Equidistribution, Discrepancy, Pseudorandom numbers

Martin Lind

Karlstad University, Sweden

KAAS

The "Main Character"

$$\eta = \left\{\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots\right\}$$

The "main character" of this talk.

The "Main Character"

$$\eta = \left\{\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \ldots\right\}$$

The "main character" of this talk.

"Supporting characters" much more interesting.

The "Main Character"

$$\eta = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots \right\}$$

The "main character" of this talk.

"Supporting characters" much more interesting.

The full story

M. Lind, "A sharp estimate of the discrepancy of a concatenation sequence of inversive pseudorandom numbers with consecutive primes", Int. J. Number Theory, to appear
("A sharp estimate of the discrepancy of a certain numerical sequence", 2021, arxiv)

Equidistribution

Definition (H. Weyl, 1916)

The sequence $\xi = \{\xi_n\}_{n=1}^{\infty} \subseteq [0,1]$ is called **equidistributed in [0, 1]** if

$$\lim_{N\to\infty} \frac{\sharp(\{\xi_1, \xi_2, \dots, \xi_N\} \cap J)}{N} = \mathrm{length}(J)$$

for every interval $J \subseteq [0,1]$.

Equidistribution

Definition (H. Weyl, 1916)

The sequence $\xi = \{\xi_n\}_{n=1}^{\infty} \subseteq [0,1]$ is called **equidistributed in** $[0,1]$ if

$$\lim_{N\to\infty} \frac{\sharp(\{\xi_1, \xi_2, \dots, \xi_N\} \cap J)}{N} = \mathrm{length}(J)$$

for every interval $J \subseteq [0,1]$.

Equidistributed $\approx$ ξ is uniformly spread out in $[0,1]$.

Equidistribution

Definition (H. Weyl, 1916)

The sequence $\xi = \{\xi_n\}_{n=1}^{\infty} \subseteq [0,1]$ is called **equidistributed in [0, 1]** if

$$\lim_{N\to\infty} \frac{\sharp(\{\xi_1, \xi_2, \ldots, \xi_N\} \cap J)}{N} = \mathrm{length}(J)$$

for every interval $J \subseteq [0,1]$.

Equidistributed $\approx$ ξ is uniformly spread out in $[0,1]$.

Equidistributed $\approx$ deterministic version of "uniformly distributed" from probability.

Equidistribution, some remarks

Equidistribution $\Rightarrow$ density. Converse false.

Equidistribution, some remarks

Equidistribution $\Rightarrow$ density. Converse false.

Classically, $[0, 1]$ is always used. Can be exchanged for $[a, b]$

("Equidistribution modulo 1" / "Gleichverteilung mod. Eins")

Equidistribution, some remarks

Equidistribution $\Rightarrow$ density. Converse false.

Classically, $[0, 1]$ is always used. Can be exchanged for $[a, b]$

("Equidistribution modulo 1" / "Gleichverteilung mod. Eins")

Equidistribution as a concept appear in much more general settings than sequences in $[0, 1]$.

Equidistribution, some remarks

Equidistribution $\Rightarrow$ density. Converse false.

Classically, $[0, 1]$ is always used. Can be exchanged for $[a, b]$

("Equidistribution modulo 1" / "Gleichverteilung mod. Eins")

Equidistribution as a concept appear in much more general settings than sequences in $[0, 1]$.

It seems to be one of the fundamental concepts of mathematics.

Weyl's criterion

$$\{\xi_n\}_{n=1}^{\infty} \text{ equidistributed in } [0,1] \Leftrightarrow \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i \xi_n} = 0$$

Weyl's Criterion, number theoretic applications

Weyl's criterion

$$\{\xi_n\}_{n=1}^{\infty} \text{ equidistributed in } [0,1] \Leftrightarrow \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i \xi_n} = 0$$

Weyl first to use **exponential sums** $\sum_{n\leq N} e^{2\pi i \xi_n}$ in number theory.

Weyl's criterion

$$\{\xi_n\}_{n=1}^{\infty} \text{ equidistributed in } [0,1] \Leftrightarrow \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i \xi_n} = 0$$

Weyl first to use **exponential sums** $\sum_{n \le N} e^{2\pi i \xi_n}$ in number theory.

