The Korean Journal of Applied Statistics (2009)
22(6), 1345–1353

The Use of Generalized Gamma-Polynomial Approximation for Hazard Functions

Hyung-Tae Ha[1]

Department of Applied Statistics Kyungwon University

(Received August 2009; accepted October 2009)

Abstract

We introduce a simple methodology, so-called generalized gamma-polynomial approximation, based on moment-matching technique to approximate survival and hazard functions in the context of parametric survival analysis. We use the generalized gamma-polynomial approximation to approximate the density and distribution functions of convolutions and finite mixtures of random variables, from which the approximated survival and hazard functions are obtained. This technique provides very accurate approximation to the target functions, in addition to their being computationally efficient and easy to implement. In addition, the generalized gamma-polynomial approximations are very stable in middle range of the target distributions, whereas saddlepoint approximations are often unstable in a neighborhood of the mean.

Keywords: Hazard function, survival function, generalized gamma-polynomial approximation, moments, convolutions, mixtures.

1. Introduction

Survival and hazard functions are often of interest among parametric functions for statistical analysis of survival data. In connection with modeling survival data, distributions such as the exponential, the gamma, the Weibull and Gompertz distributions are often utilized. The more complicated situations arise, the more complex modelings such as convolutions or finite mixtures need to be utilized. Convolutions of exponential distributions are, for instance, often used to model the distributions of waiting times in each of the progressive stages in multi-state models, see Keilson (1978), and convolution of two Weibull distributions to model progression of cancer, progression of tumor growth, and failure of organs, see Huzurbazar and Huzurbazar (1999). And mixtures of distributions have also often been used to investigate the bi-modality or multi-modality of data generated from nonhomogeneous populations. For instance, Pierce *et al.* (1979) used mixture model to relate time-to-death data to toxicant levels and other stresses.

It should be noted that although it is useful to model complex phenomenon via convolutions or mixtures from other types of distributions such as the gamma or the Weibull distributions, there are limitations to use such models since the models may not be analytically tractable. It should be

[1] Assistant professor, Department of Applied Statistics Kyungwon University, Sungnam-ci Kyunggi-Do 461-701, Republic of Korea. E-mail: htha@kyungwon.ac.kr

mentioned that the exact expression for the density or distribution functions of the convolution of gamma and Weibull distributions are very complicated. In such circumstances, it is desirable to use approximation techniques to obtain such statistical functions of interests. Saddlepoint methods, which was introduced by Daniels (1954), have been often used to approximate survival and hazard functions in the context of parametric survival analysis, see, for instance, Daniels (1982), Butler and Huzurbazar (1997), Huzurbazar and Huzurbazar (1999), Butler and Huzurbazar (2000) and Butler (2000). It should be mentioned that saddlepoint approximations for convolution of the Weibull distributions require to calculate either moment or cumulant generating function, which are not expressed in closed form. Therefore, the alternative way to utilize saddlepoint methods in such cases requires evaluation of higher-dimensional integrals.

We introduce generalized gamma-polynomial method to approximate survival and hazard functions. Since the generalized gamma-polynomial approximant basically comprises of a product of generalized gamma baseline density and a polynomial adjustment, the approximated cumulative distribution function can be expressed simple. The approximated hazard function can also be easily determined as the ratio of approximated density function and survival function. An explicit representation of the resulting approximants of hazard functions can be obtained in terms of incomplete gamma function. The generalized gamma-polynomial approximations are very stable in middle range of the target distributions, whereas saddlepoint approximations are often unstable in a neighborhood of the mean, as noted by Reid (1996).

We are aiming to illustrate that the use of generalized gamma-polynomial approximation has benefits of simplicity and accuracy to approximate the survival and hazard functions in certain probabilistic setting. Section 2 briefly reviews the generalized gamma-polynomial approximations. Section 3 and 4 illustrate the use of generalized gamma-polynomial approximation technique in approximating survival and hazard functions of convolutions and mixtures of random variables, respectively. And in both cases, we show numerical examples. Concluding remarks are described in Section 5.

