RT 1: Retail Analytics: Estimating Censored Demand

Tessa Wilkie

February 2020

Imagine a successful burger van that caters at rugby matches. The owners think that they are doing very well because they always sell out of burgers. It seems rugby fans cannot get enough of these burgers. But as the van begins to run out of burgers too often, the owners worry that burger eaters will go to other places to get food.

The owners decide to run another burger van and double the quantities they provide. At the first outing it is a bit of a disaster. They only sell about a 10th of the burgers on offer at their second van and have to throw away a vast amount of burger patties, buns and toppings.

The problem the burger van owners ran into is estimating unobserved — or censored — demand: the subject of this report. It is a big problem for retailers in general, especially those that stock perishable goods.

Deciding how much stock to buy is a major issue for retailers. Order too much and you have to throw some away, losing profits and generating waste. Order too little and you have unsatisfied customers, who may go to another retailer.

To decide how much stock to buy, retailers have to forecast demand. This becomes particularly difficult if they have sales periods where they sell out of stock. That means there is some demand that they have not satisfied. Crucially, from an operational research perspective, they do not know how much potential demand they have not satisfied: it is unobserved.

The issue with unobserved demand is that it is difficult to estimate how much demand there would have been if the retailer did not run out of stock.

This is more of a problem for retailers that stock perishable goods, because have to throw away unsold items after a short period. They want to maximise profits, so they want to make sure they have enough stock in place to satisfy demand (or a large proportion of it). But, they need to avoid ordering so many items that they have to throw a lot away because there is not enough demand to buy it while it is still in saleable condition.

Getting demand forecasting right is important, not just from the retailer's perspective, but on a global scale. Waste contributes to pollution. Overproduction is linked to climate change.

This report examines some techniques for retailers to estimate unobserved demand: known as a right-censored demand problem since the demand distribution is right-censored (we do not observe some of the right hand side of the distribution).

The report considers in detail parametric techniques — that is, those that assume an underlying statistical distribution of demand — and use that to estimate the unobserved demand.

A simulation study shows that the parametric techniques considered a good at estimating distributions from simulated data with unobserved demand. But, we consider a situation where the demand is not from an obvious parametric distribution, and show that they can run into trouble here.

1 Introduction

Shops and retailers have to decide how much stock they should order in and when. To do this they need to forecast demand for their products. This is not a straightforward problem.
One major hurdle is that if a retailer does not have enough stock to satisfy all of the demand, they never observe it all: they cannot know how much demand was out there.
This report will focus on techniques to tackle this problem: how to estimate the demand you never see. This is known as a censored demand problem.
This is a particularly pertinent problem for stocking perishable goods. One way a retailer can make sure they never run out of stock is to order far more than they imagine they could need. They could fill their warehouse.
With some products, this may not be an issue (apart from the squeeze on storage space). If a shop owner over-orders packets of toilet paper or tinned goods and doesn't sell them all in one day, the shop owner can simply keep them and hope to sell them all over the following days. For a grocer, over-ordering is a much more painful issue. They will have to throw away fruits and vegetables that are unsold after a certain amount of time because they will go off.
This is, of course, a broader problem than for just the shop owner. Overconsumption and waste are global issues, contributing to climate change, depletion of natural resources and pollution.

2 The Newsvendor Model

This report will discuss techniques to estimate censored demand in the context of the newsvendor model. The newsvendor model is a simple model used to motivate problems around inventory levels and demand forecasting. It is used to model situations where a retailer (in this case, the newsvendor) needs to order in enough stock to satisfy demand and maximise their profits. There is the added issue that the stock is perishable. In the newsvendor's case, if they do not sell all their newspapers in one day, they will be worthless the next — nobody (except, perhaps, an historian) wants to read yesterday's news.
That said, if the newsvendor does not order in enough stock they will have to turn customers away. They will not maximise their profits. They may end up losing customers over the long term if people regularly come to buy newspapers and find that there are none to buy.
This also causes problems for demand forecasting as there is unobserved demand: the people who might have bought a paper if the newsvendor had any left in stock. The newsvendor cannot see how many people might have bought if he or she had any in stock. So how can the newsvendor decide how many papers to buy in if they cannot properly estimate how much demand is out there?

