# On Stable Simulation of Random Functions over Fixed Boundaries using Biased Beta Distributions [1]

**Oscar G. Gonzalez** [2]

---

Straight standard classical geostatistics based on symmetric, unbounded *Gaussian distributions*, can fail to model physical processes whose outcomes must be restricted to a finite interval $(a, b) \subset \Re$. This paper discusses a numerically stable, approximated approach to *simulate non stationary random functions* over finite supports.

---

**KEY WORDS:** random function, beta distribution, simulation, geostatistics.

## The Problem

Consider a *non stationary random function* (Journel, 1978) $T = T(x, y, z) : \Re^3 \to (a, b) \subset \Re$ with an unknown *probability density function* $f_{T(x,y,z)}(t)$, positive over a known given support $-\infty < a(x, y, z) < t < b(x, y, z) < \infty$, and zero elsewhere, with:

$$\mu_t = \mathrm{E}(T) = f_E(x, y, z) : \Re^3 \to (a, b) \subset \Re \quad (1)$$

$$\sigma_t^2 = \mathrm{Var}(T) = f_V(x, y, z) : \Re^3 \to \Re^+ \quad (2)$$

being known functions, and an unknown covariance function:

$$\mathrm{Cov}\big(P_i(x_i, y_i, z_i), P_j(x_j, y_j, z_j)\big) : \Re^3 \times \Re^3 \to \Re \quad (3)$$

It is needed to obtain approximated, numerically stable *stochastic simulations* of $T$. The method will be useful to simulate *conditional distributions* over a geostatistical framework, and to simulate functions of random functions.

---

[1] Confidential Manuscript of December 2003, publicly released on August 2013.

[2] GeolOil Corporation, www.geoloil.com; e-mail: oscar@geoloil.com

## The Beta Distribution

Suppose that the *Beta distribution* is a good approximation for the unknown density $f_T(t)$, for which it is known $\mathrm{E}(T) = \mu_t = f_E(x,y,z)$, $\mathrm{Var}(T) = \sigma_t^2 = f_V(x,y,z)$, and the boundaries $a = a(x,y,z)$, and $b = b(x,y,z)$

The Beta density is:

$$f_T(t) = \frac{1}{B(q,r)} \frac{(t-a)^{q-1}(b-t)^{r-1}}{(b-a)^{q+r+1}} I_{(a,b)}(t); \quad \begin{array}{c} a < t < b, \\ 0 < q < \infty, \\ 0 < r < \infty \end{array} \tag{4}$$

Where $I_{\mathcal{A}}(x)$ is the indicator function: $I_{\mathcal{A}}(x) = 1$, if $x \in \mathcal{A}$, zero otherwise; and $B(q,r) = \int_0^1 x^{q-1}(1-x)^{r-1}dx$, is the *Beta function*. Random deviates at a fixed $P(x,y,z)$ location can be generated by: (Ang, A. and Tang, W., 1984)

$$t = a + (b-a)\left(\frac{u_1^{(1/q)}}{u_1^{(1/q)} + u_2^{(1/r)}}\right) \tag{5}$$

where $u_1$ and $u_2$ are drawn from independent uniform $(0,1)$ distributions.

Moments of the Beta distribution are:

$$\mu_t = \mathrm{E}(T) = a + (b-a)\left(\frac{q}{q+r}\right) \quad \in (a,b) \tag{6}$$

$$\sigma_t^2 = \mathrm{Var}(T) = (b-a)^2 \left[\frac{q\,r}{(q+r)^2\,(q+r+1)}\right] \in (0, [(b-a)/2]^2) \tag{7}$$

Hence, *method of moments* estimators for $q$ and $r$, are:

$$q = \frac{\mu}{\sigma^2}\left[\mu - \mu^2 - \sigma^2\right] \tag{8}$$

$$r = \frac{1-\mu}{\sigma^2}\left[\mu - \mu^2 - \sigma^2\right] \tag{9}$$

where $\mu$ and $\sigma^2$ are respectively, the expectation and variance of the transformed variable:

$$T^\dagger = \frac{T - a}{b - a} \in (0, 1) \tag{10}$$

for which:

$$\mu = \mu^\dagger = \mathrm{E}(T^\dagger) = \frac{\mu_t - a}{b - a} \ \in (0, 1) \tag{11}$$

$$\sigma^2 = {\sigma^\dagger}^2 = \mathrm{Var}(T^\dagger) = \left( \frac{\sigma_t}{b - a} \right)^2 \in (0, 0.5^2) \tag{12}$$

