Estimation of process parameters to determine the optimum diagnosis interval for control of defective items

Abhyuday Mandal

Department of Statistics
University of Georgia
Athens, GA 30602-1952

Joint research with Tirthankar Dasgupta

Introduction

Determination of the most economic sampling interval for control of defective items is a problem that is extremely relevant to manufacturing processes that produce a continuous stream of products at a high speed.

- Frequent inspection requires more cost, time and manpower.
- Reduced frequency of inspection may lead to the risk of rejection of a large number of items.

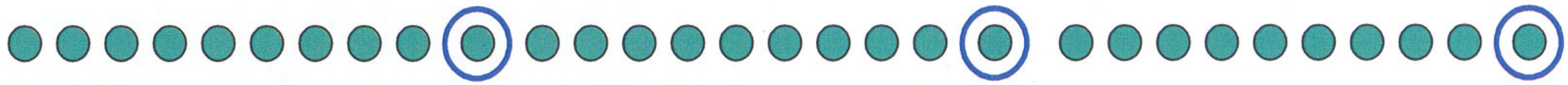
The problem of developing economically based online control methods for attributes has been addressed in detail by Taguchi (1981,1984,1985), Taguchi et al. (1989) and Nayebpour and Woodall (1993).

- The basic idea is to inspect a single item after every m units of production.
- A process adjustment is made as soon as a defective item is found.
- The value of m is to be determined to minimize the expected cost per item.

Different Cases

➢Industries producing a continuous stream of products (usually at high speed).

➢Every m^{th} product is inspected (here m = 10).

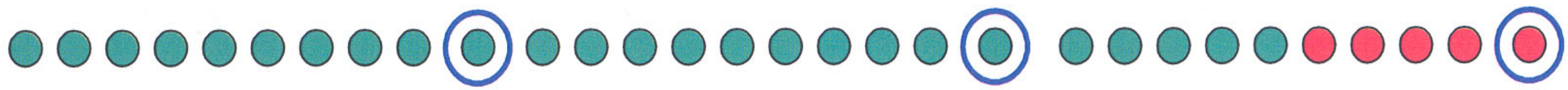


➢"In-control" process – no defectives produced.

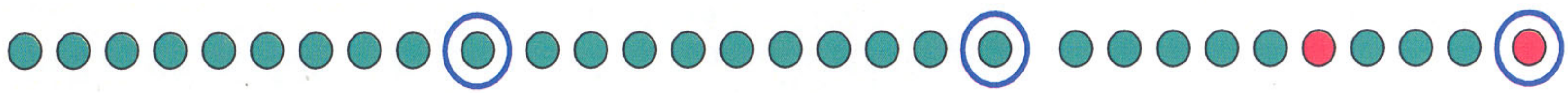
➢Process shifts to "out-of-control" state due to assignable causes and produces defectives.

Different Cases

➢Case-1 : Process shifts from 0% to 100% defectives; continues until rectification.



➢Case-2 : Process shifts from 0% to 100*π % defectives; may take long to detect.



➢Case-3 : A combination of Case-1 and Case-2.

Optimal Sampling Interval

- Consider a geometric process failure mechanism (PFM) with a parameter p. (The number of items produced before a process shift occurs, is $Geo(p)$)
- The expected loss per item ($E(L)$) is a function of p in Case I and (p,π) in Case II.
- The task of obtaining the optimal sampling interval thus consists of the following two stages :
  (i) Estimate the parameters associated with the PFM from historical data.
  (ii) Plug in these estimates into the expression for $E(L)$ and minimize it with respect to m.

The solution to the optimization problem is therefore strongly dependent on the estimate of the process parameters p and π.

