魅力度モデルと集計データによる交差弾力性について

いくつかの背景知識

マーケティング変数の弾力性

目的変数（売上数量など）Q における X の点弾力性を e_Q とする。e_Q は以下のように定義される。

$$e_Q = \frac{Q\text{の相対的変化}}{X\text{の相対的変化}} = \frac{\frac{\Delta Q}{Q}}{\frac{\Delta X}{X}} = \frac{\partial Q}{\partial X} \bullet \frac{X}{Q}$$

ここで、

$$\frac{d(\log X)}{dX} = \frac{1}{X}$$

または、

$$d(\log X) = \frac{dX}{X} = X\text{の相対的変化}$$

なので、

$$e_Q = \frac{Q\text{の相対的変化}}{X\text{の相対的変化}} = \frac{d(\log Q)}{d(\log X)}$$

でもある。

e_Q に以下の逓減型対数線形モデルを代入する。

$$Q = e^{\alpha} \prod_{i=1}^{k} X_i^{\beta_i} \varepsilon$$

$$\log Q = \alpha + \sum_{i=1}^{k} \beta_i \log(X_i) + \log \varepsilon$$

$i = 1$ として、

$$Q = e^{\alpha} X^{\beta} \varepsilon$$

$$e_Q = \frac{\partial Q}{\partial X} \bullet \frac{X}{Y} = e^{\alpha} \beta X^{\beta-1} \varepsilon \bullet \frac{X}{e^{\alpha} X^{\beta} \varepsilon} = \beta$$

よって逓減型対数線形モデルのﾊﾟﾗﾒｰﾀは独立変数の目的変数に対する点弾力性となる。

確率分布の基礎

幾何分布

成功確率が p のﾍﾞﾙﾇｰｲ試行において、はじめて成功する時点を表す確率変数 X の従う分布は以下の幾何分布で表される。

$$P(X=k)=p(1-p)^{k-1}, k=1,2,.....$$

$$X \sim Geo(p)$$

ﾍﾞﾙﾇｰｲ試行の独立同一性から、幾何分布は無記憶性を持つ。

$$P(X-x>y \mid X>x)=P(X>y)$$

指数分布

幾何分布を連続形にした分布。密度関数は、

$$f(x)=\lambda e^{-\lambda x}, x \geq 0$$

分布関数は、

$$F(x)=1-e^{-\lambda x}, x \geq 0$$

$$T \sim Exp(\lambda)$$

である。

（注）連続的な確率変数の確率分布において以下の関係がある。

$P(a \leq X \leq b)=\int_a^b f(x)$　この時 $f(x)$ を密度関数と呼ぶ。

$$F(x)=\int_{-\infty}^{x} f(y)dy=P(X \leq x), x \in \Re$$

$$\frac{d}{dx}F(x)=f(x)$$

この時 $F(x)$ を分布関数と呼ぶ。「X が x 以下の確率」である。

指数分布は無記憶性を持つ唯一の連続関数である。

密度関数を $f(t)$、分布関数を $F(t)$ とする時、

$$h(x)=\frac{f(t)}{1-F(x)} \approx \frac{1}{dt}P(t<T \leq t+dt \mid T>t)$$

をﾊｻﾞｰﾄﾞ率という。$Exp(\lambda)$ のﾊｻﾞｰﾄﾞ率は一定である。（$h(x)=\lambda$）

ﾛｼﾞｽﾃｨｯｸ分布

以下の分布関数が与えられるとき、ﾛｼﾞｽﾃｨｯｸ分布に従うという。

$F(x) = \dfrac{1}{1+e^{-a(x-b)}}, x \in \Re, a > 0$

$X \sim L(a,b)$

密度関数は、

$f(x) = \dfrac{ae^{-a(x-b)}}{1+e^{-a(x-b)}} = aF(x)(1-F(x)), x \in \Re, a > 0$ で成長曲線である。

