Learning in revenue management: Exploiting estimation of arrival rate and price response

Athanassios N. Avramidis
School of Mathematics, University of Southampton
Southampton, SO17 1BJ, UNITED KINGDOM
aa1w07@soton.ac.uk
Telephone: +44 02380 595136
Fax: +44 02380 595147
August 25, 2013

Abstract

The paper first studies dynamic pricing to maximize expected revenue of a fixed inventory of a single product under Poisson arrivals of random rate, given a Bayesian prior, and known distribution of reservation prices. For a single unit having a salvage value, we show there exists a unique revenue-maximizing price, which increases in the salvage value, provided the reservation-price hazard function is increasing. For multiple units, a discrete-time dynamic program is studied. Empirically, the optimal price increases in uncertainty, and is sensitive to the prior choice. The paper then considers a seller that knows no parameter values; all he knows is that sales arise from Poisson arrivals, where a Bernoulli random variable, independent of everything else, converts any arrival into a sale. This can represent any demand function as in Gallego and van Ryzin (1994) and Besbes and Zeevi (2009), but additional independence conditions are present here. Observing arrivals and sales at each price during part of the sale horizon, we construct estimators of the arrival rate and purchase probabilities; we refer to this process as learning. We derive the bias and mean squared error of the resulting demand-function estimator. Relative to the sale-count-only estimator of Besbes and Zeevi (2012), the summed mean squared error (across all prices) is consistently reduced, empirically. Exploitation methods based on these estimators are proposed, where the time spent learning is as Besbes and Zeevi (2012) prescribe. Empirically, the methods’ loss against the full-information optimum is competitive to the benchmark of Besbes and Zeevi (2012).
Keywords: Applied Probability; Revenue Management; Dynamic Programming; Random-Rate Poisson Process; Likelihood; Hazard Rate; Mean Squared Error.

1 Introduction

A central problem in revenue management is as follows: given an initial inventory of a product to be sold over a fixed selling horizon, the seller must decide how to adjust the price through time so as to maximize the expected revenue (under the assumption that the inventory level cannot be changed through the selling season). Talluri and van Ryzin (2004) discuss instances of this problem that range over many industries, including fashion and retail, air travel, hospitality, and leisure.

A common assumption in the literature is that the process that counts demand for the product (that is, sales, in the absence of inventory constraints) is (stationary) Poisson of rate given by the so-called demand function $\lambda(p)$, where $p$ is the price on offer (Gallego and van Ryzin, 1994). Because of the Markovian structure, the optimal dynamic pricing policy can be derived from the relevant Bellman equation.

In an attempt to relax the assumption that the demand function is known, a string of literature begins by representing the demand function as the product of an arrival rate that is unknown, times a known probability with which each possible price converts any arrival into a sale. This literature includes Lin (2006), Aviv and Pazgal (2005), Araman and Caldentey (2009), and Farias and Van Roy (2010). Farias and Van Roy (2010) assume that customers arrive according to a Poisson process; have identically distributed reservation prices; and purchase only if their reservation price exceeds the seller’s price (at their arrival time). The other papers study similar formulations. The arrival rate is given a prior distribution (gamma, two-point, or a mixture of gammas across these papers); the Bellman equation is employed, where the transition probability function requires the conditional expectation (mean) of the arrival rate (and thus demand rate) with respect to the seller’s information, which is one of arrival or sale counts. This conditional expectation is updated by Bayesian learning, meaning here application of Bayes’ rule (under the prior). Explicit analytical solution is typically difficult, and so heuristics are often constructed from the Bellman equation in continuous time (Araman and Caldentey, 2009; Farias and Van Roy, 2010), or from the corresponding dynamic program in discrete time Lin (2006). While the mixture-of-Gammas prior of Farias and Van Roy (2010) is arguably very flexible, Besbes and Zeevi (2009) note a limitation of this approach in general: the resulting optimality is only with respect to the prior. The first part of this paper lies in this area. We study a problem that shares all elements of Farias and Van Roy (2010) above, except that we do not discount the future and work in discrete time,
while they do the opposite. For the one-unit problem, we show that a finite mean reservation price ensures the maximization problem is well-posed (has a finite supremum), and that price increases in the salvage value when the reservation-price distribution has increasing hazard rate (also seen in Farias and Van Roy (2010)). For the general multi-unit problem, we formulate a dynamic program in discrete time where the arrival rate is estimated from observations: it is either the posterior Bayesian mean under an unrestricted prior, or the maximum-likelihood estimator. Our structural results are less complete than those in Farias and Van Roy (2010), but as these authors use time discounting, their results do not translate easily to our setting. We show that the price of the one-unit problem partly characterizes the pricing policy; the multi-unit price approaches the single-unit price as time approaches the horizon end, and is lower-bounded by it at all times. In examples with gamma and right-truncated gamma priors, we find that price increases in uncertainty, in agreement with these works, and we observe sensitivity to the prior, which reinforces the limitation of Bayesian optimality mentioned above.

If the demand function is known, policies that are increasingly good as the problem scale (mean demand) increases can be constructed from linear-programming (LP) models (Gallego and van Ryzin, 1994, 1997). Besbes and Zeevi (2009) and Besbes and Zeevi (2012) developed a novel approach where the seller acts against a demand function that is fully unknown. For a fixed finite price set, their algorithms involve: learning (estimation) of the unknown demand function, followed by exploitation, based on the solution to the sample analog of an LP model as above, where the demand function is replaced by its estimate from learning. Given a basic problem where capacity is \( C \), sale horizon is \( T \), and the demand function is \( \lambda() \), they consider a sequence of problems such that in the \( n \)-th problem capacity is \( nC \) and the sale horizon is \( nT \) (or, equivalently, take sale horizon \( T \) and demand function \( n\lambda() \)). They show that the fractional loss relative to the full-information optimum (regret) converges to zero as \( n \to \infty \) (for various settings, and including convergence rates). For a fixed finite price set, the theory dictates that learning time should increase as \( n^{2/3} \) (Besbes and Zeevi, 2012, Theorem 1).

In the second part of the paper we first develop new estimation methodology about the demand process in a setting more restrictive than the standard one. Specifically, we fix a finite price set, and assume (stationary) Poisson arrivals and (individual customer) purchase indicators that are Bernoulli random variables independent of the arrivals. The seller knows only the validity of this assumption—neither arrival rate nor purchase probabilities are known. This model is rich enough to generate an arbitrary demand function (observed at a fixed finite set of prices). Cope (2006) makes a similar assumption, but he only estimates purchase probabilities, not the arrival rate (the latter is unimportant in his context), and
his method is Bayesian, while ours is frequentist. While we estimate arrival rate and purchase probabilities separately, a by-product of these is an estimator of the demand function. Under a no-stock-out assumption, we derive the bias and mean squared error (MSE) of this estimator, and see that its order is inverse to the time spent learning. In our setting, under uniform and exponential reservation-price distributions, our estimator consistently has lower summed mean squared error (SMSE) (summation over all price points) than the sale-data-only estimator of Besbes and Zeevi (2012, Section 3.1) (we exploit arrivals in addition to sales, while they do not, so this is what one should expect). Based on learning as above, we develop two exploitation methods: (i) the standard LP model, fed by the new demand-function estimator; and (ii) a dynamic program (DP) of the exploitation process where the transition probability requires the arrival-rate and purchase probabilities on their own. In experiments where the learning time is chosen sensibly \((cn^{2/3})\) with a sensible \(c\), where \(n\) is the problem size), the new methods’ regret decreases with \(n\), and is typically smaller than that of Besbes and Zeevi (2012); overall, DP leads by a significant margin.

