

# A Regression based Unconstraining Demand Method in Revenue Management

JaeJune Lee<sup>a,1</sup> · Woojoo Lee<sup>a</sup> · Junghwan Kim<sup>a</sup>

<sup>a</sup>Department of Statistics, Inha University

(Received February 3, 2015; Revised February 26, 2015; Accepted February 26, 2015)

---

## Abstract

Accurate demand forecasting is a crucial component in revenue management(RM). The booking data of departed flights is used to forecast the demand for future departing flights; however, some booking requests that were denied were omitted in the departed flights data. Denied booking requests can be interpreted as censored in statistics. Thus, unconstraining demand is an important issue to forecast the true demands of future flights. Several unconstraining methods have been introduced and a method based on expectation maximization is considered superior. In this study, we propose a new unconstraining method based on a regression model that can entertain such censored data. Through a simulation study, the performance of the proposed method was evaluated with two representative unconstraining methods widely used in RM.

Keywords: unconstraining, censored data, expectation maximization, revenue management

---

## 1. 서론

항공이나 호텔 등의 서비스 분야에서 상품은 특정일에 제공되는 좌석이나 객실이고, 대략 일년 전부터 예약을 받기 시작하여 공급(또는 사용)일에 서비스를 제공한다. 이러한 산업분야의 제품은 몇 가지 특징이 있는데, 첫째, 상품은 소멸적인(perishable) 특성이 있다. 예약기간 동안 유동적인 가격으로 판매되다가 공급일 이후는 그 가치가 소멸된다. 둘째, 총 공급량은 단기간에 변경될 수 없다(fixed capacity). 본 논문에서는 항공운송 분야에 초점을 두고 수요예측의 정확도를 높일 수 있는 방법을 제시하고자 한다.

대부분의 항공사는 수입(revenue) 증대를 위해 F(1등), C(비즈니스), Y(일반) 등 항공기 객실, 특히 Y-객실의 좌석을 각각 운임이 다른 다수의 가상 예약등급(Booking Class; BC)로 구분하여 운용하는데, 이를 가격등급(Fare Class; FC)라고도 한다. 구체적으로, 예약기간 중 시점 별로 각 BC의 최대 예약가능 좌석수(Booking Limit; BL)를 설정하여 예약을 받는데, 예약요청이 설정된 BL을 초과하면 예약을 거절한다. 대부분의 항공사는 예약기간 중 20~30개 출발 전 시점(Data Capture Point; DCP)을 설정하여 각 BC 별로 예약 좌석수, BL, 평균 가격 등을 자료로 저장하여 미래 출발편의 BC 별 수요를 예측한다. 예를 들어, 특정일 미래 출발편 K-가격등급의 최종수요(출발 시점의 예약수요)는 예약기간 동안 수요발생 패턴이 유사한(예로, 성·비수기, 요일 등 고려) 다수의 기 출발편 예약자료를 이용하여 예측된다. 실제, 항공사들은 미래 출발편의 각 DCP에서 BC 별 최종수요를 예측하고, 총 공급량을 예측된 최종수요와 BC 별 가격을 반영하여 수입이 최대화되도록 각 BC의 BL을 설정하여 좌석을 관리하는 것을 수입관리(Revenue Management; RM)라고 한다.

This research was supported by Inha University.

<sup>1</sup>Corresponding author: Department of Statistics, Inha University, Younghyun-Dong, Nam-Gu, Incheon 402-751, Korea. E-mail: [jjlee@inha.ac.kr](mailto:jjlee@inha.ac.kr)

RM의 가장 핵심적인 요인은 정확한 수요예측이라고 할 수 있는데, 시스템에 저장된 기 출발편 예약 데이터에는 일부 DCP에서 수요가 BL을 초과하여 일부 수요가 예약이 거부되고 따라서 관측된 예약수가 실제 수요보다 작은 경우가 포함될 수 있다. 통계학의 관점에서 이러한 데이터는 중도절단된(censored) 데이터라고 할 수 있는데, 이런 데이터로 미래 출발편의 수요를 예측할 경우 과소예측의 문제가 발생할 수 있다. RM의 관점에서 이론적으로 발생할 수 있는 수요를 참수요라고 하고, 그것을 추정하는 것을 수요복원(unconstraining demand)이라고 한다. 수요복원 결과는 미래 출발편의 수요예측의 정확도에 중요한 요인이 된다. 기존의 연구에서 RM을 통해 약 3~7%의 수입증대 효과를 기대할 수 있는 것으로 보고되고 있다 (Belobaba, 1987; Cross, 1997; Talluri와 Van Ryzin, 2005).

본 논문의 구성은 다음과 같다. 2절에서는 중도절단된 데이터의 대표적인 복원 방법으로 알려진 Pick Up(PU)과 Expectation Maximization(EM) 방법을 살펴본다. 3절에서는 회귀모형에 기반을 둔 새로운 중도절단 수요의 복원 방법을 제안하고, 4절에서는 새로운 방법의 성능을 모의실험을 통해 기존의 대표적인 복원 방법들과 비교한다. 마지막 5절에서는 연구의 결과를 요약하고 결론을 내린다.

