選擇 모델의 開發과 엔트로피 最適化

金 炳 導*

《目 次》

Ⅰ. 序 言

情報 理論(information theory)은 無形의 개념인 정보를 數量化할 수 있는 가능성을 제시하여, 과거 50년 동안 통계학자를 중심으로 여러 분야의 학자들의 관심을 집중시켰다. 다음 장에서 보다 상세히 설명을 하겠지만 정보 이론의 핵심 개념인 엔트로피(entropy)는 어떤 주어진 確率 實驗에 어느 정도의 不確實性이 내재되어 있는가를 측정하는 척도라 정의할 수 있다. 즉, 이 불확실성을 줄이기 위해 우리는 정보를 필요로 한다.

본 연구의 첫째 목표는 최근 JASA에 발표된 Soofi(1992, 1994)의 논문을 보다 一般化시키는데 있다. Soofi는 위 논문들에서 엔트로피 極大化의 原則을 이용하여 대표적 選擇 모델(choice model)의 하나인 logit model을 도출하였다. 이는 경제 통계학자의 기존 logit model 도출 과정과는 달리 의사 결정자의 선택 행동에 대한 최소한의 가정만을 하였다는 점에서 그 의의를 찾을 수 있다. 본 연구는 Soofi의 결론에서 한걸음 더 나아가, logit model을 포함한 모든 종류의 random coefficient logit model을 엔트로피 극대화의 원칙을 이용하여 도출할 수 있음을 보였다. 이 과정에서 각 choice model이 도출되기 위해서 연구자는 어떤 종류의 부연(敷衍) 情報를 필요로 하는가를 도출하였다.

본 연구의 둘째 목표는 엔트로피 최적화의 개념을 재정립하여 연구자들이 여러 사회 현

* 서울大學校 經營大學 基金 助敎授

** 본 연구는 서울대학교 경영대학 경영연구소 연구비 지원에 의해 작성된 것임.

상에 엔트로피 개념을 적용하여 이를 분석할 때 이 개념이 바르게 적용될 수 있도록 유도하는데 있다. 엔트로피 개념은 熱力學(thermodynamics), 統計力學(statistical mechanics), 情報理論(information theory) 등 여러 학문 분야에 적용되면서 그 의미가 변천하게 되었다. 특히 사회과학 분야에 엔트로피 최적화의 개념이 적용되면서 그 적용에 많은 문제점을 야기하게 되었다(Horowitz and Horowitz 1976). 이에 엔트로피 최적화의 개념을 사회 현상에 적용하고자 하는 연구자들은 이 개념의 정확한 이해가 필수적이다.

위의 두 목표를 가진 본 연구의 구성은 다음과 같다. 다음 장에서는 엔트로피의 定義와 屬性을 간단한 예와 함께 설명한다. 제 3 장에서는 먼저 choice model의 간략한 설명과 경제 통계학자의 logit model 도출 과정을 알아본다. 그리고 동일한 logit model이 엔트로피 極大化 原則下에서는 어떻게 도출되는가를 보고 그 적용을 random coefficient logit model로 발전시킨다. 또한, 情報 指數에 대한 개념을 소개하고 交叉 엔트로피 最小化의 개념을 설명한다. 제 4 장에서는 전 장에서 소개된 엔트로피의 원리들이 實證 事例에서 어떻게 적용되는가를 보이기 위해 scanner data분석에 엔트로피의 원리를 적용하였다. 제 5 장은 본 연구의 시사점과 未來 方向을 제시하면서 본 연구를 맺는다.

Ⅱ. 엔트로피 最適化

엔트로피(entropy)란 무엇인가? 엔트로피는 원래 熱力學에서 開發된 개념이었으나 Shannon(1948)과 Jaynes(1957)에 의해 再發見되어 이후 情報理論 등 여러 학문 분야에 지대한 영향을 미치게 되었다. 이들은 엔트로피의 개념과 不確實性(uncertainty) 또는 情報(information)의 개념이 서로 밀접한 관계에 있음을 보였고, 이전까지는 常識 수준에 머물러 있던 정보 개념을 數量化하는 方法論을 제시하였다. 이후 여러 학자들에 의해 엔트로피의 개념을 정립시키고자 하는 노력이 있었는데, 본 연구에서는 Martin과 England(1981)에 의해 정립된 엔트로피 개념을 기초로 하여 본 논문이 연구하고자 하는 엔트로피의 개념을 설명하고자 한다.

엔트로피를 이해하기 위해 먼저 情報函數(information function)를 정의하자. 즉, 어떤 確率實驗(probabilistic experiment)을 시행한다고 하자. 100원짜리 동전을 던져 동전의 앞면이 나오는지 또는 뒷면이 나오는지를 보는 것이 확률 실험의 例이다. 정보 함수란 주어진 확률 실험에서 어떤 事件(event)이 發生하였을 때 이 사건이 실험자에게 주는 情報의 量을 測定하는 尺度이다. 즉, 동전을 한번 던졌을 때의 사건은 두 종류이다. 이 두 사

건이 일어날 확률이 동일하다면, 앞면이 나왔을 경우와 뒷면이 나왔을 경우 실험자가 얻는 정보의 양은 동일하다. **Martin**과 **England**는 정보 함수가 지녀야 할 가장 기본적인 두 가지 屬性을 제시하는 데 이는 다음과 같다.

첫째, $P(E_1) \geq P(E_2)$이면 $I(E_1) \leq I(E_2)$이다. 여기서 $P(E_i)$는 사건 i가 일어날 確率을 指稱하며, $I(E_i)$는 사건 i가 주는 정보의 양을 측정하는 정보 함수이다. 즉, 정보 함수는 一方向的(**monotonicity**) 性格을 갖는다. 다시 말하자면, 觀測者는 확률이 큰 사건이 발생했을 때보다 확률이 적은 사건이 발생했을 때 보다 놀라게 되는데 이는 그 사건이 가지고 있는 정보량이 많기 때문이다. 이해를 돕기 위해 福券 추첨의 예를 들어 보기로 한다. 어떤 복권이 당첨될 확률이 0.0001이고 당첨되지 않을 확률이 0.9999이라고 하자. 복권이 당첨되지 않는 사건이 발생했다면 이는 그리 놀라운 일이 아닐 것이다. 반면에, 복권이 당첨되었다면 이는 매우 놀라운 사건이며 이 사건이 가지는 정보의 가치는 크다.

정보 함수가 지녀야 할 두번째 속성은, 만약 $E_1 \perp E_2$이면 $I(E_1 \cap E_2) = I(E_1) + I(E_2)$이다. 이를 정보이론(**information theory**) 연구자들은 정보 함수의 加法的(**additivity**) 性格이라 이른다. 즉, 두 獨立된 사건을 同時에 관측했을 때 얻는 정보의 양은 각 사건을 따로 관측했을 때 얻는 정보의 양의 합과 같다는 것이다.

