A Comparison of Mersenne Twister and Linear Congruential Random Number Generators

Josh Albrecht

April 16, 2007

I. Introduction

Random numbers have a large number of uses, especially in computer programs. However, generating truly random numbers is difficult, if not impossible, and thus computer programs must settle for “pseudo-random” numbers, which are series of numbers that appear random, yet are actually generated by a deterministic algorithm. This raises a number of questions. What does it mean to appear random? What are the most important properties of random number generators (RNGs)?

One of the primary applications of RNGs is for Monte Carlo simulations. The needs of this application demonstrate clearly some of the most important demands of RNGs. First, the series of pseudo random numbers (hereafter referred to simply as random numbers) must not repeat itself too quickly, or it will not be random enough for use in certain simulations. Second, the series must be easily computed so that the RNG does not take too many resources away from the actual simulation. Finally, the series must be sufficiently random that patterns do not affect the application.

This last requirement will be the focus of the paper. Unfortunately, since randomness cannot be theoretically guaranteed, RNGs must be tested empirically to disprove the existence of specific patterns. A number of such tests exist, including the Chi-Square/Serial frequency and Kolmogorov-Smirnov tests for uniformity, runs up and down tests for independence, the autocorrelation test, and many more [1].

There are a number of different RNGs that satisfy the previous requirements, including some additive generators, some linear congruential generators, and those using the Mersenne Twister algorithm. This paper will compare the later for with the standard Java RNG.

II. Formulation of the Model/Motivation

The Mersenne Twister (MT) RNG has a number of advantages over the Java RNG, namely, that it has an extremely large period, has better equidistribution properties, and is nearly as efficient to compute [2]. It is also more theoretically interesting than the Java RNG, which is an example of a simple linear congruential generator much like those that we have studied in class [3]. The MT RNG is much too complicated to explain in full detail here, though there are a number of papers and articles describing it in varying levels of detail, which are listed in Section V. Basically, the MT algorithm through a series of bitwise XOR, AND, and shifting operations that are mathematically equivalent to operations on a twist matrix. The bitwise operations are very easy to compute, making this an ideal choice for an RNG.

While the equidistribution for both the MT and Java RNGs has been theoretically demonstrated in other works [2, 4], this paper will attempt to empirically determine the effect of this property. Testing for equidistribution can be done by the Serial Test, a generalization of the Chi-Square test to multiple dimensions. This test works by creating a number of "bins" into which randomly generated tuples are placed. The empirical count in each bin is compared to the expected value with the Chi-Square test to determine if it is significantly different.

Of course, other properties must also be tested to ensure that a given RNG is sufficiently random. A series of tests, called the Diehard Tests, have been developed to test these generators [5], and the results of running them will be given for both RNGs, along with a brief description of some of the tests.

III. Analysis: 2-4 pages, includes any programs

To test the randomness of the Java and MT RNGs, a number of experiments were performed. First, the Serial Test was run on each generator, for tuples of size 1 through 4. Each experiment created 20 bins per dimension, and generated enough values for the expected value of counts in each bin to be 10. Tuples of a size greater than 4 were unfortunately not possible given the constraints of the computational resources available. Each test was run 500 times with an alpha of 0.1. The table below shows the percent of tests that failed. The expected percent of failures for a perfect generator would be 10%.

Tuple Size:	1	2	3	4
Mersenne Twister:	8%	9%	9%	12%
Java Standard:	7%	10%	10%	11%

The above results fail to show any problems with either algorithm that cannot be explained by statistical noise. Because of this, the Diehard tests were also run for both generators. These tests are designed specifically to test the randomness of RNGs, and include such simulations as a large number of simulated craps games, a test of overlapping sums, minimum distance, and many more. The test was run on 10 files (each of 2750000 random 32-bit integers) for both algorithms. Each algorithm showed reasonable p-values for all tests with the exception of the Parking Lot test, which consistently failed for both RNGs. I believe that this particular test is broken in the current implementation however, as truly random data from [6] also failed this particular test.

These tests also fail to show any problems with either generator, leading to the formulation of a test specifically designed to show the equidistribution problem with linear congruential generators. As noted in this paper [7.], the maximum number of hyperplanes upon which the pseudo-randomly generated numbers lie is $(n!*m)$ ^ $(1/n)$ where m is the modulus of the congruential generator and n is the dimensionality of the space (the size of the tuple in the case of the Serial Test). So for example, since m=(2^48)-1 for the Java RNG, there would be at most 126 hyperplanes containing all 10-tuples generated by the Java RNG. With the naïve Serial Test above, this would require b^n total bins, where b is the number of bins per dimension, to take advantage of this fact and show a lack of equidistribution, and $(e*b)$^n total random values generated, where e is the expected value per bin.

