UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Finanzas

Estimation of Probabilities for Ordered Sets and Application to Calibration of Rating Models

Gustavo F. Serenelli y Emiliano Delfau

Marzo 2023
Nro. 849

https://ucema.edu.ar/publicaciones/doc_trabajo.php
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding ved@ucema.edu.ar

Estimation of Probabilities for Ordered Sets and Application to Calibration of Rating Models

Gustavo F. Serenelli and Emiliano Delfau *

University of CEMA,
Argentina

March 2023

Abstract

The goal of this document is to present a methodology for estimating probabilities for ordered sets. This may have several practical applications such as calibration of Rating Models, estimation of Mortality Tables or measurement of side effects related to different doze sizes. In order to do this, an Objective / Non Informative Bayesian approach is applied, through which, using a multidimensional Jeffreys prior, a posterior distribution may be inferred for each of the probabilities being estimated

keywords: Bayesian estimation, probability estimation, uninformative prior, rating calibration, low default, ordered sets, jeffreys prior.

*The opinions expressed in this document are those of the authors and do not necessarily reflect the views of the UCEMA or CEBaFi. Comments are welcome at: `gustavoserenelli@gmail.com` or `ed11@ucema.edu.ar`

Introduction

Credit Rating Calibrations may be problematic in Low Default Portfolios. In these cases, it is usual to have, for several observations of one year credit exposures, divided in different credit ratings, a binary record of their credit behaviour (Default or Not Default). Even though it is desired that, for worse credit ratings, higher Probabilities of Default (PD) should be observed, it may be the case that, given potential lack of sufficient historical information, this may not necessarily happen. In order to work with a particular example, the following hypothetical historical data is presented:

Rating	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**	**CC**	**C**
Exposures	100	100	100	100	100	100	100	100	100
Defaults	1	0	0	1	2	1	2	3	4

Table 1: Hypothetical historical default data

In this case, for example, the naive estimation of the AAA PD (1%) is higher than the naive estimation of the AA or the A PD (0%). In addition, the naive estimation of the BB PD (2%) is higher than the naive estimation of the B PD (1%). None of these situations should be tolerated in estimations of PDs for different credit ratings. This is due to the fact that these ratings are assigned to borrowers having in mind economic and financial analysis that should have some positive correlation with their PDs.

In general, the solution to this problem is to fit a monotonically increasing curve to the PDs data points. However, this approach is not objective because the functional form of the curve is chosen, in general, through a subjective analysis. This subjectivity may pose important problems to the practical use of these estimates, given that a change in the subjective choices may generate important changes in the economic inferences that may be done with these PDs.

In order to solve this problem, an Objective / Non Informative Bayesian approach is presented. Using a Montecarlo Simulation, different ordered curves of PDs are proposed and each of them is weighted with a measure of consistency with the observed defaults. Given that:

- The curves of PDs proposed are ordered (respecting the rating hierarchy)
- The same weights are used among the different PD estimations of each notch

The resulting estimated probabilities are appropriately and naturally ordered. Also, this order is achieved through an objective procedure, given that the method used to generate the scenarios of curves of PDs relies on an Objective / Non Informative prior distribution.

This paper is organized as follows:

- Summary of Bayesian Methods: A recapitulation of the Bayesian Methods usually used for parameter estimations and a description of the main Objective / Non Informative prior distributions used for this goal (Jeffreys Prior and Reference Prior) are presented.
- Derivation of Prior Distribution: For this particular Rating Calibration problem, using the methods presented in the previous chapter, a Jeffreys Prior Joint Distribution is derived.
- Simulation of Scenarios: Considering the Jeffreys Prior Joint Distribution derived in the previous chapter, a Montecarlo Model for simulating ordered PD curves is built and implemented and, as a consequence, their main results are analyzed.
- PDs estimation: With the ordered PD curves simulated in the previous chapter, the PD of each rating notch is estimated. As a consequence, for different variations of the hypothetical historical databases, these estimations are analyzed and sensibilized.

Summary of Bayesian Methods

In order to present the main Bayesian Methods, the exposition will be restricted to the problem of estimating probabilities. However, it is important to consider that these methods are naturally applied to many other estimation problems such as estimations of means, standard deviations or regression models.

In the case of estimation of a probability, its naive estimation is achieved by measuring the relative frequency of occurrence of an event, observed in a number of repetitions of the experiment. In cases of PDs estimations for low default portfolios, it may happen that there are no default occurrences at all and, as a consequence, the naive estimation of the corresponding PD would be 0. However, when using this estimation for measuring Credit Risk Allowances or Economic Capital Requirements, this would produce non positive results, which sounds non prudent.

