A New Package for the Birnbaum-Saunders Distribution

The bs package

by Víctor Leiva, Hugo Hernández, and Marco Riquelme

Background

Birnbaum and Saunders (1969a) derived an important lifetime distribution originating from a physical problem. The Birnbaum-Saunders distribution (BSD) describes the total time that passes until some type of cumulative damage produced by the development and growth of a dominant crack, surpasses a threshold, and causes a failure. Outside the field of the reliability, the BSD has been applied in quite a variety of fields. More details of applications and an extensive bibliographical revision can be found in Johnson et al. (1994) (on p. 651) and Vilca-Labra and Leiva (2006).

The BSD is defined in terms of the standard normal distribution through the random variable

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1} \right]^2,$$

where $Z \sim \mathrm{N}(0,1)$, $\alpha > 0$, and $\beta > 0$. This is denoted by $T \sim \mathrm{BS}(\alpha, \beta)$, where α is the shape parameter and β is the scale parameter. Thus, if $T \sim \mathrm{BS}(\alpha, \beta)$,

$$Z = \frac{1}{\alpha}\left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right] \sim \mathrm{N}(0,1).$$

Let $T \sim \mathrm{BS}(\alpha, \beta)$. Then, the probability density function (pdf) of T is given by

$$f_T(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right)^2} \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}, \quad (1)$$

with $t > 0$, $\alpha > 0$, and $\beta > 0$. We note that the pdf given in (1) can be written as

$$f_T(t) = \phi(a_t(\alpha,\beta)) \frac{d}{dt} a_t(\alpha, \beta),$$

where $a_t(\alpha,\beta) = \alpha^{-1}(\sqrt{t/\beta} - \sqrt{\beta/t})$ and $\phi(\cdot)$ is the pdf of $Z \sim \mathrm{N}(0,1)$. The cumulative distribution function (cdf) of T is given by

$$F_T(t) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right],$$

where $\Phi(\cdot)$ is the cdf of $Z \sim \mathrm{N}(0,1)$. Then, the quantile function (qf) of T, $t(p) = F_T^{-1}(p)$, is given by

$$t(p) = \beta\left[\frac{\alpha z_p}{2} + \sqrt{\left(\frac{\alpha z_p}{2}\right)^2 + 1}\right]^2, 0 < p < 1,$$

where z_p is the p-th percentile of $Z \sim \mathrm{N}(0,1)$. Thus, if $p = 0.5$, then $t(0.5) = \beta$, and so β is the median.

Figure 1 shows the behavior of the pdf of the BSD for some values of α. Note that as α decreases the shape of the pdf is approximately symmetrical.

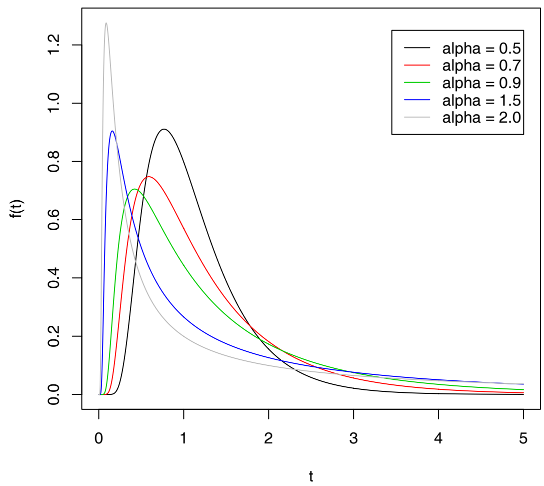

Figure 1: pdf of $T \sim \mathrm{BS}(\alpha, \beta = 1.0)$ for indicated α.

Some properties of $T \sim \mathrm{BS}(\alpha, \beta)$ are the following:
(P1) $cT \sim \mathrm{BS}(\alpha, c\beta)$, with $c > 0$.
(P2) $T^{-1} \sim \mathrm{BS}(\alpha, \beta^{-1})$.

