

doi: 10.3969/j.issn.1006-1576.2010.06.020

一种多源先验分布权重确定的新方法

王小林, 郭驰名, 程志君, 郭波

(国防科学技术大学 信息系统与管理学院, 湖南 长沙 410073)

摘要: 可靠性评估中融合多源验前信息的重要一环就是确定不同验前分布的合理权重。针对目前计算先验分布权重方法存在的缺陷, 提出了基于试验数据置信分布的方法, 以简便合理地确定权重。该方法直接与要评定的指标相联系, 从源头上减小评定的误差。示例证明, 该方法简便易行, 在可靠性试验分析等工程研究领域具有应用价值。

关键词: 多源验前信息; 权重; 信息融合; 置信分布

中图分类号: N945.17 **文献标识码:** A

A New Method for Determining Weights of Multi-Sources of Prior Distribution

WANG Xiao-lin, GUO Chi-ming, CHENG ZHI-jun, GUO Bo

(College of Management & Information System, National University of Defense Technology, Changsha 410073, China)

Abstract: An important step of multi-sources of prior information fusion is to get a reasonable weight for each prior distribution in reliability assessment. There are some different shortcomings of existing methods which calculate the weights for prior distribution, in order to determine the weights easily and rationally, it presents a new method which based on test data's confidence distribution. The method is directly linked to the targets which we want to assess, so it can reduce the assessment error from the headstream. An example shows that the method is simple and has obvious application values in the reliability analysis and other project fields.

Keywords: Multi-sources of prior information; Weight; Information fusion; Confidence distribution

0 引言

可靠性信息是可靠性评估工作的基础, 在可靠性工程中, 得到可靠性信息是多源和不确定的。如何综合利用多源可靠性信息来进行产品可靠性评估是重要的研究课题。伴随着设备的研制、定型到批量生产, 加上试验设备和手段的多样化, 使试验信息具有多种信息源; 同时, 由于技术条件或资源条件的限制, 使得现场试验数量越来越少^[1-2], 合理地解决这类有多源信息的小样本已成为热点问题。Bayes 方法能利用多源验前信息解决这类问题, 但难以确定不同验前分布在综合验前分布中的权重。

信息融合技术以其广阔的时空信息覆盖范围以及强大的信息综合和提取能力, 能有效解决制约产品可靠性评估的一些瓶颈问题。将不同的验前分布合理地融合为一个综合验前分布将涉及到信息融合问题。目前, 先验信息的融合方法主要包括: 1) 由专家确定各先验分布的权重; 2) 由各先验信息的可信度确定权重; 3) 相关函数融合法^[3]。其中, 专家确定权重的方法难以保证精度, 主观性较大; 可信度方法要主观上给出犯错的概率, 有一定的主观性; 而相关函数融合法要求有大量相关或相似产品的可用数据。即目前计算先验分布权重的方法在不同的程度上均存在缺陷, 故提出一种新方法, 即基于试

验数据得到的置信分布来确定不同先验的权重, 以合理客观地进行系统可靠性评估。

1 多源先验分布权重的确定

如图 1, 现获取了设备不同来源的可靠性信息: 先验信息 1, 先验信息 2……先验信息 n, 先将不同来源的先验信息分别作处理, 后传递到一个统一的融合中心, 在融合中心采用适当的方法综合各种信息得到综合的先验信息。利用处理后得到的综合先验信息加上试验数据通过 bayes 方法能得到相关的验后分布, 最后, 基于验后分布来进行最终的评估。



图 1 bayes 评估流程图

在融合中心对传递的先验信息进行融合, 目前多数方法是先确定先验分布的权重, 然后基于权重把多个先验信息融合成一个综合先验。遵循这个思想, 提出一种新的权重确定方法来融合先验分布。

设现要评估某设备的可靠性, 通过收集得到一些先验信息。如: 相似设备的信息、专家信息、设备试验信息等。把它们标记为: 先验信息 1, 先验信息 2……先验信息 n, 通过预处理得到各自的先验分布: $\pi_1(\lambda), \pi_2(\lambda) \dots \pi_n(\lambda)$, 其中, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ 。

收稿日期: 2010-01-23; 修回日期: 2010-03-10

基金项目: 国家青年自然科学基金 (60904002)

作者简介: 王小林 (1985-), 男, 云南人, 从事装备综合保障、可靠性分析与建模、维修决策研究。

试验数据为: X , 对不同类型的设备 X 的表现形式不一样, 如 $X = \{x_1, x_2 \dots x_n\}$ 或者 (n, s) 或 (τ, n') , 其中, x_i 表示故障时间, s 表示成功次数, τ 表示总试验时间, n' 表示失效次数。

