



中华人民共和国国家标准

GB/T 41814.1—2022

信息技术 生物特征识别校准、增强和 融合数据 第1部分：融合信息格式

Information technology—Biometric calibration, augmentation and
fusion data—Part 1: Fusion information format

(ISO/IEC 29159-1:2010, MOD)

2022-10-12发布

2023-05-01实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 符号和缩略语	2
4.1 符号	2
4.2 缩略语	2
5 符合性	2
6 融合信息格式(FIF)	3
6.1 概述	3
6.2 字节顺序	4
6.3 数值	4
6.4 融合信息头数据块	4
7 公用要素	8
7.1 概述	8
7.2 参数类型	8
7.3 参数来源	9
7.4 分布表示	10
7.5 比对次数	10
7.6 预归一化标志	10
8 第一类记录	10
8.1 目的	10
8.2 格式	11
8.3 用例(资料性)	12
9 第二类记录	12
9.1 目的	12
9.2 格式	13
9.3 用例(资料性)	14
10 第三类记录	14
10.1 目的	14
10.2 格式	14
附录 A (资料性)文件概述	17
A.1 概述	17
A.2 类型的选择	17

A.3 类型间的互用	17
A.4 扩展性	18
A.5 基于质量的融合	18
附录B (资料性) 累积分布函数示例	19
附录C (资料性) 预归一化数据的使用	20
C.1 背景	20
C.2 示例	20
C.3 最佳实践	20
附录 D (资料性) 样条函数评估源码	22
参考文献	23

前 言

本文件按照GB/T 1.1—2020《标准化工作导则第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是GB/T 41814《信息技术 生物特征识别校准、增强和融合数据》的第1部分。GB/T 41814已经发布了以下部分：

——第1部分：融合信息格式。

本文件修改采用ISO/IEC 29159-1:2010《信息技术生物特征识别校准、增强和融合数据 第1部分：融合信息格式》。

本文件与ISO/IEC 29159-1:2010相比做了下述结构调整：

- 第2章对应ISO/IEC 29159-1:2010中的第3章；
- 第3章对应 ISO/IEC 29159-1:2010中的第4章；
- 第4章对应ISO/IEC 29159-1:2010中的第5章；
- 第4章符号和缩略语分为4.1、4.2两条，分别描述符号和缩略语；
- 第5章对应ISO/IEC 29159-1:2010中的第2章。

本文件与ISO/IEC 29159-1:2010的技术差异及其原因如下：

——增加规范性引用文件 GB/T 1988—1998,用 GB/T 1988—1998规定的编码字符集替代了ASCII编码(见6.4.1、6.4.2、6.4.3),以适应我国的技术条件，增加可操作性。

本文件做了下列编辑性改动：

——表3、表4、表7、表15、表17增加表注。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：广州广电运通金融电子股份有限公司、中国电子技术标准化研究院、长春鸿达光电子与生物统计识别技术有限公司、北京眼神智能科技有限公司、厦门市熠成信息技术有限公司、上海商汤智能科技有限公司、北京中科虹霸科技有限公司、深圳市铭图创新科技有限公司、天复(东莞)标准技术有限公司、北京先见科技有限公司。

本文件主要起草人：陈建良、魏东、钟陈、王文峰、宋继伟、刘爽、杨春林、刘倩颖、石红岩、田丰、孙静、蒋慧、张玮、宋方方、校利虎、陈光、王丹丹、李清顺、王成、田立勋、杨波。

引 言

本文件定义了一种生物特征识别的融合信息格式，该格式确立了机器可读的数据格式，以描述在融合过程中输入的比对得分的统计信息。GB/T 41814《信息技术 生物特征识别校准、增强和融合数据》拟由以下部分构成。

____第1部分：融合信息格式。目的在于规定生物特征识别系统间信息融合的格式，适用于生物特征识别系统的得分级融合策略的设计和使用。

本文件定义了最常见且最易用的融合方法：得分级融合。该方法利用两个或多个系统处理个体的生物特征信息并将处理后的信息与至少一个已注册的样本进行匹配，然后输出标量比对得分。这些得分可以为真实得分(同一人)或仿冒得分(不同人)，然后设计一个融合方案将得分进行组合，以便进一步界定真实和仿冒得分之间的类边界。

