Francois Mercier and Amy Racine

M&S-CIS, Novartis Pharma AG [email: francois.mercier@novartis.com]

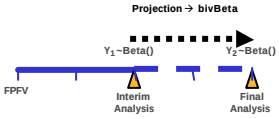
Objective

In certain clinical trials, the response variable is restricted into an interval (*e.g.* visual analogue scale VAS). This response variable may be measured repeatedly during the trial. A natural distribution for such a variable is a multivariate Beta distribution. For clinical trial simulations, it is therefore necessary to generate data from such a distribution in order to perform inference and/or predictions.
We propose hereafter to address the particular case of the bivariate Beta (*bivBeta*) distribution and present the approach (and corresponding code) we use to generate and analyze this type of data.

Methods

We assume a clinical trial in which a response variable *Y*, characterized by an interval distribution, is evaluated at both early (*Y1*) and late stage (*Y2*). The clinical trial may have been designed in such a way that at the occasion of an interim analysis, one wants to predict *Y2*, based on the available *Y1* data for decision making (e.g. futility or success of the trial).

Simulation of time trends in a random variable is frequently done within a multivariate framework where the different variables represent the state of the random variable at different point in time.



With random variables having possibly skewed interval distributions (see for instance, Figure 1), it is not recommended to use *e.g.* a truncated multivariate normal distribution for simulation. Rather, it is recommends to use multivariate Beta distributions.

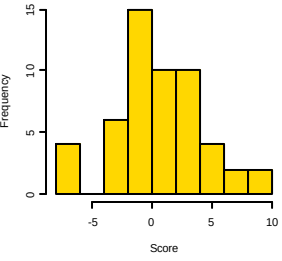


Figure 1: Distribution of a score range from -10 to 50, as an illustration of interval response variable.

The flexible Beta distribution is widely used in life sciences to describe the probability density distribution of proportions or relative frequencies of a random univariate variable.
Generation of univariate Beta-distributed random variable is straightforward, but generating pairs of correlated Beta-distributed random variables is more complex since there is no natural multivariate extension of univariate Beta distribution (Johnson and Kotz, 1976 [1]).

Following Catalani (2002)[2] it is proposed to use a Dirichlet distribution to simulate outcomes from a *bivBeta* distribution.

1/ Introduction of a shared random variable

The marginals in a Dirichlet distribution are Beta variables

Let $\{X_1, X_2, X_3\} \sim Dirichlet(a_0, a_1, a_2, a_3)$, with

$$X_i = Z_i/(Z_1+Z_2+Z_3), \quad i = 1, 2, 3$$

where Z_i's, are independent gamma variables $Z_i \sim G(shape=a_i, scale=1)$
A popular technique for generation of correlated random variables is to introduce a shared random variable:
Define

$$\begin{array}{l} Y_1 = X_1+X_3 \\ Y_2 = X_2+X_3 \end{array} \Leftrightarrow \begin{array}{l} Y_1 \sim Be(a_1+a_3, a_0+a_2) \\ Y_2 \sim Be(a_2+a_3, a_0+a_1) \end{array}$$

Set $g = (a_0+a_1+a_2+a_3)$, then we can derive the correlation coefficient([2])

$$r(Y_1, Y_2) = (-a_1a_2 + a_0a_3) / sqrt((a_1+a_3)(a_0+a_2)(a_2+a_3)(a_0+a_1))'$$

2/ Elicitation of the *alphas*

Sampling data from a bivariate density with beta marginals, with parameters, c_1, c_2 and c_3, c_4, and r, positive correlation coefficient,

Set

$$\begin{array}{l} c_1 = a_1+a_3 \\ c_2 = a_0+a_2 \\ c_3 = a_2+a_3 \end{array} \text{ and } c_4 = a_0+a_1 = c_1+c_2-c_3$$

which implies $c_1+c_2>c_3$ and $r = (-a_1a_2+a_0a_3)/sqrt(c_1c_2c_3(c_1+c_2-c_3))'$

Assuming $r>0$, we solve for $\{a_0, a_1, a_2, a_3\}$, as functions of $\{c_1, c_2, c_3, c_4, r\}$ and we obtain:

$$a_3 = r^* sqrt(c_1c_2c_3(c_1+c_2-c_3))+ c_1c_3 / (c_1+c_2)$$

it follows,

$$\begin{array}{l} a_1 = c_1 - a_3 \\ a_2 = c_3 - a_3 \\ a_0 = c_2 - c_3 + a_3 \end{array}$$

