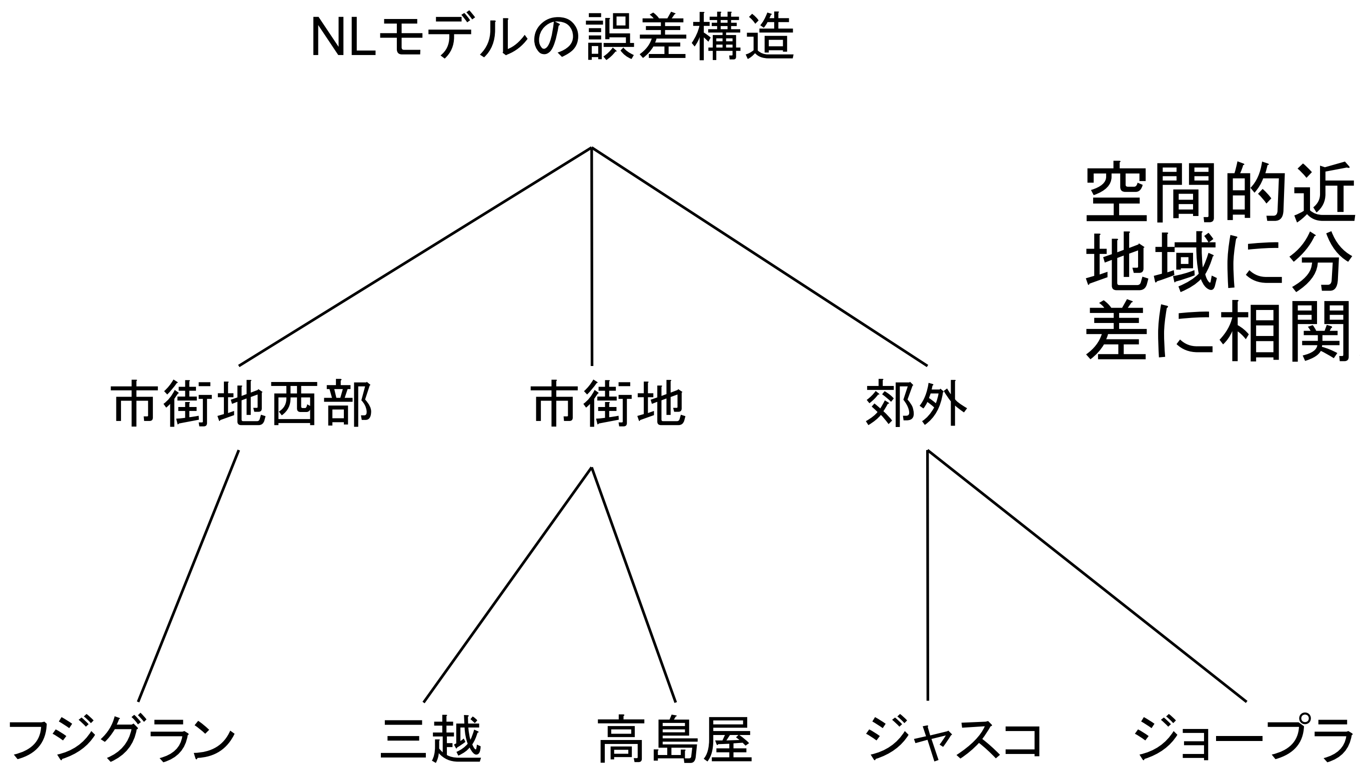
松山市内における
買物先選択行動モデル

東京大学　都市生活行動学研究室

基礎集計

月曜	0.070	0.088	0.080	0.082	0.143
火曜	**0.279**	0.088	0.120	0.072	0.119
水曜	0.140	0.147	0.160	0.113	0.167
木曜	0.093	0.118	0.100	0.134	0.071
金曜	0.070	0.088	0.080	0.103	0.167
土曜	0.186	**0.294**	0.220	0.206	**0.238**
日曜	0.163	0.176	0.240	**0.289**	0.095

ネスティッドモデルを用いた推定



空間的近接性により3つの地域に分け、それぞれの誤差に相関があると考える.

