1 Introduction

Slow-moving stock-keeping units (SKU) are common in spare parts inventory systems and are usually held to avoid the undesirable consequences of the unavailability of items when requested. These consequences could mean any damage or risk associated with not being able to operate the equipment requested. Due to the extremely low consumption rate, slow-moving spares are exceedingly inflexible with respect to overstocking. As such, overestimating the demand for these items can result in extra storage costs and complete losses when these items become obsolete. In addition, with the number of SKUs usually being very high, simple solutions, like keeping extra safety stocks for all items of a certain type, are not satisfactory. Thus, an accurate demand forecasting must be performed to ensure effective inventory investments.

However, the available records of demand are often very limited, some of which only have null values over observation intervals. For this reason, the inventory management of such items is performed on a group basis, when multiple stock-keeping units (SKUs) are used to develop the “population-averaged” lead-time demand probability distribution. This distribution is then used for each item when deciding on the replenishment order size and the re-order level. Air force inventory systems are examples that rely on this type of approach, particularly because of the large percentage of their slow-moving units kept stock as a result of extremely low demand. While accepted in practice, the described straightforward method has a clear disadvantage: due to the averaging of demand patterns, it underestimates the demand for items with larger consumption and overestimates the demand for those with lower consumption.

This chapter deals with an empirical Bayesian method to estimate the demand distribution of multiple slow-moving items in case of extremely low demand and short history of requests. An extension of the beta-binomial probability distribution is given for flexible empirical Bayesian forecasting of SKUs with zero demand records while relying on the demand for similar items.

To accurately determine the practical significance of the problem, consider the following real-life example of the UK Royal Air Force (RAF) inventory system [Eaves & Kingsman, 2004], one of the largest and most diverse inventories in the world. At the beginning of 2000, the RAF kept about 684,000 consumable line items resulting in approximately 145 million SKUs and a total stock value of 1.2 billion pounds. From the overall number of line items, about 8.5% accounted for 90% of the annual demand value (fast-moving items). Moreover, 37.3% of the SKUs had fewer than 10 demand transactions, of which 40.5% had zero
demand over the observation period of six years. In total, about 60% of the SKUs could be classified as slow-moving with extremely low consumption, for which the demand per period (one month) is binary, and the maximum lead-time demand is finite and small. It is clear that an accurate demand forecasting for items of the considered type is very important for an inventory system like that of the RAF.

This chapter is based on our previous publications [Dolgui et al., 2004; Dolgui & Pashkevich, 2008a; Dolgui & Pashkevich, 2008b; Dolgui & Pashkevich, 2008c]. The rest of this chapter is organized as follows: Section 2 reviews related literature; Section 3 deals with the problem statement; Section 4 describes the usage of the standard beta-binomial model to solve the problem; an extension of this model that incorporates prior information for the maximum probability of demand per period is presented in Section 5; parameter estimation and the corresponding Bayesian forecasting issues for the proposed model are considered in Sections 6 and 7, respectively; finally, we present our conclusion in Section 8.

2 Related Publications

Accurate demand forecasting is a central issue in successful inventory management and is critical in achieving realistic estimates of the overall service level, or a total fill rate in the case of multiple SKUs [Silver et al., 1998]. Intermittent demand analysis is a challenge for inventory control due to the specific nature of the underlying demand process, which makes the forecasting problem especially difficult [Willemain et al., 2004; Fildes & Beard, 1992]. Slow-moving demand is a significant type of intermittent demand in which a request is made for a single unit in most of the cases [Syntetos et al., 2005]. As an example, this is a common situation in service parts inventory systems [Cochran & Lewis, 2002].

