

센서링과 절단 환경에서의 경로 손실 추정 방법에 대한 비교 연구

이 경 규*, 오 성 준^o

A Comparison Study on the Path Loss Estimation in Censoring and Truncation Environments

Kyung-gyu Lee*, Seong-jun Oh^o

요 약

밀리미터파 대역은 주파수가 30GHz-300GHz이고, 파장이 10mm-1mm인 EHF (Extremely High Frequency) 대역이다. 밀리미터파 대역에서는 장애물이 있는 경우 전파 감쇠가 심하기 때문에 Line-of-Sight (LoS)가 아닌 경우 신호가 잘 잡히지 않는다. 그렇기 때문에 밀리미터파 대역에서 신호 감쇠 측정을 할 경우에 측정 장비가 noise와 구별할 수 없는 신호들이 관찰된다. 이와 같이 감쇠가 심한 환경에서 신호 감쇠 data를 보면 특정한 값에서 제한을 받는 것이 관찰된다. 특정한 값에서 제한 받는 것을 그대로 두고 일반적인 Least square로 추정을 하는 경우에는 감쇠 exponent를 과소평가 할 수도 있다. 본 논문에서는 특정한 값에서 제한을 받아도 정확한 추정이 가능한 Tobit Maximum Likelihood Estimation, Heckman Two-stage Model 그리고 Truncation Regression model의 성능 비교를 하였다.

Key Words : Regression, Estimation, Millimeter-wave, Path loss, Censoring, Truncation

ABSTRACT

The millimeter wave band is in the extremely high frequency band whose frequency and wavelength are 30-300GHz and 10-1mm respectively. When the obstacles block the propagation path which is not Line-of-Sight (LoS), due to a high propagation loss, it is hard to receive a signal in the millimeter wave band. Therefore When the path loss is measured in the millimeter wave band, the signal which is not distinguished from the noise is observed. Consequently, the path loss data which is limited in certain value is observed in the high propagation loss environment. If the original least square is implemented without taking the limitation of certain value into account, the path loss exponent may be underestimated. In this paper, the performance of Tobit Maximum Likelihood Estimation, Heckman Two-stage Model and Truncation Regression Model which can estimate properly in the censoring or truncated environments are compared.

I. 서 론

지난 십 몇 년간 2G, 3G, 4G로 무선통신기술이 발

전하면서 점차 트래픽은 증가하였다. Cisco의 Visual Networking Index^[1]의 전망에 따르면, 전 세계적인 트래픽은 스마트폰의 대중화에 따른 트래픽 증가로

* 본 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행되었음.

• First Author : Korea University Graduate School of Computer and Information Technology, dgmstt@korea.ac.kr, 학생회원

o Corresponding Author : Korea University Graduate School of Information Security, seongjun@korea.ac.kr, 종신회원

논문번호 : KICS2016-10-318, Received October 18, 2016; Revised January 3, 2017; Accepted January 31, 2017

2015년 말에 월 3.7 exabytes에 이르렀다. 또한 Internet of Things (IoT)^[11], Device to Device (D2D), Machine to Machine (M2M), Financial Technology (FinTech)의 발전에 따라서 2020년에는 월 30.6 exabytes에 도달할 전망이다. IOT, D2D, M2M, FinTech를 상용화 하려면 4G보다 더 높은 data rate가 필요하다. 이에 따라 IMT-2020으로 명명된 5G가 2015년부터 표준화가 진행되고 있다. 5G에서 높은 data rate를 얻기 위한 방법 중에 하나는 광대역을 사용하거나 MIMO^[12]의 개수를 더 늘리는 것이다. 현재 4G가 사용하는 6GHz 이하의 대역은 이미 주파수의 포화상태에 따라서, 수백 메가헤르츠 단위의 주파수 대역이 어려운 실정이다. 그렇기 때문에 5G에서는 6GHz 이상의 대역 (30GHz-300GHz)을 활용하여 통신에 사용하는 방향으로 진행 중이다.