试验数据信息和先验信息分别独自地提供关于可靠性参数的信息, 虽然它们反映的信息量大小有异, 但地位应是对等的^[4]。基于这种思想, 将试验数据视为“先验信息”构造置信分布, 并以它作为标准来确定不同先验的权重。

假设设备可靠度含有表达式 $g(\lambda)$, 该方法的基本步骤如下:

1) 收集先验信息, 并将先验信息按照适当的方法转化成各自的先验分布 $\pi_1(\lambda), \pi_2(\lambda) \dots \pi_n(\lambda)$;

2) 基于试验数据诱导出可靠性参数 λ 置信分布 $\pi_b(\lambda)$, 并将此置信分布作为真实的“先验分布”;

3) 依据上述先验信息获得先验分布 $\pi_1(\lambda), \pi_2(\lambda) \dots \pi_n(\lambda)$ 和试验数据获得的置信分布 $\pi_b(\lambda)$ 分别求 $g(\lambda)$ 的期望 $E[g(\lambda)]$, 以 $E_{\pi_b}[g(\lambda)]$ 标准求“距离” e_i : $e_i = |E_{\pi_b}[g(\lambda)] - E_{\pi_i}[g(\lambda)]|$

最后, 通过 e_i 确定不同先验的权重。一般情况下 $E[g(\lambda)]$ 是容易计算的, 如果分布复杂或者 $g(\lambda)$ 复杂, $E[g(\lambda)]$ 可以通过数值积分得到。

4) 根据权重融合多源信息得到融合先验分布, 根据 Bayes 公式得到后验分布, 基于后验分布做相应的统计推断。

$$e_i = |E_{\pi_b}[g(\lambda)] - E_{\pi_i}[g(\lambda)]| \quad (1)$$

e_i 越小, 说明先验分布 $\pi_i(\lambda)$ 和置信分布 $\pi_b(\lambda)$ 相差越小, 即先验信息和试验信息一致性较好, 那么 $\pi_i(\lambda)$ 权重应该越大。即 e_i 越小, 表示先验信息 i 提供的先验信息更能够反映设备可靠性的真实情况, 那么 $\pi_i(\lambda)$ 权重应该较大。

令 $w_i = \frac{1}{e_i}$, $w = [w_1, w_2 \dots w_n]$, 对 w 进行归一化处理可以得到不同先验的权重。即:

$$w'_i = \frac{w_i}{\sum_{j=1}^n w_j} \quad (2)$$

于是得 $w' = [w'_1, w'_2 \dots w'_n]$, 也就是得到了不同先验分布的权重, 能够对多源信息进行信息加权融合。这里令 $w_i = \frac{1}{e_i}$ 是依据 e_i 的意义: e_i 越小, $\pi_i(\lambda)$ 权重应该越大。得到权重向量 w' 后, 对先验分布函

数 $\pi_1(\lambda), \pi_2(\lambda) \dots \pi_n(\lambda)$ 进行加权融合, 得到一个综合的先验分布^[1]:

$$\pi(\lambda) = \sum_{i=1}^n w_i \pi_i(\lambda) \quad (3)$$

从权重向量的计算过程来看, 没有加入主观信息, 说明该方法是客观的, 能直观反应出不同先验分布的重要性, 计算也较为方便。

2 $g(\lambda)$ 的确定

设备的可靠度可以表示为 $R(t|\lambda)$, 设可靠度表达式中含有表达式 $g(\lambda)$, 其中 $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]$ 。 $g(\lambda)$ 为 λ 的函数表达式, 与 t 无关, 它对可靠性指标的值有关键的决定作用。因此, 只要能够对 $g(\lambda)$ 进行精确评估, 就能够对可靠性相关的指标进行精确评估, 从而提高最终评定结果的可信性。基于这个思想, 通过不同先验分布对 $g(\lambda)$ 的估计来确定不同先验的权重, 即通过“距离”来确定权重。