The remainder of this paper is organized as follows. Section 2 develops the pricing problem with random arrival rate and known reservation price distribution. Section 2.1 discusses the essential object: the Bayesian conditional mean, and alternatively the maximum-likelihood estimator, of the arrival rate with respect to arrival or sale information. Section 2.2 studies optimal pricing under this demand model. Section 2.2.1 studies a single-unit pricing problem where reservation prices have known distribution. In Section 2.2.2 we develop the dynamic program for the general multi-unit problem. Numerical illustrations appear in Section 2.2.4. The model where all demand-process parameters are unknown is in Section 3. Section 3.1 develops estimators of all the parameters and studies the estimation error. Section 3.2 develops revenue-management methods that build on the estimator above, and studies their performance experimentally.

## 2 Pricing Problem

The seller has an inventory (capacity) of \(C\) units to sell, over a time horizon \(T\), after which products have no value. Customers arrive according to a Poisson process of rate \(\lambda\), each having a reservation price for the product with tail probability \(\bar{H}\), and independent of everything else. The arrival rate is viewed as random. The seller posts a price at all times; a price of \(x\) results in a sale and revenue \(x\) with probability \(\bar{H}(x)\), and no sale otherwise. The seller seeks an optimum pricing policy. In practice, finitely many price points are often adequate or even desirable. In this case, knowledge of \(\bar{H}\) is equivalent to knowledge of a customer’s purchase probability at each price point. This assumption, common in the literature, may
be too restrictive in practice.

2.1 Learning of the Arrival Rate

Earlier models (Lin, 2006; Aviv and Pazgal, 2005; Araman and Caldentey, 2009; Farias and Van Roy, 2010) are all Bayesian and impose on \( \lambda \) a prior distribution, \( \pi_0 \), that in each paper is restricted to some parametric family. We unify and extend the approach by keeping the parametric restriction in the background and also by allowing learning via maximum likelihood, which eliminates the need for prior parameters. We begin with Bayesian learning.

Assume that arrival counts \( A_{t_j} \) are observed at times \( 0 < t_1 < \ldots < t_k = t \). Writing the likelihood of the event \( \{ \lambda = x, A_{t_j} = a_{t_j}, j = 1, \ldots, k \} \), we obtain the conditional distribution of \( \lambda \) (up to a constant):

\[
\pi_0(x) \prod_j e^{-x(t_j-t_{j-1})} \frac{[x(t_j-t_{j-1})]^{a_{t_j}-a_{t_{j-1}}}}{(a_{t_j}-a_{t_{j-1}})!} \propto \pi_0(x) e^{-xt} x^{a_t}. \tag{1}
\]

Thus, the distribution of \( \lambda \) depends on the individual \( A_{t_j} \) only through \( A_t \) (the same is true for the apparently larger information \( A_s : s \leq t \)), and the conditional mean arrival rate \( \mathbb{E}[\lambda | A_{t_j}, j = 1, \ldots, k] \) is a function \( g(t, A_t) = g(t, A_t; \pi_0) \). An essential consequence of the existence of the function \( g \) is that there exist value functions in the dynamic program (DP) describing an optimum pricing policy (Section 2.2.2) that depend only on time \( t \), remaining capacity, and \( A_t \); that is, the individual \( A_{t_j} \) do not matter.

These observations extend to the alternative that sales are observable and arrivals are not. Specifically, let \( p_t \) be the price at \( t \) and assume sale counts \( S_{t_j} \) are observed at times \( 0 = t_0 < t_1 < \ldots < t_k = t \). Conditional on \( \lambda = x \), the process \( \{ S_t : t \geq 0 \} \) is Poisson with rate function \( m_t = x \int_0^t \bar{H}(p_s)ds \) (Lin, 2006, Section 5.1). Writing the likelihood of the event \( \{ \lambda = x, S_{t_j} = s_{t_j}, j = 1, \ldots, k \} \), we obtain the conditional (marginal) of \( \lambda \):

\[
\pi_0(x) \prod_{j=1}^k e^{-x\Delta_j m} \frac{(x\Delta_j m)^{\Delta_j} \Delta_j!}{\Delta_j^s!} \propto \pi_0(x) e^{-xm_t} x^{s_t}. \tag{2}
\]

where \( m_{t_j} = \int_0^{t_j} \bar{H}(p_s)ds, \Delta_j m = m_{t_j} - m_{t_{j-1}} \), and \( \Delta_j s = s_{t_j} - s_{t_{j-1}} \). Here, the conditional of \( \lambda \) depends on the \( S_{t_j} \) only through \( S_t \); it belongs to the family (1); and \( \mathbb{E}[\lambda | S_{t_j}, j = 1, \ldots, k] = g(m_t, S_t) \). The DP state component need only include \( S_t \) and not the \( S_{t_j} \).

Explicit formulas for \( g \) may be found for a \( \pi_0 \) that is Gamma (Lin, 2006; Aviv and Pazgal, 2005), two-point discrete (Araman and Caldentey, 2009), or a finite mixture of Gammas (Farias and Van Roy, 2010). Other convenient prior families are the truncated Gamma (which contains the Gamma and Uniform Families) and the general finite discrete; we give the relevant facts for these.
Gamma prior, possibly truncated. The \( \mathcal{RG}(a,b,u) \) (right-truncated Gamma) density is \( (b^a / \Gamma(a)) x^{a-1} e^{-bx} 1_{x \leq u} \), and the mean is \( b^{-1} \gamma_L(a+1, bu) / \gamma_L(a, bu) \), where \( \gamma_L(a, u) := \int_0^u x^{a-1} e^{-x} \, dx \) is the lower incomplete gamma function. Under this prior, the posterior (1) belongs to the same family, and is \( \mathcal{RG}(a+A_t, b+t, u) \). Similarly, the \( \mathcal{LG}(a, b, u) \) (left-truncated Gamma) density at \( x \) is \( \left( b^a / \Gamma(a) \right) x^{a-1} e^{-bx} 1_{x \geq u} \), and the mean is \( b^{-1} \gamma_U(a+1, bu) / \gamma_U(a, bu) \), where \( \gamma_U(a, u) := \int_u^\infty x^{a-1} e^{-x} \, dx \) is the upper incomplete gamma function. Under this prior, the posterior is \( \mathcal{LG}(a + A_t, b + t, u) \).