30GHz-300GHz 대역은 보통 밀리미터파 대역이라고 불린다. 밀리미터파 대역에서는 고정통신, 위성들이 사용해왔지만, 거의 LoS 환경에 제한되어서 사용을 했다. 반면에 이동통신은 LOS 뿐만 아니라 Non Line-of-Sight (NLoS)에서도 사용 가능성이 있다. 밀리미터파 대역은 높은 주파수로 인하여 경로 감쇠가 심하고, 낮은 주파수와 다르게 회절이 영향이 약해진다. 그렇기 때문에 채널이 낮은 주파수 대역과는 달라고, 채널에 대한 많은 연구가 진행 중이다. 밀리미터파 대역의 측정은 세계 여러 나라에서 진행되었다. ETRI, DOCOMO, NYU, UT Austine에서 26.4, 28, 37.1, 38, 73 GHz 등의 측정이 이루어졌다^[2-4]. [2]는 100m 정도의 단거리를 측정해서 크게 드러나지 않지만, [3]과 [4]에서 NLoS 경우에 path loss가 일정 수치에서 더 이상 증가하지 않고 정체되어있는 것을 볼 수 있다. 왜냐하면 장비의 한계를 벗어나는 신호가 존재하기 때문이다.

신호 측정에 있어서, 장비의 한계를 벗어나는 신호(outage)는 존재한다. 너무 감도가 작기 때문에 잡음과 구별이 되지 않기 때문에 발생하게 된다. 신호를 실제로 측정을 하다 보면 어떤 신호 세기 이하로는 측정이 되지 않는 경우가 생긴다. 신호의 세기를 측정할 때, 송신 신호 세기와 수신 신호 세기의 차이인 신호 감쇠로 모델링을 한다^[13]. 일반적으로 신호 감쇠를 모델링을 할 때는 Least square 방법을 사용하여, 1차식으로 추정하는 경우가 많다. 신호 샘플에 outage가 없는 경우에는 Least square는 최소의 오차 값을 찾아내어, 가장 적합한 파라미터를 추정 하게 된다. 그러나 만약 outage가 많이 발생했는데도 불구하고, 이것을 무시하고 추정을 하게 되면 편향(bias)가 생긴다. 편

향이 발생하면, 실제 값을 과대평가하거나 과소평가하게 된다. 그렇기 때문에 outage를 감안하여 신호 감쇠 모델링을 하는 것이 필요하다.

II. Missing Data 이론

Outage를 인지하고, 추정을 하는 방식에는 관측되지 않은 data (missing data)를 대체(imputation)하는 방법이 있다. 먼저 관측되지 않은 data가 어떻게 발생했는지에 대해서 분류하면, MCAR (Missing Complete At Random), MAR (Missing At Random), MNAR (Missing Not At Random) 등이 있다. MCAR은 관측되지 않은 data가 완전히 무작위로 발생한다고 보는 것이고, MAR은 관측되지 않은 data가 관찰되는 data에 의해 제한되지 않는 경우이다. MNAR은 관측되지 않은 data가 관찰되는 어느 특정한 값에 의해서 제한되는 경우이다. 그러므로 신호 감쇠에 의해서 제한되는 경우는 어느 특정한 값에 의해서 제한되는 경우인 MNAR이라고 할 수 있다.

관측되지 않은 data를 포함하여 다루는 경우에는, 잃어버린 data의 값을 다른 값으로 대체한다. MNAR에서는 대체 값을 추정하는 방법에는 Censoring 방법, Truncation 방법, Sample selection 방법이 있다. Sample selection 방법은 신호 감쇠가 거리 이외의 다른 샘플에 대해서 제한을 받는다는 가정이 필요하다. 그러나 신호 감쇠는 x축인 거리 이외의 다른 변수에 의해서 제한을 받지 않는다고 가정하기 때문에 사용할 수 없는 방법이다. 본 논문에서는 Censoring 방법과 Truncation 방법을 사용한다. Censoring 방법은 관측되지 않은 data의 값이 한계 값으로 추정하는 방법이다. 예를 들어서 신호 감쇠 값이 c 이상 넘어가면 outage가 난다고 할 때, 그 값들을 모두 c 로 고정시키는 것이다. 반면에 Truncation 방법은 관측되지 않은 data의 값을 아예 관측하지 못했다고 가정을 한다. 그리고 대체 값을 0으로 추정하고, 관측되는 data만 가지고 계산하는 방법이다. 본 논문에서는 Censoring 방법이 바탕이 되는 Tobit ML estimation, Heckman two-stage model 그리고 Truncation 방법인 Truncate Regression model의 성능을 비교해 볼 것이다.

