APPLICATIONES MATHEMATICAE
29,2 (2002), pp. 135–144

Alicja Jokiel-Rokita and Ryszard Magiera (Wrocław)

THE BAYES CHOICE OF AN EXPERIMENT IN ESTIMATING A SUCCESS PROBABILITY

Abstract. A Bayesian method of estimation of a success probability p is considered in the case when two experiments are available: individual Bernoulli (p) trials—the *p-experiment*—or products of r individual Bernoulli (p) trials—the *p^r-experiment*. This problem has its roots in reliability, where one can test either single components or a system of r identical components. One of the problems considered is to find the degree $\widetilde{r}$ of the $p^{\widetilde{r}}$-experiment and the size $\widetilde{m}$ of the p-experiment such that the Bayes estimator based on $\widetilde{m}$ observations of the p experiment and $N-\widetilde{m}$ observations of the $p^{\widetilde{r}}$-experiment minimizes the Bayes risk among all the Bayes estimators based on m observations of the p-experiment and $N-m$ observations of the p^r-experiment. Another problem is to sequentially select some combination of these two experiments, i.e., to decide, using the additional information resulting from the observation at each stage, which experiment should be carried out at the next stage to achieve a lower posterior expected loss.

1. Introduction. Consider a sequence of Bernoulli trials with success probability p. To estimate p, two experiments can be performed: one can observe either an individual trial outcome (the *p-experiment*), or the product of r individual trial outcomes (the *p^r-experiment*), where r is an integer, $r \geq 2$. A total of N experiments (tests) are performed, and the problem is

- to find a pair $(\widetilde{r}, \widetilde{m})$, where $\widetilde{r}$ is the size of the alternative to the p-experiment, and $\widetilde{m}$ is the number of p-experiments in a sequence of N experiments, so that the final Bayes estimator of the parameter p gives the lowest Bayes risk;

2000 *Mathematics Subject Classification*: Primary 62L12, 62L15; Secondary 62C10.
Key words and phrases: allocation, Bayes estimation, Bayes risk, loss function, posterior expected loss, prior distribution.

• to sequentially select some combination of these two experiments so that at each stage, the information available at that stage can be used to determine which experiment to carry out at the next stage in order to achieve a lower posterior expected loss.

The p^r-experiment is a slightly disguised version of the well-studied grouped data experiment with groups of size r. This type of sampling has been shown to reduce the cost of classifying all the members of a population according to whether or not they possess a certain trait, when the incidence rate is fairly low (Dorfman (1943)). It has also been used to estimate the failure probability, $q = 1-p$, and using groups of sizes other than one can reduce the cost of testing (Sobel and Elashoff (1975)), and can lower the variance of the resulting estimator (Chen and Swallow (1990)). Reliability settings, in which components can be tested either individually, or as a system of r identical components in a series, are prime examples of situations in which group testing can be useful (Easterling and Prairie (1971)). Other group testing scenarios arise in environmental monitoring, where sample units of soil or plant matter are combined and tested for toxins. In these settings, the term "group testing" is often replaced by "composite sampling". (For a review of composite sampling methods see Lancaster and Keller-McNulty (1996).) Gastwirth and Hammick (1989), for example, applied group testing methods to estimate the prevalence of human immune virus (HIV) antibodies among subpopulations. In screening scenarios of this sort, group testing is particularly desirable, because it provides donor privacy, an issue of serious concern among individuals at risk of HIV. The "pooled testing" of Tu, Litvak, and Pagano (1995) is another example in which group testing is used to estimate HIV prevalence.

In this paper the problem is placed in a Bayesian framework with squared error estimation loss. In Section 2, the notation is presented and the problem is precisely defined. In Section 3, the allocation that minimizes the Bayes risk is derived, and in Section 4, an ad hoc adaptive sequential procedure is proposed, and on the basis of computer simulations, it is shown that this procedure gives a lower Bayesian expected loss on average of the Bayes terminal estimator.