2. Generalized Gamma-Polynomial Approximation

A general semi-parametric approach to density approximation is proposed in Ha and Provost (2007). In this section, we review the approximation technique on the use of generalized gamma baseline density function for approximating density and distribution functions, from which the approximated survival and hazard functions are obtained.

Letting X and $E(X^h)$ be a random variable whose support is the real half line and its raw moments, denoted by $\mu_X(h)$, $h = 0, 1, \ldots$, respectively, we first are interested in approximating the exact density and distribution functions of the random variable X, denoted by $f_X(x)$ and $F_X(x)$, respectively. A *generalized gamma-polynomial* density approximant of degree d, denoted by $\hat{f}_X(x;\, d)$, is

$$\hat{f}_X(x;\, d) = g(x) \sum_{i=0}^{d} \xi_i x^i, \tag{2.1}$$

where

$$g(w) \;=\; \frac{\gamma}{\beta^{\alpha\,\gamma}\Gamma(\alpha)} w^{\alpha\,\gamma-1} e^{-\left(\frac{w}{\beta}\right)^{\gamma}}\, \mathcal{I}_{(0,\infty)}(w) \tag{2.2}$$

is the generalized gamma baseline density function with three parameters $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ on letting $\mathcal{I}_A(x)$ denote the indicator function, which is equal to 1 when $x \in A$ and 0 otherwise,

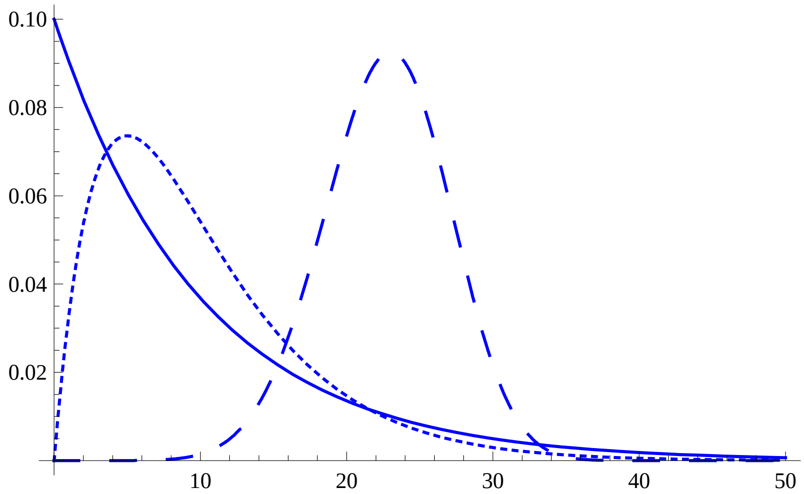

Figure 2.1. Exponential(solid), Gamma(dotted) & Weibull Densities(dashed)

and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. The generalized gamma distribution is an extensive family that contains nearly all of the most commonly used distributions, including the exponential, Weibull, lognormal and gamma. More importantly, the generalized gamma distribution family includes commonly used types of hazard function: monotonically increasing and decreasing, as well as bathtub and arc-shaped hazards. As seen in Figure 2.1, the Weibull distribution is a special case of the generalized gamma when $\alpha = 1$, the exponential distribution when $\alpha = 1$ and $\gamma = 1$ and the gamma distribution when $\gamma = 1$. The j^{th} moments of the generalized gamma baseline density function, denoted by $\mu_G(j)$, can be expressed as;

$$\mu_G(j) = \int_0^\infty x^j g(x)dx = \frac{\beta^h\,\Gamma[\alpha + h/\gamma]}{\Gamma[\alpha]}, \quad j = 0, 1, \ldots. \tag{2.3}$$

Three parameters α, β and γ of the generalized gamma baseline density function are estimated from the first three moments of the target distribution as follows:

$$\begin{aligned}\mu_X(1) &= \frac{\beta\,\Gamma[\alpha + 1/\gamma]}{\Gamma[\alpha]}\\ \mu_X(2) &= \frac{\beta^2\,\Gamma[\alpha + 2/\gamma]}{\Gamma[\alpha]}\\ \mu_X(3) &= \frac{\beta^3\,\Gamma[\alpha + 3/\gamma]}{\Gamma[\alpha]}.\end{aligned} \tag{2.4}$$

