9th INTERNATIONAL SYMPOSIUM ON MEASUREMENT AND QUALITY CONTROL (9th ISMQC)
November 21 – 24, 2007, IIT Madras

ABOUT REASONABLE NUMBER OF RANKINGS IN PREFERENCE PROFILE WHEN MEASURING QUALITY

Sergey V. Muravyov
Department of Computer-aided Measurement Systems and Metrology
Tomsk Polytechnic University, Tomsk, 634050, RUSSIA
E-mail: muravyov@camsam.tpu.ru

Abstract: *To plan the quality measuring in the form of consensus relation determination for the given m weak order relations (rankings) it is necessary to know a reasonable number of the rankings. If a ranking is produced by an expert then the number of rankings is equal to number of experts. It is proposed to estimate the expert number using simple probabilistic Bernoulli model, where m experts reveal defects (demerits) of an object. The model assumes that the more the number of an expert group participants, the less the probability of a new defect revealing. Based on this assumption, the probability decrease have been evaluated and graphically presented. The investigations allow to suppose that the number of rankings in preference profile can be from 4 to 10 for typical applications.*
Key words: *Quality Measurement; Ordinal Scale Measurement; Number of Weak Orders, Number of Experts*

1. INTRODUCTION

An estimation of object(s) quality by different criteria is frequently carried out by the groups consisting of decision makers, i.e. experts. Some examples of the expert-based quality measurement can be found in (Kershenbaum & Khvastunov, 1999; Ng & Abramson, 1992); Rossi et al., 2005).

When organizing an expert examination, it is important to know *how many experts are necessary and sufficient for obtaining thorough results of the examination procedure*. An approaches to the problem were proposed in (Kershenbaum & Khvastunov, 1999; Nielsen & Landauer, 1993; Nielsen, 1994) and other works. However, the issue still need further inquiry.

Our specific interest to the issue originates from the possibility to measure quality *in ordinal scale* by the following way. Suppose we have *m rankings* on set $A = \{a_1, a_2, ..., a_n\}$ of *n objects*. Then we have the relation set $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_m\}$, where each of *m* rankings (preference relations) $\lambda = \{a_1 \succ a_2 \succ ... \sim a_s \sim a_t \succ ... \sim a_n\}$ may include $\succ$, a strict preference relation π, and ~, an equivalence (or indifference) relation ν, so that $\lambda = \pi \cup \nu$. Such a relation λ is generally called a *weak order*. The relation set Λ is titled a *preference profile* for the given *m* rankings. We can determine a single preference relation that would give an integrative characterization of the objects. Let a subspace Π be a set of all *n*! linear (strict) order relations $\succ$ on *A*. Each linear order corresponds to one of permutations of first *n* natural numbers $\mathbf{N_n}$. We use a permutation $\beta \in \Pi$ of the alternatives $a_1, ..., a_n$ to represent the preference profile Λ and will call it *consensus ranking*.

The problem described is also the *problem of voting* or group decision where *A* is a set of *n* alternatives or candidates which are ranked by group of *m* individuals (multisensors, voters, experts, focus groups, criteria, etc.). On the basis of a distance between rankings, we can define the distance $D(\beta, \Lambda)$ from β to the profile Λ and then formulate a quality measuring problem as

$$\beta = \arg\min_{\lambda \in \Pi} D(\lambda, \Lambda) , \qquad (1)$$

where $D(\lambda, \Lambda) = \sum_{k=1}^{m} d(\lambda, \lambda_k)$.

How to find a solution of the problem (1) is described, for example, in (Barthélemy, 1989; Muravyov, 2006; Muravyov, 2007).

In this paper, we are interested in estimation of an upper bound for the number of experts (rankings) *m*. At that, we are aware of absolutely exact estimations for *m* cannot be determined. So we will search for its approximate values.

The remainder of the paper is devoted to an attempt of the problem solution.

2. PROBABILISTIC MODEL

In the paper, we use a probabilistic approach to the problem of calculating number of experts and as the examination model we also use a situation where *m* experts reveal *defects* (*demerits*) of an object under quality testing. The notation of defect has here a wide sense as the objects may be of very different origin. Particularly, by defect one can understand the failure of a product to conform to specification or the non-conformance to intended usage requirement.

Let us use the following simple model based on Bernoulli trial, see, for example (Gnedenko, 1963). Suppose we have an expert group consisting of *m* experts revealing defects independently of each other. The defect findings also are independent of whether they have been found before. Let *p* be the probability of finding the average defect by a single "average" expert (we will cal it elementary probability). Then the probability *P* of that at least one defect is revealed by *m* evaluators is defined as follows:

$$P = 1-(1-p)^m, \tag{2}$$

On the other hand, the probability P can be determined taking into account its frequency interpretation. Let D_t be the total number of defects and D_f be the number of defects that have been found at least once by m experts. Then we have

$$P \simeq \frac{D_f}{D_t}, \tag{3}$$

Finally, having regard to equations (2) and (3), we obtain the following expression for the number of defects found:

$$D_f \simeq D_t[1-(1-p)^m], \tag{4}$$

The similar model (however, referring to Poisson process) has been used in (Nielsen & Landauer, 1993; Nielsen, 1994) in order to estimate the amount of evaluation required to detect so called usability problems in a user interface design.

