A Comparison: Some Approximations for the Aggregate Claims Distribution

by K. Ranee Thiagarajah

ABSTRACT

Several approximations for the distribution of aggregate claims have been proposed in the literature. In this paper, we have developed a saddlepoint approximation for the aggregate claims distribution and compared it with some existing approximations, such as NP_2, gamma, IG, and gamma-IG mixture. Numerical comparisons are made with the exact results, in terms of relative errors, for various combinations of claim count and claim size distributions. The saddlepoint approximation gives very satisfactory accuracy, followed by gamma-IG mixture approximation in all the applications considered in this paper.

KEYWORDS

Frequency distribution, severity distribution, aggregate claims distribution, approximation methods

1. Introduction

Aggregate claim distributions have been widely discussed in the actuarial literature. For example, see Heckman and Meyers (1983), Teugels (1985), Pentikäinen (1987), Papush et al. (2001), Hardy (2004), and Reijnen et al. (2005). In the context of insurance theory, the aggregate claims can be viewed as a sum of individual claim amounts for a random number of claim counts over a fixed time-period. In other words, it can be represented as a sum (S) of individual claim amounts $X_1, X_2, \ldots, X_N$, where N is the random number of claim counts over a fixed time period. Conditional on $N = n$, the random variables $X_1, X_2, \ldots, X_N$ are assumed to be positive, mutually independent, and identically distributed. It is also assumed that the common distribution of X_i's is independent of N. One can easily write the cumulative distribution function ($c.\ d.\ f$) of aggregate claims random variable (S) as

$$CF_s(x) = P(S \le x)$$
$$= \sum_{n=0}^{\infty} P(N=n)P(X_1 + X_2 + \ldots + X_N \le x | N = n). \tag{1.1}$$

The computation of this compound distribution function or corresponding tail probability or probability density function is generally quite cumbersome. For most combinations of distributions of N and the X_i's, the exact distribution of S is not available analytically, but the above distributional values may be obtained numerically. For discrete severity distributions, one often uses the well-known recursive method introduced by Panjer (1981) to evaluate the aggregate claims distribution. For exponential severities, simple analytical results for exact probabilities can easily be obtained (Klugman et al. 2012). For other cases, the computation in (1.1) requires tedious numerical integrations. In this situation, one often prefers to use an approximate distribution to avoid the computational complexity of distributional values. Several approximations have been developed and studied by many authors in the actuarial literature. Here, we have considered only four of them, which are NP_2, gamma, IG, and gamma-IG mixture approximations. The NP_2 approximation provided by Pesonen (1969) and the gamma approximation introduced by Bohman and Esscher (1963) are well known to the actuarial community. Chaubey et al. (1998) introduced IG and gamma-IG mixture approximations to aggregate claims distribution. They compared these approximations with NP_2 and gamma approximations for some choices of the distributions of claim counts and claim sizes. The authors stated that the gamma-IG mixture approximation uniformly improves the accuracy, especially in the tails. However, this approximation requires numerical evaluation of incomplete gamma functions. Seri and Choirat (2015) have studied a number of approximations for compound Poisson processes only and discussed their findings. Alhejaili and Abd-Elfattah (2013) discussed the saddlepoint approximation, introduced by Lugannani and Rice (1980), for some stopped-sum distributions, and noted that the approximation shows a great accuracy compared to exact distribution. Recently, Thiagarajah (2017) has studied extensively the saddlepoint approximation to the tail probabilities of aggregate claims for various combinations of claim counts and claim severity distributions. The author has compared the approximate tail probabilities with the exact probabilities, and concluded that the accuracy of this approximation is quite good in all applications considered in the paper.

The paper is organized as follows: In Section 2 we present a brief description of five of the above approximations to the cumulative distribution function of aggregate claims. In Section 3, we compare the accuracy of these approximations numerically in terms of relative errors for various combinations of claim count and claim size distributions. The relative error is computed as (approximate probability – exact probability)/exact probability. If the exact probabilities are not available analytically, we obtain them through simulations. In that case, we first generate a random number N from a claim count distribution,

and then generate N random values from a claim size distribution. The aggregate claim amount (S) is taken as the sum of those N values. Each probability is based on one hundred thousand replications. The final section contains the conclusion.

2. Approximations

In this section, we provide five approximations for the distribution of the aggregate claims. We need mean, variance, skewness, or kurtosis of the aggregate claims distribution for these approximation methods. These quantities can easily be obtained from the cumulant generating function, which is the natural logarithm of the moment generating function. Let us denote the cumulant generating function (*c.g.f*) of S as $C_S(t)$. Then, we can write the mean $= \mu_S = C_S^{(1)}(0)$, the variance $= \sigma_S^2 = \mu_2(s) = C_S^{(2)}(0)$, the third central moment $\mu_3(s) = C_S^{(3)}(0)$, and the fourth central moment $\mu_4(s) = C_S^{(4)}(0) + 3(C_S^{(2)}(0))^2$. Now, we express the skewness as $\gamma_S = \dfrac{\mu_3(s)}{\sqrt{(\mu_2(s))^3}}$ and the kurtosis as $\kappa_S = \dfrac{\mu_4(s)}{(\mu_2(s))^2}$.

2.1. NP_2 approximation

Using the well-known central limit theorem, one can approximate the aggregate claims distribution by a normal distribution. This method works only for a large volume of risks. Pesonen (1969) provided the following expression, which is called normal power (NP_2) approximation, as an adjustment to the normal approximation:

$$CF(x) \approx \Phi\left[\sqrt{1+\frac{9}{\gamma_S^2}+\frac{6Z}{\gamma_S}}-\frac{3}{\gamma_S}\right], \tag{2.1}$$

where $Z = \dfrac{(x-\mu_s)}{\sigma_s}$ and γ_S is the skewness of S. The μ_S and σ_S are the mean and the standard deviation of S. Pentikäinen (1977) claims that this approximation method gives very satisfactory results, provided that the skewness of the distribution of interest is very small.

2.2. Gamma approximation

Bohman and Esscher (1963) discussed this approximation, which is based on incomplete gamma function. Here, the aggregate claims distribution is approximated by a simple gamma distribution. An improvement to the simple gamma distribution is referred to as the gamma approximation, which is given in (2.2).

$$CF(x) \approx \frac{1}{\Gamma(\alpha)} \int_0^{\alpha+Z\sqrt{\alpha}} e^{-y} y^{a-1} dy, \tag{2.2}$$

where $Z = \dfrac{(x-\mu_s)}{\sigma_s}$ and $\alpha = \dfrac{4}{\gamma_s^2}$. Seal (1977) commented that the NP_2 method should be abandoned in favor of this gamma approximation. Gendron and Crépeau (1989) claimed that this approximation provides satisfactory results when the claim size distribution is inverse Gaussian. Pentikäinen (1977) stated that both NP_2 and gamma approximations provide similar outcomes, and both are acceptable approximations when the skewness of the aggregate claims distribution is less than two.

2.3. IG approximation

Chaubey et al. (1998) proposed this approximation, which was developed by matching the moments, as in the previous two methods. They approximated the random variable S by a shifted IG(m, b) distribution by matching the first three central moments. This means

$$CF(x) \approx G_{m,b}(x - x_0), \tag{2.3}$$

where G is the cumulative probability of the shifted IG distribution. The parameters m and b can be written as $m = \dfrac{3\left[C_s^{(2)}(0)\right]^2}{C_s^{(3)}(0)}$, $b = \dfrac{C_s^{(3)}(0)}{3C_s^{(2)}(0)}$, and $x_0 = C_S^{(1)}(0) - m$. The authors claim that the approximation is almost as good as the gamma approximation.

2.4. Gamma-IG mixture approximation

Chaubey et al. (1998) also introduced this approximation as a weighted average of gamma and IG approximations. The approximation was given as

$$CF(x) \approx wCF_1(x) + (1-w)CF_2(x), \tag{2.4}$$

where $w = \dfrac{\kappa_S - \kappa_{F_2}}{\kappa_{F_1} - \kappa_{F_2}}$, and κ stands for kurtosis. This is an improvement of the accuracy of both gamma and IG approximations.

2.5. Saddlepoint approximation

Lugannani and Rice (1980) introduced a method based on the saddlepoint technique, which can be applied to continuous and discrete distributions. It is a simple and accurate approximation for distribution function, which avoids any integration. For continuous severity distributions of Xi's, the saddlepoint approximation to the cumulative distribution function of S can be written as

$$CF(x) \approx \begin{cases} \Phi(\hat{\omega}) + \phi(\hat{\omega})\left\{\dfrac{1}{\hat{\omega}} - \dfrac{1}{\hat{u}}\right\}, & x \neq \mu \\ \dfrac{1}{2} + \dfrac{C_S^{(3)}(0)}{6\sqrt{2\pi\left(C_S^{(2)}(0)\right)^3}}, & x = \mu \end{cases} \tag{2.5}$$

where Φ and φ are the respective cumulative distribution function and the probability density function of a standard normal random variable, μ is the mean of the distribution, $\hat{w} = \operatorname{sgn}(\hat{t})\sqrt{2\{\hat{t}x - C_S(\hat{t})\}}$, and $\hat{u} = \hat{t}\sqrt{C_S^{(2)}(\hat{t})}$. The saddlepoint $\hat{t} = \hat{t}(x)$ is the unique solution to the equation $C_S^{(1)}(\hat{t}) = x$. For more mathematical details of this approximation, we refer the readers to Lugannani and Rice (1980) and Daniels (1987).

3. Some compound distributions

3.1. Poisson-gamma distribution (λ, α, θ)

In this example, the number of claims has a Poisson distribution with parameter λ, and the common severity distribution is gamma with parameters α and θ. The cumulant generating function (*c.g.f.*) of S and its first derivative are as follows:

$$C_S(t) = \lambda\left[(1-\theta t)^{-\alpha} - 1\right], \quad t < 1/\theta \tag{3.1}$$

$$C_S^{(1)}(t) = \lambda\alpha\theta(1-\theta t)^{-(\alpha+1)}. \tag{3.2}$$

Expression in (3.2) yields the saddlepoint $\hat{t}$ as $\hat{t} = \dfrac{1}{\theta}\left[1 - \left(\dfrac{\lambda\alpha\theta}{x}\right)^{\frac{1}{\alpha+1}}\right]$. From (3.1), we obtain the first four moments which are given as $\mu s = \lambda\alpha\theta$; $\mu_2(s) = \lambda\alpha(\alpha+1)\theta^2$; $\mu_3(s) = \lambda\alpha(\alpha+1)(\alpha+2)\theta^3$; $\mu_4(s) = \lambda\alpha(\alpha+1)[(\alpha+2)(\alpha+3) + 3\lambda\alpha(\alpha+1)]\theta^4$. Figures 1 and 2 display the cumulative (CF) and the tail probabilities (SF) for different parameter combinations. Figures 3 and 4 illustrate the relative errors for those combinations. For all cases, the approximation (2.5) gives excellent results, followed by gamma-IG mixture approximation.