As a consequence, a Bayesian measurement can be helpful in this situation. In this case, for different PD scenarios, it can be measured how consistent each scenario is with respect to the observed data. Supposing "n" observations and "x" defaults, this measurement of consistency can be calculated by the Binomial Probability of observing "x" successes in "n" experiments with a success probability of "PD":

$$Bin(n, x, PD) = \binom{n}{x} PD^x \, (1 - PD)^{n-x} \tag{1}$$

And, as a consequence, the estimation of the PD for this scenario (let's call it $(\overline{PD})$) of "x" defaults in "n" historical expositions, can be calculated averaging each PD scenario using a weight proportional to the binomial consistency measurement:

$$\overline{PD} = \frac{\int_0^1 PD \cdot \binom{n}{x} PD^x \, (1 - PD)^{n-x} \cdot 1 \cdot dPD}{\int_0^1 \binom{n}{x} PD^x \, (1 - PD)^{n-x} \cdot 1 \cdot dPD} \tag{2}$$

In this equation, a second weight of 1 has been explicitly written. These two weights (the "binomial" weight and the "one" weight), may be interpreted as below:

1. "One" weight: Each scenario is equally likely. Nature would generate PD scenarios in a uniform way.
2. "Binomial" weight: once the PD scenario is generated, the more likely it is to observe the default data, the higher that scenario should be weighted.

In an informal interpretation this may be seen as if the universe branches in the infinite combination of scenarios of PDs and number of defaults ("x"). In these terms, intuitively, for each PD value, the product of these two weights is proportional to the amount of scenarios, consistent with that PD value and with the amount of "x" observed defaults.

The "one" weight is usually regarded as the Prior Distribution of the parameter. Also, after multiplying the two weights and dividing them by the sum of the products of both weights for each of the PD scenarios, a new distribution is obtained. This new distribution is usually called the Posterior Distribution of the parameter (in this case, the Posterior Distribution of the PD).

It can be proved that Equation 2 produces a result of $(x + 1)/(n + 2)$. This result, for large n, is closer to the naive estimation of the PD for "$x + 1$" successes than to the naive estimation of the PD for "x" successes. As a consequence, it looks as if this formula should have some kind of arrangement. In fact, it can be shown that this arrangement should be made to what has been called the "one" weight.

The "one" weight, given that it represents a uniform weight, may look as an objective choice. However, it may be proven that it is not necessarily objective. There are two observations that may be done to this choice and those are its *lack of invariance after a change of the parametrization of the problem and its non minimization of the information added to the problem.*

1. *Lack of invariance after a change of the parametrization of the problem:*

   In this context, we are working with a binomial problem represented by Equation 1. Given that this problem is generally stated in terms of the success probability ("PD"), it is usual to suppose that this PD should be uniform. However, nothing forbids the use of another parametrization of the binomial distribution. For example, defining a new variable θ, ranging from 0 to $\pi/2$, if we apply the following cosine square transformation:

   $$PD = \cos(\theta)^2 \tag{3}$$

   we can express the binomial distribution as the following:

   $$Bin(n, x, \theta) = \binom{n}{x} \cos(\theta)^{2x} \, \sin(\theta)^{2n-2x} \tag{4}$$

If the general practice would be to state the binomial problem through Equation 4, we would be tempted to assign a uniform distribution to the θ parameter. However, in this case, if we transform the problem back to Equation 1 form (using Equation 3 transformation), the PD parameter would inherit (from the θ parameter) a non uniform distribution. As a consequence, depending on the parametrization in which the problem is expressed, the prior distribution may differ. Hence, the estimation results will depend on the subjective selection of the parametrization of the problem.

To avoid this problem, there's a methodology called "Jeffrey's rule prior" (Jeffreys 1946, 1961), which states that, an invariant Prior Distribution, for a general distribution $p(\boldsymbol{x}|\boldsymbol{\gamma})$ (not necessarily binomial), with a vectorial m-dimensional random variable $\boldsymbol{x}$ and a vectorial n-dimensional parameter $\boldsymbol{\gamma}$, may be calculated as follows[1]:

$$\pi(\gamma_1, \gamma_2, ..., \gamma_n) = \det(I(\boldsymbol{\gamma}))^{1/2} \tag{5}$$

Where $I(\boldsymbol{\gamma})$ represents the Information Matrix with (i, j) element:

$$I(\boldsymbol{\gamma})_{i,j} = E_{\boldsymbol{x}|\boldsymbol{\gamma}}\left[\frac{\partial \log(p(\boldsymbol{x}|\boldsymbol{\gamma}))}{\partial \gamma_i}\frac{\partial \log(p(\boldsymbol{x}|\boldsymbol{\gamma}))}{\partial \gamma_j}\right] \tag{6}$$

In particular, applying this methodology to the binomial (one dimensional) problem, where $\boldsymbol{\gamma} = PD$ and $p(\boldsymbol{x}|\boldsymbol{\gamma})$ is equal to $\binom{n}{x}PD^x(1-PD)^{n-x}$ we obtain:

$$\frac{\partial \log(Bin(n, x, PD))}{\partial PD} = \frac{x}{PD} - \frac{n-x}{1-PD} \tag{7}$$

Hence, its information matrix is just a scalar given by the expected value of the square of equation 7, which, in turn, is equal to:

$$I(PD) = E_{x|PD} = \left[\left(\frac{x}{PD} - \frac{n-x}{1-PD}\right)^2\right]$$

$$= E_{x|PD}\left[\left(\frac{x}{PD}\right)^2\right] - E_{x|PD}\left[2\frac{x}{PD}\frac{n-x}{1-PD}\right] + E_{x|PD}\left[\left(\frac{n-x}{1-PD}\right)^2\right]$$

