·自动化技术·

测量数据的信息熵与测量误差熵研究

童 玲，陈光祎禹，吕 文

(电子科技大学自动化工程学院 成都 610054)

【摘要】以Shannon信息论为基础，详细分析了测量过程的物理机制和被测量与测量结果的映射关系，建立被测量、测量误差和测量结果的信息论数学模型。研究了用信息集合和信息熵模型表征测量数据不确定度、被测量值和测量结果的方法，以及被测量信息熵、测量误差熵和测量结果信息熵的内涵、相互关系和求解方法。并推导了典型分布下传统的不确定度理论中的不确定度、置信系数与测量信息论中信息熵的数学关系，为从传统不确定度理论到测量信息论的数据处理方法过渡做了一些前期预研。

关 键 词 测量误差熵; 测量信息熵; 测量不确定度; Shannon信息熵
中图分类号 O1-647; N32 文献标识码 A

Information Entropy and Error Entropy of Measurement Data

TONG Ling，CHEN Guang-ju，LÜ Wen

(School of Automation Engineering, University Eletronic Science and Technology of China Chengdu 610054)

Abstract The paper introduces the model of the measured quantity, error and measuring data on the base of Shannon information theory after analysing the physical mean of the measurement and the relationship between the measured quantity and the measuring results. The measuring uncertainty, the measured quantity and the measuring results are presented by the model of the information entropy. The measuring information entropy and the error entropy are defined and their relationship is gived. And the mathematic relation of the information entropy and uncertainty of same special distributions is deduced.

Key words measuring error entropy; measuring information entropy; measurement uncertainty; Shannon entropy

“ 测量 ”的经典解释为用实验的方法获取被测量值的过程。其过程的物理含义为以基准、标准或工作用测量器具给出的量值为基本单位，去比对被测量值，所获得的单位的倍数为测量结果[1]。它是用实验的方法对被测量进行量化的过程，其中的关键为“ 实验 ”和“ 量化 ”。“ 实验 ”是一个物理过程，它保证测量不是通过计算、仿真或其他手段进行，强调其客观性，从而使得“ 测量 ”成为一切科学研究和工程技术的基础。而“ 量化 ”则是通过人类发明和创造的一些方法和规则对被测量值进行离散和编码，以便于进行识别、存储、显示等处理。大多数“ 量化 ”采用十进制技术，形成了一整套自然界量值数字化方法。“ 量化 ”过程既体现了人类的主观思想—— 用人类规定的一系列基本单位(如SI制中的基本单位系列)以及导出单位和单位进制去对被测量进行编码；同时也客观体现了科学技术发展水平—— 单位的精度、对被测量的分辨率以及量程等，在任何时候都体现了当时科学技术的最高水平[2]。因此从本质上看“ 测量 ”实际上是信息论中对被观测信息(被测量值)的编码和处理过程，与其他信息编码不同的是它的编码结果是数量，多数为十进制。不同单位的编码结果不同，但它并不影响被观测信息(被测量值)的客观性(量值大小)，这是单位换算的基础。因此可以认为信息论是处理测量及其结果的一套重要理论，测量信息论是信息论总体的一个研究分支，是信息论的一个重要应用研究领域[3-5]。

1 被测量信息熵

在传统的测量理论中，被测量被视为“ 一个 ”客观存在的、不变(测量过程中)的量值，此量值具有数字上的连续性，不能为基本单位的整数分割。

而实际被测量不是一个不变的、单一量值。在

收稿日期：2007 – 03 – 19
作者简介：童 玲(1963 –)，女，在职博士生，教授，主要从事测试理论与技术方面的研究.

各种因素影响下(如环境、人为、设备等)，除掉单向漂移变化，任何一个被测量都是一个随时间变化的随机参量，即被测量本质上为一个随机过程。被测量X的实际数学模型可以表示为一个可能性分布函数——pdf随被测量量值和时间变化的连续信源集合。测量过程是追求在$p(x,t)$分布下的集合的数学期望值$\overline{X}(t)$及不确定性[6]——信息熵$H_X(t)$的操作。

根据概率论和Shannon的信息论理论，被测量的数学期望和信息熵分别为[7-8]：

$$\overline{X}(t)=\int_{-\infty}^{+\infty}xp[x,t]\mathrm{d}x \tag{1}$$

$$H(X,t)=-\int_{-\infty}^{+\infty}p(x,t)\log_2\,p(x,t)\mathrm{d}x \tag{2}$$

被测量数学期望反映了被测量的随机平均特性；而被测量信息熵则是被测量信源集合不确定性或离散性的体现。

自然界存在的所有客观量值的pdf都随时间变化，其数学期望和信息熵也随时间变化。相对于测量操作而言，有的量变化慢，有的量变化快，严格按照此模型进行测量是不可行的。经简化后的、可操作的物理模型是：被测量是一个客观存在、单一的、固定不变的量值；被测量是一个拥有一定概率分布不随时间变化的连续集合；被测量概率分布满足时间遍历条件的连续集合。