III. Regression Model

본 논문에서 분석에 사용한 신호 감쇠 model은 다음과 같다.

$$a_1 + a_2 * 10 \log(d), \quad (1)$$

α_1 은 1차식의 y절편이고, α_2 는 감쇠 exponent이다. 그리고 d 는 떨어진 거리 (m)를 나타낸다. 식 (1)은 AB model로 주파수 계수가 없는 1차식으로 나타난 간단한 신호 감쇠 model이다. 식 (1)은 일반적으로 Least square로 파라미터들을 추정한다. 이 때 관측된 값 y 는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} y &= a_1 + a_2 * 10 \log(d) + \epsilon \\ &= [1 \ 10 \log(d)] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \epsilon \\ &= x\beta + \epsilon \end{aligned} \quad (2)$$

이 때, x 는 길이 d 를 포함하고 있고, β 는 파라미터이다. ϵ 은 평균이 0, 분산이 σ^2 인 가우시안 분포의 랜덤변수이다. y 의 평균과 분산은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} E[y|x] &= E[x\beta + \epsilon] = E[x\beta] + E[\epsilon] = x\beta \\ V[y|x] &= V[x\beta + \epsilon] = V[\epsilon] = \sigma^2 \end{aligned} \quad (3)$$

y 를 Least square로 추정을 하면 다음과 같다.

$$\beta = (x^T x)^{-1} x^T y \quad (4)$$

만약 y 가 일정한 값 c 로 제한이 되지 않는다면, Least square은 가장 정확하고 치우치지 않는 (unbiased) 추정을 하게 된다. 그러나 관측 값이 특정한 값 c 에 의해서 제한을 받는 환경에서는 달라진다. 관측 값 y 의 잠재적 값인 y^* 이 특정한 값 c 에 의해 제한을 받는다면, 다음과 같이 나타낼 수 있다.

$$\begin{aligned} y &= \begin{cases} y = y^* & \text{if } y^* < c \\ y = c & \text{if } y^* \geq c \end{cases} \\ y^* &= x\beta + \epsilon, \end{aligned} \quad (5)$$

식 (5)는 y^* 가 c 보다 작을 때는 관측 값을 그대로 사용하고, y^* 이 c 보다 클 때는 c 로 대체한다. 이 경우에는 평균이 다음과 같이 나타난다.

$$E[y|x] = pr(y \geq c|x) E[y|y \geq c,x] + pr(y < c|x) E[y|y < c,x] \quad (6)$$

$E[y|y < c,x]$ 를 다시 나타내면 다음과 같다.

$$\begin{aligned} E[y|y < c,x] &= E[x\beta + \epsilon|y < c] \\ &= E[x\beta|y < c] + E[\epsilon|y < c] \\ &= x\beta + E[\epsilon|y < c] \end{aligned} \quad (7)$$

ϵ 은 가우시안 분포의 평균이 0인 랜덤변수이기 때문에 조건 평균 $E[\epsilon|y < c]$ 은 반드시 0이 아니다. 일반적인 Least square로 추정을 하게 되면, $pr(y < c)$ 는 1이 되고, $pr(y \geq c)$ 는 0이 된다. 그러면 평균은 다음과 같이 된다.

$$E[y|x] = x\beta + E[\epsilon|y < c] \quad (8)$$

식 (8)을 보면 Censoring 또는 Truncation 된 data를 처리 할 때 일반적인 Least square를 사용하면 $E[\epsilon|y < c]$ 가 0이 아니기 때문에 편향된 값을 가지게 되어 부정확하게 되는 것을 알 수 있다.

3.1 Tobit Maximum Likelihood Estimation

Tobin의 Tobit model(1958)^[5]은 Tobin의 probit model의 약자로 Censored data 환경의 기본적인 추정 모델이다. Tobit model은 식 (5)에서 나타나는 잠재적인 (Latent) 값인 y^* 에 의해서 y 가 결정되는 model을 말한다. Tobit model을 Maximum Likelihood (ML)로 추정하는 것을 Tobit Maximum Likelihood Estimation 이라고 부른다.

ML은 Least square과는 다르게 관측된 값의 분포를 통하여 추정을 한다. 관측되는 값이 모두 독립적이고, 가우시안 분포를 가진다고 가정을 하면, ML의 결과는 Least square의 결과와 같다. 하지만 Censoring 환경에서는 특정한 값 c 에서 제한이 되기 때문에 가우시안 분포가 아니다. ML에 사용될 분포는 Censoring 환경 정의에 의해서 다음과 같다.

$$f(y_i|x_i) = \begin{cases} \Pr(y_i = c|x_i) & \text{if } y_i \geq c \\ f^*(y_i|x_i) & \text{if } y_i < c \end{cases} \quad (9)$$

Censoring 환경에서 c 이상의 y 는 c 로 고정되기 때문에 $\Pr(y_i = c|x_i)$ 가 분포가 된다. 그리고 y 가 c 보다 작게 되면 그대로 가우시안 분포 $f^*(y_i|x_i)$ 를 사용한다. $\Pr(y_i = c|x_i)$ 는 확률의 기본적인 식으로 유도가능하다.

