Paper 418-2012

Modeling heavy-tailed distributions in healthcare utilization by parametric and Bayesian methods

Joseph C. Gardiner

Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI

ABSTRACT

Distributions of healthcare utilization such as hospital length of stay and inpatient cost are generally right skewed. The extremes represent legitimate observations on patients who, because of the severity of their illness and need for medical intervention, have long in-stays and incur large costs. In this context we demonstrate the application of several parametric models for fitting heavy tailed data. Both maximum likelihood and Bayesian methods are used for estimation in certain Coxian phase-type models, mixtures of exponential distributions, and for comparison, the lognormal, log-logistic, Weibull, generalized gamma and generalized Pareto —including the standard Pareto and Burr distributions. We focus on the mean and percentiles of the response, and illustrate our methods with an empirical example on fitting models to hospital stays for acute myocardial infarction in the Nationwide Inpatient Sample of the Healthcare Utilization Project. A suite of SAS® procedures is used in all computations, specifically the procedures GENMOD, LIFEREG, MCMC, NLMIXED, FMM and SEVERITY.

1. INTRODUCTION

Heavy-tailed parametric distributions have many applications in analyses of economic, financial and physical systems. The stochastic model specifies a conditional distribution of positive random variable T given exogenous covariates **x.** In health care applications T represents utilization such as hospital length of stay (LOS) or hospital cost. Right-skewness, heteroscedasticity and heterogeneity are often observed in patient samples of LOS and cost. The challenge faced by the analyst is to posit a suitable model that captures the essential features of the entire distribution and to estimate covariate effects (e.g., patient and hospital characteristics) on summary statistics such as the mean and percentiles of the distribution. Commonly used parametric families include the lognormal, log-logistic, Weibull, generalized gamma, Pareto, and Burr. The latter two distributions are used extensively in models for claim size in insurance and in models for income (Klugman *et al*, 2004).

A distribution of T is said to be heavy-tailed if $E(e^{\lambda T}) = \infty$ for all $\lambda>0$, or in other words, the moment-generating function is not finite on the positive real line (Foss *et al*, 2011).This does not preclude the existence of finite moments. Equivalently, a distribution is heavy-tailed if its survival distribution S satisfies $e^{\lambda t} S(t) \to \infty$ as $t \to \infty$ for all $\lambda>0$. In section 2 we describe the Burr, Pareto, log-logistic and lognormal distributions which are all heavy-tailed. The Weibull distribution is heavy-tailed if and only if its shape parameter <1. The gamma distribution and in particular the exponential and mixtures of exponential distributions are not heavy-tailed. Sections 3 and 4 describe our analyses of an empirical example to LOS. Additional details are in Gardiner *et* al, 2012. We conclude with a discussion and summary in section 5.

SAS Global Forum Statistics and Data Analysis

2. MODELS

We outline here some of the parametric models that can be fitted with SAS procedures.

2.1 Accelerated Failure Time (AFT) *Model*

The accelerated failure time (AFT) model log transforms T to a location-scale family $\log T = \mu + \sigma\varepsilon$ where the parameters (μ, σ) are modeled by $\mu = \mathbf{x}'\beta$, $\log\sigma = \mathbf{x}'\delta$ and the random variable ε has a fully specified distribution. The Weibull distribution $S(t) = P[T > t] = \exp(-(t/\theta)^{\gamma})$, $\theta > 0, \gamma > 0$, has AFT form with $\mu = \log\theta$, $\sigma = \gamma^{-1}$ and ε has the extreme value distribution. The lognormal, log-logistic, generalized gamma are also in the AFT class, although in the generalized gamma the distribution of ε has an additional parameter. Extensive applications are found in the fields of biostatistics and reliability. Procedures LIFEREG and RELIABILITY are the engines for analysis of the AFT model. These procedures also permit incomplete observation of the response by allowing for left, right and interval censored data. The SEVERITY procedure provides additional functionality for left and right truncated data. It does not use a location-scale formulation for the response T but requires a scale parameter in its distribution in order to model covariates. Several pre-programmed distributions are available including the Pareto, standard gamma, and inverse Gaussian distributions that cannot be expressed in AFT form. Additionally it has the capability of fitting any continuous distribution by defining their distribution and density through functions and subroutines that call the FCMP procedure.

2.2 Mixed Proportional Hazards Model

An alternative to AFT models is the class of proportional hazards (PH) models used widely in biostatistics (Lawless, 2003). The cumulative hazard function $H(t \mid \mathbf{x})$ of T is expressed as $H(t \mid \mathbf{x}) = H_0(t)\exp(\mathbf{x}'\beta)$ where $H_0(t)$ is a cumulative baseline hazard function which is left unspecified in the semi-parametric Cox regression model. Specifying $H_0(t) = t^{\gamma}$ yields the Weibull with survival distribution $S(t \mid \mathbf{x}) = \exp\left(-\exp(\mathbf{x}'\beta)t^{\gamma}\right)$. The Weibull is the only continuous distribution with both the AFT and PH forms.

To incorporate unobserved heterogeneity a positive random effect (frailty) ν, independent of $\mathbf{x}$ is incorporated as $H(t \mid \mathbf{x}, \nu) = \nu H_0(t)\exp(\mathbf{x}'\beta)$. This is the mixed proportional hazards (MPH) model. The unconditional survival function is $S(t \mid \mathbf{x}) = E\left(\exp(-\nu H_0(t)\exp(\mathbf{x}'\beta))\right)$ where the expectation is with respect to the mixing distribution of ν. With specific assumptions on ν and $H_0(t)$ we obtain some of the aforementioned distributions. For example, choosing the gamma distribution for ν (shape α and scale $1/\alpha$) so that the mean is 1 leads to $S(t \mid \mathbf{x}) = \left(1 + \alpha^{-1}H_0(t)\exp(\mathbf{x}'\beta)\right)^{-\alpha}$. The Weibull hazard $H_0(t) = t^{\gamma}$ returns the Burr distribution. With exponential mixing (α=1) we get the log-logistic distribution. The standard Pareto distribution is obtained by gamma mixing of the exponential hazard $H_0(t) = t$. Proc PHREG is the workhorse for analysis of the PH model. With the addition of a lognormal random effect ν its reach is extended to the MPH whilst maintaining the semi-parametric feature of the Cox model.