効用関数

$$U(takasi) = V_{takasi} + \varepsilon_{center} + \varepsilon_{center,takasi}$$
$$U(mituko) = V_{mituko} + \varepsilon_{center} + \varepsilon_{center,mituko}$$
$$U(f\_matu) = V_{f\_matu} + \varepsilon_{west} + \varepsilon_{west,f\_matu}$$
$$U(jyasuko) = V_{jyasuko} + \varepsilon_{south} + \varepsilon_{south,jyasuko}$$
$$U(joepla) = V_{joepla} + \varepsilon_{south} + \varepsilon_{south,joepla}$$

確定項に用いた説明変数

$$V(takasi) = \beta_1 * d_{takasi} + \beta_{10}$$

$$V(misuto) = \beta_2 * d_{mituko} + \beta_6 * age40 + \beta_9 * noncar + \beta_{11}$$

$$V(f\_matu) = \beta_3 * d_{f\_matu} + \beta_{12}$$

$$V(jyasuko) = \beta_4 * d_{jyasuko} + \beta_7 * tuesday + \beta_{13}$$

$$V(joepla) = \beta_5 * d_{joepla} + \beta_8 * woman$$

推定結果

	パラメータ	パラメータ値	t値
β１	距離パラメータ	−0.317	−2.54
β２	距離パラメータ	−0.662	−3.79
β３	距離パラメータ	−0.397	−2.89
β４	距離パラメータ	−0.384	−2.55
β５	距離パラメータ	−0.410	−2.73
β６	40才以上ダミーパラメータ	1.605	3.30
β７	火曜日ダミーパラメータ	1.058	2.96
β８	女性ダミーパラメータ	1.176	2.19
β９	車でないダミーパラメータ	−1.019	−3.02
β10	定数項	1.409	2.54
β11	定数項	0.122	0.18
β12	定数項	1.486	2.75
β13	定数項	0.919	1.46
β14	スケールパラメータ	2.523	3.22
	サンプル数	287	
	決定係数	0.26	
	修正済み決定係数	0.22	

考察

- NLモデルとMNLモデルを比べたが決定係数の値に大きな違いは見られなかった.
- ネストとしては市街地と郊外の２つに分けるよりは、空間的な近接性で3つに分けたほうが、わずかだがモデルの精度は上がった.
- しかし新たにダミー変数を導入するほうが有効であったため、ネストの組み方に改善の余地がある.

Multi nominal logit モデルの適用

- 買い物地選択行動をmulti nominal logit モデルを用いてモデル化する.

$$V_{高島\_n} = \beta_1 + \beta_3 d_{高島\_n}$$

$$V_{三越\_n} = \beta_2 + \beta_3 d_{三越\_n} + \beta_5 a_n$$

$$V_{ジョ\_n} = \beta_3 d_{ジョ\_n} + \beta_6 f_n$$

説明変数の検討(1)

- サンプルプログラムを改良する方針で説明変数を検討.
- 店までの距離そのものを説明変数として使うのではなく，距離を営業時間や売り場面積で除した値を使用.

$$d_{i\_n} \longrightarrow \frac{d_{i\_n}}{T_i * S_i}, \frac{d_{i\_n}}{S_i}, \frac{d_{i\_n}}{T_i}$$

d_{i_n} : 利用者nの施設iまでの距離

T_i : 施設iの営業時間

S_i : 施設iの敷地面積

説明変数の検討(2)

- 検討の結果，$\frac{d_{i_n}}{T}$を採用すると説明能力が向上．
- この値は，買い物可能な時間の長さを表しているといえる．
- ダミー変数についても多重共線性を考慮した上で検討．

効用関数

$$V_{takashi} = \beta_1 + \beta_5 \frac{d_{takashi\_n}}{T_{takashi} * 100}$$

$$V_{mituko} = \beta_2 + \beta_6 \frac{d_{mituko\_n}}{T_{mituko} * 100} + \beta_{10} a_n + \beta_{11} c_n$$

$$V_{f\_matu} = \beta_3 + \beta_7 \frac{d_{f\_matu\_n}}{T_{f\_matu\_n} * 100}$$

$$V_{jyasuko} = \beta_4 + \beta_8 \frac{d_{jyasuko\_n}}{T_{jyasuko} * 100} + \beta_{12} f_n + \beta_{13} t_n$$

$$V_{joepla} = \beta_9 \frac{d_{joepla\_n}}{T_{joepla} * 100} + \beta_{14} w_n + \beta_{15} a_n$$

d_{i_n} : 利用者nの施設iまでの距離[m]