Finite Discrete prior. Suppose \( \pi_0 \) is discrete with support \( \{ x_1, x_2, \ldots, x_k \} \) and respective probability masses \( q_1, q_2, \ldots, q_k \). The posterior (1) assigns probability \( w_j := q_j e^{-x_j t} x_j^a / \left( \sum_i q_i e^{-x_i t} x_i^a \right) \) to \( x_j \), for each \( j \); then, \( g(t, a_t) = \sum_{j=1}^k x_j w_j \).

In any case, \( g \) may be estimated by sampling from (1) via Markov Chain Monte Carlo, but this approach tends to be involved (Asmussen and Glynn, 2007, Chapter XIII).

A complementary approach is via maximum likelihood. Here, given arrivals as detailed above, we define \( g(t, a_t) \) as the \( x \) that maximizes the likelihood function, defined as (1) with \( \pi_0(x) = 1 \), which gives \( g(t, a_t) = a_t / t \). Alternatively, under observation of sales as detailed above, the maximizer of the likelihood ((2) with \( \pi_0(x) = 1 \)) is \( g(t, s_t) = s_t / m_t \). This approach is free of any prior parameters.

2.2 Optimal Pricing

In Section 2.2.1 we study a one-unit problem where an unsold item has a salvage value. Building on this, Section 2.2.2 studies a DP formulation of the general multi-unit problem.

2.2.1 Pricing a Single Unit with Salvage Value

Consider a single certain arrival, as above, and suppose one unit is available, with salvage value \( v \) when the customer does not purchase. (In the multi-unit problem studied in the next section, the salvage value is determined via a dynamic program.) If the seller prices at \( x \), then his expected value is

\[
\bar{H}(x) x + (1 - \bar{H}(x)) v = \bar{H}(x)(x - v) + v.
\]

For any fixed \( v \geq 0 \), we seek the supremum of \( G(x) = G(x; v) = \bar{H}(x)(x - v) \) over \( x \geq 0 \). Note that \( x < v \) is dominated by \( x = v \). It is equivalent to maximize \( \bar{G}(x) = \bar{G}(x; v) = \log[\bar{H}(x)(x - v)] \) over \( x \geq v \).

For the problem to be well-posed, we require that for all \( v < \infty \), the supremum be finite, i.e., \( \sup_x G(x; v) < \infty \). A sufficient condition for this is that \( \bar{H} \) have finite mean. We state a stronger result from which this follows.
Lemma 1  If $\bar{H}$ has finite mean, that is, $\int_0^\infty \bar{H}(x)dx < \infty$, then for any nonnegative function $g$ such that $\int_0^\infty g(x)dx = \infty$, we have $\lim_{x \to \infty} \bar{H}(x)/g(x) = 0$.

Proof. Suppose $\bar{H}(x)/g(x)$ does not converge to zero. Then there exist $\epsilon > 0$ and a sequence $\{x_n\}$, $x_n \uparrow \infty$, such that $\bar{H}(x_n)/g(x_n) \geq \epsilon$ for all $n$. This implies $\int_0^\infty \bar{H}(x)dx \geq \int_0^\infty \epsilon g(x)dx = \infty$, which is a contradiction; this proves the claim.

Taking $g(x) = 1/x$ yields

$$\lim_{x \to \infty} x\bar{H}(x) = 0.$$ 

In view of this, the function $R^*(v) := \sup_{x \geq 0} G(x; v)$ is finite for all $v \geq 0$.

Proposition 1  (a) Assume that $\bar{H}$ has finite mean and a hazard $r(x) = -d\log \bar{H}(x)/dx$ that is finite on $(0, \infty)$. Then for all $v \geq 0$, we have $R^*(v) > 0$.

(b) Assume, in addition, that $r$ is continuous and increasing on $(0, \infty)$. Then, the supremum of $\tilde{G}(\cdot; v)$ occurs at the unique $x^*$ satisfying $f(x^*) = v$, where $f(x) = x - 1/r(x)$. We have $R^*(v) = \exp(\bar{H}(f^{-1}(v))/r(f^{-1}(v)))$, where $f^{-1}$ is the function inverse to $f$. Moreover, $f$ and $f^{-1}$ are strictly increasing, and $R^*$ is decreasing.

Proof. Proof of (a). Fix any $v \geq 0$. The derivative of $\tilde{G} = \log G$ with respect to $x$ is $\tilde{G}'(x) = -r(x) + 1/(x - v)$. As $r$ is bounded on any finite interval, $\tilde{G}'(x)$ is positive for all $x$ in a right neighborhood of $v$. This proves (a).

Proof of (b). The continuity of $r$ implies continuity of $\tilde{G}'$. Observe that $\tilde{G}'$ is negative for some $x$ large enough, for otherwise we obtain a contradiction to $\lim_{x \to \infty} \log(x\bar{H}(x)) = -\infty$, which is true, as seen following Lemma 1. Thus, there exists a solution to $\tilde{G}'(x) = 0$, i.e., $f(x) = v$. The increasing property of $r$ implies that $\tilde{G}'$ is strictly decreasing, so the solution is the unique maximizer of $\tilde{G}$ and $G$. As $f$ is strictly increasing, its inverse function $f^{-1}$ exists and is strictly increasing, and the maximizer may be written $x^* = f^{-1}(v)$. The expression of $R^*$ follows on substituting the maximizer. The monotonicity of $R^*$ follows from that of $f^{-1}$, $\bar{H}$, and $r$. $\Box$

We briefly consider some special cases, including violations of the assumption on the hazard above, where the maximizer is unique and easily identified.

(a) The Weibull distribution with shape $\beta > 0$ and scale 1 can be defined via $\bar{H}(x) = e^{-x^\beta}$ for $x \geq 0$; Here, $r(x) = \beta x^{\beta - 1}$; and for any $v$, $\tilde{G}'(x) = 0$ has a unique solution. For $v = 0$, the explicit solution is $x^* = \beta^{-1/\beta}$. For $\beta = 1$, this is the exponential distribution; in this case, and now allowing an arbitrary scale (mean) $\alpha$, we have $r(x) = \alpha^{-1}$ and $x^* = \alpha + v$. 


(b) The Pareto distribution with shape $\alpha > 0$, scale $\beta > 0$, and support $(0, \infty)$ can be defined via $\bar{H}(x) = (1 + x/\beta)^{-\alpha}$. The mean is finite if and only if $\alpha > 1$, and in this case the mean is $\beta/(\alpha - 1)$. Here, $r(x) = \alpha/(\beta + x)$, and we find $x^* = (\beta + \alpha v)/(\alpha - 1)$.