Exercise $\{[n\alpha]\}_{n=1}^{\infty}$ equidistributed in $[0,1] \Leftrightarrow \alpha \notin \mathbb{Q}$.
($[x] = x - \lfloor x \rfloor$ is the *fractional part* of x.)

Measure-theoretic formulation

Empirical measure of $\{\xi_n\}_{n=1}^N$

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{\xi_n}$$

Measure-theoretic formulation

Empirical measure of $\{\xi_n\}_{n=1}^N$

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N}\delta_{\xi_n}$$

$\xi = \{\xi_n\}_{n=1}^\infty$ equidistributed in $[0,1] \Leftrightarrow$

$$\lim_{N\to\infty}\mu_N(J) = m(J), \quad \text{for every interval} \quad J \subseteq [0,1]$$

($m =$ Lebesgue measure on $[0,1]$).

Discrepancy

$$\xi = \{\xi_n\}_{n=1}^{\infty}, \qquad A_N(r) = \sharp(\{\xi_1, \xi_2, \ldots \xi_N\} \cap [0, r])$$

Discrepancy

$$\xi = \{\xi_n\}_{n=1}^{\infty}, \qquad A_N(r) = \sharp(\{\xi_1, \xi_2, \dots \xi_N\} \cap [0, r])$$

Definition

The **discrepancy** (or **star discrepancy**) of ξ is given by

$$D_N^*(\xi) = \sup_{0 \le r \le 1} \left| \frac{A_N(r)}{N} - r \right|$$

$$\xi = \{\xi_n\}_{n=1}^{\infty}, \qquad A_N(r) = \sharp(\{\xi_1, \xi_2, \dots \xi_N\} \cap [0, r])$$

Definition

The **discrepancy** (or **star discrepancy**) of ξ is given by

$$D_N^*(\xi) = \sup_{0 \le r \le 1} \left| \frac{A_N(r)}{N} - r \right|$$

Discrepancy measures equidistribution of ξ

$$\xi \text{ is equidistributed} \quad \Leftrightarrow \quad \lim_{N \to \infty} D_N^*(\xi) = 0$$

Discrepancy

$$\xi = \{\xi_n\}_{n=1}^{\infty}, \qquad A_N(r) = \sharp(\{\xi_1, \xi_2, \dots \xi_N\} \cap [0, r])$$

Definition

The **discrepancy** (or **star discrepancy**) of ξ is given by

$$D_N^*(\xi) = \sup_{0 \le r \le 1} \left| \frac{A_N(r)}{N} - r \right|$$

Discrepancy measures equidistribution of ξ

$$\xi \text{ is equidistributed} \quad \Leftrightarrow \quad \lim_{N \to \infty} D_N^*(\xi) = 0$$

Faster convergence rate of $D_N^*(\xi) \Rightarrow \xi$ "more equidistributed".

Discrepancy

Discrepancy measures how much ξ "deviates" from being uniformly spread out (equidistributed).

Discrepancy

Discrepancy measures how much ξ "deviates" from being uniformly spread out (equidistributed).

Another look at discrepancy as a deviation:

$$D_N^*(\xi) = \sup_{0 \le r \le 1} |\mu_N([0,r]) - m([0,r])|$$

"Total variation-like" distance between μ_N and m.

An example

Exercise Construct an **easy** example of a equidistributed sequence.

An example

Exercise Construct an **easy** example of a equidistributed sequence.

Solution: put together (concatenate) **blocks** of equidistant rational numbers. Denominators increases from one block to another.

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \underbrace{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}}, \dots \right\}$$

An example

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \ldots \right\}$$

The strategy actually works, ω is equidistributed!
In fact, more can be said.

An example

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \dots \right\}$$

The strategy actually works, ω is equidistributed!
In fact, more can be said.

Exercise (nice) Prove that $D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$, and that the convergence rate $N^{-1/2}$ is sharp:

$$\liminf_{N \to \infty} \sqrt{N} D_N^*(\omega) > 0$$

An example

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \ldots \right\}$$

The strategy actually works, ω is equidistributed!
In fact, more can be said.

Exercise (nice) Prove that $D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$, and that the convergence rate $N^{-1/2}$ is sharp:

$$\liminf_{N\to\infty} \sqrt{N} D_N^*(\omega) > 0$$

Shall return to ω later!

Recall the main character!

The "main character":

$$\eta = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \right.$$
$$\left. \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \ldots \right\}$$

Recall the main character!

