

プロセス改善と定量化モデルの 進化と発展 II

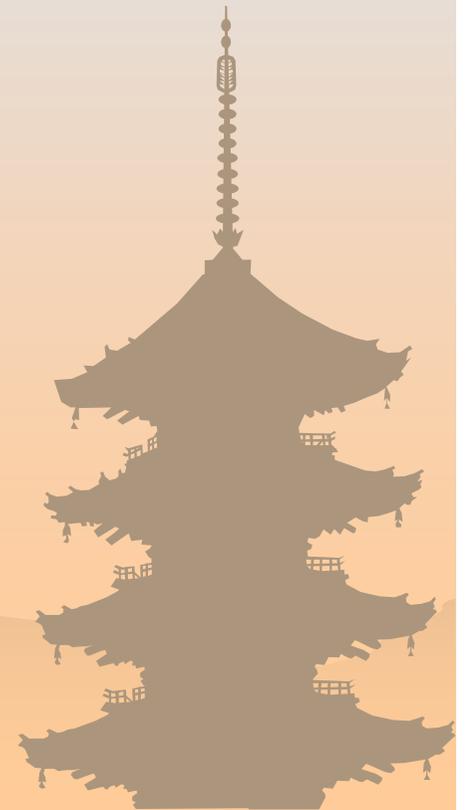
- 一般化線形モデルの適用

日立ソリューションズ
小室 睦



目次

- ❁ 効果説明モデルとその限界
- ❁ やりたいこと、問題点
- ❁ 解決のアイデア
- ❁ 一般化線形回帰による定式化
- ❁ 実データによる検証
- ❁ まとめ



効果説明モデルとその限界

○ 効果が生じるメカニズムを明らかにしている

× 効果算出方法が間接的

✓ 中間変数の使用

✓ 規模に関する再正規化

• 数値的な処理では誤差も累積する恐れがある

本発表の範囲

○ 品質向上、原価低減効果を定量的に示す

× 予測には使えない

レビュー
前倒し率
 $d(p_j)$

指数的に影響

$R(P)$

線形相関

規模(再)正規化
された欠陥密度

欠陥密度
テスト工程

やりたいこと、問題点

- ❁ テスト工程の欠陥密度 DDT とレビューの前倒し率 $d(p_j)$ を直接関係付けられないか
- ❁ 前倒し率 $d(p_j)$ は指数関数的を通して効いてくるので、対数をとるのが自然
- ❁ $\log(DDT) \sim d(p_j) + (\text{レビュー以外の影響})$
 - (記法) 右辺はRなどで用いる「モデル式」
 - 説明変数を '+' で結んで書き並べたもの
 - この形で少し実験してみたが精度のよい結果は得られなかった



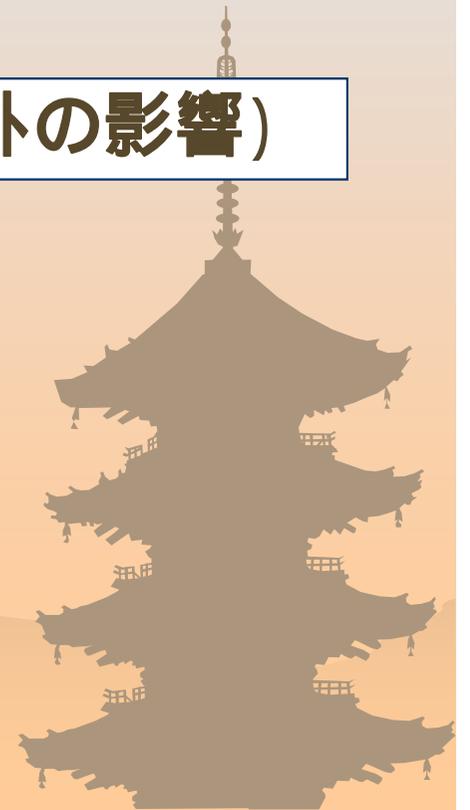
解決のアイデア

- ❁ 定義より、 $DDT = test.defect / size$
 - $test.defect$: テスト工程での欠陥摘出数、
 - $size$: 開発規模
- ❁ $\log(DDT) = \log(test.defect) - \log(size)$
 - $\log(size)$ はテスト工程に入る前の時点で確定できる
 - 説明変数側に回せば、精度向上するはず
 - 規模依存性の表現にもなる