3.2. Poisson-IG distribution (λ, μ, θ)

In this example, the claim count random variable (N) follows a Poisson distribution with parameter λ, and the common severity random variable (X) has inverse Gaussian distribution with parameters μ and θ, defined as in Klugman et al. (2012). The *c.g.f* of S and its first derivative are

$$C_S(t) = \lambda\left\{\exp\left[\frac{\theta}{\mu}\left(1-\sqrt{v}\right)\right] - 1\right\}, \quad t < \frac{\theta}{2\mu^2} \tag{3.3}$$

$$C_S^{(1)}(t) = \frac{\lambda\mu}{\sqrt{v}}\exp\left[\frac{\theta}{\mu}\left(1-\sqrt{v}\right)\right], \tag{3.4}$$

where $v = 1 - \dfrac{2t\mu^2}{\theta}$. The saddlepoint, $\hat{t} = \hat{t}(x)$, can be obtained by solving the following equation numerically:

$$\frac{\ln v}{2} + \frac{\theta\sqrt{v}}{\mu} = \frac{\theta}{\mu} - \ln\left(\frac{x}{\lambda\mu}\right). \tag{3.5}$$

Figure 1. CF: Poisson-gamma (λ, α, θ)

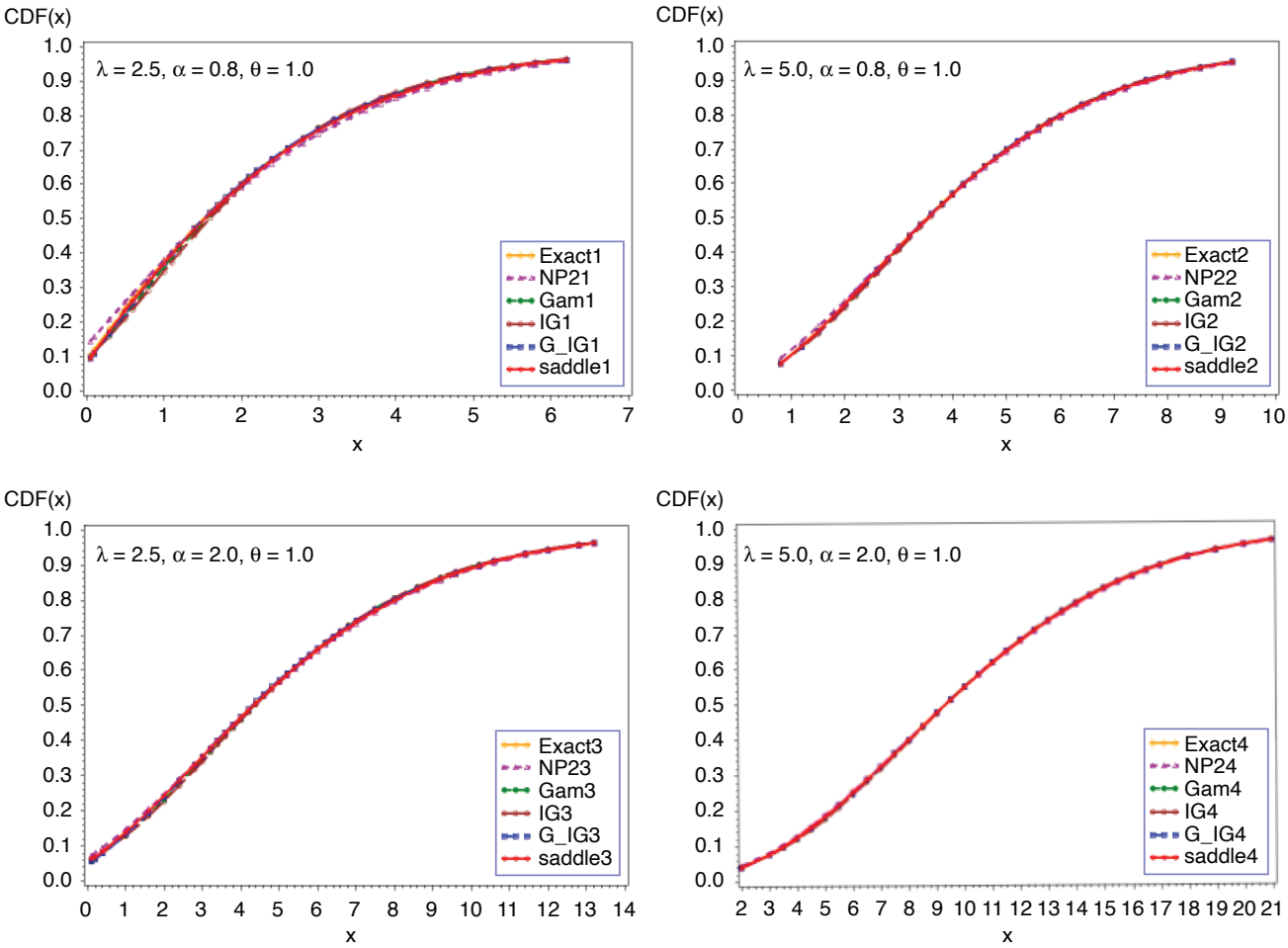


Figure 2. SF: Poisson-gamma (λ, α, θ)

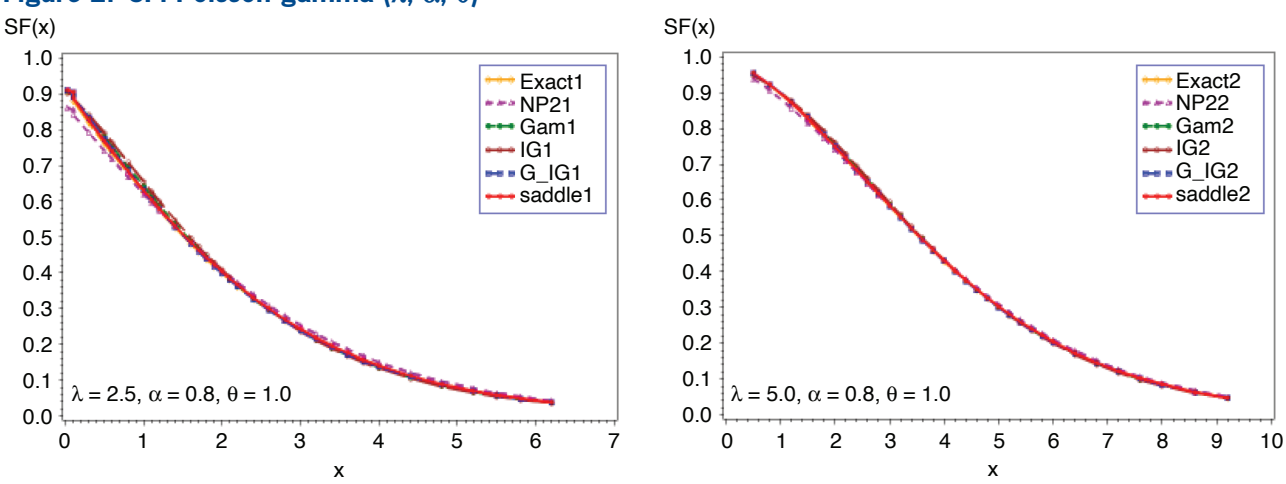


Figure 2. SF: Poisson-gamma (λ, α, θ) (*Continued*)

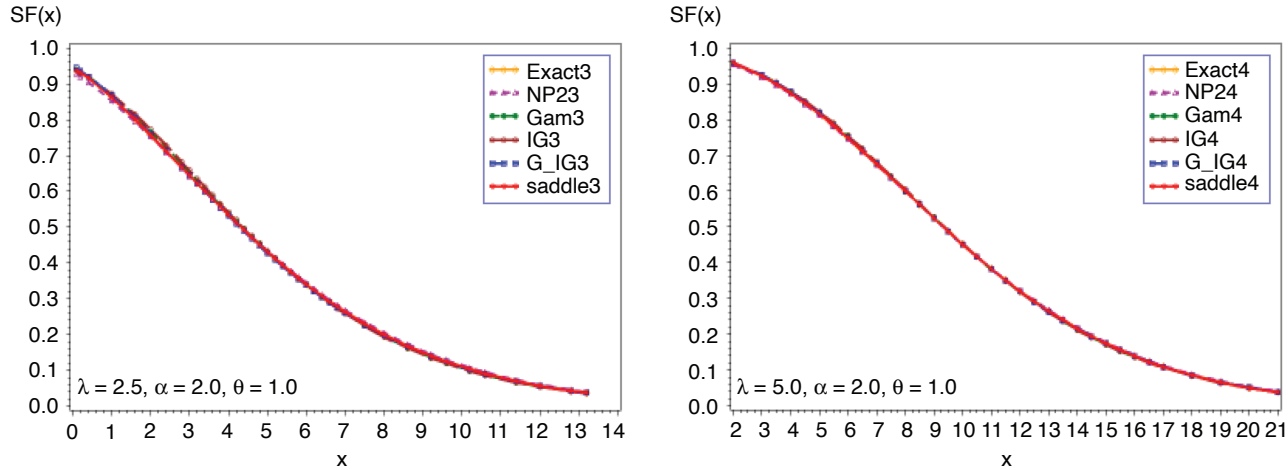


Figure 3. CF relative error: Poisson-gamma (λ, α, θ)