2.1 Definitions

We have observed and recorded sales of newspapers for n days. Let X_t be the demand for newspapers on day t where $t = 1, ..., n$. We might record a sample of sales $x_1, ..., x_n$ on days $1, ..., n$.
Let S_t be the number of newspapers in stock on day t.
The newsvendor has to decide on the evening before how many newspapers to order to offer for sale on the following morning. Lead times — that is, how long after ordering newspapers they arrive — are assumed to be zero. Stockouts, where there is no stock left to satisfy demand, and hence unobserved demand, occur when $X_t \geq S_t$. When there is a stockout, there is an element of demand that is unobserved.

Newspaper sales observed on a particular day will either fall into set A, where there is no stock-out, or set B where inventory exceeds demand and so there is censored demand on that day. The censored demand in this problem is always right-censored. That is, there is unobserved demand (if there is any) *after* a certain point in time on that day. We do not consider the case of stock being replenished intra-day and so the possibility of several periods where censored demand may occur.
The different techniques available to estimate censored demand either make a parametric assumption about the level of demand, X_t on day t, or use observed data to make non-parametric estimations. The parametric methods will estimate the true mean, μ, of the distribution of demand and the true variance, σ.
The newsvendor is concerned with with underage and overage costs: the cost to their profit if they don't have enough newspapers in stock (underage), or if they stock too many (overage). If a newsvendor buys each newspaper from their supplier for $50p$, and they can sell the newspaper for \$1, then the cost of underage is $50p$ per unit. If they have excess stock then they can sell them back to the supplier for $25p$ at the end of the day, then the cost of overage is $25p$ per unit.
Conrad (1976) states that the newsvendor can decide how many copies to order by following the following formula: order a copies over b copies, where $a < b$ when:

$$F(a) > \frac{C_u}{C_o + C_u}$$

$F()$ is the cumulative distribution function for demand, C_u denotes underage cost and C_o, overage cost.
For simplicity, this report will consider the problem where S_t is constant, and will be referred to as S.

3 Overview of methods

Parametric methods assume that there is an underlying statistical distribution to the demand. Knowledge of the distribution allows the forecaster to make inferences about the demand that they cannot see. The effectiveness of these techniques will depend heavily on how well the underlying data approximates to the assumed distribution. In this report we consider a method proposed by Conrad (1976) for estimating demand that accords to a Poisson distribution. Nahmias (1994) proposes three techniques to estimate demand for a censored normal distribution: we give a description of one of these methods below.
Agrawal and Smith (1996) recommend using the negative binomial distribution to estimate demand. Their reasons include it being a flexible distribution able to model demand for items that sell in high numbers each time period, as well as items that sell in few numbers each day. It is also more flexible than the Poisson distribution because its mean to variance relationship can be greater than one.
Berk et al. (2007) are among the researchers to consider a Bayesian approach to updating the predictive demand distribution. They propose a two-moment approximation of the posterior distribution.
Lau and Lau (1996) consider a non-parametric method to estimate the demand distribution where some observations are censored. They use a technique first proposed by Kaplan and Meier (1958), coupled with using observed hourly sales to estimate the unobserved demand on the right hand side of the distribution, and fitting a Tocher curve to estimate the whole distribution.
Sachs and Minner (2014) compares a selection of the above techniques and uses the Lau and Lau (1996) method, incorporated with a linear programming technique to allow other influencing fac-

tors on demand into the model. In this report we consider more closely the methods proposed by Conrad (1976) and Nahmias (1994).

3.1 Parametric methods

3.1.1 Normal distribution

Nahmias (1994) proposes simplified estimators for normally distributed data with right-censored demand.

Taking a sample of newspaper sales $x_1, ..., x_n$ on n days. The underlying distribution of demand (or of sales if we imagine unlimited stock of newspapers) is assumed normal with mean μ and variance σ^2 unknown. On r of the n days, demand is lower than S, the number of newspapers in stock. So on $n - r$ days demand exceeds or equals stock and there is potentially unobserved demand.