## A Biased Approximation

No all conceivable values of $\mu$ and $\sigma$ are admissible to yield positive values of $q$ and $r$. That is, from equations (8) and (9) , $\mu$ and $\sigma$ have to satisfy:

$$\mu - \mu^2 - \sigma^2 > 0 \tag{13}$$

or

$$\sigma < \sqrt{\mu\,(1 - \mu)} \tag{14}$$

Therefore, the maximum admissible value $\sigma_{\max}$ for $\sigma$ is:

$$\sigma_{\max} = \max_{0<\mu<1} \left\{ \sqrt{\mu\,(1 - \mu)} \right\} = 0.5 \tag{15}$$

Nevertheless, for practical numerical simulations, the restriction $\mu - \mu^2 - \sigma^2 > 0$, can yield values of $q$ or $r$ very close to 0. This can produce numerical stability problems in drawing outcomes from the beta distribution, as it can require a very large sample size $n$ to get good moments estimations by:

$$\widetilde{\mathrm{E}}(T^\dagger) = \bar{T}^\dagger = \frac{1}{n} \sum_{i=1}^{n} T_i^\dagger \tag{16}$$

$$\widetilde{\sigma^2}(T^\dagger) = {S^\dagger}^2 = \frac{1}{n - 1} \sum_{i=1} \left( T_i^\dagger - \widetilde{\mathrm{E}}(T^\dagger) \right)^2 \tag{17}$$

To overcome the numerical instability problem due to the closeness of $q$ or $r$ to 0, it will be introduced explicitly some *bias* by slight shifting $q$ and $r$ over some fixed threshold $\delta > 0$. That is:

$$q = \frac{\mu}{\sigma^2}\left[\mu - \mu^2 - \sigma^2\right] \geq \delta > 0 \tag{18}$$

$$r = \frac{1-\mu}{\sigma^2}\left[\mu - \mu^2 - \sigma^2\right] \geq \delta > 0 \tag{19}$$

Hence $q$ and $r$ have to satisfy:

$$q : \mu^3 - \mu^2 + \mu\,\sigma^2 + \delta\,\sigma^2 \leq 0 \tag{20}$$

$$r : \mu^3 - 2\mu^2 + \mu\,(\sigma^2 + 1) - \sigma^2\,(1 + \delta) \geq 0 \tag{21}$$

Inequalities (20) and (21) are more restrictive than (13) . These equations are algebraically equivalent, if in equation (21) , $\mu$ is replaced by $1-\mu'$. That is, it is enough to study the equality boundary behavior in the expression:

$$g(u, \sigma) \stackrel{\text{def}}{=} u^3 - u^2 + u\,\sigma^2 + \delta\,\sigma^2 = 0 \tag{22}$$

## Three Practical Approaches to Biasdness

Given the known functions $\mu_t = f_E(x, y, z)$ and $\sigma_t^2 = f_V(x, y, z)$, at a sampled set:

$$\Omega = \Big\{P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2), \ldots, P_n(x_n, y_n, z_n)\Big\} \subset \Re^3 \tag{23}$$

it is required to asses if for each point $P_i(x_i, y_i, z_i) \quad i = 1, 2, \ldots, n$ the parameters $\mu_i(P_i)$ and $\sigma_i^2(P_i)$ satisfies expressions (20) and (21) . If not, the original parameters $\mu_i$ and $\sigma_i^2$ can be biased to values $\mu_i^*$ and ${\sigma^*}_i^2$, in order to satisfy the inequalities. There are three approaches to fulfill the requirement:

1.- If major interest on an applied study, is to honor the uncertainty, then it can be held: $\sigma^* = \sigma$ fixed, and allow some shifting bias $\mu^* \neq \mu$.

