

Estimates on Returnable Packaging Material

J. Bierkens (Radboud University Nijmegen)

H. Blok (Leiden University)

M. Heydenreich (Leiden University and CWI Amsterdam)

R. Núñez Queija (University of Amsterdam and CWI Amsterdam)

P. van Meurs (Eindhoven University of Technology)*

F. Spieksma (Leiden University) J. Tuitman (KU Leuven)

Abstract

When a beer company replaces its returnable packaging materials, for example when updating the design of a bottle, it needs to know in advance how much new material will be needed. Dutch beer brewer Heineken submitted the question of estimating the returnable packaging materials to the 2013 Studygroup Mathematics with Industry. In this report, we present both stochastic flow models and a queueing model to estimate the amount of returnable packaging material present in the market. Furthermore, we give recommendations on what data to collect, and how to sample this data in an unbiased way in order to increase accuracy of the estimation.

KEYWORDS: Modelling, Markov Chain, Stochastic Differential Equation, Queueing Theory

1 Introduction

Beer companies, like Heineken, use *returnable packaging materials* (i.e., bottles, cases and kegs) multiple times. To simplify our terminology, we will throughout refer to returnable packaging materials as *bottles*, keeping in mind that all results apply to other types of materials as well. In some markets, for example in the Netherlands, customers pay a *deposit* on bottles, which is returned to them when the bottles are returned. In other markets, for example in many African countries, a *full-for-empty* system is used instead. In this system, customers return empty bottles only when purchasing full ones. Unlike in the deposit system, in the full-for-empty system any purchase of new bottles is limited by the number of empty bottles available to the customers. Therefore, customers tend to keep a much larger stock of empty bottles in a full-for-empty system than in a deposit system.

*Corresponding author

Occasionally, the returnable packaging material is changed, for example because of a new bottle design, and then the beer company needs to know how many new bottles will be needed. This is an intrinsically difficult question, because bottles might be broken or stored away and reappear only many years after they are sold. At the moment, the companies miss an efficient model for estimating the number of bottles in the market, especially in the case of a full-for-empty system. In some packaging change operations, significantly more new bottles were needed than expected. Heineken requested the 2013 Study Group Mathematics with Industry to develop a model for estimating the number of bottles in the market more accurately, and asked for recommendations on what data to collect for use in such a model.

The structure of this report is as follows. First, in Section 2 we introduce terminology and notation, and explain what data is currently available. Moreover, we present an easy way of estimating the so-called break rate, which will be an important parameter in what follows. Then we develop two different models for the number of bottles in the market: in Section 3, Markov Chains and stochastic differential equations are used, while in Section 4 a queueing model is discussed. The difference of the two approaches is discussed in Section 4. Next, in Section 5, we elaborate on how to use the sample data to get reliable parameter estimates. Finally, we summarize our findings in Section 6.

2 Problem description

2.1 Modelling and notation

In a very simplified market model, bottles are sold at the distribution center, and arrive at the market. After remaining there for some time, the empty bottles are returned to the beer company. If the number of returned bottles is not sufficient to satisfy the beer demand, new bottles have to be produced (see Figure 1).

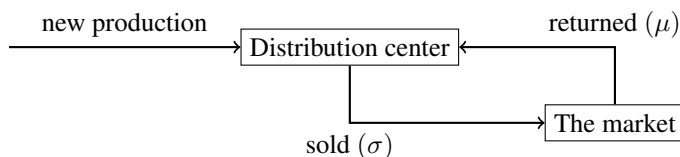


Figure 1: The flow of bottles

We aim at estimating the number of bottles currently in the market. In particular, we are interested in the number of bottles that are expected to be returned. To this end, we differentiate between different categories of bottles:

- **The returning bottles** $R(t)$

The number of bottles at time t that will be returned even in the absence of a

packaging change. Typically, these bottles are in the market for a relatively short period of time before they return to the distribution center.

- **The sleeping bottles** $S(t)$

The number of bottles at time t that will only be returned after a packaging change. These materials are temporarily stored away or used for other purposes and will turn up after a change of bottle.

- **The broken bottles** $B(t)$

The number of bottles at time t that will never return to the distribution center. They could be broken, lost or stored away permanently.

The sum $M(t) = R(t) + S(t) + B(t)$ of these three numbers represents the total number of bottles present in the market at time t . The company is especially interested in estimating $S(t)$ more accurately.

2.2 Available Data

2.2.1 Volumes

The company has kept track of the volumes of the sold as well as the returned bottles for more than 20 years for different factories. These data are collected per product on a monthly basis. We introduce the following notation:

- **Sold items** $\sigma(t)$

$\sigma(t)$ = the number of bottles sold in month t .

- **Returned items** $\mu(t)$

$\mu(t)$ = the number of bottles returned in month t .

To obtain the number of bottles accumulated in the market since the beginning of measurements, we take the sum of the number of sold items and subtract the number of returned items, that is,

$$M(t) - M(0) = \sum_{s=0}^t (\sigma(s) - \mu(s)).$$

2.2.2 Circulation Times

More recently, the company has started collecting samples of *circulation times* of bottles. The circulation time of a bottle is the time that elapses between leaving and returning to the distribution center. The data on the circulation times can be used to estimate $R(t)$, as explained in Section 3 and Section 4. As this data only represents returned bottles, it cannot be used to differentiate between sleeping and broken bottles. However, we can extract estimates for $M(t)$ and $R(t)$, and hence also for the sum $B(t) + S(t) = M(t) - R(t)$.

Currently the data is collected in a rather biased way, which results in unreliable estimates for the circulation time. In Section 5 we elaborate on how to improve the sampling.