連續 函數(**continuous function**) 중에서 위의 두 속성을 동시에 만족시키는 함수의 형태는 어떠하여야 하는가? **Khinchin**(1957)을 포함한 여러 학자들은 로그 함수만이 위의 두 속성을 동시에 만족하는 연속 함수임을 증명하였다. 그러므로, 정보 이론 학자들은 정보 함수의 형태로 (자연) 로그 함수를 쓴다. 예를 들면, $I(\pi) = -\ln \pi$.

확률 실험에서 관측자는 實驗 前까지는 어떤 사건이 실현될 것인가를 모르고 단지 각 사건이 일어날 확률만을 안다. 그러므로 주어진 확률 실험 자체의 정보의 양을 (또는 불확실성의 정도를) 측정하기 위해서는 위의 정보 함수의 期待値를 계산하여야 한다. 즉, $H(\pi) = -E\pi[\ln \pi]$. 그러므로, π의 確率 分布 函數(**probability density function**)가 연속 함수라면 정보함수의 기대치는

$$H(\pi) = -\int_{\pi} \pi \ln \pi \tag{2.1}$$

반면에 π가 離散的 變數(**discrete variable**)라면 정보함수의 기대치는

$$H(\pi) = -\sum_{i=1}^{n} \pi_i \ln \pi_i \tag{2.2}$$

위의 확률 실험에서의 정보 함수의 기대치인 $H(\pi)$가 바로 열역학론에서 쓰여 지고 있

던 엔트로피의 함수 형태와 동일하지만 그 해석은 서로 相異함을 注目하여야 한다. 본 論文에서는 엔트로피를 어떤 주어진 확률 실험에 어느 정도 불확실성이 내재되어 있는가를 측정하는 척도라고 解釋한다. 위의 100원짜리 동전을 던지는 확률 실험에서 앞면과 뒷면이 나올 확률이 각각 0.5라면 $H(\pi) = -(0.5 \times \ln 0.5 + 0.5 \times \ln 0.5) \approx 0.693$. 반면에 위의 복권의 경우는 $H(\pi) = -(0.0001 \times \ln 0.0001 + 0.9999 \approx \ln 0.9999)$ 0.001. 즉, 복권 확률 실험의 경우는 엔트로피가 낮고 따라서 불확실성이 거의 내재되있지 않다고 할 수 있다. 이 경우 우리는 어떤 사건이 발생할 것인가에 대해 거의 확실히 예측이 가능하다. 그러나, 동전 실험은 엔트로피가 높고 따라서 그 실험 자체의 불확실성이 매우 높다고 할 수 있다.

Ⅲ. 選擇 모델(Choice Models)

1. 選擇 모델의 發達

선택 모델의 起源에 대해 학자들 간의 다소의 異見은 있겠으나 心理學者였던 Thurston(1927)이 최초로 現代的 意味의 선택 모형을 提示했다고 보는 것이 妥當할 것이다. Thurston의 모델을 어떤 소비자의 자동차 購買時 brand의 선택의 예로 설명하기로 하자. 소비자는 자동차를 구매할 때 먼저 각 代案(alternative or brand)들의 效用을 계산하고 그 중 효용이 가장 큰 代案을 선택한다. 그러나, 각 대안들의 효용은 確率變數(random variable)이기 때문에 硏究者는 백 퍼센트 정확하게 소비자의 選擇 代案을 豫測할 수는 없고 단지 각 대안을 선택할 確率을 導出할 수 있다. 이와 같은 Thurston의 確率的 選擇모델(probabilistic choice model) 理論은 여러 학문 분야의 선택 모델 발전에 지대한 영향을 주게 된다. Thurston의 논문 이후 數理心理學(mathematical psychology)에서는 公理的 接近(axiomatic approach)을 함으로써 여러 확률적 선택 모델을 도출하기 위한 이론적 기틀을 만드는 데에 주력하였다(Luce 1959; Luce and Suppes 1965; Tversky 1972). 반면에 生物 統計學者(biometrician)들의 관심은 實在 實驗 狀況에서 從屬 變數(dependent variable)의 성격이 전형적인 回歸 分析法을 적용하기가 곤란할 때의 통계학적 방법론을 개발하는 것이었는데 (예를 들면, 선택모델에서 소비자의 brand 선택 변수는 종속변수이다). 그 결과 logit model을 기본 모델로 하는 여러 離散 또는 質的 反應모델(discrete or qualitative response model)의 統計學으로 발전하게 된다(Berkson 1944; Cox 1970).

1970년대에 들어서자 經濟 統計學者(econometrician)들은 경제학의 여러 문제를 檢證하는데 있어서 數理 心理學과 生物 統計學에서 개발된 선택 모델을 적용하게 된다(McFadden. 1974; Ben-Akiva and Lerman 1985). 이들의 연구는 선택모델의 이론적 측면과 反應函數 導出, 變數 推定, 豫測 등의 통계학적 문제를 하나로 묶는 틀을 제시하였다는 점에서 그 의의를 찾을 수 있다. 그 이후 선택모델은 다양한 경제학적 선택 문제에 적용되는데, 그 대표적 예를 들면 學生의 大學 선택 문제, 職業의 선택, 住居 地域 및 住宅 種類의 선택, 出産 子女數의 선택, 자동차 등 소비재 상품 구매에서 brand의 선택 등등이다. 이는 경제학의 가장 기본적인 관심사의 하나가 인간의 선택 행위를 이해하는 데에 있다는 점에서 그리 놀라운 일은 아니다. 여기서 우리는 다음 節의 주제인 엔트로피 개념을 이용한 선택 모델 도출과의 비교를 돕기 위하여, McFadden의 논문을 중심으로 경제 통계학자들이 提示하는 확률적 선택 모델을 簡略하게 敍述하고자 한다.

자동차 구매를 예로 들자면, 어떤 소비자가 자동차를 구매하기 위해 J개의 brand를 考慮한다고 하자. 특정 brand 선택을 위해 각 brand의 효용을 算出하여야 하는데, 이 소비자의 brand j에 대한 效用 函數는 다음과 같다고 하자.

$$U_j = \overline{U}_j + \varepsilon_j \tag{3.1}$$

위의 式에서 $\overline{U}_j$는 brand j의 豫測이 可能한 效用(deterministic utility)이고 ε_j는 豫測이 不可能한 效用(random utility)이다. 즉, 연구자는 어떤 소비자의 일련의 선택 행위를 관측하고 모델을 설정하여 그 소비자의 선택 행위를 설명하고자 한다. 그러나, 대부분의 경우 소비자의 선택 행위는 연구자의 모델에 의해 백 퍼센트 설명될 수는 없다. 이렇게 연구자의 모델에 의해 설명할 수 없는 부분의 효용을 random utility라 이른다. 반면에, deterministic utility는 연구자의 모델에 의해 설명되어지는 부분이므로 우리는 $\overline{U}_j$를 다음과 같이 정의할 수 있다.

$$\overline{U}_i = \beta' x_j \quad \text{where} \quad \beta = [\beta_1, \cdots, \beta_k]' \quad \text{and} \quad x_j = [x_{1j}, \cdots, x_{kj}]' \tag{3.2}$$

위 式에서 벡터 x_j는 brand j의 製品 屬性 (예를 들면, 제품 가격, 크기, 製造業體의 이름, 馬力, 등등)이나 消費者의 特性으로 (예를 들면, 소득 수준, 교육 정도, 성별, 나이, 등등) 구성된 獨立 變數이고, 벡터 b는 소비자의 구매 데이터를 觀測한 후 연구자가 推定하여야 할 一群의 母數(parameter)이다.