SAS Global Forum Statistics and Data Analysis

2.3 Coxian Phase-type Distribution

A single parametric distribution coupled with a covariate model might be insufficient to capture variability and heterogeneity. Finite mixtures of distributions have many applications where the focus is on identifying and eliciting the characteristics of the heterogeneous subgroups. Continuous mixtures as described above for the MPH model are structurally different in that they involve continuous random effects that are meant to incorporate unobserved heterogeneity.

A phase-type distribution is the distribution of the time to absorption T in a continuous time finite state homogenous Markov process. (Fackrell, 2009) Suppose there are m transient states labeled 1,2,…,m and a single absorbing state with label 'm +1' . Coxian phase-type (CPH-m) distributions result when the process begins in state 1 and only forward transitions are allowed, $1\rightarrow 2$, $2\rightarrow 3$,…, $m-1\rightarrow m$ and exit to the absorbing state $m+1$ can occur from any transient state, $1\rightarrow m+1$, $2\rightarrow m+1$,…. $m\rightarrow m+1$. The explicit form of the survival distribution is $S(t)=\mathbf{e}\exp(\mathbf{Q}t)\mathbf{1}$ where $\mathbf{1}$ is the $m\times 1$ vector whose elements are all equal to 1, $\mathbf{e}=(1,0,\ldots,0)$ ($1\times m$ vector) and $\exp(\mathbf{Q}t)=\sum_{k=0}^{\infty}(\mathbf{Q}t)^k/k!$ is the matrix exponential (Golub and van Loan, 1996). The density function is $f(t)=\mathbf{e}\exp(\mathbf{Q}t)(-\mathbf{Q}\mathbf{1})$. The $m\times m$ matrix $\mathbf{Q}$ has the transition intensities in the transient states. It is upper triangular and single-banded. It turns out that T can be represented in distribution as $T=\sum_{k=1}^{m}Z_kW_k$ with $W_k=T_1+T_2+\cdots+T_k$, the sum of independent exponential variables, $T_k\sim EXP(\lambda_k)$ with hazard rate λ_k, and $(Z_1,Z_2,\ldots,Z_m)$ is independent of $\{T_k:1\le k\le m\}$ with multinomial distribution and probabilities $(\eta_1,\eta_2,\ldots,\eta_m)$, $\sum_{k=1}^{m}\eta_k=1$. Then $S(t)=\sum_{k=1}^{m}\eta_kS_k(t)$ is a finite mixture of survival distributions, where $S_k(t)$ is the survival distribution (Erlang) of W_k. The CPH-m has $2m-1$ parameters.

To incorporate covariate effects we impose an order restriction $\lambda_1\ge\lambda_2\ldots\ge\lambda_m$ and map $(\lambda_1,\lambda_2,\ldots,\lambda_m)$ to $(\lambda_1,\nu_2,\ldots,\nu_m)$ where $\lambda_k=\lambda_1\nu_2\ldots\nu_k$, $0<\nu_k\le 1$. If we regard λ_1 as an inverse-scale parameter covariate effects may be incorporated via the model $\lambda_1=\exp(-\mathbf{x}'\beta)$. We can use the programming features of SEVERITY to fit the CPH-m.

2.4 Finite Mixture of Distributions

The survival distribution of a finite mixture is $S(t\,|\,\mathbf{x})=\sum_{k=1}^{m}\eta_kS_k(t\,|\,\mathbf{x})$ where the S_k's are component survival distributions and η_k's are mixing probabilities with $\sum_{k=1}^{m}\eta_k=1$. Covariates $\mathbf{x}$ are entered through parameterization of the mean (or a function of the mean) in S_k and in the multinomial probabilities η_k, e.g., a generalized logit model. A homogenous finite mixture has the component distributions S_k in the same parametric family and the mixing probabilities do not depend on covariates. A finite mixture distribution of exponentials (FME-m) is $S(t\,|\,\mathbf{x})=\sum_{k=1}^{m}\eta_k\exp(-\lambda_k(\mathbf{x})t)$ with $\lambda_k(\mathbf{x})=\exp(-\mathbf{x}'\beta_k)$. The CPH-$m$ is a homogenous finite mixture of (generalized) Erlang distributions with the same β_k across components but different intercepts. The first component is an exponential distribution. We can fit the FME-m using the FMM procedure.

SAS Global Forum Statistics and Data Analysis

Table 1 summarizes the functional form of the survival function for some selected distributions. Absent are the generalized gamma which is a parent for the lognormal, Weibull, and standard gamma distributions, and the generalized Pareto which is a parent for the Pareto distribution. Details are found in the documentation for LIFEREG and RELIABILITY for the generalized gamma and in SEVERITY for the generalized Pareto. For regression models covariates enter through parameterization of the scale parameter θ. The formulae for the mean $E(T \mid \mathbf{x})$ and 100×p-th percentile $t_p(\mathbf{x})$ are functions of all the parameters of the distribution. The maximum likelihood method is used in estimation. Construction of a confidence interval for these summary statistics is carried out for the log-transformed quantity and then back-transformed.