### 2.2.2 Multi-Unit Pricing

Assume first that arrivals are observable. Given an integer solution resolution, $n_S$, and a time increment $\delta > 0$ linked to it via $n_S\delta = T$, a state $(i, z, y)$ represents that time is $i\delta$, inventory is $z$, and the arrival count up to this time, denoted $A_i$, equals $y$. Truncating the support of $A_i$ above an upper limit $\bar{y}_i$, we have the finite set of states

$$\mathcal{S} = \{(i, z, y) : i \in \{0, 1, \ldots, n_S - 1\}, y \in \{0, 1, \ldots, \bar{y}_i\}, z \in \{0, 1, \ldots, C\}\}. \quad (4)$$

Let $J_{i,z,y}$ denote the value function, defined as the seller’s maximum expected revenue onward from state $(i, z, y)$. An arrival at this state gives rise to the one-unit selling problem in Section 2.2.1, with salvage value $\Delta_z J := J_{i+1,z,y+1} - J_{i+1,z-1,y+1}$. By the dynamic programming principle, the value functions are determined by backward recursion in time:

$$J_{i,z,y} = \max_x \bar{H}(x)(x - \Delta_z J) + J_{i+1,z,y+1} + (1 - q_{i,y})J_{i+1,z,y}, \quad i = n_S - 1, n_S - 2, \ldots, 0, \quad (5)$$

where $q_{i,y} = g(i\delta, y)\delta$, with $g(i\delta, y)$ being the conditional mean, or the maximum-likelihood estimate, of the arrival rate at the current state (Section 2.1), and subject to boundary conditions: $J_{n_S,z,y} = 0$ for all $z, y$, and $J_{i,0,y} = 0$ for all $i, y$. Assuming alternatively that sales are observable, only state component $y$ need be modified, so it represents the sale count up to time $i\delta$, and $q_{i,y} = g(m_i\delta, y)\delta$, with $m_i$ given below (2). For the expectations in (5) to make sense, we require

$$\max_{0 \leq i \leq n_S-1, 0 \leq y \leq \bar{y}_i} g(i\delta, y)\delta \leq 1. \quad (6)$$

This requirement is ensured later.

Following Proposition 1, the maximized term above equals $R^*(\Delta_z J)$, attained by the (optimal) price $p_{i,z,y} = f^{-1}(\Delta_z J)$.

We can characterize the DP and prices as follows.

**Proposition 2**

(a) Provided $f^{-1}$ in Proposition 1 is increasing, the price $p_{i,z,y}$ is lower-bounded by $f^{-1}(0)$ for all $(i, z, y)$ in $\mathcal{S}$.

(b) If $J_{i,z,y+1} - J_{i,z,y} \geq 0$ for all $(i, z, y)$ in $\mathcal{S}$, then the time differences $\Delta_i J_{i,z,y} := J_{i,z,y} - J_{i+1,z,y}$ are strictly positive.

**Proof.** Proof of (a). Observe that $\Delta_z J \geq 0$; this is shown by “throwing away” one unit and acting optimally thereafter. As $f^{-1}$ is increasing, we have $p_{i,z,y} = f^{-1}(\Delta_z J) \geq f^{-1}(0)$ for all $(i, z, y)$.
Proof of (b). Rearranging (5), we have $\Delta_i J_{i,z,y} = q_{i,y} (R^* (\Delta_z J) + J_{i+1,z,y+1} - J_{i+1,z,y})$ for $z > 0$. The parenthesized term is positive because of the assumption and the fact $R^* (\Delta_z J) > 0$. Together with $q_{i,y} > 0$, this proves the claim.

When $f^{-1}$ is strictly increasing (see Proposition 1), a strict concavity of value functions in $z$ (i.e., $\Delta_z J$ is strictly decreasing in $z$) would imply that $p_{i,z,y} = f^{-1} (\Delta_z J)$ is strictly decreasing in $z$, a condition that is typical in the literature.

A simple practical consequence of (a) above (when true) is that the seller prices always, regardless of anything else, at or above $f^{-1}(0)$, which is an (essentially explicit) function of the price-reservation hazard function.

2.2.3 Computation and Implementation of a DP-Based Policy

We assume given a control resolution, $n_C$, defined as the number of equidistant control times at which price updates during the sale horizon are allowed. Recall that $n_S$ (equivalently, $\delta$) and the $\bar{y}_i$ define the set of states, $\mathcal{S}$ in (4). Selection of these involves considerations (and constraints) as follows. First, the requirement (6) suggests that $\delta$ must be small enough ($n_S$ large enough). Second, in the event of an exceedance $A_i > \bar{y}_i$ during operation (recall $A_i$ is the arrival count up to time $i\delta$), our policy heuristically replaces $A_i$ by $\bar{y}_i$ (only for the purpose of pricing at that time); thus, a $\bar{y}_i$ large enough so that exceedances are infrequent is sensible; to this end, if the distribution of $A_i$ is known (e.g., under a gamma prior, it is negative binomial), then $\bar{y}_i$ may be set as an upper quantile of it (e.g., of order 0.99). Third, it is desirable that $n_S$ be an integer multiple of $n_C$; then, (near-optimal) actions at the $t$-th control time, $(t - 1)T/n_C$, are the optimal actions at states in $\mathcal{S}$ whose time index is $(t - 1) n_S/n_C$, for $t = 1, 2, \ldots, n_C$.

Under maximum-likelihood learning, the explicit form of $g$ (Section 2.1) allows explicit selections of $\bar{y}_i$ and $\delta$ that meet the requirements above, and this in turn permits expressing the computing work (speed) and required storage as functions of the primitives $C$, $T$, and $\lambda$; this is done in Section 3.2.1.

2.2.4 Numerical Illustration

Consider an example where $T = 1$; $C = 30$; arrivals are observable; reservation prices are mean-one exponential; and two priors are considered: the gamma (G) with mean one and CV $\in \{0.2, 1, 2\}$ against the right-truncated gamma (RG) found by numerically minimizing the kurtosis, with the mean and CV fixed to those of the gamma counterpart. The DP is solved with $\bar{y}_i$, the order 0.999-quantile of the underlying negative binomial, and $n_S = 128\lambda$. (Significantly larger $\bar{y}_i$ led to overflows in computing the right-truncated gamma moments.)
Figure 1: Price as a function of mean number of arrivals under gamma and right-truncated gamma priors.

Figure 1 shows price as a function of the arrival rate. Price is increasing in uncertainty, while the value $J$ (not shown) decreases; this effect is well-known from Aviv and Pazgal (2005, Figure 2) and Farias and Van Roy (2010). We see evidence of sensitivity to the prior, even when the mean and variance are fixed.

Consistently, we saw the following: value functions were concave increasing in capacity; and price was decreasing in capacity and sell horizon, and increasing in the arrival count (when every other variable was fixed). This includes examples with Pareto reservation prices with infinite variance.

To give an idea of price paths over time, we look closer at an example with $C = 5, \lambda T = 8$, gamma priors of low and high uncertainty, and $n_{c} = 80$ control points. We “sampled” five paths (implied by sampled arrival counts and reservation prices) in stratified way: in instance $i$, the demand (arrival count) is the quantile of order $(i - 1/2)/5$ of its (negative binomial) prior; this set of paths seems more representative than five random paths. Figure 2 shows such paths for two CV values. The price jumps immediately after each sale. An arrival that does not purchase causes a jump whose size is visible in the high-CV case and invisible in the low-CV case.
3 Learning and exploitation of an unknown demand

A fixed finite set of prices, \((p_i)_{i=1}^k\) is assumed, together with

**Assumption 1** Customers arrive according to a Poisson process of rate \(\lambda > 0\). Whenever the \(i\)-th price \((p_i)\) is prevailing, an arriving customer purchases with probability \(q_i\), independently of everything else.