極値分布

$X_1, X_2, ...$ を同一の分布関数 $G(x)$ を持つ確率変数の列とし、

$Y_n = \max\{X_1, X_2, ..., X_n\} - \theta \log n$

とおく。Y_n の分布関数を $F_n(x)$ とすれば、

$F_n(x) = P(\max\{X_1, X_2, ..., X_n\} \le x + \theta \log n)$
$= [G(x + \theta \log n)]^n$

たとえば $X_i \sim Exp(1/\theta)$ の時、

$F_n(x) = \left(1 - e^{-(x+\theta \log n)/\theta}\right)^n = \left(1 - \dfrac{e^{-x/\theta}}{n}\right)^n$

これは、$n \to \infty$ の時 $\exp\left\{-e^{-x/\theta}\right\}$ に収束する。確率変数の最大値におけるこのような極限分布を極値分布という。極値分布には次の三種類がある。

タイプ I：二重指数分布
タイプ II：フレッシェ分布
タイプ III：ワイブル分布

二重指数分布

分布関数は、

$F(x) = \exp\left\{-e^{-\lambda(x-\mu)}\right\}, x \in \Re, \lambda > 0$

密度関数は、

$f(x) = \lambda \exp\left\{-\lambda(x-\mu) - e^{\lambda(x-\mu)}\right\}, x \in \Re, \lambda > 0$

これを、$X \sim DE(\lambda, \mu)$ と書く。

MCI モデル

以下のように定式化される。

$$A_i = \exp\{\alpha_i\}\prod_{k=1}^{K} X_{ki}^{\beta_k}\varepsilon_i$$

$$S_i = \frac{A_i}{\sum_{j=1}^{n} A_j}$$

対数中央化すると線形になる。まず両辺の対数を取る。

$$\log S_i = \alpha_i + \sum_{k=1}^{K}\beta_k \log X_{ki} + \log\varepsilon_i - \log\left\{\sum_{j=1}^{n}\left(\exp\{\alpha_j\}\prod_{k=1}^{K} X_{kj}^{\beta_k}\varepsilon_j\right)\right\}$$

i に関し合計し、n で割ると、

$$\log\tilde{S} = \overline{\alpha} + \sum_{k=1}^{K}\beta_k \log\tilde{X}_k + \log\tilde{\varepsilon} - \log\left\{\sum_{j=1}^{n}\left(\exp\{\alpha_j\}\prod_{k=1}^{K} X_{kj}^{\beta_k}\varepsilon_j\right)\right\}$$

ここで~のあるものは幾何平均、－のつくものは平均。上の式から下の式を引くと、以下が得られる。

$$\log\left(\frac{S_i}{\tilde{S}}\right) = \alpha_i - \overline{\alpha} + \sum_{k=1}^{K}\beta_k \log\left(\frac{X_{ki}}{\tilde{X}_k}\right) + \log\left(\frac{\varepsilon_i}{\tilde{\varepsilon}}\right)$$

なお、$\exp(\alpha_i) = a_i$ とおくと

$$\log\left(\frac{S_i}{\tilde{S}}\right) = \log\left(\frac{a_i}{\tilde{a}}\right) + \sum_{k=1}^{K}\beta_k \log\left(\frac{X_{ki}}{\tilde{X}_k}\right) + \log\left(\frac{\varepsilon_i}{\tilde{\varepsilon}}\right)$$ が得られる。（？）

MNL モデル

以下のように定式化される。

$$A_i = \exp\left\{\alpha_i + \sum_{k=1}^{K}\beta_k X_{ki} + \varepsilon_i\right\}$$

$$S_i = \frac{A_i}{\sum_{j=1}^{n} A_j}$$

対数中央化で以下のように線形式となる。

$$\log\left(\frac{S_i}{\tilde{S}}\right) = \alpha_i - \overline{\alpha} + \sum_{k=1}^{K}\beta_k\left(X_{ki} - \overline{X}_k\right) + \varepsilon_i - \overline{\varepsilon}$$