Considering that:

$$E_{x|PD}\left[\left(\frac{x}{PD}\right)^2\right] = \frac{n \cdot PD \cdot (1-PD) + n^2 \cdot PD^2}{PD^2}$$

$$E_{x|PD}\left[\left(\frac{n-x}{1-PD}\right)^2\right] = \frac{n \cdot PD \cdot (1-PD) + n^2 \cdot (1-PD)^2}{(1-PD)^2}$$

$$-E_{x|PD}\left[2\frac{x}{PD}\frac{n-x}{1-PD}\right] = -2E_{x|PD}\left[\frac{n \cdot x}{PD \cdot (1-PD)}\right] + 2E_{x|PD}\left[\frac{x^2}{PD \cdot (1-PD)}\right]$$

$$= -2\frac{n^2}{(1-PD)} + 2\frac{n \cdot PD \cdot (1-PD + n^2 \cdot PD^2)}{PD \cdot (1-PD)} = +2n - 2n^2$$

Adding the three terms, the information matrix is equal to:

$$I(PD) = \frac{n}{PD \cdot (1-PD)}$$

Given that this matrix is one dimensional, the calculation of its determinant is trivial and, hence, the Jeffreys Prior, given by Equation 5, is equal to:

$$\pi(PD) = \frac{1}{\sqrt{PD \cdot (1-PD)}} \tag{8}$$

In this expression we have got rid of the constant $\sqrt{n}$ in the numerator given that it does not contribute in any way in the weighting of the PD scenarios[2].

[1] The equality holds up to a normalization constant.

[2] It can be demonstrated that the normalization constant of this distribution is $1/\pi$. The distribution obtained as a result is equal to the $Beta(1/2, 1/2)$ distribution.

With this prior distribution, the Bayesian PD estimation is:

$$\overline{PD} = \frac{\int_0^1 PD \cdot \binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD) \cdot dPD}{\int_0^1 \binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD) \cdot dPD} \tag{9}$$

Equation 9 may be solved analytically and it can be demonstrated that the result of this estimation is the following:

$$\overline{PD} = \frac{x+0.5}{n+1} \tag{10}$$

This formula has many advantages in comparison to the traditional estimation of probabilities and to the bayes estimation of probabilities based on a Uniform prior:

(a) It Represents an intermediate value between the naive estimation of the PD for "$x+1$" successes and the naive estimation of the PD for "x" successes. As a consequence, this result, calculated using Jeffreys Prior, sounds more reasonable than the result obtained using a Uniform Prior.

(b) It is also worth noting that, when $x = 0$, this Bayesian estimation of the PD is equal to $0.5/(n+1)$. Hence, this method provides an objective procedure to generate a non zero estimate of the PD for historical data with zero defaults.

Lastly, it is important to mention that, if we would have calculated the Jeffreys Prior to the Binomial Distribution, using the parametrization described by Equation 4, we would have obtained that the invariant Jeffreys Distribution for θ is equal to the Uniform Distribution[3].

2. *Non minimization of the information added to the problem:*

Given what has been described up to this point, it is clear that the Uniform Distribution does not represent an Objective Choice (having in mind that the Uniformity of the PD depends on the subjective decision related to the parametrization of the Binomial Distribution). In addition to this, it can be claimed that the Uniform Distribution does not minimize the (subjective) information added to the problem.

Following a methodology initiated in Bernardo (1979), the problem of finding the statistical prior distribution that minimizes the (subjective) information added to the inference problem can be mathematically formulated. His approach is based on finding what is called the Reference Prior. That is to say, the Prior Distribution that maximizes the Mutual Information between the observed data ($\boldsymbol{x}$ in general terms) and the Parameters ($\boldsymbol{\gamma}$ in general terms [4][5]). Having in mind that the Mutual Information (MI) of two random variables (X and Y) with joint density $f(x,y)$ and marginal densities $g(x)$ and $h(y)$ can be calculated as follows:

$$MI(X,Y) = \iint f(x,y) \cdot \log \frac{f(x,y)}{g(x) \cdot h(y)} \cdot dxdy$$

The Mutual Information of the data $\boldsymbol{x}$ and the parameters $\boldsymbol{\gamma}$ can be calculated as follows:

$$MI(\boldsymbol{x},\boldsymbol{\gamma}) = \iint f(\boldsymbol{x},\boldsymbol{\gamma}) \cdot \log \frac{f(\boldsymbol{x},\boldsymbol{\gamma})}{g(\boldsymbol{x}) \cdot h(\boldsymbol{\gamma})} \cdot d\boldsymbol{x}d\boldsymbol{\gamma}$$

In particular, for the one dimensional case of the PD estimation using the binomial distribution, this calculation reduces to the following:

$$MI(x,PD) = \sum_{x=0}^{n} \int_0^1 \binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD) \cdot \log \frac{\binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD)}{\int_0^1 \binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD) dPD \cdot \pi(PD)} dPD \tag{11}$$

In this particular case:

[3]Given everything said, this uniform distribution for θ could have been obtained by two different procedures:

(a) Calculation of the Jeffreys Distribution to Equation 4

(b) Perform a change of variables in the distribution defined by Equation 8, using the cosine square transformation described in Equation 3.

The coincidence of these two different procedures is an illustration of the above stated claim that the Jeffreys Prior is invariant to changes in the parametrization of the problem.