The "main character":

$$\eta = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots \right\}$$

We recognize the structure from ω, but in η we only have prime denominators.

Recall the main character!

The "main character":

$$\eta = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots \right\}$$

We recognize the structure from ω, but in η we only have prime denominators.

Question Why the numerators?

Recall the main character!

The "main character":

$$\eta = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots \right\}$$

We recognize the structure from ω, but in η we only have prime denominators.

Question Why the numerators?

To answer this, we need randomness!

Pseudorandom numbers

Random numbers are useful!

Problem: What/how/where is "random"?

Pseudorandom numbers

Random numbers are useful!

Problem: What/how/where is "random"?

Substitute: **pseudorandom numbers**.

Numbers generated by some algorithm (so **not random in any meaningful sense**) that appears to be random/unpredictable.

Pseudorandom numbers

Random numbers are useful!

Problem: What/how/where is "random"?

Substitute: **pseudorandom numbers**.

Numbers generated by some algorithm (so **not random in any meaningful sense**) that appears to be random/unpredictable.

Generating **good** pseudorandom numbers is a serious scientific problem!

R. R. Coveyou: "Random number generation is too important to be left to chance."

D. E. Knuth: "Random numbers should not be generated with a method chosen at random."

Pseudorandom numbers

It turns out to be sufficient to generate "random numbers" from $U[0,1]$ (uniform distribution on $[0,1]$).

Pseudorandom numbers

It turns out to be sufficient to generate "random numbers" from $U[0,1]$ (uniform distribution on $[0,1]$).

Any other distribution (e.g. normal, Poisson,...) can be obtained from $U[0,1]$ by **inverse transform sampling**.

Pseudorandom numbers

To generate "random numbers" from $U[0,1]$, one often does the following.

Pseudorandom numbers

To generate "random numbers" from $U[0,1]$, one often does the following.

(1) Generate "uniformly unpredictable" integers $m_1, m_2, \ldots m_N$ in a large interval $[0, K]$.

Pseudorandom numbers

To generate "random numbers" from $U[0,1]$, one often does the following.

(1) Generate "uniformly unpredictable" integers $m_1, m_2, \dots m_N$ in a large interval $[0, K]$.

(2) Normalize to $[0,1]$ by setting $x_n = m_n/K$ for $n = 1, 2, \dots, N$

Pseudorandom numbers

To generate "random numbers" from $U[0,1]$, one often does the following.

(1) Generate "uniformly unpredictable" integers $m_1, m_2, \dots m_N$ in a large interval $[0, K]$.

(2) Normalize to $[0,1]$ by setting $x_n = m_n/K$ for $n = 1, 2, \dots, N$

Getting the "uniformly unpredictable" integers is of course the hard part.

They are often generated arithmetically.

Simple but powerful example

Let p be a (large) prime and consider the map

$$\xi \mapsto \xi^{-1} \pmod{p}$$

defined on $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

Simple but powerful example

Let p be a (large) prime and consider the map

$$\xi \mapsto \xi^{-1} \pmod{p}$$

defined on $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

$1 \mapsto 1, \quad 2 \mapsto (p+1)/2, \quad p-1 \mapsto p-1$

Simple but powerful example

Let p be a (large) prime and consider the map

$$\xi \mapsto \xi^{-1} \pmod{p}$$

defined on $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

$1 \mapsto 1, \quad 2 \mapsto (p+1)/2, \quad p-1 \mapsto p-1$

Not unpredictable!

Simple but powerful example

Let p be a (large) prime and consider the map

$$\xi \mapsto \xi^{-1} \pmod{p}$$

defined on $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

$1 \mapsto 1, \quad 2 \mapsto (p+1)/2, \quad p-1 \mapsto p-1$

Not unpredictable!

Already $3^{-1} \pmod{p}$ is less obvious.
(Two possibilities, which it is depends on $p \pmod{3}$.)

Simple but powerful example
Let p be a (large) prime and consider the map

$$\xi \mapsto \xi^{-1} \pmod{p}$$

defined on $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

$1 \mapsto 1, \quad 2 \mapsto (p+1)/2, \quad p-1 \mapsto p-1$

Not unpredictable!

Already $3^{-1} \pmod{p}$ is less obvious.
(Two possibilities, which it is depends on $p \pmod{3}$.)

