How good is your puzzle anyway?

Computational (non?)randomness

Albert Stark

September 9, 2026

Hello! 👋

I'm Albert, a 3rd year and normally one of the organisers of these talks
If you want to see these slides or the corresponding notes I'm reading off, they'll be up on my website https://albert10032.github.io/

Structure and randomness

First-order approximation (symbols independent but with frequencies of English text).

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

Second-order approximation (digram structure as in English).

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

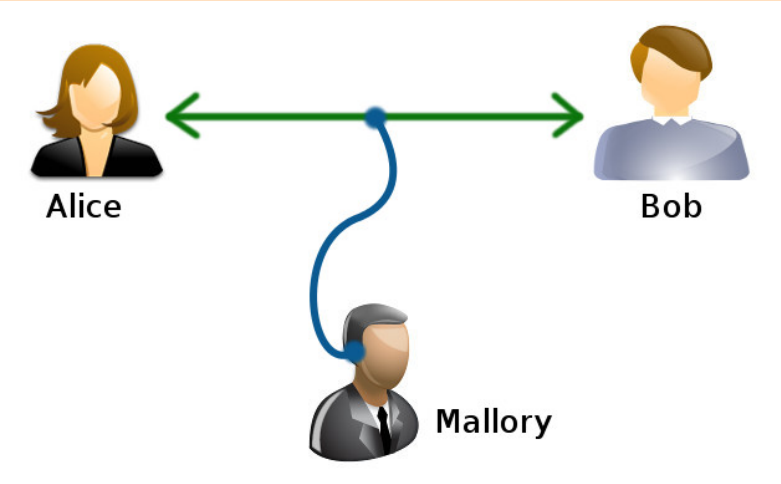
Third-order approximation (trigram structure as in English).

IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONS-TURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

Claude Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27 (1948). URL: https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf, §3

Protecting secrets

Question: How to transfer information between 2 parties, and be certain that any third party that intercepts it cannot understand it?



Answer: Encryption!

Question: Language has structure, how to design a cryptosystem that does not betray that structure?

Answer: Possible. This has to be tested.

Reducing the problem

Good encryption seems random

$\Downarrow$

Every symbol is *Independent and Identically Distributed (IID)*

Encrypting data $\leftrightsquigarrow$ Pseudo random number generation (PRNG)

How to test a PRNG?

A first idea: The χ^2 test

Do some *observations* fit an *expected* distribution?

Given X a value we are calculating

$$X = \sum_{i=0}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Compare with known value for χ_p^{k-1}, given $k-1$ degrees of freedom and confidence level p

Alone, this is insufficient!!

E.g. $\{\cdots$ 😀🪱😀🪱😀🪱😀🪱😀🪱😀🪱$\cdots\}$

Standardised tests

χ^2 is insensitive to the order of observation!

Want: A test (or collection of tests) that are *sensitive to order of observation* in many different ways

This an engineering challenge more than it is a mathematical one

First tests proposed by Knuth. A modern suite is TestU01.
We focus today on the DIEHARD tests.

Random matrix theory (1)

Question: Consider $M \in \mathrm{Mat}_n(\mathbb{F}_2), m_{i,j} \sim U(\{0,1\})$. Let $\mathcal{R}_n : \mathrm{Mat}_n(\mathbb{F}_2) \to \{0, 1, \cdots, n\}$ be the rank function.
What is the expected distribution of $\mathcal{R}$ given a large number of random matrices?

Answer: Given a new function $\mathcal{Q}_n := n - \mathcal{R}_n$, we have the following

$$P(\mathcal{Q}_n = k) = p_{k,n} = \frac{1}{2^{k^2}} \frac{\prod_{i=k+1}^{n} \left(1 - \frac{1}{2^i}\right)^2}{\prod_{i=1}^{n-k} \left(1 - \frac{1}{2^i}\right)}$$

Given to us by Fullman and Goldstein[a].

[a] Jason Fulman and Larry Goldstein. "Stein's method and the rank distribution of random matrices over finite fields". In: *The Annals of Probability* 43.3 (2015). ISSN: 0091-1798. DOI: 10.1214/13-aop889. URL: http://dx.doi.org/10.1214/13-AOP889.