Clearly, the number of experts can be easily obtained from the formula (2), i.e.

$$m = \frac{\ln(1-P)}{\ln(1-p)}. \tag{5}$$

The graph constructed by formula (1) (see Fig. 1) shows that there is some critical value m_c of m such that any $m > m_c$ does not give an essential increase of number of defects found. For example, at $p = 0.6$, there is no necessity to have more 5 experts as these five experts have found practically all defects.

One more proposition can be made of consideration of Fig. 1: *the more the number of an expert group participants, the less the probability of a new defect revealing*. The proposition has been suggested also in (Kershenbaum & Khvastunov, 1999).

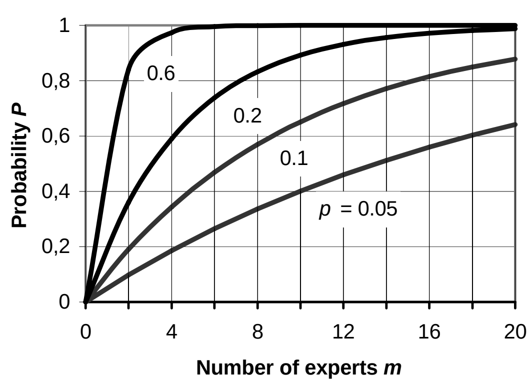

Fig. 1. Graphic presentation of the probability P depending on m for different values of p, see expression (2).

3. PROBABILITY OF NEW DEFECTS REVEALING DUE TO ADDITIONAL EXPERTS

It is interesting to investigate how the probability P will increase after adding one more evaluator to the expert group. The following formula shows how many times the probability $P(m+1) = P_1$ is greater than the probability $P(m) = P$:

$$\frac{P_1}{P} = \frac{1-(1-p)^{m+1}}{1-(1-p)^m} = 1 + \frac{p(1-p)}{1-(1-p)^m}. \tag{6}$$

It can be seen from Fig. 2 that the increase of manpower by one more expert result in a minor gain of the probability P of defects revealing. This gain becomes especially insignificant for all numbers $m > m_c = 4$. And the more probability p the more this insignificance.

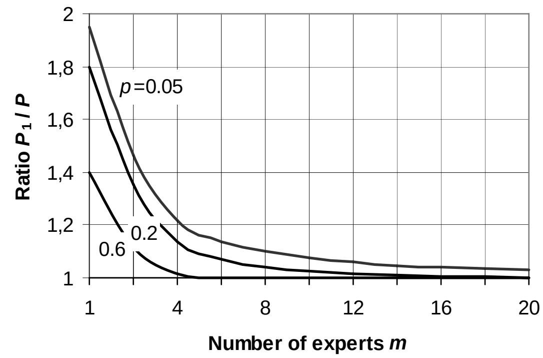

Fig. 2. Graph of the ratio P_1/P depending on m for different values of p, see expression (6).

It is worth to estimate this gain in explicit and more general form. Let α be the relative probability growth resulting from inclusion of k additional experts into the group consisting of m experts, i.e.

$$\alpha = \frac{P(m+k)-P(m)}{P(m)} = \frac{P_k - P}{P}, \tag{7}$$

where

$$P_k = 1-(1-p)^m(1-p)^k, \tag{8}$$

From (2), (7) and (8) we have

$$\alpha = (1-p)^m \frac{1-(1-p)^k}{1-(1-p)^m}. \tag{9}$$

Calculations of α are reduced in Table 1 and graphically presented in Fig. 3.

As Table 1 and Fig. 3 indicate an essential gain of probability of revealing new defects due to attraction of additional k exists only if the elementary probability p is low (see Fig.3, $p = 0.05$). In this case, the dependency $\alpha(k)$ has almost linear character. However, already at $m = 7$, doubling of found defects number ($\alpha = 100\%$) happens only where $k = 10$.

At $p = 0.5$, if the expert group has included 4 members, attraction of new expert is useless as it gives out no new data to test a quality of the object. If $p > 0.5$, one can see the loss of necessity in new experts as early as $k = 5$, though the expert group consists of a single member.

Looking at Table 1 one can see that *it is important in what fashion an expert group was set up*. Indeed, if $p = 0.05$, $m = 2$ and $k = 8$ give the relative growth $\alpha = 3.11$, whereas $m = 4$ and $k = 6$ produce only $\alpha = 1.16$, and at that the total number of evaluators is the same: $m + k = 10$. Thus, a combination of m and k with their sum fixed results in greater growth α, if $m < k$.