For slow-moving demand, the major problems are (i) the lack of order records to estimate the past consumption reliably and (ii) zero consumption over a long period of time [Mitchell et al., 1983]. Forecasting techniques developed for smooth and continuous demand are not applicable here because their assumptions of continuity and normal demand distribution are not appropriate. A number of techniques that relax these assumptions have been developed, and of these, the most widely used is Croston's method [Croston, 1972]. Willemain et al. [1994] evaluated Croston’s method under the relaxation of the assumptions on which it is based using simulated data; the authors then compared it with single exponential smoothing using industrial data. Modifications of this approach with various improvements have likewise been proposed [Johnston & Boylan, 1996; Syntelos et al., 2005]. Another way of handling the demand for slow-moving SKUs is based on the Poisson probability distribution [Dunsmuir & Snyder, 1989; Schultz, 1987; Ward, 1978; Watson, 1987]. Recently, the bootstrap technique has become popular when modeling intermittent demand [Willemain et al., 2004].

Nevertheless, in the cases when some of the inventory items under consideration have no records of past requests, the problem of demand forecasting becomes even more complex. This is because common techniques like Croston’s method and its extensions cannot be applied to estimate the probability distribution of demand.

Another challenge is the limited demand history available to model lead-time consumption, a difficulty that has been overcome through the Bayesian paradigm. First proposed by Scarf [1959] and Silver [1965], it was later successfully applied to various inventory management problems [Hill, 1997, 1999; Aronis et al., 2004]. Still, the empirical Bayesian technique, one which uses the heterogeneity among multiple inventory items to overcome the problem of short historical records, has not received enough attention in scientific literature [Bradford & Sugrue, 1990], even though it is potentially important for practical applications.

Most of the work in the Bayesian inventory modeling has been done by assuming a Poisson demand distribution with a gamma prior for the rate parameter. However, the beta-binomial distribution is a better
choice in the case of extremely low consumption, as the upper bound of the lead-time demand is finite and small. This mixture probability distribution was previously applied to a number of inventory management problems [Petrovic et al., 1989]. We explain the beta-binomial model (BBM) in this chapter. An extension of the beta-binomial probability distribution is also presented with the objective of considering additional information regarding extremely low demand probability for multiple slow-moving SKUs.

In [Dolgui & Pashkevich, 2008a], we compared the performance of the binomial (BM) and BBM models of demand forecasting for multiple slow-moving inventory items using the most common demand patterns corresponding to the uniform case, a bell-like distribution in which all values are close to zero or one, two sub-groups of items with high probability of demand in one group and low in the second one, and so on. In that paper, we concluded that the BBM model significantly decreased the holding costs required to achieve a desired service level when compared with the BM model for all tested patterns (8–66% of gain depending on the pattern). In addition, we found that the greatest gain was obtained for the U-shaped probability distribution. As mentioned earlier, we present the BBM model in this chapter as well as its extension called Extended Beta-Binomial Model (EBBM). We use the same empirical Bayesian approach and the BBM, as in our previous work, but this time, we performed a more precise estimation of the demand using an additional parameter [Dolgui & Pashkevich, 2008b].

3 Problem Description

Let us introduce the following notations:

- \( k \): Number of inventory items in the managed group
- \( n \): Number of periods with available historical demand data
- \( m \): Lead-time length
- \( B = (b_{ij}) \): Binary \( k \times n \)-matrix with past demand records
- \( b_{ij} \): Equal to one if demand has occurred for item \( i \) in period \( j \); zero otherwise
- \( s_i \): Total number of requests for inventory item \( i \) in the past \((b_{i1}, b_{i2}, \ldots, b_{in})\)
- \( p_i \): Unknown probability of a request per period for inventory item \( i \)
- \( p_{BM} \): Population-averaged probability of a request per period
- \( \pi \): Extended beta-binomial model parameter; expert’s estimate for maximum of \( \{p_i\} \)
- \( L_i \): Unknown re-order level for inventory item \( i \)
- \( c_i \): Unit holding cost for inventory item \( i \)
- \( D_i \): Random lead-time demand for inventory item \( i \)
- \( D_A \): Random population-averaged lead-time demand
- \( F \): Target weighted probability of no shortage during the lead-time
- \( B(\alpha, \beta) \): Beta-function with parameters \( \alpha \) and \( \beta \)

Assume that the inventory control is performed over a group of similar SKUs, with \( k \) being the group size. Let \( n \) denote the number of periods for which the demand \( B \) data are available, where \( B = (b_{ij}) \) is a binary \( k \times n \)-matrix and \( b_{ij} = 1 \) if the demand has occurred for item \( i \) in the period \( j \), and \( b_{ij} = 0 \) if otherwise. The demand is assumed to be binary due to the specific properties of the SKUs under consideration, that is, it is supposed that there is always a possibility to select the period length so that no more than one request occurs in each period.

