



Session 1

Statistical Applications

INSPECTION DATA USE FOR INSPECTION PROGRAM DEVELOPMENT

Yuri M. Paramonov, Andrey S. Kuznetsov

*Riga Technical University,
Lomonosova 1, Riga LV-1019, Latvia, E-mail: andreyk@hotmail.ru*

The inspection program development on the base of preliminary full-scale fatigue test results and its correction by the use of inspection data in service are considered. Since the years 1953 & 1954, when there were three consequent fatal crashes of Comet's aircraft, then up to now the problem of metal fatigue of aircraft draws attention of the entire aviation world. The first attempt to solve this problem was the use of safe-life approach. In this case an aircraft must be discarded from service when number of flight hours exceeds the so called specified life, which is defined as mean fatigue life of full-scale fatigue test divided by reliability factor. The value of reliability factor in some cases can be even more than 10 and the cost of reliability at this approach is too high. Later a new so called fail-safe (or now it is called damage tolerance (some authors call it "inspection dependent")) approach was developed, in which the main emphasis is on the inspection program. The guidelines for developing structural inspection programs, that meet the new damage tolerance regulations, are given in the ATA document, MSG-3. A key element of this approach is a probability calculation of fatigue failure.

Calculation of this probability should be based on a fatigue crack propagation model (FCPM) and the corresponding fatigue test result processing for parameters of FCPM's estimation. By the use of corresponding estimations the inspection program should be developed. The problem is that the cost of such test is very high and usually we have only test of one (seldom two) aircraft. As a result we have the observation of only one fatigue crack in specified structural significant item (SSI). This means that to provide reliability we should increase the number of inspections and increase the cost of preventive maintenance as a whole. In order to avoid this loss we can use the data of inspections for more precise estimations of FCPM's parameters.

This paper is devoted to discussion of this problem. Exponential model of fatigue crack growth was used. Modeling of times, when fatigue crack become detectable and when it growth up to critical size, was made by Monte Carlo method. Maximum likelihood method was used for parameter of fatigue crack growth estimation. It is shown, that the significant decreasing of inspection number can be achieved even if for correction of the inspection program only the information about first crack discovered in service will be used.

МЕТОДЫ ОПРЕДЕЛЕНИЯ КОЛИЧЕСТВА КЛАСТЕРОВ ПРИ КЛАССИФИКАЦИИ БЕЗ ОБУЧЕНИЯ

Ирина Яцкив, Лада Гусарова

*Институт транспорта и связи,
ул. Ломоносова 1, LV-1019, Рига, Латвия
Тел.: (+371)-7100650. Факс: (+371)-7100660, E-mail: ivl@tsi.lv*

Ключевые слова: кластерный анализ, число кластеров, правила останковки

Кластерный анализ – технология группирования объектов в неизвестные классы. Он отличается от дискриминантного тем, что неизвестны ни число кластеров, ни их характеристики. Проблема определения числа кластеров является одной из основных нерешенных до настоящего времени задач кластерного анализа (непараметрического случая классификации). Обратимся к двум наиболее используемым типам процедур кластерного анализа: иерархическим и итеративным. Для итеративных алгоритмов число кластеров является одним из входных параметров алгоритма. Для иерархических процедур характерен визуальный анализ дендрограммы и определение по ней наиболее предпочтительного числа кластеров. Такой визуальный анализ чрезвычайно затруднен при:

- 1) большом количестве рассматриваемых объектов;
- 2) неявной выраженности структуры данных. (Кластерный анализ прежде всего решает задачу привнесения структуры в данные, а если данные по сути своей однородны, то не проанализировав корректность определенного числа кластеров, можно допустить грубую ошибку в дальнейших статистических исследованиях).

Существуют различные формальные подходы, облегчающие процедуры определения предпочтительного числа кластеров [1, 2]. Эти подходы называются правилами останковки. Миллиган и Купер [3] исследовали более тридцати из них. В статье рассматриваются два локальных и глобальное правила останковки, преимущества и недостатки применения этих правил рассматриваются на примере классификации европейских стран по уровню развития транспорта.