## 2. 선행 연구의 고찰

### 2.1. 중도절단 데이터의 복원

본 논문에서는 논점을 단순화하기 위해 미래 특정일 출발편의 특정 BC에 대해 인접한 두 DCP 사이의 수요를 예측하는 문제를 다룬다. 예를 들어 인접한 두 DCP 시점으로 출발 14일 전을 DCP<sub>1</sub>, 출발 7일 전을 DCP<sub>2</sub>,  $n$ 회 기 출발편을  $j = 1, 2, \dots, n$ 라고 하고, 두 DCP에서 수집된  $n$ 개 기 출발편의 예약 좌석수와 최대 예약가능 좌석수를 각각  $(s_{1,j}, BL_{1,j}), (s_{2,j}, BL_{2,j})$ 라 표기한다. 주목할 것은  $s_{2,j}$ 의 일부는  $BL_{2,j}$ 에 의해 예약이 거부될 수 있는데, RM에서는 제한된(constrained) 수요라고 한다. 수요 복원이란 제한된 수요 자료로부터 DCP<sub>2</sub> 시점에서의 이론적 참수요(true demand)  $S_{2,j}$ 를 추정하는 것이다. 여기서  $s$ 와 BL은 고정된 수치로,  $S$ 는 확률변수로 간주한다. 마지막으로  $j$ 번째 기출발편에 대해 두 DCP 사이에 발생한 신규수요(new demand)에서  $s_{1,j}$ 의 DCP<sub>2</sub> 시점에서의 예약 취소분을 감한 양을 순수요(net demand)라고 하고  $Z_{2,j}$ 라고 표기하기로 한다. 그러면  $S_{2,j}$ 는 다음과 같이 표현된다.

$$S_{2,j} = s_{1,j} + Z_{2,j}. \quad (2.1)$$

따라서 DCP<sub>2</sub> 시점의 참수요  $S_{2,j}$ 의 추정 문제는  $s_{1,j}$ 가 상수이기 때문에 순수요  $Z_{2,j}$ 의 추정 문제와 같다.

### 2.2. 기존의 복원 방법

기존의 복원 방법은  $Z_{2,j} \sim i.i.d. N(\mu, \sigma^2), j = 1, 2, \dots, n$ 을 가정하고,  $s_{1,j}$ 와 무관하게  $Z_{2,j}$ 를 추정한다. 만일  $s_{2,j} < BL_{2,j}$ 이면 두 DCP 시점 사이의 모든 예약요청이 허용되어  $S_{2,j} = s_{2,j}$ 가 만족되고, 따라서  $S_{2,j} = s_{2,j} = s_{1,j} + Z_{2,j}$ 가 성립된다. 반면  $s_{2,j} \geq BL_{2,j}$ 이면,  $s_{2,j} - s_{1,j}$ 는 일부 수요가 예약이 거부된 우측 중도절단된 수요로서  $S_{2,j} \geq s_{2,j}$ , 즉,  $Z_{2,j} \geq s_{2,j} - s_{1,j}$ 가 만족된다. 이런 배경에서 미래 출발편의 수요예측에서는 다음의 과정으로  $Z_{2,j}$ 를 예측한다.

단계 1(복원) :  $s_{2,j} < BL_{2,j}$ 이면  $s_{2,j} - s_{1,j} = Z_{2,j}$ ,  $s_{2,j} \geq BL_{2,j}$ 이면  $Z_{2,j}$ 를 추정.

단계 2(예측) : 단계 1의  $Z_{2,j}$ 들로 미래 출발편의  $Z$ 를 예측.

지금까지 다양한  $Z_{2,j}$ 의 복원 방법이 제안되었는데, PU와 EM이 대표적인 방법으로 알려져왔다. Weatherford와 Pöhl (2002)는 모의실험을 통해 EM이 PU 보다 더 우수하다고 보고한 바 있다.

**2.2.1. PU 방법** 기 출발편 자료에서 두 DCP에서의 예약 좌석수의 차이를  $d_{2,j} = s_{2,j} - s_{1,j}$ 라 표기할 때, 중도절단되지 않은(즉,  $s_{2,j} < BL_{2,j}$ ) 기 출발편의 경우 모든 수요가 예약되어  $\hat{Z}_{2,j} =$

$Z_{2,j} = d_{2,j}$ 가 된다. 이런 기 출발편  $d_{2,j}$ 들의 평균을  $\bar{d}_0$ 로 표기할 때, PU은 중도절단된 출발편의 순수요  $Z_{2,j}$ 를  $d_{2,j}$ 와  $\bar{d}_0$  중 큰 값으로 추정한다. 즉,  $\hat{Z}_{2,j} = \max(d_{2,j}, \bar{d}_0)$ 를 적용한다. PU에 기반을 둔 미래 출발편의 순수요  $Z$ 는 모든 기 출발편  $\hat{Z}_{2,j}$ 의 평균  $\hat{Z}$ 로 예측한다. 이 과정을 요약하면 다음과 같다.