[4]Here, both the parameters ($\boldsymbol{\gamma}$) and the data ($\boldsymbol{x}$) are considered as multidimensional random variables, not necessarily of the same dimension.

[5]The intuition behind this statement would be that maximizing the Mutual Information that is obtained by knowing the parameters ($\boldsymbol{\gamma}$) or the data ($\boldsymbol{x}$), would mean minimizing the information that one adds to the data (x) by supposing a particular distribution of the parameters ($\boldsymbol{\gamma}$).

- The data ($\boldsymbol{x}$) has been replaced by the amount of defaults x
- The parameters ($\boldsymbol{\gamma}$) were replaced by the probability of default PD
- The joint density of the data and the parameter ($f(\boldsymbol{x}, \boldsymbol{\gamma})$) has been replaced by $\binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD)$
- The density of the parameters $h(\boldsymbol{\gamma})$ has been replaced by $\pi(PD)$
- The marginal density of the parameters $g(\boldsymbol{x})$ has been replaced by $\int_0^1 \binom{n}{x} PD^x (1-PD)^{n-x} \cdot \pi(PD) dPD$.

The optimization problem of finding the distribution of the parameters ($h(\boldsymbol{\gamma})$) that maximizes the above mentioned Mutual Informations may be computationally hard. However, it can be demonstrated that, for the Binomial problem stated in Equation 11, the Prior Distribution of the PD ($\pi(PD)$) that finds its maximum coincides with Jeffreys Prior (Equation 8).

Even though the Reference Prior method can be more accurate in detecting the Prior Distribution that minimizes subjectivity than Jeffreys method (given that Jeffreys method relies on invariance arguments and not on minimization of subjectivity), in the following, given its computational simplicity, we will derive our estimates using Jeffreys Method.

Derivation of the Prior Distribution

For the problem of calibrating a rating model of "m" rating notches, we would have, for each rating notch "i", an amount "n_i" of independent exposures and an amount "x_i" of defaults. Both exposures and defaults can be aggregated, respectively, in "m" dimensional vectors "$\boldsymbol{n}$" and "$\boldsymbol{x}$". Hence, the probability of observing vector "$\boldsymbol{x}$" defaults in vector "$\boldsymbol{n}$" exposures given the vector "$\boldsymbol{PD}$" of Probabilities of Default is:

$$Bin(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{PD}) = \prod_{i=1}^{m} \binom{x_i}{n_i} PD_i^{x_i} (1 - PD_i)^{n_i - x_i} \tag{12}$$

Applying Equation 6 to this equation, in order to obtain the information matrix, would mean to perform the following calculation for each i and j:

$$I(\boldsymbol{PD})_{i,j} = E_{x_i, x_j | PD_i, PD_j} \left[\left(\frac{x_i}{PD_i} - \frac{n_i - x_i}{1 - PD_i} \right) \left(\frac{x_j}{PD_j} - \frac{n_j - x_j}{1 - PD_j} \right) \right] \tag{13}$$

For $i \neq j$, given the independence of x_i and x_j, this expected value can be calculated by applying the expected value operator of each of the products. Given that the expected value of both products is 0, Equation 13 is 0 for $i \neq j$.

For $i = j$, we are back to the one dimensional case. As a consequence, we obtain that:

$$I(\boldsymbol{PD})_{i,i} = \frac{n_i}{PD_i \cdot (1 - PDi)} \tag{14}$$

Hence, the information matrix is zero outside the main diagonal and is equal to Equation 14 inside the main diagonal. As a consequence, its determinant is equal to the product of all its diagonal elements and the Jeffreys Prior would be equal to the square root of that product[6]:

$$\pi(\boldsymbol{PD}) = \prod_{i=1}^{m} \sqrt{\frac{1}{PD_i \cdot (1 - PDi)}} \tag{15}$$

Performing the cosine square transformation stated in Equation 3 to each PD_i, as in the one dimensional case, we obtain a uniform joint distribution[7]. That is to say, up to a normalization constant:

$$\pi(\boldsymbol{\theta}) = 1 \tag{16}$$

[6] Again, we have got rid of the redundant constant.

[7] This is due to the fact that the joint density obtained is a product of independent individual densities (independent in this context means that each individual density depends on one variable only). Hence, the Jacobian of this joint density reduces to the product of the independent derivatives of the individual densities above mentioned. Given what has been mentioned, the joint density expressed in terms of θ reduces to the product of the individual densities of each θ_i and, given that all of them are equal to 1, the claim is demonstrated.