위의 效用函數를 갖는 소비자는 자신의 효용을 極大化하는 brand를 선택한다. 즉, 경

제학의 기본 假定에 부합되는 合理的 消費者는 다음의 원칙에 따라 brand를 선택한다.

$$\text{Choose brand } j \text{ if } U_j > U_m \text{ for all } m \neq j \tag{3.3}$$

그러므로, 위 소비자가 brand j를 선택할 확률은 다음과 같이 계산할 수 있다.

$$P_j = \Pr(U_j > U_m \text{ for all } m \neq j) = \Pr(\varepsilon_m - \varepsilon_j < \overline{U}_j - \overline{U}_m \text{ for all } m \neq j) \tag{3.4}$$

위의 式(3.4)에서 McFadden은 ε_j의 確率分布函數(probability density function)가 Weibull 分布를 따른다고 假定하면 multinomial logit model이 導出될 수 있고, ε_j의 確率分布函數 가 正規分布를 따른다고 假定하면 multinomial probit model이 도출될 수 있음을 보였다.

Multinomial logit model

$$P_j = \frac{\exp(\overline{U}_j)}{\sum_{m=1}^{J}\exp(\overline{U}_m)} = \frac{\exp(\beta' X_j)}{\sum_{m=1}^{J}\exp(\beta' X_m)} \tag{3.5}$$

Multinomial probit model

$$P_j = \int_{-\infty}^{\overline{U}_j-\overline{U}_1}\int_{-\infty}^{\overline{U}_j-\overline{U}_2}\cdots\int_{-\infty}^{\overline{U}_j-\overline{U}_{j-1}}\int_{-\infty}^{\overline{U}_j-\overline{U}_{j+1}}\cdots\int_{-\infty}^{\overline{U}_j-\overline{U}_J} f(\eta_1, \eta_2, \cdots, \eta_{j-1}, \eta_{j+1}, \cdots, \eta_J) \tag{3.6}$$

where $f(\eta_1, \eta_2, \cdots, \eta_{j-1}, \eta_{j+1}, \cdots, \eta_1)$은 $(J-1)$ variate multivariate正規分布函數이다.

2. Logit model의 導出: 엔트로피 極大化

Logit이나 probit 등의 선택모델을 도출하는 데 있어서 위의 경제 통계학적 접근법은 모델의 誤差(random errors) 또는 예측 불가능한 효용 부분이 어떤 特定한 確率分布를 따른다고 가정한다는 점에 留意하여야 한다. 즉, random errors가 實在로는 正規分布를 따르는데 logit model을 사용하여 소비자의 선택 행위를 설명하고자 한다면 이는 model specification error를 誘發한다는 것이다. 또한, 연구자는 데이터를 분석하기 전에 random errors의 확률분포 형태를 모르기 때문에 任意的으로 어떤 確率分布函數를 가정하여야 한다. 이 절에서는 엔트로피 개념을 이용하여 선택모델을 도출하는 색다른 방법을 提示함으로써, 위의 경제 통계학적 도출 방법의 短點을 補完하는데 그 목적이 있다. 이를

위해서 Jaynes(1957)에 의해 소개된 엔트로피 極大化의 原則(**maximum entropy principle, MaxEnt principle, ME principle**)의 이해가 필요하기 때문에 이의 설명을 하기로 한다.

n 種類의 서로 다른 事件(**event or outcome**)을 가진 어떤 確率實驗을 가정하자.(예를 들면, 주사위를 던져 어떤 숫자가 나오는가를 觀測하는 확률실험은 여섯 종류의 結果 또는 事件이 존재한다.) 이 확률 실험에서 각 사건이 일어날 확률은 다음과 같다.

$$p_1,\cdots,p_n \ \ here \ \ p_1 \geq 0,\cdots,p_n \geq 0 \ \ and \ \ \sum_{i=1}^{n} p_i$$

위의 확률 실험에서 p의 확률 분포에 대해 우리에게 주어진 유일한 情報는 그 合이 1이라는 사실뿐이다. 엔트로피 極大化 原則의 바탕에 깔려있는 개념은 주어진 확률 실험의 不確實性의 程度 또는 엔트로피를 측정할 때 우리는 우리에게 주어진 정보만을 사용하여야 한다는 것이다. 이를 다르게 표현하면, 엔트로피 극대화 원칙이란 주어진 정보를 다 사용한 후 남아 있는 確率實驗의 不確實性을 極大化하여야 한다는 것이다. 위의 예에서 주어진 유일한 정보는 각 확률의 합이 1이라는 것이다. 그러므로, 이 경우 엔트로피 극대화의 문제는 확률의 합이 1이라는 制約 條件을 充足시키는 가장 不確實한 確率分布를 발견하는 것이다. 이를 충족시키는 p의 확률 분포는 uniform distribution이다.

$$p = \left(\frac{1}{n},\ \frac{1}{n},\cdots,\frac{1}{n}\right)$$

위의 **uniform distribution**은 n種類의 事件을 가지고 있는 모든 確率分布函數 중에서 가장 不確實性이 큰 (또는 엔트로피를 極大化하는) 分布函數이다. 이러한 점에서 엔트로피 極大化 原則은 라플라스의 不充分 理由의 原則을(**Laplace's principle of insufficient reason**) 一般化한 原則이라고 할 수 있다 (**Kapur and Kesavan** 1992). 즉, 라플라스의 원칙에 따르면, 위의 예에서 우리가 가지고 있는 유일한 정보는 확률의 합이 1이라는 사실 뿐인데 **uniform distribution** 以外의 다른 確率分布函數를 선택할 充分한 理由가 없기 때문에 **uniform distribution**이 가장 타당하다는 것이다.

위에서 敍述한 엔트로피 극대화 원칙을 바탕으로 意思 決定者의 選擇 問題를 설명하는 確率分布函數를 도출해 보기로 하자. 우리의 목표는 아래에서 定義할 確率實驗의 엔트로피를 극대화하는 確率分布函數를 도출하는 데에 있다. 경제 통계학자들이 도출한 선택모델과 比較를 容易하게 하기 위해 다음과 같은 意思決定 問題를 假定한다. 전체 소비자의

數는 總 I명이고, 이 소비자들은 각각 K번의 brand 선택 의사결정을 하며, J種類의 brand가 존재한다고 하자. 이 경우 I명 전체 소비자의 모든 선택 의사결정에 대한 총 엔트로피는 다음과 같다.

$$H(\pi) = -\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}\log(\pi_{ikj}) \tag{3.7}$$

위의 式(3.7)에서 는 소비자 i가 k번째 선택 의사결정에서 brand j를 선택할 확률을 의미한다. 엔트로피 극대화 원칙에 따라 위의 엔트로피를 극대화하는 確率分布函數를 도출하는데 우리는 몇 가지 制約 條件 또는 情報를 가지고 있다고 하자. 그 첫째는 우리는 確率分布函數의 도출을 원하므로 확률의 합이 1이 되어야 한다는 아주 기본적인 情報이다.

$$\sum_{j=1}^{J}\pi_{ikj} = 1 \text{ for all } i(i = 1,\ldots,I) \text{ and } k(k = 1,\ldots,K) \tag{3.8}$$

式(3.8)만이 우리가 가진 유일한 정보라면 위에서 설명하였듯이 uniform distribution이 式(3.7)을 極大化하는 確率分布函數이다.

