ゼロ切断・ゼロ過剰複合分布に基づいたソフトウェア信頼性評価

*呉 敬馳, 土肥 正 (01307065), 岡村寛之 (05000041)
広島大学 先進理工系科学研究科

1. はじめに

ソフトウェアの開発プロセスでは人的な作業が多いため、バグや不具合が発生することが多い。そのため、テストプロセスでのソフトウェアの信頼性を向上させることが重要であり、ソフトウェアの信頼性を定量的に評価する方法が必要とされる。非同次ポアソン過程モデル（NHPP モデル）はソフトウェアの定量的信頼性評価に用いられている。NHPP モデルは、ポアソン分布 (P) と二項分布 (B) の複合モデルであり、簡潔で適用性が高いため、長い間ソフトウェア信頼性モデルとして使われてきた。

NHPP モデルの拡張として、Grottke and Trivedil [1] によるゼロ切断（zero-truncated; ZT）ポアソン分布を適用したモデルが提案され、その適合性が高いことが示されている。非負の整数値を取る離散形確率分布において、ゼロ切断分布は 0 の度数が観測されず、1 以上の値しか取り得ない確率分布である。一方、ゼロ過剰（zero-infated; ZI）分布とは、一定割合の 0 の度数を過剰に見積もる混合分布である。本稿では、ZI 二項分布（ZIB）、ZT ポアソン分布（ZTP）、ZI ポアソン分布（ZIP）の組み合わせを考慮した複合分布をソフトウェア信頼性モデルに適用することを考える。文献 [1] では、B と ZTP を組み合わせることで、新しい非同次ポアソン過程モデルを提案し、フォールト検出時間分布が特異でないより一般的なソフトウェア信頼性モデルを提案している。本稿では、この考え方を ZIB を用いて拡張し、さらに適用範囲の広い信頼性モデリングの枠組みを提案する

2. NHPP に基づいたソフトウェア信頼性モデル

テスト時刻 t (≥ 0) までに各々のソフトウェアフォールトが検出される確率は独立かつ同一であり、連続形確率分布関数 $G(t)$ に従うものとする。テスト開始前にソフトウェア内に潜在する総フォールト数を N とすれば、時刻 t までに検出される総フォールト数 $\{X(t), t{>}0\}$ の確率関数は二項分布 $B(N, G(t))$ によって与えられる。N は未知であり, 母数 v (>0) をもつポアソン分布 $P(v)$ に従うと仮定する。具体的に、二項分布 $B(N, G(t))$ の確率関数は以下のように与えられる。

$$Pr(X(t) = m|N) = \binom{N}{m} G(t)^m \left[1 - G(t)\right]^{N-m}. \tag{1}$$

ポアソン分布 $P(v)$ の確率関数は以下のように与えられる。

$$Pr(N = n) = \frac{v^n}{n!} e^{-v}. \tag{2}$$

この場合、$X(t)$ の無条件確率関数は

$$\begin{aligned} &Pr(X(t) = m) \\ &= \sum_{n=0}^{\infty} Pr(X = m|N = n) Pr(N = n) \\ &= \sum_{n=0}^{\infty} \binom{n}{m} G(t)^m \left[1 - G(t)\right]^{n-m} \frac{v^n}{n!} e^{-v} \\ &= \frac{\left[vG(t)\right]^m}{m!} e^{-vG(t)} \end{aligned} \tag{3}$$

となる。

3. ゼロ切断及びゼロ過剰複合確率分布

文献 [1] ではゼロ切断ポアソン分布 (ZTP) に基づいた NHPP モデルを提案している。本稿ではさらにゼロ過剰を二項分布とポアソン分布に適応し、ZTP と異なるモデルを提案する。ゼロ切断ポアソン分布 (ZTP) の確率関数は以下のように与えられる。

$$Pr(X = m) = \frac{v^m}{m!(e^v - 1)} \quad (m \geq 1). \tag{4}$$

一方、ゼロ過剰ポアソン分布 (ZIP) の確率関数は以下のように与えられる。

$$\begin{aligned} Pr(X = m) &= w_1 + (1 - w_1)e^{-v} \quad (m = 0), \\ Pr(X = m) &= (1 - w_1)\frac{v^m}{m!} e^{-v} \quad (m \geq 1). \end{aligned} \tag{5}$$