指数分布的可靠度表达式为 $R(t) = e^{-\lambda t}$, λ 的取值对靠性指标的值有关键的决定作用, 因此取 $g(\lambda) = \lambda$ 。正太分布的可靠度表达式为 $R(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda_1)^2}{2\lambda_2^2}} dx$, 在均值为 λ_1 的条件下, 方差 λ_2 的取值对靠性指标的值有关键的决定作用, 因此取 $g(\lambda) = \lambda_2$; 在均值为 λ_1 和方差 λ_2 未知的条件下, $\lambda_1^2 / \lambda_2^2$ 的取值对靠性指标的值有关键的决定作用, 因此取 $g(\lambda) = \lambda_1^2 / \lambda_2^2$ 。Weibull 分布可靠度表达式为 $R(t) = e^{-\left(\frac{t}{\lambda_1}\right)^{\lambda_2}}$, $\left(\frac{1}{\lambda_1}\right)^{\lambda_2}$ 的取值对靠性指标的值有关键决定作用, 因此取 $g(\lambda) = \left(\frac{1}{\lambda_1}\right)^{\lambda_2}$, 其中 λ_1 为尺度参数或特征寿命, λ_2 为形状参数。Pareto 分布可靠度表达式为 $R(t) = \left(\frac{\lambda_1}{t}\right)^{\lambda_2}$, $\lambda_1^{\lambda_2}$ 的取值对靠性指标的值有关键的决定作用, 因此取 $g(\lambda) = \lambda_1^{\lambda_2}$, 其中 λ_1 为尺度参数, λ_2 为形状参数。同理可以得到其它分布的 $g(\lambda)$ 。

为了方便计算, 表 1 中列出常用分布的 $g(\lambda)$:

表 1 常用分布的 $g(\lambda)$

常用分布	$g(\lambda)$
指数分布	λ
正太分布 (方差或者均值已知)	λ
正太分布 (方差和均值未知)	$\lambda_1^2 / \lambda_2^2$
Weibull 分布	$\left(\frac{1}{\lambda_1}\right)^{\lambda_2}$
Pareto 分布	$\lambda_1^{\lambda_2}$

从 $g(\lambda)$ 的选择来看, $g(\lambda)$ 直接与要评定的指标相联系, 能从源头上减小评定的误差; 同时, 从 $g(\lambda)$ 的利用过程来看, 把先验信息直接和评定指标联系起来, 能充分挖掘出有用的信息, 进一步充分地利用先验信息。

3 置信分布的确定

将实际获得的试验数据根据置信限理论诱导出可靠性参数 λ 的一个分布 (称为置信分布)。记由数据 X 用经典统计的置信区间理论获得可靠性参数 λ 的下限 (置信度为 γ) 为 $\lambda_L(x; \gamma)$, 满足^[4-5]:

$$P(\lambda \geq \lambda_L(x; \gamma)) = \gamma \quad (4)$$

一般, 当 X 给定时, $\lambda_L(x; \gamma)$ 随 γ 的变动可生成 λ 的概率分布。记其密度为 $f(\lambda)$, 称为 λ 的置信分布。同样, 置信上限也可以诱导出 λ 的置信分布。

对于离散型变量 X , 基于上限和下限诱导出的置信分布一般稍有不同。例如, 对可靠度为 R 的成败型数据 (n, s) (n 表示试验次数, s 表示成功次数), 可得其置信分布为:

$$f(R) = \frac{dF(R)}{dR} = B_e(R; s, n-s+1) \quad (5)$$

同理, R 的上限置信密度 $B_e(R; s+1, n-s)$, 折中考虑, 在成败型试验数据场合可取产品可靠度 R 的置信分布为^[4]:

$$f(R) = B_e(R; s+0.5, n-s+0.5) \quad (6)$$

对一般连续型变量 X , 由置信下限和置信上限诱导出的置信分布是相同的。例如: 某指数产品预定失效次数为 z , 总试验时间为 τ , 指数分布产品的可靠性参数 (失效率 λ) 的置信分布如下^[5]:

$$f(\lambda) = \Gamma(\lambda | z, \tau) \quad (7)$$

4 示例

有一类设备的寿命是服从指数分布的, 现得到设备的一组完全试验数据为 (注: 失效率为 0.006 指数分布产生的随机变量): {128.813 3, 193.790 1, 76.757 2, 237.739 8, 140.013 2, 214.606 1}。同时, 得到设备的一些仿真信息, 类似设备的信息和专家信息等先验信息。这里采用了置信分布, 把先验信息转化为指数分布的共轭先验分布即伽马分布, 相应分布的参数如表 2。

由式 (7) 可得试验数据的置信分布为:

$$f(\lambda) = \frac{991.719 7^6}{\Gamma(6)} \lambda^{6-1} e^{-991.719 7\lambda} \quad (8)$$

指数分布的可靠度表达式为: $R(t) = e^{-\lambda t}$, 故令

$g(\lambda) = \lambda$, 则 $E(g(\lambda)) = 0.006 05$ 。

同理, 可以得到 $E_{\pi_i}(g(\lambda))$, 如表 2 第 2 列。由式 (1) 可得表 2 第 4 列, 由式 (2) 可得表 1 第 5 列。此时, 先验分布的权重就确定了: $w' = [0.129 732, 0.759 607, 0.110 661]$ 。从表 2 中可以看出: 先验信息 2 的权重最大, 即 $E_{\pi_i}(g(\lambda))$ 的估计值与 $E(g(\lambda))$ 相差越小, 则这个先验的分布权重就越大。

