

An application of reinforced urn processes to determining maximum tolerated dose

Maura Mezzetti^a, Pietro Muliere^b, Paolo Bulla^{b,*}

^a*Dipartimento di Studi Economico-Finanziari e Metodi Quantitativi, Università di Roma Tor Vergata, Italy*

^b*Istituto di Metodi Quantitativi, Università Bocconi, viale Isonzo 25, 20136 Milano, Italy*

Received 16 October 2006; received in revised form 26 November 2006; accepted 26 November 2006

Available online 4 January 2007

Abstract

Based on reinforced urn process introduced by Muliere et al. [2000. Urn schemes and reinforced random walks. *Stochastic Process. Appl.* 88(1), 59–78] we propose a Bayesian nonparametric approach to analyse a design determining the maximum tolerated dose (MTD) in Phase I clinical trials for new drug development when inpatient dose escalation is allowed. A predictive distribution of MTD is obtained and its point estimation may consist in the corresponding expected value.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Phase I; Maximum tolerated dose; Urn scheme

1. Introduction

Reinforced urn processes (RUPs) have been introduced by Muliere et al. (2000) as random walks on a state space of Pólya urns. They are a smart tool for characterizing some well-known Bayesian nonparametric priors, but also for predictive inference. The above-cited reference provides some examples of applications of these processes to survival analysis and to multiple state processes, while Amerio et al. (2004) deal with credit default probability estimation. In this work, we present a further application to determining *maximum tolerated dose* (MTD).

Indeed, in Phase I clinical trials, researchers test a new drug or treatment in a small group of patients for the first time in order to evaluate its safety, identify side effects and determine a therapeutically useful interval of doses. The upper end of the interval is the MTD and the lower end of the interval is the *minimal effective dose* (MED). Here, we will face the estimation of the MTD; generalization to MED will be shown straightforward.

Using RUPs a predictive structure between the individuals is defined and it is possible to obtain an estimation of the distribution function and a point estimation of MTD (or MED). Both of them are easy to compute due to Pólya sampling scheme.

*Corresponding author.

E-mail address: paolo.bulla@unibocconi.it (P. Bulla).

The rest of the paper is organized as follows. Section 2 presents some traditional approach to the determination of critical doses. In Section 3, we discuss the proposed model for the trial. In Section 4 we characterize the prior distribution through the RUP and recall some results of [Muliere et al. \(2000\)](#) for the computation of posterior and predictive distribution. In Section 5, we present applications to simulated data set. We conclude the paper with a brief discussion in Section 6.

2. Traditional approaches to MTD

A well-established approach to search for MED and MTD is based on the construction of a dose–toxicity curve. [Robbins and Monro \(1951\)](#) and later [Wu \(1985\)](#) try to estimate, through stochastic approximation, the MTD as the quantile of this curve.

[O’Quigley et al. \(1990\)](#) has developed an adaptive Bayesian design called the continual reassessment method (CRM). [Eichhorn and Zacks \(1973, 1981\)](#), [Robinson \(1978\)](#) and [Shih \(1989\)](#) studied the sequential search problem through linear regression dose–toxicity models. [Muliere and Petrone \(1993\)](#) proposed a Bayesian predictive sequential method both in parametric and nonparametric version, the latter with a Dirichlet process.

[Rosenberger \(1996\)](#) suggests using the generalized Pólya urn model for a dose–response study, while [Durham et al. \(1997\)](#) faces the Phase I problem by means of a random walk rule with a biased coin up-and-down procedure; moreover, it proves that CRM performs as well without the Bayesian machinery.

Although some of these methods perform quite well in small sample dose–response study simulations, their main drawback is just requiring an explicit dose–toxicity curve, often nonparametric, that could be artificial and complicated.

Recently, [Gasparini and Eisele \(2000\)](#) propose a curvefree method: modelling the probabilities of toxicity directly as the unknown parameter of interest, a product of beta prior distributions is considered.

Unlike the above-cited works, the method we present in this paper is, as already said, a Bayesian nonparametric approach to the determination of MTD as well as the MED when inpatient dose escalation is allowed.

The parameter of interest is the distribution function of a critical dose (MTD or MED), whilst the dose–toxicity (or dose–efficacy) curve is not explicitly estimated.