2. Notation and description of the problem. We set up the problem in its fully sequential form, although much of the development in the next sections will not use all of this notation. Let $X_{11}, X_{12}, \ldots$ be a sequence of iid Bernoulli (p) random variables that are independent of $X_{r1}, X_{r2}, \ldots$ iid Bernoulli p^r random variables. A total of N tests (experiments) will be done, where at each stage the decision to carry out a p-experiment or p^r-experiment can be made based on past observations. More precisely, an *allocation rule* is a sequence $a = (a_1, \ldots, a_N)$ such that for $k = 1, \ldots, N$, a_k

takes values 0 or 1 and is measurable with respect to $\{Z_1, \ldots, Z_{k-1}\}$, where $Z_i = a_i X_{1i} + (1 - a_i)X_{ri}$. Thus, a_i indicates the population from which the ith observation or test is sampled, with 1 indicating the p-experiment and 0 indicating the p^r-experiment. The terminal estimator must be measurable $\{Z_1, \ldots, Z_N\}$. Finally, let $m_k = \sum_{i=1}^{k} a_i$ be the total number of observations taken from p-experiments at stage k and $n_k = k - m_k$ be the total number of observations taken from p^r-experiments at stage k, where $k = 1, \ldots, N$. In the case $k = N$, the k subscript will be dropped.

When m observations from p-experiments and n observations from p^r-experiments are taken, then a sufficient statistic for the parameter p is the vector $(X_1^{(m)}, X_r^{(n)})$, where $X_1^{(m)}$ denotes the number of successes in the p-experiments, and $X_r^{(n)}$ the number of successes in the p^r-experiments. The initial goal is to find the Bayes estimator $d^*(X_1^{(m)}, X_r^{(n)})$ of the parameter p, i.e., the estimator for which

$$R(\pi, d^*) = \inf_d R(\pi, d),$$

where $R(\pi, d) = E^{\pi}[E_p L(p, d)]$ is the Bayes risk of the estimator d with respect to the prior distribution π of the parameter $p \in \mathcal{P} = (0, 1)$. The function $L(p, d)$ denotes the loss function connected with the estimator d, and it is supposed to be the squared estimation error loss function of the form

$$L(p, d) = (p - d)^2. \tag{1}$$

The choice of the quadratic loss function is reasonable if we pay attention to the error in the same degree for all values p, irrespective of whether the true value p is near to 0 or 1. Otherwise one should consider, for instance, the weights $1/p$, $1/(1 - p)$ or $1/(p(1 - p))$.

We assume that the prior distribution π is of the form

$$f_r(p; \alpha, \beta, \gamma) = \frac{1}{M_r(\alpha, \beta, \gamma)}\, p^{\alpha-1}(1 - p)^{\beta-1}(1 - p^r)^{\gamma-1} I_{(0,1)}(p), \tag{2}$$

where $\alpha, \beta, \gamma \in (0, \infty)$, $r \in \mathbb{N}$, and

$$M_r(\alpha, \beta, \gamma) = \int_0^1 p^{\alpha-1}(1 - p)^{\beta-1}(1 - p^r)^{\gamma-1}\, dp. \tag{3}$$

If observations from p- and p^r-experiments are taken, then the prior distribution above is a natural conjugate distribution, nd we will denote it by $\pi_r(\alpha, \beta, \gamma)$.

LEMMA 1. *The $M_r(\alpha, \beta, \gamma)$ function has the following properties*:

1. $M_r(\alpha, \beta, 1) = B(\alpha, \beta)$, *where $B(\alpha, \beta)$ denotes the beta function*;
2. $M_r(\alpha + 1, \beta, 1) = \frac{\alpha}{\alpha+\beta} M_r(\alpha, \beta, 1)$;
3. $M_r(\alpha, \beta + 1, \gamma) = M_r(\alpha, \beta, \gamma) - M_r(\alpha + 1, \beta, \gamma)$;