2.3 Approximation of $R(t) + S(t)$

We are left with the question of how to estimate the total number of returning materials $R(t) + S(t)$. We proceed by assuming that $S(t)$ stabilizes for sufficiently large times t . This is a reasonable assumption since the storage capacity for bottles in the market is limited and hence the number of sleeping bottles cannot grow indefinitely. Moreover, we also assume that the fraction of sold bottles that ends up broken is constant. We call this constant β and refer to it as the *break rate*. Thus $\beta M(t) \approx B(t)$ for large times t . By our assumptions, the break rate can be estimated by the number of bottles that are not returned in a long time period $[t_0, t_1]$ divided by the number of bottles sold in that time period. That is,

$$\hat{\beta} = \frac{\sum_{s=t_0}^{t_1} (\sigma(s) - \mu(s))}{\sum_{s=t_0}^{t_1} \sigma(s)} = 1 - \frac{\sum_{s=t_0}^{t_1} \mu(s)}{\sum_{s=t_0}^{t_1} \sigma(s)},$$

for $t_0, t_1, t_1 - t_0$ large enough.

Now we can write

$$R(t) + S(t) = M(t) - B(t) = M(t) - \hat{\beta}M(t) = \frac{\sum_{s=t_0}^{t_1} \mu(s)}{\sum_{s=t_0}^{t_1} \sigma(s)} M(t).$$

Remark: The first assumption on the stabilization of the number of sleeping bottles in the market might not be very realistic in fast growing markets.

3 Analysis based on Markov models

In this section we propose a simple Markovian model to model the ‘dynamics’ of a single bottle while it is with a customer. It should be considered an example. More realistic models can be constructed by introducing more states and/or more complex dynamics (e.g., dynamics that are not homogeneous in time). The models in this section are especially suited for the deposit system, and less so for the ‘full-for-empty’ system, because we do not model the fact that a customer returns bottles at the same time that he buys new ones. See also the discussion in Section 3.5.

3.1 Markov model for single unit

First let us consider an individual bottle at a customer. Suppose it behaves according to the following simple Markov dynamics. The bottle can be in either of four states, FULL, EMPTY, BROKEN and RETURNED. The states BROKEN and RETURNED are absorbing. Transitions (per unit of time) occur as in Figure 2.

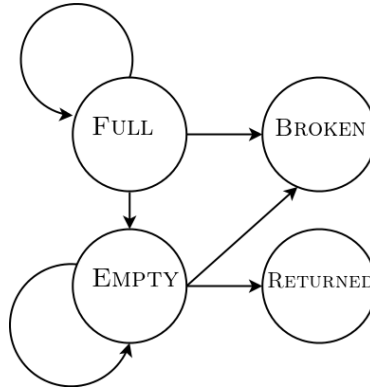


Figure 2: A Markov model. Transition rates between the states F(ull), E(mpty), B(roken), and R(eturned) are denoted by λ_{FE} , λ_{FB} , λ_{ER} , λ_{EB} .

3.2 Markov model for quantities of bottles

From the Markov model for a single bottle we can derive a model for the flow of bottles. Suppose, at time t there are a total number of F_t non-empty bottles and E_t empty bottles in the market. Let $h > 0$ be a small time step and suppose a number of $U_{t+h} - U_t$ bottles are sold during the time interval $[t, t + h)$. Denote the total number of returned bottles within $[t, t + h)$ by $Y_{t+h} - Y_t$. The variable names U for input and Y for output are chosen to correspond to the usual notation in systems theory. In systems theory, there also is the notion of state, denoted by X , corresponding in our case with the two-dimensional vector (E, F) .

Remark: Note that the number of bottles sold per unit of time is $(U_{t+h} - U_t)/h$, which for small h is equivalent to the time derivative of U_t . The same holds for Y_t . This ‘cumulative’ notation for U and Y is helpful in the continuous time limit, where trajectories of U and Y will typically be non-differentiable.

If we denote by $(F \rightarrow B)_t$ and $(E \rightarrow B)_t$ the number of full and empty bottles broken within $[t, t + h)$, respectively, $(F \rightarrow E)_t$ the number of emptied bottles and $(E \rightarrow R)_t$ the number of empty bottles returned within this interval, we obtain the following balance equations:

$$\begin{aligned}
 F_{t+h} &= F_t + U_{t+h} - U_t - (F \rightarrow B)_t - (F \rightarrow E)_t, \\
 E_{t+h} &= E_t + (F \rightarrow E)_t - (E \rightarrow B)_t - (E \rightarrow R)_t, \\
 Y_{t+h} &= Y_t + (E \rightarrow R)_t.
 \end{aligned}$$

Now for all the transitions in the Markov model, assuming independent ‘behaviour’ of individual bottles, and using that h is small, we see that e.g. $(F \rightarrow B)_t \sim \text{Bin}(F_t, \lambda_{FB}h)$. Using the normal approximation for the binomial distribution (assuming F_t and E_t are large), we find that $(F \rightarrow B)_t$ is approximately normally distributed

with mean $h\lambda_{FB}F_t$ and variance $F_t(h\lambda_{FB})(1-h\lambda_{FB})\doteq h\lambda_{FB}F_t$. Therefore, we have approximately the following *discrete time Markov model*

$$\begin{aligned} F_{t+h} &= (1 - h(\lambda_{FB} + \lambda_{FE}))F_t + U_{t+h} - U_t - \sqrt{h\lambda_{FB}F_t}\varepsilon_t^{FB} - \sqrt{h\lambda_{FE}F_t}\varepsilon_t^{FE}, \\ E_{t+h} &= E_t + h(\lambda_{FE}F_t - (\lambda_{EB} + \lambda_{ER})E_t) + \sqrt{h\lambda_{FE}F_t}\varepsilon_t^{FE} \\ &\quad - \sqrt{h\lambda_{ER}E_t}\varepsilon_t^{ER} - \sqrt{h\lambda_{EB}E_t}\varepsilon_t^{EB}, \\ Y_{t+h} &= Y_t + h\lambda_{ER}E_t + \sqrt{h\lambda_{ER}E_t}\varepsilon_t^{ER}. \end{aligned}$$

where all the ε_t^{\dots} are normally distributed with mean 0 and variance 1. Assuming the fluctuations in F and E to be relatively small, the following model is more straightforward to analyse.

