La densité, g(n), affectée aux entiers, n, doit être sommable à l'unité, , ce qui implique qu'elle décroisse plus vite que 1/n. Poser g(n)=1/ne où e est proche de 1+ me semble un brin artificiel quand une solution plus satisfaisante existe.

En fait, il faut garantir que la convergence soit la plus lente possible sans quoi on pourra toujours suspecter que l'on a introduit un biais souterrain favorisant les petits entiers. Il se fait que la limite entre divergence et convergence est connue depuis longtemps : Chaitin s'en est fait l'écho dans l'un de ses travaux relatif aux nombres Oméga où il était confronté à un problème largement apparenté "Comment tirer au sort un programme préfixe binaire afin d'alimenter une machine universelle de Turing ?".

Le critère est le suivant : il faut que g(n) soit asymptotique à

C'est la conséquence de la divergence des séries :

http://www.physinfo.org/annexes/entier/diverge.jpg

et de la convergence des séries à peine modifiées :

http://www.physinfo.org/annexes/entier/converge.jpg

Il peut paraître ardu de construire une fonction possédant cette propriété asymptotique, à la frontière de la convergence et de la divergence, mais c'est, au contraire, chose aisée. Voici la méthode que j'utilise (d'autres, asymptotiquement équivalentes, sont certainement possibles) : elle est basée sur un codage préfixe asymptotiquement optimal des entiers positifs.

L'ensemble des entiers binaires positifs est en correspondance parfaite avec l'ensemble des suites binaires, il suffit d'ôter le chiffre 1 initial, non significatif.



Un codage préfixe asymptotiquement optimal des suites amputées (donc des entiers positifs) s'obtient comme suit : on préfixe itérativement la suite amputée par le binaire qui code sa longueur (lui-même amputé de son 1 initial) jusqu'à tomber sur une étiquette 2=10b ou 3=11b (ce qui ne peut manquer de se produire) que l'on n'ampute pas. On préfixe le résultat par autant de 0 que d'étiquettes rencontrées. Les suites {}, {0} et {1} sont codées séparément à la main sous la forme préfixe 10, 110, et 111.

Le résultat des calculs est fourni en annexe dans deux formats distincts (Mathematica et HTML). Le codage optimal et la densité {n, code(sn), g(n)} des premiers entiers valent :

Length(code) = {2, 3, 3, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 10, 10, …}

et

Ce codage étant préfixe et complet, sa somme de Kraft vaut 1 ce que les calculs en annexe vérifient. De plus g(n) possède le comportement asymptotique requis. Le codage peut paraître coûteux aux petites valeurs de n mais il se rattrape aux grandes valeurs, en fait il est asymptotiquement optimal et sans biais.