The availability of an item is crucial for the system to work effectively as a stock-out leads to undesirable consequences. For example, the corresponding equipment might not operate or the unsatisfied demands could result in a backlog with a backlogging cost [Chauhan et al., 2009]. The corresponding expenses
Chapte 1 – Beta-Binomial Model and Its Generalizations in the Demand Forecasting for Multiple Slow-Moving

(backlogging cost) are difficult to estimate and are not comparable with the holding cost. Thus, here a service level criterion is employed instead of the backlogging cost, considering all items are slow-moving and with some of them having no record of demand over the entire observation period. Presumably, the replenishment ordering costs are insignificant with respect to the inventory storage costs. Therefore, the size of the replenishment order is always one.

For a group of slow-moving items, a straightforward approach estimates the average request probability as follows:

\[ p_{BM} = \sum_{i=1}^{k} \sum_{j=1}^{n} \frac{b_{ij}}{k \cdot n}, \quad (1) \]

and assigns the binomial distribution \( Bi(m, p_{BM}) \) to all SKUs in the group, where \( m \) is the lead time. Afterwards, the following problem is solved for one “aggregated item” presented as:

\[ \text{Minimize: } L, \text{ subject to: } P\{D \leq L\} \geq F. \quad (2) \]

The expression (2) presents a usual service-level problem for a single inventory item, and \( L \) is the reorder level used for all SKUs in the group. The described method will be called the BM of the lead-time demand hereafter. This approach can solve the data availability problem of the industrial data sets that are “short and wide” [Willemain et al., 2004]. The problem of having few observations \( (n) \) for each SKU is overcome by utilizing a large number of items in a group \( (k) \).

Nevertheless, the approach signified by expressions (1) and (2) does not consider the heterogeneity of item demands. This underestimates the re-order level for the items with larger consumption and overestimates the re-order level for the ones with lower consumption. Therefore, in this chapter, a more sophisticated inventory optimization approach is utilized [Grange, 1998], the aim of which is to satisfy the target weighted probability of no shortage during the lead-time over the whole group of \( k \) items. The ultimate goal is to find the unknown re-order levels \( \{L_i\} \) from the following constrained optimization problem:

\[ \text{Minimize: } L_1, \ldots, L_k \sum_{i=1}^{k} c_i L_i, \text{ subject to: } \frac{\sum_{i=1}^{k} P\{D_i \leq L_i\} \cdot E\{D_i\}}{\sum_{i=1}^{k} E\{D_i\}} \geq F, \quad (4) \]

where \( D_i \) denotes the random demand for the item \( i \) over the lead-time, \( c_i \) is the unit holding cost for this item, and \( F \) is the target weighted probability of no shortage during the lead-time. The problem (4) is \( NP \)-hard, with \( k \) decision variables in the general case.

To solve the optimization problem (4), one must know the probability distributions of the random product demands \( D_i, i = 1, 2, \ldots, k \). These distributions must be assessed using the binary \( k \times n \)-matrix \( B \) which contains the historical demand records, where \( k \) is large (hundreds or thousands), and \( n \) is small (5–10, for example).

In the approach (1) – (2), there is a common lead time \( m \) for all items; in the method presented further in this chapter, we also rely on a single lead time \( m \) for all SKUs [Dolgvi & Pashkevich, 2008b].

An accurate estimation of the probability distributions for \( \{D_i\} \) is performed through the Bayesian forecasting of these distributions while considering the history of individual requests for each item. The SKUs are now forecasted based not only on the demand for all items in the group but also on individual requests history (empirical Bayesian approach).

