オーダードレスポンスモデル

名古屋大学　山本俊行

オーダードレスポンスモデルとは？

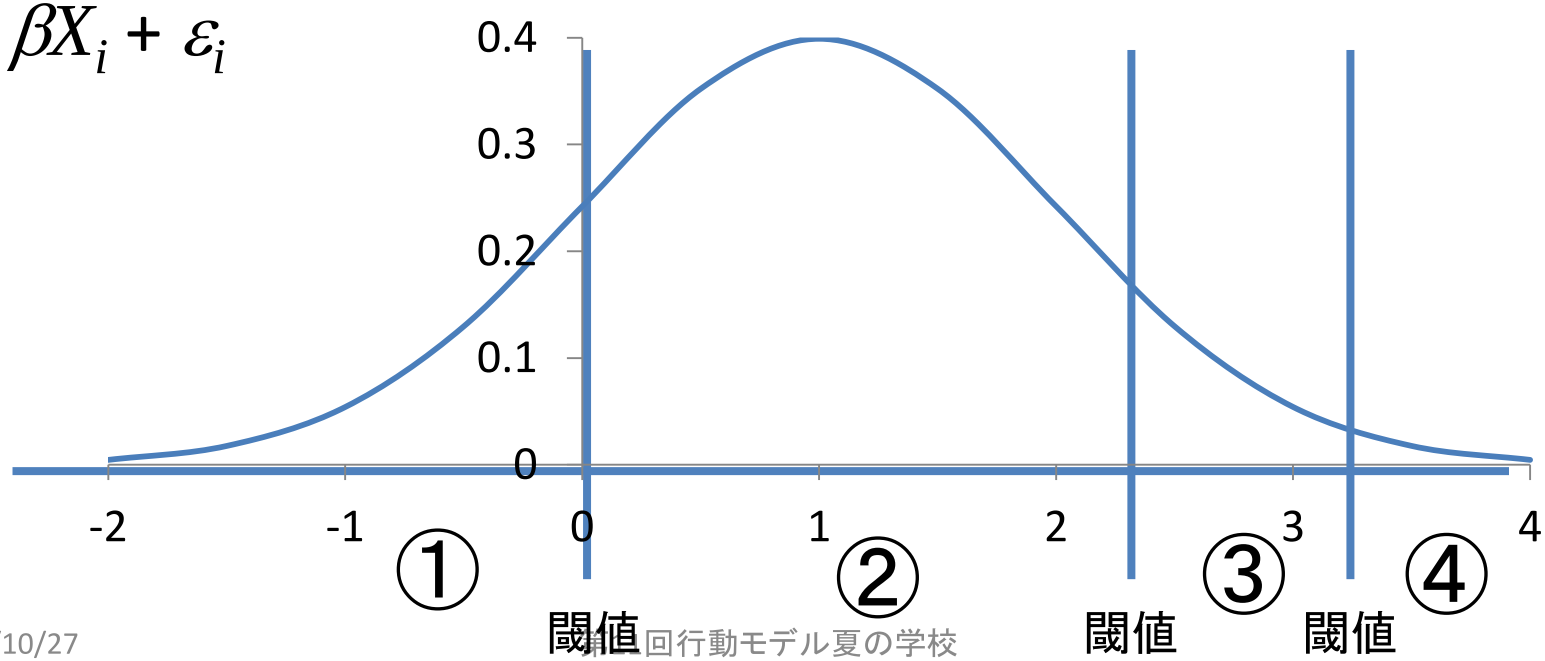
- 被説明変数が離散値として観測される
- 観測される変数の裏に連続的な（潜在）変数を想定する