Table 1: Some parametric distributions

Distribution Name	Survival Distribution	Mean and 100×p-th percentile
Weibull $\theta>0, \gamma>0$	$S(t)=\exp(-(t/\theta)^{\gamma})$, $\log\theta=\mathbf{x}'\beta, \sigma=\gamma^{-1}$.	$E(T \mid \mathbf{x})=\exp(\mathbf{x}'\beta)\Gamma(\sigma+1))$ $t_p(\mathbf{x})=\exp(\mathbf{x}'\beta)\left(-\log(1-p)\right)^{\sigma}$
Lognormal $\theta>0, \gamma>0$	$S(t \mid \mathbf{x})=\Phi\left(-\log(t/\theta)^{\gamma}\right)$, $\log\theta=\mathbf{x}'\beta, \sigma=\gamma^{-1}$.	$E(T \mid \mathbf{x})=\exp(\mathbf{x}'\beta+½\sigma^2)$ $t_p(\mathbf{x})=\exp\left\{\mathbf{x}'\beta+\sigma\Phi^{-1}(p)\right\}$
Burr, $\alpha>0$, $\theta>0, \gamma>0$	$S(t)=\left(1+(t/\theta)^{\gamma}\right)^{-\alpha}$, $\log\theta=\mathbf{x}'\beta, \sigma=\gamma^{-1}$.	$E(T \mid \mathbf{x})=\exp(\mathbf{x}'\beta)\Gamma(1+\sigma)\Gamma(\alpha-\sigma)/\Gamma(\alpha)$, $\sigma<\alpha$ $t_p(\mathbf{x})=\exp(\mathbf{x}'\beta)\left\{(1-p)^{-1/\alpha}-1\right\}^{\sigma}$
Pareto $\alpha>0$	$S(t)=\left(1+(t/\theta)\right)^{-\alpha}$, $\log\theta=\mathbf{x}'\beta$.	$E(T \mid \mathbf{x})=(\alpha-1)^{-1}\exp\left(\mathbf{x}'\beta\right)$, $\alpha>1$ $t_p(\mathbf{x})=\exp(\mathbf{x}'\beta)\left\{(1-p)^{-1/\alpha}-1\right\}$
Log-logistic $\theta>0, \gamma>0$	$S(t)=\left(1+(t/\theta)^{\gamma}\right)^{-1}$, $\log\theta=\mathbf{x}'\beta, \sigma=\gamma^{-1}$.	$E(T \mid \mathbf{x})=\exp(\mathbf{x}'\beta)\Gamma(1+\sigma)\Gamma(1-\sigma)$, $\sigma<1$ $t_p(\mathbf{x})=\exp(\mathbf{x}'\beta)\left\{p/(1-p)\right\}^{\sigma}$
†Gamma $\theta>0, \alpha>0$	$f(t)=\left(\theta\Gamma(\alpha)\right)^{-1}(t/\theta)^{\alpha-1}e^{-t/\theta}$, $\log\theta=\mathbf{x}'\beta$.	$E(T \mid \mathbf{x})=\alpha\exp(\mathbf{x}'\beta)$ $t_p(\mathbf{x})=\exp(\mathbf{x}'\beta)\Gamma^{-1}(p,\alpha)$
‡CPH-2, $\lambda_1>0$ $0<\eta_2,\nu_2<1$	$S(t)=(1-\eta_2)e^{-\lambda_1 t}+$ $\left(\frac{\eta_2}{1-\nu_2}\right)\left\{e^{-\nu_2\lambda_1 t}-\nu_2 e^{-\lambda_1 t}\right\}$ $\log\lambda_1=-\mathbf{x}'\beta$.	$E(T \mid \mathbf{x})=\exp(\mathbf{x}'\beta)\left(1+\eta_2\nu_2^{-1}\right)$ $t_p(\mathbf{x})$ is the solution to $S\left(t_p(\mathbf{x})\right)=1-p$

† $\Gamma^{-1}(p,\alpha)$ is the inverse gamma function with shape α; f is the density of the Gamma distribution.

‡ For $\nu_2=1$, $S(t)=e^{-\lambda_1 t}\left(1+\eta_2\lambda_1 t\right)$.

3. APPLICATION

We use a data set of N=11,749 hospital admissions for acute myocardial infarction drawn from the 2003 Nationwide Inpatient Sample of the Healthcare Utilization Project which samples annually healthcare utilization (LOS and hospital charges) in approximately 1,000 community-based hospitals in the US (HCUP Overview, 2009). The range of LOS is 1 to 142 days, mean 5.51 (SD=5.90). Covariates include patient age at admission (18 to 84 years), gender (37% female), and presenting comobidity as assessed by the Charlson Comorbidity Index (CCI). (Charlson *et al*, 1987) The CCI is a weighted sum of the presence of disease or medical conditions. It includes diabetes, with or without complications, renal disease, pulmonary disease, congestive heart failure, and peripheral vascular disease. The CCI is categorized into 4 subgroups: 1, 2, 3 and ≥4. These subgroups represent respectively 38%, 29%, 17%, and 16% of the sample. Patients are categorized by the most complex procedure that they underwent (percent of sample in parenthesis): CABG=coronary artery bypass surgery (12%), PTCA=percutaneous transluminal coronary angioplasty (40%), CATH=cardiac catheterization (19%), OTHER=other procedures performed (16%), or NONE=no procedure performed (13%). Additional characteristics of the sample include factors at the hospital-level: geographic region (Northeast, South, Midwest, and West), bed size (small, medium, large), and location/teaching status (urban teaching, urban non-teaching, rural). In this application we use LOS as outcome. Observations are complete: there are no censored values.

3.1 Maximum Likelihood Estimation

In table 1 the regression model for the scale parameter is $\theta(\mathbf{x}) = \theta_0 \exp(\mathbf{x}'\beta)$ or equivalently $\log\theta(\mathbf{x}) = \log\theta_0 + \mathbf{x}'\beta$ so that $\log\theta_0$ serves as intercept. For the CPH-2 distribution we use the parameterization

$$S(t) = (1-\eta_2)\exp(-\lambda_1 t) + \left(\frac{\eta_2}{1-\nu_2}\right)\{\exp(-\lambda_1\nu_2 t) - \nu_2\exp(-\lambda_1 t)\}$$
$$= \left(1-\frac{\eta_2}{1-\nu_2}\right)\exp(-\lambda_1 t) + \left(\frac{\eta_2}{1-\nu_2}\right)\exp(-\lambda_1\nu_2 t)$$

where $\lambda_1(\mathbf{x}) = 1/\theta(\mathbf{x})$ and ν_2, η_2 are additional parameters with $0<\nu_2<1$, $0<\eta_2<1$. The second line has the same form as a FME-2 except that we cannot guarantee that $\left(\frac{\eta_2}{1-\nu_2}\right)<1$.