From the moment matching technique between the moments of the target distribution and those of the generalized gamma-polynomial density approximant, we can obtain the coefficients ξ_i of the polynomial adjustment. That is, the coefficients ξ_i satisfy the equality

$$(\xi_0, \ldots, \xi_d)' = \mathcal{M}^{-1}(\mu_X(0), \ldots, \mu_X(d))'\,, \tag{2.5}$$

a prime denoting the transpose of a vector and $\mathcal{M}$ being an $(d+1)\times(d+1)$ moment matrix whose $(h+1)^{th}$ row is $(\mu_G(h), \ldots, \mu_G(h+d))'$, $h = 0, 1, \ldots, d$.

The approximate cumulative distribution function of X, denoted by $\hat{G}_X(x;\, d)$, evaluated at c_0 is then

$$\begin{aligned}\hat{G}_X(c_0;\, d) &= \int_0^{c_0} \hat{f}_X(x;\, d)\mathrm{d}x \\ &= \sum_{i=0}^{d} \xi_i \int_0^{c_0} x^i g(x)\mathrm{d}x \\ &= \sum_{i=0}^{d} \frac{\xi_i\, \beta^i\, (\Gamma(i/\gamma+\alpha) - \Gamma(i/\gamma+\alpha, (c_0/\beta)^\gamma))}{\Gamma(\alpha)}, \quad c_0 > 0, \end{aligned} \tag{2.6}$$

where $\Gamma(a,\theta) = \int_\theta^\infty t^{a-1}e^{-t}dt$ denotes the incomplete gamma function.

Once the probability density and cumulative distribution functions of X are approximated, the approximate survival and hazard functions of X, denoted by $\hat{S}_X(x;\, d)$ and $\hat{H}_X(x;\, d)$, respectively, evaluated at a positive real value c_0 as

$$\begin{aligned}\hat{S}_X(c_0; d) &= 1 - \hat{G}_X(c_0; d) \\ &= \frac{1}{\Gamma(\alpha)}\left[\Gamma(\alpha) - \sum_{i=0}^{d}\left\{\xi_i\beta^i\left(\Gamma\left(\frac{i}{\gamma}+\alpha\right) - \Gamma\left(\frac{i}{\gamma}+\alpha, \left(\frac{c_0}{\beta}\right)^\gamma\right)\right)\right\}\right]\end{aligned} \tag{2.7}$$

and

$$\begin{aligned}\hat{H}_X(c_0; d) &= \frac{\hat{f}_X(c_0; d)}{\hat{S}_X(c_0; d)} = \frac{\hat{f}_X(c_0; d)}{1 - \hat{G}_X(c_0; d)} \\ &= \frac{\gamma e^{-\left(\frac{c_0}{\beta}\right)^\gamma} \sum_{i=0}^{d} \xi_i c_0^{\alpha\gamma+i-1}\mathcal{I}_{(0,\infty)}(c_0)}{\beta^{\alpha\gamma}\left[\Gamma(\alpha) - \sum_{i=0}^{d}\left\{\xi_i\beta^i\left(\Gamma\left(\frac{i}{\gamma}+\alpha\right) - \Gamma\left(\frac{i}{\gamma}+\alpha, \left(\frac{c_0}{\beta}\right)^\gamma\right)\right)\right\}\right]}.\end{aligned} \tag{2.8}$$

We now show how the generalized gamma-polynomial approximation technique can be used in the cases of convolution and mixture of random variables and illustrate their computation and quality.

3. Convolutions

A convolution of two or more independent random variables can be expressed as a sum of those variables. Convolutions of random variables have been applied to a wide array of scientific fields. Although it is useful to model waiting times in different states via convolutions from other types of distributions such as the gamma or the Weibull distributions, there are limitations to use such models since the models may not be analytically tractable. In this section we briefly outline a convolution model and provide its moments.