T_i : 施設iの営業時間[h]

a_n : 40歳以上ダミー

c_n :自動車ダミー

f_n :フリーターダミー

t_n : 火曜日ダミー

w_n : 女性ダミー

推定結果

		パラメータ値	t値
β1	定数項	2.157	2.60
β2	定数項	4.289 (*10^(−1))	4.47 (*10^(−1))
β3	定数項	2.365	2.69
β4	定数項	2.519	3.00
β5	サービスレベル変数	−7.843 (*10^(−1))	−7.00
β6	サービスレベル変数	−9.499 (*10^(−1))	−5.28
β7	サービスレベル変数	−1.233	−6.68
β8	サービスレベル変数	−1.531	−7.28
β9	サービスレベル変数	−1.069	−5.93
β10	40歳以上ダミー	1.991	3.82
β11	自動車ダミー	−1.033	−2.42
β12	フリーターダミー	3.064	3.28
β13	火曜日ダミー	1.924	3.83
β14	女性ダミー	1.519	2.31
β15	40歳以上ダミー	1.199	2.37
	サンプル数	287	
	決定係数	0.295	
	自由度修正済み決定係数	0.259	

考察

- サービスレベル変数として導入した $\frac{d_{i\_n}}{T_i}$ の有意な説明能力が示された.
- モデルとしてもある程度の適合度があると考えられる.
- 関数形を考慮するなど, ダミー変数を導入する以上の工夫をすることが今後の課題である.

Mixed Logit Modelのパラメータ推定

効用関数が以下の式の形で示される.

$$U_a = \beta X_a + \eta_a + \nu_a$$

標準正規分布N(0,1)であり、

確率計算がopen-formの積分になるので計算が難しい.

N(0,1)に従う乱数を発生させてやり、モンテカルロ法で

シミュレーション

(ηの値をいろいろ変えて確率計算して、平均をとる)

シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)

シミュレーションに使う乱数

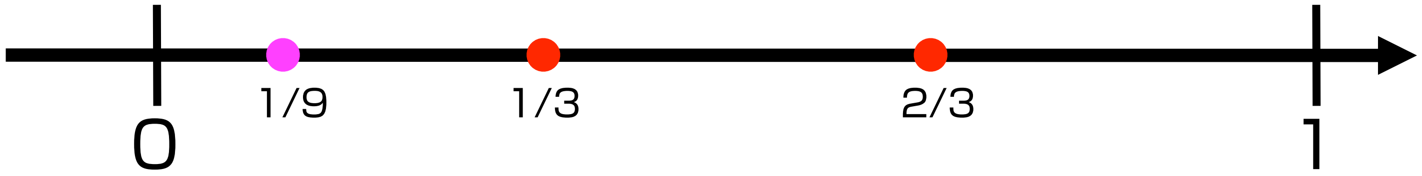
- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)

シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)

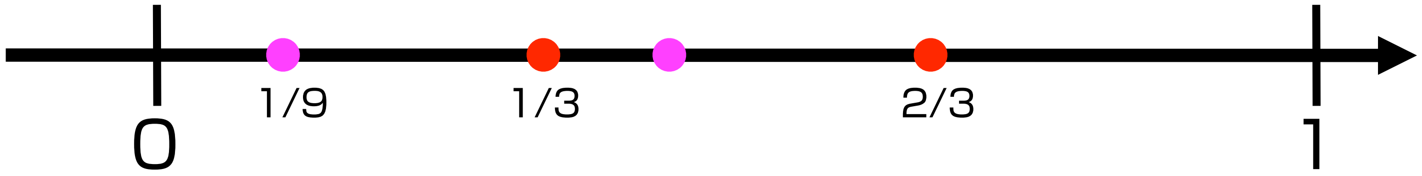
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