Even for small values of e and b, and reasonable values of n, b^n is too large to fit in the memory of a normal computer, and $(e*b)$^n is too large to compute in a reasonable amount of time, even with the MT algorithm generating almost 100 million values per second. If the test were altered such that only the bins along a single line were considered, the number of bins could be reduced to simply b. In this case, $(e*b)$^n values must still be generated, but e can be < 1. If b is sufficiently large that bins are guaranteed to fall between the hyperplanes (since there are a limited number and they are at worst equally spaced), the number of bins with a count of 0 should be statistically significantly higher.

Unfortunately, this approach is still computationally infeasible, and the attempt to empirically prove equidistributional problem with the Java RNG is thus concluded.

Instead, we turn now to the theoretical approach to the problem. The following graphs illustrate the nature of the problem:

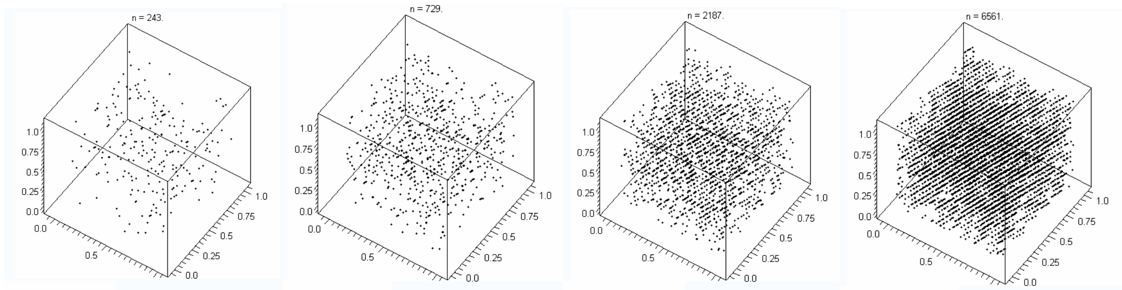

Figure 1: License for these images is included in Appendix B.

As can be seen above, a pattern clearly emerges as the number of points generated increases. The existence of any pattern in a RNG is clear evidence that it is not truly random. As demonstrated in [7], these patterns become much more severe as the dimension of the space increases for linear congruential generators, which would correspond to a decreased number of planes in the above graphs. The same problem exists for the MT RNG, but it is not until the dimension of the space is well over 600 that the problems become apparent, which is a significant improvement. Another known problem with both approaches is that the sequence generated begins to repeat after a certain number of numbers are generated. Again, however, the MT algorithm is superior in this respect, having a period of almost 2^132049 as compared to the period of 2^48 for the Java RNG.

The code for the program I created to perform the Serial Test is attached in Appendix A. Links to the other code used in this project are listed in Section V.

IV. Discussion:

Two methods for pseudo-random number generation, the standard Java linear congruential generator and the Mersenne Twister generator, have been described, tested, and compared. Though we were unable to show the shortcomings of the Java RNG empirically with the limited computing resources available, they are significant shortcomings that must be considered. We have seen that the Mersenne-Twister algorithm is superior in almost every respect, and must conclude that, when possible, it should be preferred over simpler linear congruential generators such as the Java standard.

V. References

1. http://www.cs.pitt.edu/~ramirez/cs1538/cs1538.ppt. Class slides about random numbers and tests for randomness.

2. http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html. The website of the authors of the MT algorithm. Contains original papers describing it, articles summarizing, links to implementations.

3. http://java.sun.com/j2se/1.4.2/docs/api/java/util/Random.html. The Java documentation for the Random class.

4. http://www.math.utah.edu/~beebe/java/random/README. A math professor from Utah discussing the Java RNG.

5. http://www.stat.fsu.edu/pub/diehard/. The Diehard RNG tests

6. http://www.random.org/. A source of truly random data.

7. http://www.pnas.org/cgi/reprint/61/1/25.pdf. Random Numbers Fall Mainly In The Planes

8. http://www.cs.gmu.edu/~sean/research/. Java implementation of the MT algorithm used for this project.

9. http://www1.fpl.fs.fed.us/distributions.html. Java Normal distribution, for Chi-Square test.

10. http://en.wikipedia.org/wiki/Image:Lcg_3d.gif. Source of images for Figure 1.

Appendix B.

Version 1.2, November 2002

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
* B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
* C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
* D. Preserve all the copyright notices of the Document.
* E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
* F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
* G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
* H. Include an unaltered copy of this License.
* I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
* J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
* K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
* L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
* M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
* N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

* O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.