This is the design of the trial looking for MTD. Ethical concerns suggest a sequential search for the MTD by giving individuals from a very low to gradually increasing dosage to be a reasonable procedure. Hence, the studies start at doses minimally toxic in animals and, at each stage, if the dose is safe, a cohort of patients are treated at a higher dose, with escalation continuing until an MTD is defined. Various schemes may be used with rapid augmentation at lower, presumably more tolerable, doses and slower when the doses get larger.

3. Design of the trial

Let us consider the determination of MTD. As in many practical situations, each subject i is given drug’s doses $\underline{X} = (x_0, x_1, \dots, x_k, \dots)$, in ascending order until the subject experiences the critical value of toxicity.

Let $Y_i(x_k)$ denote the level of toxicity at dose x_k for subject i . Indeed, as many practitioners agree, for each subject i , the only interesting events are of type $\{Y_i(x) > \eta\}$ and $\{Y_i(x) \leq \eta\}$ that is whether the toxicity goes over an unacceptable threshold η or not. Hence, for $x \in \underline{X}$, $Y_i(x)$ can be simply substituted by $N_i(x) = I_{\{Y_i(x) > \eta\}}$.

The i th subject is characterized by a sequence of doses $(x_{i0}, x_{i1}, \dots, x_{ik_i})$. As soon as the i th subject experiences a critical value for toxicity, e.g. (x_{ik_i}) , we stop to give him further doses and pass to the next one. For the $(i + 1)$ th subject we restart from initial dose, and continue as above until his critical value for toxicity is reached. Each subject generates a Markovian sequence of doses, since the decision to submit the patient to the dose x_k depends on the toxicity at the previous dose, x_{k-1} .

Considering all the patients, the overall experiment is characterized by the following sequence of doses:

$$\begin{aligned} \{X_n\}_{n \geq 1} &= \{x_{10}, x_{11}, \dots, x_{1k_1}, x_{20}, x_{21}, \dots, x_{2k_2}, x_{n0}, \dots, x_{n1}, \dots, x_{nk_n}, \dots\} \\ &= \{x_0, x_{11}, \dots, x_{1k_1}, x_0, x_{21}, \dots, x_{2k_2}, x_0, \dots, x_{n1}, \dots, x_{nk_n}, \dots\} \end{aligned} \quad (1)$$

and by the corresponding sequence of $N_i(x_k)$ that is zeros and ones indicating the status of not reached yet the toxicity (0) and reached the toxicity (1). In this way, the counting process $N(x) = \sum_{i=1}^n N_i(x)$, $x \in \underline{X}$, indicates the number of transition from the status 0 to 1 at dose x .

We assume that patients can be deemed exchangeable since they belong to the same population and the order in which they appear in the trial is not important. So, the data collected in this way turn out to be partially exchangeable in a twofold meaning. First, according to the definition proposed by [Diaconis and Freedman \(1980\)](#), that is, the subsequences of doses corresponding to each subject—the so-called x_0 blocks of $\{X_n\}$ as they have the form $\{x_0 = x_{j0}, \dots, x_{jk}\}$ —are exchangeable; hence, from their main result, under the recurrence hypothesis, the sequence $\{X_n\}_{n \geq 1}$ is a mixture of Markov chains. Second, the sequence of toxicities (or that of its indicators) exhibits a partial exchangeability *à la de Finetti* as the sequences of toxicities within each dose, i.e. $\{Y_n(x_k)\}_{n \geq 1}$ ($\{N_n(x_k)\}_{n \geq 1}$) are sequences of exchangeable random variables. We remark that the partial exchangeability scheme specifies the type of dependence between observations measured at the same dose, but does not specify anything on the nature of the stochastic dependence between observations measured at different dose levels.

The partially exchangeable sequence in (1) can be summarized by the sequence $\{\Theta_n\}_{n \geq 1} = (\Theta_1, \Theta_2, \dots, \Theta_n, \dots)$, Θ_i being the last coordinate of the i th x_0 block of the process $\{X_{ik}\}$ that is the smallest dose at which the i th subject experiments toxicity. As each Θ_i is a function of the corresponding x_0 block and these are exchangeable, $\{\Theta_n\}$ is infinitely exchangeable as well.