- 適用例
  - 自動車保有台数:0台，1台，．．．J台
  - 交通事故の損傷度:車両のみ，軽傷，重症，死亡

$$
\begin{aligned}
y_i &= 0 \quad \text{if } y_i^* \le 0, \\
&= 1 \quad \text{if } 0 < y_i^* \le \mu_1, \\
&= 2 \quad \text{if } \mu_1 < y_i^* \le \mu_2, \\
&\ \vdots \\
&= J \quad \text{if } \mu_{J-1} < y_i^*
\end{aligned}
$$

y_i : 観測変数
y_i^* : 潜在変数
μ_j : 閾値
β : 未知パラメータ
X_i : 説明変数ベクトル
ε_i : 誤差項（正規分布やロジスティック分布）

$$y_i^* = \beta X_i + \varepsilon_i$$



$$
\begin{aligned}
y_i &= 0 \quad \text{if } y_i^* \le 0, \\
&= 1 \quad \text{if } 0 < y_i^* \le \mu_1, \\
&= 2 \quad \text{if } \mu_1 < y_i^* \le \mu_2, \\
&\vdots \\
&= J \quad \text{if } \mu_{J-1} < y_i^*
\end{aligned}
$$

y_i:観測変数
y_i^*:潜在変数
μ_j:閾値
β:未知パラメータ
X_i:説明変数ベクトル
ε_i:誤差項

$F(\ )$:累積分布関数

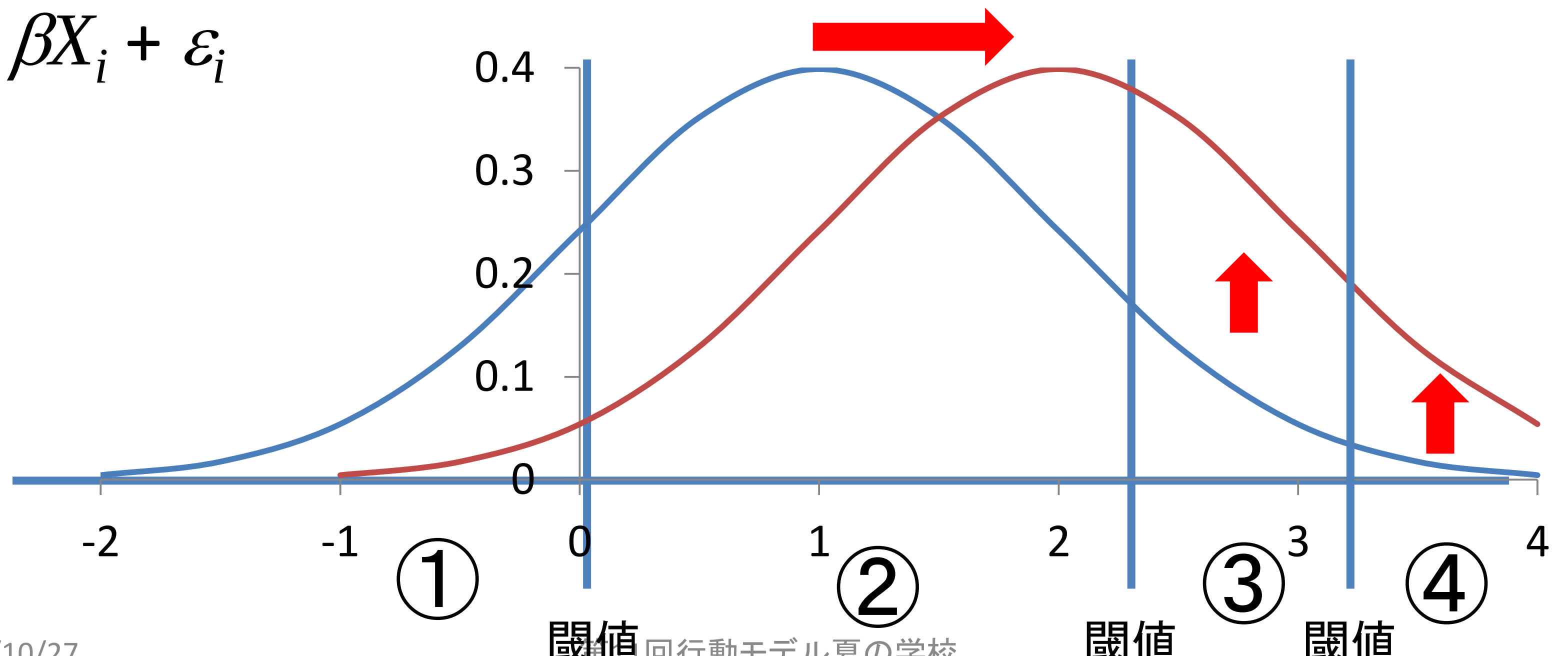
$$y_i^* = \beta X_i + \varepsilon_i$$

$$P(y_i = 0) = F(-\beta X_i),$$
$$P(y_i = j) = F(\mu_j - \beta X_i) - F(\mu_{j-1} - \beta X_i),\ j = 1,2,\ldots J-1,$$
$$P(y_i = J) = 1 - F(\mu_{J-1} - \beta X_i)$$