$$\begin{aligned} F_{t+h} &= (1 - h(\lambda_{FB} + \lambda_{FE}))F_t + (U_{t+h} - U_t) - \sqrt{h}\sigma_{FB}\varepsilon_t^{FB} - \sqrt{h}\sigma_{FE}\varepsilon_t^{FE}, \\ E_{t+h} &= E_t + h(\lambda_{FE}F_t - (\lambda_{EB} + \lambda_{ER})E_t) + \sqrt{h}\sigma_{FE}\varepsilon_t^{FE} \\ &\quad - \sqrt{h}\sigma_{ER}\varepsilon_t^{ER} - \sqrt{h}\sigma_{EB}\varepsilon_t^{EB}, \\ Y_{t+h} &= Y_t + h\lambda_{ER}E_t + \sqrt{h\lambda_{ER}E_t}\varepsilon_t^{ER}. \end{aligned} \tag{1}$$

3.2.1 Diffusion limit

By taking the $h \downarrow 0$ limit, we may write the Markov model (1) as the following system of stochastic differential equations (SDEs),

$$\begin{aligned} dF_t &= -(\lambda_{FB} + \lambda_{FE})F_t dt + dU_t - \sqrt{\lambda_{FB}F_t} dW_t^{FB} - \sqrt{\lambda_{FE}F_t} dW_t^{FE}, \\ dE_t &= [\lambda_{FE}F_t - (\lambda_{EB} + \lambda_{ER})E_t] dt + \sqrt{\lambda_{FE}F_t} dW_t^{FE} \\ &\quad - \sqrt{\lambda_{ER}E_t} dW_t^{ER} - \sqrt{\lambda_{EB}E_t} dW_t^{EB}, \\ dY_t &= \lambda_{ER}E_t dt + \sqrt{\lambda_{ER}E_t} dW_t^{ER}, \end{aligned}$$

where W_t^{\dots} are independent Brownian motions.

This (non-linear) system is rather difficult to analyse, mainly because of the presence of square roots. As before, under the assumption that the fluctuations in F and E around their average values F_{average} and E_{average} are relatively small, we could work with the linearized system of stochastic differential equations in which the randomness is additive:

$$\begin{aligned} dF_t &= -(\lambda_{FB} + \lambda_{FE})F_t dt + dU_t - \sigma_{FB} dW_t^{FB} - \sigma_{FE} dW_t^{FE}, \\ dE_t &= [\lambda_{FE}F_t - (\lambda_{EB} + \lambda_{ER})E_t] dt + \sigma_{FE} dW_t^{FE} - \sigma_{EB} dW_t^{EB} - \sigma_{ER} dW_t^{ER}, \\ dY_t &= \lambda_{ER}E_t dt + \sigma_{ER} dW_t^{ER}. \end{aligned} \tag{2}$$

Here $\sigma_{FB} = \sqrt{\lambda_{FB}F_{\text{average}}}$, etc.

Assuming U is sufficiently smooth, we may write $dU = u(t) dt$. We may write (2) in abstract form as

$$\begin{aligned} dX(t) &= A_1 X(t) dt + Bu(t) dt + \Sigma_1 dW_t, \\ dY(t) &= A_2 X(t) dt + \Sigma_2 dW_t. \end{aligned} \quad (3)$$

where

$$\begin{aligned} X(t) &= \begin{bmatrix} F_t \\ E_t \end{bmatrix}, & Y(t) &= Y_t, \\ A_1 &= \begin{bmatrix} -(\lambda_{FB} + \lambda_{FE}) & 0 \\ \lambda_{FE} & -(\lambda_{EB} + \lambda_{ER}) \end{bmatrix}, & A_2 &= [0 \quad \lambda_{ER}], \\ B &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, & \Sigma_2 &= [0 \quad 0 \quad 0 \quad \sigma_{ER}], \quad \text{and} \\ \Sigma_1 &= \begin{bmatrix} -\sigma_{FB} & -\sigma_{FE} & 0 & 0 \\ \sigma_{FB} & 0 & -\sigma_{EB} & -\sigma_{ER} \end{bmatrix} \end{aligned}$$

and W is a four-dimensional standard Brownian motion.

3.2.2 ODE approximation / fluid limit

For the mean value behaviour of either of the models (1) or (2), we obtain the system of ordinary differential equations

$$\begin{cases} \dot{f}(t) &= -(\lambda_{FB} + \lambda_{FE})f(t) + u(t), \\ \dot{e}(t) &= \lambda_{FE}f(t) - (\lambda_{EB} + \lambda_{ER})e(t), \\ y(t) &= \lambda_{ER}e(t), \end{cases}$$

where $f(t) = \mathbb{E}F_t$, $e(t) = \mathbb{E}E_t$, but $y(t) = \frac{d}{dt}\mathbb{E}Y_t$. Let $\lambda_F := \lambda_{FB} + \lambda_{FE}$ and $\lambda_E := \lambda_{EB} + \lambda_{ER}$. The solution to this system of ordinary differential equations is given by

$$\begin{aligned} f(t) &= \exp(-\lambda_F t)f(0) + \int_0^t \exp(-\lambda_F(t-s))u(s) ds, \\ e(t) &= \exp(-\lambda_E t)e(0) + \int_0^t \exp(-\lambda_E(t-s))f(s) ds, \\ y(t) &= \lambda_{ER}e(t). \end{aligned}$$

This equation of f tells us the following. As $\lambda_F \in (0, 1)$, the first term on the right hand side decays exponentially fast to zero. This means that the dependence on the initial value of the number of full bottles in the system will not matter. The second term says that the value of $f(t)$ depends on the history of the average number of the inflow of bottles (i.e. $u(s)$ for $s \in [0, t]$), where the dependence is the strongest at the most recent history.