$$\begin{aligned} \Pr(y_i = c|x_i) &= \Pr(y_i^* \geq c|x_i) = \Pr(x_i\beta + \epsilon_i \geq c) \\ &= \Pr\left(\frac{\epsilon_i}{\sigma} \geq \frac{c - x_i\beta}{\sigma}\right) = 1 - \Phi\left(\frac{c - x_i\beta}{\sigma}\right) \end{aligned} \quad (10)$$

또한 가우시안 분포 $f^*(y|x)$ 는 다음과 같이 나타낼 수 있다.

$$f^*(y_i|x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i-x_i\beta)^2}{2\sigma^2}} \tag{11}$$

$$= \frac{1}{\sigma} \phi\left(\frac{y_i-x_i\beta}{\sigma}\right)$$

ϕ 와 Φ 는 각각 표준정규분포와 표준정규누적분포이다. ϕ 와 Φ 는 다음과 같다.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{12}$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

식 (9), (10), (11)를 모두 사용하면 Tobit ML Estimation의 likelihood 함수가 나온다^[5]. Likelihood 함수는 n 개의 data가 관측되고 m 개의 data가 관측이 되지 않았을 때를 가정한다. 각 data는 모두 독립적이라고 가정한다. Tobit ML Estimation의 likelihood 함수는 다음과 같다.

$$l(y_i, x_i) = \prod f(y_i|x_i) = \prod_n \frac{1}{\sigma} \phi\left(\frac{y_i-x_i\beta}{\sigma}\right) \prod_m \left(1-\Phi\left(\frac{c-x_i\beta}{\sigma}\right)\right) \tag{13}$$

n 은 관측되는 data의 개수이고, m 은 관측되지 않은 data의 개수이다. 식 (13)에 대해서 최댓값을 가지는 β 를 찾아서 추정을 한다. Likelihood 함수는 누적분포에 대한 함수이기 때문에 비선형이다. 그렇기 때문에 일반적인 미분으로 최댓값을 찾기는 어렵다. 본 논문에서는 Matlab의 fminsearch 함수를 사용하여서 최댓값을 찾았다^[6].

3.2 Heckman Two-stage Model

Heckman two-stage model은 James Heckman의 sample selection model을 바탕으로 하고 있다^[7]. Sample selection model은 여러 개의 샘플 중에 몇 가지 샘플을 알고 있고, 나머지 샘플에 대한 정보가 없다고 가정한다. 알고 있는 몇 가지 샘플 등을 통해서 모르는 샘플들을 추정하는 방식이다. Path loss 추정과 같이 샘플이 한 가지만 있는 경우에는 sample

selection model은 censoring model과 같아진다.

Heckman two-stage model은 2가지 단계로 나뉘어져서 계산된다^[14]. 첫 번째 단계는 selection 단계로 관측된 data와 관측되지 않은 data를 모두 사용하여, 1차적인 파라미터들을 추정한다. 그리고 두 번째 단계인 outcome 단계에서는 관측된 data만 사용하여, 최종적인 파라미터 β 를 추정한다.

Selection 단계의 과정은 다음과 같이 나타낸다. 먼저 식 (5)의 환경에서 식 (7)을 유도한다. 식 (7)은 다음과 같이 나타낼 수 있다^[14].

$$E[y_i|y_i^* < c, x_i] = x_i\beta + E[\epsilon_i|y_i^* < c] \tag{14}$$

$E[y_i|y_i^* < c, x_i]$ 는 관찰되는 data의 조건부 평균이고, β 는 궁극적으로 추정해야하는 파라미터이다. $E[\epsilon_i|y_i^* < c]$ 는 잘려진 정규분포 랜덤변수의 평균이다. $E[\epsilon_i|y_i^* < c]$ 는 통계적인 결과를 통해서 구할 수 있고, 다음과 같다^[8].

$$E[\epsilon_i|y_i^* < c] = \sigma \frac{-\phi\left(\frac{x_i\hat{\beta}}{\sigma}\right)}{\Phi\left(\frac{x_i\hat{\beta}}{\sigma}\right)} = -\sigma\lambda_i \tag{15}$$

λ_i 는 inverse Mills ratio 또는 hazard rate라고 불린다^[9]. $\hat{\beta}$ 와 σ 는 모든 data가 주어졌을 때, 추정되는 확률을 가지고 probit model을 통하여 추정을 할 수 있다. Probit model은 가변수 d 의 값이 0과 1만 존재하고, 다음과 같이 나타낼 수 있다.