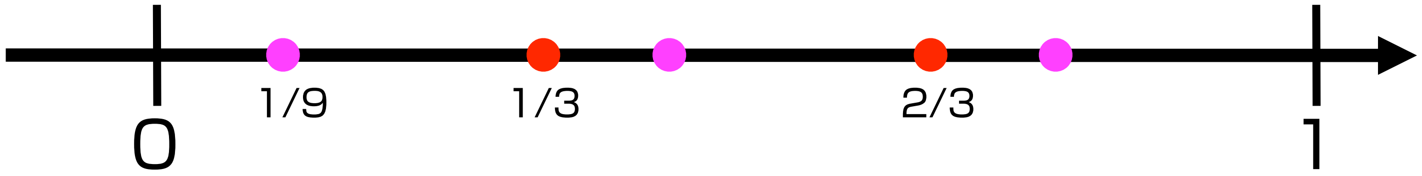
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



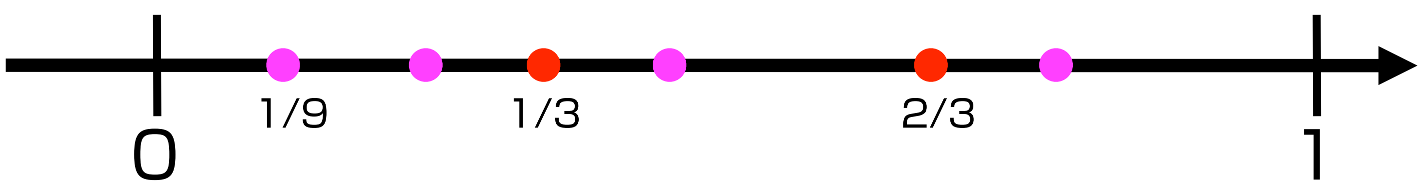
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



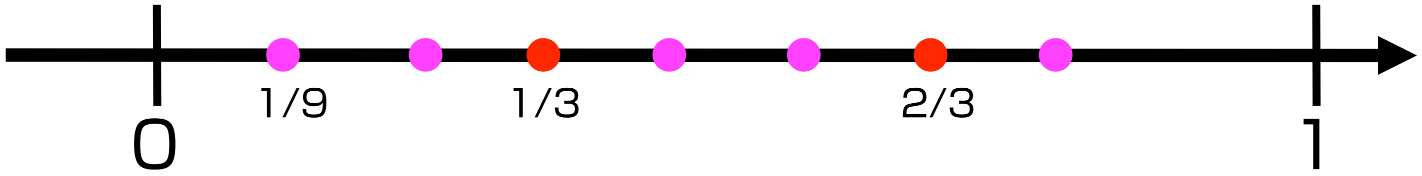
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



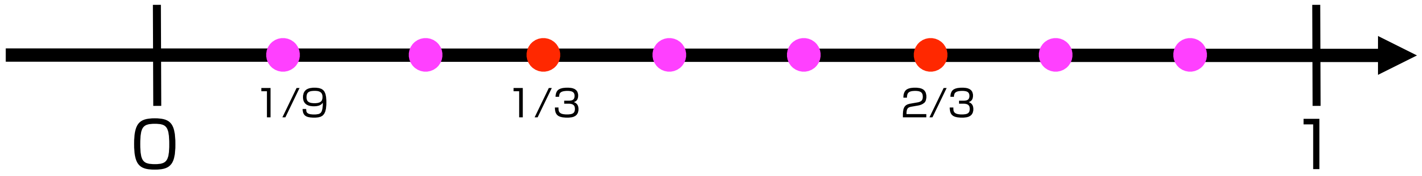
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



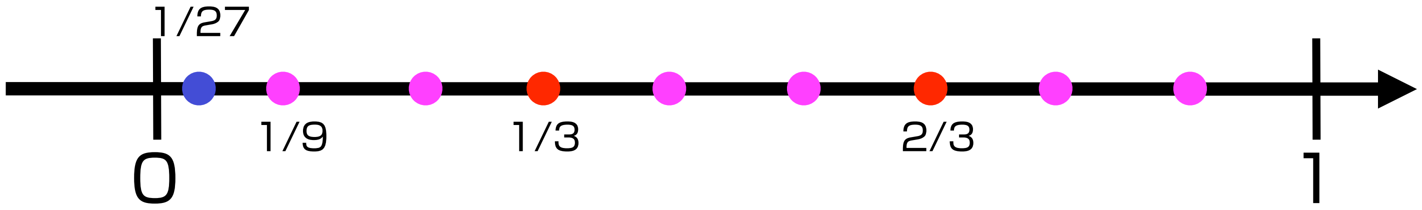
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