이제는 式(3.8)에 附加하여 다른 정보를 가지고 있다고 가정하자. 즉, 두번째 制約條件은 獨立變數들의 (確率加重)平均값을 안다는 것인데 이를 수식으로 쓰자면 다음과 같다.

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}x_{ikj} = X \text{ where } x_{ikj} = \left[x_{ijk}^{1},\ldots,x_{ikj}^{q}\right]' \text{ and } X = \left[X^{1},\ldots,X^{q}\right]' \tag{3.9}$$

위의 式(3.9)에서 우리는 q개의 獨立變數를 정의하였는데 이들 獨立變數들에 대한 정보가 소비자 選擇行動의 豫測을 가능하게 하고 그 결과 確率實驗의 불확실성을 감소시키는 것이다.

위의 式(3.8)과 式(3.9)의 두 式을 制約條件으로 하면서 式(3.7)의 엔트로피를 극대화하기 위해 문제를 풀기 위해 Lagrange function을 정의하면 이는 다음과 같다.

$$L(\pi) = -\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}\ \ln(\pi_{ikj}) + \sum_{i=1}^{I}\sum_{k=1}^{K}\left[\ln\frac{e}{Z_{ik}}\sum_{j=1}^{J}(\pi_{ikj} - 1)\right] - \beta'\left[\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}x_{ikj} - X\right]$$

위의 式에서 Z_{ik}는 式(3.8)의 制約條件에 따른 Lagrange multiplier로써 確率分布函數의 도출을 위한 常數이다. 이 常數를 統計力學에서는 partition function이라 부른다.

위의 Lagrange function을 극대화하는 確率分布函數를 구하기 위해 이의 一次 最適化 條件을 모든 $i(i = 1, \ldots, I)$, $k(k = 1, \ldots, K)$, $j(j = 1, \ldots, J)$에 대하여 구하면 다음과 같다.

$$\frac{\partial L(\pi)}{\partial \pi_{ikj}} = -\ln(Z_{ik}\pi_{ijk}) - \beta' x_{ikj} = 0 \text{ 즉, } \pi_{ijk} = e^{-\beta' x_{ikj}} / Z_{ik} \tag{3.10}$$

式(3.10)을 式(3.8)에 代入함으로써 Z_{ik}의 값을 구할 수 있고, 여기서 구한 Z_{ik}의 값과 式(3.10)으로부터 우리는 p의 確率分布函數를 도출할 수 있다.

$$\pi_{ikj} = \frac{\exp(\beta' x_{ikj})}{\sum_{m=1}^{J} \exp(\beta' x_{ikm})} \quad where \quad \beta = \left[\beta_1, \ldots, \beta_q\right]' \tag{3.11}$$

위의 式(3.11)은 경제 통계학자들이 도출한 multinomial logit model과 그 形態가 一致한다(Soofi 1992, 1994). 그러나, 위의 엔트로피 극대화 원칙을 이용한 logit model 도출은 의사 결정자의 選擇行動에 대한 最小의 假定만을 사용하였다는 데에 그 意義를 찾을 수 있다. 즉, 경제 통계학자의 도출과는 달리 연구자의 모델 誤差(random error)에 대한 어떤 確率分布상의 가정도 하지 아니하였다는 점을 注目하여야 한다. 엔트로피 극대화 원칙을 이용한 모델 도출은 연구자가 어떤 종류의 정보를 가지고 있다면 이를 수량화하여 제약 조건으로 취급한다. 즉, 확률의 합이 1이 되어야 한다는 정보만을 가지고 있을 때는 우리는 uniform distribution이 엔트로피를 극대화하는 확률분포였는데 반해, 각 독립 변수의 加重 平均값에 대한 정보를 입수함으로써 logit probability가 엔트로피를 극대화하는 확률분포가 되었다.

Soofi에 의해 試圖된 위의 logit model 도출은, 엔트로피 극대화 문제에 있어 어떤 종류의 制約條件 또는 情報가 logit model의 적용에 最小限으로 필요한가를 보여주었다. 우리는 여기서 위의 도출 과정을 보다 一般化시켜, Soofi의 logit model을 포함한 최근 經營學과 經濟統計學에서 볼 수 있는 여러 종류의 logit model을 도출해 보기로 한다. 이를 위해 式(3.9)의 두 번째 제약 조건을 약간 변형시켜 전체 소비자에 대한 독립 변수들의 加重平均대신 각 소비자 i에 대한 독립 변수들의 加重 平均을 안다고 가정하고 이를 數式으로 표현하면 다음과 같다.

$$\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}x_{ikj} = X_i \quad \text{where} \quad x_{ikj} = \left[x_{ijk}^{1},\ldots,x_{ikj}^{q}\right]' \quad \text{and} \quad X_i = \left[X_i^{1},\ldots,X_i^{q}\right]' \qquad (3.12)$$

이 새로운 제약 조건, 式(3.12)와 위의 式(3.8) 두 式을 制約條件으로 하면서 式(3.7)의 엔트로피를 극대화하는 確率分布函數를 도출하면 다음과 같다.

$$\pi_{ikj} = \frac{\exp(\beta_i' x_{ikj})}{\sum_{m=1}^{J}\exp(\beta_i' x_{ikm})} \quad \text{where} \quad \beta_i = \left[\beta_{i1},\ldots,\beta_{iq}\right]' \qquad (3.13)$$

각 소비자 i에 대한 독립 변수들의 加重 平均을 아는 순간, 우리는 각 소비자가 서로 다른 母數(parameter)를 가질 수 있는 logit model을 도출할 수 있다. 式(3.13)의 logit model을 학자들은 individual specific logit model이라 부르는데 이 경우 우리는 각각의 소비자에게 logit model을 적용하여 각 소비자의 母數를 推定하게 된다.

위의 式(3.13)에서 각 소비자가 서로 다른 母數를 가져야 한다는 사실에 대해서는 선택모델을 연구하는 학자들간에 異見이 없을 것이다. 예를 들면, 어떤 소비자는 각 brand의 가격에 敏感하게 反應하여 brand를 선택하는 반면, 어떤 소비자는 어느 한 brand에 대한 loyalty가 매우 커서 각 brand의 가격에 鈍感할 수 있다. 그러나, 위의 individual specific logit model의 문제점은 推定하여야 할 母數의 數가 너무 많다는 사실이다. 즉, 각 소비자에 대한 母數를 추정해야 할 경우, 母數 추정에 이용할 수 있는 데이터의 수가 너무 적어 지기 때문에 母數 推定値의 信賴度가 매우 낮게 되며, 극단적인 경우는 母數 추정이 불가능하게 된다(Albert and Anderson 1984; Rossi and Allenby 1993).

