EVOLUTIONARY MARKOV CHAIN MONTE CARLO IN DECRYPTION:

BENCHMARKS & COMPARISONS TO SINGLE-CHAIN MARKOV CHAIN MONTE CARLO

ARSHIA SINGH
ANLY601: ADVANCED MACHINE LEARNING

ABSTRACT

I applied both single chain and evolutionary Markov Chain Monte Carlo methods to a piece of substitution ciphered text to try to recover the mappings, and the original text. I measured the efficiency of each method via the computational time and total steps travelled needed to achieve recovery of the original text. My results did not indicate any clear advantages to using evolutionary methods over a single chain Markov Chain Monte Carlo approach, but I think it would be fruitful to continue exploring further applications of evolutionary methods to this problem, particularly via the parallelization of a population of chains.

INTRODUCTION AND PROBLEM STATEMENT

- Few studies have applied evolutionary methods to MCMC, and even fewer have conducted comparisons between the two methods in terms of computational and step efficiency
- Research has shown that single-chain MCMC is generally effective at recovering a text that has been encoded with a simple substitution cipher
- Ciphers are no longer a common method of encryption but present a problem that is easily tested and validated
- If successful, evolutionary MCMC methodology might be leveraged in other areas of stochastic simulation like crowd movements and the spread of a disease

