# PHMによるソフトウェア製品の信頼性評価手法について

西尾泰彦† (02401815), 土肥正† (01307065), 尾崎俊治† (01002265)
†広島大学工学部第二類（電気系）

## 1. はじめに

テスト段階におけるソフトウェア製品の信頼性を定量的に評価するために，ソフトウェア信頼度成長モデル [1] と呼ばれる確率モデルが数多く提案されている．しかしながら，マルコフ過程や NHPP に基づいた従来のモデルでは，ソフトウェアテストにおいて得られる多くの有益な情報（ここでは環境データと呼ぶ）を有機的に利用した信頼性評価を行うことが事実上不可能であった．

このような問題に対して、Wightman and Bendell [2] は PHM (Proportional Hazards Model) と呼ばれる統計モデルを適用することにより，ソフトウェア製品の信頼性を推定する問題を取り扱っている．本稿の目的は、ソフトウェアフォールト発見時間データを解析するために上述の PHM が如何に利用されるかについて概説するとともに，いくつかの改良された推定方法の適用可能性について言及することである．

## 2. PHM

広義のソフトウェア信頼性評価においては，テスト資源やコストだけでなく，プログラムのコード数やプログラマの数といった様々な環境データが品質に影響を及ぼすと考えるべきであろう．Cox [3] は，故障時間データに影響を及ぼす環境データ（もしくは説明変数）を考慮しながら確率モデルを特徴づけるハザード関数を決定するために PHM を提案した．まず，時刻 $t_i$ $(i = i, \cdots, n)$ で，フォールトが発見される場合において，以下のようなパラメータを定義する．

$\mathbf{Z}_i$： 時刻 $t_i$における環境データを表す説明変数のベクトル

$\dot{\beta}$： 未知のパラメータで，説明変数 $\mathbf{Z}_i$の係数ベクトル

本稿では特に，

$$\mathbf{Z}_i = (z_{i1}, z_{i2}, \cdots, z_{ij}, \cdots, z_{ip}) \tag{1}$$

$$\dot{\beta} = (\beta_1, \beta_2, \cdots, \beta_j, \cdots, \beta_p) \tag{2}$$

とする．ここで，$\beta_j$ $(j = 1, \cdots, p)$ の値が大きいほど，説明変数 $z_{ij}$ $(j = 1, \cdots, p)$ はハザード関数に大きな影響を及ぼすことに注意されたい．このとき，ハザード関数は，

$$h(t_i; \mathbf{Z}_i) = h_0(t_i)\phi(\mathbf{Z}_i, \dot{\beta}) \tag{3}$$

のように定義される．但し，$h_0(t_i)$ は基本ハザード関数であり，

$$\begin{aligned}\phi(\mathbf{Z}_i, \dot{\beta}) &= \exp(\mathbf{Z}_i, \dot{\beta}), \\ &= \exp(z_{i1}\beta_1 + z_{i2}\beta_2 + \cdots + z_{ip}\beta_q)\end{aligned} \tag{4}$$

である．

いま，$i$ 番目 $(i = 1, 2, \cdots, n)$ のソフトウェアフォールトが発見されたという状況において，ソフトウェアフォールト発見時刻 $0 = t_0 < t_1 < \cdots < t_i < \cdots < t_n$ と対応する環境データのベクトル $\mathbf{Z}_i$ が得られているものとする．このとき，$t_i$で故障する条件付き確率を

$$\frac{h(t_i; \mathbf{Z}_i)}{\sum_{j \in R(t_i)} h(t_i; \mathbf{Z}_j)} = \frac{\exp(\mathbf{Z}_i \dot{\beta})}{\sum_{j \in R(t_i)} \exp(\mathbf{Z}_j \dot{\beta})} \tag{5}$$