The distribution of the MTD will be estimated by the predictive distribution of Θ_{n+1} , given $\Theta_1, \dots, \Theta_n$, or, equivalently, the predictive distribution of $N_{n+1}(x)$ given $\{N_n(x)\}$, that is:

$$\begin{aligned} P[\Theta_{n+1} = x_k | \Theta_1, \Theta_2, \dots, \Theta_n] \\ = P[N_{n+1}(x_k) = 1 | \{N_i(x_j), j \geq 0\} | l = 1, \dots, n; N_{n+1}(x_j), j = 0, \dots, k-1]. \end{aligned}$$

Our idea is to assume a discrete beta-Stacy process with parameters $(w(x_k), b(x_k))_k$ as a prior distribution for F , the cdf of the $\Theta_i \in \underline{X}$. This idea allows us to use results from [Walker and Muliere \(1997\)](#) assuring that the posterior distribution for F is again a discrete time beta-Stacy process with jumps at $\{x_k\}$; we will compute the posterior parameters through a characterization of the beta-Stacy using the RUPs presented in the next section.

4. An urn scheme determining MTD

In this section we provide basic definitions of RUP giving theoretical foundation to the above described design. These processes are not intended to play any role in treatment allocation, but just to perform the statistical analysis of the results of the previous design. Let us start with a description of the Pólya urn.

Let U be an urn with initial composition $C = (w_0, b_0)$: that is, U contains $w_0 \geq 0$ balls of colour 0, let us say white, and U contains $b_0 \geq 0$ balls of colour 1, let us say black. The urn U is called a Pólya urn if its composition changes, when a ball is sampled, according to the following rule: every ball sampled from the urn is replaced into it along with another of the same colour (or more than one, generically m). Reinforcement is set to be $m > 0$ for every sampling. The added balls obviously *reinforce* the probability that a ball of the same colour as the one currently drawn out will be sampled in the future.

Pólya urn has a well-known Bayesian interpretation: the initial composition represents the prior guess about the occurrence of an event, the extraction reflects the outcome of the statistical experiment and the replacement of the balls is the updating of the prior knowledge. Notice that m controls the weight assigned to the data relatively to the initial opinion.

Repeating this sampling procedure over and over, an infinite sequence of random variables $\{Z_k\}$ is generated. Z_k takes value 1 if the k th ball sampled is white, 0 if it is black. By this construction, it turns out that $Z_1 \sim \text{Bernoulli}(w_0/(b_0 + w_0))$, while, for all $k \geq 1$, conditionally on N_1, \dots, N_k ,

$$Z_{k+1} = \begin{cases} 0 & \text{with probability } \frac{b_k}{b_k + w_k}, \\ 1 & \text{with probability } \frac{w_k}{b_k + w_k}, \end{cases} \quad (2)$$

where b_k is the number of black balls, w_k the number of white ones, after the extraction of the k th ball. These quantities evolve according to

$$(w_{k+1}, b_{k+1}) = \begin{cases} (w_k + m, b_k) & \text{with probability } \frac{w_k}{b_k + w_k}, \\ (w_k, b_k + m) & \text{with probability } \frac{b_k}{b_k + w_k}. \end{cases}$$

In the analysis of the Pólya urn, the interest consists in the process Z_k as well as the limit behaviour of $p_k = w_k/(w_k + b_k)$, the proportion of white balls contained in the urn.

As the sequence $\{Z_k\}$ is exchangeable, by de Finetti’s representation theorem (de Finetti, 1937), conditionally on a random variable Π , the variables Z_k are independent and identically distributed (iid) Bernoulli(Π). Moreover,

- (1) Π is a beta($w_0/m, b_0/m$);
- (2) $p_k \xrightarrow{\text{a.s.}} \Pi$, that is, the proportion of white balls converges almost surely to the random probability Π .

This classical result makes clearer our previous excursus about the relation between Pólya urn and Bayesianism.

The Pólya urn is generalized by Muliere et al. (2000) with RUPs. Let us consider a special case. Suppose to have a discrete set $x_0 < x_1 < x_2 < \dots$ and associate at each of them a Pólya urn, possibly with different initial compositions. Let us indicate with $C(x) = (w(x), b(x))$ the initial composition of the urn at dose x . A process $\{X_n\}$ is defined iteratively: set $X_0 = x_0$ and, for $n \geq 1$, if $X_{n-1} = x_i$ take a sample from the urn associated to x_i : if the ball extracted is white set $X_n = x_{i+1}$, otherwise set $X_n = x_0$.