$$d_i = \begin{cases} d_i = 1 & \text{if } y_i^* < c \\ d_i = 0 & \text{if } y_i^* \geq c \end{cases} \tag{16}$$

$$y^* = x\hat{\beta} + \epsilon,$$

Probit model의 likelihood 함수는 다음과 같이 나타난다.

$$l = (\Phi(x_i\hat{\beta}))^{d_i} (1-\Phi(x_i\hat{\beta}))^{(1-d_i)} \tag{17}$$

식 (17)을 ML을 통하여 $\hat{\beta}$ 를 추정할 수 있다. σ 는 probit model에서 따로 분리하여 추정하기 어렵다. 그래서 정규분포로 가정하고 1로 가정한다. Probit model을 통하여 $\hat{\beta}$ 를 추정하면, 그 값으로 λ_i 를 구할

수 있다. Outcome 단계로 넘어가면, 식 (14)에서 β 를 제외하고 다른 변수들은 추정이 되어있다. 관찰되는 data를 사용하여, 일반적인 Least square를 통해 β 를 추정한다. Heckman two-step estimation은 probit model을 통해서 쉽고 직관적으로 구할 수 있다. 하지만 가장한 표준편차 σ 가 부정확하다는 단점이 있다.

3.3 Truncated Regression Model

Truncated regression model에서는 Censoring 환경과는 다른 환경을 가정하게 된다. Censoring 환경이 관측되지 않은 data의 정보가 있는 반면에 Truncation 환경은 관측되지 않은 data에 대한 정보가 존재하지 않는다. Truncated regression model에 대한 가정은 다음과 같이 볼 수 있다.

$$y = \begin{cases} y = y^* & \text{if } y^* < c \\ y = 0 & \text{if } y^* \geq c \end{cases} \quad (18)$$

$$y^* = x\beta + \epsilon,$$

Censoring 환경 같은 경우에는 y_i^* 가 c 보다 크면 값이 c 로 대체가 된다. 그러나 Truncation 환경에서는 값을 모르는 것뿐만 아니라 그에 대응하는 x 축도 모른다. 신호 감쇠를 예를 들자면, outage 난 곳의 값뿐만 아니라 거리 또한 모른다고 보아야 한다. 그렇기 때문에 Truncated 환경에서는 two-stage 방법을 사용할 수 없다. Two-stage 방법은 probit model로 $\hat{\beta}$ 를 추정하여 λ_i 를 selection 단계에서 추정을 한 다음에, outcome 단계에서 새로운 변수로 집어넣어서 다시 추정을 한다. Truncated 환경에서는 관측되지 않은 data에 대한 정보가 없기 때문에 $\hat{\beta}$ 를 추정 할 수가 없다. 그래서 Truncated regression model에서는 ML을 사용한다. ML에 사용할 분포는 식 (9)의 $f(x|y^* < c)$ 의 경우와 같다. $f(x|y^* \geq c)$ 인 경우는 c 보다 큰 경우는 없기 때문에 존재하지 않는다. Truncated regression model의 likelihood 함수는 다음과 같이 나타난다^[14].

$$l(y_i, x_i) = \prod_n \frac{1}{\sigma} \frac{\phi\left(\frac{y_i - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{c - x_i\beta}{\sigma}\right)} \quad (19)$$

n 은 관측이 되는 data 값의 개수이다. Truncated regression model은 likelihood 함수의 형태는 Tobit ML estimation에서 관측되지 않은 data의 부분이 제외된 형태이다.

IV. 시뮬레이션 결과

각 model들을 비교함에 있어서 필요한 신호 감쇠 샘플은 Ray-tracing simulation을 통해서 구하였다^[10]. 본 논문에서 사용한 Ray-tracing simulation은 실제 측정값과 비교를 통해 검증을 마친 것으로 2016년 ITU-R WP3K 회의에 제출을 하였다^[15]. 또한 측정값과 simulation 값의 Root Mean Square Error는 10.9 dB로 실제 path loss의 경향을 잘 표현할 수 있다. [15]에서 실제 측정을 한 것 중에 28GHz의 NLoS만을 선별하여 Ray-tracing simulation을 진행하였다. Ray-tracing simulation으로 만든 신호 감쇠 샘플의 개수는 2349개이고, 값의 범위는 86 dB 부터 171 dB 까지다. 신호 감쇠 샘플에 관측 제한이 없는 경우 추정된 감쇠 exponent (PLE)는 5.23이다.