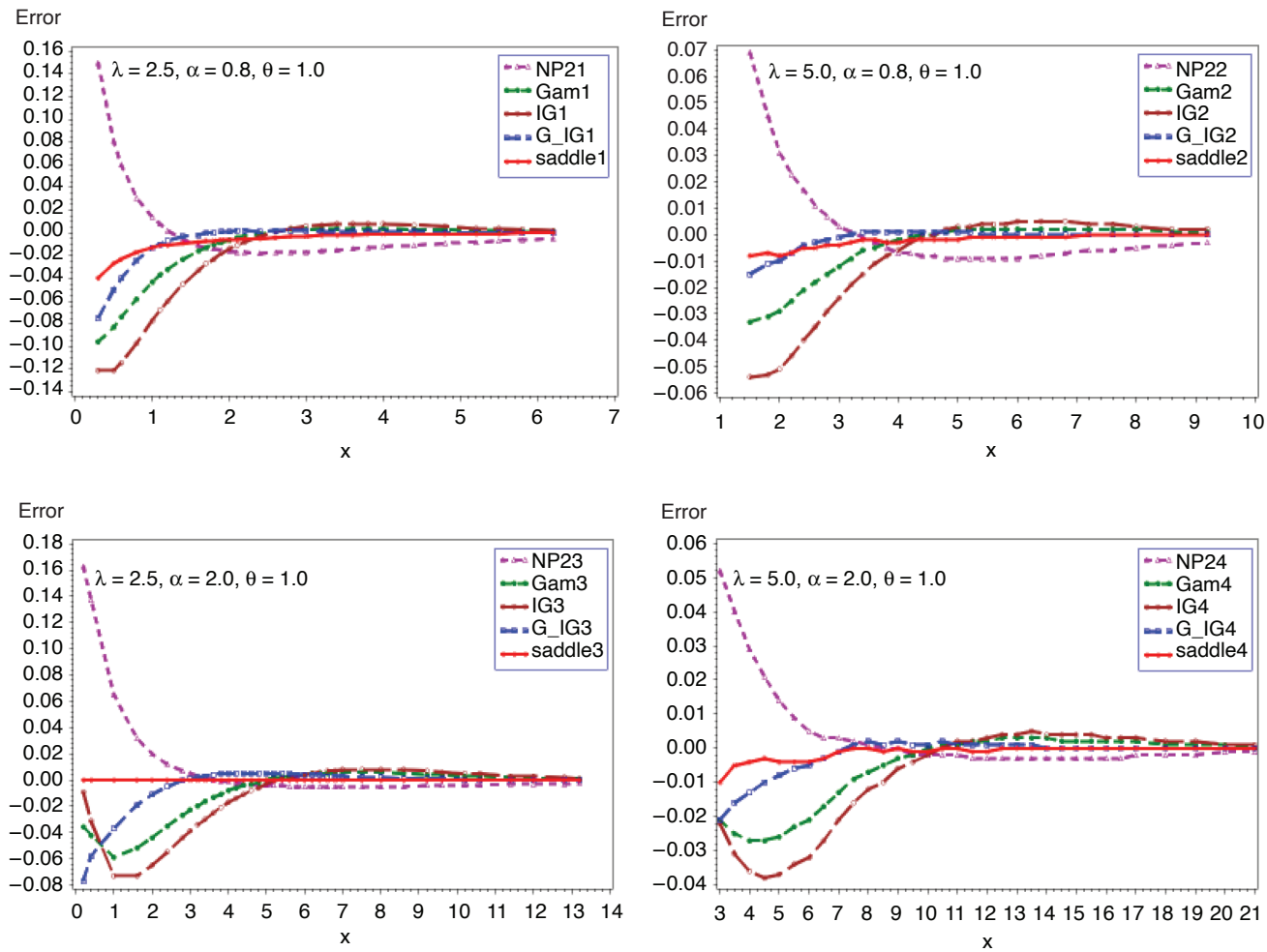
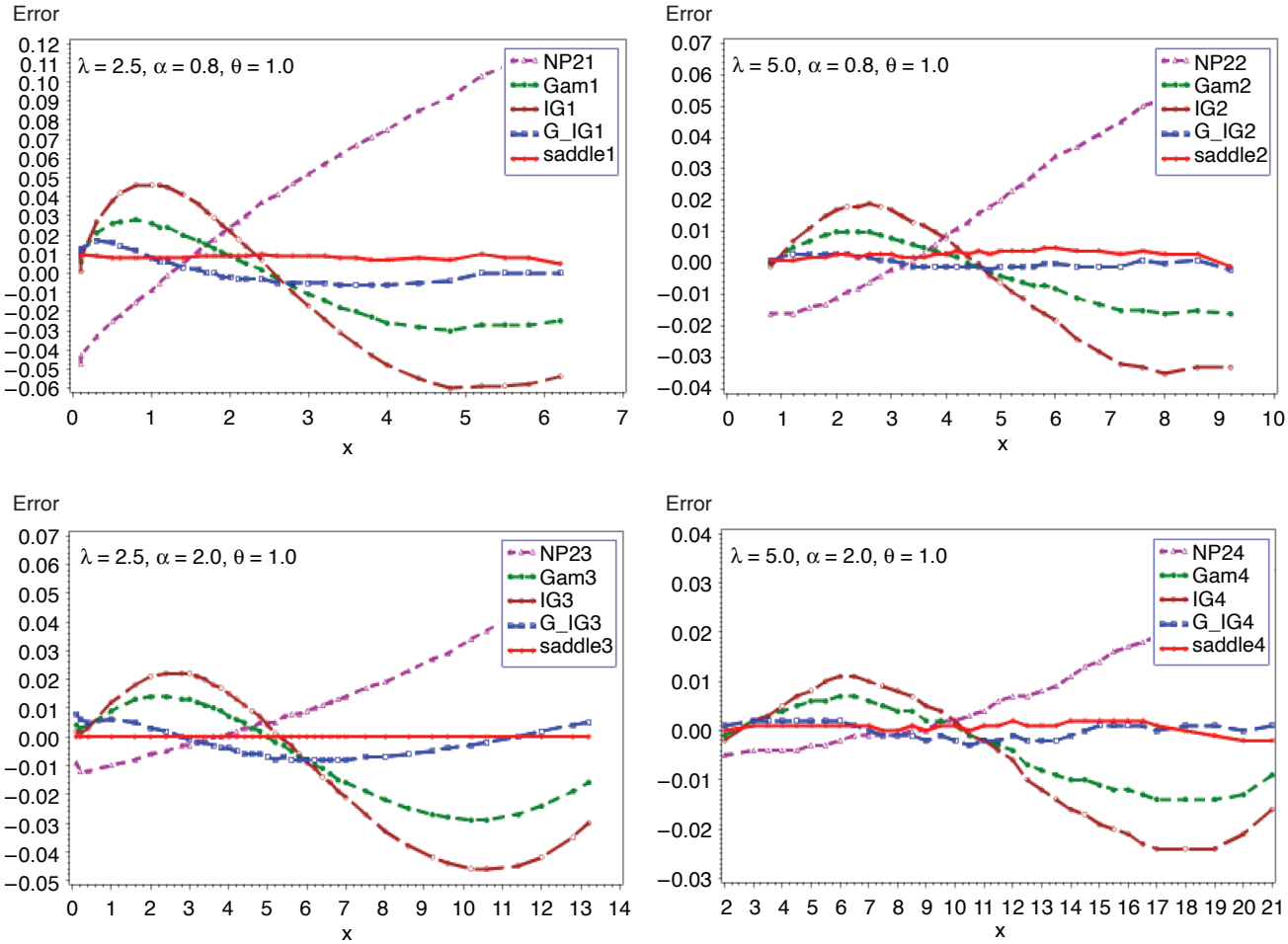


Figure 4. SF Relative error: Poisson-gamma (λ, α, θ)



Equation (3.3) yields the following quantities:

$$\mu s = \lambda\mu;$$

$$\mu_2(s) = \lambda(\mu+\theta)\mu^2/\theta;$$

$$\mu_3(s) = \lambda\left(3\mu^2+3\mu\theta+\theta^2\right)\mu^3/\theta^2$$

$$\mu_4(s) = \lambda\begin{bmatrix} 15\mu^2(\mu+\theta)+\theta^2(6\mu+\theta) \\ +3\lambda\theta(\mu+\theta)^2 \end{bmatrix}\mu^4/\theta^3.$$

Figures 5 and 6 present the cumulative and the tail probabilities for various parameter choices of frequency and severity distributions. Figures 7 and 8 display the relative errors for those cases. The approximations given in (2.4) and (2.5) are seen to have remarkably small relative errors for all cases.

3.3. Binomial-gamma distribution (*m*, *q*, α, θ)

In this example, the number of claims random variable follows a binomial distribution with parameters m and q. The common severity random variable has a gamma distribution with parameters α and θ. The *c.g.f.* of S and its first derivative are given as

$$C_S(t) = m\ln\left\{1-q+q(1-\theta t)^{-\alpha}\right\}, \quad t<1/\theta \tag{3.6}$$

$$C_S^{(1)}(t) = mq\alpha\theta\left[\frac{(1-\theta t)^{-\alpha-1}}{1-q+q(1-\theta t)^{-\alpha}}\right]. \tag{3.7}$$

Figure 5. CF: Poisson-IG (λ, μ, θ)

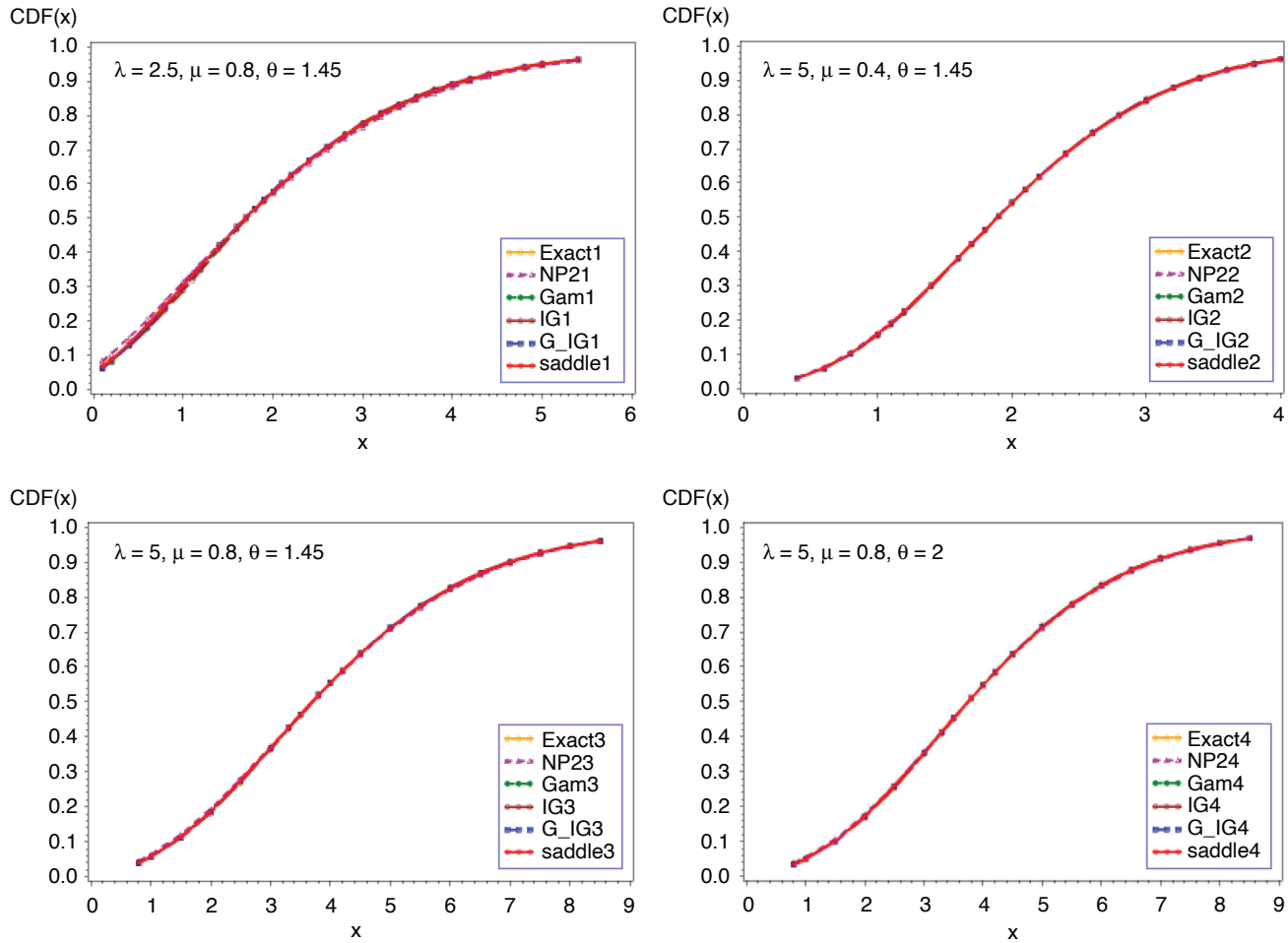
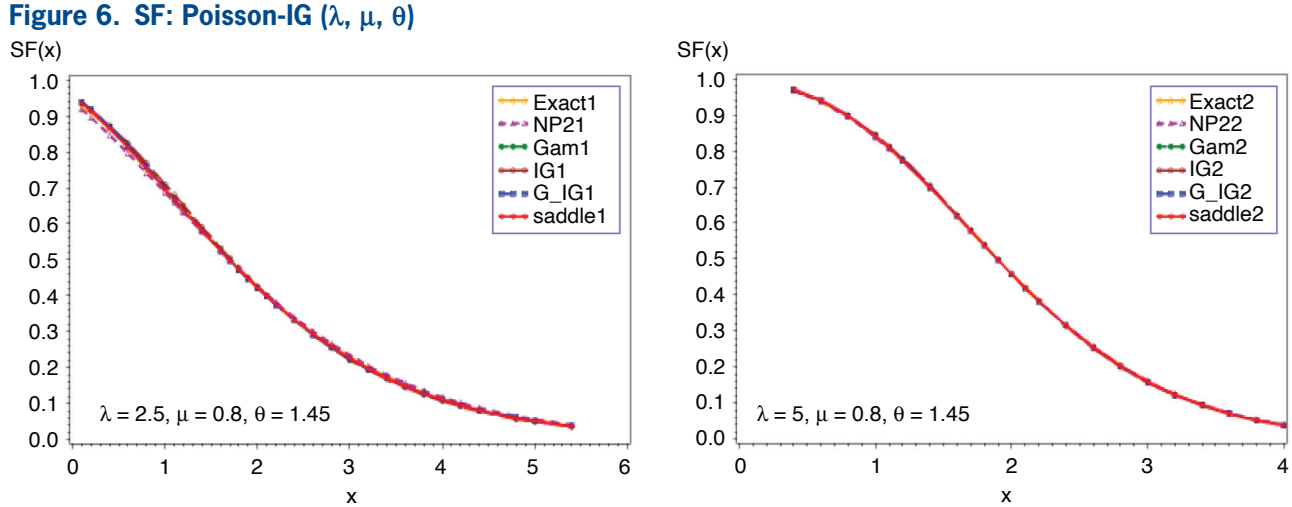