The distribution and hazard functions of a Weibull random variable W with parameters (ω, ν), denoted by $f_W(t)$ and $h_W(t)$ respectively, are

$$f_W(t) = \nu\omega^\nu\, t^{\nu-1}\exp\left(-\omega^\nu \mathrm{t}^\nu\right)\mathcal{I}_{(0,\infty)}(\mathrm{t}) \tag{3.1}$$

and

$$h_W(t) = \nu\omega^\nu t^{\nu-1}\mathcal{I}_{(0,\infty)}(t), \tag{3.2}$$

where $\nu > 0, \omega > 0$. Since the moment generating function of the Weibull distribution is not available in closed form, we can not evaluate its moments from their moment generating function. The j^{th} moments of the Weibull random variable W is denoted by $\mu_W(j)$, that is,

$$\int_0^\infty x^j f_W(x)dx \equiv \mu_W(j)\,. \tag{3.3}$$

Fortunately, the moments of the Weibull distribution is simply expressed, that is, the h^{th} moment of the Weibull distribution is

$$\mu_W(h) = \omega^{-h}\Gamma\left(1+\frac{h}{\nu}\right), \quad h = 0, 1, \ldots. \tag{3.4}$$

Let C be a random variable of a convolution of ℓ number of the independent Weibull random variables denoted by W_i, that is $C = W_1 + W_2 + \cdots + W_\ell = \sum_{j=1}^{\ell} W_i$. It should be noted that although the exact distribution functions for convolutions of Weibull distributions are analytically intractable, their exact moments can be easily determined. The h^{th} moment of the convolution denoted by $\mu_C(h)$ can be obtained as follows;

$$\begin{aligned}
\mu_C(h) &= \mathbf{E}\left[\left(\sum_{i=1}^{\ell} W_i\right)^h\right] \\
&= \sum_{\left(\sum_{i=1}^{\ell} j_i = h\right)} \binom{h}{j_1, \ldots, j_\ell} \prod_{i=1}^{\ell} \mu_{W_i}(j_i) \\
&= \sum_{\left(\sum_{i=1}^{\ell} j_i = h\right)} \binom{h}{j_1, \ldots, j_\ell} \prod_{i=1}^{\ell} \omega_i^{-j_i}\Gamma\left(1+\frac{j_i}{\nu_i}\right),
\end{aligned} \tag{3.5}$$

where ω_i and ν_i are the parameters of W_i, $\mu_{W_i}(k)$ is the k^{th} exact moment of W_i and $\sum_{i=1}^{\ell} j_i = h$. However, in fact that density approximations via generalized gamma-polynomial approximation technique require the moments of random quantity of interest, generalized gamma-polynomial approximation might be more suitable than saddlepoint methods for approximating the distribution of convolution of the Weibull random variables. It is because once the moments of the convolution could be calculated, generalized gamma-polynomial approximation would be utilized to provide approximation for its distribution function without requiring the complex integration. We now show that the generalized gamma-polynomial approximation technique is an alternative to the saddlepoint method in the case of convolution of two Weibull distributions.

Numerical Example 1

We consider an example of the convolution of two Weibull random variables where the parameters of the Weibull random variables have values $\omega_1 = 2$, $\omega_2 = 3$ and $\nu_1 = 5$, $\nu_2 = 7$. The parameters of the generalized gamma baseline density function are obtained by matching the first three moments of the convolution of two Weibull random variables to the first three moments of the generalized gamma baseline density with the parameters as shown in Equation (2.3). The estimated parameters are $\alpha = 2.42251$, $\beta = 7.48116$ and $\gamma = 2.17197$, that is, the baseline density function is

$$g(w) = 0.0000434258\, w^{4.26161}\, e^{-0.0126405\times w^{2.17197}}\, \mathcal{I}_{(0,\infty)}(w)\,. \tag{3.6}$$