Chicken First or Egg First

- In Case II, the estimator of p involves π, and incorrect estimation of π may lead to an erroneous estimate of p.
- Nayebpour and Woodall (1993) suggest obtaining an estimate of π using historical data on retrospective inspections.
- However, many companies may consider performing retrospective inspection uneconomic based on their perception about the value of π. Indeed, Nayebpour and Woodall (1993) recommend that retrospective inspection should be performed if

$$C_I \leq \pi C_D.$$

We need retrospective inspection data to estimate π and, on the other hand, an estimate of π to decide whether to perform retrospective inspection or not.

Benefits of the Proposed Methods

- If one can devise a reasonable estimation method from the data on cycle lengths, it would result in the following benefits:
  - It would prevent economic penalties resulting in incorrect estimation of the optimum inspection interval m.
  - It would assist the managers to take a better decision regarding whether to implement retrospective inspection or not.

Estimation of p and π in Case II

- Two methods have been proposed
  - Estimation using EM algorithm

    This is easy to implement, but can not be generalized to Case III.
  - The Bayesian approach

    It is very likely that in most of the cases process engineers will have some vague idea about π, which may not be good enough to check the condition $C_I \leq \pi C_D$, but may provide the analyst with a reasonable prior distribution for π.

The Statistical Model for Case II

For the i^{th} production cycle, $i = 1, 2, \ldots$, let

(i) U_i denote the number of products manufactured till the appearance of the first defect.

(ii) $X_i = [U_i/m] + 1$ denote the number of inspections from the beginning of the cycle to the first one immediately after appearance of the first defect.

(iii) Y_i denote the number of additional inspections necessary to detect the assignable cause after X_i.

(iv) l denote the number of units produced from the time a defective item is sampled until the time the production process is stopped for adjustment.

(v) $S_i = X_i + Y_i + [l/m]$ denote the number of products inspected in the cycle.

(vi) $T_i = m(X_i + Y_i) + l$ denote the total length of a cycle, or in other words, the number of products manufactured in a cycle.

(vii) C_i denote the total cost incurred in the cycle.

The Statistical Model for Case II

- We assume that U_i and Y_i, $(i = 1, 2, \ldots)$ are geometric random variables with parameters p and π so that