Figure 6. SF: Poisson-IG (λ, μ, θ)



(continued on next page)

Figure 6. SF: Poisson-IG (λ, μ, θ) (*Continued*)

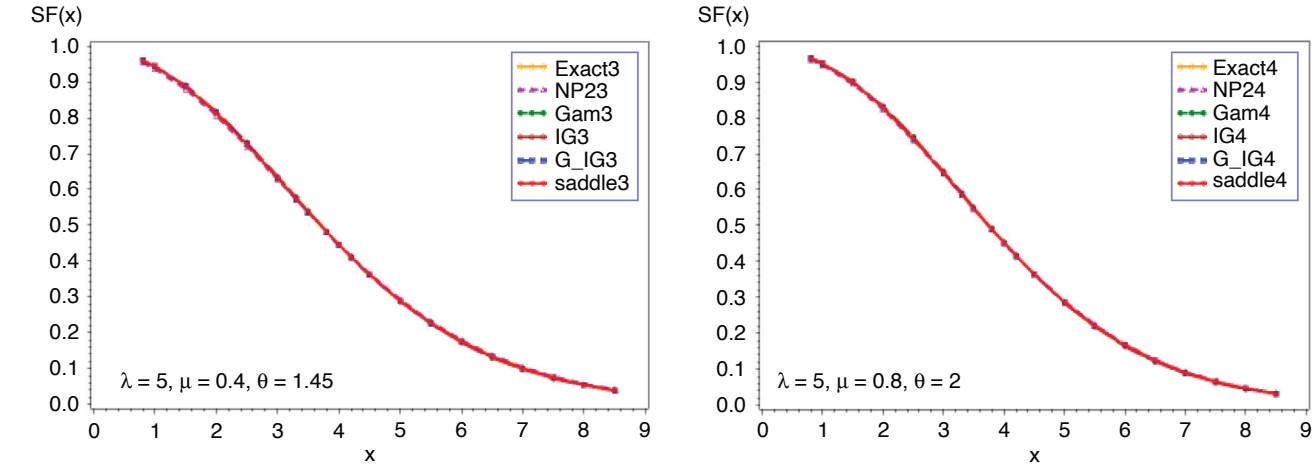


Figure 7. CF relative error: Poisson-IG (λ, μ, θ)

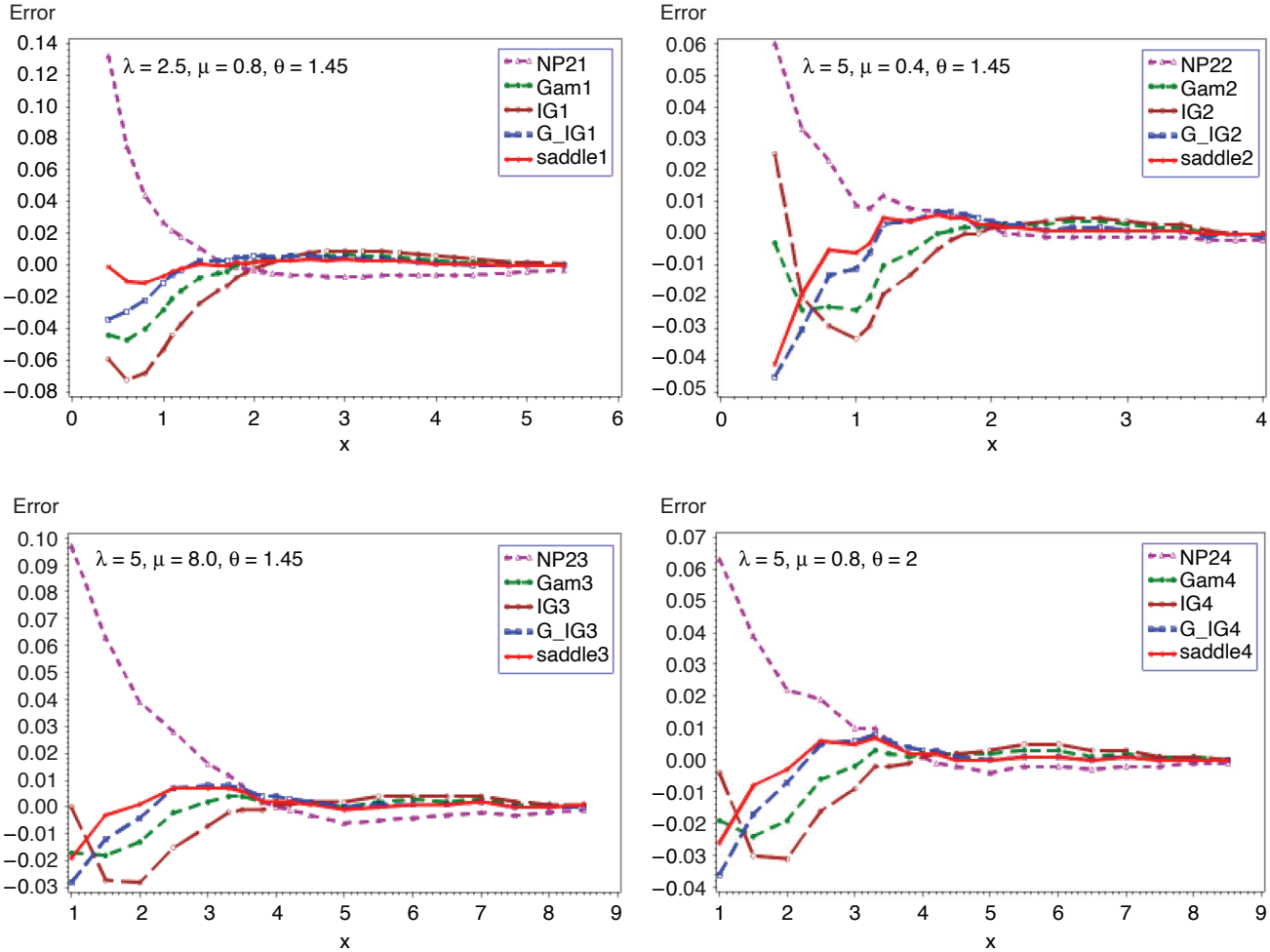
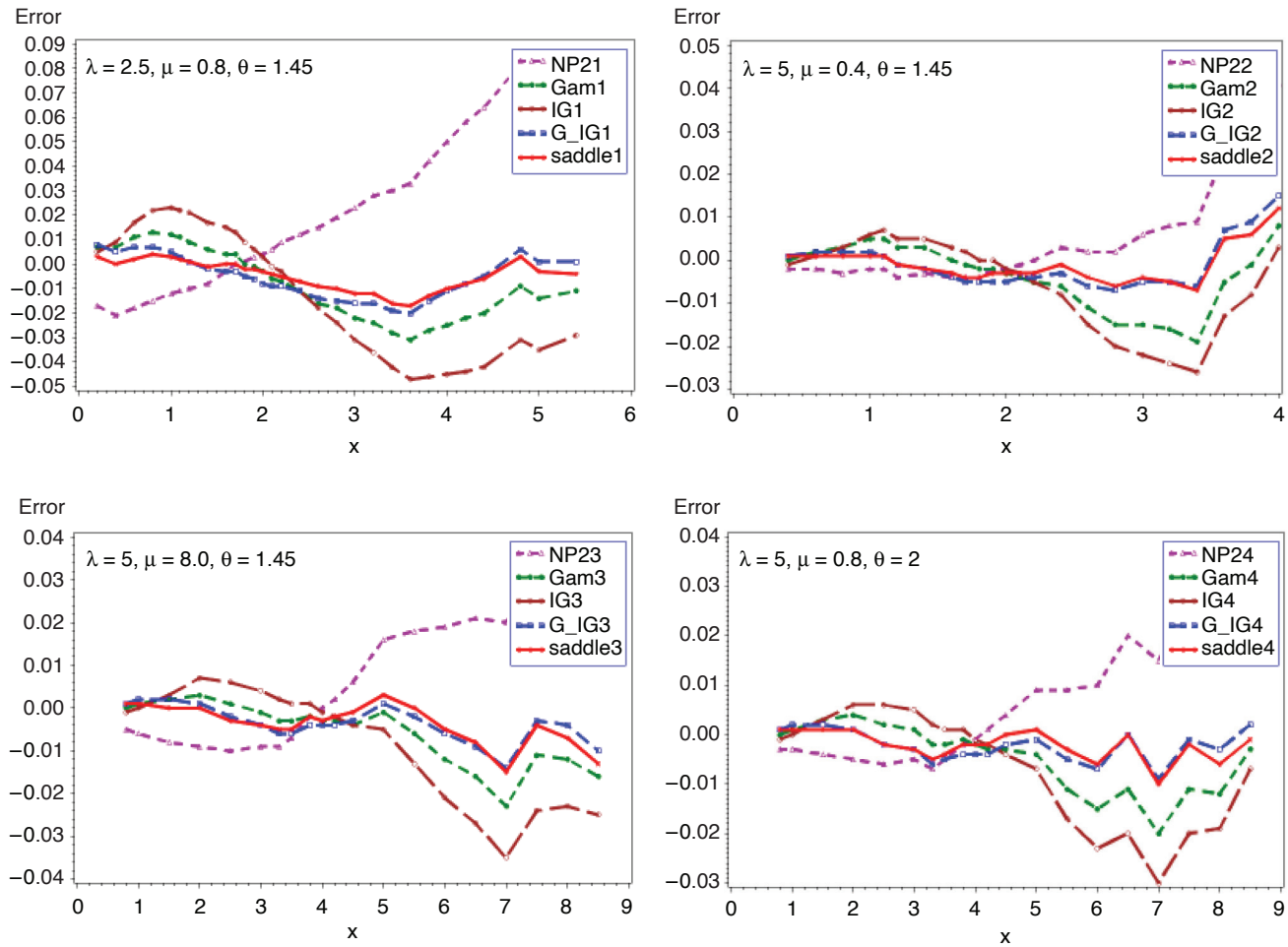


