际上本书主要是对 ICA 较基础的入门引导，并且偏重于其在生物医学工程领域的应用。本书的主要内容如下：除在绪论中对 ICA 问题的含义、提出、历史和现状作一简述外，全书内容大致可分为 5 个主要部分：

(1) 预备知识　简述、复习与扩展本书涉及的基础理论。其中主要是：①概率论及统计特征；②信息论基础知识；③信号经线性系统后对有关特征的影响。

(2) 优化判据　由于独立分量分析与盲信源分解密切相关，是个优化问题，因此首先讨论优化的判据。从信息论和统计估计理论上介绍已提出的各种判据，并将其纳入统一框架中，说明其个性与共性、联系与区别。

(3) 优化算法　这是本书的核心部分。将算法概括为三种主要类型：①成批数据的处理；②随着数据输入的自适应更新处理；③每次只提取一个关心成分的“逐次剥皮”处理。

(4) 进一步问题　其中包括时间延迟、卷积等情况和近年来引人注意的稀疏分量分析(scarse component analysis，SCA)。这些问题基本上属于正在发展的前沿问题。本书将简述其基本概念，介绍其各种处理算法及存在问题。

(5) 应用　主要结合生物医学信号的处理介绍信号增强、去噪、特征提取及图像分析等问题。具体内容包括：脑电中伪迹的去除、诱发响应的单次(或少次)提取、功能性磁共振图像上激活区的确定及面部图像的识别等。

全书论述注意适应工程技术人员的特点，力求从工程技术观点，使用工程技术术语来阐述问题。加强物理概念的说明，避免过于抽象，便于工程技术人员理解与自学。并且在

应用部分引入本研究室的一些研究成果。

全书章节安排请参看本书目录。参考文献分章列于各章章末。

由于作者知识水平的限制和成书过程的仓促，书中必有不少疏漏与错误，期待且欢迎读者们批评指正。

附录A 有关独立分量分析的若干专著

1 Te-won Lee. Independent Component Analysis. Kluwer Academic Press，1998
2 Gustavo Deco. An Information Theoretic Approach to Neural Computing. New York：Springer，1996
3 Aapo Hyvarinen，Juha Karhunen Erkki Oja. Independent Component Analysis. John Wiley and Sons，2001
4 Mark Girolami. Self-Organising Neural Networks：Independent Component Analysis and Blind Source Separation. Springer-Verlag London Limited，1999
5 Simon Haykin ed.. Unsupervised Adaptive Filtering：Volume I，Blind Source Separation. John Wiley and Sons，2000
6 Simon Haykin ed.. Unsupervised Adaptive Filtering：Volume II，Blind Deconvolution. John Wiley and Sons，2000
7 Stephen Roberts，Richard Everson ed.. Independent Component analysis：Principles and Practice. Cambridge University Press，2001
8 Mark Girolami ed.. Advances in Independent Component Analysis. Springer，2000
9 A. Cichocki，S-I Amari. Adaptive Blind Signal and Image Processing：Learning Algorithms and Applications. John Wiley and Sons，2002
10 J. V. Stone. Independent Component Analysis：A Tutorial Introduction. MIT Press，2004

参考文献

1 J. Herault，et al.. Space or time adaptive signal processing by neural network models，in Denker ed.："Neural Network for Computing". AIP Conference Proceedings，151，1986，American Institute of Physics
2 C. Jutten，et al.. Blind separation of sources，Pt. I：An adaptive algorithm based on neuromimetic architecture. Signal Processing，24(1)，1-10，1991
3 P. Comon，et al.. Blind separation of sources，Pt. II：Problem statement. Signal Processing，24(1)，11-20，1991
4 P. Comon et al.. Independent component analysis. A new concept?. Signal Processing，36(3)，287-314，1994
5 A. J. Bell，et al.. An information-maximization approach to blind separation and blind deconvolution. Neural Computation，7：1129-1159，1995

6 T. W. Lee, et al.. Independent component analysis using an extended infomax algorithm for sub-Gaussian and super-Gaussian sources. Neural Computation,11(2):409-433,1999

7 S.-I. Amari. New learning in structural parameter space—natural Riemannian gradient. Advances in Neural Information Processing Systems,9:127-133,1997

8 S.-I. Amari. Natural gradient works efficiently in learning. Neural Computation, 10(2): 251-276,1998

9 A. Hyvarinen, et al.. A fast fixed-point algorithm for independent component analysis. Neural Computation,9(7),1483-1492,1997