To interpret e , we need some more calculation. Just as in the interpretation of f , we can conclude from the first term in the right hand side of the expression for $e(t)$ that the dependence on the initial data decays exponentially. For the second term, we separate 3 cases:

$\lambda_F = \lambda_E$. In this case, the probability that a full bottle leaves the FULL state in a certain time period, equals the probability that an empty bottle leaves the EMPTY state in the same period. We get

$$\begin{aligned} e(t) - \exp(-\lambda_E t)e(0) &= \int_0^t \exp(-\lambda_E(t-s))f(s) ds \\ &= t \exp(-\lambda_E t)f(0) + \int_0^t (t-s) \exp(-\lambda_E(t-s))u(s) ds, \end{aligned}$$

from which we learn that also $e(t)$ depends on the history of $u(t)$, but with a certain delay. This expression also shows that the dependence of $f(0)$ on $e(t)$ vanishes exponentially fast in the long run, but at the start there is an increasing dependence.

$\lambda_F \neq \lambda_E$. In this case, we obtain

$$\begin{aligned} e(t) - \exp(-\lambda_E t)e(0) &= \int_0^t \exp(-\lambda_E(t-s))f(s) ds \\ &= \frac{\exp(-\lambda_F t) - \exp(-\lambda_E t)}{\lambda_E - \lambda_F} \exp(-\lambda_F t)f(0) \\ &\quad + \int_0^t \frac{\exp(-\lambda_F(t-s)) - \exp(-\lambda_E(t-s))}{\lambda_E - \lambda_F} \exp((\lambda_E - \lambda_F)s) u(s) ds, \end{aligned}$$

from which we see again that the influence of $f(0)$ decays exponentially. The second term denotes the cumulative and delayed dependence on the input stream $u(t)$.

The interpretation of the number of bottles that is expected to be returned per time unit, i.e. $y(t)$, is just a fixed fraction of $e(t)$, of which we discussed its behaviour above.

3.3 State estimation: Kalman filtering

Let us consider the stochastic model again (see (1) for the time-discrete model and (2) for the continuous in time model). Before stating how we can get information out of the data by using these stochastic models, we would like to discuss the basic theory behind Kalman filtering.

Suppose a random variable Y has a conditional distribution depending on ‘hidden state’ X and ‘input’ U ; loosely denoted as $p(Y|X, U)$. Furthermore suppose X and U are distributed according to some ‘prior’ distribution $p(X, U)$. Bayes’ formula gives

us that, given observations of U and Y , we may compute X as

$$\begin{aligned} p(X = x|Y = y, U = u) &= \frac{p(X = x, Y = y, U = u)}{p(Y = y, U = u)} \\ &= \frac{p(Y = y|X = x, U = u)p(X = x, U = u)}{\sum_x p(Y = y|X = x, U = u)p(X = x, U = u)}. \end{aligned}$$

In other words, based on the conditional distribution $p(Y|X, U)$ and the prior distribution $p(X, U)$, we may compute a ‘posterior’ distribution $p(X|Y, U)$. This posterior distribution enables us to estimate the hidden state X based on observations of U and Y .

The same idea may be applied recursively to systems of the form (3), leading to the *Kalman-Bucy filter* [4, 7], or, in discrete time, the *Kalman filter* [1, 9]. Such a filter allows us in this example to obtain estimates \hat{F}_t, \hat{E}_t of F_t and E_t , based on observations of U_t and Y_t . Kalman filters appear notationally involved, but once the dynamic model (such as (1) or (2)) is identified, implementation of such a filter is relatively straightforward. It gives us estimates for $\mathbb{E}E_t, \mathbb{E}F_t, \text{Var} E_t$ and $\text{Var} F_t$, which become more accurate for larger t . It is therefore preferable to use a data set with a long time series.

3.4 Estimation of the model parameters

To complete our model, we need to estimate the λ_{\dots} parameters. In this section we demonstrate an estimation method based on the data of the sampling.

3.4.1 Estimation through distribution of circulation times

Conditional on the eventual return of a bottle, we have to wait time $T_E \sim \exp(\lambda_{FE})$ before a bottle is being emptied, plus a time $T_R \sim \exp(\lambda_{ER})$ before the empty bottle is returned. The total waiting time $T = T_E + T_R$ is then the sum of two exponential random variables, and has a *hypoexponentially distribution* with parameters λ_{FE} and λ_{ER} .

For convenience, we write $\lambda_1 = \lambda_{FE}$ and $\lambda_2 = \lambda_{ER}$. Since the random variables T_E and T_R are independent, the mean of T is $\mathbb{E}T = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$ and the variance is $\text{Var}(T) = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}$. The distribution function of T may be computed (in case $\lambda_1 \neq \lambda_2$)

as

$$\begin{aligned}
 F_T(t) &= \mathbb{P}(T_E + T_R \leq t) = \int_0^t \mathbb{P}(T_E + T_R \leq t | T_e = s) f_{T_e}(s) ds \\
 &= \int_0^t \mathbb{P}(T_r \leq t - s) \lambda_1 \exp(-\lambda_1 s) ds \\
 &= \lambda_1 \int_0^t (1 - \exp(-\lambda_2(t - s))) (\exp(-\lambda_1 s)) ds \\
 &= 1 + \frac{1}{\lambda_1 - \lambda_2} (\lambda_2 \exp(-\lambda_1 t) - \lambda_1 \exp(-\lambda_2 t)), \quad t \geq 0.
 \end{aligned}$$

The density function is then

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (\exp(-\lambda_2 t) - \exp(-\lambda_1 t)), \quad t \geq 0.$$

In case $\lambda_1 = \lambda_2$, a similar computation gives $F_T(t) = 1 - \exp(-\lambda_1 t) - \lambda_1 t \exp(-\lambda_1 t)$ and $f_T(t) = \lambda_1^2 t \exp(-\lambda_1 t)$ for $t \geq 0$.