Figure 8. SF Relative error: Poisson-IG (λ, μ, θ)



The saddlepoint, $\hat{t} = \hat{t}(x)$, can be obtained numerically from

$$mq\alpha\theta\left[\frac{(1-\theta t)^{-\alpha-1}}{1-q+q(1-\theta t)^{-\alpha}}\right] = x. \tag{3.8}$$

The required moments obtained from (3.6) are as follows:

$$\mu s = mq\alpha\theta$$

$$\mu_2(s) = mq\alpha[1+(1-q)\alpha]\theta^2$$

$$\mu_3(s) = mq\alpha\left[\delta_1 + 2\alpha^2 q^2\right]\theta^3$$

$$\mu_4(s) = mq\alpha\begin{bmatrix}\delta_2 - 6\alpha^3 q^3 \\ +3mq\alpha\{1+(1-q)\alpha\}^2\end{bmatrix}\theta^4,$$

where $\delta_1 = (\alpha + 1)(\alpha + 2) - 3\alpha(\alpha + 1)q$ and $\delta_2 = (\alpha + 1)\{(\alpha + 2)(\alpha + 3) - \alpha(7\alpha + 11)q + 12\,\alpha^2 q^2\}$. Figures 9 and 10 present the cumulative and the tail probabilities for various choices of parameter values. Figures 11 and 12 depict the relative errors for those combinations. The outcomes are the same as in the previous applications. When $\alpha = 1$, the above distribution modifies to *binomial exponential* (m, q, θ). The *c.g.f.* of S and its derivative reduce to

$$C_S(t) = \ln\left(1+\frac{q\theta t}{1-\theta t}\right)^m, \quad t < 1/\theta \tag{3.9}$$

$$C_S^{(1)}(t) = m\theta\left[\frac{1}{1-\theta t} - \frac{1-q}{1-\theta t+q\theta t}\right]. \tag{3.10}$$

Figure 9. CF: Binomial gamma (*m*, *q*, α, θ)

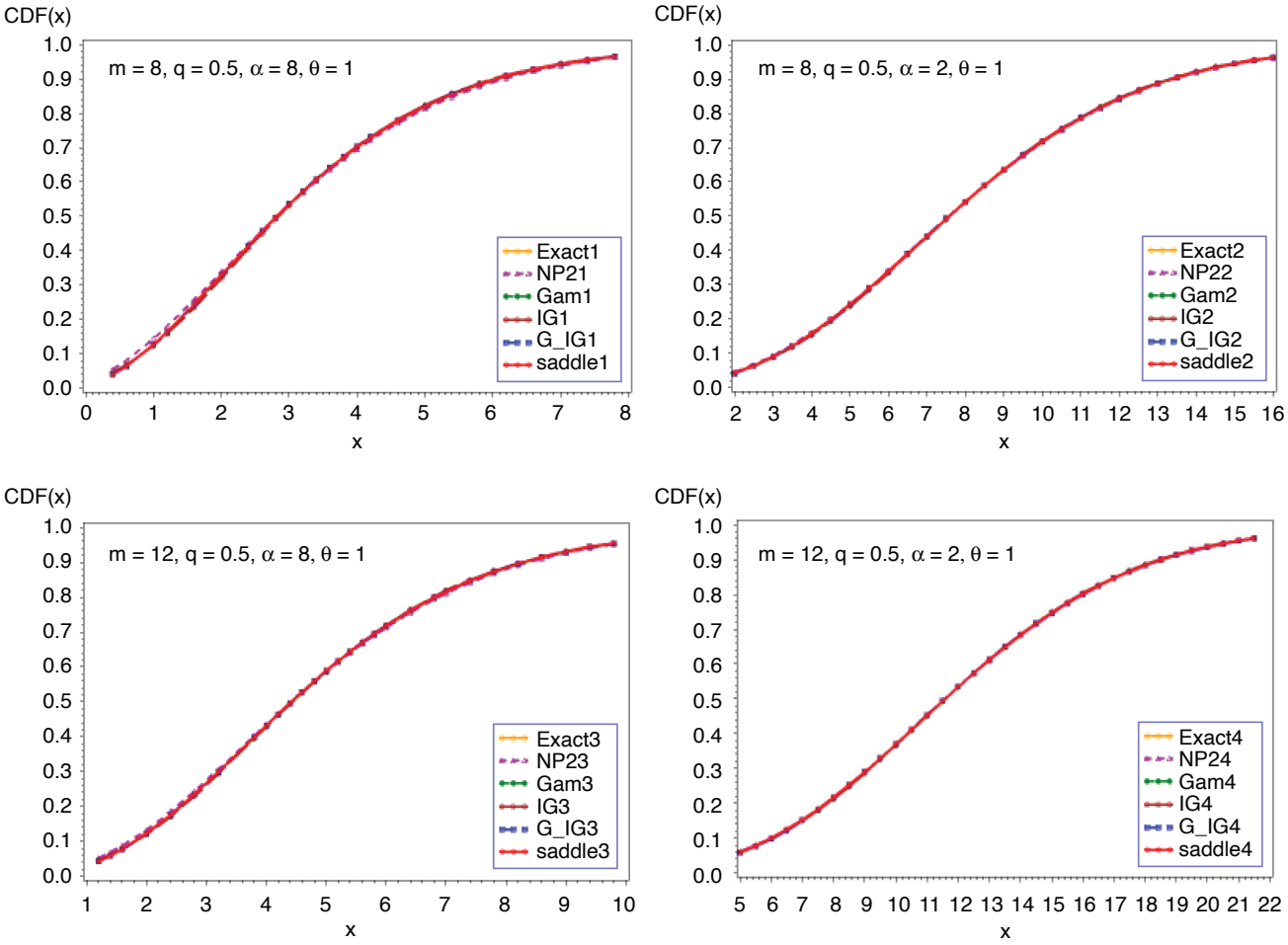


Figure 10. SF: Binomial gamma (*m*, *q*, α, θ)

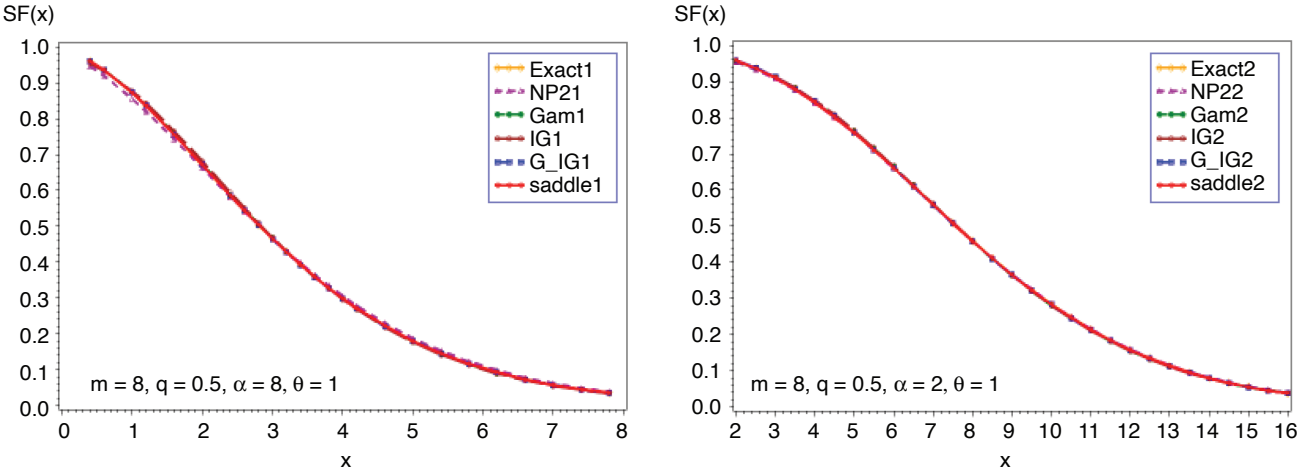


Figure 10. SF: Binomial gamma (*m*, *q*, α, θ) (*Continued*)