10 A. Hyvarinen. Fast and robust fixed-point algorithm for independent component analysis. IEEE Trans. on Neural Network,10(3),626-634,1999

11 J.-F. Cardoso. Higher order contrast for independent component analysis. Neural Computation, 11(1):157-193,1999

12 J.-F. Cardoso, et al.. Blind beamforming for non-Gaussian signal. IEE-Proc. F, 140(6): 362-370,1993

第2章 CHAPTER 2

预备知识

2.1 概述

从绪论所述可以看出：独立分量分析涉及的基础知识较宽。它不但像统计信号处理一样，需要概率论、随机过程以及线性代数的知识，而且还经常涉及信息论。因此，作为进一步讨论的准备，在这一章里将从以下几方面把所需知识简单总结一下：

(1) 概率与统计特征　由于在大学本科已有一定基础，因此本章叙述只是复习性质，并把特征函数和高阶统计量的有关知识予以加强。

(2) 信息论的有关概念　这是本章重点。首先介绍一些主要名词术语的含义及性质；然后联系信号通过线性系统，分析在输入、输出间这些参数的关系。

(3) 概率密度函数的级数展开　由于信息论的基本参数其定义大多涉及概率密度函数，而实际工作时它们一般是未知的，估计它们的计算又较复杂，因此人们往往把概率密度函数作级数展开，从而把对信息论特征参数的估计转化成对高阶统计量的估计。

2.2 概率与统计特征[9]

2.2.1 有关概率的复习

由于一些内容业已熟知，因此不再重复其定义，只把对连续随机变量的一些主要关系列举如下：

(1) 累积分布函数 $F(x)$　它是一个单调的升函数，其值由 0 至 1。由于是单调函数，所以必定具有一对一的可逆性。多维情况下，令 $\boldsymbol{x}=[x_1,\cdots,x_i,\cdots,x_M]$，则记作 $F(\boldsymbol{x})$。

(2) 概率密度函数(本书中常简记作 pdf) $p(x)$　是 $F(x)$ 对 x 的导数。

单变量：

$$p(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

多变量：

$$p(\boldsymbol{x}) = \frac{\partial F(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

也称“联合 pdf”。此式是纯量函数 F 对矢量变量 $\boldsymbol{x}$ 的微分，其含义见本章附录 B。

它们是非负函数。

(3) 边际 pdf，$p(x_i)$

$$p(x_i) = \int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}^i$$

其中 $\mathrm{d}\boldsymbol{x}^i$ 表示 $\mathrm{d}\boldsymbol{x}$ 中不包括 $\boldsymbol{x}_i$。例如，二维情况下有：

$$p(x_1) = \int p(x_1, x_2) \mathrm{d}x_2$$

$$p(x_2) = \int p(x_1, x_2) \mathrm{d}x_1$$

(4) 联合 pdf 与边际 pdf 间的关系　以二维情况为例，有：

$$p(x_1, x_2) = p(x_1) p(x_2 \mid x_1) = p(x_2) p(x_1 \mid x_2)$$

后一个因子称为条件 pdf。

(5) $\boldsymbol{x}$ 中各分量独立时联合 pdf 的特点

$$p(\boldsymbol{x}) = \prod_{i=1}^{M} p(x_i)$$

(6) 联合 pdf 的链形规则(chain rule)

$$p(\boldsymbol{x}) = \prod_{i=1}^{M} p(x_i \mid x_{i-1}, \cdots, x_1)$$

例如对三维情况有：

$$p(x_1, x_2, x_3) = p(x_1) p(x_2 \mid x_1) p(x_3 \mid x_2, x_1)$$

(7) 函数的 pdf　设 $y=f(x)$是单调函数，则有：

对单变量：$p_Y(y)=p_X[x=f^{-1}(y)]\left|\frac{\partial x}{\partial y}\right|=p_X[x=f^{-1}(y)]\Big/\left|\frac{\partial y}{\partial x}\right|$ *

对多变量：$p_Y(\boldsymbol{y})=p_X[\boldsymbol{x}=f^{-1}(\boldsymbol{y})]\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right|=p_X[\boldsymbol{x}=f^{-1}(\boldsymbol{y})]\Big/\left|\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right|$

式中$\left|\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right|$、$\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right|$是 $\boldsymbol{y}=f(\boldsymbol{x})$的相应 Jacobi 阵的行列式。例如，当 $\boldsymbol{x}=[x_1, x_2]$，$\boldsymbol{y}=[y_1, y_2]$时

$$\left|\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right| = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$$