Using the SEVERITY procedure table 2 reports maximum likelihood (ML) estimates $(\hat{\theta}_0, \hat{\beta})$ and estimates of additional non-scale parameters such as shape parameters for the CPH-2, Burr and log-logistic distributions The Burr distribution is pre-programmed. Accepting all defaults the syntax is:

```
proc severity data=losami outcdf=cdf;
LOSS LOS;
scalemodel age female type_cabg type_ptca type_cath type_other CCI2-CCI4
           region_NE region_MW region_SO LOC_RUR LOC_UNT SIZE_SML SIZE_MED;
dist burr;
run;
```

With the exception of age which is continuous, all other covariates in table 2 are dummy indicators. Together with the intercept θ_0 there are 17 parameters, excluding the shape parameters in lines 2 and 3. The log-logistic is a special case of the Burr distribution with $\alpha=1$. A Wald test or likelihood ratio test will reject the log-logistic model. The Pareto model restricts $\gamma=1$ in the Burr distribution. Fitting a Pareto distribution was problematic because of non-convergence for the shape parameter α . One reason might be noticed from the estimates of α, γ in the Burr distribution. A Wald test of $\gamma=1$ will be rejected. To compare the model CPH-2 with either the Burr or log-logistic models we use Vuong's test for strictly non-nested models (Vuong, 1989). It assumes that neither model is correct. Based on this test we might prefer the Burr or log-logistic models. The similarity of these two models is also seen in their fit statistics.

Fit statistics shown in table 2 are computed from the empirical distribution function (EDF) $F_n(t)$ and the fitted cumulative distribution function (CDF) $\hat{F}(t)$. Our KS-statistic does not apply the sample size $\sqrt{n}$ scaling factors used by SEVERITY. The CDF estimate plugs in the ML estimates for all model parameters and is the average $\hat{F}(t)=n^{-1}\sum_{i=1}^{n}F(t\,|\,\hat{\theta}(\mathbf{x}_i),\hat{\omega})$ where $\hat{\theta}(\mathbf{x}_i)=\hat{\theta}_0\exp(\mathbf{x}_i'\hat{\beta})$ and $\hat{\omega}$ denotes estimates of non-scale parameters ω. The OUTCDF= option creates a dataset of both the EDF and CDF estimates. PLOTS=(CDF PP) gives graphical output with a P-P plot.

Table 2: Maximum likelihood estimates for Coxian, Burr and Log-logistic distributions

Distribution	Coxian, CPH-2		Burr		Log-logistic	
Parameter†	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
θ_0	1.53486	0.10093	1.34376	0.06290	1.27024	0.05383
	ν_2 0.13305	0.04487	α 1.11027	0.04056	…	…
	η_2 0.00132	0.0006225	γ 2.63337	0.03720	γ 2.72608	0.02108
AGE	0.00878	0.000788	0.00863	0.000502	0.00858	0.000502
FEMALE	0.07852	0.01994	0.07738	0.01262	0.07721	0.01262
Procedure, CABG	1.16301	0.03720	1.20262	0.02324	1.21225	0.02301
PTCA	0.21516	0.03008	0.23427	0.01949	0.23912	0.01946
CATH	0.29992	0.03266	0.27796	0.02101	0.27941	0.02106
OTHER	0.48632	0.03602	0.34789	0.02448	0.33728	0.02432
Comorbidity, CCI=2	0.26551	0.02360	0.19680	0.01480	0.19181	0.01469
CCI=3	0.43590	0.02862	0.37578	0.01832	0.36993	0.01822
CCI≥4	0.62285	0.02975	0.58678	0.01895	0.58486	0.01894
Fit Statistics						
–2log L	60889		55598		55606	
KS	0.2410		0.0939		0.0950	
AD	850.12		1011.91		1022.16	
CVM	28.55		49.72		49.76	

†Covariate model includes 7 additional parameters for hospital region, location/teaching status and bed size.

CABG=coronary artery bypass surgery, PTCA=percutaneous transluminal coronary angioplasty, CATH=cardiac catheterization, OTHER=other procedures, NONE=no procedure (referent); CCI=Charlson Comorbidity Index; CCI=1 as referent. KS= Kolmogorov-Smirnov, AD=Anderson-Darling, CVM=Cramer-von Mises.

3.2 The CPH-2 and Log-logistic models

Since the log-logistic model is in the AFT class it can be estimated using the LIFEREG procedure. However, SEVERITY has the capability of fitting a continuous parametric distribution through functions and subroutines written in the FCMP procedure. We demonstrate this feature for the CPH-2 model. The syntax is in the Appendix. At a minimum we must define two functions for the probability density and cumulative distribution (or survival) of the CPH-2. Although defaults exist for initial values of the parameters $(\theta_0, \nu_2, \eta_2)$ and the regression coefficients β, initialization of $(\theta_0, \nu_2, \eta_2)$ was informed by the method of moments.

Any distribution *G* on the positive real line is said to be *well-represented by a phase-type distribution* if its first three moments can be matched with those of the phase-type distribution. Telek and Heindl (2002) supply necessary and sufficient conditions for *G* to be well-represented by a CPH-2. Using this approach (without covariates) we found suitable initial values for $(\theta_0, \nu_2, \eta_2)$ which we used in the INIT= option of the DIST statement. Lower and upper bounds of the parameters are supplied via the LOWERBOUNDS and UPPERBOUNDS functions. For output and display the functions DESCRIPTION and PARMCOUNT are useful, but not required.

3.3 The FME-2 model

A 2-component mixture of exponential distributions has survival function

$S(t \mid \mathbf{x}) = \pi_1 \exp(-\lambda_1(\mathbf{x})t) + (1-\pi_1)\exp(-\lambda_2(\mathbf{x})t)$ where $\lambda_k(\mathbf{x}) = \exp(-\mathbf{x}'\beta_k)$, k=1, 2 and $0 < \pi_1 < 1$.

By keeping covariate effects specific to each component we get a monstrous model with 35 parameters that we fitted using proc FMM (results not shown). Note that proc FMM models the mean $\mu_k(\mathbf{x})$ of the component exponential distribution as $\log \mu_k(\mathbf{x}) = \mathbf{x}'\beta_k$. An intercept is included.