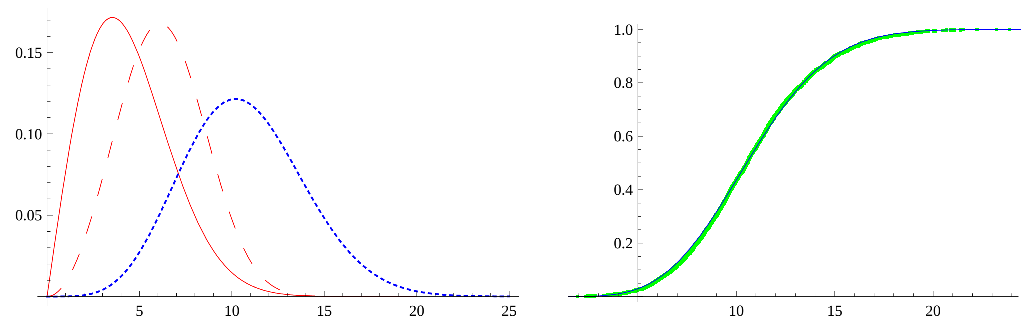

Figure 3.1. *Two Weibull(solid and dashed) & Generalized Gamma(dotted) PDFs(left panel); Simulated(dotted) & 4^{th} degree Generalized Gamma-polynomial Distribution Approximant(solid)(right panel)*

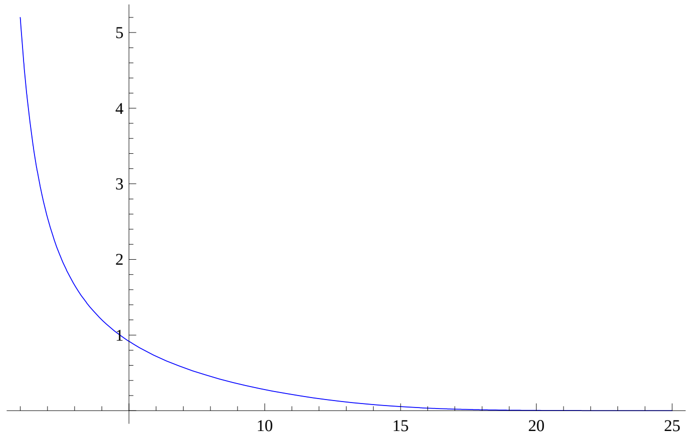

Figure 3.2. Approximated Hazard Function via 4^{th} degree Generalized Gamma-Polynomial Approximation

The exact density for two Weibull density functions and the estimated generalized gamma baseline density are shown in left panel of Figure 3.1. As determined from Equation (2.1), we could obtain the fourth-degree generalized gamma-polynomial density approximant to the density function of the convolution, which is

$$\hat{f}_X(w;4) = 0.0000434258\, w^{4.26161}\, e^{-0.0126405\, w^{2.17197}} \big(1.17392 - 0.0745696\, w + 0.0108415\, w^2 - 0.000643926\, w^3 + 0.0000133412\, w^4\big)\ . \tag{3.7}$$

The corresponding distribution approximant, as can be seen in the right panel of Figure 3.1, is in excellent agreement with the simulated distribution obtained from 3,000 replications. As can be seen in Figure 3.2, the approximated hazard function can also be easily obtained from both of the approximants for density and distribution functions of the convolution.

4. Finite Mixtures

Finite mixtures have often been used to model nonhomogeneous population related to survival analysis, for instances, in discussions of the incidence and curability of cancer, see Farewell (1977) and Langlands *et al.* (1979). In this section we briefly outline a mixed model in terms of their distributions and moments. A numerical example of approximating hazard function for a mixture of two gamma distributions will be examined by making use of the generalized gamma-polynomial approximation.

Let M be a finite mixture of ℓ number of the distributions of random variables of M_i, $i = 1, 2, \ldots, \ell$. Then M can be expressed as $M = \sum_{i=1}^{\ell} \mathcal{I}_{E_i} M_i$, where $\mathcal{I}_{E_i}$ denotes the indicator function, which is equal to 1 when an event E_i occurs and 0 otherwise. The probability density function of M, denoted by $f_M(t)$, can be expressed as follows;

$$\begin{aligned} f_M(t) &= \mathbf{Pr}(M = t) \\ &= \sum_{i=1}^{\ell} \mathbf{Pr}(M = t|E_i)\mathbf{Pr}(E_i) \\ &= \sum_{i=1}^{\ell} \phi_i f_{M_i}(t), \end{aligned} \tag{4.1}$$