Литература

1. Everitt B. Cluster Analysis, 2nd edn. Wiley, NewYork, 1993, 283 p.
2. Gordon A.D. Classification, 2nd edn. Chapman & Hall, NewYork, 1999, 256 p.
3. Milligan G.W. and Cooper M.C. An examination of procedures for determining the number of clusters in a data set, Psychometrika, 1985, 50, 159-179p.

RESAMPLING APPROACH FOR COMPLEX SYSTEMS RELIABILITY ESTIMATION

Maxim Fioshin

*Transport and Telecommunication Institute, Riga, Latvia,
Lomonosova 1, Riga, LV1019, Latvia.
Ph: (+371) 7100650. Fax (+371) 7100660. E-mail: mf@tsi.lv*

Key words: Resampling, Simulation, Estimation, Reliability

Resampling approach is alternative approach for complex systems simulation and analysis. It can be used efficiently in the case of insufficiency of primary data (Andronov and Merkurjev, 2000).

In the present work Resampling approach is applied for estimation of the complex systems reliability. The complex reliability system can be described as following. Let x_1, x_2, \dots, x_n are real variables. They can get values from the set M . Each of these variables corresponds to one element of the system.

Let operation $f_m(y_1, y_2, \dots, y_k)$, $m \leq k$, defines m -th order statistics of values y_1, y_2, \dots, y_k . For example, $f_1(y_1, y_2, \dots, y_k)$ gives minimal value from y_1, y_2, \dots, y_k , but $f_k(y_1, y_2, \dots, y_k)$ gives maximal value.

Also, relations $<$, $>$ and $=$ are defined on M . The result of two values y_i and y_j comparison is logical value, 0 or 1.

For logical values operation $P_m(z_1, z_2, \dots, z_k)$ is defined as following:

$$P_m(z_1, z_2, \dots, z_k) = \begin{cases} 1, & \text{if at least } m \text{ values from } k \text{ are equal to } 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Operation $P_1(z_1, z_2, \dots, z_k)$ defines logical "or", $P_k(z_1, z_2, \dots, z_k)$ defines logical "and". Also, "not" operation can be performed on output of this operation.

Let us construct a formula $Q(x_1, x_2, \dots, x_n)$ from given elements. Its arguments correspond to input variables x_1, x_2, \dots, x_n . Then, order statistics operations are applied to these arguments, and, maybe, some other constants from the set M . Then results are compared with each other and, maybe, with some other constants from the set M . Then operations P are made for obtained logical values. The result of a formula is logical value.

Let X_1, X_2, \dots, X_n be independent random variables with distribution functions F_1, F_2, \dots, F_n . Functions F_1, F_2, \dots, F_n are unknown, but sample populations H_1, H_2, \dots, H_n are available for each X_i .

The task is to estimate mathematical expectation θ of formula Q :

$$\theta = \int \dots \int Q(x_1, x_2, \dots, x_n) dF_1(x_1) dF_2(x_2) \dots dF_n(x_n).$$

References

1. Andronov A., Merkurjev Yu. (2000). Optimization of Statistical Sizes in Simulation. *Journal of Statistical Planning and Inference*, 85, 93-102.
2. M. Fioshin. (2001). Resampling Estimators for Hierarchical Sequential-Parallel Systems Reliability. *Journal "Transport and Telecommunication" 2002, Vol.3.* 123-126.

TRANSPORT AIDS STATE CONTROL PRINCIPLES IN CRITICAL CONDITIONS

Eugene Kopytov, Vladimir Labendik

*Transport and Telecommunication Institute, Riga, LV-1019, Latvia
Lomonosova iela 1, Riga, LV-1019, Latvia
Phone: +(371) 7100590, fax: +(371)7100660, E-mail: kopitov@tsi.lv,*

Nikolay Kuznetsov

*Scientific and Technical Centre "Techcon", St. Petersburg, Russia
vitehkon@spb.cityline.ru*

Key words: aircraft diagnostics, computer diagnostic systems, neural networks

Complication of the modern technique and the growth of the sizes of transport aids leads to more difficult sequences of their accidents and catastrophes. At the same time, when the decision should be taken during grains of second the so-called human factor starts acting – the crew is not able to analyze all incoming information about the operation of technique and to take a proper decision.