Given n i.i.d. observations t_1, \dots, t_n of a hypoexponentially distributed random variable, we can estimate the parameters λ_1 and λ_2 in two ways:

- (i) By maximizing the (log)-likelihood function

$$l(\lambda_1, \lambda_2) = \sum_{i=1}^n \ln \left(\frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (\exp(-\lambda_2 t_i) - \exp(-\lambda_1 t_i)) \right), \quad \lambda_1, \lambda_2 > 0$$

with respect to λ_1 and λ_2 . This will always provide estimates for λ_1 and λ_2 , but needs to be performed numerically.

- (ii) By the method of moments: choose λ_1 and λ_2 so that the sample mean and variance match the computed expectation and variance. Let $\hat{\sigma}^2$ denote sample variance and $\hat{\mu}$ denote sample mean. Write $a_1 = 1/\lambda_1$ and $a_2 = 1/\lambda_2$. By the above expressions for mean and variance of T , we find the conditions $a_1^2 + a_2^2 = \hat{\sigma}^2$, and $a_1 + a_2 = \hat{\mu}$. This results in the expression $a_{1,2} = \frac{1}{2}\hat{\mu} \pm \sqrt{\frac{1}{2}\hat{\sigma}^2 - \frac{1}{4}\hat{\mu}^2}$, so that

$$\lambda_{1,2} = \frac{1}{a_{1,2}} = \left(\frac{1}{2}\hat{\mu} \pm \sqrt{\frac{1}{2}\hat{\sigma}^2 - \frac{1}{4}\hat{\mu}^2} \right)^{-1}.$$

Note that these estimates become non-sensical in case $\frac{1}{2}\hat{\sigma}^2 - \frac{1}{4}\hat{\mu}^2 < 0$ or if $\sqrt{\frac{1}{2}\hat{\sigma}^2 - \frac{1}{4}\hat{\mu}^2} \geq \frac{1}{2}\hat{\mu}$. This means that we require

$$\frac{1}{2}\hat{\mu}^2 \leq \hat{\sigma}^2 < \hat{\mu}^2,$$

which will not hold in all cases. This is a limitation of the method of moments, whereas likelihood maximization will provides an estimate for all cases. It is

however also an indication that the proposed model does not need to be a perfect fit for the observed data. In Figure 3, the frequency data of bottle circulation times is compared with the best hypoexponential fit, using the method of moments.

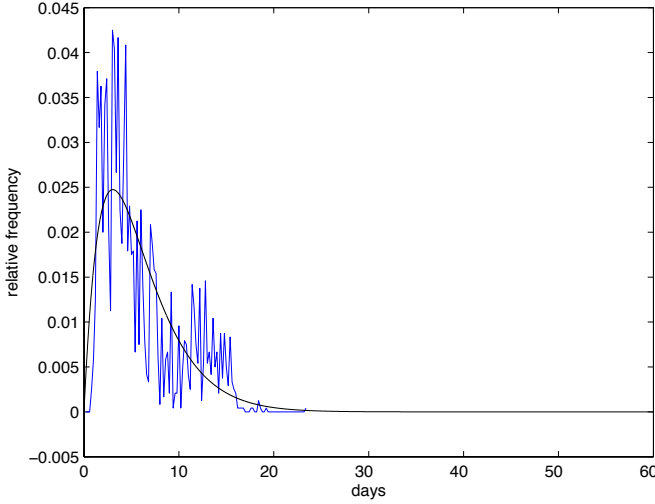


Figure 3: Frequency data of circulation times and the hypoexponential fit according to the method of moments.

3.4.2 Estimation based on stationarity assumption – 2 state model

Consider (1) in a stationary regime. For simplicity assume $h = 1$ and all transition probabilities $\lambda_{..} \ll 1$. We assume the number of bottles sold in a time interval $[t, t+1)$ equals $U(t+h) - U(t) \sim N(\mu_U, \sigma_U^2)$. Furthermore suppose $F \sim N(\mu_F, \sigma_F^2)$, $E \sim N(\mu_E, \sigma_E^2)$, and $Y(t+1) - Y(t) \sim N(\mu_Y, \sigma_Y^2)$. By the discrete time equations (1) (with $h = 1$), we immediately find

$$\mu_F = \frac{\mu_U}{\lambda_{FE} + \lambda_{FB}}, \quad \mu_E = \frac{\lambda_{FE}\mu_F}{\lambda_{EB} + \lambda_{ER}}, \quad \mu_Y = \lambda_{ER}\mu_E,$$

giving

$$\mu_Y = \frac{\lambda_{ER}\lambda_{FE}}{(\lambda_{EB} + \lambda_{ER})(\lambda_{FE} + \lambda_{FB})} \mu_U$$

and thus providing an equation for the unknown parameters $\lambda_{..}$ in terms of means μ_U and μ_Y , which can be estimated by the respective sample averages.

3.4.3 Estimation based on stationarity assumption – 1 state model

By a more involved analysis concerning covariances, extra equations may be obtained. We will explain this idea for a simplified model with only one recurrent state. It should in principle be possible, but more involved, to carry out the same analysis for the two-state model.

Consider the situation in which, within a single time step, a bottle can be broken (rate per unit time λ_B) or returned (rate λ_R). See Figure 4.