이와 같은 individual specific logit model의 문제점을 補完함과 同時에 소비자의 相異한 反應 程度를 인정하는 선택모델의 개발이 최근 십여년 동안 선택모델을 연구하는 학자들간에 많은 관심을 집중시켰다. 이 努力의 대표적인 모델이 random coefficient logit인데 이를 엔트로피 극대화 원칙에 따라 도출해 보기로 한다. 이를 위해 式(3.12)의 두 번째 제약 조건을 약간 변형시켜 각 소비자 i에 대한 독립 변수들의 加重 平均 자체를 아는 대신 그 確率分布를 안다고 가정하자. 이 정보를 數式으로 표현하면 다음과 같다.

$$\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}x_{ikj} = \tilde{X} \quad \text{where} \quad x_{ikj} = \left[x_{ikj}^{1},\ldots,x_{ikj}^{q}\right]' \quad \text{and} \quad \tilde{X} = \left[X_i^{1},\ldots,X_i^{q}\right]' \qquad (3.14)$$

위의 式(3.14)에서 벡터 $\tilde{X}$는 確率變數의 벡터(a vector of random variable)이다. 예를

들면, 確率變數 벡터 $\tilde{X}$는 그 平均이 μ이고 variance-covariance matrix가 Ω인 multivariate normal distribution이라고 가정하자. 이 경우 우리는 각 소비자 i에 대한 獨立變數들의 加重 平均값은 모르지만, 그 값이 우리가 그 성격을 알고 있는 multivariate normal distribution으로부터 無作爲로 抽出된 값과 같다는 정보를 가지고 있다는 것이다. 즉, 그 정보의 量이 式(3.12)의 경우와 같이 많지는 않지만 式(3.9)으로 주어진 정보의 量보다는 많은 경우이다. 새로운 制約條件인 式(3.14)와 위의 式(3.8)의 制約條件을 만족하면서 式(3.7)의 엔트로피를 극대화하는 確率分布函數를 도출하면 다음과 같다.

$$\pi_{ikj} = \frac{\exp(\tilde{\beta}'X_{ikj})}{\sum_{m=1}^{J}\exp(\tilde{\beta}'X_{ikm})} \quad \text{where } \tilde{\beta} = \left[\tilde{\beta}_1, \ldots, \tilde{\beta}_q\right]' \tag{3.15}$$

위의 式(3.15)에서 벡터 $\tilde{\beta}$는 random coefficient로써 그 分布函數는 $\tilde{X}$의 分布 函數에 따라 결정된다. 예를 들면, 만약 式(3.14)에서 주어진 의 分布函數가 multivariate normal distribution이라면 $\tilde{\beta}$도 역시 multivariate normal distribution이 된다. 이와 같은 형태의 random coefficient logit model에서 우리는 소비자 개개인의 母數를 推定하는 대신에 確率分布函數의 母數를 推定한다(Albert and Chib 1993). 반면에, 만약 式(3.14)에서 주어진 $\tilde{X}$의 分布函數가 Snumber of mass-point distribution이라고 하면, 母數 벡터 $\tilde{\beta}$의 分布函數는 Snumber of mass-point distribution이 된다. 이와 같은 형태의 logit model은 마케팅을 중심으로 많이 쓰이는 選擇모델인데(Kamakura and Russell 1989) 이의 여러 가지 변형된 형태에 대하여는 紙面 관계상 省略하기로 한다.

3. 情報에 의한 不確實性의 減少

第 2 節의 例에서 볼 수 있었듯이 어떤 確率實驗에서 확률의 合이 1이라는 사실 이외에 아무 정보를 가지고 있지 않다면, 엔트로피를 극대화하는 確率分布函數는 uniform distribution이다. 그러므로, 이 때의 엔트로피는 다음과 같다.

$$H(\pi_0) = -\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}\log(\pi_{ikj}) = -\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{J}\log\left(\frac{1}{J}\right) = I\cdot K\cdot\log J \tag{3.16}$$

우리는 logit model의 도출에서 볼 수 있었듯이, 새로운 정보를 入手함으로써 위 確率實驗의 엔트로피를 減少시킬 수 있었다. 예를 들면, 式(3.9)로 표현된 q개의 獨立 變數

들의 加重 平均값에 대한 정보를 가지고 있다면 전형적인 logit model인 式(3.11)을 도출할 수 있었다. 이와 같은 logit probability로 표현된 確率分布函數는 위 確率實驗의 엔트로피를 $H(\pi_0)$로부터 $H(\pi)$로 減少시킨다. 이 때 엔트로피 감소의 폭이 q개의 獨立 變數들의 加重 平均값에 대한 情報의 價値를 測定한다고 할 수 있다. 즉, 情報理論을 연구하는 학자들은, 어떤 특정 모델에 의한 確率實驗의 不確實性 減少의 程度를 測定하는 尺度로써 情報指數(Information Index)라는 척도를 다음과 같이 定義한다.

$$I(\pi) = \frac{H(\pi_0) - H(\pi)}{H(\pi_0)} = 1 - \frac{H(\pi)}{H(\pi_0)} \tag{3.17}$$

위의 情報指數는 $0 \leq I(\pi) < 1$ 의 성격을 가지고 있는데, 예를 들어 $I(\pi) = 0$라면 새로운 附加情報를 가진 모델이 주어진 確率實驗의 不確實性을 전혀 감소시키지 못하였다는 것이다. 즉, 이 새로운 情報의 價値는 0이라는 것이다. 반면에, $I(\pi) \approx 1$의 경우는 $H(\pi) \approx 0$을 意味하므로, 새로운 모델이 確率實驗의 결과를 거의 백 퍼센트 豫測한다는 것을 意味한다. 자동차 구매의 예로 돌아가자면, 연구자가 logit model을 사용하여 각 소비자의 선택 brand를 예측한다고 하자. 만약 Logit model의 母數를 推定한 후 계산한 $H(\pi)$의 값이 0에 가까웠다면 이 의미는 연구자의 logit model이 각 소비자가 선택하는 자동차 brand를 거의 백 퍼센트 정확하게 예측했다는 것을 意味한다. 이러한 의미에서, 情報指數는 연구자의 모델의 適合度(goodness-of-fit)를 測定할 수 있는 尺度인 셈이다. 이는 線型回歸分析에서의 R^2의 개념이나 McFadden의 Pseudo-R^2와 類似한 개념으로 解釋할 수 있다 (Soofi 1992).