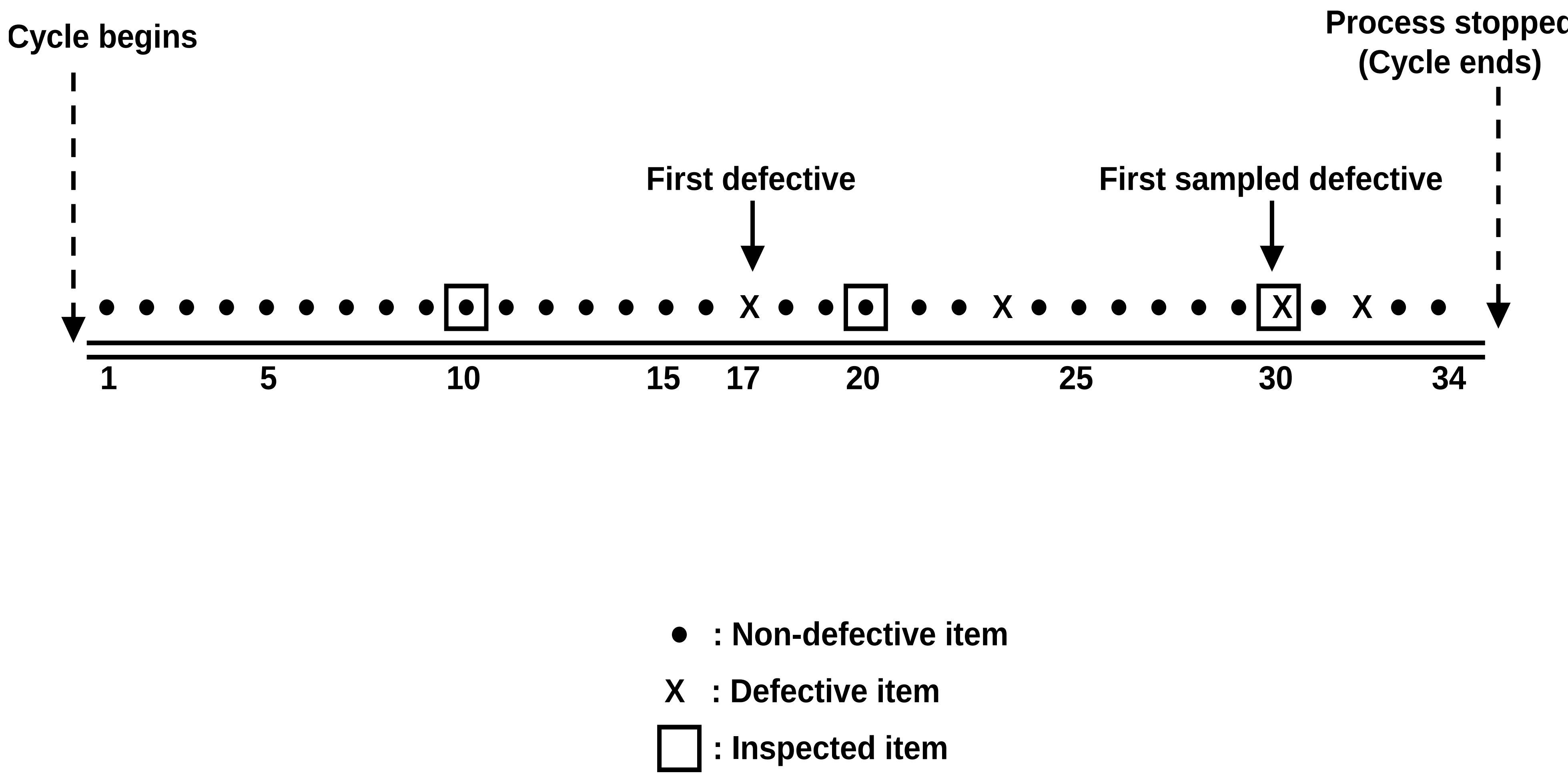
$$
\begin{aligned}
P(U_i = u) &= pq^{u-1}, \qquad u = 1, 2, \ldots \qquad \text{where} \quad q = 1 - p, \\
P(Y_i = y) &= \pi(1 - \pi)^y, \qquad y = 0, 1, 2, \ldots
\end{aligned}
$$

- It readily follows that the pmf of X_i is given by

$$
\begin{aligned}
P(X_i = x) &= P\Big((x-1)m < U_i \le mx\Big) \\
&= q^{(x-1)m}(1 - q^m), \qquad\qquad x = 1, 2, \ldots
\end{aligned}
$$

An Illustrative Example

- Suppose the current sampling interval m is 10.
- The 17^{th} item is the first defective item, after which the process starts producing $100\pi\%$ defectives.
- The 20^{th} item is non-defective; hence, the second inspection is unable to detect the assignable cause.
- The defect appears in the 30^{th} item and is detected.
- However, the process can be stopped after 4 more items have been manufactured, i.e., only after the 34^{th} item. Thus, in this cycle, $U = 17, X = 2, Y = 1, l = 4, T = 34, S = X + Y = 3$.


Cycle begins
Process stopped
(Cycle ends)
First defective
First sampled defective
X
X
X
X
1
5
10
15
17
20
25
30
34
: Non-defective item
X : Defective item
: Inspected item

Some Comments

Note that Case I and Case III can also be explained with the above process model.

- In Case I, $\pi = 1$ and hence $Y = 0, S = X$.
- In Case III, we have two possibilities –
  - either the minor assignable cause (after which the process starts producing $100\pi\%$ defectives) or
  - the major assignable cause (after which the process starts producing 100% defectives) appears first.

  Thus, in this case $U = min(U_1, U_2)$ where $U_1 \sim Geometric(p_1)$ and $U_2 \sim Geometric(p_2)$. Consequently, $U \sim Geometric(p)$ where $p = p_1 + p_2 - p_1 p_2$.