Other than the validity of Assumption 1, the seller knows neither \(\lambda\) nor the \(q_i\). A Poisson process representing demand (sales under no inventory constraints) of rate \(\lambda_i\) at price \(p_i\), which is the standard assumption in the literature, has a representation as in Assumption 1: take arrival rate \(\max_i \lambda_i\) and purchase probability \(\lambda_i / \max_j \lambda_j\) at price \(p_i\). The standard assumption is indifferent to arrivals, while in contrast Assumption 1 requires that purchases occur as the thinning of arrivals via Bernoulli purchase indicators that are independent from arrivals; thus, Assumption 1 is more restrictive. But on the positive side, if it holds, we can expect more accurate estimation of the vector \((\lambda q_i)_{i=1}^k\) by using both arrival and sale counts (not constrained by stock-outs) compared to using only sale counts; Proposition 3 below is the basis for this claim. Assumption 1 may be tenable when there is a large number of users that desire the product independently of others and of past prices, for example in e-commerce (Cope, 2006)); it seems less defensible when a significant number of customers adapt their arrivals, or purchase decisions, to past prices.
3.1 Estimation Method

We estimate separately the arrival rate and purchase probabilities as follows.

**Learning Method A (τ).** Set the learning interval as \([0, \tau]\), and set \(t = \tau/k\). During time \((i-1)t\) to \(it\), for \(i = 1, 2, \ldots, k\), price at \(p_i\) provided inventory is positive, and record the arrival count, \(A_i\), and the sale count, \(S_i\). If a stock-out occurs at any time, post an “infinite” price, and stop sales. In any case, set

\[
\hat{\lambda} := \frac{\sum_{i=1}^{k} A_i}{\tau}, \quad \hat{q}_i := \frac{S_i}{A_i} I_{A_i > 0}
\]  

as estimators of \(\lambda\) and \(q_i\), respectively, where \(I_{A_i > 0}\) is the indicator function of the event \(A_i > 0\). Based on these, we have the estimator \(\theta_A := (\hat{\lambda}, \hat{q}_i)\) of \((\lambda, q_i)\).

The estimator’s (normalized) bias and mean squared error (MSE) are given in (9) and (12) below, respectively, at each \(i\) separately. Proofs are in the Appendix.

**Proposition 3** Let \(X_\lambda\) denote a Poisson(\(\lambda\)) random variable. Put \(\rho := \mathbb{P}(X_{\tau/k} > 0)\), and note \(\rho - 1 = -\exp(-\lambda \tau/k)\). Define \(h(\lambda) := \mathbb{E}[X_{\lambda}^{-1} I_{X_\lambda > 0}]\) and

\[
c_{1,\lambda,\tau,k} := h\left(\frac{\lambda \tau}{k}\right) + 2\rho + (k - 1)^{-1}, \quad c_{2,\lambda,\tau,k} = \frac{k - 1}{k}(1 - 2\lambda) + \lambda \tau \left(\frac{k - 1}{k}\right)^2.
\]  

Under Assumption 1, and provided no stock-out occurs, we have:

\[
B := \frac{\mathbb{E}[\hat{\lambda} \hat{q}_i] - \lambda q_i}{\lambda q_i} = \frac{k - 1}{k}(\rho - 1) = \frac{k - 1}{k} e^{-\lambda \tau/k}
\]

\[
\mathbb{E}[\text{Var}(\hat{\lambda} \hat{q}_i | A)] = q_i(1 - q_i) \left[\lambda^2 \left(\frac{k - 1}{k}\right)^2 h\left(\frac{\lambda \tau}{k}\right) + \lambda \tau^{-1} \frac{k - 1}{k} c_{1,\lambda,\tau,k}\right]
\]

\[
\leq q_i(1 - q_i)\lambda \tau^{-1} \left[c_{0} k \left(\frac{k - 1}{k}\right)^2 + c_{1,\lambda,\tau,k} \frac{k - 1}{k}\right],
\]

\[
\mathbb{E}[(\hat{\lambda} \hat{q}_i - \lambda q_i)^2] = \mathbb{E}[\text{Var}(\hat{\lambda} \hat{q}_i | A)] + q_i^2 \lambda \tau^{-1} (1 - e^{-\lambda \tau/k}) c_{2,\lambda,\tau,k}
\]

where \(A = (A_1, A_2, \ldots, A_k)\), and \(c_0 := \sup_{\lambda > 0} [h(\lambda)/\lambda] < \infty\).

The finiteness of \(c_0\) above is assumed so as to avoid technicalities. Using Monte Carlo simulation, we estimated the maximizer of \(h(\lambda)/\lambda\) as 4.18, and \(c_0 \approx 1.296\). The expressions simplify as \(\lambda \tau/k\) (mean number of arrivals during a single-price learning period) increases; this condition seems natural, and is also prescribed in Besbes and Zeevi (2009, Proposition 1) where both \(\tau\) and \(k\) are free to move. In this case, \(e^{-\lambda \tau/k}\) is negligible, so \((1 - e^{-\lambda \tau/k}) c_{2,\lambda,\tau,k}\) is well-approximated as one; the MSE is of order \(\tau^{-1}\); and the bias (9) is well-approximated as zero. Separately, the summed mean squared error (SMSE) \(v(\theta_A) := \sum_{i=1}^{k} \mathbb{E}[(\hat{\lambda} \hat{q}_i - \lambda q_i)^2]\) is also explicit via (10)-(12).
In comparison, consider the estimator of Besbes and Zeevi (2009, 2012):

\[
\tilde{\lambda}_q := \frac{S_i}{\tau/k}, \quad \theta_S = (\tilde{\lambda}_q i)_{i=1}^k.
\] (13)

The same no-stock-out assumption is needed to obtain the error in closed form; it implies \( S_i \sim \text{Poisson}(\lambda q \tau / k) \), and thus \( v(\theta_S) = \sum_{i=1}^k \lambda q_i (\tau / k)^{-1} \). This is of order \( \tau^{-1} \), but has a different multiplier. With the exponential in (12) set to zero for simplicity, it can be seen that \( \text{MSE}(\tilde{\lambda}_q_i) > \text{MSE}(\tilde{\lambda}_q_i) \) for small enough \( q_i \), but by a small factor, while otherwise we have the opposite, and the margin increases with \( q_i \). The two estimator’s error is reported for examples in Section 3.2.3.