4. 交叉 엔트로피의 極小化

이 절의 목표는 엔트로피 개념이 적용될 수 있는 또 다른 중요한 分野를 紹介하면서 交叉 엔트로피 極小化라는 原則을 설명하고자 한다. 第 2 節에서 論議된 確率實驗의 경우는 연구자가 確率分布 自體에 대한 아무런 정보를 가지고 있지 않았고, 이 경우 우리는 주어진 制約條件 下에서 엔트로피를 極大化하는 確率分布를 찾았다. 그러나, 연구자가 事前에 確率分布 自體에 대한 어떤 정보를 가지고 있는 경우가 있다. 맥주 구매의 경우 소비자의 brand 선택의 예를 들어 보자. 어떤 연구자가 맥주 구매 의사를 가지고 있는 無作爲로 抽出된 1,000명의 소비자 개개인이 어떤 brand를 선택하는가를 예측하려 한다고 하자. 맥주 소비자와 市場에 대한 아무 정보가 없다면 라플라스의 不充分 理由의 原則에 따라

모든 소비자에게 uniform distribution을 적용하여야 한다. 즉, 시장에 J종류의 brand가 존재한다면, 모든 소비자의 각 brand를 선택할 確率은 $1/J$이다. 以後 각 소비자의 구매 행위를 상당 기간 동안 觀測함으로써 수집된 데이터에 logit model등의 選擇모델을 적용하여 각 소비자의 brand선택을 보다 정확히 예측할 수 있다.

그러나, 위의 예에서 選擇모델을 적용하기 전에 연구자에게 確率分布 自體에 대한 어떤 정보가 있는 경우가 있다. 각 brand에 대한 市場 全體의 市場 占有率이 그 대표적인 예이다. 이와 같은 사전 정보는 여러 經路를 통해 얻을 수 있는데 다른 地域의 市場 經驗을 利用하는 것도 또 한가지 방법이 될 수 있다. 이 경우 선택 모델을 적용할 때 어떻게 시장 점유율이라는 사전 정보를 이용할 수 있을까?

위의 경우와 類似한 사전 정보를 統計學에서는 prior information이라 부르는데 이런 문제를 주로 다루는 統計學 分野가 Bayesian statistics이다. 交叉 엔트로피 極小化 原則의 이해를 돕기 위해 여기서 Bayesian statistician들이 사전 정보를 取扱하는 방법을 간략하게 설명하기로 한다.

먼저 어떤 假設에 대한 事前確率(prior probability)을 $p(H|I_0)$이라 하자. 여기서 I_0는 事前 情報를 의미한다. 우리는 새로운 정보 또는 데이터를 수집함에 따라 이 事前 確率을 修整하여야 하는데, 이 때 데이터로부터 얻는 假設에 대한 情報는 likelihood function 또는 $p(y|H)$로 요약할 수 있다. 이 式에서 y는 데이터를 의미한다. 그러므로, 데이터를 觀測한 後의 가설에 대한 事後確率(posterior probability)은 Bayes' theorem에 의해 다음과 같이 얻을 수 있다.

$$p(H|I_0,\ y) = p(H|I_0) \times p(y|H) \tag{3.18}$$

그러면, 엔트로피의 概念下에서 이와 같은 確率分布에 대한 事前 情報를 어떻게 이용할 수 있을까? 이 경우 쓰이는 엔트로피의 개념이 交叉 엔트로피 極小化의 原則 (minimum cross-entropy principle)이다(Shore and Johnson 1980). 이 原則의 核心은 만약 연구자가 確率分布에 대한 事前 情報를 가지고 있다면, 이 사전 정보와 데이터에서 얻을 수 있는 確率分布에 대한 情報의 差異를 極小化하여야 한다는 것인데, 우리는 여기서 交叉 엔트로피 極小化의 原則과 Bayesian statistics의 接近法에 類似性이 있다는 점을 발견할 수 있다.

第 2 節의 엔트로피 極大化 原則에서 든 例로 돌아가서, 이제 우리가 確率 分布에 대한 事前 情報가 있다고 가정하자. 즉, 소비자 i가 k번째 選擇 意思 決定에서 brand j를 선택

할 事前確率이 p_{ikj}라 하자. 만약 이 事前 情報가 市場 全體의 各 brand의 市場 占有率이라면 p_{ikj}는 brand j의 市場 占有率이 된다. 이 경우 p_{ikj}와 $_{pikj}$(소비자 i가 k번째 선택 의사결정에서 brand j를 선택할 확률)의 差異의 모든 i와 j와 k에 대한 合은 아래의 式(3.19)와 같이 Kullback-Leibler discrimination information function으로 測定될 수 있다.

$$KL(\pi : p) = \sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{j=1}^{J}\pi_{ikj}\ln\left(\frac{\pi_{ikj}}{p_{ikj}}\right) \tag{3.19}$$

위의 式(3.8)과 式(3.9)의 두 式을 制約條件으로 하면서 式(3.19)의 KL(π : p)를 極小化하는 確率分布函數를 구하면 이는 다음과 같다.

$$\pi_{ikj} = \frac{p_{ikj}\exp(\beta' x_{ikj})}{\sum_{m=1}^{J} p_{ikm}\exp(\beta' x_{ikm})} \quad where \quad \beta = \left[\beta_1, \ldots, \beta_q\right]' \tag{3.20}$$

위의 式(3.20)에서 만약 確率分布에 대한 아무 事前 정보가 없다면 라플라스의 不充分理由의 原則에 따라 $p_{ikj} = 1/J$가 되는데, 이 경우 $p_{ikj} = 1/J$를 위의 式(3.20)에 代入하면 우리는 式(3.11)과 同一한 式을 얻을 수 있다. 이 점에서 交叉 엔트로피 極小化의 原則은 엔트로피 極大化의 原則을 一般化한 原則이라고 할 수 있다. 그러나 이 경우 以外의 어떤 사전 정보가 존재한다면 式(3.20)의 變形된 logit model을 데이터에 적용함으로써 事後確率分布를 도출할 수 있어, Bayesian statistician들과 類似한 統計的 推論을 할 수 있다.

Ⅳ. Scanner Data分析의 例

우리는 前 章에서 情報 價値를 數量化하는 手段으로써의 엔트로피의 개념을 설명하였는데 이 章에서는 이를 實證 事例에서 어떻게 이용할 수 있는가를 보기로 한다.

이를 위해 우리는 A.C. Nielsen이 제공한 ERIM scanner panel데이터를 사용하기로 한다. Nielsen이 ERIM scanner panel data를 수집한 과정을 아주 간략히 설명하자면, 먼저 Nielsen은 美國 國民을 代表할 수 있는 도시 몇 개를 선택하여 각 도시마다 10,000명 정도의 居住 市民을 無作爲로 抽出하였다. 選擇된 家口들을 ERIM panelists라 부르는데 이들에게 일종의 전자 ID카드를 지급하여 이들로 하여금 물건을 구매할 때마다 소매점에

ID카드를 제출하도록 하였다. 각 小賣店들과 電算 通信網이 연결된 Nielsen은 각 ERIM panelist가 어떤 물건을 언제 얼마의 가격에 구입하였는지의 정보를 쉽게 얻을 수 있다. 이와 함께 소매점 및 제조업체의 여러 마케팅 활동의 정보도 同一 通信網을 이용하여 電送받을 수 있다.