Simulation of scenarios

Given Equation 16, it looks easier to simulate scenarios related to $\boldsymbol{\theta}$ rather than $\boldsymbol{PD}$. However, it is important to have in mind that each pair of consecutive PDs (PD_i and PD_{i+1}) should respect the following relation:

$$PD_i < PD_{i+1} \tag{17}$$

Or, in terms of θ (given the monotonically inverse relation between PD and θ defined in Equation 3):

$$\theta_i > \theta_{i+1} \tag{18}$$

Considering the constraint described in Equation 18, the Montecarlo Simulation Method proposed:

- Starts simulating a scenario for θ_1,
- Then it will generate a scenario for θ_2, restricted to the condition that θ_2 should be smaller than θ_1,
- Then it will generate a scenario for θ_3, restricted to the condition that θ_3 should be smaller than θ_2,
- And this procedure continues to operate recursively until it reaches the simulation of the last variable (θ_m)

Given this procedure, we will need to know the following distributions:

- Marginal distribution of θ_1
- Distribution of θ_2 conditioned on the occurrence of θ_1
- Distribution of θ_3 conditioned on the occurrence of θ_2
- ...
- Distribution of θ_m conditioned on the occurrence of θ_{m-1}

Instead of deriving these distributions through analytic procedures, we will visually analyze its densities for $m = 2$ and $m = 3$ cases and, considering its simple joint distribution given in Equation 16, we will easily extrapolate our conclusions to the general m-dimensional case. Also, in order to simplify this calculation, we will work with variables restricted to the $[0, 1]$ interval and, then, the simulation will be adjusted in order to represent the correct interval ($[0, \frac{\pi}{2}]$).

For the $m = 2$ case, the admissible region of $\boldsymbol{\theta}$ is the whole sets of pairs (θ_1, θ_2) such that $\theta_1 > \theta_2$. That region can be represented through the following graph:

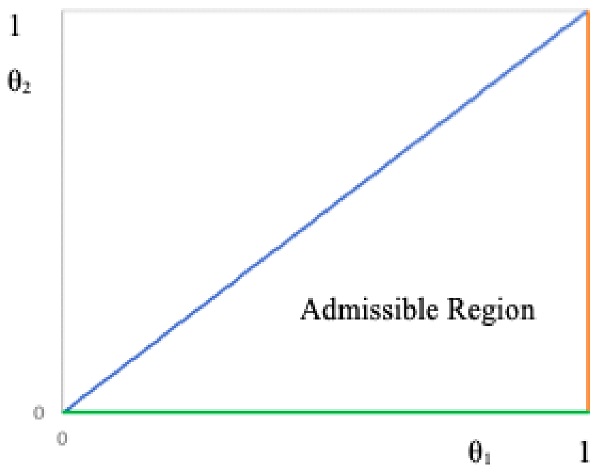

Figure 1: Admissible Region

Given that each of the points of the Admissible Region has the same probability (given the uniformity of the Prior Distribution of $\boldsymbol{\theta}$), there are more admissible scenarios for high θ_1 than for low θ_1. As we move from left to right in the graph, the amount of admissible scenarios grows linearly. Hence, we can easily see that the Cumulative Distribution of θ_1 ($CD(\theta_1)$) is equal to the area of the triangle accumulated from the zero $x-value$

up to the θ_1 $x - value$ (this area is equal to $\frac{\theta_1^2}{2}$) divided by the total area of the admissible region (this total area is equal to $1/2$):

$$CD(\theta_1) = \theta_1^2$$

Once a simulation of θ_1 is obtained, we will know that :

- the simulation of θ_2 should be lower than the simulated value of θ_1 and that,
- given Equation 16, each scenario is equally likely.

As a consequence, the conditioned Cumulative Distribution of θ_2 is[8]:

$$CD(\theta_2|\theta_1) = \frac{\theta_2}{\theta_1}$$

For the $m = 3$ case, we can generalize what we have obtained from the $m = 2$ case. The admissible region would be a 3-dimensional volume. For the particular case of $\theta_3 = 0$, the (θ_1, θ_2) admissible region would be the same as the one represented above. For a generic positive θ_3 value, the corresponding (θ_1, θ_2) admissible region would be also a similar right triangle but whose bottom left hand side point is (θ_3, θ_3) and whose upper right hand side point is $(1, 1)$. As a consequence, we can represent the 3-dimensional admissible region as follows:

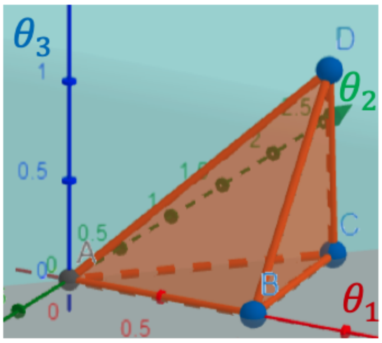

Figure 2: Admissible Region: 3-dimensional

Moving from left to right in the x direction (that is to say, in the θ_1 direction), increases the (θ_2, θ_3) admissible region. For each θ_1 value, the corresponding (θ_2, θ_3) 2-dimensional admissible region has an area of $\frac{\theta_1^2}{2}$. As a consequence, the density of θ_1 should be proportional to this area and the integral of this area should be proportional to the Marginal Cumulative Density of θ_1. Given that the integral of $\frac{\theta_1^2}{2}$ is proportional to θ_1^3 and that θ_1^3 is equal to 1 when θ_1 is equal to 1, we can conclude that:

$$CD(\theta_1) = \theta_1^3$$

Having determined a particular scenario for θ_1 we are back to the two dimensional problem, with the only difference that our admissible (θ_2, θ_3) region should be restricted to $\theta_2 < \theta_1$ and $\theta_3 < \theta_1$. As a consequence, the area of the total triangle is $\frac{\theta_1^2}{2}$. Scaling appropriately what has been mentioned for the $m = 2$ cases, we can conclude that:

$$CD(\theta_2|\theta_1) = \left(\frac{\theta_2}{\theta_1}\right)^2$$

And:

$$CD(\theta_3|\theta_2) = \left(\frac{\theta_3}{\theta_2}\right)$$

Extrapolating these ideas to the m-dimensional case, we can obtain the following marginal and conditional cumulative distributions:

$$CD(\theta_1) = \theta_1^m \tag{19}$$

$$CD(\theta_{i+1}|\theta_i) = \left(\frac{\theta_{i+1}}{\theta_i}\right)^{m-i}$$

[8]In simple terms, this distribution is uniform and accumulates 100% probability when $\theta_2 = \theta_1$