比对得分的分布取决于生物特征比对子系统。得分范围和分布形状可能随系统的不同区别很大。常见的两种融合方法如下。

____在基于分类的过程中，将现有比对得分直接组合输出一个决策或得分。

____在基于归一化的过程中，融合之前需将每个得分转换到一个公共区间。基于平均值和标准偏差等统计参数的简单归一化方法有时比较有效，而复杂归一化方法需要了解整个得分分布的详细信息。本文件定义了一种融合信息格式(Fusion Information Format,FIF),用于灵活支持任何一种常见的转换形式。本文件通过建立一种标准的数据交换方法来支持一种可帮助实现生物特征识别系统整合的模块化方法。该方法中，比对和融合算法仍受各自知识产权保护。因此，本文件设想了一种应用的可能性：利用两种或更多基本的获取和比对技术(例如，手型和指纹)各生成一个得分，然后将这些得分输入已使用此处定义的相应FIF 实例进行初始化的融合模块。

概念上的多模态融合过程中各记录的逻辑角色见图1。

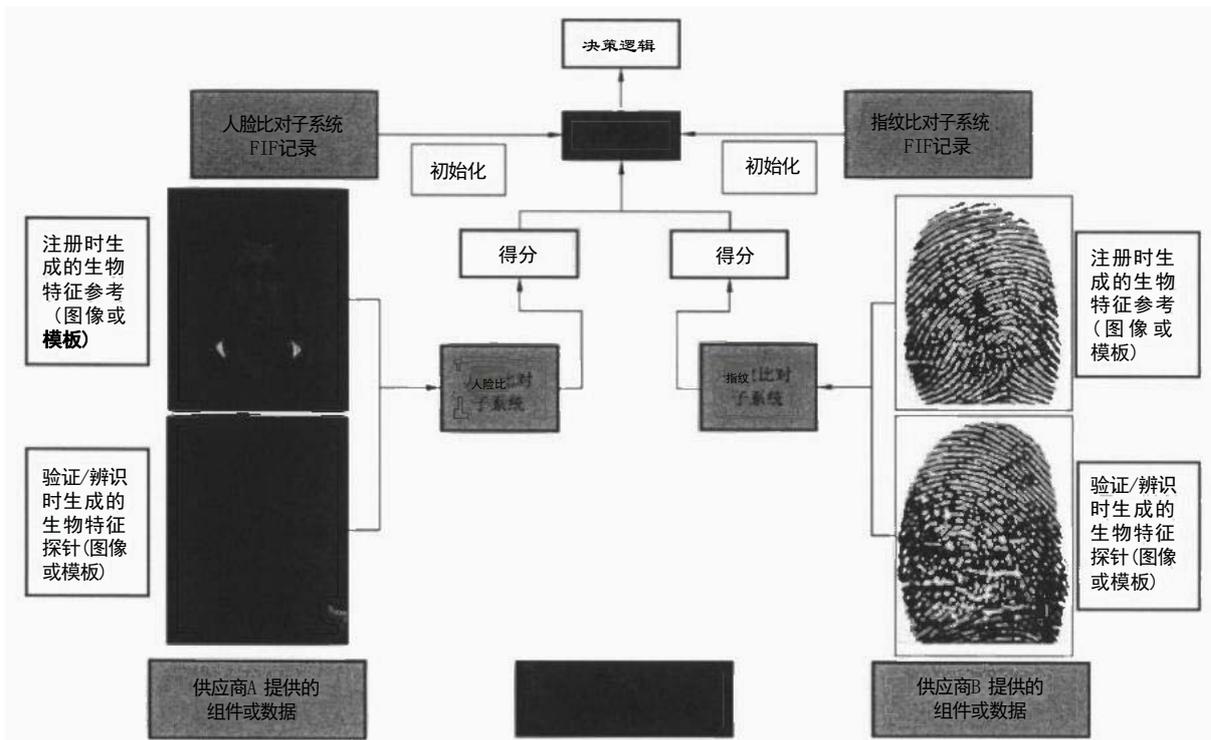


图 1 融合信息格式应用示意图

信息技术 生物特征识别校准、增强和融合数据 第1部分：融合信息格式

1 范围

本文件规定了一种机器可读的生物特征识别的融合信息格式。
本文件适用于描述在融合过程中输入的比对得分的统计信息。
本文件不适用于以下内容：
——比对得分归一化过程的标准化；
——融合过程的标准化或定义。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 1988—1998 信息技术信息交换用七位编码字符集(eqv ISO/IEC 646:1991)