Random matrix theory (2)

Index	b_0	b_1	$\cdots$	$b_{n\times n}$	$b_{n\times n+1}$	$\cdots$	$b_{M-2(n\times n)+1}$	$\cdots$	$b_{M-(n\times n)}$
Stream	0	1	$\cdots$	1	0	$\cdots$	1	$\cdots$	0
M_1	0	1	$\cdots$	1					
M_2		1	$\cdots$	1	0				
$\cdots$									
M_M							1	$\cdots$	0

$R_i := \mathcal{R}(M_i) \rightsquigarrow \chi^2$ test with the previous dist.

To hedge against constructive conspiracies we run the test on multiple different sizes of matrix; $32 \times 32, 31 \times 31$ and 6×8

Bitstreams

Question: Given a string of 0,1 taken as an *"alphabet"* consider a subset of overlapping *"words"*. What would we expect the distribution of words of length l in a larger string of length $n >> l$ to be, assuming each letter was iid?

Stream	$0010000111110010010110000\cdots$
W_1	00100001111100100101
W_2	01000011111001001011
W_3	10000111110010010110
$\vdots$	

Answer: Something normal probably??? (Yes)

Fact: In a stream of length $2^{21} - 19$ it is known that
$J := \#\{20 \text{ letter words } \underline{\text{not}} \text{ observed}\} \rightsquigarrow J \sim N(141909, 428)$

Break input into 20 of these $2^{21} - 19$ bit blocks and repeat on each for extra confidence

Count the ones (🍚)

Question: Given a particular byte, what is the probability distribution for the number of ones in that byte?

Answer: $P(\#1s = n) = \frac{1}{2^8}\binom{8}{n} = \left\{\frac{1}{256}, \frac{8}{256}, \frac{28}{256}, \frac{56}{256}, \frac{70}{256}, \frac{56}{256}, \frac{28}{256}, \frac{8}{256}, \frac{1}{256}\right\}$

	A	B	C	D	E
# ones	0, 1, 2	3	4	5	6, 7, 8
P	37/256	56/256	70/256	56/256	37/256

Count the ones (🍚🍚)

Stream	00100001111100100101100011011101000011010010110110101110 $\cdots$							
Letters	A	D	B	E	B	C	D	$\cdots$
F_1	A	D	B	E	B			
F_2		D	B	E	B	C		
F_3			B	E	B	C	D	

$n = 256000$ overlapping words $\rightsquigarrow \chi^2$ test on a uniform distribution.

Note: a.) Byte-stream not bit-stream!
b.) As these things go, $n = 256000 \nleftrightsquigarrow$ 255kb, compare to last test 2mb

Birthday spacings (🎂)

Question:

Answer:

$P(\text{noone shares a bday}) = \binom{n}{m}\frac{m!}{n^m} \rightsquigarrow P(\text{At least 2 people share a bday}) = 1 - \binom{n}{m}\frac{m!}{n^m}$.

About 23 people needed for $P > 0.5$ in a year with 365 days

Birthday spacings (🎂 🎂)

Question: With the same setup, if we create a new list of all the spacings between birthdays what is the expected distribution of this "inter-birthday interval"?

$$B = \{1, 100, 84, 31, 33, 2, 1, 33, 4, \cdots\} \xrightarrow[\text{sort}]{} B_{\text{ordered}} = \{1, 1, 1, 2, 4, 7, 8, 12, 12, 13, \cdots\}$$
$$I_{B_{\text{ordered}}} = \{0, 0, 1, 2, 3, 1, 4, 0, 1, \cdots\}$$
$$J_{n,m}(x) = \#\text{ of } x\text{s in } I_{B_{\text{ordered}}}$$

Answer: $J_{n,m} \sim Poisson\left(\frac{m^3}{4n}\right)$

Birthday spacings (🎂 🎂 🎂)

The DIEHARD tests imagine a year with $n = 2^{24}$ days and $m = 2^9$ birthdays.
In this case take a stream of 2^9 32 bit integers as the input ($\approx$ 16kb!)