그림 1에 보면 신호 감쇠 샘플을 가지고 5개의 model들을 추정해놓은 것을 볼 수 있다. 검은색 실선이 신호 감쇠의 관측 한계 (Truncation point)이고, 값은 140 dB 이다. 관측 한계를 넘어서서 관측이 되지 않는 신호 감쇠들 (Unobserved Path loss)은 빨간색 x로 표시되어있다. 관측이 가능한 신호 감쇠 (Observed Path loss)는 파란색 점으로 표시되어있다. 첫 번째 model인 Complete Least square는 관측이 되는 것과 되지 않는 모든 신호 감쇠를 사용하여 일반적인 Least square를 한 것이다. 그러므로 Complete Least Square는 가장 오차가 적고, 정확한 model이라고 할 수 있다. 그 다음 model들로는 Tobit ML estimation, Heckman two-stage model 그리고 Truncated regression model이 있다. 이 model들은 각각의 가정된 환경에 따라서 추정을 한다. 마지막으로 Truncated

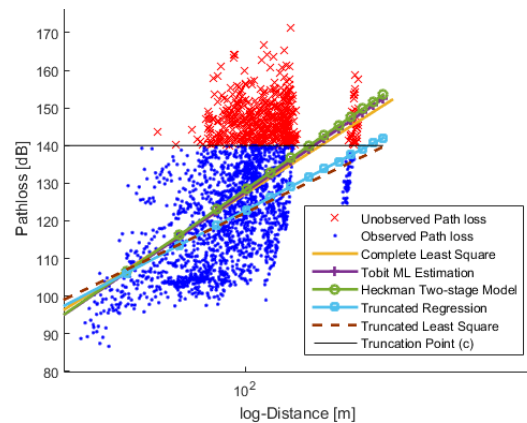


그림 1. Path loss sample과 Regression model 예시
Fig. 1. Example of the path loss sample and the regression models

Least square은 관측된 신호 감쇠만을 가지고 Least square를 한 경우이다. 관측 제한을 가정하지 않은 model이기 때문에 가장 오차가 크다.

그림 2와 그림 3은 그림 1의 관측되는 신호 감쇠 data percentage에 대해 RMSE와 PLE를 비교한 것을 나타낸 것이다. 관측 한계가 낮을수록 더 관측되는 신호 감쇠 data percentage는 떨어진다. 반대로 관측 한계가 높을수록 관측되는 신호 감쇠 data percentage가 높아지고, 각각의 추정치 원래 값에 더 가까워진다. RMSE의 계산은 다음과 같이 나타낸다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (x_i \beta_{CLS} - x_i \beta)^2} \quad (20)$$

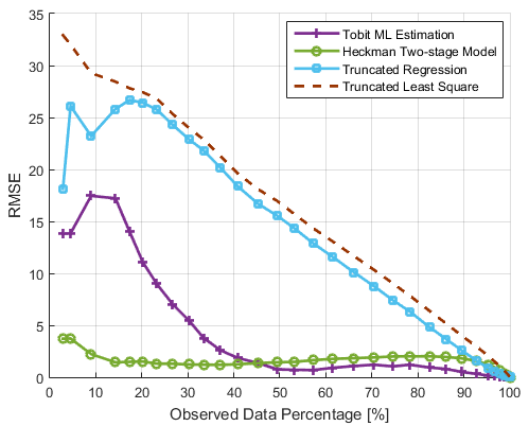


그림 2. 관측되는 data percentage에 따른 RMSE 비교
Fig. 2. Comparison of RMSE in terms of the observed data percentage

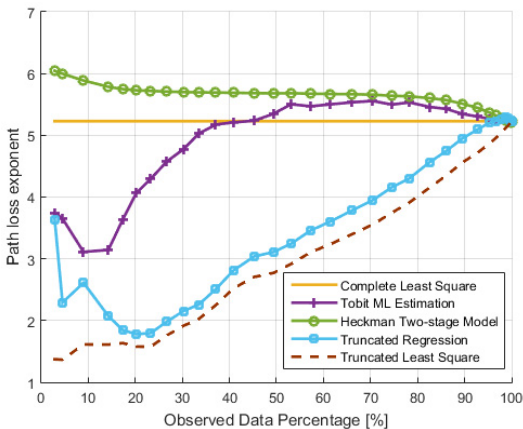


그림 3. 관측되는 data percentage에 따른 Path loss exponent 비교
Fig. 3. Comparison of the path loss exponent in terms of the observed data percentage

β_{CLS} 는 Complete Least square의 파라미터이고, β 는 그 외의 다른 model들의 파라미터이다. N 은 model의 샘플의 개수이다.