우리는 이 時点에서 먼저 우리의 목표가 ERIM panelist의 一般 購買行爲의 분석에 있지 않다는 점을 밝혀둔다. 이 章의 우리의 목표는 단지 엔트로피의 개념을 어떻게 實證事例에서 이용할 수 있는가의 例를 드는데 있다. 그러므로, 우리는 美國 Missouri州의 Springfield市에 居住하는 ERIM panelist중에서 100명을 무작위로 선택하여 이들의 約 2年 동안의 tuna fish can 購買行爲를 분석하였다. Tuna fish can제품에는 수십 종류의 다양한 brand가 존재하지만 우리는 편의상 市場 占有率이 가장 높은 5種類의 brand만을 선택하였는데 이들 5個社의 brand가 전체 시장의 90% 정도의 市場 占有率을 차지하고 있다.(ERIM scanner panel data에 대한 보다 상세한 설명은 Kim, Blattberg, and Rossi(1995)를 參考하라.) 무작위로 선택된 100명의 家口에서 이들 上位 5個 brand의 市場 占有率 및 平均 價格은 다음 〈표 1〉과 같다.

위의 100명의 소비자의 구매 데이터에 multinomial logit model을 적용하면, 소비자 i가 k번째 選擇 意思決定에서 brand j를 선택할 확률은 다음과 같이 數式으로 표현할 수 있다.

$$p_{ikj} = \frac{\exp(\alpha_j + \beta \ln x_{ikj})}{\sum_{m=1}^{J} \exp(\alpha_m + \beta \ln x_{ikm})} \tag{4.1}$$

위의 式(4.1)에서 x_{ikj}는 소비자 i가 k번째 選擇 意思決定을 할 때의 brand j의 價格을 나타내며, $\alpha_j (j = 1, \ldots, J)$와 β는 주어진 데이터로부터 推定하여야 할 母數(parameter)

〈표 1〉 市場 占有率 및 平均 價格

	시장 점유율	평균 가격
Starkist Water	40.8	$0.70
Chicken-of-the-Sea Water	17.2	$0.74
Store Brand Water	8.1	$0.62
Starkist Oil	16.9	$0.74
Chicken-of-the-Sea Oil	17.0	$0.71

이다. 아래의 〈표 2〉는 100명의 소비자에 대한 母數 推定 結果를 요약한 표이다.

〈표 2〉의 추정에서 이용한 總 消費者의 數는 100명이고 총 選擇 意思決定의 數는 428번이다. 우리는 두 종류의 logit model을 推定하였는데 두 번째 行의 結果는 價格 情報를 사용하지 않은 추정 결과이고 세 번째 行의 結果는 價格 情報를 사용한 추정 결과이다. 〈표 2〉에서 괄호 속의 숫자는 각 推定値의 標準誤差(standard errors of estimates)를 나타내고, 우리는 logit model의 identification을 위해 Chicken-of-the-Sea Water의 intercept를 0으로 固定시켰다. 다른 用語로 表現하면, 다른 母數들은 0으로 固定된 Chicken-of-the-Sea Water의 intercept의 값에 相對的으로 推定되었음을 밝혀 둔다.

주어진 데이터의 情報 價値를 測定하기 위해 우리는 먼저 소비자 구매 데이터를 관측하기 전의 엔트로피를 式(3.16)에 의해 계산하면 $H(\pi_0) = 688.8$이다. 다음 段階로 購買時 各 brand의 가격은 모르지만 소비자가 어떤 brand를 구매하였는가의 정보는 가지고 있는 경우 우리는 두 번째 行의 logit model을 適用할 수 있고, 이 경우의 $H(\pi)$를 계산하면 $H(\pi) = 629.2$이 되어 아무 정보가 없을 때에 비해 不確實性이 約 9% 減少하였다. 즉, 이 model의 情報指數는 0.09이다. 반면, 購買時 각 brand의 가격을 관측할 수 있다면 우리는 세 번째 行의 logit model을 적용할 수 있고, 이 경우의 $H(\pi)$는 507.7이 되어 아무 정보가 없을 때에 비해 不確實性이 約 26% 감소되었다. 즉, 세 번째 行의 logit model의 情報指數는 26% 이다. 이와 같은 두 model간의 情報指數의 큰 差異는 각 brand의 가격에 대한 정보가 이 確率實驗의 確率分布函數를 이해하는데 결정적 역할을 한다는 것을 보여준다. 즉, 이 시장의 소비자들은 tuna fish can을 선택할 때 각 brand에

〈표 2〉 Logit model의 母數 推定 結果

	Model without price	Model with price
Intercept (α_j)	0.86(0.14)	0.94(0.15)
Starkist Water	0.00	0.00
Chicken-of-the-Sea Water	−0.75(0.18)	−1.57(0.21)
Store Brand Water Starkist Oil	−0.02(0.22)	−0.15(0.19)
Chicken-of-the-Sea Oil	−0.01(0.29)	−0.97(0.23)
log price (β)	—	−4.26(0.24)
Loglikelihood	−629.16	−507.73
Total number of households:	100	
Total number of observations:	428	

의 brand loyalty가 相對的으로 낮아 購買時 각 brand의 價格 條件에 의해 brand를 선택하는 경우가 많다는 것을 간접적으로 示唆한다.

우리는 위의 情報指數의 槪念은 旣存의 소비자 데이터를 이용하여 새로운 소비자의 購買行爲를 예측하는 경우에도 이용할 수 있다. 즉, 위에서 분석한 100명의 소비자 購買行爲를 바탕으로 다른 소비자들의 선택 brand를 예측하고자 할 때 우리는 旣存 소비자들의 各 brand 市場 占有率을 事前情報로써 이용할 수 있음을 留念하여야 한다. 이 경우 우리는 交叉 엔트로피 極小化 原則의 이용이 가능하다. 또는, 旣存 소비자 데이터로부터 推定된 logit model의 모든 母數 推定値를 事前情報로 이용하는 방법도 고려할 수 있다.

例를 들자면 위에서 설명한 ERIM scanner panel data로부터 100명의 새로운 소비자를 無作爲로 抽出하여 이들의 brand 選擇行動을 豫測하기로 하자. 새롭게 抽出된 100명의 소비자들의 總 購買 回數는 503回이었다. 만약 確率分布에 대한 아무런 정보가 없다면, 總 엔트로피는 式(3.15)에 의해 $H(\pi_0) = 807.9$가 된다. 다음 段階로, 만약 以前에 분석한 100명의 소비자로부터 얻은 정보 중에서 各 brand의 市場 占有率만을 사용한다고 하면, $H(\pi) = 729.3$이 되어 確率分布의 不確實性을 約 10% 감소시킬 수 있다. 즉, 다른 소비자로부터 얻은 各 brand의 市場 占有率은 0.10의 情報指數를 가진다. 반면에 旣存 소비자 100명의 데이터로부터 推定된 logit model의 모든 母數 推定値를 이용하면 H(p)는 645.5로 더욱 낮아져 이때의 情報指數는 約 20%이다. 마지막으로, 기존 데이터로부터 推定된 logit model의 모든 母數 推定値로 계산한 事前確率分布와 함께 새로운 100명의 데이터에 式(3.20)의 變形 logit model을 적용하면 엔트로피가 더욱 낮아져 $H(\pi) = 511.3$이 되어 우리는 約 37%의 不確實性을 줄일 수 있었다.