Take a "small chunk" $\{k, k+1, \ldots, k+m\} \subset \mathbb{Z}_p$ and consider its image under the inverse map. Will look rather unpredictable!

Pseudorandom numbers

Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0,1]$.

Pseudorandom numbers

Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0,1]$.

Take $p = 1667$

Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0,1]$.

Take $p = 1667$
Let $\zeta_n = (800 + n)^{-1} \pmod{1667}$ for $n = 1, 2, \ldots 20$

Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0,1]$.

Take $p = 1667$
Let $\zeta_n = (800 + n)^{-1} \pmod{1667}$ for $n = 1, 2, \ldots 20$
Normalize to $[0,1]$ by taking $x_n = \dfrac{\zeta_n}{p}$ for $n = 1, 2, \ldots 20$

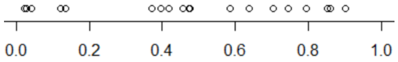
Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0, 1]$.

Take $p = 1667$

Let $\zeta_n = (800 + n)^{-1} \pmod{1667}$ for $n = 1, 2, \ldots 20$

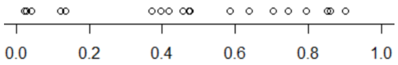
Normalize to $[0, 1]$ by taking $x_n = \dfrac{\zeta_n}{p}$ for $n = 1, 2, \ldots 20$



Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0, 1]$.

Take $p = 1667$
Let $\zeta_n = (800 + n)^{-1} \pmod{1667}$ for $n = 1, 2, \ldots 20$
Normalize to $[0, 1]$ by taking $x_n = \dfrac{\zeta_n}{p}$ for $n = 1, 2, \ldots 20$

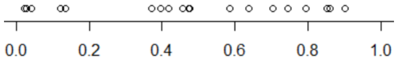


Looks rather random! (Clusters and holes)

Practical test

Provide a (pseudo)random sample $\{x_1, x_2, \ldots, x_{20}\}$ from $U[0, 1]$.

Take $p = 1667$
Let $\zeta_n = (800 + n)^{-1} \pmod{1667}$ for $n = 1, 2, \ldots 20$
Normalize to $[0, 1]$ by taking $x_n = \dfrac{\zeta_n}{p}$ for $n = 1, 2, \ldots 20$



Looks rather random! (Clusters and holes)

Can also perform a statistical test.

Pseudorandom numbers

Want to test if $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a random sample from $U[0, 1]$ (null hypothesis).

Pseudorandom numbers

Want to test if $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a random sample from $U[0, 1]$ (null hypothesis).

The Kolmogorov-Smirnov test (nonparametric test).

Pseudorandom numbers

Want to test if $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a random sample from $U[0,1]$ (null hypothesis).

The Kolmogorov-Smirnov test (nonparametric test).

Empirical distribution function

$$\hat{F}_{\mathbf{x}}(t) = \frac{\sharp(\{x_n \in \mathbf{x} : x_n \leq t\})}{N}$$

Want to test if $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a random sample from $U[0,1]$ (null hypothesis).

The Kolmogorov-Smirnov test (nonparametric test).

Empirical distribution function

$$\hat{F}_{\mathbf{x}}(t) = \frac{\sharp(\{x_n \in \mathbf{x} : x_n \leq t\})}{N}$$

Test statistic

$$T = \sup_{0 \leq t \leq 1} |\hat{F}_{\mathbf{x}}(t) - t|$$

(Observe that $F(t) = t \quad (0 \leq t \leq 1)$ is the distribution function of a random variable $X \sim U[0,1]$.)

Pseudorandom numbers

Want to test if $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a random sample from $U[0,1]$ (null hypothesis).

The Kolmogorov-Smirnov test (nonparametric test).