4. $M_r(\alpha+1,\beta,\gamma) = M_r(\alpha,\beta,\gamma) - M_r(\alpha,\beta+1,\gamma)$;
5. $M_r(\alpha,\beta,\gamma+1) = M_r(\alpha,\beta,\gamma) - M_r(\alpha+r,\beta,\gamma)$;
6. $$M_r(\alpha,\beta,\gamma) = \sum_{i=0}^{\gamma-1} \binom{\gamma-1}{i} (-1)^i M_r(\alpha+ir,\beta,1)$$
$$= \sum_{i=0}^{\gamma-1} \binom{\gamma-1}{i} (-1)^i B(\alpha+ir,\beta);$$
7. $$M_r(\alpha+1,\beta,\gamma) = \sum_{i=0}^{\gamma-1} \binom{\gamma-1}{i} (-1)^i M_r(\alpha+ir,\beta,1)\,\frac{\alpha+ir}{\alpha+ir+\beta}$$
$$= \sum_{i=0}^{\gamma-1} \binom{\gamma-1}{i} (-1)^i B(\alpha+ir,\beta)\,\frac{\alpha+ir}{\alpha+ir+\beta}.$$

3. The minimum Bayes risk allocation. The following lemma gives the form of the Bayes estimator of the parameter p with respect to the prior distribution $\pi_r(\alpha,\beta,\gamma)$ in the case when m observations from p-experiments and n observations from p^r-experiments ($m+n=N$) are taken.

LEMMA 2. *Under the loss function given by* (1), *if m observations from p-experiments and n observations from p^r-experiments are taken, then the estimator*

$$d^*(X_1^{(m)}, X_r^{(n)})$$

$$= \frac{M_r(\alpha+X_1^{(m)}+rX_r^{(n)}+1,\beta+m-X_1^{(m)},\gamma+n-X_r^{(n)})}{M_r(\alpha+X_1^{(m)}+rX_r^{(n)},\beta+m-X_1^{(m)},\gamma+n-X_r^{(n)})} \tag{4}$$

$$= \frac{\displaystyle\sum_{i=0}^{\gamma+n-X_r^{(n)}-1} \frac{B(\alpha+X_1^{(m)}+rX_r^{(n)}+ir,\beta+m-X_1^{(m)})}{(-1)^i(\gamma+n-X_r^{(n)}-1-i)!\;i!}\,\frac{\alpha+X_1^{(m)}+rX_r^{(n)}+ir}{\alpha+\beta+rX_r^{(n)}+ir+m}}{\displaystyle\sum_{i=0}^{\gamma+n-X_r^{(n)}-1} \frac{B(\alpha+X_1^{(m)}+rX_r^{(n)}+ir,\beta+m-X_1^{(m)})}{(-1)^i(\gamma+n-X_r^{(n)}-1-i)!\;i!}} \tag{5}$$

is Bayes with respect to the prior distribution $\pi_r(\alpha,\beta,\gamma)$, and its Bayes risk is

$$R(\pi_r(\alpha,\beta,\gamma),d^*) \tag{6}$$
$$= \frac{1}{M_r(\alpha,\beta,\gamma)} \sum_{i=0}^{m}\sum_{j=0}^{n} \binom{m}{i}\binom{n}{j} \Big[M_r(\alpha+i+rj+2,\beta+m-i,\gamma+n-j) - \frac{M_r^2(\alpha+i+rj+1,\beta+m-i,\gamma+n-j)}{M_r(\alpha+i+rj,\beta+m-i,\gamma+n-j)} \Big].$$

Proof. The probability of observing x_1 successes in the sequence of p-experiments, and x_r successes in the sequence of p^r-experiments, given p, is equal to

$$P(X_1^{(m)} = x_1,\, X_r^{(n)} = x_r \,|\, p) = \binom{m}{x_1}\binom{n}{x_r} p^{x_1+rx_r}(1-p)^{m-x_1}(1-p^r)^{n-x_r},$$

where $x_1 \in \{0,1,\dots,m\}$, $x_r \in \{0,1,\dots,n\}$. Hence, the density function of the posterior distribution of the parameter p is of the form (2) with $\alpha := \alpha + x_1 + rx_r$, $\beta := \beta + m - x_1$, $\gamma := \gamma + n - x_r$. Under the loss function given by (1), the Bayes estimator $d^*(X_1^{(m)}, X_r^{(n)})$ of p is equal to the expected value of the posterior distibution, so it is of the form (4), and by items 6 and 7 of Lemma 1, it is of the form (5).