如果 $\boldsymbol{y}$ 与 $\boldsymbol{x}$ 是线性关系 $\boldsymbol{y}=\boldsymbol{M}\boldsymbol{x}$，则$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}}=\boldsymbol{M}$，因而此时有：

$$p_Y(\boldsymbol{y}) = p_X(\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{y}) / |\boldsymbol{M}|$$

如果 $\boldsymbol{y}=\boldsymbol{M}\boldsymbol{x}$ 是正交归一的线性变换 $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}=\boldsymbol{I}$(也就是 $\boldsymbol{M}^{\mathrm{T}}=\boldsymbol{M}^{-1}$)，此时 $|\boldsymbol{M}|=1$，因而 $p_Y(\boldsymbol{y})=p_X(\boldsymbol{x})$。即：正交归一变换不改变 pdf。

---

* 此处 pdf 加下标记作 p_X，p_Y 是为了区别两者是不同函数。

2.2.2　特征函数

特征函数的定义如下。

对单变量(也就是一维)情况：

$$\phi(\omega)=\int p(x)\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}x=\mathrm{E}[\mathrm{e}^{\mathrm{j}\omega x}] \tag{2-1a}$$

对多变量(也就是多维)情况：

设 $\boldsymbol{x}=[x_1,x_2,\cdots,x_M]^{\mathrm{T}},\boldsymbol{\omega}=[\omega_1,\omega_2,\cdots,\omega_M]^{\mathrm{T}}$

则

$$\boldsymbol{\Phi}(\boldsymbol{\omega})=\int p(\boldsymbol{x})\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{x}=\mathrm{E}[\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}}] \tag{2-1b}$$

(2-1b)式的积分当然是多重积分。(2-1a)式实际是(2-1b)式的退化形式。

由定义可见特征函数是以 $p(\boldsymbol{x})$为 pdf 对 $\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}}$求得的均值；它也可以理解成 $p(\boldsymbol{x})$的反演傅氏变换。

对 $\phi(\boldsymbol{\omega})$取对数称为“第二特征函数 $\psi(\boldsymbol{\omega})$”，

对单变量情况有：

$$\psi(\omega)=\log\phi(\omega) \tag{2-2a}$$

对多变量情况有：

$$\psi(\boldsymbol{\omega})=\log\phi(\boldsymbol{\omega}) \tag{2-2b}$$

以下作进一步分析讨论时为了简化分析推导，在(2-1)式和(2-2)式中用 Laplace 算子 s 代替 $\mathrm{j}\omega$，把它们改写成：

$$\phi(s)=\int p(x)\mathrm{e}^{sx}\mathrm{d}x \tag{2-1c}$$

$$\phi(\boldsymbol{s})=\int p(\boldsymbol{x})\mathrm{e}^{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} \tag{2-1d}$$

式中 $\boldsymbol{s}=[s_1,s_2,\cdots,s_N]^{\mathrm{T}}$

$$\psi(s)=\log\phi(s) \tag{2-2c}$$

$$\psi(\boldsymbol{s})=\log\phi(\boldsymbol{s}) \tag{2-2d}$$

下面讨论特征函数的一些性质。

1. 特征函数在 $s=0$ 处的 Taylor 级数展开

由(2-1c)式不难得到：

单变量：

$$\phi(s)=\sum_{n=0}^{\infty}\frac{s^n}{n!}m_n \tag{2-3a}$$

式中 $m_n=\left.\dfrac{\mathrm{d}^n\phi}{\mathrm{d}s^n}\right|_{s=0}=\int x^n p(x)\mathrm{d}x=\mathrm{E}(x^n)$ 是 x 的 n 阶矩。

证明

$$\begin{aligned}\phi(s)&=\sum_{n=0}^{\infty}\frac{s^n}{n!}\left[\frac{\mathrm{d}^n\phi}{\mathrm{d}s^n}\right]_{s=0}=\sum_{n=0}^{\infty}\frac{s^n}{n!}\left.\left[\frac{\mathrm{d}^n}{\mathrm{d}s^n}\int p(x)\mathrm{e}^{sx}\mathrm{d}x\right]\right|_{s=0}\\&=\sum_{n=0}^{\infty}\frac{s^n}{n!}\int x^n p(x)\mathrm{d}x=\sum_{n=0}^{\infty}\frac{s^n}{n!}m_n\qquad(m_0=1)\end{aligned}$$