where $f_{M_i}(t)$ is the probability density function of M_i and ϕ_i is the probability that E_i occurs, that is, $\phi_i = \mathbf{Pr}(E_i)$. ϕ_i can also be considered as the weight for the random variable of M_i, which satisfy $0 \leq \phi_i \leq 1$ and $\sum_{i=1}^{\ell} \phi_i = 1$. In the case that M_i are gamma variates with parameters $(\alpha_i,\ \beta_i)$, the density function of M is expressed as

$$f_M(t) = \sum_{i=1}^{\ell} \frac{\phi_i}{\Gamma(\alpha_i)\beta_i^{\alpha_i}} t^{\alpha_i - 1} e^{-\frac{t}{\beta_i}} \mathcal{I}_{(0,\infty)}(t). \tag{4.2}$$

And the j^{th} integer raw moment of M, denoted by $\mu_M(j)$, can be obtained as

$$\begin{aligned} \mu_M(j) &= \int_0^{\infty} x^j f_M(x) dx \\ &= \int_0^{\infty} x^j \sum_{i=1}^{\ell} \frac{\phi_i}{\Gamma(\alpha_i)\beta_i^{\alpha_i}} x^{\alpha_i - 1} e^{-\frac{x}{\beta_i}} dx \\ &= \sum_{i=1}^{\ell} \phi_i \left(\int_0^{\infty} x^j \frac{1}{\Gamma(\alpha_i)\ \beta_i^{\alpha_i}} x^{\alpha_i - 1} e^{-\frac{x}{\beta_i}} dx \right) \\ &= \sum_{i=1}^{\ell} \phi_i \beta_i^j \prod_{k=1}^{j} (\alpha_i + j - k)\,. \end{aligned} \tag{4.3}$$

We now show that the generalized gamma-polynomial approximation technique can be applied to approximate the hazard function of a mixture of two gamma distributions.

Numerical Example 2

We consider an example where a mixture of two gamma random variables with parameters $(5, 8)$ and $(2, 3)$, and weights $\phi_1 = 4/5$ and $\phi_2 = 1/5$. From matching the first three moments of the mixture of two gamma distributions to the first three moments of the generalized gamma baseline density with the parameters α, β and γ as shown in Equation (2.3), we obtain the parameters of the generalized gamma baseline density function, that is, $\alpha = 1.09856$, $\beta = 33.4493$ and $\gamma = 1.5$, that is,

$$g(t) = 0.004847\, t^{0.647837}\, e^{-0.00516916 \times t^{1.5}}\ \mathcal{I}_{(0,\infty)}(t)\,. \tag{4.4}$$

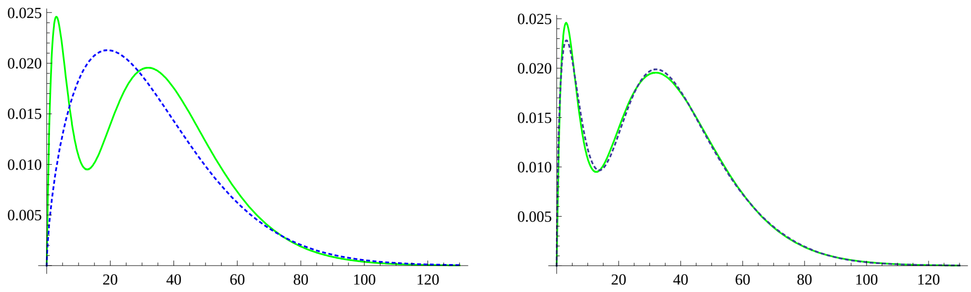

Figure 4.1. *Exact (solid) & Generalized Gamma(dotted) Density Functions(left panel); Exact Density Function(solid) & 10^{th} degree Generalized Gamma-polynomial Density Approximant(dotted)(right panel)*