Some Comments

- The sequence $(T_1, C_1), (T_2, C_2), \ldots$ represents a renewal reward process (Ross, 1996). Thus, by the renewal reward theorem, the long-term expected loss per product $E(L)$ converges to $\frac{E(C)}{E(T)}$, where $E(C_i) = E(C)$ and $E(T_i) = E(T)$ for $i \geq 1$.

- Under the geometric PFM with a given p, explicit expressions for $E(C)$ and $E(T)$ can be computed (Nayebpour and Woodall, 1993) and $E(L)$ can be expressed as a convex function of m for given p and π.

- The optimum sampling interval is to obtained as

$$m^* = \arg\min E\{L(m, \hat{p}, \hat{\pi})\}$$

Estimation of p and π in Case II

- Suppose, we observe N production cycles and have the data on on the number of products inspected in each cycle $s_1, s_2, \ldots, s_N$.

- The objective is to estimate π and p from this dataset.

  PROPOSITION 1: The log-likelihood function of p and π is given by

$$\begin{aligned} & \log L(p, \pi; s_1, s_2, \ldots, s_N) \\ = \quad & N \log \pi + N \log(1 - q^m) - N \log|1 - \pi - q^m| + \sum_{i=1}^{N} \log|(1-\pi)^{r_i} - q^{m r_i}|, \end{aligned}$$

$$\text{where } r = s - [l/m].$$

- Clearly, the log-likelihood does not yield a straightforward expression for the MLE.

- Thus, one has to use numerical methods to solve the optimization problem. However, owing to the complex nature of the nonlinear function, its direct optimization is not very easy.

Estimation of p and π in Case II

- Note that, in this problem, the observed data $s_1, s_2, \ldots, s_N$ are realizations of the random variable $S = X + Y$.
- If it was possible to observe X and Y separately, it would have been possible to estimate p and π without much difficulty.
- Thus, in a sense, the data we observe here is incomplete. EM algorithm (Dempster et al. 1977) is a popular way of parameter estimation for such kind of problems and it is possible to simplify the optimization considerably using the EM algorithm.

Estimation using EM algorithm

- Suppose, instead of $s_1, s_2, \ldots, s_N$, we had observed the complete data $(x_1, y_1), \ldots, (x_N, y_N)$. Then, after observing the complete data, the log-likelihood function would have been

$$\log L(\mathbf{x}, \mathbf{y}; \theta) = N\left(\log \pi + \log(1-q^m) - m \log q\right) + m \log q \sum_{i=1}^{N} x_i + \log(1-\pi) \sum_{i=1}^{N} y_i$$

- PROPOSITION 2 : Let $G_X(\theta)$ and $G_Y(\theta)$ denote respectively the conditional expectations of X and Y, given $S = s$. Then,

  (i) $G_X(\theta) = \frac{1}{1-\psi} - \frac{r\psi^r}{\psi(1-\psi^r)}$, where $\psi = \frac{q^m}{1-\pi}$ and $r = s - [l/m]$.

  (ii) $G_Y(\theta) = \frac{\phi(1-\phi^{r-1})}{(1-\phi)(1-\phi^r)} - \frac{(r-1)\phi^r}{1-\phi^r}$, where $\phi = \frac{1}{\psi}$ and $r = s - [l/m]$.

Estimation using EM algorithm

(I) (E-step) Compute $Q(\theta,\theta^{(k-1)}) = E_{\theta^{(k-1)}}\left(\log L(\mathbf{x},\mathbf{y}|\theta,\mathbf{s})\right)$

$$= N\left(\log\pi + \log(1-q^m) - m\log q\right) + m\log q \sum_{i=1}^{N} G_{X_i}\left(\theta^{(k-1)}\right) + \log(1-\pi)\sum_{i=1}^{N} G_{Y_i}\left(\theta^{(k-1)}\right)$$

(II) (M-step) Find $\theta^{(k)}$ such that $\theta^{(k)} = \arg\max_\theta Q\left(\theta,\theta^{(k-1)}\right)$.