2 测量结果信息熵

代表测量结果的数据构成测量的另一个信息集合：测量结果信息集合。它与被测量信源集合不同的是：(1) 它不是一个独立存在的信息集合，所含的信息内容(测量结果数据)和信息量(与不确定性相关的测量结果信息熵)由被测量信源和测量过程(测量原理、技术、设备以及人员等)决定。(2) 它是一个离散集合，离散的最小间距与测量分辨率有关，同时也正是由于其离散特性，使得被测量信源的部分信息丢失(作为连续集合的信源的处于阶梯之间的信息丢失)，表现在信息量减少，信息熵变化。由于测量结果信息集合由双重不确定因素决定，其各种特性远比被测量信源复杂，用信息论的方法能准确地获得测量结果信息集合的各种特性参数，并从中分析出有关被测量信源的相关特性。合理地评价测量质量和测量系统是测量信息论的主要内容之一。

测量结果信息集合是离散的具有一定概率分布的集合。此集合的不确定性可用信息熵来表示。测量结果信息集合不是一个独立存在的信息集合，它是以信源集合的内容及其分布为条件而存在的。二者之间的关系充分反应了测量过程的质量，因此在测量信息论中用表示二者之间关系的信息量参数来表示测量的质量。

设测量结果集合由数据 $\{y_0,y_1,\cdots,y_n\}$ 构成，其概率空间为：

$$Y:\begin{bmatrix}y_0 & y_1 & \cdots y_i & \cdots y_{n-1}\\ p_0 & p_1 & \cdots p_i & \cdots p_{n-1}\end{bmatrix}\quad\sum_{i=0}^{n-1}p_i=1 \tag{3}$$

则测量结果信息熵为：

$$H(Y)=-\sum_{j=0}^{m-1}p(y_j)\log_2\,p(y_j) \tag{4}$$

3 被测量集合与测量结果集合的关系及其测量误差熵

如图1所示，信源集合为连续集合，其元素的概率分布为$p(x)$，信源集合中的任何一个值都会对结果集合中的所有元素产生影响，是结果集合的条件。针对信源集合中的某个值x，结果集合的条件熵为：

$$H(Y/x)=-\sum_{j=0}^{m-1}p(y_j/x)\log_2\,p(y_j/x) \tag{5}$$

以整个信源集合为条件的结果集合的条件熵为：

$$\begin{aligned}H(Y/X)&=-\int_s p(x)H(Y/x)\mathrm{d}x=\\&-\int_s p(x)\sum_{j=0}^{m-1}p(y_j/x)\log_2\,p(y_j/x)\,\mathrm{d}x=\\&-\int_s\sum_{j=0}^{m-1}p(x,y_j)\log_2\,p(y_j/x)\,\mathrm{d}x\end{aligned} \tag{6}$$

式中 $p(x)\mathrm{d}x$为信源集合取x的概率；$p(x,y_j)=p(x)p(y_j/x)$。

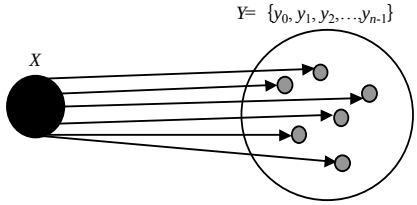

图1 连续分布信源集合与结果集合的关系

若测量结果信息集合的信息熵为$H(Y)$，则被测量集合和测量结果集合交互熵为：

$$H(X\cdot Y)=H(Y)-H(Y/X) \tag{7}$$

式中 $H(X\cdot Y)$代表通过测量从被测量信源集合传递到测量结果信息集合的“真信息”的信息熵；$H(Y/X)$则代表由于测量引入的误差和干扰信息的信息熵，称为误差熵。如果整个测量没有任何误差和干扰，则测量结果信息集合中的所有信息都来自于信源集