The translation to the CPH-2 model restricts $\lambda_2(\mathbf{x}) = \nu_2 \lambda_1(\mathbf{x})$ where $0 < \nu_2 < 1$. This equates the covariate effects and results in the same CPH-2 model with $\pi_1 = 1 - \eta_2(1-\nu_2)^{-1}$ provided the bounds on π_1 are satisfied. The syntax is

```
proc fmm data=Losami gconv=0;
model LOS=age female type_: CCI2-CCI4 region_: LOC_: SIZE_:
            /k=2 dist=exponential link=log
      equate=effects(age female type_: CCI2-CCI4 region_: LOC_: SIZE_:);
restrict int 1, int -1< -1;
probmodel/cl;
run;
```

Table 3 summarizes the results of ML estimation: $\log \nu_2 = \log \mu_1(\mathbf{x}) - \log \mu_2(\mathbf{x}) = \text{intcpt1} - \text{intcpt2}$. Also logit Prob refers to $\log(\pi_1 / (1-\pi_1))$. Hence $\nu_2 = \exp(0.4284 - 2.4455) = 0.1330$, $\eta_2 = (1-\pi_1)(1-\nu_2)$ $= (1-.9985)(1-.1330) = .00131$. This results in the same point estimates. The small difference in standard errors is because SEVERITY makes by default an adjustment to the estimated covariance matrix: $\mathbf{V} = (n/d)\mathbf{H}^{-1}$ where $\mathbf{H}$ is the Hessian, n=#observations, d=n–#parameters. The adjustment can be turned off with the VARDEF=N option in the proc severity statement.

SAS Global Forum Statistics and Data Analysis

Table 3: Maximum likelihood estimates for 2-component exponential model with equated covariate effects in components

Parameter	Estimate	StdError
Intcpt 1	0.4284	0.06571
Intcpt 2	2.4455	0.3421
Logit Prob	6.4855	0.4994
AGE	0.008783	0.000788
FEMALE	0.07852	0.01992
Procedure, CABG	1.1630	0.03717
PTCA	0.2152	0.03005
CATH	0.2999	0.03264
OTHER	0.4863	0.03599
Comorbidity, CCI=2	0.2655	0.02358
CCI=3	0.4359	0.02860
CCI≥4	0.6229	0.02973

†Covariate model includes 7 additional parameters for hospital region, location/teaching status and bed size.

CABG=coronary artery bypass surgery, PTCA=percutaneous transluminal coronary angioplasty, CATH=cardiac catheterization, OTHER=other procedures, NONE=no procedure (referent); CCI=Charlson Comorbidity Index; CCI=1 as referent.

3.4 Estimation of the mean and percentiles

Given a covariate profile **x** the formulae shown in table 1 may be used to estimate the mean and percentiles after estimation of all model parameters, and subsequently 95% confidence intervals (CI). Although our methods depend on the asymptotic distribution of the ML estimates, it is more accurate to transform the corresponding CIs for $\log E(T \mid \mathbf{x})$ and $\log t_p(\mathbf{x})$. For the CPH-*m* and FME-*m* distributions we do not have an explicit expression for the percentiles $t_p(\mathbf{x})$. Instead a $100(1-\delta)\%$ confidence interval for $t_p(\mathbf{x})$ is obtained from the values of t that satisfy

$$-z_{1-\delta/2} \le \frac{g\left(\hat{S}(t \mid \mathbf{x})\right) - g(1-p)}{\sqrt{Var\left\{g\left(\hat{S}(t \mid \mathbf{x})\right)\right\}}} \le z_{1-\delta/2}$$

where $\hat{S}(t \mid \mathbf{x})$ is the ML estimator of the survival distribution and $z_{1-\delta/2}$ is the $100\times(1-½\delta)$ percentile of the standard normal distribution. The transformation g is either the arcsine-square-root, log, or log(–log), and the variance is approximated by $\left[g'\left(\hat{S}(t \mid \mathbf{x})\right)\right]^2 Var\left\{\hat{S}(t \mid \mathbf{x})\right\}$ where g' is the derivative of g. If g is the identity function then a transformation is not applied.

For the CPH-2 the mean LOS is $\mu(\mathbf{x}) = \theta(\mathbf{x})\left(1 + \frac{\eta_2}{\nu_2}\right)$. We use the NLMIXED procedure to estimate the parametric function $\log \mu(\mathbf{x}) = \log \theta_0 + \mathbf{x}'\beta + \log\left(1 + \exp(\log \eta_2 - \log \nu_2)\right)$ and obtain a 95% CI for $\mu(\mathbf{x})$ by back transformation. The NLMIXED procedure uses a COXIAN_PDF subroutine written in FCMP. Percentile estimates $\hat{t}_p$ are obtained from the survival estimate $\hat{S}(t \mid \mathbf{x})$ by solving

SAS Global Forum Statistics and Data Analysis

the equation $\hat{S}(\hat{t}_p \mid \hat{\theta}(\mathbf{x}), \hat{\nu}_2, \hat{\eta}_2) = 1 - p$. The SOLVE statement in proc MODEL can be used for this purpose. To obtain 95% CIs we could use the aforementioned approach with the complementary log(–log) transformation. Because $\hat{S}(t \mid \mathbf{x}), \sqrt{Var\,\hat{S}(t \mid \mathbf{x})}$ are computed only for values of t in the input data set in NLMIXED, it might be necessary to compute the above ratio at additional values of t in order to derive a more accurate CI.

4. BAYES METHODS

Recent enhancements to the several SAS procedures permit Bayesian analyses. For distributions in the AFT class LIFEREG could be used and GENMOD for models formulated as generalized linear models. Proc MCMC is dedicated to fitting Bayesian models. The likelihood may be constructed using the original untransformed data, or the log-transformed data. This affects the numerical value of the DIC statistic, the difference is a constant $(= \sum_{i=1}^{n} \log LOS_i)$. There is a slight difference in the sampling strategy used by MCMC and by LIFEREG and GENMOD. MCMC applies the Metropolis-Hastings algorithm to draw samples from the target distribution via adaptive rejection sampling that uses the normal distribution as the proposal distribution. The distributions considered in LIFEREG and GENMOD exploit the log-concavity of the posterior densities to construct a proposal distribution. There is an additional sampling step if log-concavity is not met.