단계 1(복원) :  $s_{2,j} < BL_{2,j}$ 이면  $\hat{Z}_{2,j} = d_{2,j}$ ,  $s_{2,j} \geq BL_{2,j}$ 이면  $\hat{Z}_{2,j} = \max(d_{2,j}, \bar{d}_0)$ .

단계 2(예측) :  $\hat{Z} = \sum_{j=1}^n \frac{\hat{Z}_{2,j}}{n}$ .

**2.2.2. EM 방법** EM 방법은 Dempster 등 (1977)의 불완전 자료에 대한 EM 알고리즘을 응용한 것으로, 중도절단된 항공 예약자료에 대한 최적의 복원 방법으로 보고되었다 (Zeni, 2001; Weatherford와 Pölt, 2002). EM 방법은 총 2단계로 구성되는데, 먼저 E-단계에서 중도절단된 값은 추정값으로 대체된다. 이후 M-단계에서는  $Z_{2,j} \sim i.i.d. N(\mu, \sigma^2)$ ,  $j = 1, 2, \dots, n$ 의 정규분포 가능도 함수를 최대화한다. 이러한 과정은 반복추정된 가능도 함수가 더 이상 커질 수 없을 때까지 반복된다. 구체적으로,  $s_{2,j} < BL_{2,j}$ 인 경우는  $\hat{Z}_{2,j} = d_{2,j}$ 로,  $s_{2,j} \geq BL_{2,j}$ 에 해당하는 경우  $\hat{Z}_{2,j}$ 는  $Z_{2,j}$ 의 평균  $\mu$ 와 표준편차  $\sigma$ 를 이용하여 다음과 같은  $d_{2,j}$  보다 큰 조건부기대값으로 추정한다.

$$\hat{Z}_{2,j} = E(Z_{2,j} | Z_{2,j} > d_{2,j}) = \mu^{(i)} + \sigma^{(i)} \cdot \frac{\phi(\lambda_j^{(i)})}{1 - \Phi(\lambda_j^{(i)})},$$

여기서  $\lambda_j^{(i)} = (d_{2,j} - \mu^{(i)})/\sigma^{(i)}$  이고,  $\phi(\cdot)$ 와  $\Phi(\cdot)$ 는 각각 표준정규 확률변수의 확률밀도함수와 누적분포함수이다. 이후 M-단계에서는 E-단계로부터 구한  $\hat{Z}_{2,j}$ 들로부터  $\mu^{(i)}$ 와  $\sigma^{(i)}$ 를 갱신하고, 갱신된 값들로부터  $\hat{Z}_{2,j}$ 을 다시 추정한다. 이러한 과정은 추정량  $\mu^{(i)}$ 와  $\sigma^{(i)}$ 이 수렴될 때까지 반복 수행된다.

### 3. 회귀모형 기반 수요복원 방법

2.1절에서 DCP<sub>1</sub> 시점의 예약 좌석수가  $s_{1,j}$ 일 때, DCP<sub>2</sub> 시점의 이론적 참수요는  $S_{2,j} = s_{1,j} + Z_{2,j}$ 로 표현됨을 제시하였다. DCP<sub>1</sub>과 DCP<sub>2</sub> 사이의 신규수요를  $N_{2,j}$ , 그리고  $s_{1,j}$ 의 DCP<sub>2</sub> 시점에서의 예약 취소분을  $C_{2,j}$ 라고 하자. 이때 두 DCP 사이의 예약 취소 비율을  $p$ 라고 하고  $C_{2,j} \sim \text{Bin}(s_{1,j}, p)$ 를 가정하면, 순수요  $Z_{2,j}$ 는 신규수요에서 예약 취소분을 감한 양으로 표현될 수 있으므로 다음의 식이 성립한다.

$$Z_{2,j} = N_{2,j} - C_{2,j}.$$

위에서  $E(N_{2,j}) = \beta_0$ 를 가정하면  $E(Z_{2,j}) = \beta_0 - p \cdot s_{1,j}$ 가 성립하고,  $\beta_1 = 1 - p$ 를 예약 유지 비율로 정의하면 DCP<sub>2</sub> 시점의 참수요  $S_{2,j}$ 에 대한 모형으로 다음의 회귀모형을 고려할 수 있다.

$$S_{2,j} = \beta_0 + \beta_1 s_{1,j} + \epsilon_j \sim N(\beta_0 + \beta_1 s_{1,j}, \sigma^2), \quad \epsilon_j \sim i.i.d. N(0, \sigma^2). \quad (3.1)$$

본 논문에서는 식 (3.1)의 회귀모형 분포로부터 중도절단된 값을 복원하는 방법을 고려하였고, 이를 회귀모형 기반 수요복원 방법(이하 Reg)이라 부르기로 한다. 주목할 것은 기존의 복원 방법에서는 순수요  $Z_{2,j}$ 가 DCP<sub>1</sub> 시점의 예약 좌석수  $s_{1,j}$ 에 무관하다고 가정하는데 비하여, Reg는 예약 유지 비율을 반영한 보다 과학적인 방법으로 기존의 복원방법들과는 상당한 차이가 있을 수 있다는 것이다. 본 논문에서는 새로운 복원 방법의 성능을 모의실험을 통해 기존의 대표적인 복원 방법(PU, EM)과 비교하는 과정에서  $S_{2,j} \sim N(\beta_0 + \beta_1 s_{1,j}, \sigma^2)$ 의 분포를 고려하여 모의자료를 생성하여 분석한다. 참고로 Weatherford와 Pölt (2002)는 중도절단된 기 출발편 데이터의 수요복원 방법들에 대한 비교 연구에서 모의실험 결과 EM 방법이 가장 우수하다고 보고한 바 있다.