2.- If major interest on an applied study, is to honor the exactitude, then it can be held: $\mu^* = \mu$ fixed, and allow some shifting bias $\sigma^* < \sigma$

3.- If it is desired not to shift very much $\mu$ and $\sigma$, then a hybrid approach can allow some shifting bias to both parameters $\mu^* \neq \mu$ and $\sigma^* < \sigma$

## Case I: Honoring Uncertainty

The equality bound of equation (22) can be studied using the *Cardan's formula* (Condon, 1967) for the third degree polynomial in $u$, by identifying:

$$a_o x^3 + a_1 x^2 + a_2 x + a_3 = 0 \tag{24}$$

$$a_o = 1; \quad a_1 = -1; \quad a_2 = \sigma^2; \quad a_3 = \delta\,\sigma^2; \quad \text{and} \quad x = u \tag{25}$$

By putting:

$$x = y - a_1/(3a_o) \tag{26}$$

equation (24) is transformed into

$$y^3 + ay + b = 0 \tag{27}$$

If it is set:

$$D = (b^2/4) + (a^3/27) \tag{28}$$

and

$$\cos(3\phi) = -(b/2)/\sqrt{-a^3/27} \tag{29}$$

then, in the case where $a < 0$ and $D < 0$, the roots are:

$$y_1 = 2\sqrt{-a/3}\ \cos(\phi) \tag{30}$$

$$y_2 = (-y_1/2) + \sqrt{-a}\ \sin(\phi) \tag{31}$$

$$y_3 = (-y_1/2) - \sqrt{-a}\ \sin(\phi) \tag{32}$$

Replacing (25) into (24) , and setting $x = y - a_1/(3a_o)$:

$$y^3 + \underbrace{\left(\sigma^2 - \frac{1}{3}\right)}_{a<0} \, y \; + \; \underbrace{\left[\sigma^2 \left(\delta + \frac{1}{3}\right) - \frac{2}{27}\right]}_{b} = 0 \tag{33}$$

Applying now (28) :

$$D = \frac{1}{4}\left[\sigma^2 \left(\delta + \frac{1}{3}\right) - \frac{2}{27}\right]^2 + \frac{1}{27}\left(\sigma^2 - \frac{1}{3}\right)^3 \; = \; D(\sigma, \delta) \tag{34}$$

Values for $D < 0$, depends upon values on $\sigma$ and $\delta$. Once fixed the known $\sigma$, a maximum admissible value for $\delta$, is found by setting $D = 0$ and arranging equation (34) in terms of a second order polynomial in $\delta$. That is:

$$\left(\frac{1}{4}\,\sigma^4\right)\delta^2 \; + \left(\frac{1}{6}\,\sigma^4 - \frac{1}{27}\,\sigma^2\right)\delta \, + \\ \left(\frac{1}{36}\,\sigma^4 - \frac{1}{81}\,\sigma^2 + \frac{1}{729} + \frac{1}{27}\left[\sigma^2 - \frac{1}{3}\right]^3\right) \; = \; 0 \tag{35}$$

Therefore:

$$\delta \;=\; \left(\frac{2}{27\,\sigma^2} - \frac{1}{3}\right) \;\pm\; \frac{2}{\sigma^2}\left(\frac{1}{3}\left[\frac{1}{3} - \sigma^2\right]\right)^{3/2} \tag{36}$$

for which only the positive root is admissible, then:

$$\delta_{\max} = \frac{2}{27\,\sigma^2} - \frac{1}{3} + \frac{2}{\sigma^2}\left(\frac{1}{3}\left[\frac{1}{3} - \sigma^2\right]\right)^{3/2} \;=\; \delta_{\max}(\sigma) \tag{37}$$

Hence, given $\sigma$, any selection of the threshold value $\delta$ in the interval

$$\delta \in (0, \delta_{\max}) \subset \Re^+ \tag{38}$$

warranties $D < 0$. Among the three Cardan's roots $y_1$, $y_2$, and $y_3$, only the root $y_2$ is admissible.

Now, given $\mu$ and $\sigma$, plain values of $q$ and $r$ are computed using the moment methods by equations (8) and (9) . If $q, r > \delta$, then is not necessary to follow any shifting procedure, since the beta simulated outcomes would be stable.