But, till the present time on the transport aids the airborne control and diagnostic systems are working in similar modes in normal and critical conditions, and in the aircraft a half of information incoming during the flight is written and stored in the so-called 'black boxes'. The examination of the flight accidents starts with the search of these 'black boxes' and the further analysis of the emergency operation of technique after the accident.

For the increase of the reliability of the aircraft operation it is offered to improve the principle of the processing of the flying information received in the computer diagnostic system with the help of the highlighting of the 'critical' modes of the information processing in their work.

The precondition for the performance of these offers is a global development of the satellite communication and network of the worldwide information system Internet, as well as the application of more miniaturization of the airborne computer devices and the application of more developed software supply including the artificial intelligent systems with the use of the so-called neural networks [1, 2].

When the extreme deviations are reached, the engine must be switched off, as it is known. It is offered to implement a double deviation control on the board – both extreme and dangerous (a little bit lower than extreme) deviations. In case of the achievement of the dangerous deviations the integrated on-ground and airborne decision-taking system about the engine state during the flight in the critical situations must switch on.

This system should have 3 levels:

airborne system of engine state control according to the extreme readings of the measured parameters;

on-ground part of computer diagnostic system processing the given flight information in the real time mode and comparing these readings with the set ones of the previous flights (evaluation of trend parameters);

database of system for the comparison of the statistical information of the given engine and the fleet of the one-type aviation engines for the evaluation of the connection of the critical deviations of parameters with the dangerous malfunctions. The example of the block

scheme of the structural interaction of the integrated on-ground and airborne decision-taking system levels is presented.

References

1. Kopytov E., Kabelev N. Cluster Analysis and Artificial Neural Network Applications in Time-Critical Aviation Tasks. *Proceedings of the Nordic-Baltic Transport Research Conference*, vol. 2 (5), April 13-14, 2000, Riga, pp. 53-57.
2. Kopytov E., Labendik V., Kabelev N. Neural Network Applications in Aircraft Engine Diagnostics. In: *Scientific Proceedings of the Scientific-technical Union of Mechanical Engineering*, vol. 2 (57). Third International Congress "Mechanical Engineering Technologies'01", June 24-26, 2001, Sofia, Bulgaria, pp. 558 – 560.

RESAMPLING VERSUS A DEPENDENCE IN INPUT STATISTICAL DATA

A.M.Andronov

*Transport and Telecommunication Institute
Lomonosov Str. 1, Riga, LV-1019, Latvia
e-mail: lora@mailbox.riga.lv*

Methods of statistical data analysis supposes usually that given data are realizations of mutual independent random variables. In fact these data can be dependent variables. It complicates the statistical analysis. We show, how resampling approach works in this case.

Let us consider a population of n objects. Each object is described by m parameters $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_m$, that are mutual independent random variables. We suppose that for various objects vectors $\tilde{\mathbf{X}}^{(1)} = (\tilde{X}_1^{(1)}, \tilde{X}_2^{(1)}, \dots, \tilde{X}_m^{(1)})$, $\tilde{\mathbf{X}}^{(2)} = (\tilde{X}_1^{(2)}, \tilde{X}_2^{(2)}, \dots, \tilde{X}_m^{(2)})$, ... are independent and identically distributed ones.

Now we measure the parameters of the j -th object and get vector $\mathbf{X}^{(j)} = (X_1^{(j)}, X_2^{(j)}, \dots, X_m^{(j)})$ instead of $\tilde{\mathbf{X}}^{(j)} = (\tilde{X}_1^{(j)}, \tilde{X}_2^{(j)}, \dots, \tilde{X}_m^{(j)})$, $j = 1, 2, \dots, n$. This difference can be explained for example by measurement mistakes. We suppose that: 1) $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots$ are mutual independent random vectors; 2) $X_1^{(j)}, X_2^{(j)}, \dots, X_m^{(j)}$, $j = 1, 2, \dots, n$ are dependent random variables; 3) marginal distribution of $X_i^{(j)}$ coincide with the distribution of $\tilde{X}_i^{(j)}$ for each $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$.