Less formally, beginning from the urn corresponding to x_0 , balls are sequentially sampled from urns associated to x_0, x_1, x_2, \dots until a black ball is obtained. At this time, the urns are updated following the usual Pólya rule, each of them with m balls, and the sequential sampling starts once again. The sequence $\{X_n\}$ is a RUP.

If $\lim_{n \rightarrow \infty} \prod_{i=1}^n w(x_i)/(w(x_i) + b(x_i)) = 0$, the process $\{X_n\}$ is said to be recurrent, i.e. $P[X_n = 0 \text{ for infinitely many } n] = 1$.

In this case, we can define the sequence of stopping times $\{\tau_n\}$ as $\tau_0 = 0$ and $\tau_n = \inf\{n > \tau_{n-1} : X_n = 0\}$ and $\Theta_n = X_{\tau_{n-1}}$ for $n = 1, 2, \dots$. The most important result about this kind of RUP is stated in the following proposition.

Proposition 1. *If the RUP $\{X_n\}$ is recurrent, the infinite sequence $\{\Theta_n\}$ is exchangeable; that is, there exists a random distribution function F such that, conditionally on F , the random variables of the sequence $\{\Theta_n\}$ are iid with distribution F . Moreover, for $x_{i+1} \in \underline{X}$, the random mass assigned by F to the subset $\{x_0, x_1, \dots, x_{i+1}\}$ is equal to*

$$1 - \prod_{k=1}^{i+1} (1 - W_k),$$

where W_1, W_2, W_3, \dots are independent random variables and W_{i+1} has a beta distribution with parameters $w(x_i)/m$ and $b(x_i)/m$ for $i = 0, 1, 2, \dots$. That is, the law of F is that of a beta-Stacy process.

The beta-Stacy process has been defined by Walker and Muliere (1997) as a nonparametric prior for distribution functions. They proved also that the posterior distribution for F is again a discrete time beta-Stacy process with jumps at $\{x_k\}$ and parameters $(w(x_k) + m \times w_k, b(x_k) + m \times b_k)_k$ where $w_j = \sum_{i=1}^n I_{[N_i(x_j)=0]}$, $b_j = \sum_{i=1}^n I_{[N_i(x_j)=1|N_i(x_{j-1})=0]}$ and $N_i(x)$ is a counting process referring to the i th sequential sampling such that it is 0 until the x in which the black ball is sampled and 1 thereafter.

Referring to MTD search and interpreting white balls as the status *not reached yet the critical value of toxicity* and the black ones as *reached the critical value*, different urn compositions translate prior probability of experiencing a critical value of toxicity.

When the first subject is given dose x_0 , the information we have about the probability of experiencing toxicity is given simply by *a priori* information about the composition of the urn; that is, this probability will be equivalent to the probability of sampling a black ball, while the probability of “no toxicity” will be equal to that of sampling a white ball:

$$P(\Theta_1 = x_0) = P(N_1(x_0) = 1) = \frac{b(x_0)}{b(x_0) + w(x_0)}, \quad (3)$$

$$P(\Theta_1 > x_0) = P(N_1(x_0) = 0) = \frac{w(x_0)}{b(x_0) + w(x_0)}. \quad (4)$$

Following the previous scheme, when a subject not experiencing the critical value of toxicity at x_0 is observed, the composition of the urn for the next subject will be updated as $C^1(x_0) = (w(x_0) + m, b(x_0))$, whilst if a subject experiencing toxicity at x_0 is observed, the updating will be $C^1(x_0) = (w(x_0), b(x_0) + m)$ (to ease the exposition we indicate as C^i the composition updated after the observation of the i th subject).

If the first subject did not experience toxicity, he is given the successive dose and so on. Since nobody has been given those doses yet:

$$\begin{aligned} P(\Theta_1 = x_k) &= P(N_1(x_0) = \dots = N_1(x_{k-1}) = 0, N_1(x_k) = 1) \\ &= \frac{b(x_k)}{b(x_k) + w(x_k)} \prod_{j=0}^{k-1} \frac{w(x_j)}{b(x_j) + w(x_j)}. \end{aligned} \quad (5)$$

Finally, if the first subject experiences toxicity at dose x_k , it will turn out that $C^1(x_k) = (w(x_k), b(x_k) + m)$.