Problem (4) is solved with respect to each item type based on their past requests. Each class contains items with the same demand level (number of requested items) in the past and is assigned its own lead-time demand distribution, as opposed to the binomial model with one aggregated item. The number of decision
variables is reduced to a maximum of \( n + 1 \) which makes the exhaustive search for (4) applicable (since \( n \) is small).

A separate note should be made in assessing the quality of forecasting. As the records of demand are limited and may contain many zeroes, the performance of the forecasting approach must be evaluated not for each item, but over a group of items under review. The shortage of any item has a key consequence, and it is impossible to compare the outcomes of shortages for different items \textit{a priori}. It is also difficult to state which items are more important than others. In addition, the simulations show that global system performances are consistent with practical considerations. Indeed, for this type of inventory system, the average performance is more important than accuracy for a particular item [Hopp & Spearman, 1995].

Let us now present an overview of the forecasting concept employed in a BBM.

4 BBM for Slow-Moving Items with Low Demand and Short History

This section provides a framework for the solution of the problem being considered by relying on ideas culled from longitudinal statistical data analysis [Diggle et al., 2002], that have been successfully used to solve a number of problems with similar data structure in other application areas [Pashkevich & Kharin, 2004]. We would like to point out that the beta-binomial distribution, which serves as a foundation in our approach, is often justified in inventory management in a slightly different manner. It is usually supposed that a demand distribution for a single inventory item is of this kind because the “request probabilities” of the consumers have a beta distribution, and the number of consumers is limited and small (otherwise, a negative binomial distribution is used). Here, we use the beta-binomial distribution because there is heterogeneity of demand within the group of SKUs, similar to the approach proposed by Bradford and Sugrue [1990] for the mixed Poisson distribution.

The BBM [Collet, 2002], as applied to the lead-time demand forecasting problem under consideration, is formulated using three assumptions: \( A_1 \), \( A_2 \), and \( A_3 \).

The first assumption (\( A_1 \)) is that the probability \( p_i \) of an item \( i \) being ordered during the overall review interval is invariant with respect to period number \( j \):

\[
(A_1) \ P\{b_{ij} = 1\} = p_i, \forall j = 1, 2, \ldots, n.
\]

This assumption of the stationary demand over the review period is realistic from a practical point of view, as in our case in which the consumption rate is low, and the number of periods with available data \( n \) is small. The latter can be interpreted as the overall review period being negligibly small when compared with the demand dynamics.

The second assumption (\( A_2 \)) introduces both the relation and the heterogeneity among the SKUs by supposing that the demand probabilities \( \{p_i\} \) for all \( i \) are drawn from the same beta distribution [Johnson et al., 1995] with the parameters \( \alpha \) and \( \beta \):

\[
(A_2) \ L\{p_i\} = B(\alpha, \beta), i = 1, 2, \ldots, k.
\]

This statistical distribution will be called the prior distribution hereafter. This assumption allows creating a population-averaged demand model of SKUs, which will later be used for the Bayesian forecasting of the lead-time demand probability distribution for each particular item. The beta distribution is very flexible and can represent different heterogeneities in the demand pattern and take a variety of shapes depending on the values of the parameters \( \alpha \) and \( \beta \) [Collet, 2002]. The probability density function is expressed as follows:

\[
f_{p_i} = (y|\alpha, \beta) = y^{\alpha-1}(1-y)^{\beta-1}/B(\alpha, \beta).
\]
For example, if $\alpha$ and $\beta$ are both less than or equal to 1, the distribution will be U- or J-shaped. These shapes represent polarized distributions, in which some items have small request probabilities and others have large, although few SKUs can be found in between. On the other hand, if $\alpha$ and $\beta$ are both large, the distribution will resemble a spike so that all items have more or less the same request probabilities. If the values of $\alpha$ and $\beta$ are just a little larger than 1, then the beta distribution looks like an inverted U or like the central part of the normal curve. If $\alpha$ and $\beta$ are both equal to one, the distribution becomes uniform. If one of the parameters is larger than 1 and the second is smaller, the probability density function is L-shaped.