Our aim is an evaluation of mathematical expectation

$$\theta = E\psi(\tilde{\mathbf{X}}) = E\psi(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_m)$$

where ψ is known function of m arguments.

For this we dispose of sample population $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}$. We consider various estimators of θ and investigate these efficiency.

APPROXIMATION OF FATIGUE CURVE OF COMPOSITE MATERIAL

A.Yu. Paramonova, M.A. Kleinhof, Yu.M. Paramonov

*Transport and Telecommunication Institute
Lomonosov Str. 1, Riga, LV-1019, Latvia
E-mail: rauprm@junik.lv*

Key words: strength, fatigue life, composite

The composite material is now in wide use in different types of structure, for example, in airframe and jet engines. This work is motivated by the need to develop and present quantitative fatigue – life information used in design of such structures. Fatigue data are usually presented in the form of a S-N curve, a log-log plot of cyclic stress or strain S versus the mean or median fatigue life N , which is expressed in cycles to failure. A great number of articles is devoted to this problem. Wide discussion on this matter took place on F.G. Pascual's and W.Q. Meeker's paper in *Technometrics*, November 1999 [1]. Great number of references are given there in References and Additional References. So although for two last decades the main attention of fatigue study is paid to the use of fracture mechanics, nevertheless the study of S-N curve is very important for brittle materials and especially for composite materials.

Some simple probabilistic model for fatigue life of composite material consisting mainly of parallel rigid components (strands) is offered. This model, which is founded on the use of Markov chain theory, allows to get S-type curve of internal stress growth and, finally, to see the connection between static strength distribution parameters and S-N fatigue curve (Wholer curve). The expectation value and variance of fatigue life can be calculated. The model is too simple and does not provide numerical coincidence with experimental fatigue test data if we use static strength distribution parameters without correction. But we can fit this model to actual fatigue datasets by the use of it as nonlinear regression model of S-N fatigue curve. Finally we can predict consequence of static strength parameter changes by the use of considered model.

Several numerical examples were considered. It is shown, that the offered model gives satisfactory result for fatigue curve approximation.

References

1. Pascual F.G. and Meeker W.Q., "Estimating Fatigue Curves With the Random Fatigue-Limit Model," *Technometrics*, v.41, 277-302 (1999).
2. Paramonov Yu.M. and Kleinhof M.A., "Simple statistical model of composite material strength". *Aviation*, #5. Vilnius: Technika, 2000, pp.128-132.

ОЦЕНКА И ПРЕДСКАЗАНИЕ КАЧЕСТВА КЛАССИФИКАЦИИ ДЛЯ АЛГОРИТМОВ КЛАССИФИКАЦИИ С ОБУЧЕНИЕМ

Анатолий Рессин

*Институт транспорта и связи,
ул. Ломоносова 1, LV-1019, Рига, Латвия
Тел.: (+371)-7100650. Факс: (+371)-7100660, E-mail: aar@tsi.lv*

Ключевые слова: Классификация с обучением, матрица классификации, перекодировка классов.

При использовании алгоритмов классификации с обучением ([1]) возникает вопрос о качестве материала, на котором происходит обучение. Как известно (см., например, [2] с дальнейшей библиографией), существует множество различных функционалов для оценки качества классификации. Как правило, эти методы ограничиваются лишь оценкой качества полученного разбиения объектов на классы. В данной работе предложен метод, который позволяет предсказать эффективность применения того или иного алгоритма классификации к исследуемым объектам по результатам, полученным при обучении. Для этого, сначала вводится понятие групповой неразличимости классов по отношению к классифицирующему алгоритму. Используя это понятие, показывается, что предложенный метод инвариантен по отношению к перекодировке классов, а также представлена геометрическая интерпретация предлагаемого метода.

Литература

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерностей: Справочное издание / Под ред. С.А. Айвазяна. – Москва: Финансы и статистика, 1989. – 607 с.
2. Мандель И.Д. Кластерный анализ. – Москва: Финансы и статистика, 1988. – 176 с.