Nahmias (1994) defines: $p = \frac{r}{n}$, $z = \Phi^{-1}(p)$, $\bar{x}_r = \frac{1}{r}\sum_{i=1}^{r} x_i$, and sample variance $s_r^2 = \frac{1}{(r-1)}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$.

Then he states a large sample estimator for the mean is:

$$\hat{\mu} = \bar{x}_r + \frac{\hat{\sigma}\phi(z)}{p},$$

and a large sample estimator for the variance is:

$$\hat{\sigma}^2 = \frac{s_r^2}{(1 - z\phi(z)/p - [\phi(z)/p]^2}.$$

The $\phi(z)$ denotes the probability density function of the standardised normal distribution evaluated at z, and Φ, the cumulative density function of the standardised normal.

Nahmias (1994) notes that these large sample estimators sacrifice a little in accuracy over, for example, maximum likelihood estimators. But prefers them for their computational simplicity. In a Monte Carlo simulation, setting n at $10, 40, 70$ and 100, Nahmias (1994) found that only when $n = 10$ and S was low (so there was a large amount of unobserved demand), was there a "substantial difference" between the large sample estimators and the more accurate, but computationally more burdensome, Maximum Likelihood Estimator method.

3.1.2 Poisson distribution

Conrad (1976) finds estimators for the situation where underlying demand is assumed to be distributed according to a Poisson distribution with rate parameter λ:

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!},$$

where $x = 0, 1, 2, ...$ and $\lambda > 0$. Conrad (1976) states that the assumption of sales occurring as a Poisson process is appropriate if we assume customers arrive randomly through the time that the newsvendor is open.

Defining, for $S \geq 0$:

$$P(S; \lambda) = \sum_{x=0}^{S} \frac{\lambda^x e^{-}\lambda}{x!}.$$

Then $\hat{\lambda}$, the estimated value for the true parameter λ can be found by solving:

$$\left(\sum_{i=1}^{r} x_i - mn\right)[1 - P(S-1; \hat{\lambda})] + \hat{\lambda}(n-r)[1 - P(S-2; \hat{\lambda})] = 0.$$

This equation has to be solved numerically. To make this easier, by cutting down the size of the space to be searched, Conrad (1976) shows that $\lambda > \frac{\sum_{i=1}^{r} x_i}{r}$. Conrad (1976) says that this method is easily adaptable to situations, where, for example, demand is different on different days of the week.
For example, a newsvendor selling the *Financial Times* may see different demand at the weekend where workers in the financial sector do not feel the need to read about financial news before getting into the office in the morning.
In this situation, Conrad (1976) advises that one can pick observations for a particular day of the week at random and have a Poisson model of demand for this set. One issue in this is that it may take a long time to get a large enough sample. We would need almost two years to get data for 100 Sundays.

4 Simulation Study

To illustrate two of the parametric methods discussed above, this report includes a short simulation study. concerning the assumption of normally distributed demand Nahmias (1994) and demand approximating to a Poisson density Conrad (1976).
It will show how the methods perform in the ideal case when the underlying distribution of demand matches those assumed by the methods.
It will look at how well the methods perform depending on how early the sample is truncated.
It will also explore a situation where the assumed demand does not quite match the simulated sample.

4.1 Normal Demand Assumed

We imagine a newsvendor who observes his sales on 100 days, $n = 100$. For the first part of the simulation we assume the underlying demand is normally distributed with mean 100 and variance, 100. We discard any negative observations. We also round simulated normal variables to the nearest integer value, to represent whole units to order.
We set S, the number of newspapers our newspaper seller has in stock, so that: in case A — S matches the 75th quantile of the normal distribution (rounded to the nearest integer); Case B — S matches the 50th quantile of the normal distribution; Case C — S matches the 25th quantile of the normal distribution.
We calculate the estimated mean and variance according to Nahmias (1994)'s approximate method. We replicate each 1000 times and take the average. We then compare this to the mean and variance of the true distribution.