GB/T 26237.1—2010 信息技术生物特征识别数据交换格式第1部分：框架” (ISO/IEC 19794-1:2006,MOD)

注：GB/T 26237.1—2010被引用的内容与 ISO/IEC 19794-1:2006被引用的内容没有技术上的差异。

GB/T 28826.1—2012 信息技术公用生物特征识别交换格式框架第1部分：数据元素规范 (ISO/IEC 19785-1:2006,MOD)

注：GB/T 28826.1—2012被引用的内容与ISO/IEC 19785-1:2006被引用的内容没有技术上的差异。

IEEE 754-2008 IEEE浮点数运算标准2(IEEE Standard for Floating-Point Arithmetic)

3 术语和定义

GB/T 26237.1—2010界定的以及下列术语和定义适用于本文件。

3.1

生物特征样本 **biometric sample**

在生物特征项提取之前的生物特征特性的模拟表示或数字表示。

注：生物特征采集设备是配备单个组件的生物特征采集子系统。

3.2

累积分布函数 **cumulative distribution function**

某变量取值小于或等于某数值的概率。

1)ISO/IEC 19794-1的最新版本为2011年修订的版本，本文件涉及的相关内容没有技术性的修订。

2)IEEE 754的最新版本为2019年修订的版本，本文件涉及的相关内容没有技术性的修订。

3.3

真实得分 genuine score

通过对同一人的两份样本进行比对而获得的比对得分。

3.4

仿冒得分 impostor score

通过对来自不同人的两份样本进行比对而获得的比对得分。

3.5

位置参数 location parameter

分布位置的一般度量。

注：位置参数并不一定是分布的平均值。

3.6

概率密度函数 probability density function

累积分布函数的导数。

3.7

尺度参数 scale parameter

分布宽度的一般度量。

注：尺度参数的分布通常不是方差或标准差。

3.8

比对得分 comparison score

生物特征比对子系统的标量输出。

注：本文件中的比对得分一般用于距离(得分越低，表示样本来自同一人的似然性越大)和相似度得分(得分越高，表示样本来自同一人的似然性越大)。6.4.9中定义的得分意义类型明确表达了二者之间的区别。

4 符号和缩略语

4.1 符号

在二进制记录结构的定义表格中，状态栏中的符号“M”表示该字段为必选字段。符号“0”表示可选字段，意味着该字段对应的字节可能为空。因为前面的字段中记录了可选字段的存在与否，在所有情况下，都可以解析记录。

4.2 缩略语

下列缩略语适用于本文件。

CBEFF: 公用生物特征交换格式框架(Common Biometric Exchange Format Framework)

CDF: 累积分布函数(Cumulative Distribution Function)

ECDF: 经验累积分布函数(Empirical Cumulative Distribution Function)

FAR: 错误接受率(False Accept Rate)

FMR: 误匹配率(False Match Rate)

FIF: 融合信息格式(Fusion Information Format)

ID: 标识符(Identifier)

PDF: 概率密度函数(Probability Density Function)

5 符合性

如果记录符合第6章的所有规范性要求，则符合本文件。对第8章、第9章或第10章三章中任何

一章的符合性都应表现为对第7章的各条的符合性。

6 融合信息格式(FIF)

6.1 概述

6.1.1 记录结构

FIF 记录用于支持多模态生物特征识别和决策支持系统的模块化。结构格式见表1。

注：应用将创建本文件的配置文件。默认的配置文件将显式调用其中一个记录类型。

表 1 FIF 记录结构

融合信息头数据块	第一类记录 (见6.1.3)	和/或	第二类记录 (见6.1.4)	和/或	第三类记录 (见6.1.5)
----------	-------------------	-----	-------------------	-----	-------------------

6.1.2 头结构

FIF 结构定义了记录的格式，并指示记录内容。结构格式见表2。

表 2 融合信息头数据块结构

融合信息头数据块	格式标识符	版本号	记录长度	生物特征类型	比对子系统产品ID	数据库ID	注册数据库质量	验证数据库质量	得分意义	类型实例数
6.4	6.4.2	6.4.3	6.4.4	6.4.5	6.4.6	6.4.7	6.4.8	6.4.8	6.4.9	6.4.10
25字节	4字节	4字节	4字节	3字节	4字节	2字节	1字节	1字节	1字节	1字节