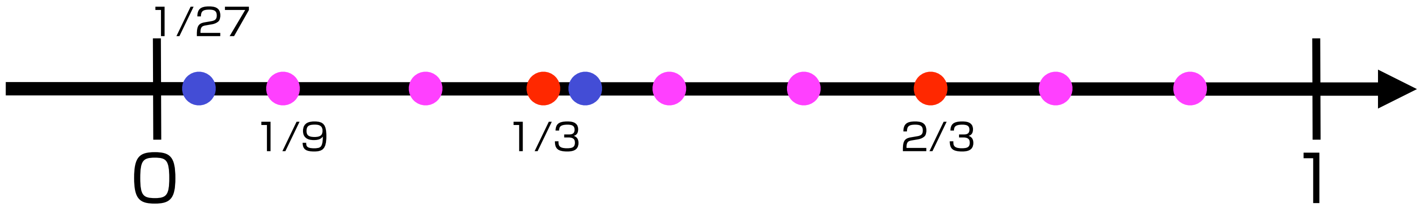
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



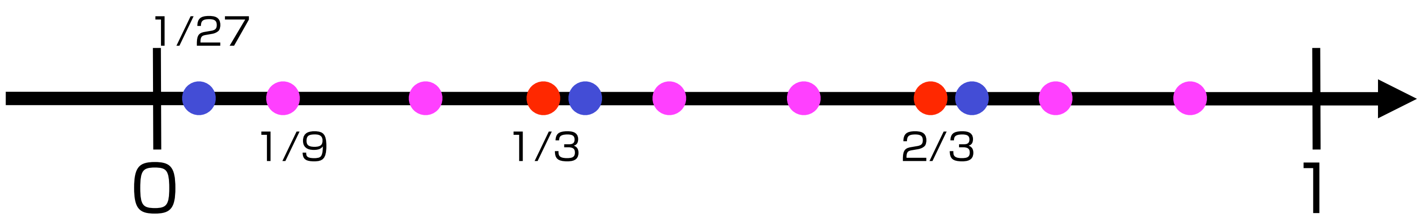
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