Case	Estimated Mean	Estimated Variance	S
A	101.4654	99.9988	107
B	100.0309	102.4257	100
C	102.4347	99.9062	93

Table 1: Results for normally distributed demand

We can see that the process is reasonably accurate, even for a very low level of observed sales data.

But if we don't get the right underlying distribution (which is entirely possible in the real world), we can get less accurate results.
If we repeat the process, using the Nahmias (1994) method that assumes normally distributed demand, but this time where the underlying demand is actually distributed according to a Poisson distribution with parameter $\lambda = 100$, then we retain very accurate estimation of the mean, but the variance is underestimated in cases where we have more highly censored data.

Case	Estimated Mean	Estimated Variance	S
A	99.8641	96.0895	107
B	99.7142	93.7973	100
C	99.2857	90.0670	93

Table 2: Results for analysis of Poisson distributed demand that is assumed to be normally distributed

4.2 Poisson Distributed Demand

Secondly, we assume Poisson demand and examine the procedure proposed by Conrad (1976). We simulate 100 days of demand with counts from a Poisson distribution ($n = 100$) with parameter $\lambda = 10$. We replicate 1000 times.
The results are presented below. As before, in case A we set the stock level, S, at the 75th quantile of the Poisson distribution, Case B at the 50th and Case C at the 25th.

Case	$\hat{\lambda}$	S
A	9.9983	12
B	10.0090	10
C	10.0310	8

Table 3: Results for Poisson-distributed demand

We can see in Table 3 that the Conrad (1976) method is also impressively accurate. Accuracy tails off slighty when S is small.

4.2.1 A Bimodal Example, or How to Break the System

A major problem with using parametric methods is that our newsvendor might not know what distribution the demand belongs to, and may mis-specify it. The demand may also not accord to any known distribution.
We saw in Table 1 that the Nahmias (1994) method was very accurate when estimating the distribution of normally distributed data that had been right-censored. However, if we attempt to apply the procedure to non-normally distributed data, the wheels start to fall off.
To illustrate this, we simulate a bimodal distribution, created taking samples from two normal distributions. The first 50 are sampled from a Normal distribution with mean 100 and standard deviation, 5. The second 50 from a Normal distribution with mean 125 and standard deviation, 2. With bimodal data, we might assume that the newsvendor never orders enough to see many sales

generated from the second peak of demand. Here we assume that they have $S = 120$ newspapers in stock, so if they sell 120 newspapers on a day then there is unobserved demand.