Reg 복원 방법에서는 중도절단된( $s_{2,j} \geq BL_{2,j}$ ) 기 출발편에 대한 DCP<sub>2</sub> 시점의 참수요  $S_{2,j}$ 를 다음과 같이 추정한다.

$$\hat{S}_{2,j} = E(S_{2,j} | S_{2,j} > s_{2,j}) = \mu_j^{(i)} + \sigma^{(i)} \cdot \frac{\phi(\lambda_j^{(i)})}{1 - \Phi(\lambda_j^{(i)})}.$$

새로운 복원 방법은 EM 방법과는 다르게 회귀모형을 이용하여 기 출발편, 즉 각  $s_{1,j}$ 마다 다른 평균값  $\mu_j = \hat{\beta}_0 + \hat{\beta}_1 s_{1,j}$ 을 수요복원에 적용한다. 위에서  $\mu_j^{(i)} = \hat{\beta}_0^{(i)} + \hat{\beta}_1^{(i)} s_{1,j}$ ,  $\lambda_j^{(i)} = (s_{2,j} - \mu_j^{(i)})/\sigma^{(i)}$ 이고,  $\phi(\cdot)$ 와  $\Phi(\cdot)$ 는 각각 표준정규 확률변수의 확률밀도함수와 누적분포함수이다. 추정된  $\hat{S}_{2,j}$ 로부터  $\mu_j^{(i)}$ 와  $\sigma^{(i)}$ 가 갱신되고, 이로부터 다시  $\hat{S}_{2,j}$ 가 추정된다. 이러한 과정은 추정량  $\hat{\beta}_0^{(i)}, \hat{\beta}_1^{(i)}, \sigma^{(i)}$ 가 수렴될 때까지 반복 수행된다. 최종 추정된  $\hat{S}_{2,j}$ 로부터  $s_{2,j} < BL_{2,j}$ 인 경우는  $\hat{Z}_{2,j} = d_{2,j}$ 로,  $s_{2,j} \geq BL_{2,j}$ 에 해당하는 경우는  $\hat{Z}_{2,j} = \hat{S}_{2,j} - s_{1,j}$ 로 추정된다.

## 4. 모의실험

### 4.1. 모의실험 설계

3절에서 제안된 새로운 복원 방법인 Reg의 성능을 모의실험을 통해 기존의 대표적인 복원 방법(PU, EM)과 비교하고자 하였다. 모의실험은 EM 방법이 가장 우수하다고 주장한 Weatherford와 Pölt (2002)의 모의실험 설계를 기반으로 계획되었다. DCP<sub>1</sub>의 예약 좌석수  $s_{1,j}$ 는 정규모집단  $N(25, 5^2)$ 로부터 100회( $n = 100$ )의 기 출발편 자료를 생성하였다. 이는 2년(약 100주)간 같은 요일에 출발할 유사한 수요발생 패턴을 갖는 예약자료에 해당되는 것이다.

DCP<sub>1</sub>과 DCP<sub>2</sub> 사이의 새로 발생하는 수요  $N_{2,j}$ 는 i)  $N(10, 2^2)$ , ii)  $N(25, 2^2)$ , iii)  $N(25, 5^2)$ 의 3가지 분포를 고려하였다. 여기서 ii)는 i) 대비 평균 수준이 2.5배 증가하였을 때의 변화를 살펴보기 위함이고, iii)은 i) 대비 평균 수준은 물론 표준편차의 수준이 2.5배 증가하였을 때의 변화를 살펴보기 위함이다.

DCP<sub>2</sub>의 이론적 예약 좌석수  $S_{2,j}$ 는 보다 논리적인 모의자료의 생성을 위해 식 (3.1)의 회귀모형으로부터 생성하였다. 본 논문에는 예약유지비율  $\beta_1$ 으로 1.0, 0.9, 0.8, 0.7의 4가지를 고려하였는데, 이는 예약 취소 비율 0~30%를 상정한 것이다. DCP<sub>2</sub> 시점의 최대 예약가능 좌석수  $BL_{2,j}$ 는 Weatherford와 Pölt (2002)을 참고하여 다음의 분포로부터 생성하였다.

$$BL_{2,j} \sim N(\beta_0 + \beta_1 s_{1,j} + \theta_c \sigma, \sigma^2), \quad \epsilon_j \approx i.i.d. N(0, \sigma^2). \quad (4.1)$$

식 (4.1)의  $BL_{2,j}$ 는 식 (3.1)의  $S_{2,j}$  분포와 비교했을 때, 분산은 같으나 평균 수준만 다른 분포이다. 여기서  $\theta_c$ 를 적절히 정함으로써 원하는 수준의 중도절단된 비율을 갖는 임의자료를 생성할 수 있다. 중도절단된 비율로는 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95의 11가지를 고려하였고, 100회의 기출발편에 대한 DCP<sub>2</sub> 시점의 예약좌석수  $s_{2,j}$ 는  $S_{2,j} < BL_{2,j}$ 이면  $s_{2,j} = S_{2,j}$ ,  $S_{2,j} \geq BL_{2,j}$ 이면  $s_{2,j} = BL_{2,j}$ 로 계산하였다.