The n-th moment of the BSD is given by $\mathbb{E}(T^n) =$

$$\beta \sum_{j=0}^{n} \binom{2n}{2j} \sum_{i=0}^{j} \binom{j}{i} \frac{(2(n-j+i))!}{2^{n-j+i}(n-j+i)!} \left(\frac{\alpha}{2}\right)^{2(n-j+i)}.$$

Thus, the mean, the variance, and the variation, skewness, and kurtosis coefficients are, respectively,

(i) $\mu = \frac{\beta}{2}(\alpha^2 + 2)$,

(ii) $\sigma^2 = \frac{\beta^2}{4}(5\alpha^4 + 4\alpha^2)$,

(iii) $\gamma = \frac{\sqrt{5\alpha^4 + 4\alpha^2}}{\alpha^2 + 2}$,

(iv) $\delta = \frac{44\alpha^3 + 24\alpha}{\left(\sqrt{5\alpha^2+4}\right)^3}$, and

(v) $\kappa = 3 + \frac{558\alpha^4 + 240\alpha^2}{(5\alpha^2+4)^2}$.

We note if $\alpha \to 0$, then $\delta \to 0$ and $\kappa \to 3$. The coefficients γ, δ, and κ are invariant under scale, that is, these indicators are functionally independent of the scale parameter, β.

 ISSN 1609-3631

A useful function in lifetime analysis is the failure rate (fr), defined by $h_T(t) = \frac{f_T(t)}{1-F_T(t)}$, where $f_T(\cdot)$ and $F_T(\cdot)$ are a pdf and its cdf, respectively; see, e.g., Johnson et al. (1994)(on p. 640). The BSD does not have an monotonically increasing failure rate. This is initially increasing until its critical point and then it decreases until becomes stabilized in a positive constant (not in zero), as happens with the failure rate of other distributions like the lognormal model. Specifically, if $t \to \infty$, then $h_T(t) \to (2\alpha^2\beta)^{-1}$. Other indicators of aging are the failure rate average (fra), the reliability function (rf), and the conditional reliability function (crf).

Figure 2 shows the behavior of the fr of the BSD for some values of α. Note that as α decreases the shape of the fr is approximately increasing.

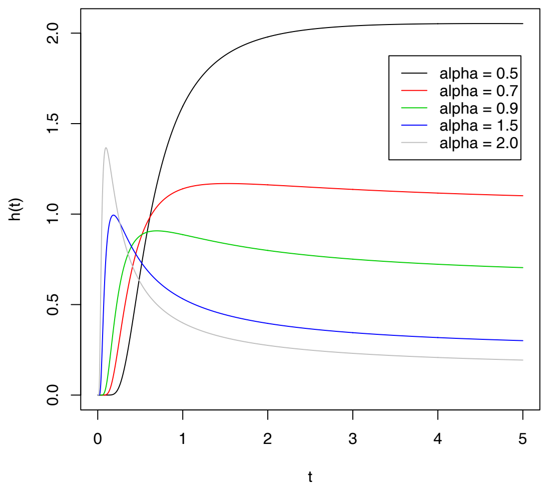

Figure 2: fr of $T \sim \text{BS}(\alpha, \beta = 1.0)$ for indicated α.

This article is separated into three parts. In the second part, the main functions of the **bs** package and some illustrative examples are presented. Finally, some concluding remarks are made in the third part.

The bs functions

The **bs** package contains basic probabilistic functions, reliability indicators, and random number generators from the BSD. In order to deal with the computation of the pdf, the cdf, and the qf, the commands `dbs()`, `pbs()`, and `qbs()`, respectively, are used. The following instructions illustrate these commands.

```
> dbs(3,alpha=0.5,beta=1.0,log=FALSE)
[1] 0.02133878

> pbs(1,alpha=0.5,beta=1.0,log=FALSE)
[1] 0.5

> qbs(0.5,alpha=0.5,beta=2.5,log=FALSE)
[1] 2.5
```

For reliability analysis, in order to calculate the fr, the fra, the rf, and the crf, we have implemented the functions `frbs()`, `frabs()`, `rfbs()`, and `crfbs()`, respectively. For obtaining random numbers, we have developed the following three generators based on: (G1) the relationship between the standard normal distribution (SND) and the BSD (see (Chang and Tang, 1994a)); (G2) the relationship between the sinh-normal distribution (SHND) and the BSD (see (Rieck, 2003)); (G3) the relationship between the inverse Gaussian distribution (IGD) and the BSD (Johnson et al. (1994), p. 658). For these three generators, the functions `rbs1()`, `rbs2()`, and `rbs3()` are used, respectively. Also, we have developed a common function, named `rbs()`, for obtaining random numbers from the suitable generator depending on the desired setting. The function `rbs()` selects the most appropriate method automatically. For details about effectiveness and efficiency of these three generators, and in order to select the most suitable one see Leiva, Sanhueza et al. (2006). The following instructions illustrate these commands:

```
> rbs3(n=6,alpha=0.5,beta=1.0)
[1] 0.7372910 0.4480005 1.8632176
[4] 0.9728011 1.2675537 0.2252379

> rbs(n=6,alpha=0.5,beta=1.0)
[1] 0.5905414 1.1378133 1.1664306
[4] 1.3187935 1.2609212 1.8212990
```

Another group of functions related to estimation, graphical analysis, and goodness-of-fit for the BSD are also available.

