一般化ガンマ分布を適用したソフトウェア信頼性モデルのパラメータ推定

岡村寛之 (01013754)，安藤光昭，土肥正 (01307065)
広島大学大学院工学研究科情報工学専攻

1. はじめに

コンピュータシステムに高度な信頼性が求められる現在の社会において，ソフトウェア信頼性の定量的な評価は重要な意味をもつ．一般にソフトウェアの信頼性は，ソフトウェア内に潜在するフォールトの出現頻度によって計られる．よく知られている手法は，フォールト発見事象を確率過程を用いて表現する手法である．これらの確率過程はソフトウェア信頼度成長モデルと呼ばれ，現在までに数多くのモデルが提案されている．

有限フォールトをもつソフトウェア信頼度成長モデルを統一的に扱う枠組みとして，一般化順序統計量モデルが知られている [1]．一般化順序統計量モデルは非同次ポアソン過程 (NHPP) によって表現され，ソフトウェアのフォールト発見時刻に関する確率分布の選択によって任意に従来のソフトウェア信頼度成長モデルを表現することができる．しかしながら，実際に信頼性評価を行う場合には，複数のモデルの中から評価対象となるシステムに対して最適なモデルを選択する作業が必要となる．そこで本稿では，フォールト発見時刻が一般ガンマ分布に従うモデルを提案する．一般ガンマ分布はパラメータを変化させることで他のさまざまな分布パターンを表現することができる．すなわち，一般ガンマ分布を用いたソフトウェア信頼度成長モデルについて有効なパラメータ推定手法を提案することができれば，多くのモデルを統一的に扱うことが可能となる．

一般化順序統計量モデルに対するパラメータ推定には，一般的に最尤法が用いらてきたが，従来の方法では安定して望ましい解を得ることが困難であった．これに対し，EM アルゴリズムは最尤推定値を安定して精度よく算出できることが知られている [2]．本稿では，一般ガンマ分布を用いたソフトウェア信頼度成長モデルについて，EM アルゴリズムを基本とした推定手法を提案する．

2. 一般化順序統計量モデルに対するパラメータ推定手法

一般化順序統計量モデルは，次のような仮定のもとで議論される．

(仮定 A) ソフトウェアのテスト中にソフトウェア故障が発生した場合，その原因となるフォールトは瞬間的に発見・除去される．

(仮定 B) プログラム中に含まれる初期フォールト数 N_0 は平均 $\omega\ (>0)$ のポアソン分布に従う．

(仮定 C) ソフトウェア故障はそれぞれ独立かつ時間に関してランダムに発生し，それぞれのフォールト発見時刻は確率分布 $F(t)$，密度関数 $f(t)$，$t \geq 0$ に従う非負で連続の確率変数によって記述される．

以上の仮定より，時刻 t までに発見されるフォールト数 $N(t)$ の確率分布は

$$\Pr\{N(t)=m\} = \frac{\{\omega F(t)\}^m}{m!}\exp\{-\omega F(t)\} \qquad (1)$$

と表現される．

今，フォールト発見時刻がパラメータ $\boldsymbol{\theta}$ をもつ確率分布に従う場合の，一般化順序統計量モデルについて考える．フォールト発見時刻を総数 N_0 が観測不可能な確率変数列 $\{X_i, i=1,\ldots,N_0\}$ として定義すると，$N_0 = n$ という条件のもとで，$m < n$ に対して不完全なフォールトデータ $\mathcal{D}_t = (x_1,\ldots,x_m)$ が得られたとき，一般化順序統計量モデルにおける期待対数尤度関数は

$$\begin{aligned}\log L_{\mathrm{E}}(\omega,\boldsymbol{\theta}|\mathcal{D}_t) &= \mathrm{E}[N_0|\mathcal{D}_t;\omega,\boldsymbol{\theta}]\log\omega - \omega \\ &\quad + \mathrm{E}\left[\sum_{i=1}^{N_0}\log f(X_i;\boldsymbol{\theta})\middle|\mathcal{D}_t;\omega,\boldsymbol{\theta}\right] \qquad (2)\end{aligned}$$