Being a conjugate prior for the binomial distribution, the beta distribution is capable of flexibly describing the heterogeneity of demand probabilities for the items within the group [Collet, 2002]. It is also capable of providing computationally simple forecasting expressions.

The third assumption $(A_3)$ is that the random demand probabilities $\{p_i\}$ are independently drawn from the beta distribution, and are thus independent in total:

$(A_3)$ Probabilities $p_1, p_2, \ldots, p_k$ are i.i.d. random variables.

Under the assumptions $A_1 - A_3$ and with the lead time equal to $m$ periods, the population-averaged lead-time demand probability distribution for every item of the group is the beta-binomial with the parameters $m, \alpha,$ and $\beta$ [Collet, 2002]:

$$P\{D_A = x\} = \binom{m}{x} \cdot \frac{B(\alpha + x, \beta + m - x)}{B(\alpha, \beta)}, \quad x = 0, 1, \ldots, m,$$

where $D_A$ denotes the population-averaged lead-time demand.

We can then divide all SKUs into $(S + 1)$ classes (or sub-groups) based on the previous requests $\{s_i\}$, $S = \max(s_1, s_2, \ldots, s_k)$. Afterwards, we can deal with $S + 1$ aggregated items. Each class contains items with the same number of demanded items in the past. Under the proposed assumptions, the maximum possible value of $S$ is equal to $n$. For each aggregated item, we will apply the above-formulated BBM with the empirical Bayesian approach.

The major advantage of the Bayesian approach in forecasting is that it leads to the adjustment of the population-averaged probability distribution (6) using observations specific to a particular item. As the beta distribution is conjugate prior to the binomial model, the posterior distribution of the demand probability $p_i$ for the item $i$ is also beta but with shifted parameters:

$$L(p_i \mid s_i) = B(\alpha + s_i, \beta + n - s_i), \quad s_i = \sum_{j=1}^{n} b_{ij},$$

where $s_i$ is the total demand over the observed period for the $i$-th SKU. Hence, the forecast of the lead-time demand $D_i$ for the item $i$ also follows the beta-binomial distribution but with the parameters modified according to (7):

$$P\{D_i = x \mid s_i\} = \binom{m}{x} \cdot \frac{B(\hat{\alpha} + s_i + x, \hat{\beta} + m + n - s_i - x)}{B(\hat{\alpha} + s_i, \hat{\beta} + n - s_i)}, \quad x = 0, 1, \ldots, m,$$

where the hats denote the estimates of the corresponding parameters (see Section 6).

Expression (8) can be used to calculate the re-order levels for the SKUs in the corresponding sub-group. Estimation of the model parameters is performed using explicit expressions based on the method of moments (MM) or the iterative numerical estimators based on the maximum likelihood (ML) (Tripathi et al., 1994).

The procedure of modeling the demand distribution for the group of slow-moving items with low demand can be outlined as follows:
- Obtain the historical demand data for the group of items as a \((k \times n)\) binary matrix \(B\).
- Estimate the parameters \(\alpha\) and \(\beta\) using the matrix \(B\).
- Divide SKUs in \((S+1)\) sub-groups, with each sub-group containing items with the same number of requests in the past.
- Compute the Bayesian posterior lead-time demand probability distribution \((8)\) for each sub-group \(i\) based on the demand \(s_i\) observed for the corresponding items.

Finally, problem \((4)\), which balances the weighted probability of no shortage during the lead-time and the holding costs, takes the following form:

\[
\text{Minimize: } L_1, ..., L_s \sum_{i=1}^{k} r_x^i \cdot \bar{L}_x, \quad \text{subject to: } \frac{\sum_{x=0}^{s} r_x^i P\{\bar{D}_x \leq \bar{L}_x\} \cdot E\{\bar{D}_x\}}{\sum_{x=0}^{s} r_x^i E\{\bar{D}_x\}} \geq F, \tag{9}
\]