### 3.2 Exploitation After Learning

We will examine, experimentally, the performance of two novel exploitation methods based on the estimators above. Inventory level and sale horizon are denoted \( C \) and \( T \) respectively. The methods are based on ideas in Besbes and Zeevi (2009) and Besbes and Zeevi (2012). A basic problem with capacity \( C \) and sale horizon \( T = 1 \), together with an integer \( n \), generate a new problem, of size \( n \), with capacity \( C_n = nC \); sale horizon \( T_n = n \); and time spent learning \( \tau = \tau_n = cn^{2/3} \), where \( c \) is a constant, and \( \tau_n < n \). The growth of \( \tau_n \) is guided by Besbes and Zeevi (2012, Theorem 1), but this theory does not guide clearly about \( c \). The time spent learning increases with \( n \), but the learn fraction \( \tau_n/n = cn^{-1/3} \) decreases.

One method, based on dynamic programming, is discussed in Section 3.2.1. The other method is a simple variant of that in Besbes and Zeevi (2012, Section 3.1); these are discussed in Section 3.2.2.

#### 3.2.1 Dynamic Programming after Learning: Method and Computation

The method is as follows.

1. **Learning.** Method A(\( \tau \)).

2. **Policy Determination.** Specify the control resolution \( n_C \), so \( \epsilon = (T - \tau)/n_C \) is the time between adjacent control times (Section 2.2.3). Time increment \( \delta \), integer \( n_S \) so that \( n_S \delta = T - \tau \), and upper limits \( \hat{y}_i, i = 0, 1, \ldots, n_S - 1 \) are specified later. At time \( \tau \), solve a DP similar to (5), modified as follows: (i) time only “covers” the exploitation interval \([\tau, T]\), and the capacity is \( C_{\text{rem}} = C - S_L \), where \( S_L \) is the number of sales during learning; (ii) a state \((i, z, y)\) represents that time is \( \tau + i\delta \), capacity is \( z \), and the arrival count since \( \tau \) (that is, during \([\tau, \tau + i\delta]\)) is \( y \); (iii) with \( A = \sum_{j=1}^k A_j \) the arrival count during learning (Method A), the conditional arrival rate at a state \((i, \cdot, y)\)
is 

\[ g(i\delta, y) = (A + y)/(\tau + i\delta) \]  
(maximum likelihood, Section 2.1); and (iv) purchase probabilities in (5) are replaced by estimates \( \hat{g}_t \) from Method A. The DP solver outputs upper limits \( \hat{y}_t \) and price indices \( j_{t,z,y}^* \) (in \( \{1, 2, \ldots, k\} \)) for \( t \in \{1, \ldots, n_C\}, \) \( y \in \{0, 1, \ldots, \hat{y}_t\}, \) \( z \in \{0, 1, \ldots, C_{\text{rem}}\}, \) where time index \( t \) represents (control) time \( \tau + (t - 1)\epsilon. \) (Each \( \hat{y} \) is an appropriate \( \hat{y}_t. \)) For convenience, policy determination is fully detailed as Algorithm 1.

3. Exploitation. For each \( t = 1 \) to \( n_C: \) at time \( \tau + (t - 1)\epsilon, \) observe the capacity (inventory level) \( z, \) and let \( A \) be the arrival count since time \( \tau; \) set the price as \( p_{j_{t,z,y}} \), where \( y = \min(A, \hat{y}_t) \), until time \( \tau + i\epsilon. \) We use \( n_C = 20 \) in all numerical results later, but none of our discussion requires this.

The policy determination step requires upper limits \( \hat{y}_i, i = 0, 1, \ldots, n_S - 1, \) and time increment \( \delta \) so that the transition probabilities in (5) are small enough. Our selection is based on the following.

Lemma 2 Let \( A \) be arrival count observed during the learning interval \([0, \tau].\) Put 

\[ g(i\delta, y) = (A + y)/(\tau + i\delta). \]  

If \( \hat{y}_i = i \) for \( i = 0, 1, \ldots, n_S - 1, \) and if \( \delta \) satisfies

\[ \frac{\delta A}{\tau} < 1, \]  

then

\[ \max_{0 \leq i \leq n_S - 1} \max_{0 \leq y \leq \hat{y}_i} g(i\delta, y) \delta = \frac{A + n_s - 1}{\tau + (n_S - 1)\delta} < 1. \]

Proof. For the \( g \) as specified, the inner maximum occurs at \( y = \hat{y}_i = i, \) so it is \( (A+i)\delta/(\tau+i\delta). \) Then, the maximum of the latter occurs at \( i = n_S - 1 \) (as the derivative with respect to \( i \) is positive when (14) holds), and its value is \( (A + n_s - 1)\delta/(\tau + (n_S - 1)\delta); \) this proves the equality. The inequality then follows from (14).

In view of (14), we set \( \delta = (2\lambda)^{-1}, \) where \( \lambda = \hat{\lambda} + 2\sqrt{\hat{\lambda}/\tau} \) is an (approximate) 95%-confidence upper bound on \( \lambda. \)

Computing and storage requirements. The work of Algorithm 1 is linear in the number of states, which is \( \sum_{t=0}^{n_S-1} C_{\text{rem}}(\hat{y}_t + 1) = C_{\text{rem}}n_S(n_S + 1)/2. \) Exploitation requires storage of price indices for each state corresponding to the control resolution; the number of such states is \( \sum_{t=1}^{n_C} C_{\text{rem}}(\hat{y}_t + 1) = C_{\text{rem}}\sum_{t=1}^{n_C} [(t - 1)(n_S/n_C) + 1] = C_{\text{rem}}[n_S(n_C - 1)/2 + n_C]. \) These may become onerous as the system size increases. To quantify this, consider sizing via \( n \) in the usual way: \( C_n = nC, \) \( T_n = n, \) \( \tau_n = cn^{2/3}, \) \( n \to \infty, \) but keep \( n_C \) fixed. Then, with \( a_n \sim b_n \) meaning \( a_n/b_n \to 1, \) we have: \( T_n - \tau_n \sim n, \) \( \hat{\lambda} = \hat{\lambda}_n \to \lambda, \) and \( n_S = n_{S,n} \sim 2n\lambda; \) and
we expect $C_{\text{rem}} \sim nC$ (since $S_L/n = S_{L,n}/n \to 0$ from Besbes and Zeevi (2012, Theorem 1)). Then, the number of states is $\sim 2C\lambda^2 n^3$, and the number of price indices to be stored is $\sim (n_C - 1)C\lambda n^2$ when $n_C > 1$, or $C_{\text{rem}} n_C \sim nC$ when $n_C = 1$. In summary: computation work (policy determination) is cubic in $n$; storage necessary for exploitation is quadratic in $n$ and linear in $n_C - 1$ when $n_C > 1$, or linear in $n$ when $n_C = 1$.