となる．

EM アルゴリズムによるパラメータの推定値は，不完全なデータ $\mathcal{D}_t$ のみが観測されているもとでの期待対数尤度，すなわち式 (2) を最大にする値を更新値とすることによって得られる．具体的に，$p+1$ ステップの更新式は次のようになる．

$$\hat{\omega}_{p+1} = \mathrm{E}[N_0|\mathcal{D}_t;\hat{\omega}_p,\hat{\boldsymbol{\theta}}_p], \qquad (3)$$

$$\hat{\boldsymbol{\theta}}_{p+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\left\{\mathrm{E}\left[\sum_{i=1}^{N_0}\log f(X_i,\boldsymbol{\theta})\middle|\mathcal{D}_t;\hat{\omega}_p,\hat{\boldsymbol{\theta}}_p\right]\right\}. \qquad (4)$$

3. 一般ガンマ分布を適用したモデルに対する推定

一般ガンマ分布の確率密度関数は，パラメータ $\alpha>0$, $\beta>0$, $\gamma\geq 0$, $k>0$ を用いて

$$f(t;\alpha,\beta,\gamma,k) = \frac{k(t-\gamma)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}\exp\left\{-\left(\frac{t-\gamma}{\beta}\right)^k\right\} \qquad (5)$$

によって与えらる．ここで，本稿では β を尺度パラメータ，α, k をそれぞれ第 1，第 2 形状パラメータ，γ をシフトパラメータと呼ぶ．一般ガンマ分布は，式 (5) において $\alpha=1$ とするとワイブル分布，$\alpha=0.5$, $k=2$, $\gamma=0$ とすると半正規分布の確率密度関数を表す．

フォールト発見時刻 X_i が一般ガンマ分布に従うとき，式 (5) のすべてのパラメータを同時に推定するためには複雑な非線形方程式を解く必要がある．そこで本稿では，まず $\gamma=0$ とし，α, k を固定した場合における ω, β に対する EM アル

ゴリズムの更新式を導出し，その結果を用いて形状パラメータ α, k を推定する手法を提案する．

尺度パラメータ β に対する EM アルゴリズムの $p+1$ ステップの更新式は

$$\hat{\beta}_{p+1}^{k}=\frac{\mathrm{E}\left[\sum_{i=1}^{N_0} X_i{}^k \middle| \mathcal{D}_t;\hat{\omega}_p,\hat{\beta}_p\right]}{\mathrm{E}\left[N_0 \middle| \mathcal{D}_t;\hat{\omega}_p,\hat{\beta}_p\right]\alpha} \tag{6}$$

と表すことができる．ここで，任意の可測関数 $h(\cdot)$ に関して

$$\mathrm{E}\left[\sum_{i=1}^{N_0} h(X_i) \middle| \mathcal{D}_t;\hat{\omega}_p,\hat{\beta}_p\right] = \sum_{i=1}^{m} h(x_i)+\hat{\omega}\int_{x_m}^{\infty} h(u)f(u;\hat{\boldsymbol{\theta}})du \tag{7}$$

が成立する [2]．式 (3)，(6) に式 (7) を適用すると，EM アルゴリズムの更新式は

$$\hat{\omega}_{p+1} = m+\hat{\omega}_p\overline{F}(x_m;\alpha,\hat{\beta}_p,k), \tag{8}$$

$$\hat{\beta}_{p+1}^{k} = \frac{\sum_{i=1}^{m} x_i{}^k+\alpha\hat{\omega}_p\hat{\beta}_p^k\overline{F}(x_m;\alpha+1,\hat{\beta}_p,k)}{\alpha\{m+\hat{\omega}_p\overline{F}(x_m;\alpha,\hat{\beta}_p,k)\}} \tag{9}$$

によって与えられる．ここで，

$$\overline{F}(x;\alpha,\beta,k)=\int_{x}^{\infty} f(u;\alpha,\beta,k)du \tag{10}$$

である．

次に，形状パラメータ α, k に対する推定手法を以下のように提案する．この手法は，対数尤度を最大にすることに基づいている

対数尤度を表す関数を $\mathcal{L}(\cdot)$ と表記する．

手順 1 変化幅 $r_1=1$ とし，$m=1$ とする．

手順 2 初期値 $(\alpha_{m,1},\ k_{m,1})$ を選び，$l=1$ とする．

手順 3 $d_e=\{(\alpha_{m,l}\pm r_m,\ k_{m,l}\pm r_m),\ (\alpha_{m,l}\pm r_m,\ k_{m,l}\mp r_m)\mid \alpha_{m,l}-r_m>0, k_{m,l}-r_m>0\}$ として d_e の各値について ω, β を推定し，それぞれの対数尤度を計算する．

