Fader et al.: “Counting Your Customers” the Easy Way

김희영

1. 논문의 목적

"Parameter estimation이 쉬운 미래 구매 예측 모형 생성"

기존의 주 모형인 Pareto/NBD (Schmittein et al., 1987)는 Parameter estimation이 어렵다는 문제가 존재

2. 논문 서술 과정

3. BG/NBD Assumptions

4. Model Development
at the Individual level

5. Model Development
for a Randomly Chosen Individual

6. Simulation & Empirical Analysis

3. BG-NBD Assumptions

가정 1.

While active, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ:

$$f(t_j | t_{j-1}; \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad t_j > t_{j-1} \geq 0.$$

*Poisson process: 시간에 따라 발생하는 이벤트(예: 구매)의 횟수와 관련된 확률과정으로, Poisson process를 따르는 두 연속적인 사건 사이의 시간 분포는 지수분포를 따름

3. BG-NBD Assumptions

가정 2.

Heterogeneity in λ follows a gamma distribution with pdf

$$f(\lambda|r,\alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}, \quad \lambda > 0.$$

거래고객의 이질성을 표현하기 위하여 transaction rate λ가 gamma distribution으로부터 발생했다고 가정함

3. BG-NBD Assumptions

가정 3.

After any transaction, a customer becomes inactive with probability p. Therefore the point at which the customer "drops out" is distributed across transactions according to a (shifted) geometric distribution with pmf

$$\text{P(inactive immediately after } j\text{th transaction)}$$
$$= p(1-p)^{j-1}, \quad j = 1,2,3, \ldots$$

p :고객이 inactive 해질 확률로, 거래 발생 직후 고객의 inactive 여부가 결정됨

3. BG-NBD Assumptions

가정 4.

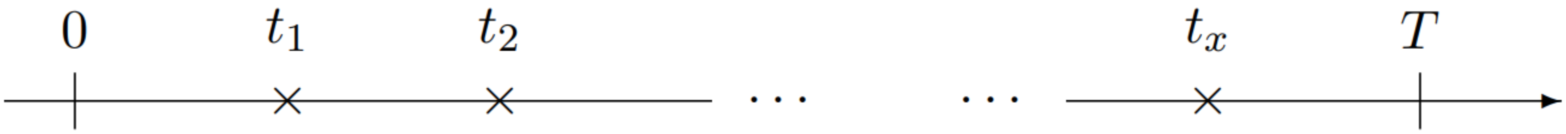
Heterogeneity in p follows a beta distribution with pdf

$$f(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}, 0 \leq p \leq 1,$$

where B(a, b) is the beta function, which can be expressed in terms of gamma functions: B(a, b) = Γ(a)Γ(b)/Γ(a + b).

거래고객의 이질성을 표현하기 위하여 dropout rate p가
beta distribution으로부터 발생했다고 가정함

4. Model Development at the Individual level



$(0, T]$ 기간동안 x 번 구매한 고객의 각각의 구매 시점이 $t_1, t_2, \ldots, t_x$.
이 경우의 individual-level likelihood function을 구하고자 함

각 거래시점 $t_1, \ldots, t_x$와 기간의 끝 T 시점의 가능도를 곱한 결합확률을
likelihood function*으로 정함

*Likelihood function:
주어진 자료에 대하데이터가 얻어질 가능성을 파라미터에 대한 함수로 나타낸 것

4. Model Development at the Individual level

Likelihood function at the Individual level

$$L(\lambda, p|X = x, T) = (1-p)^x \lambda^x e^{-\lambda T} + \delta_{x>0} p(1-p)^{x-1} \lambda^x e^{-\lambda t_x}$$

where $\delta_{x>0} = 1$ if $x > 0$, 0 otherwise.

$P(X(t) = x|\lambda, p)$:t 시점 동안 x 번의 구매가 발생할 확률

$$P(X(t) = x|\lambda, p) = (1-p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} + \delta_{x>0} p(1-p)^{x-1}[1 - e^{-\lambda t} \Sigma_{j=0}^{x-1} \frac{(\lambda t)^j}{j!}]$$

$E(X(t)|\lambda, p)$:t 시점 동안 평균 구매 횟수

$$E(X(t)|\lambda, p) = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t}$$

$*\ X(t)$:A random variable of the number of transactions occuring in a time period of length t (with a time origin of 0)

5. Moving to a Randomly Chosen Individual

$$L(r, \alpha, a, b|X = x, t_x, T) = \int_0^1 \int_0^\infty L\,(\lambda, p|X = x, t_x, T) f\,(\lambda|r, \alpha) f\,(p|a, b) d\lambda dp$$

transaction rate λ와 the dropout probability p 는 관찰되지 않는 파라미터이므로, likelihood function을 λ와 p 분포에 대해 평균을 취함

$$LL(r, \alpha, a, b) = \Sigma_{i=1}^{N} ln\left[L\left(\gamma, \alpha, a, b|X_i = x_i, t_{x_i}, T_i\right)\right]$$

Maximum Likelihood Estimation을 통해 Log-likelihood 함수를 최대화하는 γ, α, a, b를 찾음

5. Moving to a Randomly Chosen Individual

적합된 파라미터 γ, α, a, b로부터
다음과 같은 값들을 구할 수 있음

t 시점 동안 x 번의 구매가 발생할 확률

$$P(X(t)=x|r,\alpha,a,b) = \frac{B(a,b+x)}{B(a,b)}\frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+t}\right)^r\left(\frac{t}{\alpha+t}\right)^x + \delta_{x>0}\frac{B(a+1,b+x-1)}{B(a,b)} \cdot \left[1-\left(\frac{\alpha}{\alpha+t}\right)^r\left\{\sum_{j=0}^{x-1}\frac{\Gamma(r+j)}{\Gamma(r)j!}\left(\frac{t}{\alpha+t}\right)^j\right\}\right]. \quad (8)$$

t 시점 동안 평균 구매 횟수

$$E(X(t)|r,\alpha,a,b) = \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\left(r,b;a+b-1;\frac{t}{\alpha+t}\right)\right], \quad (9)$$

과거 관찰행동(x, t_x, T)을 기반한 기대 거래횟수

$$E(Y(t)|X=x,t_x,T,r,\alpha,a,b) = \frac{\frac{a+b+x-1}{a-1}\left[1-\left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x}{}_2F_1\left(r+x,b+x;a+b+x-1;\frac{t}{\alpha+T+t}\right)\right]}{1+\delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}}. \quad (10)$$

6. Simulation & Empirical Analysis

Table1 Summary of Simulation Results

	MAPE	Penetration	Average purchase frequency
Worst 10 worlds	5.29	26%	2.6
Other 71 worlds	2.32	43%	3.8

MAPE:

평균 절대 비율 오차. $\frac{1}{n}\Sigma_{i=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$, A_t는 실제값, F_t는 예측값

Penetration:

적어도 1번 이상의 구매가 일어난 고객의 비율. $1 - P(X = 0)$

Average purchase frequency:

구매자 중 평균구매건수. $E(X)/\left(1 - P(X = 0)\right)$

Figure 3 Conditional Expectations