モデルの制約

- 説明変数の変化が全ての観測変数値の生起確率に影響する

$$y_i^* = \beta X_i + \varepsilon_i$$

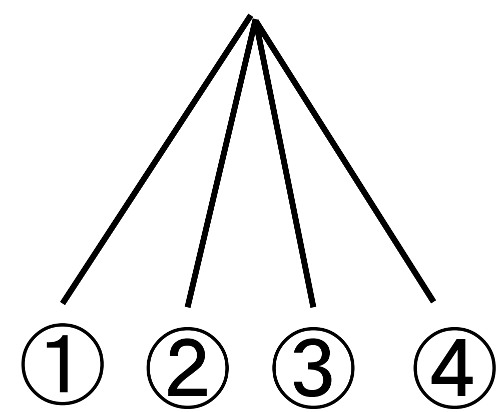


不都合な例と対応

- 交通事故の例では，エアバッグは
  - 死亡確率を減らす（潜在変数が左向きに動く）
  - 車両のみ→軽傷の効果（右向き）

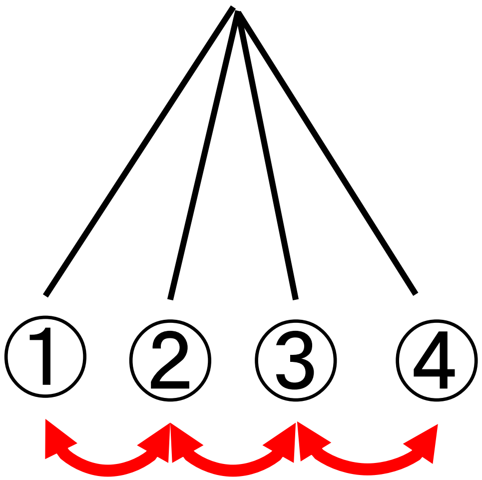
  の効果を同時に持つ

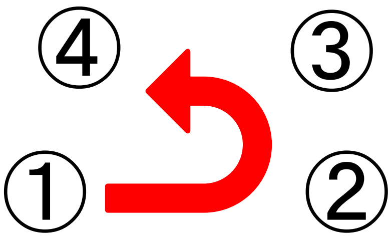
- 対応：離散選択モデルの適用
  - ○自由度は向上
  - ×観測変数の序数性は無視