多变量：类似地在 $\boldsymbol{s}=0$ 处对(2-1d)作 Taylor 展开，可得：

$$\phi(\boldsymbol{s}) = \sum_{n_1,n_2,\cdots,n_M=0}^{\infty} \frac{s_1^{n_1} s_2^{n_2} \cdots s_M^{n_M}}{n_1!n_2!\cdots n_M!} M_{n_1,n_2,\cdots,n_M} \tag{2-3b}$$

式中 $M_{n_1,n_2,\cdots,n_M} = \left.\dfrac{\partial^n \phi(\boldsymbol{s})}{\partial \boldsymbol{s}_1^{n_1}, \partial \boldsymbol{s}_2^{n_2}, \cdots, \partial \boldsymbol{s}_M^{n_M}}\right|_{s_1=s_2=\cdots=s_M=0}$，且 $n=n_1+n_2+\cdots+n_M$。

不难看出(2-3a)式是(2-3b)式在 $n_1=n, n_2=\cdots=n_M=0$ 及 $\boldsymbol{s}=s$ 条件下的退化形式。2.2.3 小节将要说明 $M_{n_1,n_2,\cdots,n_M}$ 实际上是多变量的联合矩。(2-3)式说明 $\phi(\boldsymbol{s})$可以表示成各阶联合矩(单变量时为各阶矩)的代数和，因此特征函数又被称为“矩的生成函数”(moment generating function)。

对第二特征函数也可作类似的 Taylor 级数展开。展开结果只要把(2-3)式中的 ϕ 换成 ψ 即可：

单变量：

$$\psi(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} k_n \quad (k_0 = 0) \tag{2-4a}$$

式中

$$k_n = \left.\frac{\mathrm{d}^n \psi(s)}{\mathrm{d}s^n}\right|_{s=0}$$

多变量：

$$\psi(s) = \sum_{n_1,n_2,\cdots,n_M=0}^{\infty} \frac{s_1^{n_1} s_2^{n_2} \cdots s_M^{n_M}}{n_1!n_2!\cdots n_M!} K_{n_1,n_2,\cdots,n_M} \tag{2-4b}$$

式中

$$K_{n_1,n_2,\cdots,n_M} = \left.\frac{\partial^n \psi(s)}{\partial s_1^{n_1}, \partial s_2^{n_2}, \cdots, \partial s_M^{n_M}}\right|_{s_1=s_2=\cdots=s_M=0}$$

同样不难看出(2-4a)式是(2-4b)式在 $n_1=n, n_2=\cdots=n_M=0$ 及 $\boldsymbol{s}=s$ 条件下的退化形式。2.2.3 小节将要说明：单变量下 k_n 是 n 阶累计量；多变量下 $K_{n_1,n_2,\cdots,n_M}$ 是各阶的联合累计量。(2-4)式表明 $\psi(\boldsymbol{s})$可以表示成各阶联合累计量的代数和。因此 $\psi(\boldsymbol{s})$又被称为“累计量的生成函数”(cumulant generating function)。

2. 特征函数的一些性质

对 $\phi(s)$有：

(1) 移位特性 如 $y=x+a$，则 $\phi_y(s)=\mathrm{e}^{sa}\phi_x(s)$ (2-5)

(2) 比例特性 如 $y=ax$，则 $\phi_y(s)=\phi_x(as)$ (2-6)

(3) 叠加特性 当 x_1、x_2 互相独立时，如 $y=x_1+x_2$，则 $\phi_y(s)=\phi_{x_1}(s)\cdot\phi_{x_2}(s)$ (2-7)

(1)、(2)两特性的证明可根据定义直接得出，特性(3)的证明如下：

概率论中已知当 $y=x_1+x_2$ 时有 $p(y)=p(x_1)*p(x_2)$(符号 * 表示卷积)。对此式作反演傅氏变换(FT)，根据卷积定理便得：

$$\phi_y(x) = \phi_{x_1}(s)\cdot\phi_{x_2}(s)$$

对 $\psi(s)$有：

(1) 移位特性 如 $y=x+a$，则 $\psi_y(s)=sa+\psi_x(s)$ (2-5)′

(2) 比例特性 如 $y=ax$，则 $\psi_y(s)=\psi_x(as)$ (2-6)′

(3) 叠加特性 当 x_1、x_2 互相独立时，如 $y=x_1+x_2$，则 $\psi_y(s)=\psi_{x_1}(s)+\psi_{x_2}(s)$ (2-7)′