Denote by $\pi^* = \pi_r(\alpha + x_1 + rx_r, \beta + m - x_1, \gamma + n - x_r)$ the posterior distribution of p. The Bayesian expected loss is equal to the variance of the posterior distribution, and is of the form

$$(7)\quad E^{\pi^*} L(p, d^*) = \frac{M_r(\alpha + x_1 + rx_r + 2, \beta + m - x_1, \gamma + n - x_r)}{M_r(\alpha + x_1 + rx_r, \beta + m - x_1, \gamma + n - x_r)} - \frac{M_r^2(\alpha + x_1 + rx_r + 1, \beta + m - x_1, \gamma + n - x_r)}{M_r^2(\alpha + x_1 + rx_r, \beta + m - x_1, \gamma + n - x_r)}.$$

The distribution of the statistic $(X_1^{(m)}, X_r^{(n)})$ is

$$P(X_1^{(m)} = x_1,\, X_r^{(n)} = x_r) = \frac{1}{M_r(\alpha,\beta,\gamma)}\binom{m}{x_1}\binom{n}{x_r} \times \int_0^1 p^{\alpha+x_1+rx_r-1}(1-p)^{\beta+m-x_1-1}(1-p^r)^{\gamma+n-x_r-1}\,dp.$$

Hence, the Bayes risk $R(\pi_r(\alpha,\beta,\gamma), d^*)$ (the expected value of the Bayesian expected loss) of the Bayes estimator $d^*(X_1^{(m)}, X_r^{(n)})$ is given by (6). ∎

If the prior parameters α, β, γ of the distribution $\pi_r(\alpha,\beta,\gamma)$ are known, and the sample size N is fixed, then the Bayes risk of the Bayes estimator d^* depends on the number m of p-experiments and on the size r of the alternative experiment to the p-experiment.

Write $R_N^{\alpha,\beta,\gamma}(r,m) = R(\pi_r(\alpha,\beta,\gamma), d^*(X_1^{(m)}, X_r^{(N-m)}))$ to emphasize that we will consider the Bayes risk for d^* as a function of r and m.

Proposition 1. *For given α, β, γ and N, there exists an r_0 such that for all $r > r_0$, $R_N^{\alpha,\beta,\gamma}(r,m)$ is a decreasing function of m.*

Set $R_N^{\alpha,\beta}(r,m) = R_N^{\alpha,\beta,1}(r,m)$. If $\gamma = 1$, i.e., when the prior distribution is a beta distribution $B(\alpha,\beta)$, we might be interested in finding a pair $(\widetilde{r}, \widetilde{m})$

for which

$$R_N^{\alpha,\beta}(\widetilde{r},\widetilde{m}) = \inf_{(r,m)\in I} R_N^{\alpha,\beta}(r,m),$$

where $I = \mathbb{N} \times \{0, 1, \ldots, N\}$.

Let us remark that $R_N^{\alpha,\beta,\gamma}(r,N) = R_N^{\alpha,\beta,\gamma}(2,N)$ for all $r \in \{2,3,\ldots\}$. Hence for $\widetilde{m} = N$ we assume that $\widetilde{r} = 2$. Unfortunately, we cannot give an explicit general formula for such an optimal pair $(\widetilde{r},\widetilde{m})$. Applying numerical evaluation leads to the following assertions.

REMARKS 1. 1. The optimal pair $(\widetilde{r},\widetilde{m})$ depends not only on the parameters α, β of the prior distributions, but also on the sample size N: $(\widetilde{r},\widetilde{m}) = (r(\alpha,\beta,N), m(\alpha,\beta,N))$.

2. For $\beta \geq \alpha$, we have $\widetilde{r} = 2$. In this case the value of $\widetilde{m}$ depends on α, β and N.

3. For $\beta > 2\alpha$, the optimal $\widetilde{m}$ equals N, i.e., we should take observations only from the p-experiment.