In order to estimate the shape (α) and scale (β) parameters from the BSD, we have implemented the following three methods based on: (E1) the likelihood method (MLE) and the mean-mean estimator (see (Birnbaum and Saunders, 1969b)); (E2) a graphical method (GME), which permits estimation of α and β by the least square method (see (Chang and Tang, 1994b)); (E3) the modified moment method (MME) (see (Ng et al., 2003)). For E1, E2, and E3, the functions `est1bs()`, `est2bs()`, and `est3bs()` are used. Next, two examples related to the use of these commands are presented. The first example is based on simulated data and the second example is from a real data set.

Example 1. In order to carry out simulation studies, we develop the functions `simul.bs.gme()`, `simul.bs.mle()`, and `simul.bs.mme()`. These functions generate random samples, estimate parameters, and establish goodness-of-fit. The samples of size n, one for each method (G1, G2, or G3), are generated by using `rbs1()`, `rbs2()`, and `rbs3()`, respectively. The estimations, one for each method (E1, E2, or E3), are obtained by using `est1bs()`, `est2bs()`, and `est3bs()`, respectively. The goodness-of-fit method is based on the statistic of Kolmogorov-Smirnov (KS), which is available by means of the function `ksbs()`. The generated observations by means of G1, G2, and G3 are saved as slots of the R class `simulBsClass`, which are named `sample1`, `sample2`, and `sample3`, respectively. Furthermore, the results of the simulation study are saved in a fourth slot of this class, named `results`. Thus, for instance, the instruction

```
> simul.bs.mle(100,0.5,1.0)
```

simulates three samples of size $n = 100$ from a population $T \sim \mathrm{BS}(\alpha = 0.5, \beta = 1.0)$, one for each method (G1, G2, or G3), computes the MLES's for α and β, and establish goodness-of-fit for each sample. The results can be saved in the variable `resultsOfsimul.bs.mle` by using the instruction

```
> resultsOfsimul.bs.mle<-
                   simul.bs.mle(100,0.5,1.0)
```

Hence,

```
> resultsOfsimul.bs.mle@results
```

retrieves the slot `results`, which gives the following summary of the simulation study:

```
        MLE(alpha) MLE(beta)  KS      p-value
Sample1  0.66456    1.06143   0.06701 0.76032
Sample2  0.43454    0.93341   0.09310 0.35135
Sample3  0.60549    0.86124   0.08475 0.46917
```

In addition,

```
> resultsOfsimul.bs.mle@sample1
```

retrieves the slot `sample1`, which only shows the generated sample associated with G1 and does not show the estimates neither the goodness-of-fit.

Next, we give an example of real data in order to illustrate the functions `est1bs()`, `est2bs()`, and `est3bs()`.

Example 2. Birnbaum and Saunders (1969b) reported a data set (see Table 1) corresponding to fatigue life (T) measures in cycles ($\times 10^{-3}$) of $n = 101$ aluminum coupons (specimens) of type 6061-T6. These specimens were cut in a parallel angle to the rotation way and oscillating to 18 cycles per seconds. The coupons were exposed to a pressure with maximum stress of 31.000 psi (pounds per square inch). All specimens were tested until failure.

The **bs** package brings these loaded data, so that if we input the command `psi31`, the data will be ready to use them. Thus,

```
> data <- psi31
```

saves the data set in the variable `data`. The instruction

```
> estimates <- est1bs(data)
```

computes the MLE's for α and β and uses the mean-mean-estimate (see (Birnbaum and Saunders, 1969b)) as initial estimator for β. Then, the following result is obtained:

```
> estimates
$beta.starting.value [1] 131.8193

$alpha [1] 0.1703847

$beta [1] 131.8188

$converge [1] "TRUE"

$iteration [1] 2
```

The estimations of α and β can be saved in the variables `alpha` and `beta`, that is,

```
> alpha <- estimate$alpha
> beta <- estimate$beta
```