オーダードGEV，MXLモデル

- 隣り合った選択肢間の誤差項に相関を持たせる
- 序列性というより近接性の表現
  - 必ずしも全体の大小関係を表現しない？





大小関係

序数性を保持した対応

- 閾値の構造化
- 観測変数値別のパラメータ
- 繰り返し2項選択モデル

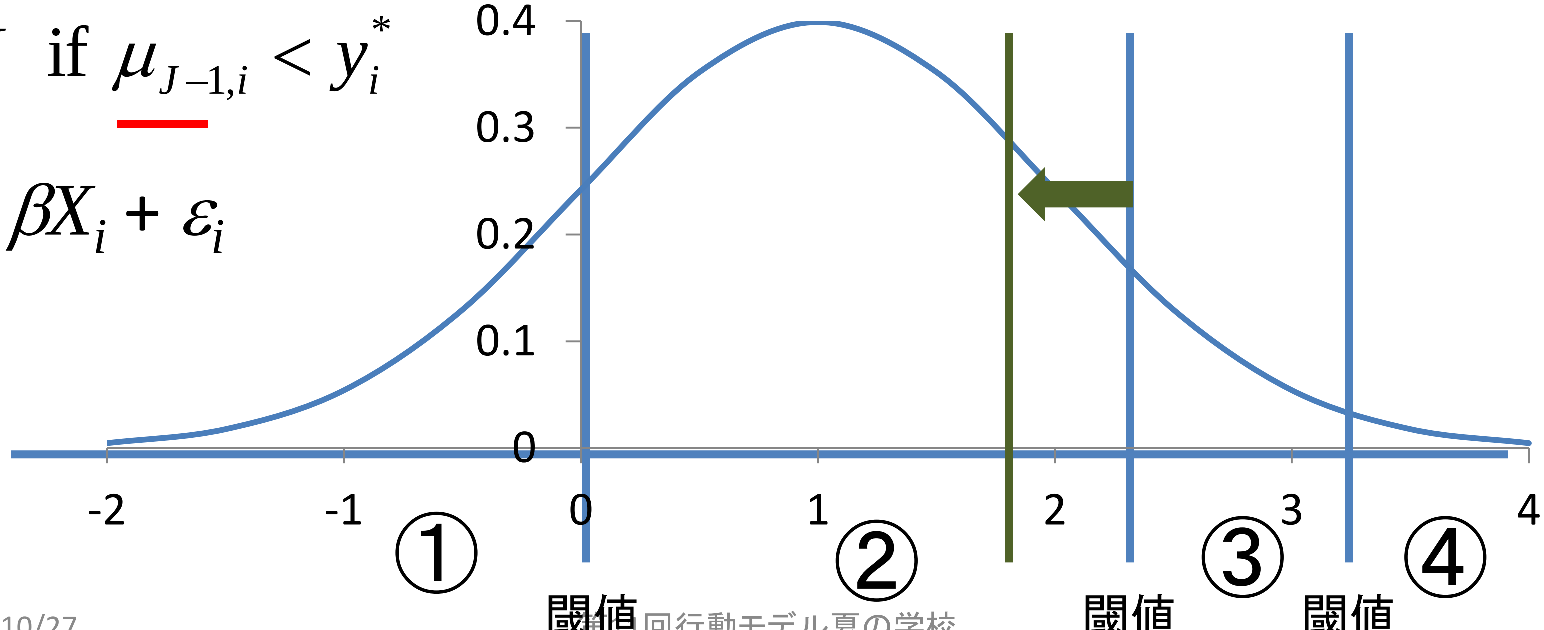
- 閾値の構造化

$$
\begin{aligned}
y_i &= 0 \quad \text{if } y_i^* \le 0, \\
&= 1 \quad \text{if } 0 < y_i^* \le \underline{\mu_{1i}}, \\
&= 2 \quad \text{if } \underline{\mu_{1i}} < y_i^* \le \underline{\mu_{2i}}, \\
&\;\vdots \\
&= J \quad \text{if } \underline{\mu_{J-1,i}} < y_i^*
\end{aligned}
$$