のように定義する．これにより，未知パラメータ$\dot{\beta}$に対する尤度関数は

$$L(\dot{\beta}) = \Pi_{i=1}^{n} \frac{\exp(\mathbf{Z}_i, \dot{\beta})}{\left[\sum_{k \in R(t_i)} \exp(\mathbf{Z}_k, \dot{\beta})\right]^{d_i}} \tag{6}$$

となる．ここで，$d_i$ は時刻 $t_i$ でフォールトが同時に発生する個数を表し，$R(t_i)$ は時刻 $t_i$ におけるリスクを表す．式 (6) を未知パラメータ $\beta_j$ $(j = 1, \cdots, p)$ で微分することによって，$\phi(\mathbf{Z}_i, \dot{\beta})$ を推定することが可能となる．

## 3. 基本ハザード関数の推定

基本ハザード関数 $h_0(t)$ を決定するための最も単純な方法は，多くの信頼度成長モデルにおいてなされているように，$h_0(t)$ に適当な関数形を仮定することによりパラメトリック推定量を求めることである [2]．ここでは，具体的な形状を仮定することなく，基本ハザード関数を推定するためのノンパラメトリック推定法について述べる．

### (3.1) 方法1

Breslow [4] は，$\dot{\beta}$の最尤推定値 $\hat{\dot{\beta}}$ が得られたとき，基本ハザード関数を

$$h_{01}(t_i) = \frac{d_i}{(t_i - t_{i-1}) \sum_{m \in R(t_i)} \exp(\mathbf{Z}_m \hat{\dot{\beta}})} \tag{7}$$

のように表現することを提案している．これにより，基本信頼度関数 $S_0(t)$ ならびに累積基本ハザード関数は，

$$S_0(t) = \Pi_{i: t_i \le t}\left(1 - \frac{d_i}{\sum_{m \in R(t_i)} \exp(\mathbf{Z}_m \hat{\dot{\beta}})}\right), \tag{8}$$



© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

$$H_0(t) = \sum_{i:t_i \le t} \frac{d_i}{\displaystyle\sum_{m \in R(t_i)} \exp(\mathbf{Z}_m \hat{\beta})} \tag{9}$$

となる．

### (3.2) 方法 2

文献 [5] では，$\beta$の最尤推定値 $\hat{\beta}$ が得られた場合，基本信頼度関数 $S_0(t)$，信頼度関数

$$S(t_i) = [S_0(t_i)]^{\exp(\mathbf{Z}_i \hat{\beta})} \tag{10}$$

ならびに累積ハザード関数

$$H(t_i) = -\ln S(t_i) \tag{11}$$

の推定法について議論している．まず，$\alpha_i$ $(i = 1, \cdots, n)$ という時刻 $t_i$ $(i = 1, \cdots, n)$ においてフォールトが発生しない確率を表すパラメータの推定を行い，フォールト発見時刻 $t_i$ $(i = 1, \cdots, n)$ における基本ハザード関数の推定値をもとに信頼度関数を推定することを考える．

はじめに，$z'_{ij}$ を $i$ 番目のフォールト発見時間に関する環境データの規格化変数とする．すなわち，

$$z'_{ij} = \frac{z_{ij} - \bar{z}_j}{\sigma_j} \tag{12}$$

とする．ただし，$\bar{z}_j, \sigma_j$ は $j$ 番目の環境データの平均と標準偏差を表す．このとき，新しい説明変数を

$$z^*_{ij} = \begin{cases} 0 & (z'_{ij} \le 0) \\ 1 & (z'_{ij} > 0) \end{cases}$$

のように定義し，新しい説明変数のベクトルを

$$\mathbf{Z}^*_i = (z^*_{i1}, z^*_{i2}, \cdots, z^*_{ij}, \cdots, z^*_{ip}) \tag{13}$$