In order to perform montecarlo simulations, it is usual to work with the inverses of the Cumulative Distributions. Considering Equations 19, these are straightforward:

$$\theta_1 = CD\left(\theta_1\right)^{1/m}$$

$$\theta_{i+1} = CD\left(\theta_{i+1}|\theta_i\right)^{1/(m-i)}\theta_i \tag{20}$$

As a consequence, the Montecarlo Simulation Process will be the following:

1. Propose "m" $[0,1]$ independent uniform random numbers.
2. Replace them in $CD\left(\theta_1\right)$ and $CD\left(\theta_{i+1}|\theta_i\right)$ variables in Equations 20 in order to obtain the simulated vector $\boldsymbol{\theta}$.
3. Given that these parameters are restricted to the $[0,1]$ interval, and we need variables in the $[0,\frac{\pi}{2}]$ interval, we multiply the simulated vector $\boldsymbol{\theta}$ by $\frac{\pi}{2}$.
4. Calculate the $\boldsymbol{PD}$ vector applying the cosine square transformation stated in Equation 3 to each element of the simulated vector $\boldsymbol{\theta}$.

In order to analyze the simulated results, below there may be found a simulated marginal distribution for each PD for a 9-dimensional problem:

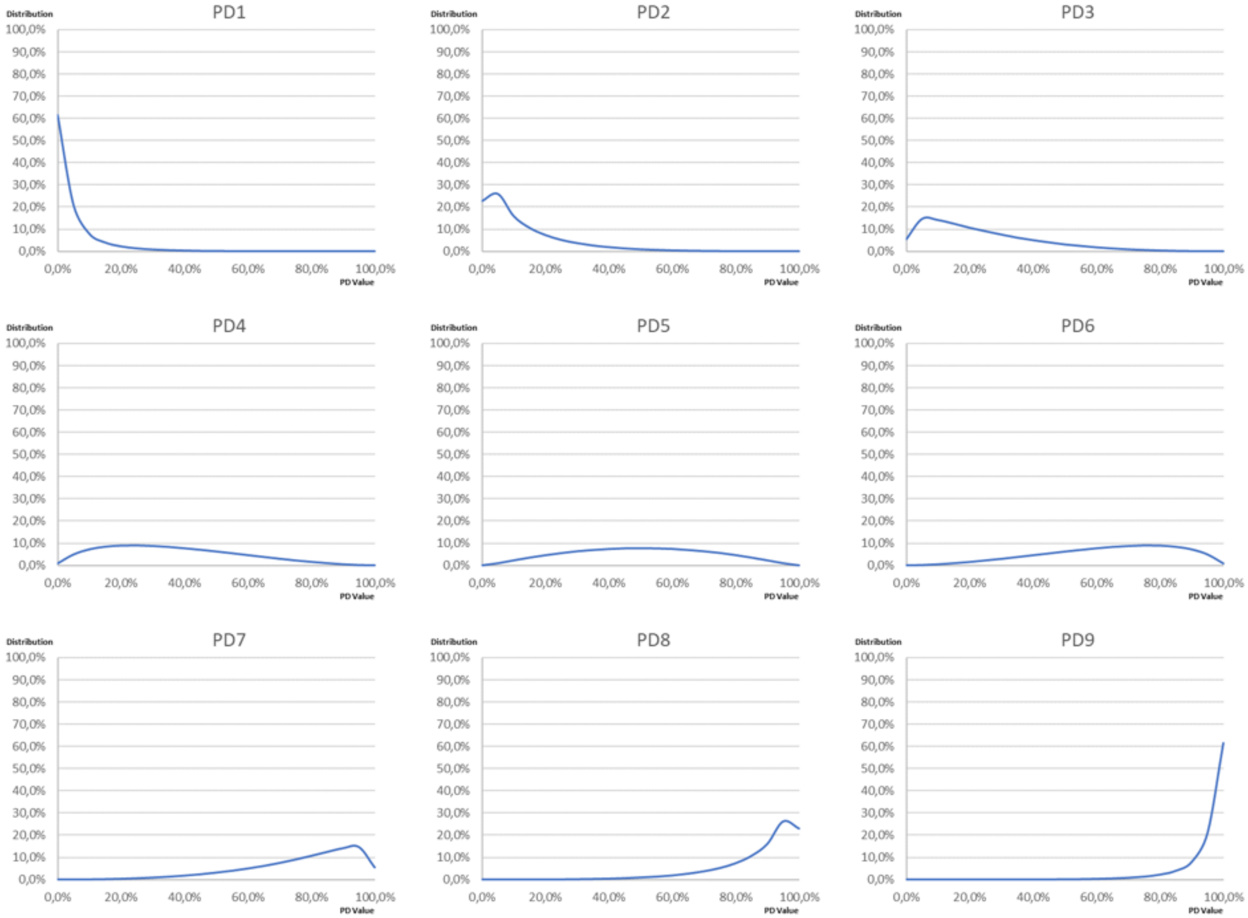

Figure 3: PD's marginal distribution: 9-dimensional problem

As it can be seen, PD_5 is a symmetrical distribution. The rest of the distributions are related in pairs. The distribution of PD_1 is a mirror transformation of the distribution of PD_9. The same happens between PD_2 and PD_8, PD_3 and PD_7 and PD_4 and PD_6. These symmetry properties are desired, given the uniformity of the joint density given in Equation 16.