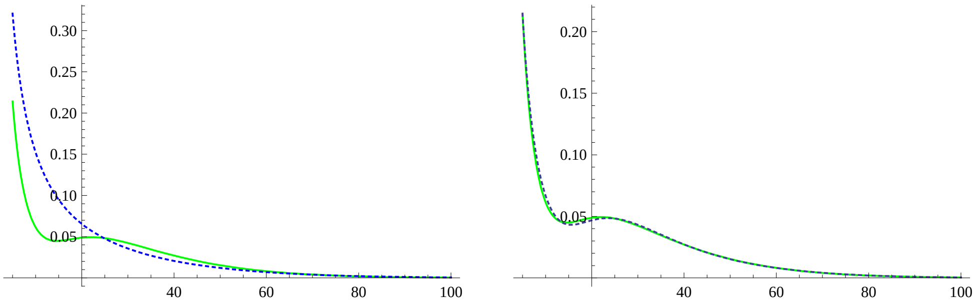

Figure 4.2. *Exact(solid) & Generalized Gamma(dotted) Hazard Functions(left panel); Exact Hazard Function(solid) & 10^{th} degree Generalized Gamma-polynomial Hazard Approximant(dotted)(right panel)*

The exact density for the mixture of two gamma distributions and the estimated generalized gamma baseline density are shown in left panel of Figure 4.1. As determined from Equation (2.1), we could obtain the tenth-degree generalized gamma-polynomial density approximant to the density function of two gamma mixture, which, as can be seen in the right panel of Figure 4.1, is in excellent agreement with the exact target functions.

The left panel of Figure 4.2 shows the approximated hazard function on the basis of the generalized gamma baseline density(dotted) with the exact hazard function(solid). The approximated hazard function, which was obtained from Equation (2.8) by making use of 10 moments of the target distribution, is also shown in Figure 4.2. As can be seen in both Figures, the approximations so obtained for density and hazard functions of the mixture of two gamma distributions are in excellent agreement with the target functions.

5. Concluding Remarks

Generalized gamma-polynomial approximation provides approximation accuracy in the entire range of the target survival and hazard functions of convolutions as well as their density and distribution functions. It should also be noted that the generalized gamma-polynomial approximation can also provide the explicit representation for the approximated density and hazard functions. The higher degree of the polynomial adjustment is recommended in order to obtain a suitable approximation in the cases that the exact density function to be approximated is more irregular or more precision is required. The symbolic computational package *Mathematica* was utilized for obtaining the survival and hazard function approximants for the numerical examples.

Acknowledgements

The author wishes to express his sincere thanks to the anonymous referees for valuable suggestions that improved the presentation of the material. This was supported by the Kyungwon University Research Fund.

References

Butler, R.W. (2000). Reliabilities for feedback systems and their saddlepoint approximation, *Statistical Science*, **15**, 279–298.

Butler, R. W. and Huzurbazar, A. V. (1997). Stochastic network models for survival analysis, *Journal of the American Statistical Association*, **92**, 246–257.

Butler, R. W. and Huzurbazar, A. V. (2000). Bayesian prediction of waiting times in stochastic models, *The Canadian Journal of Statistics*, **28**, 311–325.

Daniels, H. E. (1954). Saddlepoint approximations in statistics, *The Annals of Mathematical Statistics*, **25**, 631–650.

Daniels, H. E. (1982). The saddlepoint approximation for a general birth process, *Journal of Applied Probability*, **19**, 20–28.

Farewell, V. T. (1977). A model for a binary variable with time-censored observations, *Biometrika*, **64**, 43–46.

Ha, H. T. and Provost, S. B. (2007). A viable alternative to resorting to statistical tables, *Communication in Statistics: Simulation and Computation*, **36**, 1135–1151.

Huzurbazar, S. and Huzurbazar, A. V. (1999). Survival and hazard functions for progressive diseases using saddlepoint approximations, *Biometrics*, **55**, 198–203.

Keilson, J. (1978). Exponential spectra as a tool for the study of server-systems with several classes of customers, *Journal of Applied Probability*, **15**, 162–170.

Langlands, A. O., Pocock, S. J., Kerr, G. R. and Gore, S. M. (1979). Long-term survival of patients with breast cancer: a study of the curability of the disease, *British Medical Journal*, **2**, 1247–1251.

Pierce, D. A., Stewart, W. H. and Kopecky, K. J. (1979). Distribution-free regression analysis of grouped survival data, *Biometrics*, **35**, 785–793.

Reid, N. (1996). Likelihood and higher order approximations to tail areas, *The Canadian Journal of Statistics*, **24**, 141–165.