Inflated mixture models: Applications to multimodality in loss given default.

Mauro Ribeiro de O. Júnior
@MauroRdeOJr

Caixa Econômica Federal (Brazilian bank)

Department of Statistics
Federal University of Sao Carlos, Brazil

Thursday 27 August 2015

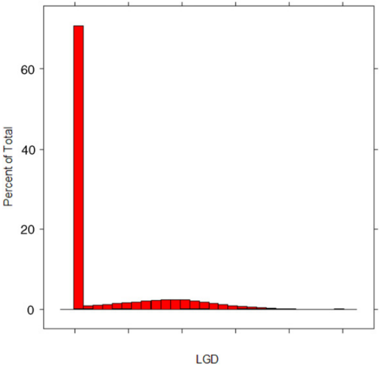
Holyrood Park

It may give us some idea about what is modality...

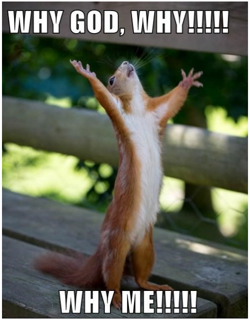
Agenda

Illustrative examples of LGD distributions I

- Tong, Edward NC, Christophe Mues, and Lyn Thomas. "A zero-adjusted gamma model for mortgage loan loss given default."International Journal of Forecasting 29.4 (2013): 548-562.



Why zero-adjusted?



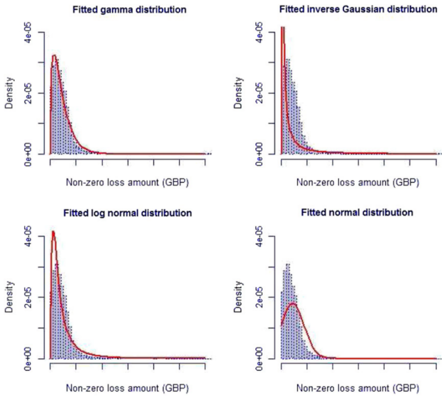
- Because the support of the gamma distribution does not include the zero point, i.e., the support is $(0, +\infty)$.

How zero-adjusted?

$$f_{i_{gamma}}(y;\vartheta) = \begin{cases} \delta_0, & \text{if} \quad y = 0 \\ (1-\delta_0) f_{gamma}, & \text{if} \quad 0 < y < +\infty, \end{cases}$$

where f_{gamma} follows a gamma distributions.

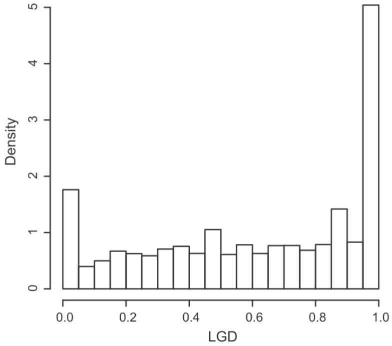
We've got more?



- Inverse Gaussian distribution, with support $(0, +\infty)$.
- Log normal distribution, with support $(0, +\infty)$.
- (truncated) Normal Gaussian distribution, with support $[0, +\infty)$.

Illustrative examples of LGD distributions II

- Calabrese, Raffaella. "Downturn loss given default: Mixture distribution estimation."European Journal of Operational Research 237.1 (2014): 271-277.



Agenda

Who is beta? and Mixture? and Inflated Mixture?