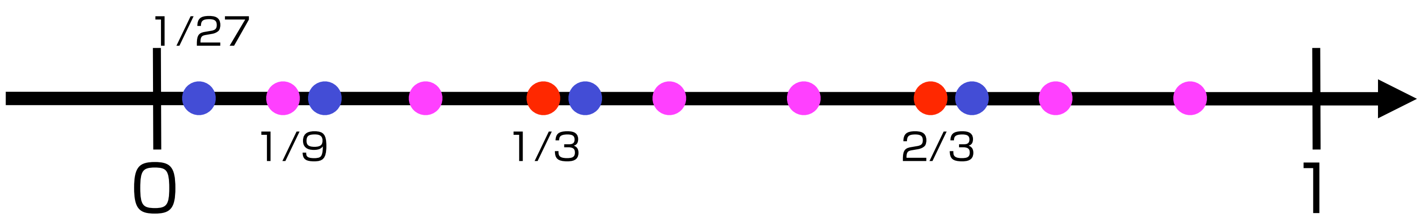
シミュレーションに使う乱数

- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



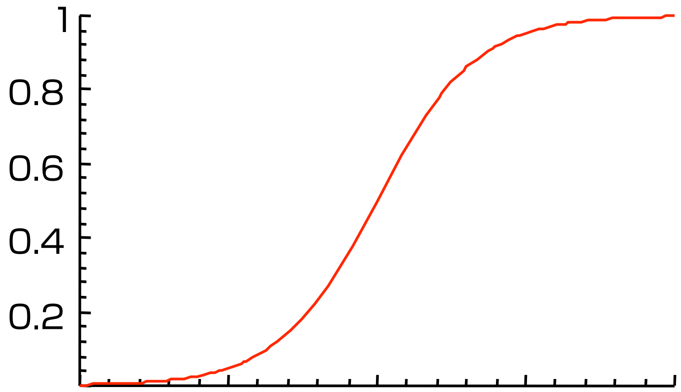
シミュレーションに使う乱数

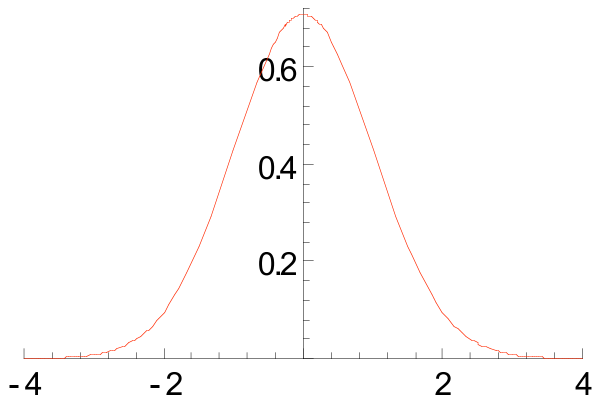
- ハルトン数列
  - 乱数ではないが、必要な値域を効率よくカバーでき、パラメータ推定を行うときに乱数を使用したときより試行回数を減らせる
  - 例:p=3のとき(pは2,3,5…と必要に応じて増やす)