JASPIC合宿研究会
参加中に気がついた

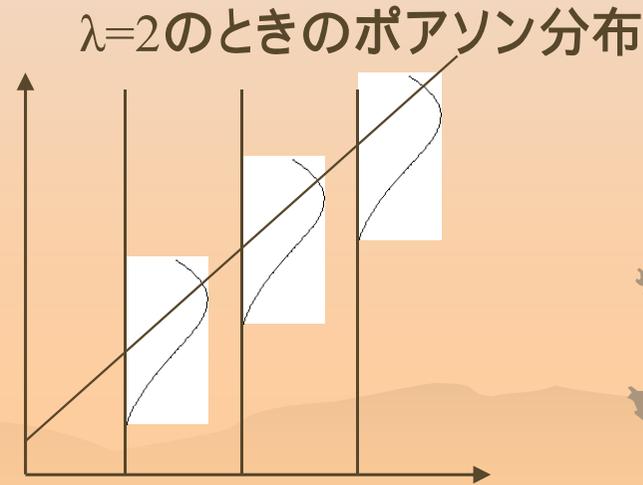
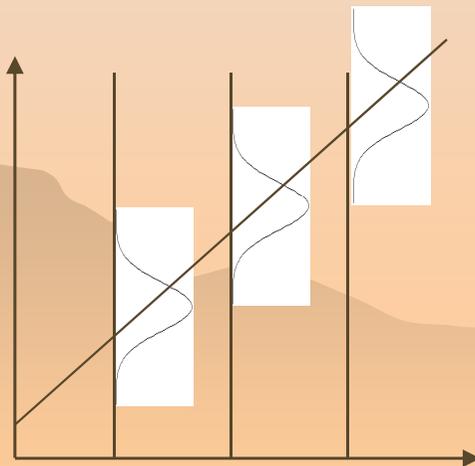
$$\log(test.defect) \sim \log(size) + d(p_j) + (\text{レビュー以外の影響})$$

教訓: JASPICの合宿は役に立つ



一般化線形回帰

- ❁ このアイデアが浮かんだ少し後にPACTの藤田さんの講義で、一般化線形回帰について教えていただいた。
- ❁ 通常の線形回帰
 - 最小自乗近似→残差に「等分散の正規分布」を仮定
- ❁ 一般化線形回帰
 - 残差が指数型分布族にしたがうことを仮定
 - 正規分布、ポアソン分布、二項分布などを含む族
 - どの分布を用いるかによって標準的なリンク関数を用意されている
 - 目的変数の変換関数
 - あてはまりの良さの尺度として最小自乗のかわりに(対数)「尤度」を用いる



一般化線形回帰による定式化

- ❁ 欠陥数である $test.defect$ 、あるいはその残差はポアソン分布にしたがうと期待される
- ❁ ポアソン分布用の標準的なリンク関数は「対数関数」
- ❁ 割り算を避けるテクニックとして 'offset' が用意されている
 - 割り算の分母を係数が1に固定された説明変数として扱う

残差がポアソン分布にしたがう一般化線形回帰とみる

$$\text{❁ } \log(test.defect) \sim \log(size) + d(p_j) + (\text{レビュー以外の影響})$$

↑
リンク関数

↑
係数を1に固定すればoffsetになる
固定しなければ規模依存性を表わす

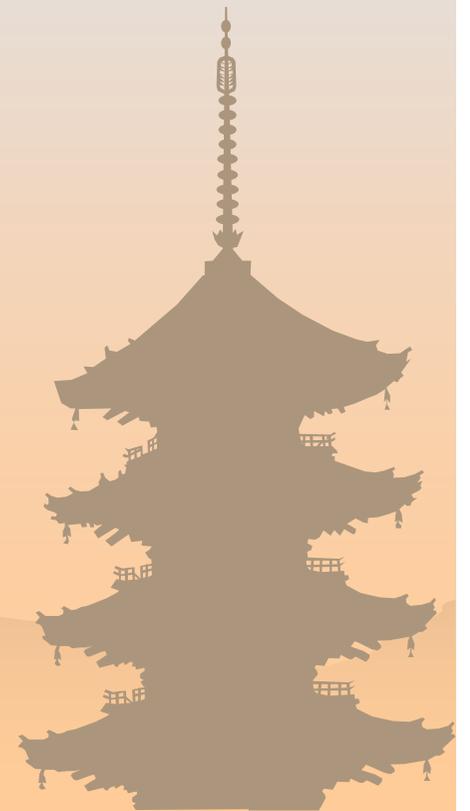
実データによる検証

工程 p	レビュー以外の説明変数	$d(p)$ のp値	擬似 R^2 (McFadden)	データ数	$d(p)$ の係数
BS 基本設計	母体規模(対数)	0.046	96.7%	10	-2.43
FS 機能設計	顧客側の体制リスク 開発環境のリスク	6.6×10^{-8}	96.6%	34	-2.13
DS 詳細設計	母体規模(対数) システムアーキテクチャの明確性リスク	0.037	91.8%	30	-0.83
P プログラミング	要件の明確性リスク 納期に関するリスク	3.4×10^{-5}	92.6%	34	-1.86

- ❁ リスクに係る説明変数はダミー変数(0/1)
- ❁ 「過分散」があり残差の分布は擬ポアソンとした

まとめ

- ❁ 2つの異なる視点が実質的に同じモデルを導く
 - レビューに関する考察
 - 経験、ソフトウェア工学
 - データに関する考察
 - ポアソン分布に対する一般化線形回帰
- ❁ 実データによる検証
 - 効果説明モデルのときより精度は向上
 - 規模を説明変数にまわしたことが大きい
- ❁ (教訓) JASPICの合宿は役に立つ



END