6.1.3 第一类记录结构

FIF 结构定义了记录的格式，并指示了记录内容。结构格式见表3。

表 3 第一类记录结构

第一类记录结构	类型	分布表示	仿冒分布			真实分布		
			比对次数	位置	尺度	比对次数	位置	尺度
8.2	8.2.3	8.2.4	8.2.5	8.2.6	8.2.7	8.2.8	8.2.9	8.2.10
26字节或50字节	1字节	1字节	4字节	10字节	10字节	4字节	10字节	10字节

注：仿冒分布数据或真实分布数据存在时，有效长度为26字节；两者均存在时，有效长度为50字节。

6.1.4 第二类记录结构

FIF 结构定义了记录的格式，并指示了记录内容。结构格式见表4。

表 4 第二类记录结构

第二类记录结构	-	类型	分布表示	仿冒分布CDF	真实分布CDF
9.2		9.2.3	9.2.4	9.2.5	9.2.6
(16N+13)字节或(32N+24)字节		1字节	1字节	(16N+11)字节	(16N+11)字节

注：N表示(x, F(x))对的数量。

6.1.5 第三类记录结构

FIF 结构定义了记录的格式，并指示了记录内容。结构格式见表5。

表 5 第三类记录结构

第三类记录结构	=	类型	分布表示	仿冒分布CDF	真实分布CDF
10.2		10.2.3	10.2.4	10.2.5	10.2.6
(16N-18)字节或(32N-38)字节		1字节	1字节	(16N-20)字节	(16N-20)字节

6.2 字节顺序

在 FIF 记录内，所有定义明确的数据块中的所有多字节值应按照大端格式存储。也就是说，任何多字节量的较高有效字节存储在存储器中比较低有效字节存储地址低。

示例：1025 (210+1) 的存储形式：00000100b为第一个字节，00000001b为第二个字节。

6.3 数值

除非另有规定，否则在本文件定义类型中的所有数值均为定长的无符号整数。数值的文本表示见表6。

本文件给出的所有数值均为十进制，除非以“0x”开头表示十六进制，或以“b”后缀表示二进制。

表 6 数值的文本表示

示例值	基数	十进制值
1010b	2	10
39	10	39
0xF5	16	245

双精度数字应符合IEEE 754-2008 相关规定。

注：IEEE 754-2008的相关规范可能不足以避免数值误差。

6.4 融合信息头数据块

6.4.1 通则

表7中的融合信息头数据块应作为所有FIF 记录的第一个数据块。

表 7 融合信息头数据块

属性	状态	大小/字节	有效值	说明
格式标识符	M	4	0x46494600	带NULL终止符的符合GB/T 1988—1998要求的字符串“FIF”
版本号	M	4	0x30313000	带NULL终止符的符合GB/T 1988—1998要求的字符串“010”
记录长度	M	4	$1 \leq L \leq 2^{12} - 1$	整条记录的长度L(单位：字节)

表 7 融合信息头数据块(续)

属性	状态	大小/字节	有效值	说明
生物特征类型	M	3	$0 \leq t \leq 0x080000$	记录源模态t
比对子系统产品ID	M	4	≥ 0	当前供应商指定的ID
数据库ID	M	2	≥ 0	当前供应商指定的ID
注册样本质量	M	1	[0~100], 254, 255	用于计算比对得分统计数据的样本总体质量
验证样本质量	M	1	[0~100], 254, 255	
得分意义	M	1	0 或 1	距离或相似度。见6.4.9
类型实例数	M	1	$1 \leq N \leq 4$	字段值不得为0
注：M表示该字段本身为必选。				