where \(r_x^i = \sum_{i=1}^{k} c_i \cdot I(s_i = x)\), and \(I(.)\) is the unit function that takes the value of one if the argument condition holds, and zero if otherwise. The variables \(\bar{L}_x\) and \(\bar{D}_x\) denote the re-order level and random lead-time demand for the sub-group \(x\), respectively. As aforementioned, the sub-group \(x\) is defined as a subset of SKUs with exactly \(x\) requests for the last \(n\) periods. As in our case the past demand record is very short, the problem \((9)\) can be efficiently solved using a simple Branch-and-Bound algorithm with the following condition for cutting unpromising branches for an incomplete solution \(\{\bar{L}_0, \bar{L}_1, ..., \bar{L}_q\}, q < S\):

\[
\frac{\sum_{x=0}^{q} r_x^i \cdot P\{\bar{D}_x \leq \bar{L}_x\} \cdot E\{\bar{D}_x\} + \sum_{x=q+1}^{s} r_x^i \cdot E\{\bar{D}_x\}}{\sum_{x=0}^{s} r_x^i \cdot E\{\bar{D}_x\}} < F. \tag{10}
\]

### 5 Generalizations in the Case of Extremely Low Demand

Although the BBM explained in the previous section provides a reasonable framework for handling demand for slow-moving SKUs with short and zero demand histories, it does not consider the fact that the demand is very low for all items. In this section, we present an extension of the beta-binomial model that can consider this information. First, a corresponding probability distribution is presented. Afterwards, parameter estimation issues are considered. Finally, we present an empirical Bayesian forecasting procedure for the new model, which ensures mean square optimal prediction.

The following extension to the assumption \(A_2\) is proposed. Suppose that there exists an expert estimate \(\pi \in (0, 1)\) of the maximum demand probability per period for the considered group of SKUs, the prior distribution of the demand probabilities then becomes the following special case of the generalized beta distribution:

\[
f_{\mu_i}(y|\alpha, \beta, \pi) = \frac{y^{\alpha-1}(\pi - y)^{\beta-1}}{B(\alpha, \beta) \cdot \pi^{\alpha+\beta-1}}. \tag{11}
\]

The parameter \(\pi\) is assumed to be small due to the specific properties of the considered demands. In the following, we present a population-averaged lead-time demand probability distribution under prior \((11)\), and then show the expressions for its mathematical expectation and variance.

The population-averaged lead-time demand probability distribution for the EBBM with prior distribution \((11)\) can be expressed as a weighted sum of the shifted beta-binomial probabilities presented as:

\[
P\{D_A = x\} = \sum_{l=x}^{m} w_{xl}(\pi, m) \cdot P_0(l|m, \alpha, \beta), \quad x = 0, 1, ..., m, \tag{12}
\]
where the weights \( w_{d,l} \) are computed as:

\[
w_{d,l}(\pi,m) = \begin{cases} \binom{l}{d} \cdot \pi^x (1 - \pi)^{l-x}, & 0 \leq x \leq l, \\ 0, & l < x \leq m, \end{cases}
\] (13)

and \( P_0(l \mid m, \alpha, \beta) \) denotes the beta-binomial probability with the corresponding parameters:

\[
P_0(l \mid m, \alpha, \beta) = \binom{m}{l} \cdot \frac{B(\alpha + l, \beta + m - l)}{B(\alpha, \beta)}. \tag{14}
\]

The mathematical expectation and variance of the developed probability distribution (12) are calculated as follows:

\[
E\{D_A\} = \pi \cdot \frac{m\alpha}{\alpha + \beta}, \quad V\{D_A\} = \pi^2 \cdot \frac{m\alpha(\alpha + \beta + m)}{(\alpha + \beta)^2(\alpha + \beta + 1)} + \pi(1 - \pi) \cdot \frac{m\alpha}{\alpha + \beta}. \tag{15}
\]

Proofs of the results presented above and in subsequent sections of this chapter can be found in our previous work [Dolgui & Pashkevich, 2008b].