Note: a.) Again observe that we are performing multiple levels of analysis to reinforce the power of this test.
b.) Also note that we only need about 16kb which is *tiny!* small enough that it can be performed in blocks of itself to try and catch poorly padded regions if we wish.

(CS)PRNGs in practice (1)

A prototypical example of one is just chucking some bit rotations and XORs on the local time and hoping it works (bad 🍺 🥤 🥤 🥤 🥤), or using something like a Mersenne Twister (better 🍺 🍺 🥤 🥤 🥤)

Linear, therefore breakable

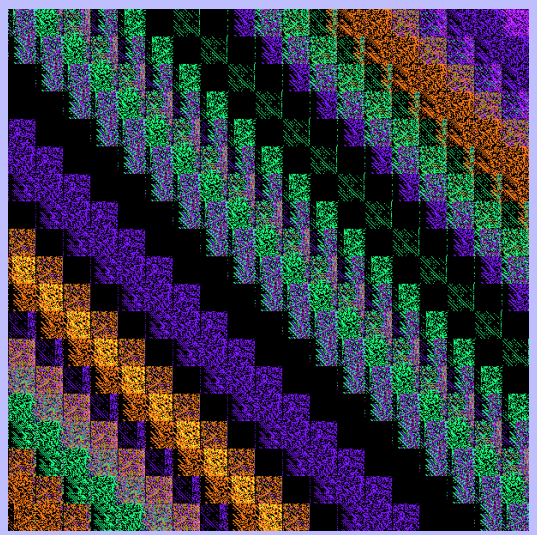


Figure: Generated by JuliaPoo[a] using 24-bit MT19937 Mersenne Twister

[a] JuliaPoo. *MT19937*. 2021. URL: https://juliapoo.github.io/art/2021-08-26-mt.html.

(CS)PRNGs in practice (2)

If you have a already created a good cipher (as verified by tests like this) then you can build a good PRNG by recursively calling the cipher on its own output using the same key in a process known as "counter mode"

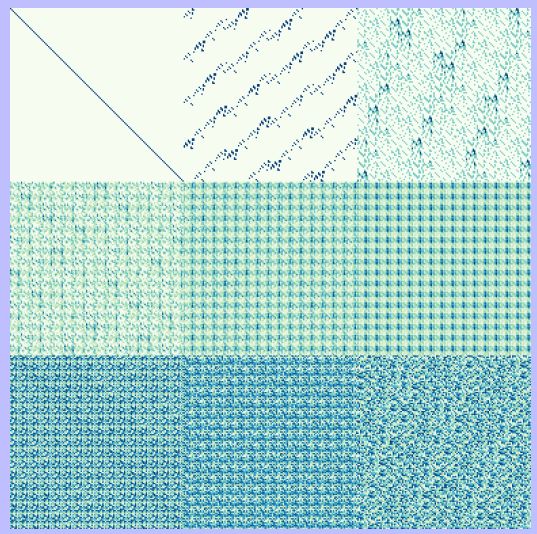


Figure: Generated by JuliaPoo using AES-128[a]

[a] JuliaPoo. *AES-128 Diffusion*. 2022. URL: https://juliapoo.github.io/art/2022-01-09-aes-diffusion.html.

Kleptography

There are however some who work against this...

Question: If you want to know what everyone is saying, what do you do?

You control how they say it!

fin.

May your ciphertexts look like noise, your keys be computationally infeasible and your standards untampered with by those who don't respect you

Thank you :)

Bibliography

 Fulman, Jason and Larry Goldstein. "Stein's method and the rank distribution of random matrices over finite fields". In: *The Annals of Probability* 43.3 (2015). ISSN: 0091-1798. DOI: 10.1214/13-aop889. URL: http://dx.doi.org/10.1214/13-AOP889.

 JuliaPoo. *AES-128 Diffusion*. 2022. URL: https://juliapoo.github.io/art/2022-01-09-aes-diffusion.html.

 — .*MT19937*. 2021. URL: https://juliapoo.github.io/art/2021-08-26-mt.html.

 Shannon, Claude. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27 (1948). URL: https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.