$$\underline{\mu_{ji} = \gamma_j X_i}$$

γ_i : 未知パラメータ

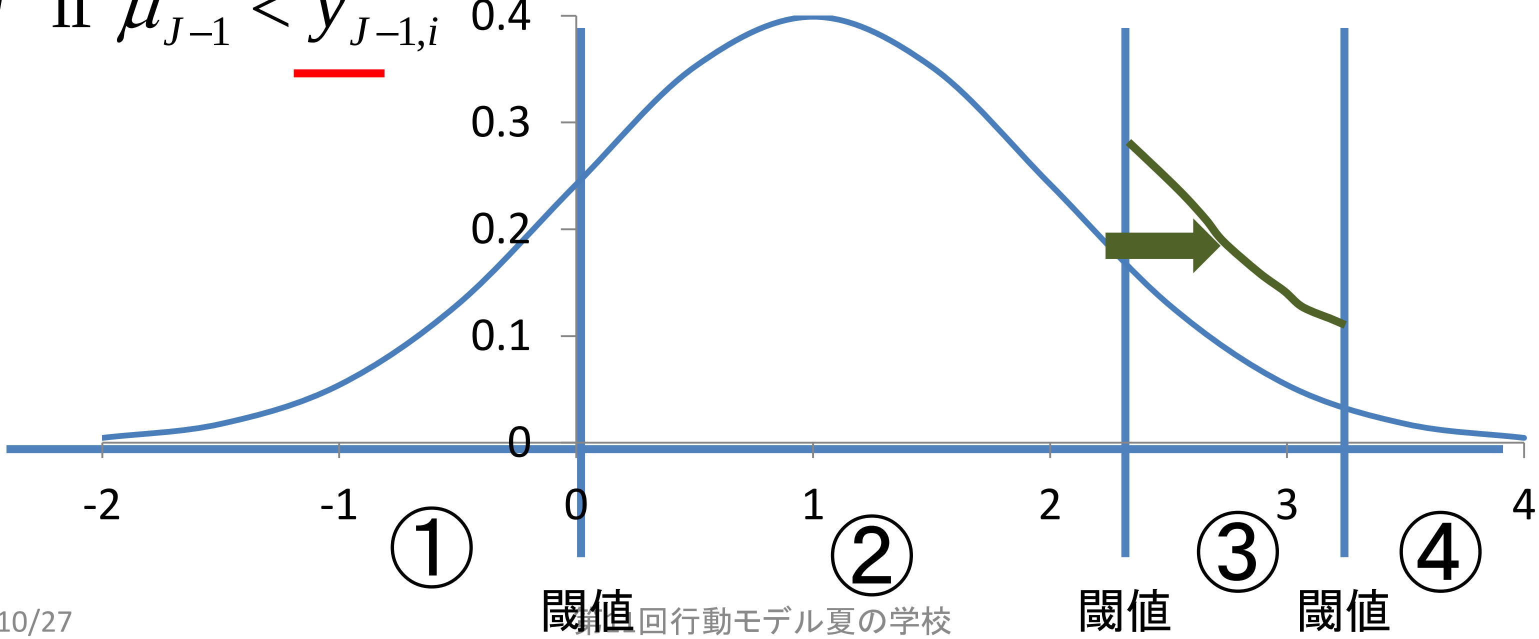
$$y_i^* = \beta X_i + \varepsilon_i$$



- 観測変数値別のパラメータ

$$
\begin{aligned}
y_i &= 0 \quad \text{if } \underline{y^*_{0i}} \le 0, \\
&= 1 \quad \text{if } 0 < \underline{y^*_{0i}},\ y^*_{1i} \le \mu_1, \\
&= 2 \quad \text{if } \mu_1 < \underline{y^*_{1i}},\ y^*_{2i} \le \mu_2, \\
&\ \vdots \\
&= J \quad \text{if } \mu_{J-1} < \underline{y^*_{J-1,i}}
\end{aligned}
$$

$$y_{ji}^* = \underline{\beta_j} X_i + \varepsilon_i$$



閾値の構造化と観測変数値別のパラメータの効果は一緒

- 閾値の構造化

$$P(y_i = 0) = F(-\beta X_i),$$
$$P(y_i = j) = F(\gamma_j X_i - \beta X_i) - F(\gamma_{j-1} X_i - \beta X_i),\ j = 1,2,\ldots J-1,$$
$$P(y_i = J) = 1 - F(\gamma_{J-1} X_i - \beta X_i)$$

- 観測変数値別のパラメータ

$$P(y_i = 0) = F(-\beta_0 X_i),$$
$$P(y_i = j) = F(\mu_j - \beta_j X_i) - F(\mu_{j-1} - \beta_{j-1} X_i),\ j = 1,2,\ldots J-1,$$
$$P(y_i = J) = 1 - F(\mu_{J-1} - \beta_{J-1} X_i)$$