ここで、w_1 は任意パラメータであり、$0 \leq w_1 \leq 1$ である。ゼロ過剰二項分布 (ZIB) の確率関数は以下のように与えられる。

$$Pr(X=m) = w_2 + (1-w_2)(1-G(t))^n (m=0),$$
$$Pr(X=m) = (1-w_2)\binom{n}{m} G(t)^m \times (1-G(t))^{n-m} \quad (m \geq 1). \tag{6}$$

同様に、w_2 は任意パラメータであり、$0 \leq w_2 \leq 1$ である。本稿では、P、ZTP、ZIP と B、ZIB の組み合わせである 6 つの NHPP モデルを考える。

表 1: 数値例に用いた確率分布関数 $G(t)$

モデル	G(t)
Exp	$G(t) = 1 - e^{-bt}$
Gamma	$G(t) = \int_0^t \frac{c^b s^{b-1} e^{-cs}}{\Gamma(b)} ds$
Pareto	$G(t) = 1 - \left(\frac{c}{t+c}\right)^b$
TruncNormal	$G(t) = \frac{F(t)-F(0)}{1-F(0)}$, $F(t) = \frac{1}{\sqrt{2\pi b}} \int_t^{-\infty} e^{-\frac{(s-c)^2}{2b^2}} ds$
LogNormal	$G(t) = \frac{1}{\sqrt{2\pi b}} \int_{log(t)}^{-\infty} e^{-\frac{(s-c)^2}{2b^2}} ds$
TruncLogist	$G(t) = \frac{F(t)-F(0)}{1-F(0)}$, $F(t) = \frac{1}{1+e^{-\frac{t-c}{b}}}$
LogLogist	$G(t) = \frac{1}{1+e^{-\frac{log(t)-c}{b}}}$
TruncEVMax	$G(t) = \frac{F(t)-F(0)}{1-F(0)}$, $F(t) = e^{-e^{-\frac{t-c}{b}}}$
LogEVMax	$G(t) = e^{-e^{-\frac{log(t)-c}{b}}}$

4. 数値例

従来の NHPP モデルに基づいたソフトウェア信頼性モデルと本稿で新たに提案する 6 つのモデルに対する適合性比較を行う。本稿では 8 セットのグループデータで 6 種類のモデルに対する AIC を計算し、適合性を比較した。ここで、AIC は以下のように与えられる。

$$AIC = -2LLF + 2 \times (\text{the number of parameters}). \tag{7}$$

ここで、LLF は最大対数尤度である。確率分布関数 $G(t)$ は 9 種類の代表的な分布関数とする (表 1)。表 2 では従来の NHPP モデルと我々が提案した 6 つのモデルの中で、最小の AIC の値を示している。多くの場合、NHPP モデルよりも提案モデルの方が小さい AIC を与えており、従来の NHPP モデルよりも高い適合性能を示していることがわかる。

表 2: 数値例の結果 (AIC)

Data Set	NHPP	OUR BEST MODEL
DS1	73.053	**69.917**
DS2	61.695	**56.139**
DS3	87.275	**87.257**
DS4	**51.057**	51.124
DS5	**29.535**	29.910
DS6	108.831	**107.299**
DS7	**123.265**	127.134
DS8	117.475	**110.131**

5. まとめ

ソフトウェア信頼性モデリングについて，ゼロ切断・ゼロ過剰複合分布を用いることで古典的な NHPP モデルとは異なるフレームワークを提案した。また、実データを用いた適合性評価により、本稿で提案したソフトウェア信頼性モデルの有効性を示した。今後は、これらのモデルの予測性能について調べる予定である。

参考文献

[1] M. Grottke and K. S. Trivedi, "On a method for mending time to failure distributions," 2005 International Conference on Dependable Systems and Networks (DSN'05), 2005, pp. 560-569, doi: 10.1109/DSN.2005.72.