Empirical distribution function

$$\hat{F}_{\mathbf{x}}(t) = \frac{\sharp(\{x_n \in \mathbf{x} : x_n \leq t\})}{N}$$

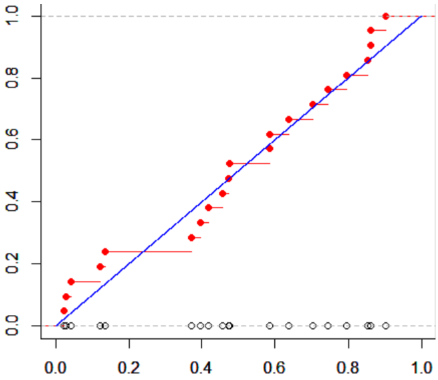
Test statistic

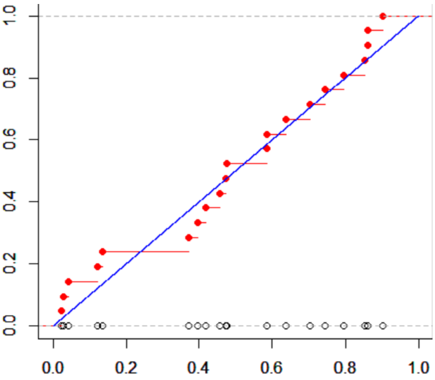
$$T = \sup_{0 \leq t \leq 1} |\hat{F}_{\mathbf{x}}(t) - t|$$

(Observe that $F(t) = t \quad (0 \leq t \leq 1)$ is the distribution function of a random variable $X \sim U[0,1]$.)

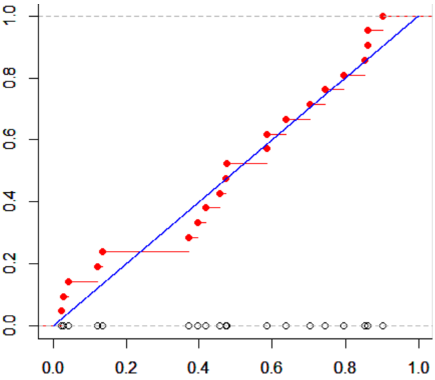
Reject H_0 if T is larger than tabulated critical value.

Pseudorandom numbers





Interesting (?) observation: the test statistic is the star discrepancy of the sample: $T = D_N^*(\mathbf{x})$.



Interesting (?) observation: the test statistic is the star discrepancy of the sample: $T = D_N^*(\mathbf{x})$.

Additional testing necessary to guarantee quality of pseudorandom numbers.

Back to ω and η

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots \right\}, \quad D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

Back to ω and η

$$\omega = \left\{\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots\right\}, \quad D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

Two possible factors that slow down the convergence rate:

1. ω contains many "duplicates", e.g. $1/2 = 2/4 = 3/6$ etc.;
2. the terms of ω within each block is ordered increasingly, i.e. $1/5, 2/5, 3/5, 4/5$.

Back to ω and η

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots \right\}, \quad D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}} \right)$$

Two possible factors that slow down the convergence rate:

1. ω contains many "duplicates", e.g. $1/2 = 2/4 = 3/6$ etc.;
2. the terms of ω within each block is ordered increasingly, i.e. $1/5, 2/5, 3/5, 4/5$.

The first issue is easily solved: only prime denominators in the blocks.

$$\left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \underbrace{\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}}, \dots \right\}$$

Back to ω and η

$$\omega = \left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots \right\}, \quad D_N^*(\omega) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

Two possible factors that slow down the convergence rate:

1. ω contains many "duplicates", e.g. $1/2 = 2/4 = 3/6$ etc.;
2. the terms of ω within each block is ordered increasingly, i.e. $1/5, 2/5, 3/5, 4/5$.

The first issue is easily solved: only prime denominators in the blocks.

$$\left\{ \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \underbrace{\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}}, \dots \right\}$$

Issue with order within each block remains!

More about η

The ordering issue: impose within each block of η the inversive pseudorandom order:

More about η

The ordering issue: impose within each block of η the inversive pseudorandom order:

$$\left\{\frac{1^{-1}}{p}, \frac{2^{-1}}{p}, \frac{3^{-1}}{p}, \ldots \frac{(p-1)^{-1}}{p}\right\},$$

where the inverse is $\pmod{p}$.

The elements (except first and last in the block) "jump around"!

More about η

The ordering issue: impose within each block of η the inversive pseudorandom order:

$$\left\{\frac{1^{-1}}{p}, \frac{2^{-1}}{p}, \frac{3^{-1}}{p}, \dots \frac{(p-1)^{-1}}{p}\right\},$$

where the inverse is $\pmod{p}$.

The elements (except first and last in the block) "jump around"!