PDs estimation

The last step of the procedure is to perform the Bayesian PD estimation. Analytically, for an m-dimensional rating calibration, the estimation of each PD should be done through the following calculation:

$$\overline{PD_i} = \frac{\int_0^1 \int_0^{PD_1} \dots \int_0^{PD_{m-2}} \int_0^{PD_{m-1}} PD_i \cdot \prod_{i=1}^m \binom{x_i}{n_i} PD_i^{x_i}(1-PD_i)^{n_i-x_i} \cdot \prod_{i=1}^m \sqrt{\frac{1}{PD_i \cdot (1-PD_i)}} \cdot dPD_m dPD_{m-1} \dots dPD_2 dPD_1}{\int_0^1 \int_0^{PD_1} \dots \int_0^{PD_{m-2}} \int_0^{PD_{m-1}} \prod_{i=1}^m \binom{x_i}{n_i} PD_i^{x_i}(1-PD_i)^{n_i-x_i} \cdot \prod_{i=1}^m \sqrt{\frac{1}{PD_i \cdot (1-PD_i)}} \cdot dPD_m dPD_{m-1} \dots dPD_2 dPD_1} \quad (21)$$

This is an m-dimensional integral that cannot be easily solved analytically[9]. However it can be approximated through the implementation of the simulation procedure stated before. This approximation works as follows:

$$\overline{PD_i} = \frac{\sum_\Omega \boldsymbol{PD}(\omega)_j \cdot \prod_{i=1}^m \binom{x_i}{n_i} \boldsymbol{PD}(\omega)_i^{x_i}(1-\boldsymbol{PD}(\omega)_i))^{n_i-x_i}}{\sum_\Omega \prod_{i=1}^m \binom{x_i}{n_i} \boldsymbol{PD}(\omega)_i^{x_i}(1-\boldsymbol{PD}(\omega)_i))^{n_i-x_i}} \quad (22)$$

In Equation 22 Ω represents the set of all simulated scenarios, ω represents one particular simulated scenario and $\boldsymbol{PD}(\omega)_i$ represents the PD simulated for the i^{th} notch and scenario ω.

The implementation of this procedure will be performed in a 9 dimensional context, with the data shown in Table 2. Below there can be found a comparison between the results of an implementation of Equation 22 using 1.000.000.000 scenarios of PD curves[10], the naive estimation of each PD and the one dimensional bayesian estimation stated in Equation 10 of each PD:

Notch	Naive	1D Bayes	9D Bayes Montecarlo	# Defaults	# Exposures
AAA	1,00%	1,49%	0,20%	1	100
AA	0,00%	0,50%	0,35%	0	100
A	0,00%	0,50%	0,59%	0	100
BBB	1,00%	1,49%	1,00%	1	100
BB	2,00%	2,48%	1,50%	2	100
B	1,00%	1,49%	2,00%	1	100
CCC	2,00%	2,48%	2,77%	2	100
CC	3,00%	3,47%	3,93%	3	100
C	4,00%	4,46%	5,95%	4	100

Table 2: 9D Bayes Montecarlo

As it is expected, neither the Naive estimation, nor the 1D Bayes estimation produces ordered estimates. As it is also expected, given their theoretical coincidences, the 9D Bayes Montecarlo method presented in this document is closer to the 1D Bayes estimates than to the Naive ones. The 9D Bayes estimate differs the most from the rest of estimates in notch C. This may be due to the fact that, given the order imposed to the scenarios, this estimation of the C notch PD leaves a bigger place for the rest of estimates of PDs to accommodate appropriately to the data set.

As a final analysis, given this difference in the estimation of the C notch PD, the amount of defaults was changed, just for this notch, from 4 to 9. The results obtained are the following.

Notch	Naive	1D Bayes	9D Bayes Montecarlo	# Defaults	# Exposures
AAA	1,00%	1,49%	0,20%	1	100
AA	0,00%	0,50%	0,36%	0	100
A	0,00%	0,50%	0,61%	0	100
BBB	1,00%	1,49%	1,04%	1	100
BB	2,00%	2,48%	1,57%	2	100
B	1,00%	1,49%	2,12%	1	100
CCC	2,00%	2,48%	3,03%	2	100
CC	3,00%	3,47%	4,65%	3	100
C	9,00%	9,41%	9,69%	9	100

Table 3: "C" notch changed: 9D Bayes Montecarlo

As it can be seen, given that there is more difference between the observed rates of default for CC and the C notches, the 9D Bayes estimate for the C notch reduces almost all its difference from the corresponding

[9]To be more precise, this is a set of m different m dimensional integrals, one for each of the m rating notches.
[10]The Python implementation of this algorithm may be found in Annex 1.