After the n subjects are observed, at a dose level x_k , we get for the $(n + 1)$ th subject:

$$\begin{aligned} P(\Theta_{n+1} = x_k | \Theta_1, \dots, \Theta_n) &= \frac{b(x_k) + m \times b_k}{b(x_k) + m \times b_k + w(x_k) + m \times w_k} \\ &\quad \times \prod_{j=0}^{k-1} \frac{w(x_j) + m \times w_j}{b(x_j) + m \times b_j + w(x_j) + m \times w_j} \end{aligned} \quad (6)$$

and

$$P(\Theta_{n+1} > x_k | \Theta_1, \dots, \Theta_n) = \prod_{j=0}^k \frac{w(x_j) + m \times w_j}{b(x_j) + m \times b_j + w(x_j) + m \times w_j}, \quad (7)$$

where w_j and b_j are defined as before.

Eqs. (6) and (7) specify the predictive distribution of MTD. Finally, as usual, a point estimation of the MTD is given by the expected value of the this predictive distribution,

$$\hat{\Theta}_{n+1} = \sum_k x_k P(\Theta_{n+1} = x_k | \Theta_1 = \theta_1, \dots, \Theta_n = \theta_n), \quad (8)$$

where the probabilities are given in (6).

4.1. Minimum effective dose

The same trial described above can be adapted to finding MED. Once again, each subject i is given drug's doses $\underline{z} = (z_0, z_1, \dots, z_h, \dots)$; the escalation is stopped when the subject experiments efficacy, that is, some threshold δ is overcome. Let $Z_i(z_h)$ be the random variable measuring the efficacy at dose z_h for subject i . Repeating escalation for different subjects, a partial exchangeable sequence $\{Z_n\}$ (analogous to the previous $\{X_n\}$) is obtained; this sequence can be replaced by the sequence $\{A_n\} = (A_1, A_2, \dots)$, that is, exchangeable and with a beta-Stacy process as de Finetti measure.

We consider another family of urns, independently from the previous one, with initial compositions $R(k) = (g_k, r_k)$: the new urn contains $r_k \geq 0$ balls of colour 1, corresponding to the status *reached the efficacy level*, and $g_k \geq 0$ balls of colour 0, corresponding to the status *not reached the efficacy level yet*. The initial

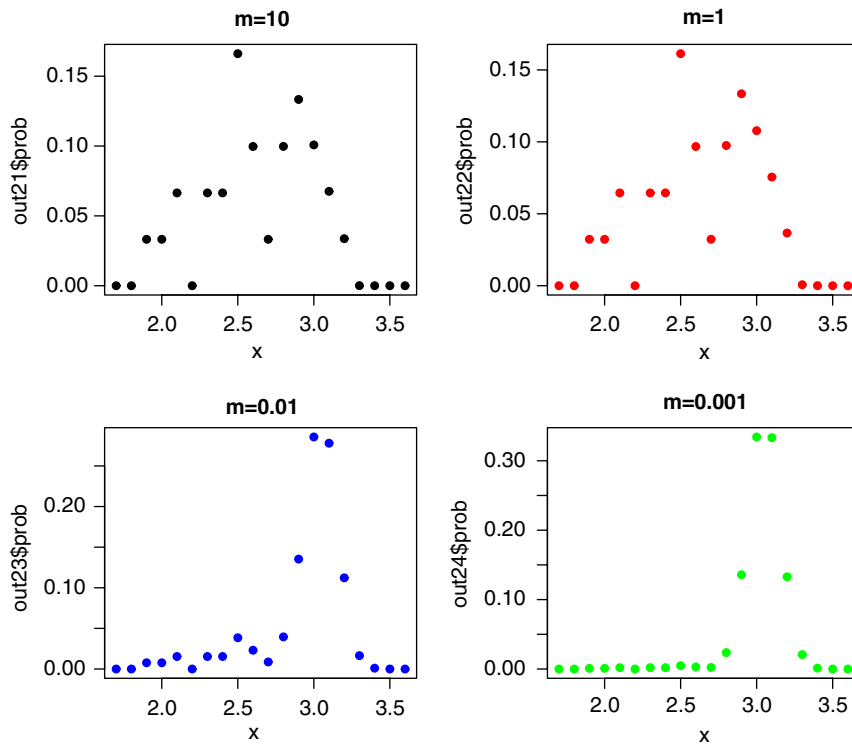


Fig. 1. Predictive distribution of MTD.

compositions of the urns, different at each dose level, correspond to the prior probability of experiencing efficacy. The posterior distribution is again a discrete time beta-Stacy process with jumps at $\{z_h\}$ and parameters can be computed as shown in the MTD case.