Equating $\frac{\partial Q(\theta,\theta^{(k-1)})}{\partial\pi}$ to zero, we obtain

$$\pi^{(k)} = \frac{N}{N+\sum_{i=1}^{N} G_{Y_i}\left(\theta^{(k-1)}\right)}$$

Also,

$$p^{(k)} = 1-\left(1-\frac{m_c}{\overline{T}-\left(l+\frac{m_c(1-\pi^{(k)})}{\pi^{(k)}}\right)}\right)^{1/m_c}$$

The Bayesian approach: Prior for π

- Suppose, based on their past experience and/or pilot study, the process engineers are able to specify a reasonable range for π as $[\pi_L, \pi_U]$.
- Recalling that π must satisfy $\pi > \pi_{bound}$, we would assign a negligibly small mass of the prior distribution below π_{bound}.
- Therefore, we can elicit a $Beta(\alpha_\pi, \beta_\pi)$ prior for π where the hyperparameters can be obtained by solving

$$\frac{\Gamma(\alpha_\pi+\beta_\pi)}{\Gamma(\alpha_\pi)\Gamma(\beta_\pi)}\int_0^{\pi_{bound}} \pi^{\alpha_\pi-1}(1-\pi)^{\beta_\pi-1}d\pi = \varepsilon$$
$$\frac{\Gamma(\alpha_\pi+\beta_\pi)}{\Gamma(\alpha_\pi)\Gamma(\beta_\pi)}\int_{\max(\pi_{bound},\pi_L)}^{\pi_U} \pi^{\alpha_\pi-1}(1-\pi)^{\beta_\pi-1}d\pi = 1-\gamma_\pi$$

- It is clear that the first implies that there is a negligibly small probability ε that π will be less than π_{bound}, and the second equation ensures that the probability of π lying beyond the stated interval equals a pre-assigned value γ_π. ($1-\gamma_\pi$ can be interpreted as the *degree of belief*.)

The Bayesian approach: Prior for p

- Even when there is no available prior information on p, we can elicit a prior distribution for p based on the knowledge of π.

- Let p_L and p_U be the lower and upper limits of p obtained by substituting π_U and $\max(\pi_{bound}, \pi_L)$ respectively in the MLE estimate of p.

- The hyperparameters α_p, β_p of a suitable Beta prior distribution for p may be obtained by solving

$$\frac{\Gamma(\alpha_p+\beta_p)}{\Gamma(\alpha_p)\Gamma(\beta_p)} \int_0^{(p_L+p_U)/2} p^{\alpha_p-1}(1-p)^{\beta_p-1} dp = \frac{1}{2}$$

$$\frac{\Gamma(\alpha_p+\beta_p)}{\Gamma(\alpha_p)\Gamma(\beta_p)} \int_{p_L}^{p_U} p^{\alpha_p-1}(1-p)^{\beta_p-1} dp = 1-\gamma_p$$

- The first equation implies that the median of the distribution is taken at the mid point of the interval $[p_L, p_U]$. Interpretation of the second equation is clearly the same as before.

A Simulated Example: Comparison of the Methods

Consider a process where we have, $p = 0.000339$, as in the Case I example of Nayebpour and Woodall (1993). Let $m_c = 500$, $\pi = 0.10$ and $l = 0$. We simulate 200 production cycles from the above process, thereby generating data of the form $s_1, s_2, \ldots, s_{200}$.

Regarding the available prior information, we consider the following two situations:

- The process engineer, from his experience, states "When the process goes out of control, it produces at most 15% defectives on an average. I have no idea about p".
- The process engineer states,"As per my experience, the appropriate range for π is $12 \pm 5\%$. p usually doesn't exceed 0.0005."