合，全部为“真实”信息，则有:

$$H(Y/X)=0，H(X\cdot Y)=H(Y) \tag{8}$$

但必须指出的是，由于测量过程中的各种因素(如量值离散化)，并不是所有被测量信源集合的信息全部到达测量结果集合中，有一部分信息被“丢失”，此时测量结果信息集合的信息熵等于交互熵，但并不等于被测量信源集合的信息熵。因此当被测量信源的信息熵已知时，测量结果的信息熵与标准源的信息熵之差即为被测系统误差熵，其中并不包含由于测量导致的信息损失所引起的信息熵变化。这部分信息熵为：

$$H(X/Y)=H(X)–H(X\cdot Y) \tag{9}$$

在没有信息“丢失”时(测量系统分辨率很高时，可作此近似)，有$H(X/Y)=0$，此时$H(X)=H(X\cdot Y)$，则误差熵为：

$$H(Y/X)=H(Y)–H(X) \tag{10}$$

即若已知被测量信源集合的信息熵$H(X)$，通过测量获得的结果信息集合的信息熵为$H(Y)$，二者之差即为测量导致的误差熵，即测量的不确定性。通常在检定测量仪器或系统时，用已知标准源测量被检定仪器或系统的误差熵。这其中包含人为和环境因素，但一般误差熵是在标准人员(检定员)和标准环境(计量室)前提下给出。在用标准测量仪器或系统测量被测量值时，由于测量系统的误差熵已知，测量结果信息集合的信息熵减去误差熵即可得到被测源的信息熵(前提是“丢失”的信息熵可忽略)。

如在经典测量模型中，被测量被视为单一、不变量值，被测量信息熵为0，则测量误差熵即为测量结果信息熵。它代表了测量结果的分布状况，表现为测量结果的不确定性[9-10]。在几种典型分布中，误差熵与标准不确定度有如下关系[11]：

$$H(Y/X)=H(Y)=\log_2(A\sigma) \tag{11}$$

式中　σ为标准不确定度；A为与分布函数有关的因子(正态分布为$\sqrt{2\pi e}$；均匀分布为$\sqrt{12}$；指数分布为$\sqrt{2e}$)。若将A视为扩展因子，则以bit为单位的扩展不确定度即为误差熵。

5　结 束 语

以Shannon信息论为理论基础的测量信息论是以信息熵为研究核心的一套现代测量数据和测量系统评价理论。它摒弃了传统的测量数学模型(如真值、误差等)，代之以集合、分布、信息熵、信息传递等现代信息论模型。在以模块化测量为发展趋势的测量仪器和系统的研究中，测量技术与信息技术和计算机技术的融合是现代测量技术研究的核心，而测量信息伦将为其提供强有力的理论支撑。尽管测量信息论的研究还处于起步阶段，但不可否认的是它必将成为信息论的一个重要研究分支。

参　考　文　献

[1] 林洪桦. 现代测量误差分析及数据处理[J]. 第6版. 计量技术, 1997, (6): 41-45.

[2] RICHTER D. Advanced mathematical tools in metrology[C]//V. Series on Advances in Mathematics for Applied Sciences. Singapore: World Scientific Publishing Company, 2001, 57: 93-104.

[3] LÜ Wen, TONG Ling, CHEN Guang-ju. The Maximum entropy method (MEM) in the measuring data processing[J]. Journal of Electronic Science and Technology of China, 2004, 2: 22-24.

[4] 童　玲, 陈光礻禹. 被测量信息熵、测量误差熵及其关系[J]. 仪器仪表学报，2004, 25(增刊): 821-824.

[5] SRIVASTAVA Y N, VITIELLO G, WIDOM A. Quantum measurements, information and entropy production[J]. Int. J. Mod. Phys., 1999, B12: 3369-3382.

[6] WEST B J. Measurement[J]. Information and Uncertainty. Mathematics & Computers in Simulation, 1987, 29(3-4): 169-189.

[7] 金振玉. 信息论[M]. 北京: 北京理工大学出版社, 1991.

[8] 捷莫尼科夫. 信息工程基础[M]. 北京: 机械工业出版社, 1985.

[9] MASRY E. Probability density estimation from sampled data[J]. IEEE Trans. Inform. Theory, 1983, IT-29: 697-709.

[10] QING Ping, WANG Zhong-yu. On non-statistic uncertainty in dynamic measurement[C]//Proc. 1st International Symposium on Instrumentation Science and Technology. Luoyan, China: [s.n.], 1999: 228-232.

[11] PRONZATO L, THIERRY E. A minimum-entropy estimator for regression problems with unknown distribution of observation errors[C]//In Bayesian Inference and Maximum Entropy Methods in Science and Engineering MaxEnt 2000, A. Mohammad-Djafari, Ed.: [s.n.], 2001: 169-180.

编　辑　漆　蓉