5. Simulation

We simulate a scenario allowing us to verify the new procedure to estimate the MTD. Each subject may be given drug's doses from $x = 1.6$ to 3.6 , regularly spaced with step 0.1 . We generate data supposing the real distribution of the MTD to be the discretization on the doses values of an $N(2.8, 0.2)$. A sample of 20 patients is considered, the predictive distribution of MTD is easily computed by (6) and then a point estimation is obtained.

According to Bayesian inference, we need to specify the prior parameters, that in this case are the values for the number of white and black balls at each dose level, (prior probability to experience toxicity), and fix the value for the reinforcement in the urn scheme m .¹ Formal indications about how to elicit the prior can be found in Walker and Muliere (1997): briefly, if the prior guess on MTD is summarized by a distribution $G(k) = P[\Theta = x_k] \forall k$, put $b(x_k) = c_k G(k)$, $w(x_k) = c_k(1 - \sum_{j=0}^k G(j))$ where the c_k 's play the role of precision of prior. Higher values of m with respect to c_k 's involve higher weight of the guess in posterior distribution, while relatively lower values imply higher weight of observations.

We choose to centre the prior on the discretization of an $N(3.1, 0.1)$, putting $c_k = 1, \forall k$.

Fig. 1 shows the predictive distribution of MTD obtained for four different values of m (10,1,0.01 and 0.001). According to the above discussion, the first case reproduces the empirical probability function of the MTD, while the last mimics the prior. The others melt this two opposites, with weight depending on m .

¹Note that, even though the idea of balls is very evocative, from a mathematical point of view the parameters $b(x_k)$, $w(x_k)$ and m must not necessarily be integer numbers, but just positive real.