НЕКОТОРЫЕ ЗАДАЧИ ТОЧЕЧНОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Виктор Люмкис

*Институт транспорта и связи,
Ломоносова 1, Рига, LV-1019, Латвия
E-mail: vlyumkis@yahoo.com*

Ключевые слова: оценивание, методы моментов и максимального правдоподобия

Классическими методами точечного оценивания, как известно, являются метод моментов и метод максимального правдоподобия [1]. Однако их применение нередко приводит к значительным вычислительным трудностям, что заставляет рассматривать проблему как отдельную самостоятельную задачу. Ряд конкретных практических задач такого рода рассматривается в настоящей работе. В одной из них находится совместная плотность двух случайных величин X и Y , распределенных соответственно по нормальному и равномерному распределениям. Демонстрируется возможность решения системы уравнений метода моментов в пакете Matchcad 7.0 [4].

Другой задачей, требующей достаточно сложной вычислительной работы, является нахождение по методу максимального правдоподобия оценок параметров α, β логистического распределения с функцией распределения вида

$$F(x) = \frac{1}{1 + e^{-(\alpha + \beta x)}}, \quad -\infty < x < +\infty.$$

Решение системы уравнений метода максимального правдоподобия для нахождения оценок параметров $\hat{\alpha}, \hat{\beta}$ было вновь проведено в пакете Matchcad 7.0 с помощью блока Given, Find.

Отдельного рассмотрения требует и задача оценки параметров по данным двух выборок наблюдений, полученных, например, по результатам n наблюдений до и m наблюдений после проведения некоторого контролируемого мероприятия (КМ) [2]. Более точно задача формулируется следующим образом (упоминание о подобной задаче имеется в [1]): пусть имеются две независимых выборки объемов n и m из двух нормальных распределений с одинаковыми математическими ожиданиями μ и дисперсиями соответственно $\lambda\sigma^2$ и σ^2 . Найти оценки метода максимального правдоподобия для указанных параметров.

В настоящей работе приводятся как аналитические выкладки для решения этой задачи, так и ее численная реализация по конкретным выборочным данным с использованием пакета Matchcad 7.0. Представлены также некоторые дополнительные свойства выведенных оценок.

Литература

1. Кендалл М., Стьюарт А. Статистические выводы и связи, Москва, Наука, 1973.
2. Парамонов Ю. М. Методы математической статистики в задачах, связанных с оценкой и обеспечением надежности усталостной долговечности авиационных конструкций, Рига, Ркии га, 1992.
3. Свешников А.А. Сборник задач по теории вероятностей, математической статистике и теории случайных функций, Москва, Наука, 1970.
4. Шор Я.Б. Статистические методы анализа и контроля качества и надежности, Москва, 1962.

ОБ ОДНОМ КЛАССЕ МЕТОДОВ РЕГУЛЯРИЗАЦИИ

Шариф Гусейнов

*Институт математики, Латвийская Академия наук
Площадь Академии 1, Рига LV-1524, Латвия*

*Институт транспорта и связи,
ул. Ломоносова 1, LV-1019, Рига, Латвия
Тел.: (+371)-7100650. Факс: (+371)-7100660
E-mail: sharif@one.lv (для коротких сообщений),
sharif@netmail.lv (для длинных сообщений)*

Моделируя многих задач физики, экономики, статистики, актуарной математики, теории полезности и принятия решений, теории игр и исследования операций и т.д., часто мы приходим к операторному уравнению первого рода. Как правило, эти уравнения относятся к некорректным задачам. Для решения достаточно широких класс некорректных задач в [1] предложен следующий итерационный метод:

$$\begin{aligned} z_0 &= 0, \\ z_{n+1} &= z_n + \theta \cdot (u_p - Az_p), \quad n \geq 0 \end{aligned} \quad (1)$$

где $0 < \theta < \frac{2}{\|A\|}$.