The beta distribution

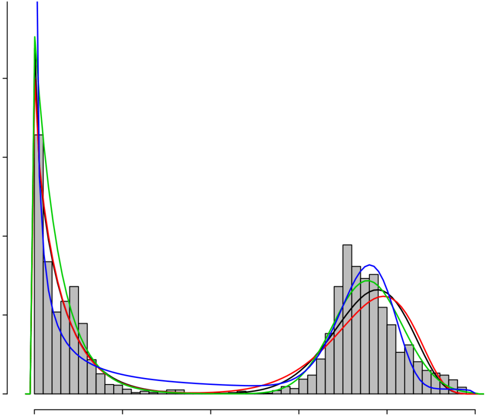
The well-known beta distribution has 2 parameters, mean $\mu \in (0, 1)$, and precision $\phi > 0$. Its density function is defined for $y \in (0, 1)$, where $\Gamma(.)$ is the gamma function:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{(\mu\phi-1)}(1-y)^{(1-\mu)\phi-1}.$$

The mixture of beta distributions

Given $f_1(y\,;\mu_1,\phi_1)$ and $f_2(y\,;\mu_2,\phi_2)$, two beta distributions, we set a mixture of two beta distributions, now with 5 parameters!:

$$f_{m_2b}(y\,;\pi,\mu_1,\phi_1,\mu_2,\phi_2)=\pi f_1(y\,;\mu_1,\phi_1)+(1-\pi)f_2(y\,;\mu_2,\phi_2).$$



The inflated mixture of beta distributions

The Y distribution is said to be an inflated (in zeros and ones) mixture of two beta distributions, with a 7-parameter $\vartheta = (\delta_0, \delta_1, \pi, \mu_1, \phi_1, \mu_2, \phi_2)$, if its density function is given by:

$$f_{im_2b}(y;\vartheta) = \begin{cases} \delta_0, & \text{if} \quad y = 0 \\ (1-\delta_0-\delta_1) f_{m_2b}, & \text{if} \quad 0 < y < 1 \\ \delta_1, & \text{if} \quad y = 1, \end{cases}$$

where f_{m_2b} follows a mixture of two beta distributions.

The first moment (mean) of Y, given by

$$E[Y] = (1-\delta_0-\delta_1)(\pi\mu_1 + (1-\pi)\mu_2) + \delta_1.$$

Parameter estimation

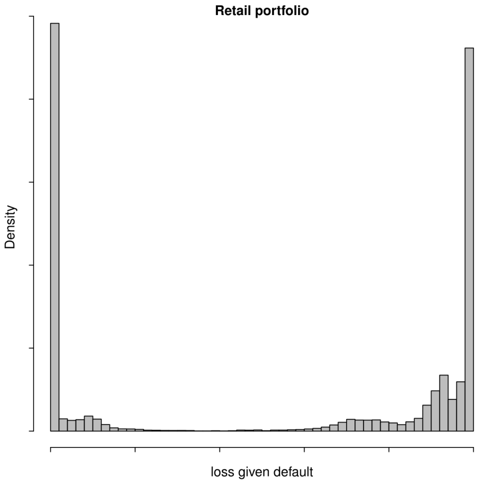
Parameter estimation is performed by straightforward use of maximum likelihood estimation (MLE).

Despite the existence of competitive methods, see for example expectation-maximization (EM) algorithm, our simulation studies support simple application of MLE approach.

The likelihood function of the inflated mixture model f_{im_2b}, with a vector 7-parameter $\vartheta = (\delta_0, \delta_1, \pi, \mu_1, \phi_1, \mu_2, \phi_2)$, is based on a sample of n observations, $\mathcal{D} = \{y_i\}$, independent and identically distributed:

$$L(\vartheta; \mathcal{D}) \propto \prod f_{im_2b}(y_i; \delta_0, \delta_1, \pi, \mu_1, \phi_1, \mu_2, \phi_2).$$

Brazilian bank non-performing retail data

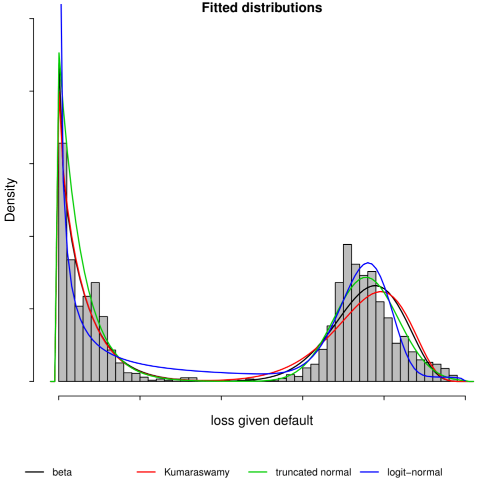


Brazilian bank non-performing retail data

Tabela: Summary of observed LGD data.