## 6 Parameter Estimation

To determine the parameters of the proposed EBBM, two traditional approaches are employed. First, the explicit estimators based on the MM are presented. The corresponding estimates can then be introduced as the initial approximation for maximum likelihood estimators, which are based on the numerical optimization routines. The parameter \( \pi \) is assumed to be known. The raw data for the parameter estimation are the binary \( s_1, s_2, \ldots, s_k \) presented in expression (7).

The MM estimators for the parameters \( \alpha \) and \( \beta \) of the probability distribution (12) with the known parameter \( \pi \) can be expressed as:

\[
\alpha = \frac{\lambda \cdot \bar{s}'}{\pi}, \quad \beta = \frac{\lambda \cdot (\pi - \bar{s}')}{\pi}, \quad \lambda = (n - 1) \cdot \frac{\bar{s}' (\pi - \bar{s}')}{\nu' - \bar{s}' (1 - \bar{s}')} - 1, \quad \bar{s}' = \frac{s'}{n}, \quad \nu' = \frac{\nu}{n}, \tag{16}
\]

where \( \bar{s}' \) and \( \nu' \) are the mean and variance of the sample \( \{s_1, s_2, \ldots, s_k\} \).

To obtain the maximum likelihood estimates, the following optimization problem must be solved:

\[
\max_{\alpha, \beta} : L(\alpha, \beta) = \prod_{i=1}^{k} \left( \sum_{l=s_i}^{n} w_{s_i,l}(\pi) \cdot \frac{n}{l} \cdot \frac{B(\alpha + l, \beta + n - l)}{B(\alpha, \beta)} \right). \tag{17}
\]

This can be accomplished using a modification of the steepest descent method.

Following a standard path, we can obtain the initial approximation via the MM, and then apply a numerical optimization algorithm to compute the maximum likelihood estimates of the parameters \( \alpha \) and \( \beta \).

If there is no available expert estimate for the parameter \( \pi \), it can be estimated from the available data on past requests in the following way. One approach is to implement a grid search over a reasonable range of values for \( \pi \) by estimating the parameters \( \alpha \) and \( \beta \) for each value and then selecting the estimates that lead to the best fit based on \( \chi^2 \)-statistics. Another approach, which is more precise but prone to overfitting, is to jointly estimate the parameters \( \alpha, \beta, \) and \( \pi \) by maximizing the likelihood function (17) with respect to all three parameters. To surmount overfitting in the latter method, the final estimate of \( \pi \) can be selected as the value which is greater than the obtained maximum likelihood estimate and is reasonable from a practical point of view.
7 Bayesian Forecasting for EBBM

Once the model parameters are estimated, an empirical Bayesian forecasting approach [Winkler, 2003] can be applied to obtain the individual probability distribution of the lead-time demand for each inventory item. In the following, we present the posterior distribution of the demand probability under the assumption of prior (11). The proofs are provided in our previous work [Dolgui & Pashkevich, 2008b].

For prior (11), the posterior distribution of the demand probability can be represented as a weighted sum of the shifted generalized beta distributions:

\[ h_{p_i}^g(y|s_i) = \sum_{l=0}^{n} \omega_{s_i l}(\pi, n) \cdot \frac{y^{\alpha + l - 1} (1 - y)^{\beta + n - l - 1}}{B(\alpha + l, \beta + n - l)}. \]  

(18)

The corresponding mean-square-error optimal point forecast is computed as follows:

\[ \hat{p}_{g_i}^s(s_i) = \sum_{l=0}^{n} \omega_{s_i l}(\pi, n) \cdot \frac{\pi (\alpha + l)}{\alpha + \beta + n}. \]  

(19)

The weights have the following form:

\[ \omega_{rl}(\pi, n) = \left( \sum_{j=0}^{n} w_{rl}(\pi, n) \cdot P_0(l|m, \alpha + r, \beta + n - r) \right) \cdot w_{rl}(\pi, n) \cdot P_0(l|m, \alpha, \beta), \]  

(20)

where \( w_{rl} \) is defined by (13).

Once the distribution of the demand probability \( p_i \) is known, the next step is to develop the posterior lead-time demand probability distribution, which adjusts the population-averaged consumption for a particular item with respect to its own demand history. We now show expressions for the corresponding probability row and the mean and variance of the random lead-time demand. The proofs are presented in our previous work [Dolgui & Pashkevich, 2008b].