表 2 先验分布参数与权重确定表

分布参数 (α, β)	$E(g(\lambda))$	e_i	w'_i	
置信分布	(6, 991.719)	0.006 05	0	—
先验分布 1	(12, 1 633.908)	0.007 344	0.001 294	0.129 732
先验分布 2	(17, 2 916.448)	0.005 829	0.000 221	0.759 607
先验分布 3	(19, 4 190.811)	0.004 533	0.001 517	0.110 661

$$\text{由式 (3) 可知, } \pi(\lambda) = 0.129 732\pi_1(\lambda) + 0.759 607\pi_2(\lambda) + 0.110 661\pi_3(\lambda) \quad (9)$$

如图 2 所示, 实线表示 $\pi(\lambda)$, 即融合后的综合先验; 点线 (先验密度 2) 表示 $\pi_2(\lambda)$, 它们的图形重叠区域较大, 直观上看, 先验信息 2 的权重应该较大, 与表 2 的数据相吻合, 反应了中确定权重的方法是有效的。

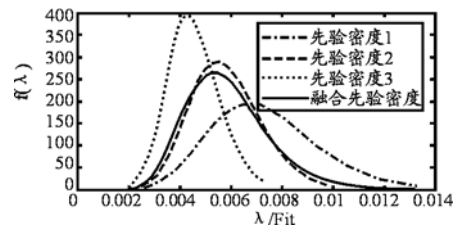


图 2 先验密度曲线

$$\text{记: } m_i(X) = \int_A \pi_i(\lambda) * L(X | \lambda) d\lambda, \quad m(X) = \int_A \pi(\lambda) * L(X | \lambda) d\lambda,$$

$$\pi_i(\lambda | X) = \frac{\pi_i(\lambda) * L(X | \lambda)}{\int_A \pi_i(\lambda) * L(X | \lambda) d\lambda}.$$

$$\text{则: } \pi(\lambda | X) = \frac{\sum_{i=1}^n w_i * m_i(X) * \pi_i(\lambda | X)}{m(X)} \quad (10)$$

其中 $n = 3$, $\pi_1(\lambda | X) = \Gamma(\lambda; 18, 2 625.627 7)$, $\pi_2(\lambda | X) = \Gamma(\lambda; 23, 3 908.167 7)$, $\pi_3(\lambda | X) = \Gamma(\lambda; 25, 5 182.530 7)$ 。

因此, 综合验后分布为:

$$\pi(\lambda | X) = 0.0921\Gamma(\lambda; 18, 2 625.627 7) + 0.822\Gamma(\lambda; 23, 3 908.167 7) + 0.0859\Gamma(\lambda; 25, 5 182.530 7) \quad (11)$$

由式 (11) 可以看出, 综合验后分布是有各验后分布加权得到的, 只是它们的权重是在先验权重的基础上进行了重新分配。它们的趋势是一样的: 先验权重, 后验权重就大。

从图 3 中可知, 通过试验数据的修正, $\pi_2(\lambda | X)$ 与综合验后分布 $\pi(\lambda | X)$ 重合区域增加了, 它的权重

较 $\pi_2(\lambda)$ 而言, 由 0.759 607 增加到 0.882。

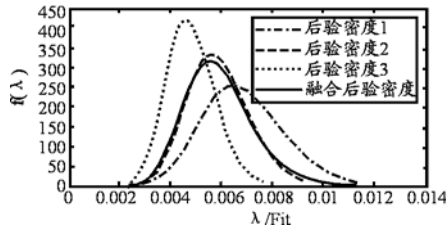


图 3 验后分布曲线

得到综合验后分布后, 在平方损失的条件下, 得失效率 λ 的点估计为 $\hat{\lambda} = E_{\pi(\lambda|X)}(\lambda)$, 即 $\hat{\lambda} = 0.005\ 883$ 。

$\hat{\lambda}$ 与 0.006 相比相差 0.000 117, 说明了该方法的精确性。此时, 除了得到失效率的点估计, 还可得到它的区间估计, 进一步得到其它可靠性指标的估计。

5 结论

(上接第 59 页)