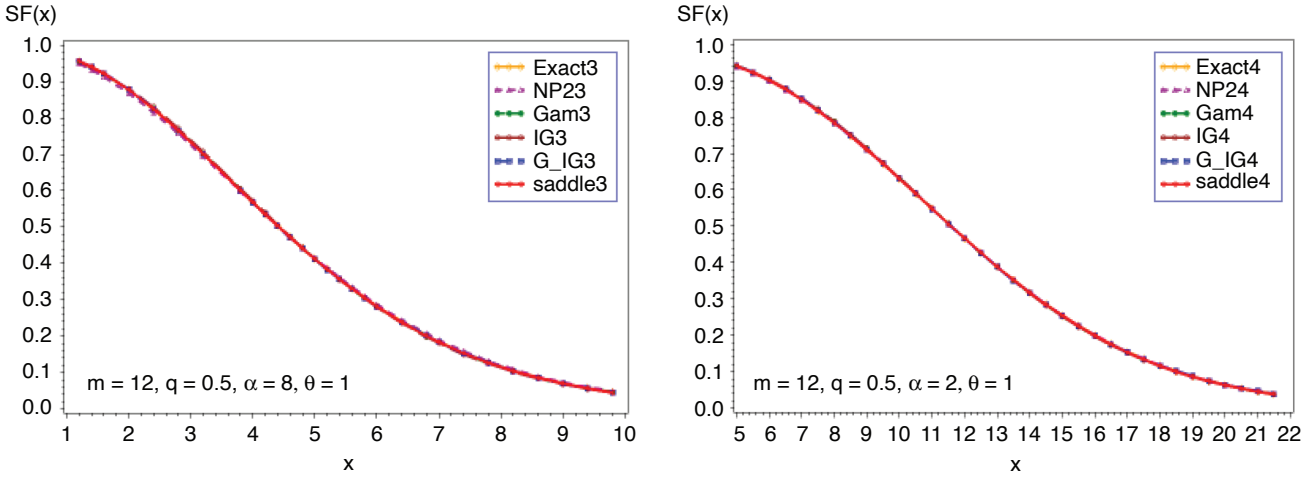


Figure 11. CF relative errors: binomial gamma (*m*, *q*, α, θ)

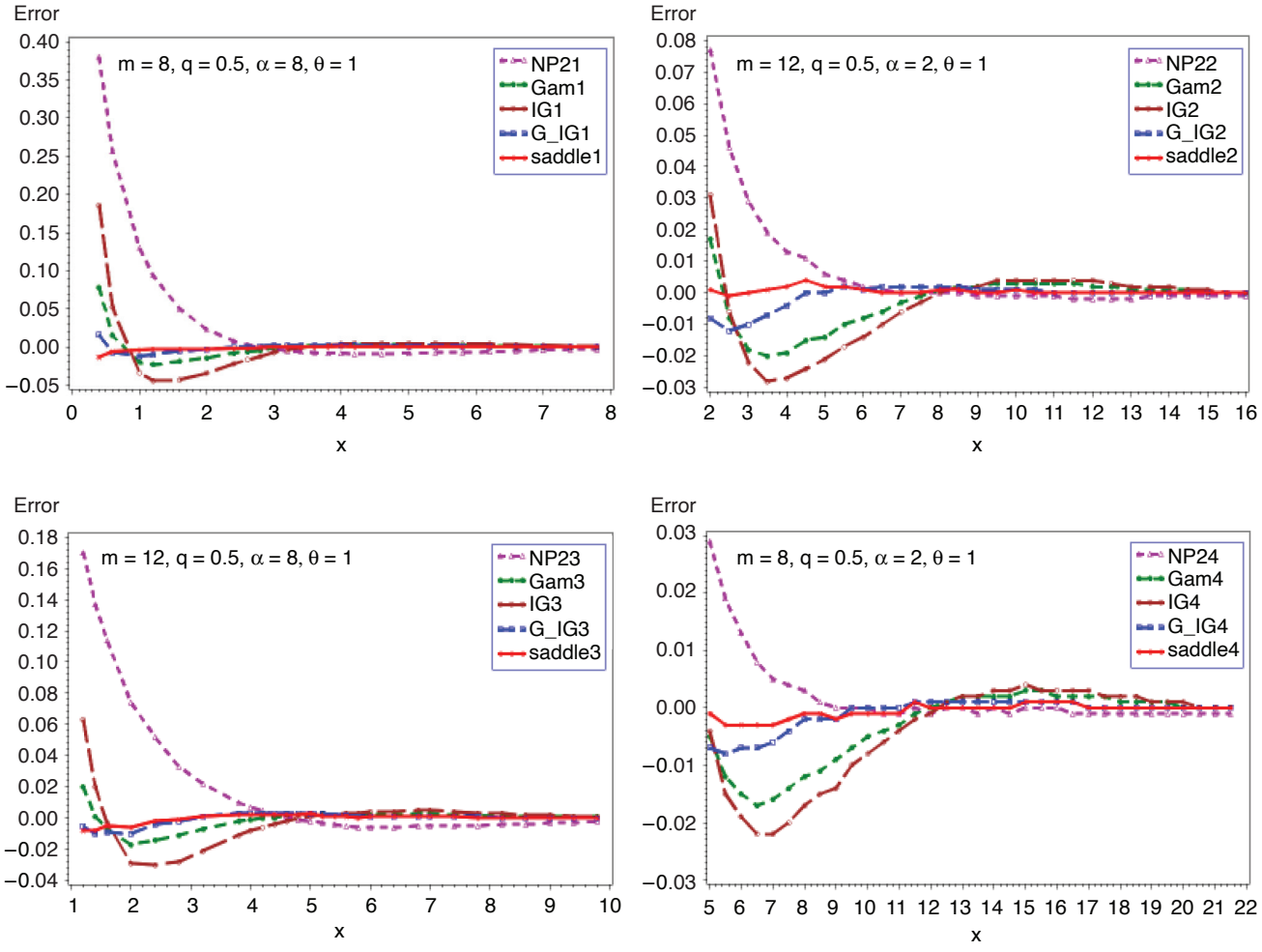
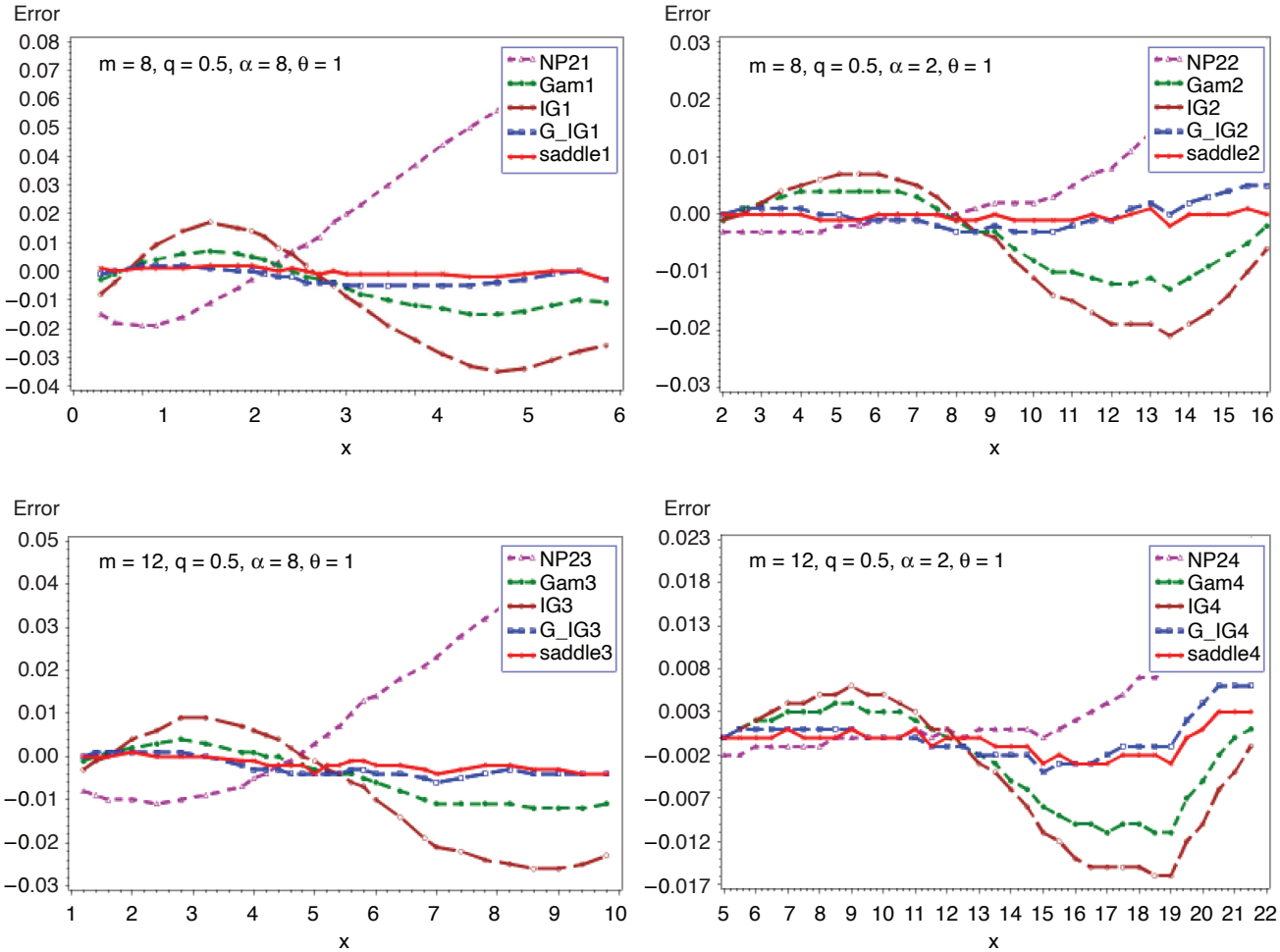


Figure 12. SF Relative errors: binomial gamma (m, q, α, θ)



The saddlepoint equation leads to the explicit solution

$$\hat{t} = \frac{(2-q) - \sqrt{q^2 + \dfrac{4mq(1-q)\theta}{x}}}{2(1-q)\theta}. \tag{3.11}$$

$$\mu s = mq\theta$$

$$\mu_2(s) = m\left[1-(1-q)^2\right]\theta^2$$

$$\mu_3(s) = 2m\left[1-(1-q)^3\right]\theta^3$$

$$\mu_4(s) = 3m\left[2\left\{1-(1-q)^4\right\} + mq^2(2-q)^2\right]\theta^4.$$

The analytical expression for the cumulative probability of S, which can easily be obtained from (1.1), is

$$CF(x) = 1 - \sum_{n=1}^{m} \binom{m}{n} q^n (1-q)^{m-n} \sum_{j=0}^{n-1} \frac{(x/\theta)^j \exp(-x/\theta)}{j!}. \tag{3.12}$$

Cumulative and tail probabilities, along with the relative errors of the approximations, are presented in Figures 13 and 14. The exact probabilities are obtained from (3.12). As can be seen in the previous examples, the results in Figures 13 and 14 indicate that the saddlepoint approximation gives very satisfactory accuracy.