6.4.2 格式标识符

格式标识符(4字节)应以GB/T 1988—1998规定的三个字符“FIF”开头，以标识符合本文件要求的记录，后跟一字节0作为NULL 字符串终止符。

6.4.3 版本号

版本号(4字节)应由GB/T 1988—1998规定的三个字符组成，后跟一字节0作为NULL 字符串终止符。该值表示本文件的版本。

第一个和第二个字符表示主版本号，第三个字符表示次版本号。
版本号应为0x30313000,即“010”——主版本号1,次版本号0。

6.4.4 记录长度

记录长度块(4字节)应为以字节为单位的整个记录的组合长度。该字段的值应为融合信息头数据块长度(25字节)加上随后第一类、第二类或者第三类记录的长度。

6.4.5 生物特征类型

该数值(3字节)应取自GB/T 28826.1—2012中6.5.6的CBEFF 生物特征识别模态的枚举。通过该字段，应用可识别出FIF 实例所代表的生物特征识别模态。

示例：对于指静脉识别模态下输出的得分，该字段的值为0x040000。

6.4.6 比对子系统产品ID

对于生成 FIF 记录所述得分信息的比对子系统算法(即用于将生物特征探针和生物特征参考进行比对的算法)，其ID(4字节=2字节+2字节)应根据表8进行记录。表中两个值均为GB/T 28826.1—2012中描述的CBEFF 产品 ID。

表 8 CBEFF产品ID

属性	配发机构	长度	说明
产品ID(存储于第一个2个字节)	生物特征识别注册机构	2字节	见GB/T 28826.1—2012
版本号(存储于第二个2个字节)	供应商	2字节	

注1:不一定要产品ID属性,因为统计信息中已包括融合模块所需的所有数据。但是,该字段对于支持版本控制和缓存非常有用。

注2:仅凭产品ID或许不足以执行得分归一化。但它对于使用本文件所述记录的应用仍然有用。

6.4.7 数据库 ID

供应商比对于子系统在生成计算FIF 中各类记录的字段4和5中所含数据的得分时使用的数据库ID(2 字节)为供应商指定标识符,应根据表9进行记录。该字段可用于如下情况:

- 比对于子系统供应商利用数据库ID 来特别指明用于FIF 记录内容计算的数据的来源;
- 应用框架或需求文档利用数据库ID 来硬性规定供应商必须采用某特定条目(比如数据库ID 为 5 5 的MINEX POEBVA指纹校准数据库)。

注: ISO/IEC JTC1 SC37针对于生成本文件记录的其他数据库制定了一个注册流程。数据库 ID注册表包括表9 所列数据库。

表 9 数据库ID

数据库ID	分布	模态
0	未指定	—
1	未知	—
2	复合	
3~15	预留	—
16	FERET人脸数据库 http://www.itl.nist.gov/iad/humanid/colorferet/	人脸
17	耶鲁人脸数据库 http://cvc.yale.edu/projects/yalefaces/yalefaces.html	人脸
18	卡内基梅隆大学PIE数据库 http://www.ri.cmu.edu/projects/project_418.html	人脸
19	AR数据库 http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html	人脸
20	FRGC I数据库 http://face.nist.gov/frvt/	人脸
21	FRGC II数据库 http://face.nist.gov/frvt/	人脸
22	FRVT 2002 HCInt(即FRVT 2006低清)数据库 http://face.nist.gov/frvt/	人脸
23	FRVT 2006高清数据库 http://face.nist.gov/frvt/	人脸
24	FRVT 2006超高清数据库 http://face.nist.gov/frvt/	人脸
25	FVC 2000	指纹
26	FVC 2002 DB1 http://bias.csr.unibo.it/fvc2002/	指纹
27	FVC 2002 DB2 http://bias.csr.unibo.it/fvc2002/	指纹
28	FVC 2002 DB3 http://bias.csr.unibo.it/fvc2002/	指纹
29	FVC 2002 DB4(合成) http://bias.csr.unibo.it/fvc2002/	指纹
30	FVC 2004 DB1 http://bias.csr.unibo.it/fvc2004/	指纹
31	FVC 2004 DB2 http://bias.csr.unibo.it/fvc2004/	指纹
32	FVC 2004 DB3 http://bias.csr.unibo.it/fvc2004/	指纹
33	FVC 2004 DB4(合成) http://bias.csr.unibo.it/fvc2004/	指纹
34	FVC 2006 DB1 http://bias.csr.unibo.it/fvc2006/	指纹
35	FVC 2006 DB2 http://bias.csr.unibo.it/fvc2006/	指纹

表 9 数据库 ID (续)