Also, by using the invariance property of the MLE's, we obtain estimations for the mean, the variance, and the variation, skewness, and kurtosis coefficients. The function `indicatorsbs()` computes the MLE's for α, β, μ, σ^2, γ, δ, and κ. Thus, the instruction

```
> indicatorsbs(data)
```

gives the following results:

```
The MLE's are:
Alpha = 0.1703847
Beta  = 131.8188
Mean = 143.0487
Variance = 522.753
Variation coefficient = 0.170967
Skewness coefficient = 0.5103301
Kurtosis coefficient = 3.43287
```

Another five data sets, called `psi21`, `psi26`, `bearings`, `timerepair`, and `biaxial`, used frequently in the literature of this topic have also been incorporated in the **bs** package.

Next, the implemented graphical functions will be described. Firstly, the command `graphdbs()` allows visualization of five different pdf's from the BSD, but the legend of this must be added. Figure 1 was plotted with this command by means of the instructions

```
> graphdbs(0.5,0.7,0.9,1.5,2.0,1.0,1.0,1.0,
  1.0,1.0)
```

and

```
> legend(3.45,2,c("alpha = 0.5","alpha = 0.7",
  "alpha = 0.9","alpha = 1.5","alpha = 2.0"),
  lty=c(1,1,1,1,1), col=c(1,2,3,4,8))
```

Table 1: Lifetimes of aluminum specimens exposed to a maximum stress of 31.000 psi.

70	90	96	97	99	100	103	104
104	105	107	108	108	108	109	109
112	112	113	114	114	114	116	119
120	120	120	121	121	123	124	124
124	124	124	128	128	129	129	130
130	130	131	131	131	131	131	132
132	132	133	134	134	134	134	134
136	136	137	138	138	138	139	139
141	141	142	142	142	142	142	142
144	144	145	146	148	148	149	151
151	152	155	156	157	157	157	157
158	159	162	163	163	164	166	166
168	170	174	196	212			

Figure 2 was plotted with the function `grapfrbs()`, which is similar to `graphdbs()`, but uses the visualization of five different fr from the BSD. Secondly, the command `qqbs()` allows drawing a quantile-quantile (Q-Q) plot to check graphically if a set of observations follows a particular BSD. This command also incorporates a straight line Q-Q that sketches a line passing through the first and the third quartile of the theoretical BSD. The instruction

```
> qqbs(data,alpha,beta,line=TRUE)
```

gives a Q-Q plot for `psi31`. Figure 3 shows this graph.

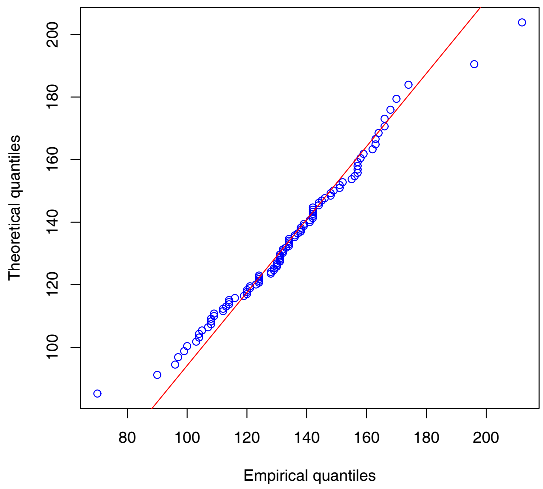

Figure 3: Q-Q plot for the BSD for psi31.

Thirdly, the command `gmbs()` allows simultaneously estimating α and β by GME and drawing the Q-Q type plot by Chang and Tang (1994b). The estimations are summarized in Table 2.

Figure 4 shows a graph produced by using the instruction `gmbs(data)`.

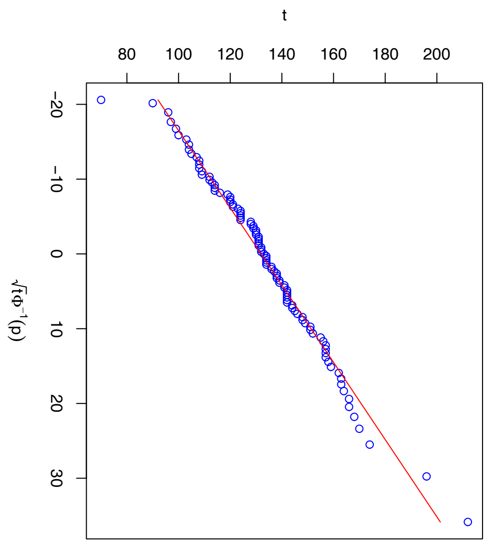

Figure 4: fitted plot for the BSD by using graphical method for psi31.