그림 2에서는 Tobit ML estimation과 Heckman two-stage model이 교차하는 것이 보인다. 관측되는 data percentage가 50%인 지점으로, 이 지점을 전후로 하여 Tobit ML estimation의 RMSE가 Heckman two-stage model 보다 더 좋아진다. 이것은 Heckman two-stage model에 사용된 probit model 때문에 나타나는 현상이다. Probit model은 data들을 모두 0과 1로 치환하고, CDF로 추정을 한다. 그래서 관측된 값의 개수의 변화에 대해 크게 바뀌지 않는다. 반대로 관측한 것이 많아지는 경우에도 느리게 변화한다. 그래서 일정한 구간 이후에는 Tobit ML estimation의 성능이 더 좋아진다. Truncation 환경은 Censoring 환경보다 정보가 적기 때문에 관측이 되지 않은 신호 감쇠가 많을수록 성능이 낮을 수밖에 없다. Truncated regression model과 Truncated Least square은 같은 조건임에도 불구하고, Truncated regression model이 성능이 더 잘 나온다.

그림 3에서는 관측되는 data percentage 변화에 따라서 PLE가 변화하는 것을 보여준다. Complete Least square는 모든 data를 알고 있다고 가정했기 때문에, 관측되는 data에 상관없이 5.23의 PLE를 유지한다. Tobit ML estimation과 Heckman two-stage model은 78 %에서 각각 5.5와 5.6을 나타내었다. Truncated regression model과 Truncated Least square은 각각 4.3과 3.9를 나타내었다. 관측되는 data가 95 %로 증가하면, Truncated Least square를 제외한 나머지 model들은 PLE의 차이가 0.03정도 밖에 나지 않는다. 확실히 신호 감쇠가 관측이 되지 않은 것을 고려하지 않고, Least square를 하면 편향이 심해지는 것을 볼 수 있다.

세 가지 모델의 Truncated Least square 대비 성능은 표 1에 나와 있다. 각 모델의 성능은 Truncated Least square의 RMSE보다 얼마나 더 적은 RMSE를 가지는 지를 percentage로 나타내었다. Observed data percentage가 98% 이상에서는 Truncated Least square의 RMSE가 작기 때문에, 그 이하에서 의미가 있다. 표 1은 일반적인 Least square로 추정을 하는 것보다 세 가지 모델을 사용하면 얻을 수 있는 정확도를 볼 수 있다. Tobit ML estimation은 극단적으로 data가 적은 경우를 제외하고 성능이 가장 좋다. Heckman two-stage model은 관측되는 data가 적을 때 가장 정

표 1. Truncated Least square 대비 RMSE 향상 percentage (%)
Table. 1. Percentage of RMSE improvement compare to Truncated Least square (%)

Truncation point (dB)	Observed data percentage (%)	Tobit ML	Heckman two-stage	Truncated regression
120	38.0	87	94	5
130	57.3	95	88	10
140	78.3	84	74	19
145	87.7	85	58	31
150	95.2	92	41	56
155	98.3	93	35	67
160	99.4	97	43	68

확하다. Probit model의 성질로 인하여, RMSE가 일정하게 유지되어 성능 향상이 크지 않다는 것은 단점이다. Truncated regression model은 Truncated Least square 보다 좋기는 하지만, 큰 성능의 기대는 어렵다. 그러나 관측되지 않은 data에 대한 정보가 없을 경우에도 사용이 가능하다는 이점이 있다. 대체적으로 세 가지 모델들 모두 Truncated Least square 보다 더 좋은 성능을 보여준다.

V. 결 론

본 논문에서는 path loss가 특정한 값에 의해서 제한을 받을 때 기존의 model의 문제점을 보완할 수 있는 모델 세 가지를 소개했다. Censoring 기반의 Tobit ML estimation과 Heckman two-stage model 그리고 Truncation 기반의 Truncated regression model이다. Tobit ML estimation은 Tobit model에서 truncated normal 분포를 적용시켜 ML로 추정한다. Tobit ML estimation은 관측된 data가 50% 이상에서 성능이 가장 좋았다. Heckman two-stage model은 두 단계로 추정을 한다. 첫 번째 selection 단계에서는 probit model을 통하여 보정할 파라미터를 찾아낸다. 그리고 outcome 단계에서 그 값으로 보정하여 Least square로 추정을 한다. Heckman two-stage model은 probit model로 인해서 관측된 data 개수에 따라 성능이 크게 변하지 않는다. 마지막으로 Truncated regression model은 다른 두 모델과는 다르게 관측되지 않은 data에 대한 정보가 없다고 가정한다. Truncated normal 분포를 적용시켜 관측되지 않은 data에 대한 분포는 제외하고, ML로 추정한다. 실제로 omni-antenna로 밀리미터파를 측정하는 경우 관측 한계는 160 dB