Здесь z_p есть точное решение уравнения первого рода

$$\begin{aligned} Az &= u, \\ z &\in H, \quad u \in AH \end{aligned} \quad (2)$$

при точно заданном $u = u_p \in H$, т.е. $Az_p = u_p$, H – гильбертово пространство, оператор $A: H \rightarrow H$ является линейным, самосопряженным, положительным и вполне непрерывным оператором. Там же доказана сходимост ь итерационного процесса (1) к решению уравнения (2) в норме исходного гильбертова пространства H . Если в уравнение (2) вместо точно заданной правой части $u = u_p \in H$ имеется $u = u_\delta$ так, что

$$\|u_p - u_\delta\| \leq \delta,$$

то, как показано в [1], итерационный процесс

$$\begin{aligned} z^{\delta}_0 &= 0, \\ z^{\delta}_{n+1} &= z^{\delta}_n + \theta \cdot (u_\delta - Az^{\delta}_n), \quad n \geq 0 \end{aligned} \quad (3)$$

сходится к точному решению z_p уравнения (2) в норме исходного гильбертова пространства H и справедливо неравенство

$$\|z^{\delta}_n - z_p\| \leq \|z_n - z_p\| + \varepsilon(n) \cdot \delta,$$

где

$$\varepsilon(n) = \theta \cdot \sum_{k=0}^{n-1} \|E - \theta \cdot A\|^k,$$

E - единичный оператор, а z_n определяется по итерационной формуле (1).

Итак, $z_n^\delta \rightarrow z_p$ в норме пространства H при $n(\delta) \rightarrow \infty$, но скорость сходимости при этом может быть сколь угодно малой, и для оценки ее надо делать предположение об истокорпредставимости точного решения (которого мы не знаем).

В [2] доказана следующая теорема:

ТЕОРЕМА. Пусть известно, что точное решение z_p уравнения (2) истокорпредставимо, т.е.

$$z_p = A^\sigma s, \quad \sigma > 0.$$

Тогда справедливо неравенство

$$\|z_p - z_n^\delta\| \leq \sigma^\sigma \cdot (n \cdot \theta \cdot e)^{-\sigma} \cdot \|s\| + n \cdot \theta \cdot \delta,$$

где $0 < \theta < \frac{2}{\|A\|}$.

Здесь z_n^δ вычисляется по формуле(3).

В данной работе доказывается, что для оценивания скорости сходимости итерационного процесса (3) достаточно сильное дополнительное требование об истокорпредставимости точного решения не потребуется, если применять энергетическую норму и при условии $0 < \theta \leq \frac{4}{3 \cdot \|A\|}$ для итерационного процесса (3)

справедлива следующая теорема:

ТЕОРЕМА. Итерационный процесс

$$\begin{aligned} z_0^\delta &= 0, \\ z_{n+1}^\delta &= z_n^\delta + \theta \cdot (u_\delta - A z_n^\delta), \quad n \geq 0 \end{aligned} \tag{3}$$

сходится в энергетической норме пространства H , если выбирать число итераций n из условия

$$\sqrt{n} \cdot \delta \xrightarrow{n \rightarrow \infty, \delta \rightarrow 0} 0.$$

Кроме того, при условии

$$0 < \theta \leq \frac{4}{3 \cdot \|A\|}$$

для итерационного процесса (3) справедливы оценки погрешности

$$\|z_p - z_n^\delta\|_A \leq (2 \cdot n \cdot \theta \cdot e)^{\frac{1}{2}} \cdot \|z_p\|_A + \left(\frac{4}{3} \cdot n \cdot \theta\right)^{\frac{1}{2}} \cdot \delta, \quad n \geq 1,$$

$$\|z_p - z_n^\delta\|_A \leq (2 \cdot n \cdot \theta \cdot e)^{\frac{1}{2}} \cdot \|z_p\|_A + \left(\frac{13}{20} \cdot n \cdot \theta\right)^{\frac{1}{2}} \cdot \delta, \quad n \geq 2.$$

Литература

1. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – Москва: Наука, 1986.
2. Вайникко Г.М. Об одном классе методов регуляризации при наличии априорной информации о решении. – Тарту, 1984.