We also have implemented commands for computing methods of goodness-of-fit and criteria of model selection from the BSD. These functions are `ksbs()`, `sicbs()`, `aicbs()`, and `hqbs()`, which calculate the KS test, the Schwartz information criterium (SIC), the Akaike information criterium (AIC), and the Hannan-Queen information criterium (HQC). The command `ksbs()` gives also a comparative graph of the cdf and the empirical distribution function. Next, the results for `psi31` are presented in Table 3.

Finally, by using methods of goodness-of-fit based on moments, we have implemented the β_1-β_2 and δ_2-δ_3 charts; see Barros (2006). These charts-of-fit are particularly useful when various data sets are collected. For example, in environmental sciences we frequently found random variables measured hourly at different sampling points. Then, for every year we have a monthly data set available at each sampling point. Thus, it is possible to compute esti-

Table 2: Estimates of α and β by using `gmbs(data)`, for the data in Table 1.

$\widehat{\alpha}$	$\widehat{\beta}$	R^2
0.1686	131.9224	0.9775

Table 3: Table 3: SIC, AIC, HQC, and KS test for psi31.

SIC	AIC	HQC	KS (p-value)
4.573125	4.547233	4.557715	0.085 ($p = 0.459$)

mations for β_1 (the square of skewness coefficient), β_2 (the kurtosis coefficient), δ_2 (the square of variation coefficient), and δ_3 (the skewness coefficient), for every month at each sampling point. In this way, the pairs $(\widehat{\beta}_1, \widehat{\beta}_2)$ and $(\widehat{\delta}_2, \widehat{\delta}_3)$ are plotted inside the theoretical β_1-β_2 and δ_2-δ_3 charts, respectively. The functions `fitbetabs()` and `fitdeltabs()` have been developed to add these charts to the scatter plot of the pairs $(\widehat{\beta}_1, \widehat{\beta}_2)$ and $(\widehat{\delta}_2, \widehat{\delta}_3)$, respectively. In order to illustrate this methodology, we have simulated twelve (12) samples by using the command `rbs(n=30,alpha,beta=1.0)`, with $\alpha = 0.2(0.2)2.4$, which can represent, for example, daily data for every month during one calendar year. These result have been saved in the matrix `samplei`, with $i = 1, \ldots, 12$. Next, the estimations of δ_2, δ_3, β_1, and β_2 are obtained and saved in the vectors `x` and `y`. Finally, the data sets have been fitted to the charts. Thus, by using the instructions

```
> plot(x,y,xlab=expression(beta*1),
  ylab=expression(beta*2),col=4)
> fitbetabs(0.2,2)
> fitbetabs(2.4,3)
```

and

```
> plot(x,y,xlab=expression(delta*2),
  ylab=expression(delta*3),col=4)
> fitdeltabs(0.2,2)
> fitdeltabs(2.4,3)
```

we have obtained the graphs in Figures 5 and 6:
BS$(\alpha = 0.2, \beta = 1.0)$ in red;
BS$(\alpha = 2.4, \beta = 1.0)$ in green;
and pairs (x, y) in blue]:

Concluding remarks

In order to analyze data coming from the BSD, we have developed the new **bs** package, which implements several useful commands. The created functions are related to probability and reliability indicators, estimation, goodness-of-fit, and simulation. In addition, some commands related to graphical tools were also developed. We think that in the future, the **bs** package can be improved by incorporating other functions related to the estimation for censored data, regression and diagnostic methods, as well as generalizations of the BSD. The theoretical aspects of this last part have been published by Galea et al. (2004), Díaz-García and Leiva (2005), Vilca-Labra and Leiva (2006), Leiva, Barros et al. (2006).

Acknowledgments

This study was carried out with the support of research projects FONDECYT 1050862, FANDES C-13955(10), DIPUV 42-2004, and DIPUCM 200601, Chile. This article was partially written during the time that Dr. Víctor Leiva was a Visiting Scholar in the University of North Carolina (UNC) at Chapel Hill, USA. The authors are thankful to the Associate Editor and to the referee for his truly helpful suggestions which led to a much improved presentation of this article.