A Simulated Example: Comparison of the Methods

Estimates of p and π are obtained using the EM algorithm and the Bayesian method. Altogether we have the following seven cases.

1. EM algorithm based estimation.
2. Bayesian estimation using uniform prior and MCMC.
3. Bayesian estimation using uniform prior with posterior mode as the estimate.
4. Bayesian estimation using a tight Beta prior ($\gamma_\pi = \gamma_p = 0.05$) and MCMC.
5. Bayesian estimation using a tight Beta prior ($\gamma_\pi = \gamma_p = 0.05$) with posterior mode as the estimate.
6. Bayesian estimation using a flatter Beta prior ($\gamma_\pi = \gamma_p = 0.25$) and MCMC.
7. Bayesian estimation using a flatter Beta prior ($\gamma_\pi = \gamma_p = 0.25$) with posterior mode as the estimate.

A Simulated Example: Comparison of the Methods

- The prior distributions for cases $2-7$ corresponding to the two situations are shown in Table 1.
- Note that the hyperparameters for the beta priors are derived taking $\varepsilon = 0.001$.

Table 1: Prior distributions for p and π for the Bayesian methods

Methods	Situation 1		Situation 2	
	Prior for p	Prior for π	Prior for p	Prior for π
2 and 3	$U[0.0002, 0.0009]$	$U[0.07, 0.15]$	$U[0.0002, 0.0005]$	$U[0.07, 0.17]$
4 and 5	$Beta(9, 16355)$	$Beta(25, 196)$	$Beta(20, 56500)$	$Beta(20, 144)$
6 and 7	$Beta(3, 5000)$	$Beta(19, 125)$	$Beta(6.75, 18550)$	$Beta(13, 76)$

- The simulation is repeated 100 times and the results are summarized next.

Summary of Simulation Output in Case II

(True values: $p = 0.000339, \pi = .10$)

Available info	Method				Estimate of p		Estimate of π	
					$mean(\hat{p})$	$sd(\hat{p})$	$mean(\hat{\pi})$	$sd(\hat{\pi})$
$0 \leq \pi \leq 0.15$ No info on p	EM algorithm				0.000356	0.000109	0.1043	0.0170
	Bayesian	uniform		MCMC	0.000367	0.000097	0.1008	0.0119
				MAP	0.000592	0.000029	0.0839	0.0049
		Beta	Tight	MCMC	0.000374	0.000057	0.0981	0.0073
				MAP	0.000589	0.000045	0.0819	0.0055
			Flat	MCMC	0.000306	0.000067	0.1127	0.0097
				MAP	0.000591	0.000049	0.0826	0.0051
$0.07 \leq \pi \leq 0.17$ $0 < p \leq 0.0005$	EM algorithm				0.000297	0.000097	0.1228	0.0307
	Bayesian	uniform		MCMC	0.000322	0.000050	0.1076	0.0127
				MAP	0.000235	0.000034	0.1372	0.0214
		Beta	Tight	MCMC	0.000269	0.000033	0.1228	0.0085
				MAP	0.000248	0.000007	0.1219	0.0052
			Flat	MCMC	0.000313	0.000022	0.1076	0.0063
				MAP	0.000231	0.000027	0.1388	0.0182

Discussions on Simulation Output in Case II

1. The MAP estimates are seen to be poor in both situations, irrespective of the prior distributions. There are multiple modes in the posterior distribution. Convergence of the optimization algorithm is seen to depend on the initial choices of the parameters.

2. The EM algorithm based estimates are good in situation 1, but the algorithm converges to a local maxima in situation 2. As explained by Wu (1983), if the log likelihood has several (local or global) maxima and stationary points, convergence of the EM algorithm depends on the choice of starting point.

3. The Bayes' estimates obtained using MCMC perform better and are more more robust to the varying levels of available preliminary information on the parameters. The performance is not very sensitive to the choice of prior except in the case of the tight Beta prior in situation 2. This is possibly a consequence of placing almost the entire mass of the prior in the stated range with mean at the center.