정도로 볼 수 있다. 본 논문에서 제시한 sample은 관측 거리가 150m 정도 밖에 되지 않아서, 160 dB에서 관측되는 data percentage가 99.4 %이다. 99.4 %에서 세 가지 모델과 일반적인 Least square의 성능이 크게 차이가 없기 때문에 일반적인 Least square가 더 쉽고 간단한 모델이다. 그러나 관측 거리가 늘어나서 관측되는 data percentage가 70-90 %로 줄어드는 경우, 세 가지 모델을 사용하는 것이 더 정확한 추정을 이끌어 낼 수 있다.

References

- [1] Cisco, *Visual Networking Index*(2016), Retrieved Feb., 01, 2016, from <http://www.cisco.com>
- [2] H.-K. Kwon, M.-D. Kim, and Y.-J. Chong, "Implementation and performance evaluation of mmWave channel sounding system," in *Proc. IEEE APS/URSI*, pp. 1011-1012, Jul. 2015.
- [3] K. Kitao, T. Imai, N. Tran, N. Omaki, Y. Okumura, M. Inomata, M. Sasaki, and W. Yamada, "Path loss prediction model for 800 MHz to 37 GHz in NLOS microcell environment," *PIMRC2015*, pp. 414-418, Hong Kong, China, Sept. 2015.
- [4] G. R. MacCartney, Jr., T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 GHz and 73 GHz for ultra-dense 5G wireless networks (Invited)," *IEEE Access*, vol. 3, pp. 2388-2424, Dec. 2015.
- [5] J. Tobin, "Estimation of relationships for limited dependent variables," *Econometrica*, vol. 26, pp. 24-36, 1958.
- [6] C. Gustafson, et al., "Statistical modeling and estimation of censored pathloss data," *IEEE Wireless Commun. Lett.*, vol. 4, no. 5, Oct. 2015.
- [7] J. J. Heckman, "Sample selection bias as a specification error," Unpublished Manuscript, Rand Corporation, Apr. 1976.
- [8] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 1st Ed.,

Wiley, 1994.

- [9] J. P. Mills, "Table of the ratio: Area to bounding ordinate, for any portion of normal curve," *Biometrika*, vol. 18, no. 3/4, pp. 395-400, 1926.
- [10] K.-G. Lee, et al., "Propagation characteristics of suburban environments using hybrid ray-tracing simulation," in *Proc. IEEE VTC*, Montreal, Canada, Sept. 2016.
- [11] Y. Kim, Y. Jeon, and I. Chong, "Device objectification and orchestraion mechanism for IoT intelligent service," *J. KICS*, vol. 38C, no. 1, pp. 19-32, Jan. 2013.
- [12] B. Shim and B. Lee, "Evolution of MIMO technology" *J. KICS*, vol. 38A, no. 8, pp. 712-723, Aug. 2013.
- [13] K.-T. Kim and H.-D. Park, "Empirical propagation path loss model for ATC telecommunication in the concourse environment" *J. KICS*, vol. 38A, no. 9, pp. 765-772, Sept. 2013.
- [14] R. Breen, *Regression Models: Censored, Sample Selected, or Truncated Data*, Thousand Oaks: Sage Publications, 1996.
- [15] Republic of Korea, *Verification of ray-tracing simulation through the measurements*, Document 3K/54, ITU-R WP 3K meeting, ITU, Geneva, Jun. 2016.

이 경 규 (Kyung-gyu Lee)



2015년 2월 : 고려대학교 정보통신대학 컴퓨터·통신공학부 학사

2015년 3월~현재 : 고려대학교 일반대학원 컴퓨터·전파통신공학과 석·박사통합과정

<관심분야> 채널 모델, Ray-tracing, 공유 연구

오 성 준 (Seong-jun Oh)



1991년 2월 : 한국과학기술원 전기 및 전자공학과 학사

1994년 2월 : 한국과학기술원 전기 및 전자공학과 석사

2000년 9월 : University of Michigan, Ann Arbor, EE:Systems PhD

2000년 9월~2007년 8월 : Ericsson Wireless Communications, 와 Qualcomm CDMA Technologies 근무

2007년 9월~현재 : 고려대학교 정보보호대학원 교수

<관심분야> 무선네트워크, 통신 표준화, 모델 구현