본 논문에서는 100회의 기 출발편 자료를 갖는 모의실험 자료를 1,000번 생성하여 비교기준의 평균을 통해 기존의 대표적인 복원 방법(PU, EM)과 새로운 복원 방법(Reg)의 성능을 비교하였다. 비교기준으로는 식 (4.2)의 DCP<sub>2</sub> 시점의 참수요  $S_{2,j}$ 에 대해 다음의 MAE(Mean Absolute Error)를 고려하였다.

$$MAE = \frac{1}{100} \sum_{j=1}^{100} |S_{2,j} - \hat{S}_{2,j}|. \quad (4.2)$$

가장 작은 MAE를 갖는 방법이 참수요에 가까운 값을 추정하는 것이므로 가장 성능이 좋다고 볼 수 있다.

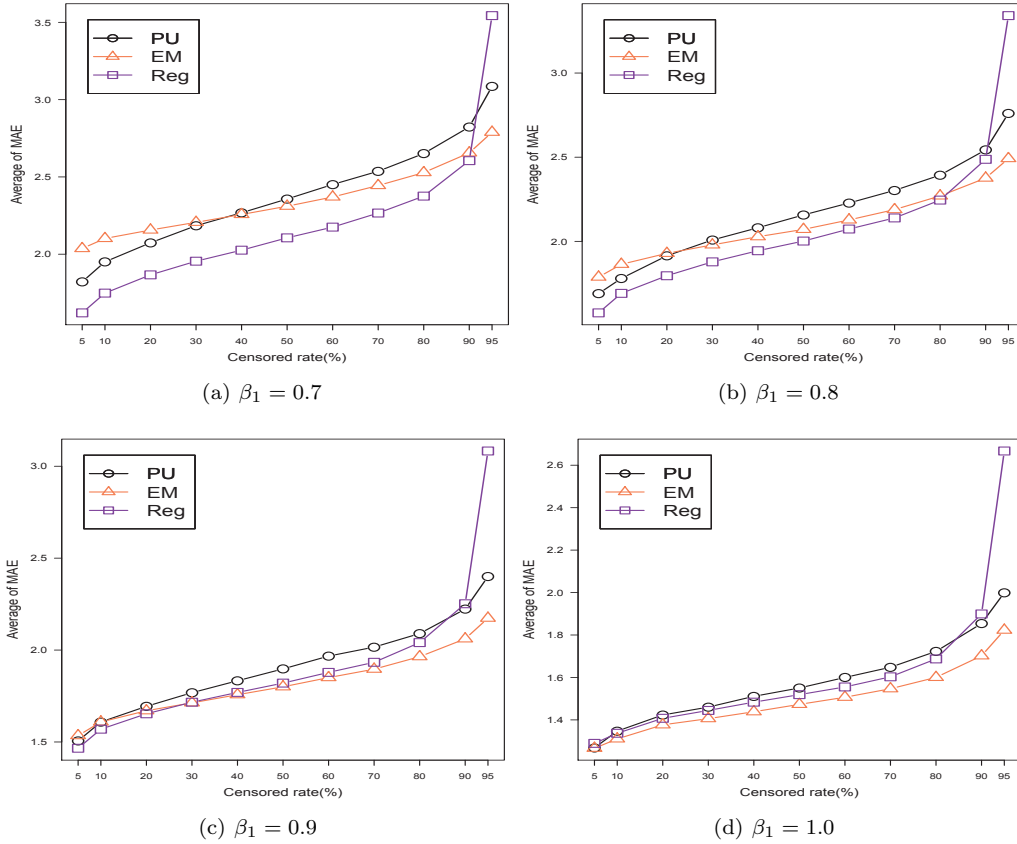


Figure 4.1.  $N_{2,j} \sim N(10, 2^2)$ .

#### 4.2. 모의실험 결과

모의실험 결과는 Figure 4.1~Figure 4.3에 요약되었고, 여기서 세로축은 MAE의 평균을 의미한다. 가장 아래쪽에 위치할수록 평균적으로 MAE가 가장 작은 방법이므로 가장 성능이 좋다고 말할 수 있다. 모의실험 결과로부터 다음의 사실을 알 수 있다.