4. The variance of $\hat{p}$ corresponding to almost each method is generally seen to be less in situation 2 as compared to situation 1, which shows that, as expected, with better and more accurate prior information, one can obtain more efficient estimates.

Estimation of parameters in Case III

- Case III, discussed by Nandi and Sreehari (1997), deals with a scenario where there are two types of assignable causes, termed minor and major, and their appearances follow geometric patterns with parameters p_1 and p_2.
- Occurrence of a major assignable cause leads to a situation like Case I, where all subsequent items produced are defective.
- A minor assignable cause leads to a situation like Case II, i.e., the process starts producing 100π% defective products following the occurrence of such a cause.
- Although Nandi and Sreehari (1997) derived expressions for the expected loss, they completely ignored the estimation of p_1, p_2 and π.

The Bayesian estimation of p_1, p_2 and π

Note that, for this case we can use exactly the same notations as in Case II if we define $p = p_1 + p_2 - p_1 p_2$, i.e., $q = q_1 q_2$, where $q_i = 1 - p_i$ for $i = 1, 2$.

To develop the Bayesian algorithm we need the following result:

PROPOSITION 3 : The probability mass function of S is given by

$$\begin{aligned} P(S=s) &= (1-q^m)\Big[\alpha(1-\pi)(\pi q_2^m + 1 - q_2^m)\Delta\left(\frac{\{(1-\pi)q_2^m\}^{r-1} - q^{m(r-1)}}{(1-\pi)q_2^m - q^m}\right) \\ &+ q^{m(r-1)}\big(1-\Delta\alpha(1-\pi)\big)\Big], \qquad s = 1, 2, \ldots \end{aligned}$$

where

$$r = s - [l/m], \quad p = p_1 + p_2 - p_1 p_2, \quad \text{and} \quad \Delta = q_2 . \frac{1-q_1}{1-q_1^m} . \frac{q_2^m - q_1^m}{q_2 - q_1}.$$

We assume that $p_i \sim Beta(\alpha_i, \beta_i), i = 1, 2$ and $\pi \sim Beta(\alpha_3, \beta_3)$. Then Proposition 3 leads to the following corollary.

The Bayesian estimation of p_1, p_2 and π

$$
\begin{aligned}
\log g(p_1, p_2, \pi | s_1, \ldots, s_N) &\propto N \log(1 - q^m) \\
+ \quad & \sum_{i=1}^{N} \log \left[\alpha(1-\pi)(\pi q_2^m + 1 - q_2^m)\Delta \frac{\{(1-\pi)q_2^m\}^{r_i - 1} - q^{m(r_i - 1)}}{(1-\pi)q_2^m - q^m} + q^{m(r_i-1)}\{1 - \Delta\alpha(1-\pi)\} \right] \\
+ \quad & \sum_{j=1}^{2} (\alpha_j - 1)\log p_j + (\alpha_3 - 1)\log \pi + \sum_{j=1}^{2} (\beta_j - 1)\log q_j + (\beta_3 - 1)\log(1-\pi)
\end{aligned}
$$

- Assuming that we have some lower and upper bounds for each of the three parameters, the hyperparameters α_i, β_i, $i = 1, 2, 3$ can be obtained in the same way as discussed before.
- If the engineers are more or less certain about the limits and are unable to say anything more about the prior distributions, uniform priors could be a possible choice again.
- Considering the complication involved in finding the posterior modes, we only use MCMC methods to simulate the posterior density of each parameter.

Simulation results

We consider the same numerical example as the one used by Nandi and Sreehari (1997) where $\pi = 0.10$, $p_1 = 0.001$ and $p_2 = 0.002$. As in Case II, we generate 200 cycles in each simulation of the process.