For the log-logistic model in AFT form the covariate vector $\boldsymbol{\beta}$ includes an intercept. Priors were specified as $\boldsymbol{\beta} \sim N(0, c\,\mathbf{I}_{17}), c = 1E6$, scale $\sigma \sim Gamma(shape = 1E-3, iscale = 1E-3)$. Starting values were their MLEs in table 2. Proc LIFEREG was used to obtain a posterior sample for $\theta = (\boldsymbol{\beta}, \sigma)$ of size B=10000 by thinning 50000 iterations after a burn-in of 2000. The syntax is

```
proc lifereg data=losami;
model los=dage female type_cabg type_ptca type_cath type_other CCI2-CCI4
      region_NE region_MW region_SO LOC_RUR LOC_UNT SIZE_SML
      SIZE_MED/dist=llogistic;
bayes seed=81011 nmc=50000 nbi=2000 thin=5 outpost=post_sampleLL initialmle
      coeffprior=normal(var=1E6) scaleprior=gamma(shape=1E-3, iscale=1E-3);
run;
```

We performed a Bayes analysis of the CPH-2 model using MCMC. Our model has structural parameters $(\theta_0, \nu_2, \eta_2)$ and the regression coefficient vector $\boldsymbol{\beta}$. We chose priors $\boldsymbol{\beta} \sim N(0, c\,\mathbf{I}_{16}), c = 1E6$, $\nu_2 \sim uniform(0,1)$, and $\eta_2 \sim uniform(0,1)$. For θ_0 we experimented with the prior (i) lognormal, $\log \theta_0 \sim N(0, 1E6)$ and (ii) inverse Gamma, $1/\theta_0 \sim IGamma(shape = 0.30, scale = 3.33)$. The joint prior of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \theta_0, \nu_2, \eta_2)$ was taken as the product of the four priors. For all parameters starting values were their MLEs in table 2. In (i) convergence was unsatisfactory for some parameters even after thinning 50000 iterations by 5 and a burn in of 5000. In (ii) we increased the number of iterations to 150000, retaining 30000 posterior samples after thinning by 5. The burn in was 5000. Posterior summaries for the comparable parameters were approximately the same in both scenarios, but trace-autocorrelation-density plots were better in (ii). Results are summarized in table 4 and figure 1.

All statistics are based on random samples $\{\boldsymbol{\theta}^{(b)} : 1 \le b \le B\}$ drawn from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$ of $\boldsymbol{\theta}$ given the data $\mathbf{y}$. For example, the posterior mean is calculated as $\bar{\boldsymbol{\theta}} = B^{-1}\sum_{b=1}^{B}\boldsymbol{\theta}^{(b)}$ and an equal-tail 95% credible interval for a one-dimensional θ is the interval spanned by the 2.5-th and 97.5-th percentiles of the sample. The 95% highest posterior density (HPD) interval is derived as the smallest in width amongst 95% credible intervals.

Because non-informative priors are used for parameters, the posterior means are as expected, close to their ML counterparts (compare tables 3 and 4). The theoretical posterior means for the CPH-2 distribution is extremely difficult to derive in closed-form. We conjecture that improved accuracy could be achieved with good approximations of integrals (such as the Laplace approximation).

Table 4: Posterior Summaries for the CPH-2 distribution

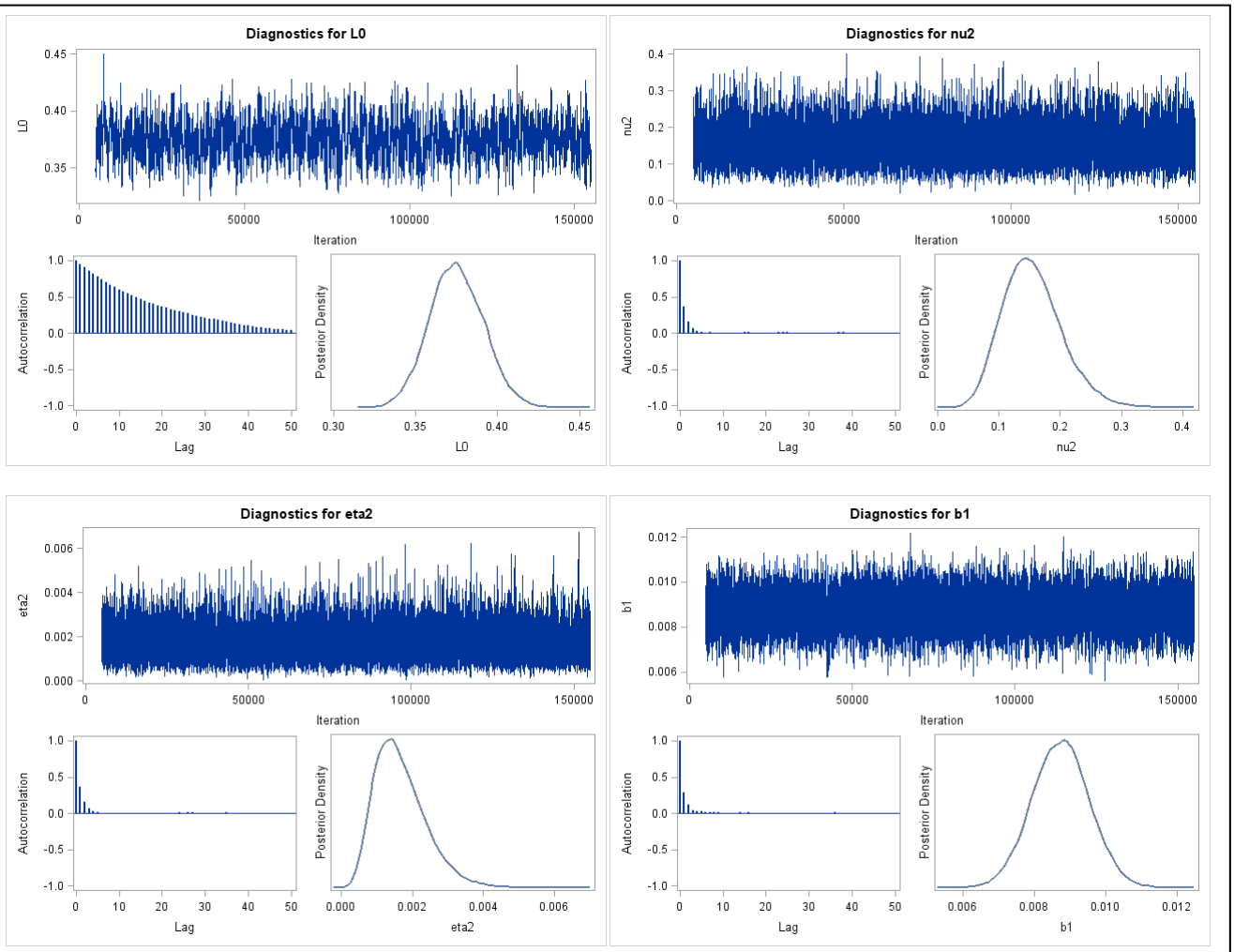
Parameter	Mean	STD DEV	Percentiles			Posterior Intervals			
			25%	50%	75%	Equal-Tail		HPD	
$1/\theta_0$	0.3751	0.0163	0.3639	0.3747	0.3862	0.3441	0.4077	0.3442	0.4078
ν_2	0.1549	0.0481	0.1208	0.1511	0.1843	0.0731	0.2613	0.0646	0.2491
$\eta_2 \times 10$	0.0167	0.00724	0.0115	0.0157	0.0209	0.00579	0.0335	0.00432	0.0309
AGE*	0.00874	0.000798	0.00820	0.00875	0.00928	0.00718	0.0103	0.00721	0.0103
FEMALE	0.0800	0.0198	0.0666	0.0797	0.0933	0.0413	0.1193	0.0411	0.1197
Procedure, CABG	1.1683	0.0373	1.1430	1.1687	1.1937	1.0956	1.2415	1.0944	1.2399
PTCA	0.2199	0.0301	0.1995	0.2201	0.2401	0.1605	0.2786	0.1631	0.2810
CATH	0.3041	0.0326	0.2819	0.3039	0.3260	0.2408	0.3687	0.2406	0.3684
OTHER	0.4895	0.0362	0.4649	0.4897	0.5138	0.4176	0.5602	0.4195	0.5619
Comorbidity, CCI=2	0.2674	0.0236	0.2516	0.2675	0.2834	0.2206	0.3132	0.2216	0.3142
CCI=3	0.4380	0.0285	0.4188	0.4382	0.4570	0.3817	0.4940	0.3827	0.4946
CCI≥4	0.6256	0.0298	0.6054	0.6258	0.6452	0.5667	0.6838	0.5678	0.6842