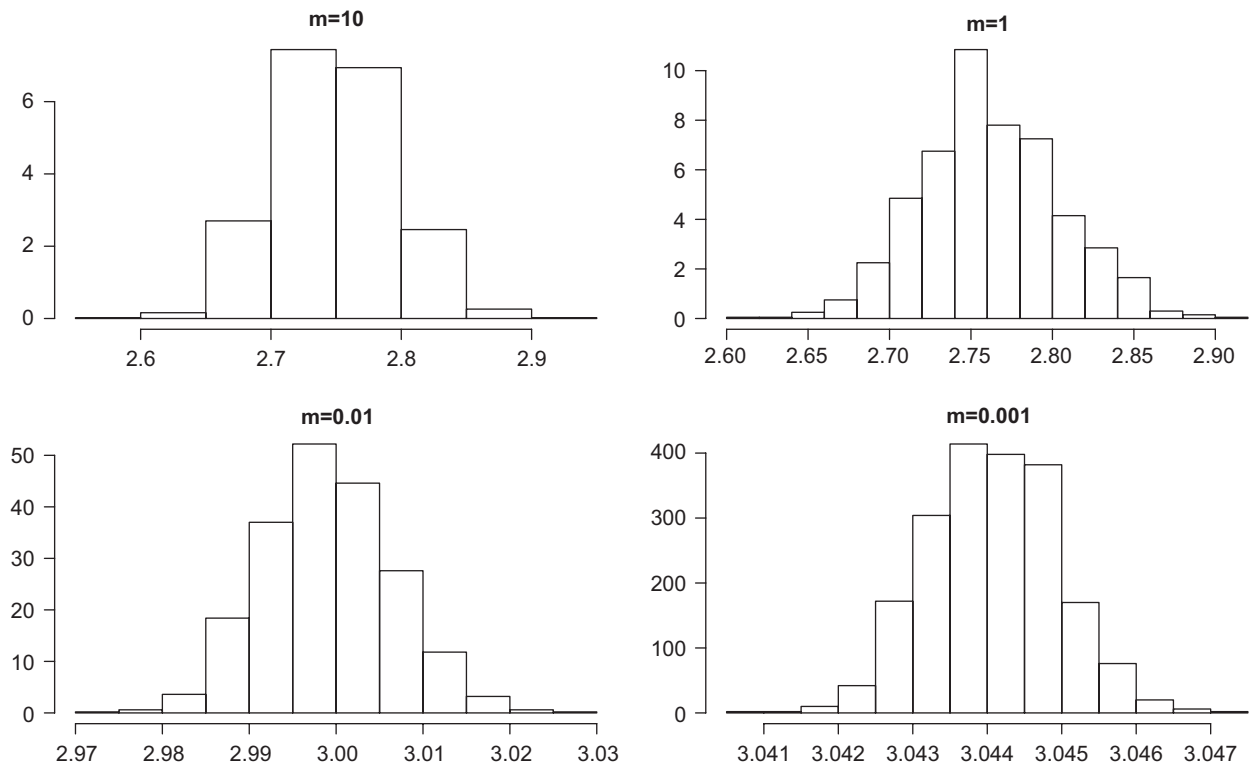


Fig. 2. MTD estimation by expected value.

In order to show that the method we propose does not require huge samples to perform well, a thousand of replications of a rather small sample of 20 patients are generated. The result is quite good.

Fig. 2 displays the histograms for the estimated MTD. As we expect, the lower the reinforcement with respect to the number of initial balls, the closer the posterior estimation to the prior values, 3.1.

6. Conclusions

We have proposed a simple design for determining MTD as well MED: subjects are given increasing drug's doses until toxicity is shown.

A particular case of RUPs provides, in a Bayesian nonparametric framework, theoretical support to this scheme and enlightens how a beta-Stacy process is a natural choice as prior. Moreover, the urn approach makes easy the computation of the updated parameters.

As shown in Muliere et al. (2000) more general classes of nonparametric priors as neutral to the right or tailfree can be characterized by RUP, hence it should be possible to derive from these processes alternative trial design adapted to possibly different situations.

References

- Amerio, E., Muliere, P., Secchi, P., 2004. Reinforced urn processes for modelling credit default distribution. *Internat. J. of Theoret. and Appl. Finance* 7 (4).
- de Finetti, B., 1937. La prévision, ses lois logiques, ses sources subjectives. *Ann. Inst. Poincaré* 7, 1–68.
- Diaconis, P., Freedman, D., 1980. De Finetti's theorem for Markov chains. *Ann. Probab.* 8, 115–130.
- Durham, S.D., Flournoy, N., Rosenberger, W.F., 1997. A random walk rule for phase I clinical trials. *Biometrics* 53, 745–760.
- Eichhorn, B.H., Zacks, S., 1973. Sequential search of an optimal dosage. *I. J. Amer. Statist. Assoc.* 68, 594–598.
- Eichhorn, B.H., Zacks, S., 1981. Bayes sequential search of an optimal dosage: linear regression with both parameters unknown. *Comm. Statist. A—Theory Methods* 10 (10), 931–953.

- Gasparini, M., Eisele, J., 2000. A curve-free method for phase I clinical trials. *Biometrics* 56 (2), 609–615.
- Muliere, P., Petrone, S., 1993. A Bayesian predictive approach to sequential search for an optimal dose: parametric and nonparametric models. *J. Italian Statist. Soc.* 2, 349–364.
- Muliere, P., Secchi, P., Walker, S.G., 2000. Urn schemes and reinforced random walks. *Stochastic Process. Appl.* 88 (1), 59–78.
- O’Quigley, J., Pepe, M., Fisher, L., 1990. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 46 (1), 33–48.
- Robbins, H., Monro, S., 1951. A stochastic approximation method. *Ann. Math. Statist.* 22, 400–407.
- Robinson, J.A., 1978. Sequential choice of an optimal dose: a prediction intervals approach. *Biometrika* 65, 75–78.
- Rosenberger, W.F., 1996. New directions in adaptive designs. *Statist. Sci.* 11 (2), 137–149.
- Shih, W.J., 1989. Prediction approaches to sequentially searching for an optimal dose. *Biometrics* 45, 623–628.
- Walker, S., Muliere, P., 1997. Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* 25 (4), 1762–1780.
- Wu, C.-F.J., 1985. Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* 80 (392), 974–984.