手順 4 α, k の更新値を

$$(\alpha_{m,l+1},\ k_{m,l+1})=\underset{d_e}{\operatorname{argmax}}\{\mathcal{L}(d_e)\}$$

とし，$(\alpha_{m,l+1},\ k_{m,l+1})=(\alpha_{m,l},\ k_{m,l})$ ならば，$(\alpha_m^*,\ k_m^*)=(\alpha_{m,l},\ k_{m,l})$ として手順 5 へ進み，そうでなければ $l=l+1$ として手順 3 に戻る．

手順 5 $|\mathcal{L}(\alpha_m^*,\ k_m^*)-\mathcal{L}(\alpha_{m-1}^*,\ k_{m-1}^*)|<\epsilon$ (ϵ は任意の数) ならば $(\alpha_m^*,\ k_m^*)$ とそのときの ω, β を推定値とし終了する．そうでなければ，$r_{m+1}=r_m/2$, $m=m+1$ として手順 3 に戻る．ただし，$\mathcal{L}(\alpha_0^*,\ k_0^*)=0$ とする．

表 1: DS1，データ数 70 個．

手法	$(\hat{\omega},\ \hat{\beta},\ \hat{\alpha},\ \hat{k})$	対数尤度
EM	(141.4, 38.39, 1.692, 0.512)	-259.92
MLE	(328.2, -44.42, 4.416, 0.048)	—

表 2: DS2，データ数 50 個．

手法	$(\hat{\omega},\ \hat{\beta},\ \hat{\alpha},\ \hat{k})$	対数尤度
EM	(59.08, 77.60, 0.715, 1.223)	-325.42
MLE	(89.22, 95.86, 0.415, 1.846)	-348.08

表 3: DS3，データ数 80 個．

手法	$(\hat{\omega},\ \hat{\beta},\ \hat{\alpha},\ \hat{k})$	対数尤度
EM	(110.2, 400.2, 0.723, 1.443)	-822.50
MLE	(-222.4, -209.8, 0.981, 1.171)	—

4. 数値例

本稿で提案した EM アルゴリズムを用いた推定手法を，実際のフォールトデータを用いて数値的に検証する．フォールトデータとして，文献 [1] に記載されている DS1(データ数 86)，DS2(データ数 136)，DS3(データ数 207) を使用し，準 Newton 法を用いた従来の最尤法 (MLE) と，今回提案した手法による推定結果との比較を行う．どちらの手法においても，初期値を $(\hat{\omega},\ \hat{\beta},\ \hat{\alpha},\ \hat{k})=(50,\ 100,\ 1,\ 1)$ とした．また今回提案した手法においては，各ステップの EM アルゴリズムの繰返回数を 10 回に設定した．

表 1～3 は，DS1 において 70 個，DS2 において 50 個，DS3 において 80 個のフォールトデータが得られた時点でそれぞれパラメータの推定を行った結果である．これらの表はパラメータ $\hat{\omega}$, $\hat{\beta}$, $\hat{\alpha}$, $\hat{k}$ の推定結果と，対応する対数尤度を示している．表 1 と表 3 に示すように，従来法による推定ではパラメータに対する制約条件に違反した推定値を算出する場合がある．これに対して，提案手法による推定では，常に制約条件を満たした最尤推定値を算出している．また，ここに掲載した以外のデータに関しても同様な結果が得られたことから，EM アルゴリズムを用いた推定手法により，安定した最尤推定値を得ることが可能であると言える．

5. まとめと今後の課題

本稿ではとくに，フォールト発見時刻に一般ガンマ分布を適用した場合における EM アルゴリズムを用いた推定手法について論じた．今後の課題としては，EM アルゴリズムを用いたパラメータの区間推定手法を開発することがあげられる．

参考文献

[1] H. Joe, Statistical inference for general order statistics and nonhomogeneous Poisson process software reliability models, *IEEE Trans. Software Eng.*, Vol. 11, pp. 1485-1490, 1989.

[2] 岡村 寛之, 渡部 保博, 土肥 正, 尾崎 俊治, EM アルゴリズムに基づいたソフトウェア信頼性モデルの推定, 電子情報通信学会論文誌 (A), Vol. J85-A No. 4, pp. 442-450, 2002.