- B 为螺旋线的螺旋角度 (rad);
- X 为水平方向的运动距离 (mm);
- R 为圆柱半径 (mm)。

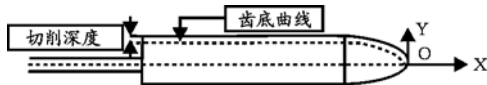


图 2 圆柱球头型旋转锉

通过以上分析可知, 在 B 轴摆好角度的情况下, 只需要 X 轴和 A 轴联动便可形成圆柱部分的刀刃曲线, 且这 2 个轴的运动是线性关系。

对于旋转锉的球头部分, 其刀刃曲线是球面螺旋线, 通过分析其齿底曲线 (图 2 中虚线) 应为椭圆曲线。采用拟和的办法, 把齿底曲线近似分解为大圆和小圆两部分。故球面螺旋线的形成应该是在 XOY 平面内, 工件是绕圆弧的圆心转动和绕自身轴线转动的合成运动。如图 2, 以旋转锉的头部为圆点建立直角坐标系, 给出圆弧的起点坐标、终点坐标和圆心, 便可计算出圆弧的插补数据。由于拟和圆弧的圆心不一定和 C 轴的旋转中心重合, 所以需要直线轴的平移来配合。其数学关系推导如下:

$$cs = \text{delta}C \tag{2}$$

$$xs = -(LenX * (1 - \cos(\text{delta}C)) - (r_b + LenY) * \sin(\text{delta}C)) \tag{3}$$

$$ys = -R - (LenY + LenX * \sin(\text{delta}C)) - LenX * \cos(\text{delta}C) \tag{4}$$

$$a = 180 * \text{tg} * \text{delta}X / \text{PI} / R0 \tag{5}$$

$$\text{tg} = \tan(B) \tag{6}$$

$$\text{delta}X = R * \cos(C + \text{delta}C) \tag{7}$$

$$R0 = R * \sin(C + \text{delta}C) \tag{8}$$

其中: PI 为圆周率; xs、ys、cs、a 为各轴的运动

通过示例可以看出, 该方法能合理地融合不同的验前分布, 得到融合先验分布, 是合理的, 并且具有简便易行的特点。同时, 该方法直接与要评定的指标相联系, 从源头上减小了评定的误差。以指数分布为例, 该方法还可以推广到其它的分布, 在可靠性试验分析等工程研究领域具有良好的应用前景和推广价值。

参考文献:

- [1] 柴建, 师义民, 等. 多源验前信息下先验分布的融合方法[J]. 科学技术与工程, 2005, 5(20): 1479-1481.
- [2] Schabe H. Combining failure rate data from various sources[J]. Microelectron Reliability, 1996, 36(1): 47-54.
- [3] 柴建, 师义民, 等. 多源验前信息下的 Bayes 信息融合方法[J]. 数理统计与管理, 2007, 27(2): 308-311.
- [4] 方良海. 产品可靠性评估中的多源信息融合技术研究[D]. 合肥工业大学研究生院, 2006: 43-45.
- [5] 周源泉, 翁朝曦. 可靠性评定[M]. 北京: 科学出版社, 1990: 58-61.

动坐标值 (mm); LenX、LenY 为旋转中心在坐标系中的坐标 (mm); tg、R0、deltaX 为计算坐标数据 a 的中间变量值。

在实际编程计算中, 计算出各段圆弧的插补数据, 手动给出 C 轴的旋转坐标值, 便可根据上述公式计算出各轴的运动坐标, 结合式 (1), 即可形成球型旋转锉的球头运动模型。通过计算得出的数据会自动嵌入到系统内部, 经过数控系统的解析, 便可同步输出, 达到控制五坐标联动, 形成刀刃轨迹。

3 结束语

根据刀具的几何模型, 并结合机床的实际坐标结构, 采用运动分析的方法确定形成刀具形面所需的机床坐标运动及各坐标之间的运动关系。该方法已成功应用于旋转锉数控系统, 且和五轴工具磨床配套成功。利用该方法加工出来的产品质量高, 使用寿命长, 可以很好地解决球型旋转锉的加工问题。

参考文献:

- [1] 张世昌, 李国勤. 旋转锉四坐标数控刃磨技术的研究[J]. 工具技术, 2001, 35(3): 16-18.
- [2] 张得礼, 周来水. 旋转锉四轴联动磨削运动模型和成型理论研究[J]. 东南大学学报(自然科学版), 2003, 33(3): 324-327.
- [3] 刘杰华. 刀具精确设计理论与实践. 北京: 国防工业出版社, 2005.
- [4] 于峰, 夏广岚, 徐智. 球头旋转锉的几何建模和精确成型理论的研究[J]. 佳木斯大学学报(自然科学版), 1999, 17(2): 110-114.
- [5] 张民惠, 朱沛洪. 旋转锉螺旋刃加工时机床运动的分析[J]. 机械科学与技术, 1994, 5(1): 101-104.