シミュレーションに使う乱数

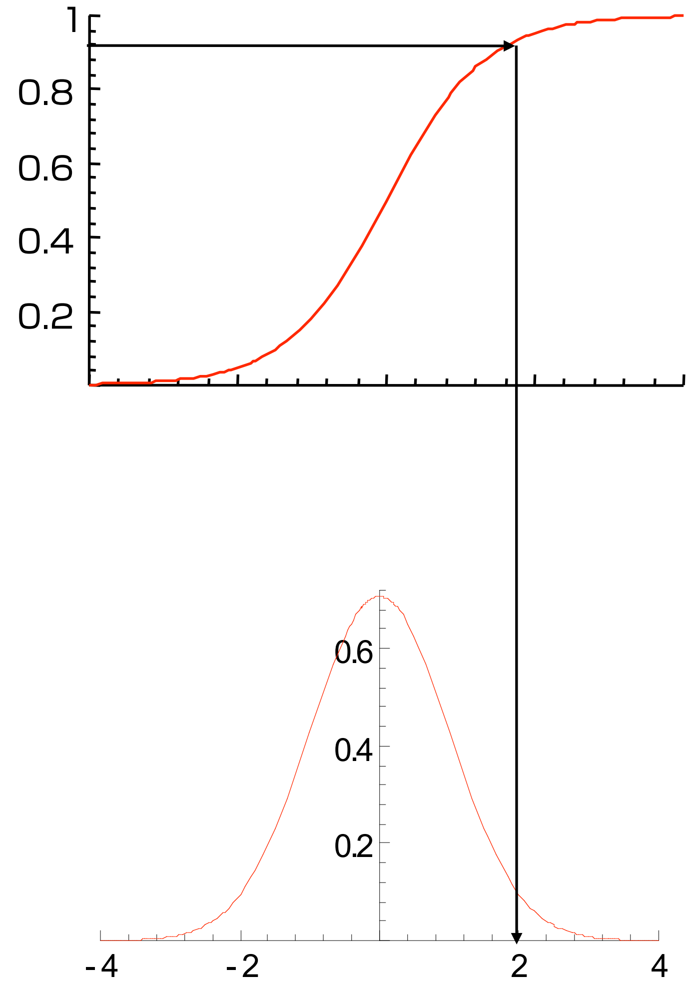
- 一様分布の0〜1の数列ができるので正規分布化





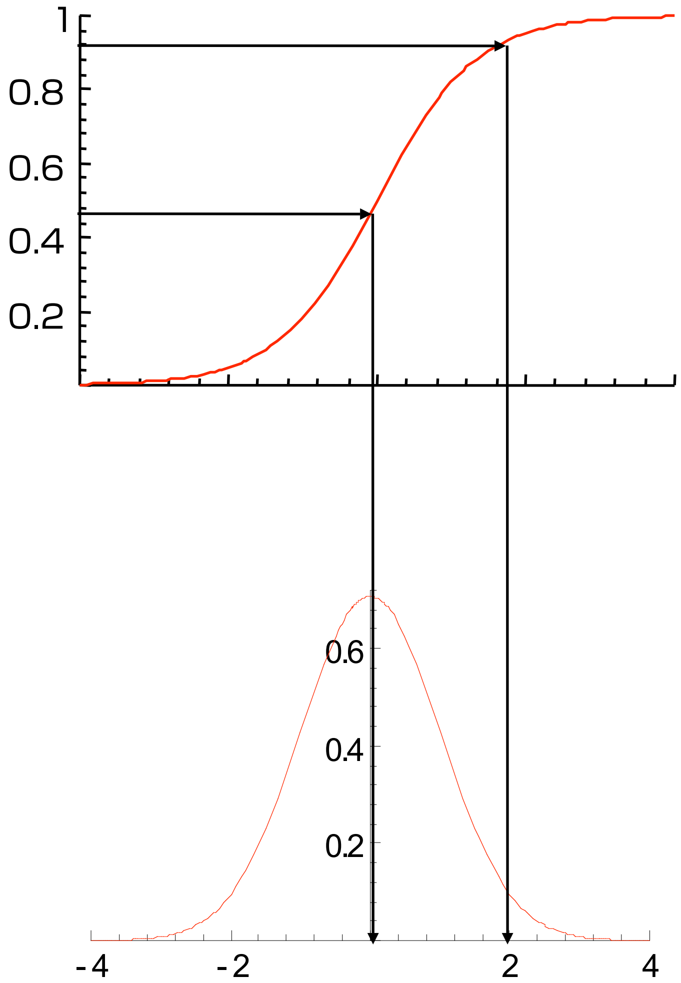
シミュレーションに使う乱数

- 一様分布の0〜1の数列ができるので正規分布化



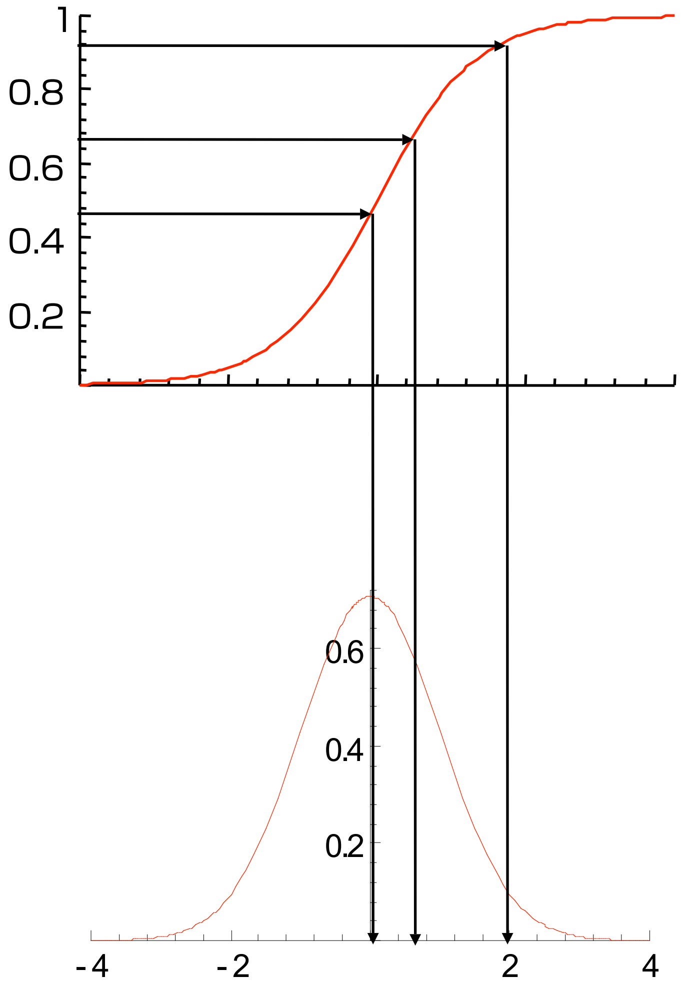
シミュレーションに使う乱数

- 一様分布の0〜1の数列ができるので正規分布化



シミュレーションに使う乱数

- 一様分布の0〜1の数列ができるので正規分布化



Mixed Logit Modelのパラメータ推定

- ηをいろいろ変えて尤度を計算

$$U_{\mathrm{a}} = \beta X_{\mathrm{a}} + \eta_{\mathrm{a}} + \nu_{\mathrm{a}}$$
$$U_{\mathrm{b}} = \beta X_{\mathrm{b}} + \eta_{\mathrm{b}} + \nu_{\mathrm{b}}$$

$$P(\mathrm{a}|\eta) = \frac{e^{\beta X_{\mathrm{a}} + \eta_{\mathrm{a}}}}{e^{\beta X_{\mathrm{a}} + \eta_{\mathrm{a}}} + e^{\beta X_{\mathrm{b}} + \eta_{\mathrm{b}}}}$$

- ηがN(0,1)に従うように乱数を生成しているので、P(a¦η)の平均は真のP(a)に近づく
  (ηを動かす積分と似たようなもの)

Mixed Logit Modelのパラメータ推定

- アルゴリズム実装時にはηに乱数を代入することは、ηという説明変数が増えた状態に似ている
- (例) 100件のデータに対して100回ηを変える場合
  - 100×100で10000回効用を計算する
  - この10000件の対数尤度を最大にするようにしてパラメータ推定をする

なんとかRに移植できた

- …ということをGAUSS用のMixed Logit Modelのプログラムや佐々木先生の資料を読みながらなんとか理解して、移植することができました

モデル立式

- 中心市街地にあるフジグラン松山，伊予鉄高島屋，三越　と　郊外にあるジャスコ松山，ジョープラの間に，誤差間の相関があると仮定し，Mixed Logit Modelに組み込む.
- 基礎集計の結果から，考慮できそうなダミー変数を導入する.