とする．これより，説明変数 $\mathbf{Z}^*_i$ を用いて，基本信頼度関数の推定を行う．

式 (13) を用いて，信頼度関数は $[S_0(t)]^{\exp(\mathbf{Z}^*_i, \hat{\beta})}$ となる．また，パラメータ $\alpha_i$ を用いると，時刻 $t_i$ での信頼度関数は

$$S_0(t_i) = \Pi_{i=1}^{n-1} \alpha_i \tag{14}$$

となる．よって，以下の尤度関数

$$\begin{aligned} L(\alpha, \hat{\beta}) \;\; = \;\; & \Pi_{i=1}^{n} \{\Pi_{m \in F(t_i)} [1 - (\alpha_i)^{\exp(\mathbf{Z}^*_m, \hat{\beta})}]\} \\ & \times \{\Pi_{q \in R(t_i) - F(t_i)} (\alpha_i)^{\exp(\mathbf{Z}^*_q, \hat{\beta})}\} \end{aligned} \tag{15}$$

を定義し，対数尤度 $l = \ln[L(\alpha, \hat{\beta})]$ を $\alpha_i$ $(i = 1, \cdots, n)$ で偏微分することにより

$$\frac{\partial l}{\partial \alpha_i} = 0 \tag{16}$$

を得る．但し，

$F(t_i)$ : 時刻 $t_i$ に同時に発見されたソフトウェアフォールトの集合，

$R(t_i) - F(t_i)$ : 時刻 $t_i$ において発見されなかったソフトウェアフォールトの集合

とする．これより，式 (16) を $\alpha_i$ $(i = 1, \cdots, n)$ について解くことによって，パラメータ $\alpha_i$ $(i = 1, \cdots, n)$ の推定値

$$\hat{\alpha}_i = \exp\left[\frac{-d_i}{\sum_{q \in R(t_i)} \exp(\mathbf{Z}^*_q, \hat{\beta})}\right] \tag{17}$$

を得ることができる．よって，基本信頼度関数の推定値は

$$\hat{S}_0(t) = \Pi_{i:t_i < t} \hat{\alpha}_i \tag{18}$$

となり，最終的に信頼度関数

$$\hat{S}(t_i; \mathbf{Z}^*_i) = [\hat{S}_0(t_i)]^{\exp(\mathbf{Z}^*_i, \hat{\beta})} \tag{19}$$

が得られる．よって，式 (19) を式 (11) に代入することにより，累積ハザード関数を求めることが可能となる．

## 4. まとめ

本稿では，ソフトウェア製品の信頼性を評価するために，PHM を適用するための理論的枠組みについて議論した．実際のフォールト発見時間データに基づいた PHM の検証は文献 [2] において十分なされていないのが現状である．実際のデータを解析した結果は研究発表会当日に報告させて頂く予定である．

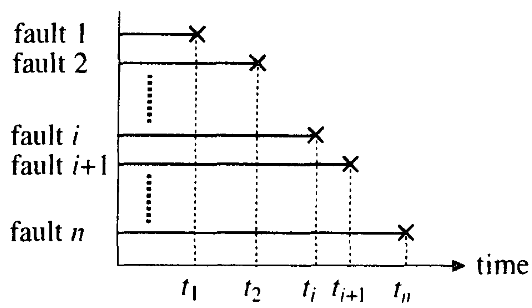


図 1: フォールト発見時間の概念図

## 参考文献


[1] 山田茂，「ソフトウェア信頼性モデル　基礎と応用」，日科技連，東京，(1994).

[2] D. W. Wightman and A. Bendell, "Proportional hazards modelling of software failure data", *Software Reliability: State of the Art Report*, A. Bendell and P. Mellor (eds.), pp. 229-263, Pergamon, Oxford (1986).

[3] D. R. Cox, "Regression models and life tables", *J. Roy. Statist. Soc.*, **B-34**, pp. 187-220, (1972).

[4] N. Breslow, "Covariate analysis of censored survival data", *Biometrics*, **30**, pp. 89-99, (1974).

[5] B. D. Bunday, *Statistical Methods in Reliability Theory and Practice*, Ellis Horwood, NY, (1991).





© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．