- Figure 4.1을 보았을 때,  $\beta_1$ 의 값이 1에서 멀어질수록, 즉 예약 취소 비율이 높아질수록 Reg의 MAE값이 PU와 EM에 비해 더 낮게 나타나는 것을 알 수 있다. 따라서 예약 유지 비율  $\beta_1$ 를 고려할 경우, Reg이 다른 방법(PU, EM)에 비해 더 합리적인 방법이라고 판단된다. 그러나  $\beta_1$ 의 값이 1에 가까워질수록 EM의 MAE 평균이 가장 작게 나타나는 것은 EM이 식 (2.1)의 관계에서 볼 수 있듯이 기본적으로  $\beta_1 = 1$ 을 가정하는 방법이기 때문인 것으로 판단된다. 또한 이는 기존의 복원 방법 중 EM이 가장 우수했다고 보고한 Weatherford와 Pölt (2002)의 결과와도 일치한다.
- 그리고 Figure 4.1에서  $\beta_1$ 과 무관하게, 중도절단된 비율이 95%와 같은 극단적인 경우에는 Reg의 MAE 평균이 급격히 상승하는 것을 확인할 수 있는데 이는 다음과 같은 이유일 것으로 판단된다. 한 세트의 모의자료는 100개의 관측치를 가지고 있는데 중도절단된 비율이 95%일 경우, 정상적인 관측치는 약 5개이다. Reg는 추정치를 계산하는 과정에서 3개의 모수를 추정하게 되는데 이는 PU(1개)와 EM(2개)보다 더 많은 모수를 추정하는 것이다. 상대적으로 더 많은 모수를 추정해야 하는 상황에서 정상적인 관측치의 개수가 매우 희소하므로 Reg의 복원 능력이 보다 낮게 나타난 것으로 판단

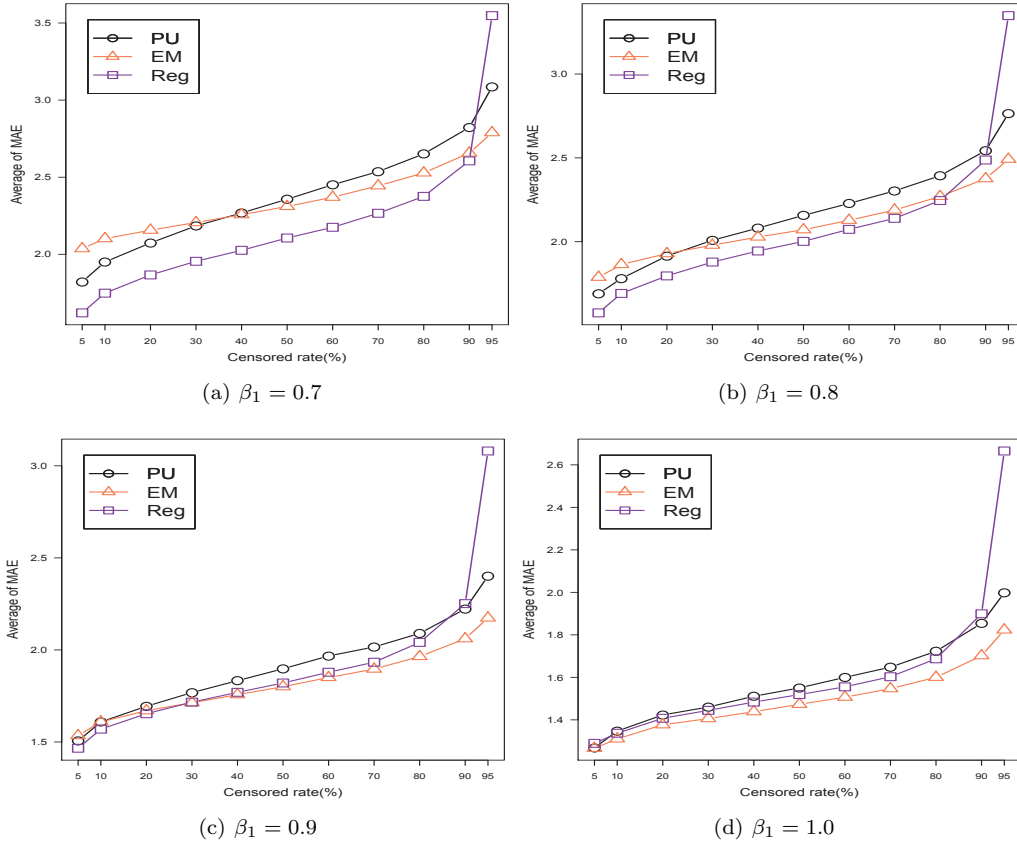


Figure 4.2.  $N_{2,j} \sim N(25, 2^2)$ .

된다. 그러나 일반적으로 발생하는 중도절단된 비율이 90% 이상인 경우가 많지 않을 것으로 판단되므로 중도절단된 비율이 90% 미만인 경우에는 예약 유지 비율  $\beta_1$ 를 고려한 Reg의 성능이 EM 보다 더 좋거나 대등하게 나타나므로 여전히 합리적인 방법이라고 판단된다. 본 논문에서는 예약 유지 비율  $\beta_1$ 를 1.0, 0.9, 0.8, 0.7의 4가지만을 고려하였는데, 실제로  $\beta_1$ 가 0.7 이하로 더 낮아질 경우, EM보다 Reg의 성능이 눈에 띄게 좋아지는 것을 확인하였다.