Algorithm 1: SolveDynamicProgram

Input: sell horizon $T$; prices $(p_i)_{i=1}^k$; learning time $\tau < T$; arrival count during learning $A$; purchase-probability (estimates) $(\hat{q}_i)_{i=1}^k$; remaining capacity $C_{\text{rem}}$; number of control points $n_C$; Notation: $C = \{1, 2, \ldots, C_{\text{rem}}\}$

Output: Upper limits $\bar{y}_i$ and price indices $\bar{y}_{i,z,y}$, $t \in \{1, \ldots, n_C\}, y \in \{0, 1, \ldots, \bar{y}_1\}, z \in C$

Data: Constant $\hat{q} \leftarrow 2$; Constant $\alpha \leftarrow 2$

1. $\lambda \leftarrow A/\tau$; $\bar{\lambda} \leftarrow \lambda + \alpha \sqrt{\lambda/\tau}$ /* estimate and 95%-confidence bound on $\lambda$ */
2. $\delta \leftarrow (\hat{q} \bar{\lambda})^{-1}$ /* $(\lambda)$-time increment satisfying (14) */
3. $n_S \leftarrow (T - \tau)/\delta$; $n_S \leftarrow \lfloor n_S/n_C \rfloor n_C$ /* solution conforming to $n_C$ */
4. $\delta \leftarrow (T - \tau)/n_S$ /* time increment */
5. for $i \leftarrow n_S$ to 0 do
6.   $\bar{y}_i \leftarrow i$ /* upper limits */
7.   $\bar{J}_{i,0,y} \leftarrow 0, y \in \{0, \ldots, \bar{y}_i\}$ /* capacity-boundary values */
8. end
9. $J_{n_S,z,y} \leftarrow 0, z \in C, y \in \{0, 1, \ldots, n_S\}$ /* time-boundary values */
10. for $i \leftarrow n_S - 1$ to 0 do
11.   for $y \leftarrow 0$ to $\bar{y}_i$ do
12.     $g_{i,y} \leftarrow (A + y)/(\tau + i\delta)$ /* conditional arrival rate */
13.     $\Delta_z J \leftarrow J_{i+1,z,y+1} - J_{i+1,z-1,y+1}, z \in C$
14.     $J_{i,z,y} \leftarrow g_{i,y}\delta[\max_j (\hat{q}_j(p_j - \Delta_z J)) + J_{i+1,z,y+1}] + (1 - g_{i,y}\delta)J_{i+1,z,y}, z \in C$
15.     $i_{i,z,y} \leftarrow \arg\max_j (\hat{q}_j(p_j - \Delta_z J)), z \in C$
16.     if $\mod (i,n_S/n_C) = 0$ then
17.       $t \leftarrow (i/(n_S/n_C)) + 1$
18.       $\bar{y}_i \leftarrow \bar{y}_i$ /* upper limit at control point $t$ */
19.     end
20.     $\bar{J}_{i,z,y} \leftarrow i_{i,z,y}, z \in C$ /* actions at control point $t$ */
21. end
22. end

3.2.2 Other Exploitation Methods

The method of Besbes and Zeevi (2012, Section 3.1) (method BZ) is as follows:

1. Learning. Identical to Method A ($\tau$), except that the estimator $\theta_S = (\hat{q}_i)_{i=1}^k$ in (13) is obtained, and no observation of arrivals is made.
2. **Policy Determination and Exploitation.** At time $\tau$, let $(t_1, \ldots, t_k)$ be the solution to the linear program (LP)

$$\max \left\{ \sum_{i=1}^{k} (\hat{\lambda} q_i t_i) p_i : \sum_{i=1}^{k} t_i \leq T - \tau, \sum_{i=1}^{k} \hat{\lambda} q_i t_i \leq C - S_L, t_i \geq 0 \right\}$$

(15)

where $S_L$ accounts for the amount of capacity sold during learning, and $S_L = n^{2/3} (\log n)^{1/2}$ for the size-$n$ problem (based on large-$n$ analysis). An alternative $S_L$ would be the sale count up to time $\tau$, but we follow the authors’ original method. The exploitation policy, as it applies to our non-network setting, is as follows. Any LP basic optimal solution has either one or two positive $t$’s; in the former case, apply, until stock-out or expiration, the price that corresponds to the unique $t$ that is positive; in the latter case, let $i$ and $j$ be the indices of those $t$’s that are positive; apply, until stock-out or expiration, price $p_i$ for a length of time $t_i$ only, and price $p_j$ at other times.

A simple alternative (a “Linear Program with Arrivals observed”, or LPA) is as follows:

1. **Learning.** Method A ($\tau$) produces estimator $\theta_A$, following (7).

2. **Policy Determination and Exploitation.** As in Method BZ, except that in the LP (15) the estimator $\theta_S = (\hat{\lambda} q_i)_{i=1}^{k}$ is replaced by $\theta_A$.

### 3.2.3 Experimental Comparison

Our examples are adapted from Besbes and Zeevi (2009, Section 6.2). The allowable prices are the midpoints of the division of $[0, 10]$ into $k$ equal intervals. Underlying (unknown to the seller), the arrival rate $\lambda$ and purchase probabilities are one of: (i) $\lambda = 30$ and Uniform$(0,10)$ reservation prices ($q_i = 1 - p_i/10$); or (ii) $\lambda = 10 \exp(1)$ and mean-1 Exponential reservation prices ($q_i = \exp(-p_i)$). We refer to (i) and (ii) as the linear (L) and exponential (E) demand function, respectively.

We consider only $n \in \{8, 16, 32, 64\}$, and a number $k$ of prices between 3 and 5. We set $c = 10^{-1/3} \approx 0.464$, so $\tau_n$ increases from 1.856 to 7.424, while the learn fraction decreases from 0.232 to 0.116. As explained earlier, the DP policy-determination work is a cubic function of $n$, and for this reason larger problems could not be solved quickly enough to give meaningful results. Our largest problem has $C = 20$, $k = 5$, and $n = 64$ (capacity is $C_{64} = 1280$); here, the number of states averaged 8 billion (8$x$10$^9$), and policy determination took 23 minutes on average (MATLAB implementation on an Intel Core i5 vPro processor).

The optimum revenue, $J^*$, is determined by a simplified version of (5): state component $y$ is removed; the arrival rate $\lambda$ replaces the conditional rates $g()$; and the maximized term has exact purchase probabilities, $\max_{1 \leq i \leq k} q_i (p_i - \Delta_j J)$. Each method is simulated many
times, and its expected revenue \( J \) is estimated by averaging (if no sales are made during learning, which is not rare for smaller \( n \), zero revenue is credited, which is conservative). The loss (regret) \( 1 - J/J^* \) captures revenue performance relative to the unknown optimum; smaller loss indicates a better method. For methods BZ and LPA, our loss estimates are very accurate (standard error is at most \( 2.6 \times 10^{-3} \) and is not reported). For method DP, the number of replications is smaller; the standard error is somewhat larger, and is reported.

Table 1 contains results. The main performance measure is the estimated loss for each of the three methods. All methods perform well in the sense that loss decreases as the problem size increases. Loss differences between methods are typically a significant number of DP standard errors. Typically, DP leads the other two by a significant margin, while LPA leads BZ by a smaller margin. Another significant observation is about the estimation errors \( v(\theta_S) \) and \( v(\theta_A) \): the latter is consistently lower. In principle, \( \theta_A \) and \( \theta_S \) are biased estimators of the demand function, and our SMSE formulas (Section 3.1) are also biased, because stock-outs during learning are possible; but all the biases are negligible if such stock-outs are rare, and this is the case in all our examples (stock-out frequency rounded to three decimals is zero); a learn fraction that is modest is helpful in this sense. The “biased” SMSEs we report are more informative (accurate) than the unbiased average SMSEs that we could compute because the latter have significant sampling error.