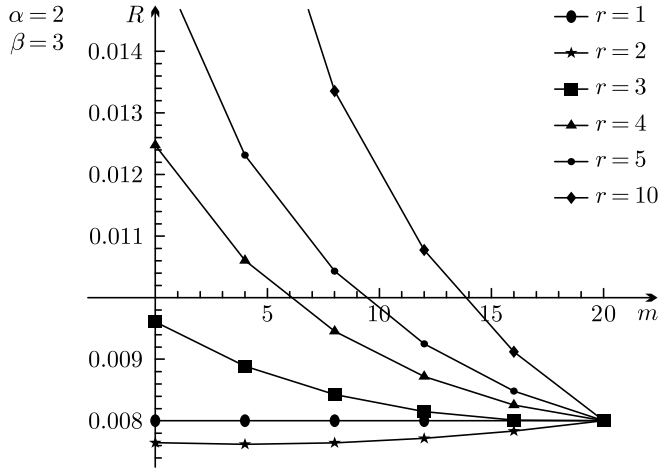
4. For $\beta < \alpha$, the optimal $\widetilde{m}$ is 0, i.e., we should take observations only from the $p^{\widetilde{r}}$-experiment. In this case the value of $\widetilde{r}$ depends on α, β and N.

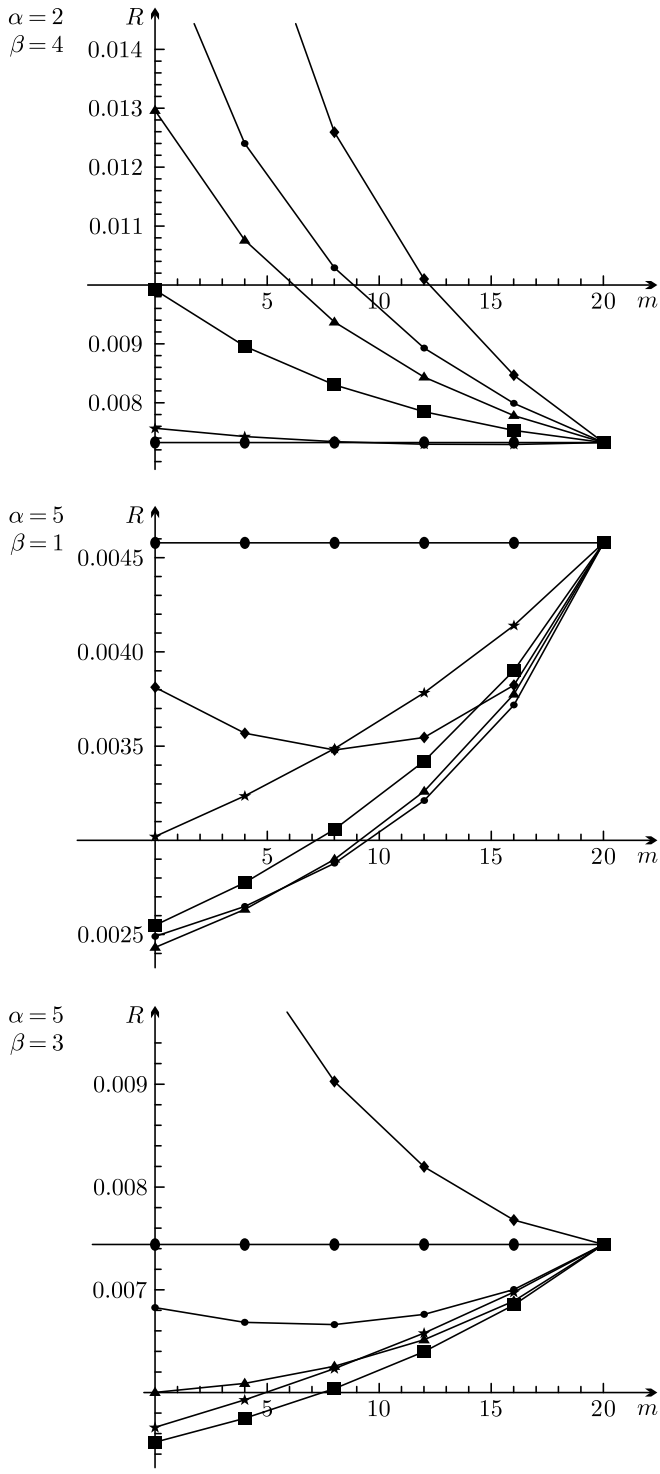
5. If $\widetilde{r} > 2$, then $\widetilde{m} = 0$.