- Figure 4.2를 살펴보면 Figure 4.1과 거의 같다. 실제로 두 모의실험 결과는 거의 차이가 없다. 즉,  $N_{2,j}$ 의 분포가  $N(10, 2^2)$ 에서  $N(25, 2^2)$ 로 평균 수준이 2.5배 증가하더라도 복원 방법의 성능에 미치는 영향력은 거의 없다고 판단된다. 이 모의실험으로부터 Reg 복원 방법의 성능은 신규수요  $N_{2,j}$ 의 평균의 변화에는 거의 영향을 받지 않음을 확인할 수 있다.
- 반면, Figure 4.3의 결과는 앞의 Figure 4.1과 Figure 4.2와는 차이가 있다. Figure 4.3은  $N_{2,j}$ 의 분포가  $N(10, 2^2)$ 에서  $N(25, 5^2)$ 로 평균 수준은 물론 표준편차의 수준까지 2.5배로 증가한 경우이다. 이 경우 PU는 가장 좋지 않고, EM과 Reg의 결과가 거의 비슷하게 나타나는 것을 알 수 있다. 또한 Figure 4.3에서는 예약 유지 비율  $\beta_1$ 이 1에서 떨어질수록 Reg이 EM 보다 더 좋아지기는 하지만, 그 변화의 정도가 Figure 4.1과 Figure 4.2에 비해서는 미미하다. 이 경우에도 본 논문에는 실지 않았지만,  $\beta_1$ 가 0.7 이하로 더 낮아질 경우, EM보다 Reg의 성능이 눈에 띄게 좋아지는 것을 확인하였다. 따라서 DCP<sub>2</sub> 시점의 신규수요  $N_{2,j}$ 의 표준편차의 크기가 복원 성능에 미치는 영향력은 예약 유

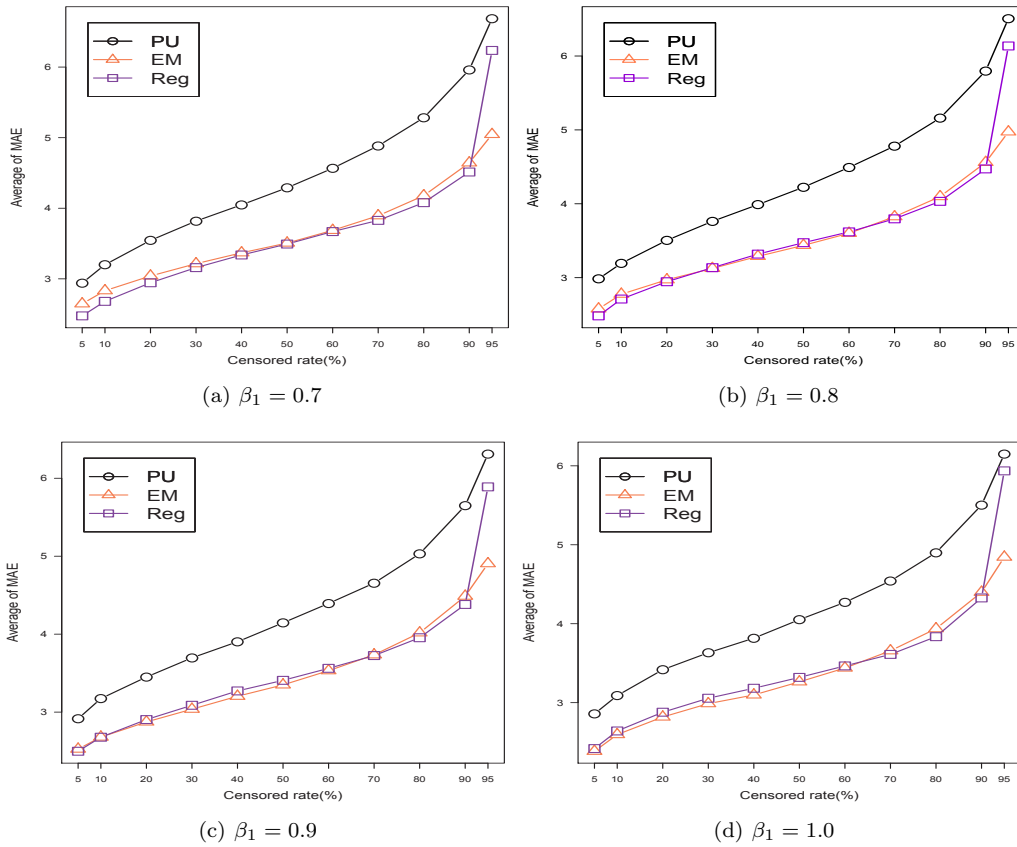


Figure 4.3.  $N_{2,j} \sim N(25, 5^2)$ .

지 비율  $\beta_1$  못지않게 중요한 것으로 판단된다.

- 참고로 본 논문에는  $N_{2,j}$ 의 분포가  $N(10, 2^2)$ 에서  $N(10, 5^2)$ 로 표준편차의 수준만 2.5배로 증가한 경우의 결과는 신지 않았지만, Figure 4.3의 결과와 거의 같게 나옴을 확인하였다. 즉, Reg 복원 방법의 성능은 신규수요  $N_{2,j}$ 의 평균의 변화에는 거의 영향을 받지 않음을 재차 확인하였다.