3.4. NB-exponential distribution (r, β, θ)

In this example, frequency distribution is negative binomial with parameters r and β, and the common severity distribution is exponential with

Figure 13. CF and relative errors: binomial exponential (*m*, *q*, θ)

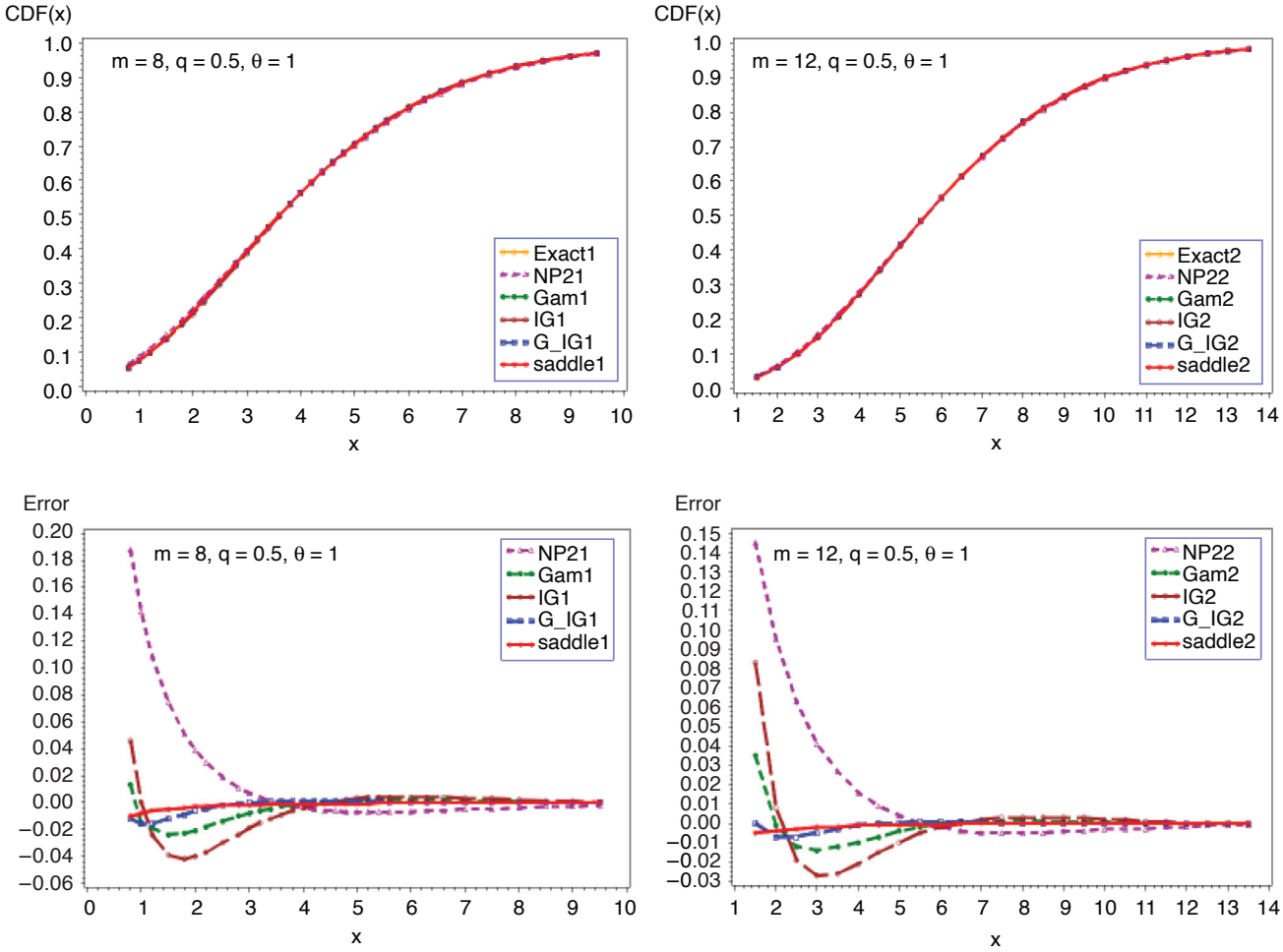
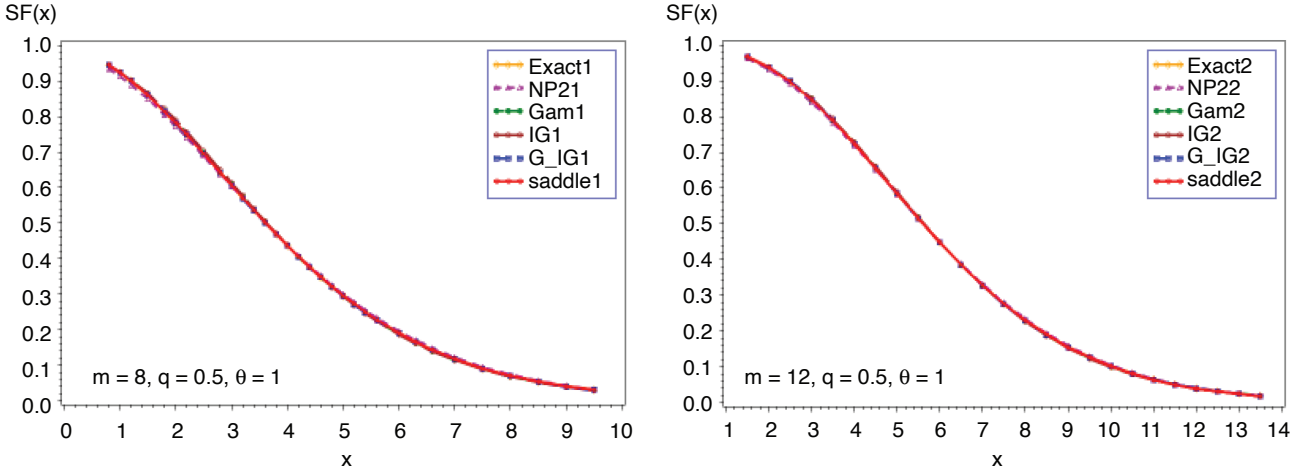
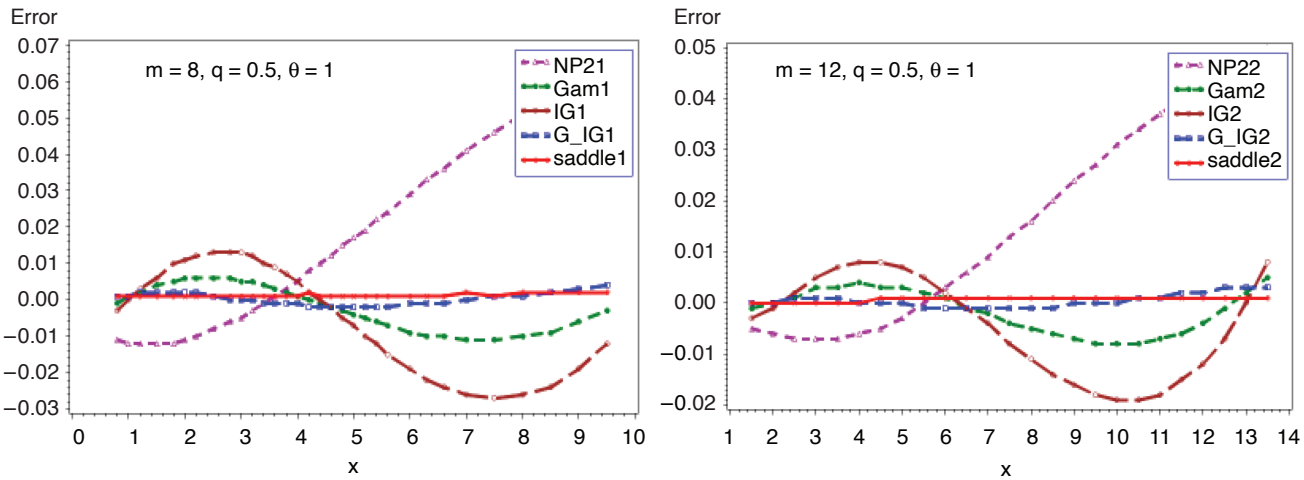


Figure 14. SF and relative errors: binomial exponential (*m*, *q*, θ)



(continued on next page)

Figure 14. SF and relative errors: binomial exponential (*m*, *q*, θ) (*Continued*)



parameter θ. The *c.g.f.* of S and its first derivative can be written as

$$C_S(t) = r\ln\left(\frac{1-\theta t}{1-\theta t-\beta\theta t}\right) \tag{3.13}$$

$$C_S^{(1)}(t) = \frac{r\beta\theta}{(1-\theta t)(1-\theta t-\beta\theta t)}. \tag{3.14}$$

Equation (3.14) yields an explicit saddlepoint solution:

$$\hat{t} = \frac{(\beta+2)-\sqrt{\beta^2+\dfrac{4r\beta\theta(1+\beta)}{x}}}{2\theta(1+\beta)}. \tag{3.15}$$

Equation (3.13) yields the following moments:

$$\mu s = r\beta\theta$$

$$\mu_2(s) = r\beta(\beta+2)\theta^2$$

$$\mu_3(s) = 2r\left[(1+\beta)^3-1\right]\theta^3$$

$$\mu_4(s) = 3r\left[2\left\{(1+\beta)^4-1\right\}+r\beta^2(\beta+2)^2\right]\theta^4.$$

The following analytical expression for the cumulative probability can be obtained from (1.1):

$$CF(x) = 1-\sum_{n=1}^{r}\binom{r}{n}\left(\frac{\beta}{1+\beta}\right)^n\left(\frac{1}{1+\beta}\right)^{r-n}$$

$$\sum_{j=0}^{n-1}\frac{(x/\theta(1+\beta))^j\, e^{-x/\theta(1+\beta)}}{j!}. \tag{3.16}$$

Figures 15 and 16 present the comparison of all five approximations. The exact probability is computed based on the analytical expression given in (3.16). The outcomes in Figures 15 and 16 reveal that the approximation given in (2.5) outperforms the other four, followed by the gamma-IG mixture approximation provided in (2.4). Corresponding results for geometric-exponential distribution (β, θ) can be obtained by letting $r = 1$.