이와 같이 엔트로피 槪念은 어떤 情報가 確率實驗에서 不確實性을 줄이는데 있어 얼마나 價値가 있는가를 數量化할 수 있다는 점과 確率分布에 대한 事前情報를 容易하게 새로운 정보와 結合할 수 있다는 점에 있어 그 長點이 있다. 이러한 점에서, 경영학의 여러 문제에 엔트로피 槪念을 적용함으로써 경영학 문제의 이해를 深化할 수 있기를 바란다.

Ⅴ. 結 言

오래 전부터 여러 統計學者들은 Fisher(1921)의 논문을 始初로 情報 槪念의 數量化를 試圖하였다. Fisher가 定義한 情報 개념은 아주 좁은 의미의 정보 개념이지만 最初로 情報 槪念의 數量化를 試圖하였다는데 그 意義가 있다. 그에 따르면, 우리는 데이터로부터

母數를 推定한 後 推定値의 標本分布(sampling distribution of the estimate)를 도출할 수 있는데, 情報는 이 標本分布의 分散의 inverse로 定義할 수 있다는 것이다. 즉, 推定値의 分散이 크다면 데이터가 母數에 대한 情報를 거의 가지고 있지 않다고 解釋할 수 있다는 것이다. 이와 같은 Fisher의 정보 개념은 以後 엔트로피 개념을 그 中心 理論으로 하는 情報 理論에 의해 그 적용 범위가 확대되어 갔다.

이에 본 논문은 情報 理論의 核心 概念인 엔트로피 極大化 原則을 再照明하여 보다 많은 경영학 연구자들이 이 分野를 쉽게 接近할 수 있도록 노력하였다. 또한 최근 엔트로피 極大化의 原則을 이용하여 logit model을 도출한 Soofi의 논문을 보다 一般化시켜, Soofi가 도출한 logit model을 包含한 모든 種類의 random coefficient logit model이 엔트로피 極大化의 原則에 따라 도출될 수 있음을 보였다.

아울러 본 연구의 결과는 앞으로 여러 방향으로 발전시킬 수 있는 餘地가 있음을 밝혀 둔다. 첫째로 經營學과 經濟學에서 볼 수 있는 모든 種類의 選擇모델을 엔트로피 極大化의 原則에 따라 도출해 보는 것이 매우 意味가 있는 연구 方向이라 하겠다. 예를 들면, probit model이나 nested logit model등을 도출하기 위해서는 어떤 種類의 附加 制約條件이 必要한가? 이 논문에서 볼 수 있듯이 選擇모델 중에서 logit model이 필요로 하는 附加 情報의 量이 가장 적은 모델이기 때문에 probit model을 도출하기 위해서는 logit model이 필요로 하는 정보 이외의 어떤 附加 情報가 필요할 것이다. 이러한 導出 過程에서 우리는 이제까지의 연구에서는 볼 수 없었던 다른 종류의 選擇모델을 개발할 수 있는 가능성도 있음을 注目하여야 한다.

둘째로, 엔트로피 極大化의 原則을 적용할 수 있는 여러 經營學的 問題들을 발견하여 既存의 方法論과 比較해 보는 것도 興味로운 未來의 硏究 方向이다. 엔트로피의 既存의 여러 분야에의 應用에 대해서는 Horowitz and Horowitz(1976)와 Kapur and Kesavan (1992)에 비교적 잘 요약이 되어 있다. 그러나, 본 논문의 scanner data의 例에서 볼 수 있듯이 아직도 이의 새로운 適用 分野를 찾는 것은 그리 어렵지 않은 듯 하다. 이에 본 논문이 보다 많은 經營學 硏究者들이 이 분야에 關心을 가질 수 있는 契機가 되기를 바라는 바이다.

參考文獻

Albert, A. and J.A. Anderson (1984), “n the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, 71, 1, 1-10.

Albert, J.H. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of American Statistical Association*, 88, 669-79.

Ben-Akiva, M. and S. R. Lerman (1985), *Discrete Choice Analysis: Theory and Applications to Travel Demand*, Cambrige: The MIT Press.

Berkson, J. (1944), "Applications of the Logistic Function to Bioassay," *Journal of the American Statistical Association*, 39, 357-65.

Cox, D. R. (1970), *Analysis of Binary Data*, Methuen, London.

Fisher, R.A. (1921), "On Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, Series A, 309-68.

Hauser, J.R. (1978), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: an Information-Theoretic Approach," *Operations Research*, 26, 406-21.

Horowitz A.R. and I. Horowitz (1976), "The Real and Illusory Virtues of Entropy-Based Measures for Business and Economic Analysis," *Decision Sciences*, 7, 4, 121-36.

Jaynes, E.T. (1957), "Information Theory and Statistical Mechanics," *Physical Review*, 106, 4, 620-30.

Jaynes, E. T. (1982), "On the Rationale of Maximum Entropy Methods," *Proceedings of the IEEE*, 70, 939-52.

Kamakura, W. and G.J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26, 379-90.

Kapur, J.N. and H.K. Kesavan (1992), *Entropy Optimization Principles with Applications*, San Diego: Academic Press, Inc.

Khinchin, A.I. (1957), *Mathematical Foundations of Information Theory*, New York: Dover.

Kim, B., R.C. Blattberg, and P. Rossi (1995), "Modeling the Distribution of Price Sensitivity andImplications for Optimal Retail Pricing," *Journal of Business & Economic Statistics*, 13, 3, 291-303.

Luce, R. (1959), *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

Luce, R. and P. Suppes (1965), Preference, Utility and Subjective Probability," in *Handbook of Mathematical Psychology (Vol. 3)* R. Luce, R. Bush, and E. Galanter, eds., New York: Wiley.

Martin, N.F. G. and J.W. Englnad (1981), *Mathematical Theory of Entropy, Vol. 12 of Encyclopedia of Mathematics and Its Applications*, Mass.: Addison-Wesley.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. P. Zarembka, New York: Academic Press, 105-42.

Rossi, P.E. and G.M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30, 2 171-82.

Ryu, H. K. (1993), "Maximum Entropy Estimation of Density and Regression Functions"' *Journal of Econometrics*, 56, 397-440.

Shannon, C.E. (1948), "A Mathematical Theory of Communication," *Bell System Tech. J.*, 27, 379-423, 623-659.

Shore, J.E. and R. W. Johnson (1980), "Axiomatic Derivation of the Principle of Maximum Entropy

and the Principle of Minimum Cross-Entropy," *IEEE Transaction on Information Theory*, 26-37.

Soofi, E.S. (1992), "A Generalizable Formulation of Conditional Logit with Diagnostics," *Journal of the American Statistical Association*, 87, 412-16.

Soofi, E. S. (1994), "Capturing the Intangible Concept of Information," *Journal of the American Statistical Association*, 89, 1243-54.

Thurston, L. (1927), "A Law of Comparative Judgement," *Psychological Review*, 34, 273-86.

Tversky, A. (1972), "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 281-99.