In experiments where the learning time was much shorter or much longer, performance tended to suffer, as we expect from Besbes and Zeevi (2009). Intuitively, in the former case, exploitation suffers due to inaccurate estimates; in the latter case, performance suffers due to suboptimal pricing during too long a period.

4 Conclusion

This paper studies two types of (non-network) revenue management problems. In the first type, all model parameters are known to the seller, except the arrival rate. A dynamic program of the pricing problem is proposed, where learning of the arrival rate, via the Bayesian-posterior-mean or the maximum-likelihood estimator, is encoded in the state variable. The discrete-time DP is not particularly amenable to give structural insights, but it is a basis for computing good policies. We observed effects that have emerged in related literature: the price increases in the seller’s arrival-rate uncertainty, captured in a prior; and the price is sensitive to the prior, even when the prior’s mean and variance are fixed.

In the second type (Section 3), all model parameters are unknown to the seller, and his only knowledge is the validity of Assumption 1, that individual arrival times and Bernoulli purchase indicators are all independent. In this setting, we proposed estimators of the arrival
<table>
<thead>
<tr>
<th>Dmd</th>
<th>k</th>
<th>C</th>
<th>n</th>
<th>Estimated Loss</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(v(\theta_S))</td>
<td>(v(\theta_A))</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>72.7</td>
<td>42.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>45.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>28.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>18.2</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>20</td>
<td>8</td>
<td>0.141</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>0.047</td>
</tr>
<tr>
<td>L</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>202.0</td>
<td>101.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>127.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>50.5</td>
</tr>
<tr>
<td>L</td>
<td>5</td>
<td>20</td>
<td>8</td>
<td>0.158</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>0.095</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>8.60</td>
<td>7.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>2.15</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>0.226</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>0.076</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>18.28</td>
<td>15.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>11.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>4.57</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>31.14</td>
<td>25.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>19.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>12.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>7.79</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>20</td>
<td>8</td>
<td>0.315</td>
<td>0.294</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>0.126</td>
</tr>
</tbody>
</table>
rate of customers, and the customers’ price response, from arrival and sale counts, and noted that together they yield a novel estimator of the demand function. Under the assumption of no-stock outs, we derived the estimator’s bias and mean squared error explicitly, and we observed that the new estimator was more accurate than the natural sales-data-only estimator in all our examples. It appears that this property can be shown to hold under mild conditions. We then proposed revenue management methods of the learning/exploitation type and studied them experimentally. The experimental results suggest that learning (estimation), followed by exploitation, is a promising approach. The burden of parameter estimation is contained within the method we proposed. Particularly encouraging are the performance gains, especially of the dynamic programming approach, over the benchmark that exploits only sale counts. The superiority of the DP method over the linear programming approach that is based on the same information (LPA) suggests that its more reactive nature is an advantage, provided the underlying dynamics are estimated to some sensible accuracy in a learning phase. We caution that our approach hinges on the assumption of independence of customers’ price response from past prices; significant applications may expose this assumption as problematic.

Appendix

Proof of Proposition 3. Write \( E[X] \) for \( E[X] \) and \( E^2[X] \) for \( (E[X])^2 \). Put \( X = \hat{\lambda}q_i, \mu = \lambda q_i, B := (EX - \mu)/\mu \). We have \( E(X - \mu)^2 = E[X^2] - \mu^2(1 + 2B), E[X^2] = E[E[X^2|A]], \) and \( E[X^2|A] = \text{Var}(X|A) + E^2[X|A], \) which give

\[
E[(\hat{\lambda}q_i - \lambda q_i)^2] = E[\text{Var}(\hat{\lambda}q_i|A)] + E[E^2[\hat{\lambda}q_i|A]] - (\lambda q_i)^2(1 + 2B). \tag{16}
\]

Assumption 1 states the independence of the \( A \)'s, and moreover, that the \( S_j \) are, conditionally on \( A \), sums of independent Bernoulli random variables. Under the no-stock-out condition, we additionally have \( A_j \sim \text{Poisson}(\lambda \tau/k) \) and \( S_j \sim \text{Binomial}(A_j, q_j) \) for each \( j \). The conditional binomial law of \( S_i \) yields

\[
E[\hat{\lambda}q_i|A] = \hat{\lambda}E[q_i|A] = \hat{\lambda}q_i1_{A_i>0}, \tag{17}
\]

\[
\text{Var}(\hat{\lambda}q_i|A) = \hat{\lambda}^2\text{Var}(q_i|A) = \hat{\lambda}^2A_i^{-1}q_i(1 - q_i)1_{A_i>0}. \tag{18}
\]

Writing \( E[\hat{\lambda}q_i] = E[E[\hat{\lambda}q_i|A]] \), and taking the outer expectation (of the right side of (17)) where the independence of the \( A \)'s is used, the bias result (9) is obtained. Turning to (10), note that (18) gives the conditional variance as \( q_i(1 - q_i) \) times \( \hat{\lambda}^2A_i^{-1}1_{A_i>0}; \) here, write \( \hat{\lambda}^2 = \tau^{-2}[(\sum_{j\neq i}A_j)^2 + 2A_i\sum_{j\neq i}A_j + A_i^2] \) and take expectation:

\[
E[\hat{\lambda}^2A_i^{-1}1_{A_i>0}] = \tau^{-2} \left[ \lambda \tau\frac{k - 1}{k} \left( 1 + \lambda \tau\frac{k - 1}{k} \right) h \left( \frac{\lambda \tau}{k} \right) + 2\lambda \tau\frac{k - 1}{k} \rho + \lambda \tau \right], \tag{19}
\]

19
by the independence of the A’s. This proves (10), and (11) follows by definition of \( c_0 \) there. The MSE result (12) will follow from (16), provided that \( \mathbb{E}\left[\mathbb{E}^2[\hat{\lambda^2}|A]\right] - (\lambda q_i)^2(1 + 2B) \) (with \( B \) in (9)) is identified with the second term in the right side of (12); the outer expectation above is the expectation of \( q_i^2 \) times \( \hat{\lambda}^2 1_{A_i > 0} \), by (17); write \( \hat{\lambda}^2 \) as above, and take expectation:

\[
\mathbb{E}[\hat{\lambda^2}1_{A_i > 0}] = \tau^{-2}\left[ \rho \lambda \tau \frac{k-1}{k} \left( 1 + \lambda \tau \frac{k-1}{k} \right) + 2\lambda \tau \frac{k-1}{k} \lambda \tau \frac{k-1}{k} + \frac{\lambda \tau}{k} \left( 1 + \frac{\lambda \tau}{k} \right) \right], \tag{20}
\]

using again the independence of the A’s. Elementary operations complete the identification, and thus prove (12).

References