The following procedure is proposed to enough improve numerical stability for those cases when $q < \delta$, $r < \delta$, or $q, r < \delta$:

If $q < \delta$ and $q \leq r$; bias $\mu$ to the corresponding cardano root $\mu^* = u_2$ related to $y_2$. The new pair of shifted moments $(\mu^*, \sigma)$ will produce slightly shifted parameters $(q^* > \delta, r^*)$ to a biased beta distribution, which should produce more stable outcomes.

Symmetrically, if $r < \delta$ and $r < q$; bias $\mu$ to $\mu^* = 1 - u_2$. The new pair of shifted moments will produce slightly shifted parameters $(q^*, r^* > \delta)$.

After studying some numerical sensibility analysis, the following heuristic expression yielded reasonable stable outcomes for the biased beta distribution:

$$\delta \;=\; (\delta_{\max} - \epsilon)\; I_{[\delta_{\max},\infty)}(\lambda) \;+\; \lambda\; I_{(0,\delta_{\max})}(\lambda) \tag{39}$$

Where $\epsilon > 0$ is a small machine dependent value to ensure a value of $\delta$ slightly smaller than $\delta_{\max}$. Values of $\lambda \cong 1.5$, yield reasonable stable results when drawing biased outcomes for the beta distribution for most cases. Values of $\lambda \cong 0^+$ can yield unsatisfactory very high sample variances on $\widetilde{\sigma_T^2}$, while values of $\lambda > 1.5$ can introduce unnecessary higher bias on $q$ and $r$.

## Case II: Honoring Exactitude

In this case, the expectation value $\mu^* = \mu$ will be held fixed, while the variance ${\sigma^*}^2 < \sigma^2$ will be shifted. The inequalities in the expressions (20) and (21) can be rewritten as:

$$q : {\sigma^*}^2 \leq \frac{\mu^2(1-\mu)}{\mu+\delta} < \sigma^2 \tag{40}$$

$$r : {\sigma^*}^2 \leq \frac{\mu\,(2\mu-\mu^2-1)}{\mu-\delta-1} < \sigma^2 \tag{41}$$

If it is defined:

$$\tau^2 = \tau^2(\mu,\delta) \stackrel{\text{def}}{=} \min\left\{\frac{\mu^2(1-\mu)}{\mu+\delta}, \frac{\mu\,(2\mu-\mu^2-1)}{\mu-\delta-1}\right\} < \sigma^2 \tag{42}$$

Then, any selection of

$${\sigma^*}^2 \in (0,\tau^2) \subset \Re^+ \tag{43}$$

will warranty numerically stable beta outcomes, since this procedure yield $q^*, r^* > \delta$.

## Case III: Hybrid Approach

Although the two former cases would improve numerical stability, they can create for some instances, biases greater than desired. A hybrid approach would allow slight bias in both values $\mu^*$ and $\sigma^*$. The following procedure is proposed:

1.- Choose a fraction $\nu \in (0,1)$ to bias the standard deviation. Then, bias the standard deviation to:

$$\sigma^* = \tau + \nu\,(\sigma-\tau) \;\in\; (\tau,\sigma) \tag{44}$$

2.- Bias the expectation $\mu^*$ accordingly to the procedure of case I.

## Geostatistical Application: Properties Summaries

Suppose that for a set of $k$ independent random functions,

$$\{T_i = T_i(x,y,z)\}_{i=1}^{k} \quad \mathrm{Cov}(T_i(P_u); T_j(P_v)) = 0 \ \forall i \neq j, \text{ and } \forall P_u, P_v \tag{45}$$

their expectation and variances functions are known.

$$\mu_i = \mu_i(x,y,z) \tag{46}$$

$$\sigma_i^2 = \sigma_i^2(x,y,z) \tag{47}$$

In practical applications, under some approximated geostatistical assumptions, (which usually are not compatible with the non stationarity approach), given a set of sampled points, heuristic *conditional means and variances* can be roughly approximated by the use of *estimation techniques*, like *kriging*. (Cressie, 1993). Then, the a-priory physical knowledge of the boundary functions $a(x,y,z)$ and $b(x,y,z)$ is incorporated to fully approximate the distribution at a fixed desired location.