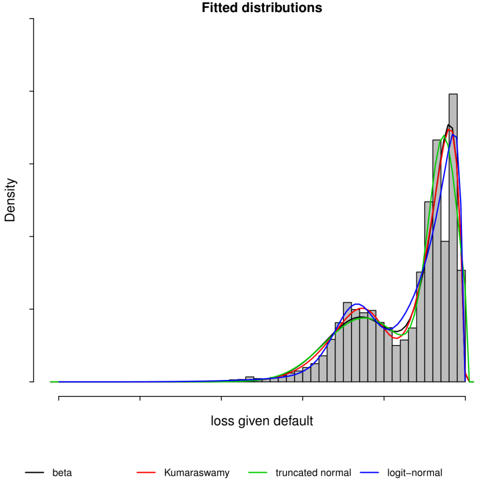
Portfolio	Qtd	Mean	Median	SD	#0	#1
1	15,295	0.52195	0.7272	0.4746	5722	6634
2	22,951	0.59814	0.9093	0.4596	8349	8398
3	440	0.32945	0.7466	0.4004	232	44
4	2,991	0.72060	0.9175	0.3810	510	265

Portfolio 1

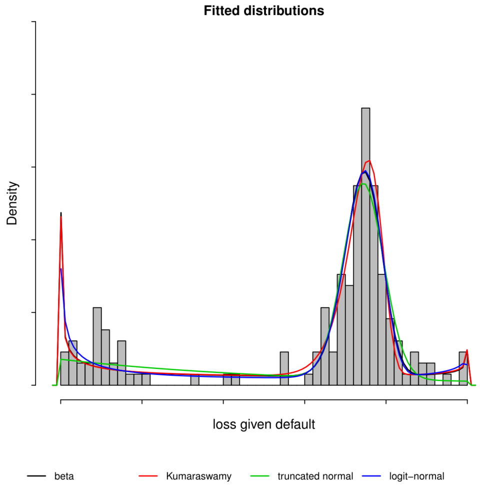


Fitted distributions
Density
loss given default
beta
Kumaraswamy
truncated normal
logit−normal

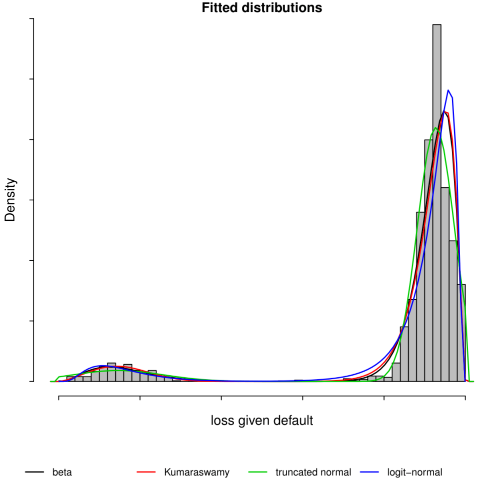
Portfolio 2


Fitted distributions
Density
loss given default
beta
Kumaraswamy
truncated normal
logit−normal

Portfolio 3


Fitted distributions
Density
loss given default
beta
Kumaraswamy
truncated normal
logit−normal

Portfolio 4


Fitted distributions
Density
loss given default
beta
Kumaraswamy
truncated normal
logit−normal

Brazilian bank non-performing retail data

Tabela: Summary of observed LGD data.