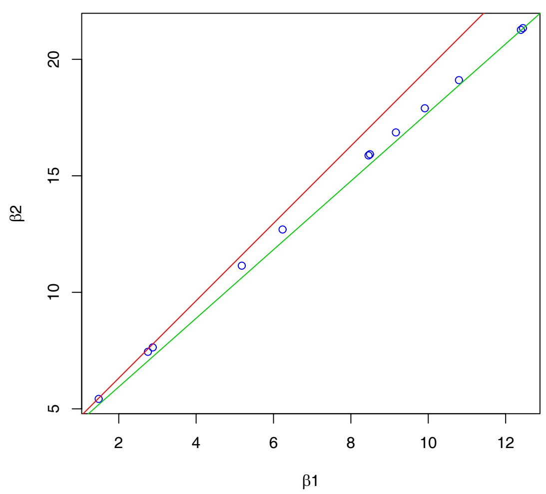

Figure 5: chart β_1-β_2 for twelve simulated samples.

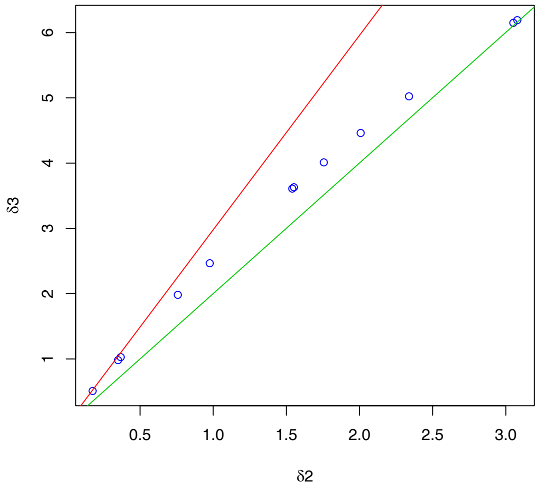

Figure 6: chart δ_2-δ_3 for twelve simulated samples.

Bibliography

M. Barros, V. Leiva, and G.A. Paula. Moments of the generalized Birnbaum–Saunders distribution: specification and application. *Submitted to Publication*, 2006.

Z.W. Birnbaum and S.C. Saunders. A new family of life distributions. *Journal of Applied Probability*, 6(2): 637-652, 1969a.

Z.W. Birnbaum and S.C. Saunders. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6(2):328-347, 1969b.

D.S. Chang and L.C. Tang. Random number generator for the Birnbaum-Saunders distribution. *Computational and Industrial Engineering*, 27(1-4):345-348, 1994a.

D.S. Chang and L.C. Tang. Graphical analysis for Birnbaum-Saunders distribution. *Microelectronics and Reliability*, 34(1):17-22, 1994b.

J.A. Díaz-García and V. Leiva. A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128(2):445-457, 2005.

M. Galea, V. Leiva, and G.A. Paula. Influence diagnostics in log-Birnbaum-Saunders regression models. *Journal of Applied Statistics*, 31:9, 1049-1064, 2004.

N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions: Volume 2*. Wiley, New York, 1994.

V. Leiva, M. Barros, G.A. Paula, and M. Galea. Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. Accepted in *Computational Statistics and Data Analysis*, 2006.

V. Leiva, A. Sanhueza, P.K. Sen, and G.A. Paula. Random number generators for the generalized Birnbaum-Saunders distribution. *Submitted to Publication*, 2006.

H.K.T. Ng, D. Kundub, and N. Balakrishnan. Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis*, 43(2):283-298, 2003.

J.R. Rieck. A comparison of two random number generators for the Birnbaum-Saunders distribution. *Communications in Statistics - Theory and Methods*, 32(5):929-934, 2003.

F. Vilca-Labra and V. Leiva. A new fatigue life model based on the family of skew-elliptical distributions. *Communications in Statistics: Theory and Methods*, 35(2):1-16, 2006.

Víctor Leiva
Departamento de Estadística, CIMFAV
Universidad de Valparaíso, Chile
victor.leiva@uv.cl; victor.leiva@yahoo.com

Hugo Hernández
Departamento de Estadística
Universidad de Valparaíso, Chile
hhp_uv@yahoo.com

Marco Riquelme
Instituto de Ciencias Básicas
Universidad Católica del Maule, Chile
mriquelm@ucm.cl