### 5. 결론

본 논문에서는 예약요청이 미리 정해진 최대 예약가능 좌석수(BL)을 초과하여 일부 수요의 예약이 거부된(중도절단된) 자료로부터 수집된 출발편에 대해 실제 발생했을 것으로 기대되는 이론적 수요를 복원하는 개선된 방법을 제시하였다. 기존에 제안된 복원 방법은 예측 시점(DCP)의 예약 좌석수에 무관하게 참수요를 추정하는 방법이었으나, 새로운 방법인 Reg는 회귀모형에 기반을 두어 예약된 좌석수 중에 예약을 취소하는 패턴을 고려하는 방법이라는 점에서 현실을 반영한 개선된 복원 방법이라고 할 수 있다. 본 논문에서는 모의실험을 통해 기존의 대표적인 복원 방법(PU, EM)과 복원 능력을 비교하였고, 예약 유지 비율을 고려한 Reg이 대부분의 경우에 더 좋은 결과를 보임을 확인하였다. 특히, 예약 유지 비율이 100%가 아닐 때, Reg 방법에 의한 참수요의 추정치가 PU와 EM에 비해 참값에 더 가깝게 나타남을 확인하였다. 그리고 신규수요  $N_{2,j}$ 의 분포에 따라 모의실험 결과에 일부 차이가 있었는데,  $N_{2,j}$ 의

평균 수준의 변화에는 세가지 복원 방법(PU, EM, Reg)의 복원 성능이 거의 영향을 받지 않았으나, 표준편차가 커질수록 Reg와 EM의 차이가 작아지게 됨을 확인하였다. 따라서 신규수요  $N_{2,j}$ 의 표준편차의 크기가 복원 성능에 미치는 영향력은 예약 유지 비율 못지않게 중요하다는 것을 확인하였다. 마지막으로 극단적으로 중도절단된 비율(95%)에서 Reg의 복원 능력이 불안정하게 나타났으나, 이러한 경우에 한하여 Reg와 EM을 융합하여 선택적으로 복원하는 방법으로 이러한 문제를 해결할 수 있을 것으로 생각된다. 요약하면 Reg의 성능은 기존의 복원 방법(PU, EM) 이상의 복원 능력을 갖추고 있고, 또한 두 DCP 시점 사이에서의 좌석수의 관계를 예약 유지 비율을 고려한 회귀모형으로 설명한, 보다 더 합리적인 방법으로 생각할 수 있다. 이는 기존의 항공 운송 분야의 수입관리(RM) 시스템의 개선에 긍정적인 영향을 미칠 것으로 예상되므로 본 논문의 Reg를 RM에 적용하는 것을 제안해 보고자 한다.

## References

- Belobaba, P. P. (1987). *Air Travel Demand and Airline Seat Inventory Management*, PhD thesis, MIT, Cambridge, MA, USA.
- Cross, R. C. (1997). *Revenue Management: Hard-Core Tactics for Market Domination*, Cassel, New York, USA.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **39**, 1–38.
- Talluri, K. T. and Van Ryzin, G. J. (2005). *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers, Boston.
- Weatherford, L. R. and Pölt, S. (2002). Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues, *Journal of Revenue and Pricing Management*, **1**, 234–254.
- Zeni, R. H. (2001). *Improved Forecast Accuracy in Revenue Management by Unconstraining Demand Estimates from Censored Data*, PhD thesis, Rutgers University, Newark, NJ, USA.



# 수입관리에서 회귀모형 기반 수요 복원 방법

이재준<sup>a,1</sup> · 이우주<sup>a</sup> · 김정환<sup>a</sup>

<sup>a</sup>인하대학교 통계학과

(2015년 2월 3일 접수, 2015년 2월 26일 수정, 2015년 2월 26일 채택)

---

## 요약

정확한 수요예측은 수입관리(RM)에서 중요한 요소이다. 기 출발편 예약 데이터는 미래 출발편의 수요를 예측하는데 이용되는데, 이 중 일부 데이터에는 예약 요청이 거부된 경우가 포함된다. 거부된 예약 요청은 통계학적 관점에서 중도절단된 것으로 해석될 수 있으며, 이러한 중도절단된 수요를 복원하는 것은 미래 출발편의 참수요 예측을 위해 중요한 사안이다. 현재까지 여러 복원방법들이 소개되었으며, Expectation Maximization 방법이 가장 우수하다고 알려져있다. 본 연구에서는 중도절단된 자료를 복원할 수 있는 회귀모형 기반의 새로운 수요복원 방법을 제시하였다. 그리고 모의실험을 통해 제안된 새로운 방법의 성능을 RM에서 대표적으로 사용되는 두 가지 복원방법들과 비교하였다.

주요용어: 복원, 중도절단된 데이터, Expectation Maximization, 수입관리

---

---

이 논문은 인하대학교 교내연구비의 지원을 받아 수행되었습니다.

<sup>1</sup>교신저자: (402-751) 인천광역시 남구 용현동 235, 인하대학교 통계학과. E-mail: jjlee@inha.ac.kr