Portfolio	Qtd	Mean	Median	SD	#0	#1
1	15,295	0.52195	0.7272	0.4746	5722	6634
2	22,951	0.59814	0.9093	0.4596	8349	8398
3	440	0.32945	0.7466	0.4004	232	44
4	2,991	0.72060	0.9175	0.3810	510	265

Tabela: Expected mean LGD by portfolio and by model.

Portfolio	Expected lgd			
	beta	Kumaraswamy	Truncated normal	Logit-normal
1	0.52205	0.52173	0.52187	0.52272
2	0.59810	0.59817	0.59814	0.59789
3	0.33187	0.33129	0.32993	0.33088
4	0.72072	0.72021	0.72060	0.71841
Relative difference %	**0.1912%**	0.1167%	0.0330%	0.0590%

Agenda

Our contribution: The regression model version

$$f_{im_2b}(y;\vartheta) = \begin{cases} \delta_0, & \text{if} \quad y = 0 \\ (1-\delta_0-\delta_1) f_{m_2b}, & \text{if} \quad 0 < y < 1 \\ \delta_1, & \text{if} \quad y = 1, \end{cases}$$

$$\begin{cases} (\delta_{0i}, \delta_{1i}) & = & \left(\frac{e^{\mathbf{x}_{1i}^\top \beta_1}}{1+e^{\mathbf{x}_{1i}^\top \beta_1}+e^{\mathbf{x}_{2i}^\top \beta_2}}, \frac{e^{\mathbf{x}_{2i}^\top \beta_2}}{1+e^{\mathbf{x}_{1i}^\top \beta_1}+e^{\mathbf{x}_{2i}^\top \beta_2}} \right) \\ \mu_{1i} & = & \frac{e^{\mathbf{x}_{3i}^\top \beta_3}}{1+e^{\mathbf{x}_{3i}^\top \beta_3}} \\ \mu_{2i} & = & \frac{e^{\mathbf{x}_{4i}^\top \beta_4}}{1+e^{\mathbf{x}_{4i}^\top \beta_4}}, \end{cases}$$

Note that the inflated mixture of beta regression model can be viewed as an extension of the inflated beta regression model introduced by Ospina, R. & Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. Computational Statistics & Data Analysis, 56(6), 1609-1623.

Application data

- We consider a sample of the portfolio 1, containing 5000 retail loans.
- We consider two covariates made available by the bank.
- Let $(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3})$, where $\mathbf{x_1}$ is the interceptor parameter, i.e., $\mathbf{x_1} = 1$, and two others are real covariates.
- $\mathbf{x_2}$ represents two group of clients according to the behavioural risk presented. The bank has its behaviour score model and segregated their customers into two groups, roughly, $\mathbf{x_2} = 0$ to customers with poor credit risk and $\mathbf{x_2} = 1$ with better credit risk.
- $\mathbf{x_3}$ is related to the loan characteristics. The loan classified as $\mathbf{x_3} = 0$, represents a group of loans with term relatively shorter than the group with $\mathbf{x_3} = 1$.

Application data

Thus, we use the following setting of the link functions:

$$
\begin{cases}
\delta_{0i} & = & \dfrac{e^{(\beta_{11}\mathbf{x}_{1i}+\beta_{12}\mathbf{x}_{2i}+\beta_{13}\mathbf{x}_{3i})}}{1+e^{(\beta_{11}\mathbf{x}_{1i}+\beta_{12}\mathbf{x}_{2i}+\beta_{13}\mathbf{x}_{3i})}+e^{(\beta_{21}\mathbf{x}_{1i}+\beta_{22}\mathbf{x}_{2i}+\beta_{23}\mathbf{x}_{3i})}} \\
\delta_{1i} & = & \dfrac{e^{(\beta_{21}\mathbf{x}_{1i}+\beta_{22}\mathbf{x}_{2i}+\beta_{23}\mathbf{x}_{3i})}}{1+e^{(\beta_{11}\mathbf{x}_{1i}+\beta_{12}\mathbf{x}_{2i}+\beta_{13}\mathbf{x}_{3i})}+e^{(\beta_{21}\mathbf{x}_{1i}+\beta_{22}\mathbf{x}_{2i}+\beta_{23}\mathbf{x}_{3i})}} \\
\mu_{1i} & = & \dfrac{e^{(\beta_{31}\mathbf{x}_{1i}+\beta_{32}\mathbf{x}_{2i}+\beta_{33}\mathbf{x}_{3i})}}{1+e^{(\beta_{31}\mathbf{x}_{1i}+\beta_{32}\mathbf{x}_{2i}+\beta_{33}\mathbf{x}_{3i})}} \\
\mu_{2i} & = & \dfrac{e^{(\beta_{41}\mathbf{x}_{1i}+\beta_{42}\mathbf{x}_{2i}+\beta_{43}\mathbf{x}_{3i})}}{1+e^{(\beta_{41}\mathbf{x}_{1i}+\beta_{42}\mathbf{x}_{2i}+\beta_{43}\mathbf{x}_{3i})}}
\end{cases}
$$