4. Conclusions

The purpose of this paper is to compare the accuracy of the saddlepoint approximation, introduced by Lugannani and Rice (1980), with four other approximations to the distribution of aggregate claims. The approximate cumulative probabilities and tail probabilities have been computed for several compound distributions, and are compared with the exact results in terms of relative errors. Based on the results in Figures 1–16, it is clear that the saddlepoint approximation gives very satisfactory accuracy, followed by the gamma-IG mixture approximation in all of the examples considered. The gamma and the IG approximations behave in a similar manner. The NP_2 approximations consistently produce higher relative errors compared to the other four. Suppose claim amounts distribution is gamma or inverse Gaussian. The saddlepoint approximation is simple and easy to

Figure 15. CF and relative errors: NB-exponential (r, β, θ)

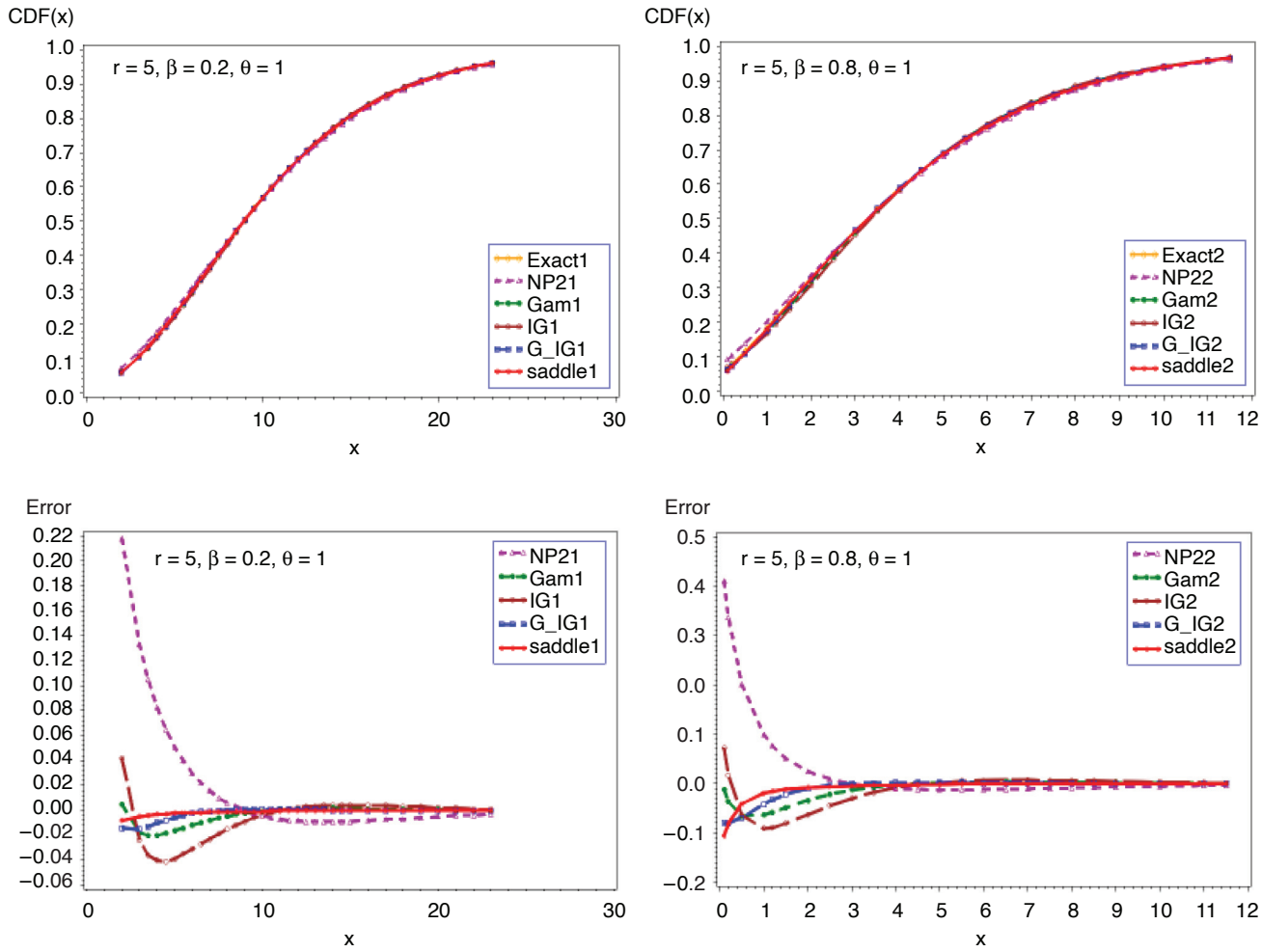
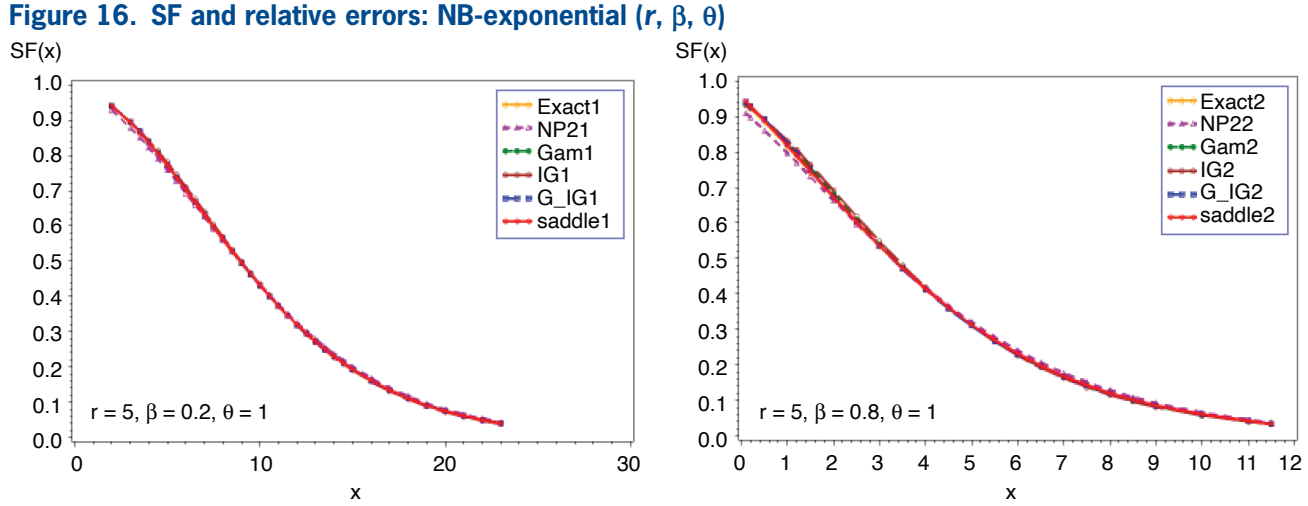
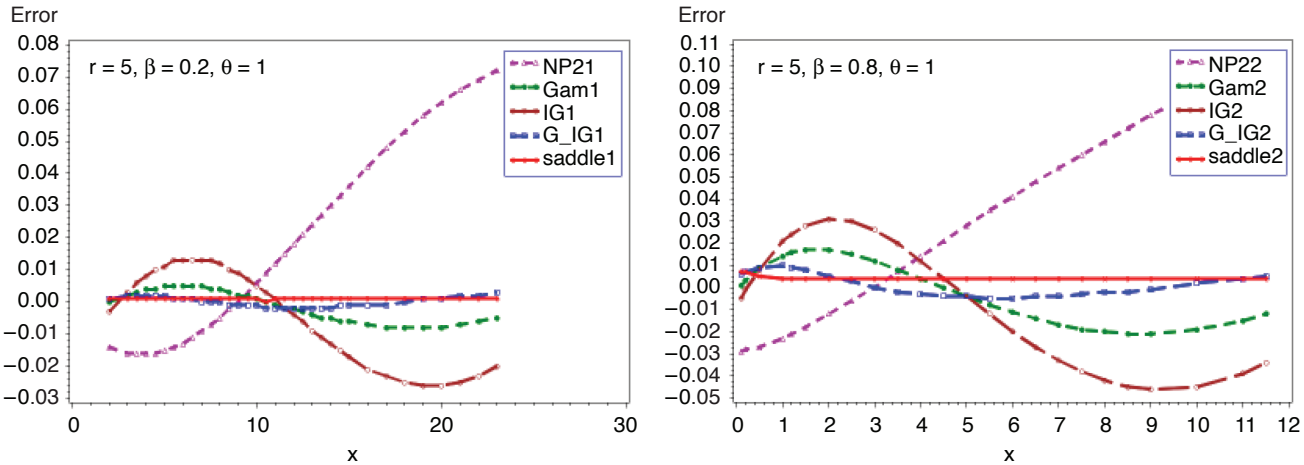


Figure 16. SF and relative errors: NB-exponential (r, β, θ)



(continued on next page)

Figure 16. SF and relative errors: NB-exponential (r, β, θ) (*Continued*)



compute with great accuracy, requiring only the first three derivatives of the cumulant generating function.

Acknowledgment

Author would like to thank the editor and three anonymous reviewers for their constructive comments/suggestions that greatly improved the manuscript.

References

Alhejaili, A. D., and E. F. Abd-Elfattah, "Saddlepoint Approximations for Stopped-sum Distributions," *Communications in Statistics: Theory and Methods* 42, 2013, pp. 3735–3743.

Bohman, H. and F. Esscher, "Studies in Risk Theory with Numerical Illustrations," *Scandinavian Actuarial Journal* 46, 1963, pp. 173 – 225.

Chaubey, Y. P., and J. Garrido, and S. Trudeau, "On the Computation of Aggregate Claims Distributions: Some New Approximations," *Insurance: Mathematics and Economics* 23, 1998, pp. 215–230.

Daniels, H. E., "Tail Probability Approximations," *International Statistical Review* 55, 1987, pp. 37–48.

Gendron, M., and H. Crépeau, "On the Computation of the Aggregate Claim Distribution when Claims are Inverse Gaussian," *Insurance: Mathematics and Economics* 8, 1989, pp. 251–258.

Hardy, M. R., "Approximating the Aggregate Claim Distribution," *Encyclopedia of Actuarial Science* I, 2004, pp. 76–86.

Heckman, P. E., and G. G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *Proceedings of the CAS* LXX, 1983, pp. 22–61.

Klugman, S.A., H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, New York: John Wiley, 2012.

Lugannani, R., and S. O. Rice, "Saddlepoint Approximations for the Distribution of the Sum of Independent Random Variables," *Advances in Applied Probability* 12, 1980, pp. 475–490.

Panjer, H. H., "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin* 12, 1981, pp. 22–26.

Papush, D. E., G. S. Patrik, and F. Podgaits, "Approximations of the Aggregate Loss Distribution," *Casualty Actuarial Society Forum* Winter, 2001, pp. 175–186.

Pentikäinen, T., "On the Approximation of the Total Amount Claims," *ASTIN Bulletin* 9, 1977, pp. 281–289.

Pentikäinen, T., "Approximate Evaluation of the Distribution Function of Aggregate Claims," *ASTIN Bulletin* 17, 1987, pp. 15–39.

Pesonen, E., "NP- Approximation of Risk Processes," *Scandinavian Actuarial Journal* 3, 1969, pp. 63–69.

Reijnen, R., W. Albers, and W. C. M. Kallenberg, "Approximations for Stop-loss Reinsurance Premiums," *Insurance: Mathematics and Economics* 36, 2005, pp. 237–250.

Seal, H. L. (1977). Approximations to risk theory's $f(x, t)$ by means of the gamma distribution. *ASTIN Bulletin* 9: pp. 213–218.

Seri, R. and C. Choirat, "Comparison of Approximations for Compound Poisson Processes," *ASTIN Bulletin* 45, 2015, pp. 601–637.

Teugels, J. L., "Approximation and Estimation of Some Compound Distributions," *Insurance: Mathematics and Economics* 4, 1985, pp. 143–153.

Thiagarajah, K., Approximate Tail Probabilities of Aggregate Claims," *International Journal of Statistics and Economics* 18, 2017, pp. 1–11.