数据库ID	分布	模态
36	FVC 2006 DB3 http://bias.csr.unibo.it/fvc2006/	指纹
37	FVC 2006 DB4(合成) http://bias.csr.unibo.it/fvc2006/	指纹
38	NIST专用数据库27 http://fingerprint.nist.gov	潜指纹
39	NIST专用数据库29 http://fingerprint.nist.gov	指纹
40	MCYT指纹样本分库 http://atvs.ii.uam.es/bbdd_EN.html	指纹
41	MCYT签名样本分库 http://atvs.ii.uam.es/bbdd_EN.html	签名
42	BANCA图像 http://www.ee.surrey.ac.uk/banca/	人脸
43	BANCA视频 http://www.ee.surrey.ac.uk/banca/	声音
44	BANCA高质量 http://www.ee.surrey.ac.uk/banca/	声音
45	BANCA低质量 http://www.ee.surrey.ac.uk/banca/	声音
46	NIST语者验证 http://www.nist.gov/speech/tests/spk/2005/	声音
47	NIST语者验证 http://www.nist.gov/speech/tests/spk/2006/	声音
48	CASIA虹膜 http://www.nlpr.ia.ac.cn/english/irids/irisdatabase.htm	虹膜
49	Bath虹膜 http://www.irisbase.com	虹膜
50	ICE 2005 http://iris.nist.gov/ice/	虹膜
51	ICE 2006 http://iris.nist.gov/ice/	虹膜
52	NIST MINEX DHS2校准集 http://fingerprint.nist.gov	指纹
53	NIST MINEX POE校准集 http://fingerprint.nist.gov	指纹
54	NIST MINEX DOS校准集 http://fingerprint.nist.gov	指纹
55	NIST MINEX POEBVA校准集 http://fingerprint.nist.gov	指纹
56	TURBINE GUC100 http://www.nislabs.no/guc100	指纹
57~2047	保留	—
2048~65535	由供应商分配；供应商之间可能非公开也不具有唯一性	
<p>注1:虽然上述部分数据库并不公开,但仍将其列述于上表,便于将来校准工作的开展。</p> <p>注2:虽然上述部分数据库不易获取,但仍将其列述于上表,以供本文件使用者参考之用。</p> <p>注3:融合系统的性能与所使用的数据库稍有关联,原因在于样本匹配时从一个数据库中获取的比对得分分布通常有别于取自另一个数据库中的比对得分分布。有关分布的稳定性,见附录C。</p>		

6.4.8 数据库质量

数据库中样本的质量可能已经整合形成一个标量数据库质量值。两个(1字节)质量汇总值应分别存储于两个连续字段；第一个值面向参考样本，第二个值面向探针样本。值0应表示可能的最低质量；值100应表示可能的较高质量。该字段的取值范围，见表10。

表10 数据库质量值

取值	说明
0~100	有效赋值
254	未赋值，原因为尚未计算出具体值
255	未赋值，原因为指配质量得分失败

6.4.9 得分意义

生物特征比对子系统的原始输出可为距离或相似度。距离值越小，表示得分为真实得分的似然性越大；而相似度值恰好相反。该字段长度为1字节，用于表明得分意义，其值见表11。

表 11 得分意义编码

意义	取值(1字节)
距离	0
相似度	1

注：此处的“距离”不一定意味着值具有度量属性。

6.4.10 类型实例数

类型实例数数据块(1字节)数量应为FIF记录中第一类、第二类或第三类记录的数量。FIF记录中可能会包含或不包含每种类型的一个实例，但必定会包含其中的一个实例。因此，一条记录中可能携带1到3个类型实例。

示例：如果一条记录中含有第一类实例和第三类实例，则该字段的值为2。

注：在应用配置(或等效说明)中指明一种类型。

7 公用要素

7.1 概述

本章定义了通用数据结构，以支持本文件的第一类、第二类和第三类记录。本章还以表格形式列举出了所有值，并对这些值分别进行了说明。

要注意在某些表格中，第一列标题为“序号”，并包含一个数值。这些数值仅方便行参考，不宜将这些值包含在标准二进制记录的实例中。

7.2 参数类型

分布相关的量对应的整数标识符见表12。

表12 各统计量的标识符

类型	描述	备注
0	未指定	生产者由于疏忽未指定(通常不可接受)
1	未知	未知量
2	平均值	-

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/538055101075006110>