The results

The results corroborate the finding that longer term loans held by lower credit risk clients have a much lower loss given default that the remaining group.

Tabela: Summary of average LGD estimated by the inflated mixture of two beta regression model

Portfolio 1	Subgroups		Qtd	Observed	Estimated
mean lgd = 0.5216	$\mathbf{x}_2 = 0$	$\mathbf{x}_3 = 0$	1,266	0.6325	0.6383
		$\mathbf{x}_3 = 1$	1,269	0.6324	0.6255
	$\mathbf{x}_2 = 1$	$\mathbf{x}_3 = 0$	1,234	0.4262	0.4265
		$\mathbf{x}_3 = 1$	1,231	0.3890	0.4101

The results: estimated parameters

Tabela: MLE results for the inflated mixture of two beta regression models

Parameter	Estimative (est)	Standard error (se)	$\|est\|/$ se
β_{11}	0.1890	0.0707	2.6731
β_{12}	0.8122	0.0809	10.0333
β_{13}	0.0851	0.0802	1.0599
β_{21}	0.9171	0.0641	14.3059
β_{22}	-0.2448	0.0786	3.1132
β_{23}	0.0013	0.0777	0.0173
π	0.5784	0.0828	6.9789
β_{31}	-0.7320	0.1005	7.2790
β_{32}	-0.2577	0.1086	2.3727
β_{33}	0.0728	0.1079	0.6750
ϕ_1	1.1757	0.0538	21.8377
β_{41}	1.1076	0.0293	37.7084
β_{41}	-0.1125	0.0349	3.2177
β_{41}	-0.0508	0.0351	1.4478
ϕ_2	62.8578	0.1110	566.0437

Selected Reference

- Calabrese, R.
  Downturn loss given default: Mixture distribution estimation.
  European Journal of Operational Research, 237(1), 271–277, 2014.
- Ospina, R. & Ferrari, S. L.
  A general class of zero-or-one inflated beta regression models.
  Computational Statistics & Data Analysis, 56(6), 1609–1623, 2012.
- Tong, Edward NC, Christophe Mues, and Lyn Thomas.
  A zero-adjusted gamma model for mortgage loan loss given default.
  International Journal of Forecasting, 29(4), 548–562, 2013.

Contact

- mauroexatas@gmail.com

Oliveira, M.R.[1], Louzada, F.[2], Pereira, G.H.A.[3], Moreira, F.[4], Calabrese, R.[5],
Inflated Mixture Models: Applications to Multimodality in Loss Given Default (July, 2015). Available at SSRN: http://ssrn.com/abstract=2634919

Thank you!

[1]Department of Credit Modelling at Caixa Econômica Federal Bank and Department of Statistics at Federal University of São Carlos, Brazil
[2]Department of Statistics at University of São Paulo, Brazil
[3]Department of Statistics at Federal University of São Carlos, Brazil
[4]University of Edinburgh Business School, UK
[5]University of Essex Business School, UK