*Age centered at 64.56 years; DIC(unlogged response)=60928; Effective #parameters=18.85;
†Covariate model includes 7 additional parameters for hospital region, location/teaching status and bed size.
CABG=coronary artery bypass surgery, PTCA=percutaneous transluminal coronary angioplasty, CATH=cardiac catheterization, OTHER=other procedures, NONE=no procedure (referent);
CCI=Charlson Comorbidity Index; CCI=1 as referent.

The current experimental version (in SAS 9.3) of proc FMM has some Bayesian capabilities. However, in the Bayes context it does not support the RESTRICT statement and EQUATE= option that we used to fit the CPH-2 model in section 3.3. Proc FMM performed very well in a Bayes analysis of a simpler FME-2 model with fewer covariates (age and gender from table 4) but we encountered problems with more complex covariate models.

```
proc fmm data=losami gconv=0 seed=30812 ;
model LOS=dage female/k=2 dist=exponential link=log;
bayes initial=mle nbi=2000 nmc=10000 betapriorparms=(0,1E6);
run;
```

Figure 1: Trace, autocorrelation and posterior density plots of selected parameters of the CPH-2 distribution



L0=1/ θ_0 ; b1=(AGE−64.56).

5. DISCUSSION

We demonstrated the use of SAS procedures to fit parametric models for heavy-tailed distributions, in particular the Burr and log-logistic distributions. For comparison we fitted the Coxian 2-phase (CPH-2) distribution which is similar to a 2-component mixture of exponential distributions. Parameter estimation is via the maximum likelihood method. Where feasible, a Bayesian analysis can be carried out with in-built features in LIFEREG and FMM. To estimate the CPH-2 model we used the programming features in SEVERITY, and for Bayesian analysis the MCMC procedure. All models allow examination of the influence of covariates on the mean and percentiles of the distribution by positing a covariate model for the underlying scale parameter of the distribution.

SAS Global Forum Statistics and Data Analysis

Our empirical example was on hospital LOS for patients with a primary diagnosis of acute myocardial infarction (AMI). Previously we used this data set in a Bayesian analysis that determines dynamically the number of phases *m* in a series of CPH-*m* models. (Tang *et al*, 2012) In general a CPH-*m* distribution is a finite mixture of Erlang distributions. In our example the CPH-2 was rewritten as a proper mixture of two component exponentials. However, it cannot be guaranteed that the mixing coefficients are between 0 and 1, although they sum to 1. This allows for the possibility of fitting a series of CPH-*m* models using the feature in FMM to rank multiple models.

Although our application is focused on a completely observed non-negative continuous response *T*, the methods would extend to situations where *T* might be left or right censored, left or right truncated, or both. Estimation of parameters is based on constructing the likelihood function in four parts: for the subsample that is fully observed (not censored or truncated); for observed responses from a truncated sample; for censored data from a truncated sample; and for censored data from a sample that is not truncated. The SEVERITY procedure can handle this situation.

Our data set of AMI hospital stays has several covariates. Computational time for Bayes analyses performed in MCMC increases appreciably with the complexity of the parametric distribution and the number of covariates. Good starting values of parameters in ML estimation can aid considerably in this respect. With a completely observed response *T*, starting values for the covariate parameters are readily obtained from a linear regression model on the logged response. In all our models covariates enter through a single index function for the underlying scale θ in the distribution of *T*. In SEVERITY starting values for non-scale parameters are informed by moment matching, percentile matching or ML methods, as feasible. For the CPH-2 we used moment-matching following methods for well-representation of a general non-negative distribution by a member in CPH-*m*. (Bobbio *et al*, 2005; Osogami and Harchol-Balter, 2006) For some heavy-tailed distributions the data might not support moment-matching.

Parametric distributions are indispensible when interest lies in quantifying the influence of covariates on some summary features of the distribution. Currently the state of empirical research is heavily concentrated on the impact of covariates on the mean. With heavy tailed distributions covariate effects on the mean of the distribution might be less important than their influence on the tails of the distribution. Upper percentiles are more appropriate. We found proc NLMIXED very convenient in this regard, after having obtained the model's ML estimates through SEVERITY, and using these estimates are starting values in NLMIXED. In future enhancements to SAS procedures it is of value to develop methods that can identify characteristics exerting influence on the tail of heavy tailed distributions. Such methods will have application in several fields where heavy tailed distributions are typical. For example with budgetary constraints facing healthcare expenditure, in studies of healthcare resource use methods for identifying drivers of high utilization could have useful policy and interventional implications.