For some illustration of the remarks above see the figures below.

It follows from the conclusions above that we do not need to search for an optimal pair $(\widetilde{r},\widetilde{m})$ in the case $\beta > 2\alpha$, because then $(\widetilde{r},\widetilde{m}) = (2,N)$. In the case when $\alpha \leq \beta \leq 2\alpha$, we just look for an optimal $\widetilde{m}$, since $\widetilde{r} = 2$. For $\beta < \alpha$, we can limit ourselves to finding the optimal $\widetilde{r}$, because then $\widetilde{m} = 0$.

Figures: The Bayes risk R as a function of the number of p-experiments for various values of α, β and r




α = 2
β = 4
R
0.014
0.013
0.012
0.011
0.009
0.008
5
10
15
20
m
α = 5
β = 1
R
0.0045
0.0040
0.0035
0.0025
5
10
15
20
m
α = 5
β = 3
R
0.009
0.008
0.007
5
10
15
20
m

In order to emphasize the need of searching for the optimal $(\widetilde{r}, \widetilde{m})$, consider

$$\varrho(\alpha, \beta, N) := \frac{R_N^{\alpha,\beta}(\widetilde{r}, \widetilde{m})}{R_N^{\alpha,\beta}(2, N)},$$

the ratio of the Bayes risk under the optimal approach and the Bayes risk under the "conservative" approach, i.e., taking only observations from the p-experiment.

REMARKS 2. The function $\varrho(\alpha, \beta, N)$ has the following properties:

1. If $\beta \leq 2\alpha$, then for fixed α and β, it is a decreasing function of N; for fixed α and N, it is an increasing function of β; for fixed β and N, it is a decreasing function of α.

2. If $\beta > 2\alpha$, then $\varrho(\alpha, \beta, N) = 1$.

For illustration we give some values of the ϱ function: $\varrho(3, 2, 10) = 0.804$, $\varrho(3, 1, 10) = 0.681$, $\varrho(17, 1, 10) = 0.379$, $\varrho(5, 1, 10) = 0.572$, $\varrho(5, 1, 20) = 0.531$.

4. Sequential allocation minimizing the posterior expected loss. Applying the optimal allocation procedure described in the previous section gives us the possibility of minimizing the mean value of the posterior expected loss, i.e., the Bayes risk. In this approach we do not use any additional information from the observations. In contrast, in the sequential approach, at each stage, this additional information may be used to decide which experiment should be carried out at the next stage. Sequential decisions between two experiments for estimating a success probability were applied by Hardwick, Page and Stout (1998). They have shown, among other things, that the a_r-cut allocation minimizes the asymptotic mean squared error of the maximum likelihood estimator of the parameter p, and that the adaptive a_r-cut procedure, replacing the MLE with the Bayes estimator, is asymptotically Bayes. We have applied this adaptive procedure and two other procedures in computer simulations to check if they give, on average, a lower posterior expected loss of the final Bayes estimator.

The adaptive a_r-cut procedure applied with the Bayes estimator is as follows: use the Bayes estimator to estimate the parameter p, and if this estimator is below (above) the cutpoint a_r, then observe an individual (product) trial at the next stage. The cutpoint a_r is the unique root in $(0, 1)$ of the equation in p,

$$p^r(1 - p^r) + r^2 p^{r-1} - 1 = 0,$$

which is obtained by equating to zero the difference between the Fisher information about p contained in a single observation of the p-experiment

and in a single observation of the p^r-experiment (see Hardwick, Page and Stout (1998)). Some of the values of a_r (first reported by Loyer (1983)) are as follows: $a_2 = 0.333$, $a_5 = 0.536$, $a_{10} = 0.679$, $a_{20} = 0.792$, $a_{50} = 0.892$, $a_{100} = 0.937$.