The result is

$$\eta = \left\{\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \underbrace{\frac{1}{7}, \frac{4}{7}, \frac{5}{7}, \frac{2}{7}, \frac{3}{7}, \frac{6}{7}}, \frac{1}{11}, \frac{6}{11}, \frac{4}{11}, \frac{3}{11}, \frac{9}{11}, \frac{2}{11}, \frac{8}{11}, \frac{7}{11}, \frac{5}{11}, \frac{10}{11}, \dots\right\}$$

The main problem

Now that we know the construction of η, I formulate the main problem that I solved

The main problem

Now that we know the construction of η, I formulate the main problem that I solved

Problem

Compute the exact asymptotic behaviour of $D_N^*(\eta)$.

The main problem

Now that we know the construction of η, I formulate the main problem that I solved

Problem

Compute the exact asymptotic behaviour of $D_N^*(\eta)$.

Motivation: silly curiosity

The main result

Theorem (The Skarphyttan Theorem, 2021)

For $N \geq 3$

$$D_N^*(\eta) \leq \frac{2}{\sqrt{N \ln(N)}}.$$

Moreover, the rate is sharp:

$$\liminf_{N\to\infty} \sqrt{N \ln(N)} D_N^*(\eta) \geq \frac{1}{2}$$

The main result

Theorem (The Skarphyttan Theorem, 2021)

For $N \geq 3$

$$D_N^*(\eta) \leq \frac{2}{\sqrt{N \ln(N)}}.$$

Moreover, the rate is sharp:

$$\liminf_{N\to\infty} \sqrt{N \ln(N)} D_N^*(\eta) \geq \frac{1}{2}$$

Who/what is Skarphyttan?

The main result

Main result, some remarks

The improvement in rate of $D_N^*(\eta) = \mathcal{O}\left(\frac{1}{\sqrt{N\ln(N)}}\right)$ compared to $D_N^*(\omega) = \mathcal{O}(N^{-1/2})$ is due to the pseudorandom ordering of the elements in η.

Main result, some remarks

The improvement in rate of $D_N^*(\eta) = \mathcal{O}\left(\frac{1}{\sqrt{N \ln(N)}}\right)$ compared to $D_N^*(\omega) = \mathcal{O}(N^{-1/2})$ is due to the pseudorandom ordering of the elements in η.

On the other hand, it is interesting to note the following.

(**Law of the iterated logarithm for** D_N^*)
If $\xi = \{\xi_n\}_{n=1}^{\infty}$ is a random sequence (i.e. $\xi_n \sim U[0,1]$), then almost surely

$$D_N^*(\xi) = \mathcal{O}\left(\sqrt{\frac{\ln(\ln(N))}{N}}\right).$$

Main result, some remarks

The improvement in rate of $D_N^*(\eta) = \mathcal{O}\left(\frac{1}{\sqrt{N \ln(N)}}\right)$ compared to $D_N^*(\omega) = \mathcal{O}(N^{-1/2})$ is due to the pseudorandom ordering of the elements in η.

On the other hand, it is interesting to note the following.

(Law of the iterated logarithm for D_N^*)
If $\xi = \{\xi_n\}_{n=1}^{\infty}$ is a random sequence (i.e. $\xi_n \sim U[0,1]$), then almost surely

$$D_N^*(\xi) = \mathcal{O}\left(\sqrt{\frac{\ln(\ln(N))}{N}}\right).$$

So our notion of "pseudorandom" is quite far away from "really random"!

Proof technique

If time permits: I want to say something about the proof.

Mainly to illustrate the last ingredient of the argument: asymptotics for prime numbers.

Proof technique

Want to estimate $D_N^*(\eta)$. Here, p_n is the n-th prime.

Proof technique

Want to estimate $D_N^*(\eta)$. Here, p_n is the n-th prime.

1. Writing $N \approx \sum_{n=1}^{m} p_n$ for some m and using the "triangle inequality for discrepancy", I get

$$ND_N^*(\eta) \leq I + II;$$

Proof technique

Want to estimate $D_N^*(\eta)$. Here, p_n is the n-th prime.

1. Writing $N \approx \sum_{n=1}^{m} p_n$ for some m and using the "triangle inequality for discrepancy", I get

$$N D_N^*(\eta) \leq I + II;$$

2. the first term can be estimated as $I \leq \sum_{n=1}^{m-1} p_n$;

Proof technique

Want to estimate $D_N^*(\eta)$. Here, p_n is the n-th prime.