The following two levels of prior information are considered:

1. "Strong" (Reasonably accurate information) :

$$0.001 \leq p_1 \leq 0.003, 0 < p_2 \leq 0.002, 0.07 \leq \pi \leq 0.13$$

2. "Weak" (Moderately accurate information) :

$$0 < p_1 \leq 0.005, 0 < p_2 \leq 0.003, 0 < \pi \leq 0.25$$

In order to study the sensitivity of the method with respect to the choice of the prior distributions, we consider the following three priors:

(i) Beta priors tightly distributed in the stated intervals with $\gamma_{p_1} = \gamma_{p_2} = \gamma_\pi = 0.05$.

(ii) Flatter Beta priors with $\gamma_{p_1} = \gamma_{p_2} = \gamma_\pi = 0.25$.

(iii) Uniform priors in the stated intervals.

Case III: Summary of Simulation Results

100 simulations were carried out for each of the 6 cases. Each estimate is the median of its simulated posterior distribution (10,000 MCMC iterations with burn-in of 1000).

Level of information	Prior Distribution		π	p_1	p_2
STRONG	High belief Beta $(\alpha_1 = 22, \beta_1 = 9980$	Mean	0.10071	0.00209	0.00085
	$\alpha_2 = 4, \beta_2 = 4000,\ \alpha_3 = 40, \beta_3 = 450)$	sd	0.00430	0.00023	0.00014
$0.001 \leq p_1 \leq 0.003$	Low belief Beta $(\alpha_1 = 5, \beta_1 = 2500$	Mean	0.09643	0.00209	0.00091
$0 \leq p_2 \leq 0.002,$	$\alpha_2 = .55, \beta_2 = 400,\ \alpha_3 = 12, \beta_3 = 110)$	sd	0.01186	0.00035	0.00024
$0.07 \leq \pi \leq 0.13.$	Uniform	Mean	0.09783	0.00205	0.00092
		sd	0.00751	0.00033	0.00022
WEAK	High belief Beta $(\alpha_1 = 3, \beta_1 = 1375$	Mean	0.09843	0.00217	0.00093
	$\alpha_2 = 2, \beta_2 = 1500,\ \alpha_3 = 3, \beta_3 = 21)$	sd	0.02809	0.00050	0.00020
$0 \leq p_1 \leq 0.005$	Low belief Beta $(\alpha_1 = 1, \beta_1 = 300$	Mean	0.09992	0.00242	0.00086
$0 \leq p_2 \leq 0.003,$	$\alpha_2 = .75, \beta_2 = 375,\ \alpha_3 = 1, \beta_3 = 5)$	sd	0.04610	0.00072	0.00033
$0 \leq \pi \leq 0.25.$	Uniform	Mean	0.09397	0.00246	0.00094
		sd	0.04086	0.00081	0.00026

Comments on Simulation Results

1. The method works satisfactorily, even when the prior information is not quite accurate.

2. However, as expected, the accuracy of prior information increases the efficiency of the parameter estimates. This is supported by the fact that with “strong” prior information the variances of the estimators are much less that with “weak” information.

3. The method does not seem to be very much sensitive to the choice of the nature of prior distribution, as also seen in Case II. If the prior information on π is of the form $\pi_0 \pm \delta$, then it might be easy to elicit a Beta prior with mean close to π_0. On the other hand, if it is simply of the form of an interval π_L, π_U, it would be more pragmatic to consider a uniform prior.

Summary and Conclusions

- Noting that the estimation problem is trivial for Case I, we highlight the problems associated with the estimation of the process parameters in Case II (p and π) and Case III (p_1, p_2 and π).
- We propose two different estimation procedures to resolve the aforementioned problems. One is based on the Bayesian approach and the other based on the EM algorithm.
- We propose some concrete guidelines for eliciting a prior distribution from the available information.
- One interesting area of future research in this area is to develop a generic framework with k types of assignable causes that would have Case I, Case II and Case III as special cases. This is encountered in several industrial situations.

Thank you