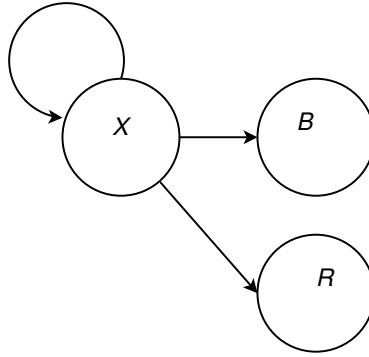


Figure 4: A simple Markov model with one absorbing state, used for determination of model parameters.

In a time interval $[t, t + 1)$ a number of $U_{t+1} - U_t$ bottles is bought, independent of the number of unreturned bottles X_t . Assuming stationarity of the randomness as before, we have the following simple Markov system:

$$X_{t+1} = (1 - \lambda_B - \lambda_R)X_t + U_{t+1} - U_t - \sigma_B \epsilon_t^B - \sigma_R \epsilon_t^R,$$

$$Y_{t+1} - Y_t = \lambda_R X_t + \sigma_R \epsilon_t^R.$$

We further simplify this model by assuming $\sigma_B = \lambda_B \mu_X$, $\sigma_R = \lambda_R \mu_X$, since the variance of a $\text{Bin}(n, \lambda)$ random variable is proportional to $n\lambda(1 - \lambda) \doteq n\lambda$ for small λ . This model has three unknowns $\lambda_B, \lambda_R, \mu_U$. Using the same argument as before, we can relate the empirical means $\hat{\mu}_Y$ and $\hat{\mu}_U$ of μ_Y and μ_U through the equality $\mu_Y = \frac{\lambda_R}{\lambda_B + \lambda_R} \mu_U$. Furthermore,

$$\sigma_X^2 + \mu_X^2 = \mathbb{E} X_{t+1}^2 = \mathbb{E} \left[\left\{ (1 - \lambda_B - \lambda_R)X_t + U_{t+1} - U_t - \sigma_B \epsilon_t^B - \sigma_R \epsilon_t^R \right\}^2 \right]$$

$$= (1 - \lambda_B - \lambda_R)^2 (\sigma_X^2 + \mu_X^2) + \sigma_U^2 + \mu_U^2 + \sigma_B^2 + \sigma_R^2,$$

or equivalently

$$\{1 - (1 - \lambda_B - \lambda_R)^2\} (\sigma_X^2 + \mu_X^2) = \sigma_U^2 + \sigma_B^2 + \sigma_R^2.$$

Therefore

$$\begin{aligned}\mathbb{E}(Y_{t+1} - Y_t)^2 &= \sigma_R^2 + \lambda_R^2 \mathbb{E}X_t^2 = \sigma_R^2 + \frac{\lambda_R^2 (\sigma_U^2 + \sigma_B^2 + \sigma_R^2)}{1 - (1 - \lambda_B - \lambda_R)^2} \\ &= \lambda_R^2 \left(\mu_X^2 + \frac{\sigma_U^2 + \mu_X^2 (\lambda_B^2 + \lambda_R^2)}{1 - (1 - \lambda_B - \lambda_R)^2} \right).\end{aligned}$$

Finally, by similar reasoning,

$$\rho = \mathbb{E}[(Y_{t+1} - Y_t)(U_t - U_{t-1})] = \lambda_R(\mu_U^2 + \sigma_U^2),$$

where the quantity on the lefthand side may be estimated from the data as

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n y_i u_{i-1},$$

where u_i and y_i are the observed sales and returns in time period i , respectively. To summarize we have obtained three equations that relate the unknowns λ_B , λ_R and μ_X in terms of $\hat{\sigma}_U^2$, $\hat{\mu}_U$, $\hat{\sigma}_Y^2$, $\hat{\mu}_Y$, and $\hat{\rho}$.

3.5 Discussion

The above results are for illustration purposes. A more detailed analysis should be performed to determine which simple model might adequately describe the dynamics of the system. From such a model, equations should be derived as in the last section which estimate system parameters from observed statistics. Then state estimation may be performed on-line to compute actual estimates of number of full and empty bottles in the system, using Kalman filtering, based on observations of sold bottles and returned bottles, or perhaps using observations of circulation times.

In the full-for-empty system, in which full bottles are only sold once the same number of empty bottles is returned, extra modelling is necessary. A simple Markov model might than model the behaviour of a customer, who in a time period may drink a unit, do nothing, or buy a bottle and thus also return bottles.

4 A queueing model for the number of bottles in the market

In Section 3, we described a rather detailed modelling approach to the description of bottles in the market. We now discuss alternative models from queueing theory that have been thoroughly investigated in the literature and are by now well understood. These models stem from different modelling assumptions on the demand process (a compound Poisson process) and particularly focus on fluctuations of the demand rate in time. Another important difference is that these models only describe bottles that will be returned; the rate at which bottles break should be discounted in the demand

process. That is, if the demand rate is x and a fraction p of the bottles are broken, then $(1 - p)x$ is the effective demand rate of bottles that will be returned.

In this section we describe the infinite server model from queueing theory that serves as a building block to model the number of bottles out in the market. The following is required as *input* for this model: The effective demand rate (function) and the distribution of d_t , the circulation time (cf. the definition in Section 2). The output is a distribution for the number of bottles that are simultaneously out in the market.

4.1 Constant demand rate, fixed circulation time distribution

We start by assuming that the demand has a constant rate and can be modeled by a Poisson process. Specifically, we start by assuming that the number of bottles purchased in an interval of t days has a Poisson distribution with an average of λt bottles, where λ is the daily average demand rate. Letting D denote a generic random variable having the distribution of the circulation time of a bottle we find that N –the number of bottles out in the market (in the stationarity regime)– has a Poisson distribution with mean $\lambda \mathbb{E}D$. The distribution of D enters the calculation only through its mean; this is usually referred to as “insensitivity” toward the circulation time distribution. The main requirement is that the time out in the market for each bottle is independent of that of all other bottles and shares the distribution of D . This result is standard in queueing theory and can be found in any textbook, see for example [5].