1. Writing $N \approx \sum_{n=1}^{m} p_n$ for some m and using the "triangle inequality for discrepancy", I get

$$ND_N^*(\eta) \leq I + II;$$

2. the first term can be estimated as $I \leq \sum_{n=1}^{m-1} p_n$;
3. using general discrepancy estimates due to Niederreiter for inversive congruential generators (i.e. for the map $\zeta \mapsto \zeta^{-1}$ on $\mathbb{Z}_p^*$ for fixed p), the second term can be estimated as $II \leq C\sqrt{p_m}\ln^2(p_m)$.

Proof technique

I thus arrive (essentially) at

$$ND_N^*(\eta) \le \sum_{n=1}^{m-1} p_n + C\sqrt{p_m}\ln^2(p_m)$$

Proof technique

I thus arrive (essentially) at

$$ND_N^*(\eta) \leq \sum_{n=1}^{m-1} p_n + C\sqrt{p_m}\ln^2(p_m)$$

The above can be massaged into the desired estimate **if** I have some knowledge of the asymptotics of primes.

Last ingredient: size of primes

$\pi(x) = \sharp(\{\text{primes} \le x\})$

Last ingredient: size of primes

$\pi(x) = \sharp(\{\text{primes} \leq x\})$

The PNT (Hadamard, de la Vallée Poussin 1896)

$$\pi(x) \sim \frac{x}{\ln(x)}.$$

Last ingredient: size of primes

$\pi(x) = \sharp(\{\text{primes} \leq x\})$

The PNT (Hadamard, de la Vallée Poussin 1896)

$$\pi(x) \sim \frac{x}{\ln(x)}.$$

Equivalently

$$m\text{-th prime } p_m = m \ln(m)(1 + o(1)).$$

Last ingredient: size of primes

$\pi(x) = \sharp(\{\text{primes} \leq x\})$

The PNT (Hadamard, de la Vallée Poussin 1896)

$$\pi(x) \sim \frac{x}{\ln(x)}.$$

Equivalently

$$m\text{-th prime } p_m = m \ln(m)(1 + o(1)).$$

I also need the asymptotic behaviour of $\sum_{n=1}^{m} p_n$.

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

$$\sum_{n=1}^{m} p_n \approx \sum_{n=1}^{m} n \ln(n) \approx \int_1^m x \ln(x) dx$$

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

$$\begin{aligned}\sum_{n=1}^{m} p_n &\approx \sum_{n=1}^{m} n\ln(n) \approx \int_1^m x\ln(x)dx \\ &= \frac{m^2}{2}\ln(m) - \int_1^m \frac{x}{2}dx\end{aligned}$$

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

$$\begin{aligned}
\sum_{n=1}^{m} p_n &\approx \sum_{n=1}^{m} n\ln(n) \approx \int_1^m x\ln(x)dx \\
&= \frac{m^2}{2}\ln(m) - \int_1^m \frac{x}{2}dx \\
&= \frac{m^2}{2}\ln(m)\left(1 - \frac{C_1}{\ln(m)} + \frac{C_2}{m^2\ln(m)}\right)
\end{aligned}$$

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

$$\sum_{n=1}^{m} p_n \approx \sum_{n=1}^{m} n\ln(n) \approx \int_1^m x\ln(x)dx$$
$$= \frac{m^2}{2}\ln(m) - \int_1^m \frac{x}{2}dx$$
$$= \frac{m^2}{2}\ln(m)\left(1 - \frac{C_1}{\ln(m)} + \frac{C_2}{m^2\ln(m)}\right)$$

Fantastically, the above heuristic actually works!

Last ingredient: size of primes

Heuristic argument (Skarphyttan has no Internet)

$$\begin{aligned}\sum_{n=1}^{m} p_n &\approx \sum_{n=1}^{m} n\ln(n) \approx \int_1^m x\ln(x)dx \\ &= \frac{m^2}{2}\ln(m) - \int_1^m \frac{x}{2}dx \\ &= \frac{m^2}{2}\ln(m)\left(1 - \frac{C_1}{\ln(m)} + \frac{C_2}{m^2\ln(m)}\right)\end{aligned}$$

Fantastically, the above heuristic actually works!

$$\sum_{n=1}^{m} p_n = \frac{m^2}{2}\ln(m)(1+o(1))$$

(see e.g. Landau's "Handbuch")