モデル式

$$
\begin{aligned}
V_{\text{takashi}} &= \beta_1 X_{\text{takashi}} + \eta_1 + \beta_6 D_{\text{Holiday}} + \nu_{\text{takashi}} \\
V_{\text{mituko}} &= \beta_2 X_{\text{mituko}} + \eta_1 + \beta_7 D_{\text{Forty}} + \nu_{\text{mituko}} \\
V_{\text{fmatu}} &= \beta_3 X_{\text{fmatu}} + \eta_2 + \nu_{\text{fmatu}} \\
V_{\text{jyasuko}} &= \beta_4 X_{\text{jyasuko}} + \eta_2 + \beta_8 D_{\text{Friday}} + \nu_{\text{jyasuko}} \\
V_{\text{joepla}} &= \beta_5 X_{\text{joepla}} + \eta_2 + \beta_9 D_{\text{Female}} + \nu_{\text{joepla}}
\end{aligned}
$$

パラメータ推定結果

	パラメータ値	t値
takashi定数項	1.640	2.327
mituko定数項	0.140	0.163
fmatu定数項	1.530	2.052
joepla定数項	1.019	1.490
jyasuko火曜ダミー	1.744	**3.686**
mituko40歳以上ダミー	1.772	**3.472**
joepla女性ダミー	1.416	2.159
空間的相関パラメタ（中心）	0.121	0.253
空間的相関パラメタ（郊外）	0.227	0.576
takashi距離	-0.854	**-7.950**
mituko距離	-1.231	**-6.314**
fmatu距離	-0.981	**-6.689**
jyasuko距離	-0.911	**-6.072**
joepla距離	-0.928	**-6.092**
takashi休日ダミー	0.452	1.640

サンプル数：287

ρ^2値： 0.244

修正済みρ^2値： 0.209

考察

- Mixed Logit Modelを用いたが，今回は有意な説明能力を示すことが出来なかった.
- 分類された空間同士の相関が十分になかった？
- また，推定プログラム作成に時間がかかり，モデルが吟味できなかったことも原因に挙げられる.