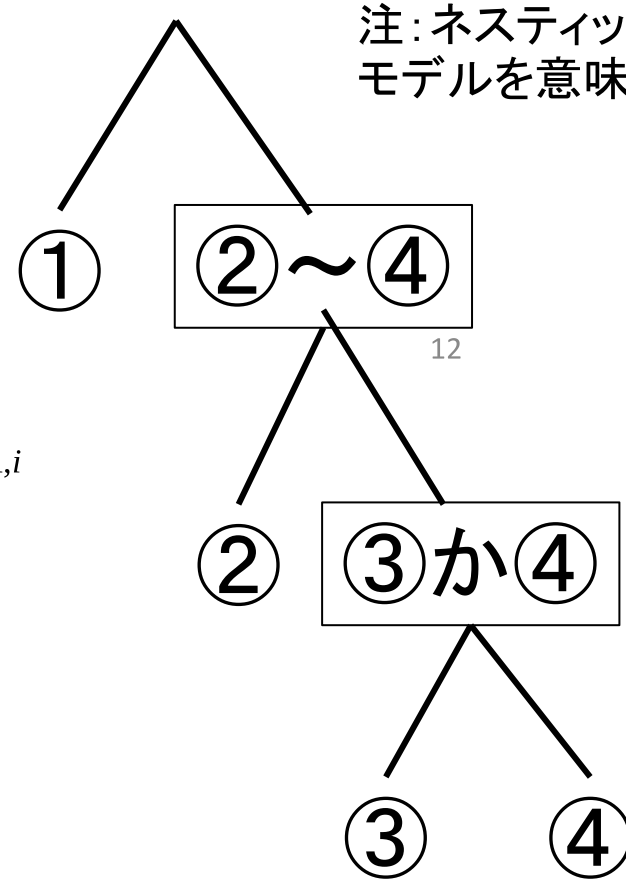
序数性を保持した対応

- 繰り返し2項選択モデル

$$
\begin{aligned}
y_i &= 0 \quad \text{if } y_{0i}^* \le 0, \\
&= 1 \quad \text{if } 0 < y_{0i}^* \ \text{ and } \ y_{1i}^* \le 0, \\
&= 2 \quad \text{if } 0 < y_{0i}^*, \ 0 < y_{1i}^* \ \text{ and } \ y_{2i}^* \le 0, \\
&\ \vdots \\
&= J \quad \text{if } 0 < y_{0i}^*, \ 0 < y_{1i}^*, \cdots, \text{and } \ 0 < y_{J-1,i}^*
\end{aligned}
$$

$$
\begin{aligned}
y_{0i}^* &= \underline{\beta_0} X_i + \underline{\varepsilon_{0i}} \\
y_{1i}^* &= \underline{\beta_1} X_i + \underline{\varepsilon_{1i}} \\
&\ \vdots \\
y_{J-1,i}^* &= \underline{\beta_{J-1}} X_i + \underline{\varepsilon_{J-1,i}}
\end{aligned}
$$

注：ネスティッドロジット
モデルを意味しない



- 繰り返し2項選択モデルの選択確率式

$$
\begin{aligned}
y_i &= 0 \quad \text{if } y_{0i}^* \le 0, \\
&= 1 \quad \text{if } 0 < y_{0i}^* \text{ and } y_{1i}^* \le 0, \\
&= 2 \quad \text{if } 0 < y_{0i}^*,\ 0 < y_{1i}^* \text{ and } y_{2i}^* \le 0, \\
&\ \vdots \\
&= J \quad \text{if } 0 < y_{0i}^*,\ 0 < y_{1i}^*, \cdots, \text{and } 0 < y_{J-1,i}^*
\end{aligned}
$$

$$
\begin{aligned}
y_{0i}^* &= \beta_0 X_i + \varepsilon_{0i} \\
y_{1i}^* &= \beta_1 X_i + \varepsilon_{1i} \\
&\ \vdots \\
y_{J-1,i}^* &= \beta_{J-1} X_i + \varepsilon_{J-1,i}
\end{aligned}
$$

$$
P(y_i = 0) = F(-\beta_0 X_i),
$$

$$
P(y_i = j) = \prod_{k=0}^{j-1} \{F(\beta_k X_i)\} F(-\beta_j X_i),
$$

$$
P(y_i = J) = \prod_{k=0}^{J-1} \{F(\beta_k X_i)\}
$$

序数性を保持した対応

- 閾値の構造化
- 観測変数値別のパラメータ
  – 誤差項は観測変数値（閾値）間で同一

- 繰り返し2項選択モデル
  – パラメータと誤差項のいずれもが観測変数値別

- 誤差項が同一の場合も独立の場合もありなら，その中間で，相関を持つ場合もあり得る
  – 交通事故データでやってみましたが相関はあまり有意ではありませんでした．

頻度モデル

- 被説明変数は離散値として観測される
- 観測される変数の裏に連続的な（潜在）変数を想定する

ポアソン分布

$$\Pr(y_i) = \lambda_i^{y_i} \exp(-\lambda_i) / y_i\,!$$

$$\lambda_i = \exp(\beta X_i)$$

y_i：観測変数
λ_i：潜在変数
β：未知パラメータ
X_i：説明変数ベクトル

- $\ln \lambda_i = \beta X_i$ はオーダードレスポンスモデルの潜在変数に相当 $\ln \lambda_i = \beta X_i \left(= y_i^*\right)$