1D Bayes estimate. The conclusion would be that, with bigger gaps among the observed default rates of the different rating notches, there would be more coincidence between the 1D Bayes estimates and the 9D Bayes estimates.

Final Remarks

The naive methods for estimating probabilities need the use of a certain amount of subjectivity. This subjectivity may pose important problems to the practical use of these estimates, given that a change in the subjective choices may generate important changes in the economic inferences that may be done with these probabilities. The Bayesian methods presented in this document offer an objective procedure for estimating probabilities for both low default portfolios and ordered rating notches.

There are numerical challenges when applying this methodology to some particular contexts. For example, when there is a big amount of rating notches, the amount of scenarios needed to reach a reasonable level of precision may be unfeasible under actual computational limits. As a consequence, some adjustments, taking into account the particular characteristics of the problem under analysis, may be needed in order to acquire a sufficient level of precision.

Annex 1: Python Algorithm for the 9D Bayes Montecarlo Estimation

```
"IMPORT LIBRARIES"
import numpy
from numpy import random
import math

"DEFINE DATASETS"
X=[1,0,0,1,2,1,2,3,9]
N=[100,100,100,100,100,100,100,100,100]

"INTIALIZE VARIABLES"
PDEst=X*0
Esc=1000000000
tita = numpy.zeros((Esc,9))+1
PRODBIN=numpy.zeros((Esc,1))
PDEst=numpy.zeros((9,1))
DENEst=numpy.zeros((9,1))

"SIMULATION MODEL"
"IT STARTS FROM LAST NOTCH BECAUSE, FOR NUMERICAL REASONS, IT IS NEEDED TO"
"LIMIT THE LAST THETA VALUE TO 0,5 AT MOST"
for i in range(9):
    t=8-i
    x = random.rand(Esc,1)
    if t==8:
        tita[:,8]=pow(x[:,0],1/(8+1))*0.5
    else:
        tita[:,t]=pow(x[:,0],1/(t+1))*tita[:,t+1]
PD=pow(numpy.cos(math.pi/2-tita*math.pi/2),2)
BINOMIAL=PD*0

"WEIGHT CALCULATION"
for i in range(9):
    t=8-i
    BINOMIAL[:,i]=PD[:,i]**X[i]*(1-PD[:,i])**(N[i]-X[i])
PRODBIN[:,0]=numpy.prod(BINOMIAL,1)

"PD ESTIMATION"
for i in range(9):
    PDEst[i]=numpy.sum(PD[:,i]*PRODBIN[:,0])/numpy.sum(PRODBIN[:,0])
```

As it is noted in the Python code, the implemented algorithm started from the last rating notch and not from the first. This was needed because, the simulation of the last θ variable was truncated to be, at most, 0,5. Even though this produces a conceptual bias in the estimates, the binomial probabilities of the scenarios left behind were negligible. Hence their contribution to the total estimates were also negligible. In return, keeping those scenarios would have generated that there were very little scenarios of the θ_9 variable around the observed value. As a consequence, the estimation of the PD_9 parameter would have had a much bigger statistical error.

References

Basel Committee on Banking Supervision (2004), *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework section 444 to 485*, Bank for International Settlements

Basel Committee on Banking Supervision (2005), *An explanatory note on the Basel II IRB risk weight function*, Bank for International Settlements

Basel Committee on Banking Supervision (2011), *Basel III: A global regulatory framework for more resilient banks and banking systems.*, Bank for International Settlements

Berger, J. O. and Bernardo, J. M. (1992c). *On the development of reference priors (with discussion).* In Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 35–60. Oxford Univ. Press

Bernardo, J. M. (1979). *Reference posterior distributions for Bayesian inference (with discussion).* J. Roy. Statist. Soc. Ser. B 41 113–147. [Reprinted in Bayesian Inference (N. G. Polson and G. C. Tiao, eds.) 229–263. Edward Elgar, Brookfield, VT, 1995.]

Bernardo, J. M. (2005). *Reference analysis.* In Handbook of Statistics 25 (D. K. Dey and C. R. Rao, eds.) 17–90. North-Holland, Amsterdam.

IASB (2014):*International Financial Reporting Standard 9 Financial Instruments.* Including Implementation Guidance and Basis for Conclusions (IFRS 9, final standard)

Jeffreys, H. (1946). *An invariant form for the prior probability in estimation problems.* Proc. Royal Soc. of London, Series A, 186, 453-461.

Jeffreys, H. (1961). *Theory of Probability*, 3rd Edition, Clarendon Press, Oxford.

Kullback, S. (1959). *Information Theory and Statistics*, 2nd Edition. Dover, New York.

Kullback, S. and Leibler, R. A. (1951). *On information and sufficiency.* Ann. Math. Statist. 22 79–86.

Shannon, C. E. (1948). *A mathematical theory of communication.* Bell System Tech. J. 27 379–423, 623–656.

Tasche, D., & Pluto, K. (2005), *Estimating Probabilities of Default for Low Default Portfolios*, Deutsche Bundesban