For the EBBM based on prior (11), the probability distribution of the lead-time demand \( D_i \) for the inventory item \( i \) is expressed as a weighted sum of the lead-time demand probability distributions for the EBBM (12) with shifted parameters, where the weights are given by expression (20):

\[ P\{D_i = x|s_i\} = \sum_{r=s_i}^{n} \omega_{s_i r}(\pi, n) \cdot \sum_{l=x}^{m} w_{xl}(\pi, m) \cdot P_0(l|m, \alpha + r, \beta + n - r), x = 0, 1, ..., m. \]  

(21)

The mean demand and its variance are computed as follows:

\[ E\{D_i|s_i\} = m \pi \cdot \sum_{r=s_i}^{n} \omega_{s_i r}(\pi, n) \cdot \frac{\alpha + r}{\alpha + \beta + n}, \]  

(22)

\[ V\{D_i|s_i\} = \sum_{r=s_i}^{n} \omega_{s_i r} \left( \pi^2 \cdot \frac{m^{[2-]} \alpha^{[2+]}(\alpha + \beta)^{[2+]}}{\alpha + \beta} \cdot m \pi \cdot \frac{m \alpha}{\alpha + \beta} \right) - E^2\{D_i|s_i\}, m^{[2-]} = m \cdot (m - 1). \]  

(23)

These results can be used to estimate the re-order points for the group of slow-moving SKUs with extremely low demand and with the past records characterized by a large percentage of zeros and short observation period. Its major advantage over the classic BBM is the additional parameter \( \pi \), which allows for a more precise specification of prior demand probability distributions.
8 Conclusions

This chapter considered the problem of modeling lead-time demand for multiple slow-moving inventory items when the available demand history is very short, and a large percentage of items have only zero records.

An empirical Bayesian forecasting approach was chosen to predict the lead-time demand probability distribution. As the demand can be considered binary, and the past history records are short, the BBM can be successfully employed. This model is very flexible and can represent heterogeneities of item demands.

Furthermore, an extension of the BBM which considered the prior information on the maximum expected probability of demand per period was presented. Parameter estimation and Bayesian forecasting routines were proposed for this model as well.

To be applicable for inventory systems, our forecasting techniques should be integrated in an appropriate inventory control model, with an example provided for reference. An optimization algorithm based on Branch and Bound approach can optimally solve this model.

Our experience in the use of these models has led to the following conclusions:

- For the inventory control problem considered in this paper, the standard binomial model tends to overestimate the inventory needed to achieve the target weighted probability of no shortage during the lead-time and is not reliable with respect to the actual probability of no shortage.

- The beta-binomial model demonstrates stable and robust performance, ensures lower holding costs when compared with the BM, and guarantees the actual probability of no shortage during the lead-time being consistent with the selected target.

- The EBBM ensures considerably lower holding costs than the BBM and also satisfies the target weighted probability of no shortage during the lead-time, while the additional computational complexity over the BBM is not significant.

Usually, items in the same group are supposed to have a “related” demand, which means that they possess similar properties. For example, these items can be spare parts from the same assembly unit or items that are historically known to have correlated demands. Thus, this aspect reinforces the “borrow-strength” properties of the empirical Bayesian method used in our approach, consequently improving its performance.

The presented approaches are valid not only for the problems mentioned in the introduction, such as military inventory systems, but also for a wide class of spare part inventory systems as those in automobile and aircraft industries, among others.

Although the assumption of the same lead-time for all items is often realistic for the type of inventory system considered, the results presented are not restricted by this assumption and can be directly extended to the case of different lead-times. The assumption that no more than one request can occur in each period poses some restrictions on the applicability of our methods. However, in most of the considered cases, this assumption was acceptable, thereby illustrating the importance and complexity of the problems involved.

References


Chapter 1 – Beta-Binomial Model and Its Generalizations in the Demand Forecasting for Multiple Slow-Moving