Table 1. Values of the relative probability growth $\alpha(k)$ for different numbers m, see expression (9).

	k	$m = 1$	$m = 2$	$m = 4$	$m = 7$
$p = 0.05$	0	0	0	0	0
	1	0.95	0.46	0.21	0.11
	2	1.85	0.90	0.42	0.22
	3	2.71	1.32	0.63	0.33
	4	3.52	1.72	0.81	0.42
	5	4.30	2.09	0.99	0.52
	6	5.03	2.45	1.16	0.61
	7	5.73	2.79	1.32	0.70
	8	6.39	3.11	1.48	0.78
	9	7.02	3.42	1.62	0.86
	10	7.62	3.71	1.76	0.93
$p = 0.5$	0	0	0	0	0
	1	0.5	0.17	0.03	0.004
	2	0.75	0.25	0.05	0.006
	3	0.87	0.29	0.06	0.007
	4	0.94	0.31	0.06	0.007
	5	0.97	0.32	0.06	0.008
	6	0.98	0.33	0.07	0.008
	7	0.99	0.33	0.07	0.008
	8	1.0	0.33	0.07	0.008
	9	1.0	0.33	0.07	0.008
	10	1.0	0.33	0.07	0.008
$p = 0.8$	0	0	0	0	0
	1	0.2	0.03	0.0013	0
	2	0.24	0.04	0.0015	0
	3	0.25	0.04	0.0016	0
	4	0.25	0.04	0.0016	0
	5	0.25	0.04	0.0016	0
	6	0.25	0.04	0.0016	0
	7	0.25	0.04	0.0016	0
	8	0.25	0.04	0.0016	0
	9	0.25	0.04	0.0016	0
	10	0.25	0.04	0.0016	0

The number k can be easily determined in explicit form using expression (8). That is

$$k = \frac{\ln(1-P_k)}{\ln(1-p)} - m = m\frac{\ln(1-P_k)}{\ln(1-P)} - m \,, \qquad (10)$$

In practice, the number k can be calculated on the assumption of desirable or critical value of P_k known. Surely, the elementary probability p should also be given or estimated.

The probability p is usually recommended to be found approximately by results of practical expert examinations. (Nielsen & Landauer, 1993) have proposed to use the following estimations:

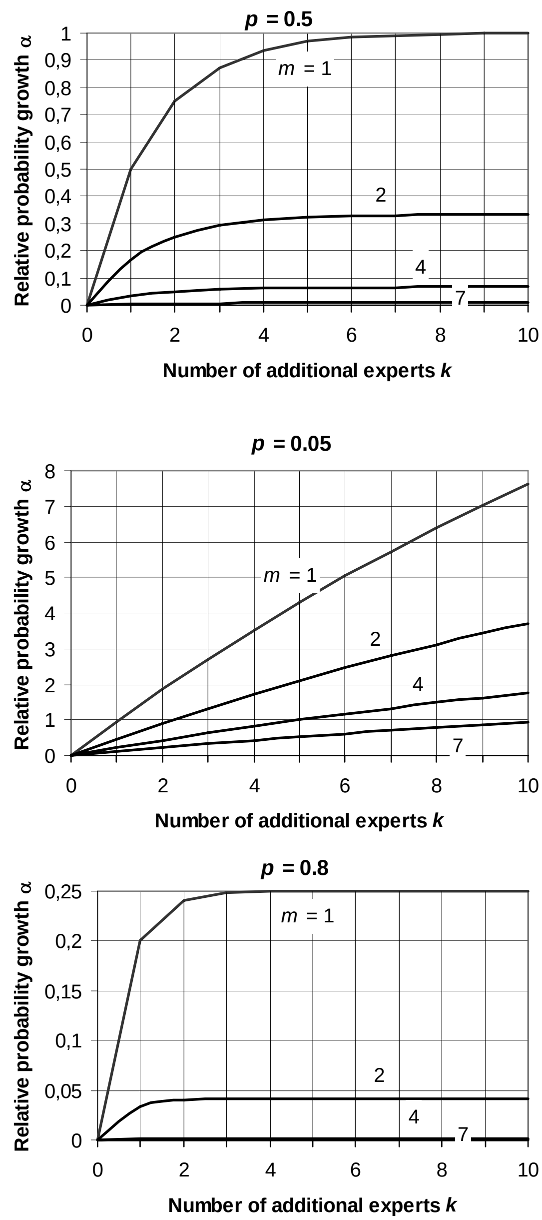

Fig. 3. Graphs (for $p = 0.05$, 0.5 and 0.8) of the relative probability growth α depending on number of additional evaluators k for different numbers m, see expression (9).

$$D_t \simeq \frac{D_f(1)}{p} \qquad (11)$$

and

$$p \simeq 2 - \frac{D_f(2)}{D_f(1)} \,, \qquad (12)$$

where $D_f(1)$ and $D_f(2)$ are *average* numbers of different defects found by a single expert and a pair of experts correspondingly. Expressions (11) and (12) are easily derived from (4).