In reality, both the demand rate and the circulation time is season-dependent and/or may have a certain non-stationary trend. We discuss these in the following two paragraphs.

4.2 Time varying demand rate, fixed circulation time distribution

Assume now that the demand rate fluctuates over time. At time t it is $\lambda(t)$, i.e., the demand process is a time-varying Poisson process with rate function $\lambda(t)$. Still the number of bottles out in the market has a Poisson distribution, but this is no longer insensitive to the circulation time distribution. Now, the number of bottles out in the market at time t has a Poisson distribution with mean

$$\mathbb{E}N(t) = \int_0^\infty \mathbb{P}(D > v) dv = \int_0^\infty \left(\int_{t-v}^t \lambda(u) du \right) f_D(v) dv, \quad (4)$$

where $f_D(t)$ is the density of the distribution of the circulation time D ; see for example Theorem 1 of [3]. For a constant demand rate, we recover from equation (4) that $\mathbb{E}N(t) = \lambda \mathbb{E}D$.

4.3 Time varying demand rate, time varying exponential circulation time distribution

If we restrict on the generality of the circulation time distribution and assume it has an exponential distribution, then a rather classical paper by [2] generalizes the previous to

the case where the mean of the circulation time distribution may fluctuate over time. Again the number of bottles in the market at time t is Poisson with mean $J(t)I(t)$ where

$$J(t) = e^{-\int_0^t \nu(u) du}, \quad (5)$$

with $1/\nu(t)$ representing the mean circulation time at time t , and

$$I(t) = \int_0^t \frac{\lambda(u)}{J(u)} du. \quad (6)$$

Of course, if we choose a fixed exponential circulation time by setting $\nu(t) = \frac{1}{\mathbb{E}D}$ in (5) and setting $f_D(t) = \frac{1}{\mathbb{E}D} e^{-\frac{t}{\mathbb{E}D}}$ in (4), we obtain the same result.

For the data available from Heineken, this is probably the most useful model. In case more is known about the characteristics of the demand function and circulation time distribution, a useful extension to the above can be found in [6]. The model there allows for time varying non-exponential demand and circulation time distributions (specifically, they allow for phase type distributions).

5 Sampling the circulation time

A key quantity to understand is the expectation for the circulation time of bottles (see Section 2.2.2 for a definition of circulation time). We describe a method to obtain this value, together with a confidence interval depending on the sample size. We also discuss how other fluctuations in the beer market, like seasonality, can be incorporated in the method to improve the estimations.

The statistical theory of estimation, sampling, and confidence intervals is well-developed. In line with this theory, we consider the circulation times of individual bottles as *random variables* with *identical* distribution. Shape or parameters of this distribution are obtained by means of sampling, that is, computation of the circulation time for a small number of bottles (a sample), and extrapolation of the findings for this sample to the entire population of bottles.

The practical side of sampling is easy: When a bottle is returned, the expiry date on the label allows calculation of the production date, which in turn gives a sound estimate of the time of sale. Together with the time of return of the bottles, this allows a fairly exact computation of the circulation time. However, a high-volume or even continuous computation of circulation times in this way is expansive and impractical. Therefore, we first discuss in this section the required sample sizes in order to guarantee a certain confidence limit for the parameters. Subsequently, we discuss how seasonality and other artifacts of the beer market can be incorporated in order to improve the estimation.

A standard assumption to facilitate the statistical analysis is (mutual) independence of the circulation time of the bottles in a sample. Therefore, the choice of the sample should be made as random as possible.

5.1 Batch sizes

We address now the question how many bottles should be sampled in order to guarantee a certain accuracy of the circulation time. Suppose we have a sample of N different bottles with circulation times X_1, X_2, \dots, X_N . Under the assumption that X_1, \dots, X_N are *independently and identically distributed* (i.i.d.), we use the sample to infer on the (unknown) distribution of circulation time. In parametric statistics, one assumes a certain *family of distributions* indexed by a finite-dimensional parameter space. It is then sufficient to estimate these parameters.

Usefulness of this approach is crucially relying on a decent choice of the family of distributions. The often used *normal family* identifies a 95%-confidence interval for the mean as all points at distance smaller than 1.96 times sample standard deviation from the sample mean.

In light of Section 3.4, it seems most reasonable to choose as model the hypoexponential distribution with two parameters. Derivation of confidence limits in closed form, such as for the normal family, seems impossible for this model. Nevertheless, *bootstrapping* provides a theoretically not very pleasing, yet very efficient, practical method to determine confidence interval by means of Monte Carlo simulation. This works as follows. Start by estimating the two parameters of the hypoexponential distribution by using either of the methods (i) or (ii) in Section 3.4. Use then a statistical software package to generate a high number, say 1000, of i.i.d. random variables with this distribution using the estimated parameters, and sort them from smallest to largest (call them Y_1, \dots, Y_{1000}). The $(1 - \alpha)\%$ -confidence interval for the circulation time as given by the bootstrap is the interval $[Y_{\alpha/2 \times 1000}, Y_{(1-\alpha/2) \times 1000}]$.

5.2 Seasonality

An artifact of the beer market is seasonality. Sales show a certain seasonal peak, typically located in summer, when people drink more beer than in other times of the year. Particularly in *full-for-empty* systems for returnable packaging materials, it is tempting to believe that customers operate with more bottles during peak time, and store some bottles elsewhere throughout the rest of the year. If this reasoning is true, then seasonality has an impact on the circulation time (at the start of the peak, customers bring the stored bottles, which yields a higher circulation time).