頻度モデル

- 交通事故分析やアクティビティ分析において適用される

- 適用例
  - 各リンクの交通事故の回数
  - 1週間のトリップ頻度

非観測異質性がある場合

- 負の二項分布モデル

$$\Pr(y_i) = \lambda_i^{y_i} \exp(-\lambda_i) / y_i!$$

$$\lambda_i = \exp(\beta X_i + \psi_i)$$

ψ_i:ガンマ分布にしたがう誤差項（分散1/θ）

$$\Pr(y_i) = \frac{\Gamma(y_i + \theta)}{y_i!\Gamma(\theta)} \left(\frac{\lambda_i}{\lambda_i + \theta} \right)^{y_i} \left(\frac{\theta}{\lambda_i + \theta} \right)^{\theta}$$

頻度が複数の事象の合計だったら？

- 有り得る例
  - 個々に異なる原因の交通事故をまとめて回数を数えている
  - 異なる目的のトリップを区別せず数えている

- 多変量頻度モデル

多変量頻度モデル

- 各事象が独立の場合

$$\Pr\left(y_{i1}, y_{i2}, \ldots, y_{iK}\right) = \prod_{k=1}^{K} \left\{ \lambda_{ik}^{\ y_{ik}} \exp\left(-\lambda_{ik}\right) \Big/ y_{ik}! \right\}$$

$$= \frac{y_{iT}!}{\prod_{k=1}^{K} y_{ik}!} \prod_{k=1}^{K} \left(\frac{\lambda_{ik}}{\lambda_{iT}} \right)^{y_{ik}} \frac{\lambda_{iT}^{\ y_{iT}} \exp\left(-\lambda_{iT}\right)}{y_{iT}!}$$

$$y_{iT} = \sum_{k=1}^{K} y_{ik}, \quad \lambda_{iT} = \sum_{k=1}^{K} \lambda_{ik}$$

- 非観測異質性があり各事象が独立の場合

$$\Pr\left(y_{i1}, y_{i2}, \ldots, y_{iK}\right) = \prod_{k=1}^{K}\left\{\frac{\Gamma\left(y_{ik}+\theta\right)}{y_{ik}!\Gamma(\theta)}\left(\frac{\lambda_{ik}}{\lambda_{ik}+\theta}\right)^{y_{ik}}\left(\frac{\theta}{\lambda_{ik}+\theta}\right)^{\theta}\right\}$$

- 非観測異質性が各事象間で同一の場合

$$\Pr\left(y_{i1}, y_{i2}, \ldots, y_{iK}\right) = \frac{y_{iT}!}{\prod_{k=1}^{K} y_{ik}!}\prod_{k=1}^{K}\left(\frac{\lambda_{ik}}{\lambda_{kT}}\right)^{y_{ik}}$$

$$\times\frac{\Gamma\left(y_{iT}+\theta\right)}{y_{iT}!\Gamma(\theta)}\left(\frac{\lambda_{iT}}{\lambda_{iT}+\theta}\right)^{y_{iT}}\left(\frac{\theta}{\lambda_{iT}+\theta}\right)^{\theta}$$

非観測異質性が相関を持つ場合

- 非観測異質性が独立の場合と同一の場合があるなら，相関がある場合も考えられる

- でも，クローズドフォームはないみたい．なので，数値積分？あるいはCML (composite marginal likelihood)推定？
  - それは今後の課題ということで．．．

参考文献

- オーダードレスポンスモデル
  - Savolainena, P.T., Fred L. Mannering, F.L., Lordc, D. and Quddus, M.A. (2011) The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis & Prevention, Vol. 43, pp. 1666-1676.
- 多変量頻度モデル
  - Kockelman, K.M. and Krishnamurthy, S. (2004) A new approach for travel demand modeling: linking Roy’s Identity to discrete choice. Transportation Research Part B, Vol. 38, pp. 459–475.
  - 山本俊行，赤佐浩一，倉内慎也，森川高行：大規模小売店舗の中心市街地への出店が買い物行動に及ぼす影響の分析，土木計画学研究・講演集，Vol. 30, CD-ROM, 2004.