Other ways of choosing the experiment at stage k may be based on the minimization of the Bayes risk of the Bayes estimator with respect to the posterior distribution of the parameter p after $k-1$ observations. Two procedures can be considered. At stage k, take an observation from the p-experiment if the Bayes risk of the Bayes estimator after one observation of this experiment is less than after one observation from the p^r-experiment, i.e., if

$$\frac{M_r^2(\alpha_k+1,\beta_k+1,\gamma_k)}{M_r(\alpha_k,\beta_k+1,\gamma_k)}+\frac{M_r^2(\alpha_k+2,\beta_k,\gamma_k)}{M_r(\alpha_k+1,\beta_k,\gamma_k)}-\frac{M_r^2(\alpha_k+1,\beta_k,\gamma_k+1)}{M_r(\alpha_k,\beta_k,\gamma_k+1)} - \frac{M_r^2(\alpha_k+r+1,\beta_k,\gamma_k)}{M_r(\alpha_k+r,\beta_k,\gamma_k)}>0,$$

where α_k, β_k and γ_k are the parameters of the posterior distribution after $k-1$ observations. Otherwise, take an observation from the p^r-experiment. This procedure will be called the *one-step procedure.* Another one, at stage k, takes an observation from the p-experiment if the Bayes risk of the Bayes estimator after $N-k$ observations of this experiment is less than after $N-k$ observations from the p^r-experiment, i.e., if

$$\sum_{i=0}^{N-k}\binom{N-k}{i}\Bigg[M_r(\alpha_k+ir+2,\beta_k,\gamma_k+N-k-i) - \frac{M_r^2(\alpha_k+ir+1,\beta_k,\gamma_k+N-k-i)}{M_r(\alpha_k+ir,\beta_k,\gamma_k+N-k-i)} - M_r(\alpha_k+i+2,\beta_k+N-k-i,\gamma_k) + \frac{M_r^2(\alpha_k+i+1,\beta_k+N-k-i,\gamma_k)}{M_r(\alpha_k+i,\beta_k+N-k-i,\gamma_k)}\Bigg]>0,$$

where α_k, β_k and γ_k are the parameters of the posterior distribution after $k-1$ observations. Otherwise, take an observation from the p^r-experiment. This procedure will be called the $(N-k)$*-step procedure.* Computer simulations lead to the following conclusions for these three procedures.

- the one-step procedure and the a_r-cut procedure give similar gains on average; on the other hand, the $(N-k)$-step procedure gives the lowest gain;
- the gain of these procedures depends on the value of the parameters α and β of the prior distribution and in general it is greater for small values of α and β;
- the gain is greater for large N.

References

C. L. Chen and W. H. Swallow (1990), *Using group testing to estimate a proportion, and to test the binomial model*, Biometrics 46, 1035–1046.

R. Dorfman (1943), *The detection of defective members of large populations*, Ann. Math. Statist. 14, 436–440.

R. G. Easterling and R. R. Prairie (1971), *Combining component and system information*, Technometrics 13, 271–280.

J. Gastwirth and P. Hammick (1989), *Estimation of the prevalence of a rare disease, preserving the anonymity of the subject by group testing: application to estimating the prevalence of AIDS antibodies in blood donors*, J. Statist. Plan. Inference 22, 15–27.

J. Hardwick, C. Page and Q. F. Stout (1998), *Sequentially deciding between two experiments for estimating a common success probability*, J. Amer. Statist. Assoc. 96, 1502–1511.

V. A. Lancaster and S. Keller-McNulty (1996), *A review of composite sampling methods*, Technical Report 96-01, Kansas State Univ. Dept. of Statistics.

M. W. Loyer (1983), *Bad probability, good statistics, and group testing for binomial estimation*, Amer. Statist. 37, 57–59.

M. Sobel and R. M. Elashoff (1975), *Group testing with a new goal, estimation*, Biometrika 62, 181–193.

X. M. Tu, E. Litvak and M. Pagano (1995), *On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening*, ibid. 82, 287–297.

Institute of Mathematics
Wrocław University of Technology
50-370 Wrocław, Poland
E-mail: arokita@im.pwr.wroc.pl
magiera@im.pwr.wroc.pl

Received on 11.7.2001;
revised version on 18.2.2002 (1586)