We propose to carry out a statistical test whether the null hypothesis H_0 : “Seasonality has no significant effect on circulation time” can be rejected in favor of the alternative hypothesis H_1 : “Seasonality does have a significant effect on circulation time”. A possible test could go as follows. Gather data at several moments in the year, e.g. monthly, bi-monthly or quarterly. Call k the number of measurements per year, and K the total number of measurements. Record the following data:

Y_n	circulation time at time n
$\bar{Y} = 1/K \sum_{n=1}^K Y_i$	mean circulation time,
X_n	sales volume at time n ,
$n = 1, 2, \dots, K$	time.

We are now considering the general linear model in centralized form:

$$Y_n - \bar{Y} = \beta_1(X_n - X_{n-1}) + \beta_2\left(X_n - \sum_{i=1}^k X_{n-i}\right) + \varepsilon_n, \quad n = 1, \dots, K, \quad (7)$$

where the error terms ε_n ($n = 1, \dots, K$) are i.i.d. normally distributed. In this model, β_1 ‘explains’ derivation in circulation time by an upward or downward sales trend (mimicking the beginning or end of a peak period). Further, $X_n - \sum_{i=1}^k X_{n-i}$ is large if we are in a peak, and small otherwise. Hence, β_2 simply relates circulation time to peaks. Of course, the above model could easily be adapted to account for other effects, for example, incorporating long-term trends in sales. Mind that the Y_n itself are sample means, which justifies our normality assumption for the ε_n .

With this model at hand, our earlier described null hypothesis can be sharpened as

$$H_0: \beta_1 = \beta_2 = 0.$$

In order to test the hypothesis, we use multilinear regression obtaining regression sum of squares (RSS) and error sum of squares (ESS). Dividing both by their corresponding degree of freedom (2 for RSS, $K-2$ for ESS), and comparing the ratio of these two with the $F(2, n-2)$ -distribution obtains the p -value associated with the data. This is known in the statistical literature as ANOVA (‘analysis of variance’). Further tests could be imposed if the null hypothesis has been rejected at the desired level of confidence.

A somewhat simpler approach would concentrate only on the ‘peak’ phenomenon. This time, we take only two samples per year, one at the beginning of the peak, and the other one at the end of the peak. We estimate parameters for each of these two samples separately, and compare how ‘distant’ they are using the methods described in the next subsection. This method is simpler than the one described before, but sheds no light on other (possible) temporal dependencies.

5.3 Handling different distribution channels

Circulation time may very well depend on the distribution channel. The main difference we shall consider here are the channels bars/restaurants on the one hand, and private customers on the other hand. Similar to the discussion in the previous subsection, we suggest a statistical test to investigate the issue. The general setup is somewhat easier this time. We are in the situation where we have two samples, with two estimated sets of parameters, and now we want to test whether they are “significantly different” in order to reject the null hypothesis H_0 : “There is no difference in circulation time parameters for different distribution channels.” The F -test (ANOVA) is the right one under the assumption of normality. However, there are also generally applicable non-parametric tests, for details we refer to Section 11.2 in [8].

5.4 Unreadable labels – an indicator for a long circulation time

A practical problem that occurs at the sampling procedure, is that the expiration date is not readable on some bottles. This is due to damage to the label, or a completely

removed label. It is likely that these bottles have a longer circulation time than the bottles with readable expiration dates, so it would bias the statistics if one leaves these bottles out of the sample. Although there are methods available to reduce the bias compared to leaving these bottles out of the sample (for example, the EM algorithm), we did not look into this any further.

6 Conclusion

In Section 2.3, we discussed a way of estimating the break rate of bottles from data that is currently available. This is already an interesting result in itself, but can also be used to identify some parameters in the stochastic flow models (see (1) and (3)) and in the queueing model (see Section 4.3). In Section 3 and Section 4, we then studied two different kinds of models for the number of bottles in the market. All of our models are quite simple, and might therefore not be very accurate in practice. Further (statistical) research is required to test their accuracy. It may be necessary to increase model complexity by dropping some assumptions. However, we note that this may result in a) lack of explicit solutions of the models, b) more unknown parameters that need to be estimated and/or c) increased computational effort.

In Section 5 it is discussed how to obtain the expected circulation time of a bottle in the market from the data, together with a confidence interval. As this value may depend on the time in the year at which the sampling has been done, we also discussed a method to test whether it is reasonable to assume that this seasonality is of little importance.

References

- [1] D. P. Bertsekas. *Dynamic programming and optimal control. Vol. I.* Athena Scientific, Belmont, MA, third edition, 2005. ISBN 1-886529-26-4.
- [2] T. Collings and C. Stoneman. The $m/m/\infty$ queue with varying arrival and departure rates. *Operations Research*, 24(4):760–773, 1976.
- [3] S. G. Eick, W. A. Massey, and W. Whitt. The physics of the $m_t/g/\infty$ queue. *Operations Research*, 41(4):731–742, 1993.
- [4] W. H. Fleming and R. W. Rishel. *Deterministic and stochastic optimal control.* Springer-Verlag, Berlin, 1975. Applications of Mathematics, No. 1.
- [5] D. Gross and C. M. Harris. *Fundamentals of Queueing Theory.* Wiley New York, 1998.
- [6] B. L. Nelson and M. R. Taaffe. The $ph_t/ph_t/\infty$ queueing system: Part i – the single node. *INFORMS journal on computing*, 16(3):266–274, 2004.

- [7] B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. ISBN 3-540-04758-1. An introduction with applications.
- [8] J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury press, 2007.
- [9] R. F. Stengel. *Optimal control and estimation*. Dover Publications Inc., New York, 1994. ISBN 0-486-68200-5. Corrected reprint of the 1986 original.