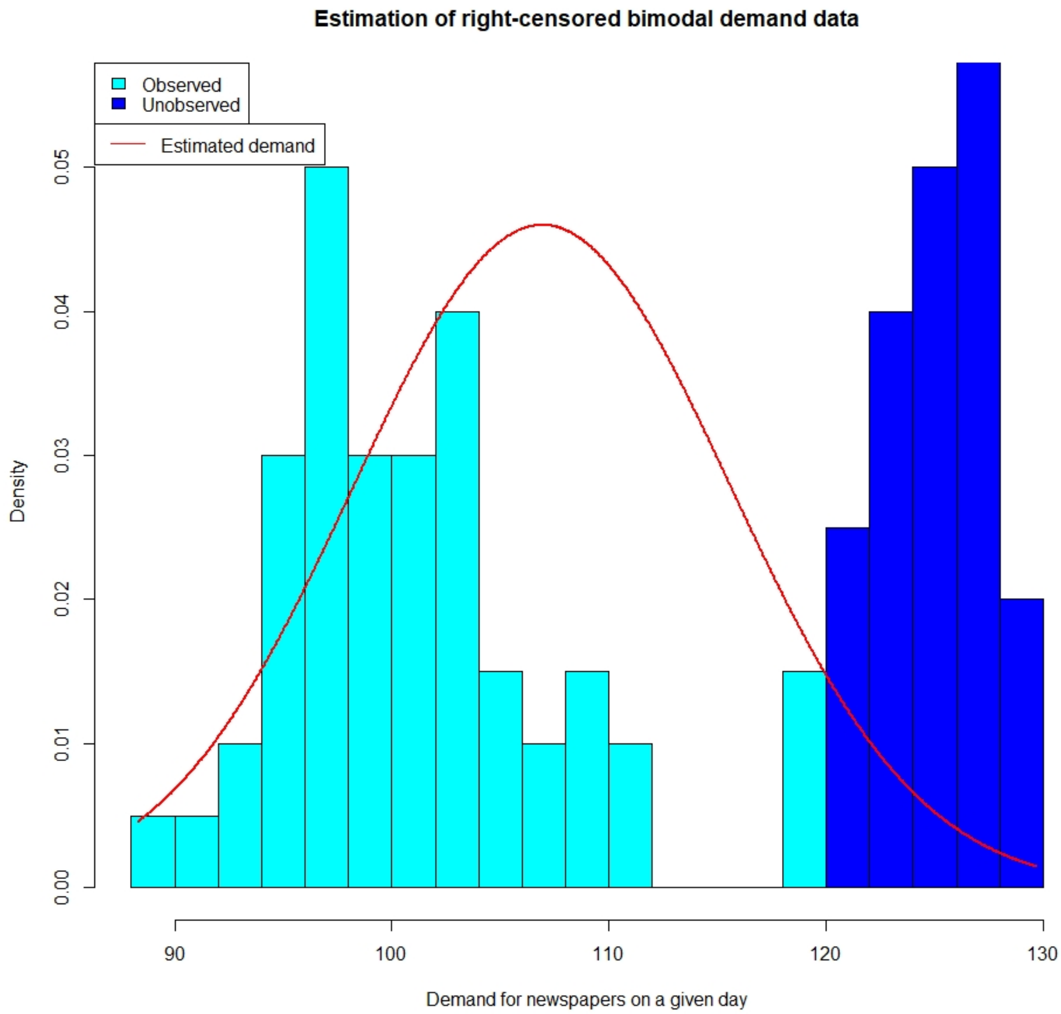

Figure 1: A bimodal example

Using the method of Nahmias (1994), we estimate that the demand fits a normal distribution with mean approximately 107 and variance approximately 75. The density is plotted with a red line in Figure 1. As we can see, the method misses a large peak in demand. A newsvendor using this to work out unobserved demand would grossly underestimate the amount of newspapers he should keep in stock. This illustration is, of course, an extreme case. But not all that an unlikely one. To avoid this working with non-parametric methods, and hourly demand data, is one option (see Sachs and Minner (2014) and Lau and Lau (1996)).

5 Conclusions and Further Work

The success of parametric techniques depend heavily on how well the model selected fits the actual demand. One issue that could be run into is where daily demand follows a bimodal distribution.

Our newsvendor might see a peak in sales in the morning and evening rush hours, as commuters pass his stand and pick up a paper to read on the way into, or home from, the office. The estimate of unseen demand might vary greatly if he runs out of copies at 12pm, or at 9pm. For more discussion on this, see Jain et al. (2015).

Sachs and Minner (2014) note that further work could include considering models where substitution — swapping an out of stock product for a similar one that is in stock — occurs.

References

Agrawal, N. and Smith, S. A. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics (NRL)*, 43(6):839–861.

Berk, E., Gürler, Ü., and Levine, R. A. (2007). Bayesian demand updating in the lost sales newsvendor problem: A two-moment approximation. *European Journal of Operational Research*, 182(1):256–281.

Conrad, S. (1976). Sales data and the estimation of demand. *Journal of the Operational Research Society*, 27(1):123–127.

Jain, A., Rudi, N., and Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, 63(1):134–150.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Lau, H.-S. and Lau, A. H.-L. (1996). Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research*, 92(2):254–265.

Nahmias, S. (1994). Demand estimation in lost sales inventory systems. *Naval Research Logistics (NRL)*, 41(6):739–757.

Sachs, A.-L. and Minner, S. (2014). The data-driven newsvendor with censored demand observations. *International Journal of Production Economics*, 149:28–36.