As a supplementary remark, it can be found to be useful the following form of expression (2):

$$P = p\sum_{i=0}^{m-1}(1-p)^i \qquad (13)$$

To make sure of validity of (13) it is enough to look at Table 2, where expressions of *P* for particular *m* have been obtained after simple removal of parentheses:

Table 2. Towards justification of expression (13).

m	*P*
1	p
2	$p(1+(1-p))$
3	$p(1+(1-p)+(1-p)^2)$
4	$p(1+(1-p)+(1-p)^2+(1-p)^3)$
…	…
m	$p(1+(1-p)+(1-p)^2+...+(1-p)^{m-1})$

4. CONCLUSION

It would not be out of place to notice that the probabilistic model discussed above is based on assumptions that hardly characterize real situation of the expert examination. In fact, probabilities of different defects revealing are different and often dependent on each other. In this relation, developments of new models more exactly describing the situation are welcome.

Nevertheless, the discussed model allows to obtain interesting and useful recommendations on selection of number of an expert group members when measuring quality. This way, after consideration of results presented in the text, we can conclude that the reasonable number of experts (or rankings in preference profile) can be from 4 to 10 for typical applications.

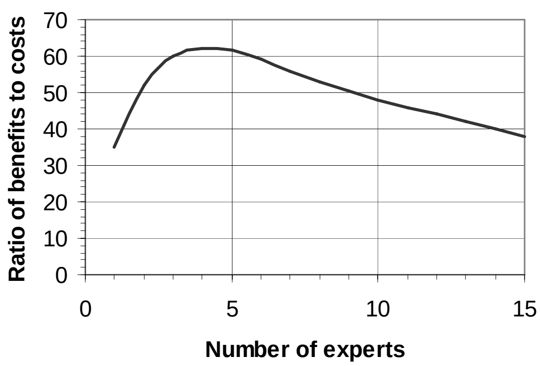

Fig. 4. Ratio of benefits to costs of expert examinations depending on number of evaluators (Nielsen, 1994).

If additionally to take into account some economical considerations, for example, to evaluate costs for attraction of experts and corresponding profit from revealed defects then the expert number estimated could be refined. It was done in (Nielsen, 1994) for the case of heuristic evaluation of software interface. Taken from the paper curve in Fig. 4 shows how many times the benefits are greater than the costs of a sample project. The optimal number of evaluators in this example is four, with benefits that are 62 times greater than the costs.

5.REFERENCES

Barthélemy, J.P.; Guenoche, A. & Hudry, O. (1989). Median linear orders: heuristics and a branch and bound algorithm, *European Journal of Operational Research*, Vol. 42, 313-325.

Gnedenko, B.V. (1963). *The Theory of Probability*. Chelsea Publishing Co.

Kershenbaum, V.Ya. & Khvastunov, R.M. (Eds.) (1999) *Qualimetry methods in Mechanical Engineering*, Tekhnoneftegaz, ISBN 5-93157-002-0, Moscow.

Muravyov, S.V. (2006) Qualitative measurement uncertainty assessment and ensuring. *Proceedings of the 9th International Conference "Quality Management and Organisational Development"* (QMOD 2006), pp. 636-647, ISBN 0-9553349-0-0, Liverpool, 9-11 August 2006, John Moores University, Liverpool, UK.

Muravyov, S.V. (2007). Rankings as ordinal scale measurement results. *Metrology and Measurement Systems*, Vol. XIII, No. 1, 9-24.

Ng, K.-C. & Abramson, B. (1992). Consensus Diagnosis: A Simulation Study, *IEEE Transactions on systems, man, and cybernetics*, Vol. 22, No. 5, (September/October, 1992), 916-928.

Nielsen, J. & Landauer, T.K. (1993). A mathematical model of the finding of usability problems. *Proceedings of ACM/IFIP INTERCHI'93 Conference*, pp. 206-213, ISBN 90-5199-133-9, Amsterdam, The Netherlands, 24-29 April 1993, IOS Press, Amsterdam.

Nielsen, J. (1994). Heuristic evaluation. In: *Usability Inspection Methods*, Nielsen, J. & Mack, R.L. (Eds.), 55-62, John Wiley & Sons, ISBN 0-471-01877-5, New York, NY.

Rossi, G.B.; Crenna, F. & Panero, M. (2005). Panel or jury testing methods in a metrological perspective, *Metrologia*, Vol. 42, 97–109.