ACKNOWLEDGEMENTS

This research was supported in part by the Agency for Healthcare Research & Quality under Grant 1R01 HS14206.

REFERENCES

Bobbio A, Horvath A, Telek M. Matching three moments with minimal acyclic phase type distributions. *Stochastic Models.* 2005;21(2-3):303-326.

Charlson ME, Pompei P, Ales KL, Mackenzie CR. A new method of classifying prognostic co-morbidity in longitudinal-studies - Development and Validation. *Journal of Chronic Diseases.* 1987;40(5):373-383.

Fackrell M. Modelling healthcare systems with phase-type distributions. *Health Care Management Science.* 2009;12(1):11-26.

Foss S, Korshunov D, Zachary S. *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer-Verlag; 2011.

Gardiner JC, Tang X, Luo Z, Ramamoorthi RV. Fitting heavy-tailed distributions to healthcare utilization by parametric and Bayesian methods. 2012. *Manuscript in review.*

Golub G, Van Loan C. *Matrix computations. 3rd Edition.* The Johns Hopkins University Press; 1996.

HCUP Overview: Healthcare Cost and Utilization Project (HCUP). www.hcup-us.ahrq.gov/overview.jsp. Agency for Healthcare Research and Quality, 2009.

Klugman SA, Panjer HH, Wilmot GE. *Loss Models: From Data to Decisions.* Wiley & Sons; 2004.

Lawless JF. *Statistical Models and Methods for Lifetime Data, 2nd Edition.* Wiley & Sons; 2003.

Osogami T, Harchol-Balter M. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluation.* 2006;63(6):524-552.

Tang X, Luo Z, Gardiner JC. Modeling hospital length of stay by Coxian phase-type regression with heterogeneity. *Statistics in Medicine.* 2012. Early view 23 FEB 2012. DOI: 10.1002/sim.4490.

Telek M, Heindl A. Matching moments for acyclic discrete and continuous phase-type distributions of second order. *International Journal of Simulation Systems, Science & Technology.* 2002;3 (3-4):47-57.

Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica.* 1989;57(2):307-333.

CONTACT INFORMATION

Joseph C. Gardiner
Division of Biostatistics, Department of Epidemiology and Biostatistics
Michigan State University, East Lansing MI 48824
jgardiner@epi.msu.edu

SAS Global Forum Statistics and Data Analysis

APPENDIX

To fit the CPH-2 distribution the following FCMP program should be first invoked and saved in the OUTLIB= library. The primary components are the COXIAN_PDF and COXIAN_CDF functions.

```
proc fcmp library=sashelp.svrtdist outlib=work.sevexmpl.models;
function COXIAN_DESCRIPTION() $32;
        length model $32;
        model = "COXIAN Distribution";
        return(model);
    endsub;

    function COXIAN_PARMCOUNT();
        return(3);
    endsub;

/*--------------Coxian PDF----------------*/
      function COXIAN_PDF(t,theta, nu2, eta2);
 lam1=(1/theta); lam2=lam1*nu2;

  z1=exp(-lam1*t); z2=exp(-lam2*t);
  if lam1>lam2 then ft=(1-eta2)*lam1*z1+(eta2*lam1*lam2/(lam1-lam2))*(z2-z1);
else ft=.;

 return (ft);
endsub;

/*-------------------Coxian CDF --------------------*/
      function COXIAN_CDF(t,theta, nu2, eta2);
       lam1=(1/theta); lam2=lam1*nu2;

  z1=exp(-lam1*t); z2=exp(-lam2*t);
  if lam1>lam2 then St=(1-eta2)*z1+ (eta2/(lam1-lam2))*(lam1*z2-lam2*z1);
else St=.;

 return (1-St);
endsub;

/*-------------------Coxian LOWERBOUND --------------------*/
      subroutine COXIAN_LOWERBOUNDS(theta, nu2, eta2);
      outargs theta, nu2, eta2;
      theta=0; nu2=0; eta2=0;
      endsub;

/*-------------------Coxian UPPERBOUND --------------------*/
subroutine COXIAN_UPPERBOUNDS(theta, nu2, eta2);
      outargs theta, nu2, eta2;
      theta=.; nu2=1; eta2=1;
      endsub;

quit;
```

SAS Global Forum Statistics and Data Analysis

Maximum likelihood estimation of the CPH-2 can be performed using the following syntax which calls the SEVERITY procedure and the above utilities saved in the library CMPLIB=work.sevexmpl. The data set name is losami with the response LOS declared in the LOSS statement.

Some options in the SEVERITY statement are also shown. In particular the OUTCDF= creates a dataset of the estimated cumulative distribution function (CDF) and the empirical distribution function (EDF).

```
options cmplib=work.sevexmpl;

ods output statisticsoffit=fitstat;
proc severity data=losami print=(statistics estimates initialvalues)
                        covout outcdf=cdf outest=est;
LOSS LOS;
dist Coxian(init=(theta=5.3 nu2=.10 eta2=.001));
run;
```

To fit a model with covariates an additional SCALEMODEL statement is required. The model is $\theta(\mathbf{x}) = \theta_0 \exp(\mathbf{x}'\beta)$ where $\mathbf{x}$ are the covariates listed below.

```
scalemodel age female type_: CCI2-CCI4 region_: LOC_: SIZE_:;
```

Because of the equivalence in parameterization between the CPH-2 and FME-2, proc FMM can be used to fit a 2-component exponential mixture model with the same covariate effects for the mean parameters of the two components. Intercepts differ. The data set is Losami, the response is LOS and there are 7 covariates. The model has 19 parameters.

```
proc fmm data=Losami gconv=0;
model LOS=age female type_: CCI2-CCI4 region_: LOC_: SIZE_:
            /k=2 dist=exponential link=log
      equate=effects(age female type_: CCI2-CCI4 region_: LOC_: SIZE_:);
restrict int 1, int -1< -1;
probmodel/cl;
run;
```

To fit a general 2-component mixture to exponentials, with component-specific covariates, the EQUATE= option should be removed. Finally, note that the PROBMODEL statement serves to generate the mixing probability estimate and 95% confidence interval on the logit scale.