Results

1/ Data generation

The Beta distribution parameters *c1, c2* and *c3* can be derived from the mean and standard deviation of *Y1* as follows (note: *c4 = (c1+c2) – c3*):

$$c_1 = \left[\hat{m}_{Y1}^2 \times \frac{(1-\hat{m}_{Y1})}{\hat{s}_{Y1}^2}\right] - \hat{m}_{Y1}$$
$$c_2 = c_1 \times \left(\frac{1}{\hat{m}_{Y1}} - 1\right)$$
$$c_3 = (c_1 + c_2) \times \hat{m}_{Y2}$$

In the context of the clinical trial described earlier, depending on the sample mean at time 1 and time 2, sample variance and expected coefficient of correlation, we would obtain the following Beta parameters estimates, in the two below examples (Table 1 and Figure 2):

Ex. 1	m_1=0.4, m_2=0.5, s_1^2=0.2, r=0.7	c_1= 2 c_2= 3 c_3= 2.5 c_4= 2.5	a_0= 2.357 a_1= 0.143 a_2= 0.642 a_3= 1.857
Ex. 2	m_1=0.6, m_2=0.7, s_1^2=0.2, r=0.5	c_1= 3 c_2= 2 c_3= 3.5 c_4= 1.5	a_0= 1.161 a_1= 0.339 a_2= 0.839 a_3= 2.661

Table 1: Elicitation of alpha values in two specific examples

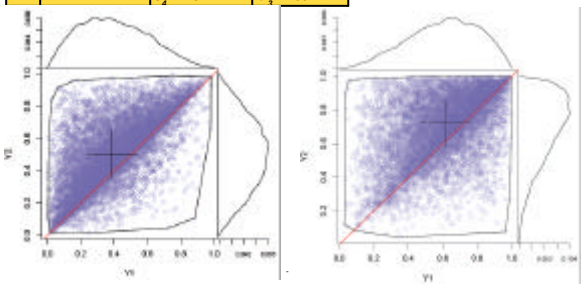
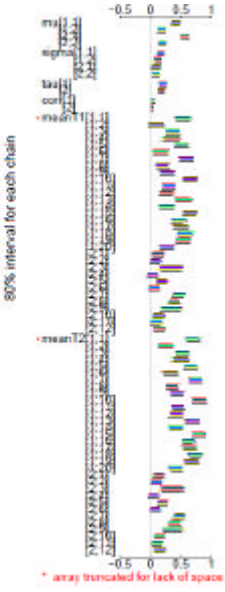
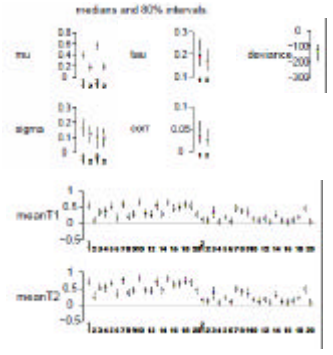


Figure 2: Scatterplots and densityplots of bivariate Beta distributed variables. Red line represents x=y and cross represents median Y1 and median Y2.

2/ Data analysis

To analyze the *bivBeta* data, a Bayesian approach was used considering the distribution of the endpoint as bivariate Normal, after *logit* transformation. As an illustration, we simulated a simple clinical trial scenario involving one placebo group and one active treatment group. The simulated data corresponded to 20 patients per group, with Beta distributed data at two occasions T1 and T2. The output of this analysis is provided hereafter:





In a second example (not reported here) we simulated a scenario where Y1 (response at T1) were complete but Y2 (response at T2) were incomplete. This scenario could, for instance, correspond to the situation observed at an interim point during a clinical trial.

Using the same data set as described above but replacing 50% of the original data at T2 by missing values, we could fit the same model and obtain similar parameter estimates (but more variability in the meanT2 estimations) [Results available on demand].

Conclusion

We present a method to generate bivariate Beta distributed data, where:

1. Beta variables are derived from Dirichlet distribution,
2. Dirichlet distribution parameters are obtained in case of correlated variables.

Although it was not possible to fit models for bivariate Beta variables directly (in WinBUGS), we could fit models considering these variables as bivariate Normal, after appropriate transformation.

References

[1] Johnson, N.L., and Kotz, S. (1976) Distributions in statistics: Continuous MultivariateDistributions. Wiley, New York.
[2] Catalani, M. (2002) Sample from a couple of positively correlated variates http://arxiv.org/abs/math/0209080v1