Typical summaries properties computations, like means, variances, quantiles, and probabilities at a fixed point $P(x,y,z)$, does not require the use of a *covariance function* for a random function $T$, but only the density $f_{T(x,y,z)}(t)$. (Nevertheless, in spite of theoretical incompatibilities, a stationary variogram had to be fitted to the set of sampled points to compute the conditional expectations and variances). Then, at a fixed point $P(x,y,z)$, any function $f(\bullet) : \Re^k \to \Re$,

$$T_f = T_f(x,y,z) = f\big(T_1(x,y,z), T_2(x,y,z), \ldots, T_k(x,y,z)\big) \tag{48}$$

can be studied by drawing $n$ outcomes of the random vector:

$$\big(T_1(x,y,z), T_2(x,y,z), \ldots, T_k(x,y,z)\big) \tag{49}$$

and then, build the random outcome set:

$$\Omega(x,y,z) = \left\{ T_{f\,i} \right\}_{i=1}^{n}$$

to fast compute summaries like expectations, variances, quantiles, and probabilities.

## Geostatistical Application: Realizations of a Random Function

Given a set of sampled points:

$$\Omega_o = \Big\{ t_1(x_1, y_1, z_1), t_2(x_2, y_2, z_2), \ldots, t_n(x_n, y_n, z_n) \Big\} \tag{50}$$

It is desired to simulate conditional realizations of the random function $T(x, y, z)$.

A heuristic algorithm, parallel to *Sequential Gaussian Simulation* (SGS), (Deutsch, 1998), can be follow to define an approximated *Sequential Biased Beta Simulation* (SBBS):

1.- Define a random path that visits each node of a desired grid, out of $\Omega_o$.

2.- Draw the first point of the random path, $P_{\alpha_1}(x_{\alpha_1}, y_{\alpha_1}, z_{\alpha_1})$

3.- Using the set $\Omega_o$, follow the approximated procedure of the former section to draw a value $T_{\alpha_1}(x_{\alpha_1}, y_{\alpha_1}, z_{\alpha_1})$

4.- Update the set of the sampled points $\Omega_o$, with the set of simulated points: $\Omega_1 = \Omega_o \cup \{T_{\alpha_1}(x_{\alpha_1}, y_{\alpha_1}, z_{\alpha_1})\}$

5.- Draw the next location $P_{\alpha_2}$ of the random path. Using now the set $\Omega_1$, draw a value $T_{\alpha_2}(x_{\alpha_2}, y_{\alpha_2}, z_{\alpha_2})$

6.- Update the set $\Omega_2 = \Omega_1 \cup \{T_{\alpha_2}(x_{\alpha_2}, y_{\alpha_2}, z_{\alpha_2})\}$

7.- Sequentially repeat the procedure until all locations of the random path are visited and simulated.

Although in some cases, SGS and SBBS might yield similar results, only the SBBS takes into account variable user defined physical boundaries $a(x, y, z)$ and $b(x, y, z)$.

## CONCLUSIONS

A numerically stable, approximated approach to simulate stationary random functions over variable finite supports, was discussed. The procedure enables fast computing of properties summaries like means, variances, quantiles and probabilities at fixed locations, and a heuristic conditional realizations procedure have been proposed.

## REFERENCES

Ang, A., and Tang, W., 1984, *Probability Concepts in Engineering Planning and Design, Volume II-Decision, Risk and Reliability.*
Wiley, New York, p. 285-286.

Cressie, Noel A. C, 1993, *Statistics for Spatial Data.*
Wiley, New York, p. 119-134.

Condon, E. U., and Odishaw, Hugh, 1967, *Handbook of Physics, Second edition.*
McGraw-Hill, New York, p. 1-16.

Deutsch C., and Journel A., 1998, *GSLIB: Geostatistical Software Library and User's Guide.*
Oxford University Press, New York, p. 125-146.

Jounel A., and Huijbregts Ch., 1987, *Mining Geostatistics.*
Academic Press, London, p. 30-35.