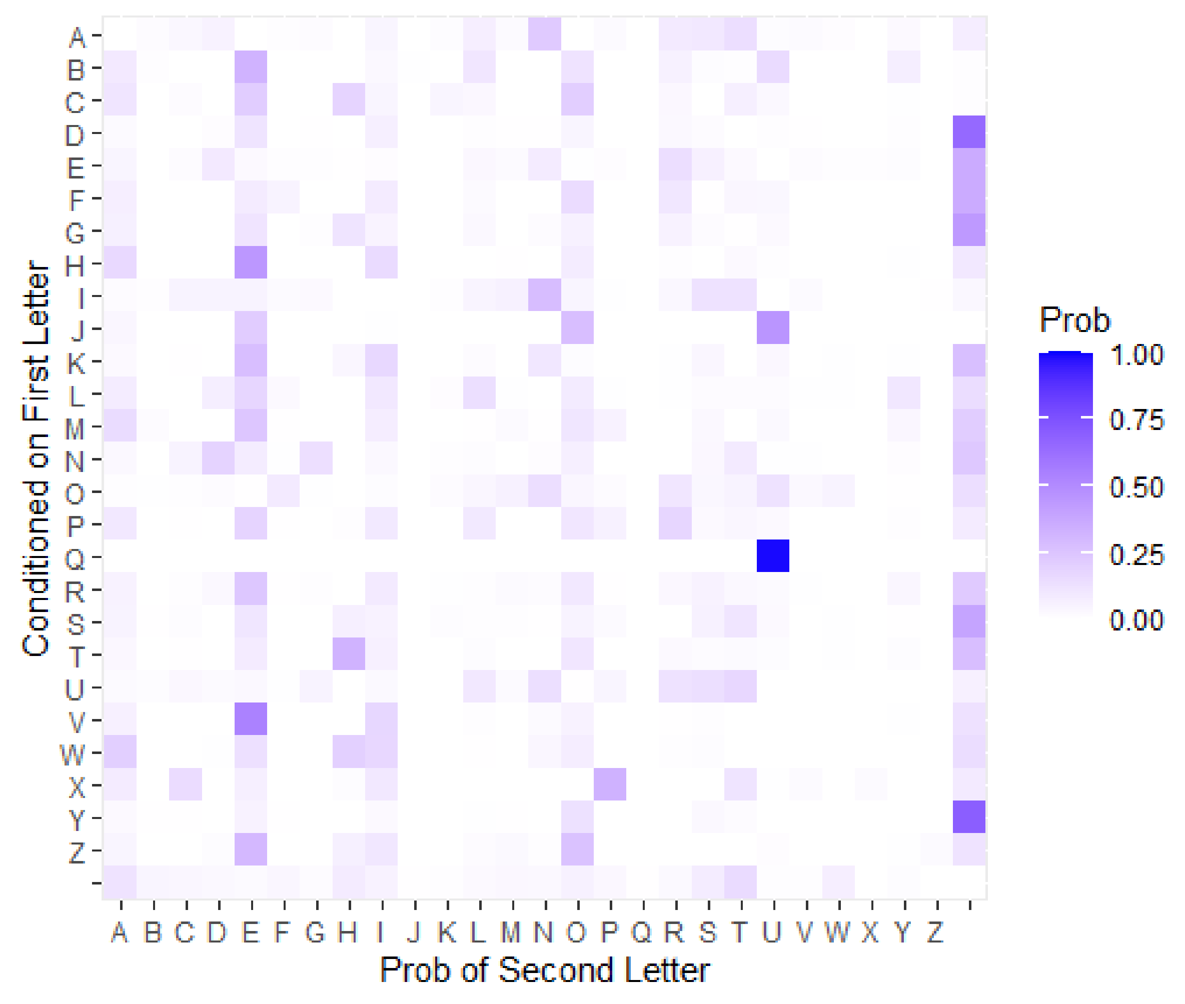
RELATED WORK

- Diaconis (2009) explores the use of single-chain MCMC in the space of cryptography. Given a text that has been encrypted using a series of substitution codes, he initializes a proposal based on a sample from a probability distribution modeled on letter frequency in a reference text. The acceptance criteria is based on the frequency of letter pairings in the reference text (for example, the pairing "AS" is more probable than "ZK").
- Drugan and Thierens (2003) summarized research on developments in the space of evolutionary MCMC and categorize these algorithms into two categories:
  - Family competitive algorithms: two chains exchange information and recombine
  - Population-driven algorithms: proposal distribution adapts based on the entire current population
- Chen and Rosenthal (2010) go further than substation-based encryption to show that Markov Chain Monte Carlo can be used to successfully decrypt transposition ciphers and even substitution-plus-transposition ciphers
- Alsharafat (2015) explores the use of evolutionary algorithms to create strong encryption methods through crossover and mutation-based ciphers. However, the paper doesn't explore the use of these methods to decrypt an encrypted text with an unknown key.
- Garg (2010) looks at the theoretical application of evolutionary algorithms as a mechanism for cryptanalysis but does not apply them or evaluate their effectiveness at decryption versus other methods.

DATA SOURCES

- Full text of War and Peace by Leo Tolstoy (65,219 lines/8.2 MB)
- A paragraph of text from Great Expectations by Charles Dickens:

  "The marshes were just a long black horizontal line then, as I stopped to look after him; and the river was just another horizontal line, not nearly so broad nor yet so black; and the sky was just a row of long angry red lines and dense black lines intermixed. On the edge of the river I could faintly make out the only two black things in all the prospect that seemed to be standing upright; one of these was the beacon by which the sailors steered, like an unhooped cask upon a pole, an ugly thing when you were near it; the other, a gibbet, with some chains hanging to it which had once held a pirate. The man was limping on towards this latter, as if he were the pirate come to life, and come down, and going back to hook himself up again. It gave me a terrible turn when I thought so; and as I saw the cattle lifting their heads to gaze after him, I wondered whether they thought so too. I looked all round for the horrible young man, and could see no signs of him. But now I was frightened again, and ran home without stopping.“
- Using a reference text leads to more flexibility than many NLP models



TRANSITION PROBABILITY MATRIX

"U" is very likely to follow the letter "Q"

"Y" and "D" are more likely to appear at the end of a word than other letters

the letters "A" and "T" are marginally more likely to appear at the beginning of a word

- "THE MARSHES WERE JUST A LONG BLACK HORIZONTAL LINE THEN AS I STOPPED TO LOOK AFTER HIM AND THE RIVER WAS JUST ANOTHER HORIZONTAL LINE NOT NEARLY SO BROAD NOR YET SO BLACK AND THE SKY WAS JUST A ROW OF LONG ANGRY RED LINES AND DENSE BLACK LINES INTERMIXED ON THE EDGE OF THE RIVER I COULD FAINTLY MAKE OUT THE ONLY TWO BLACK THINGS IN ALL THE PROSPECT THAT SEEMED TO BE STANDING UPRIGHT ONE OF THESE WAS THE BEACON BY WHICH THE SAILORS STEEREDLIKE AN UNHOOPED CASK UPON A POLEAN UGLY THING WHEN YOU WERE NEAR IT THE OTHER A GIBBET WITH SOME CHAINS HANGING TO IT WHICH HAD ONCE HELD A PIRATE THE MAN WAS LIMPING ON TOWARDS THIS LATTER AS IF HE WERE THE PIRATE COME TO LIFE AND COME DOWN AND GOING BACK TO HOOK HIMSELF UP AGAIN IT GAVE ME A TERRIBLE TURN WHEN I THOUGHT SO AND AS I SAW THE CATTLE LIFTING THEIR HEADS TO GAZE AFTER HIM I WONDERED WHETHER THEY THOUGHT SO TOO I LOOKED ALL ROUND FOR THE HORRIBLE YOUNG MAN AND COULD SEE NO SIGNS OF HIM BUT NOW I WAS FRIGHTENED AGAIN AND RAN HOME WITHOUT STOPPING“

- "ZFX HIKGFXG MXKX LPGZ I CASY OCINV FAKTBASZIC CTSX ZFXS IG T GZAUUXE ZA CAAV IQZXK FTH ISE ZFX KTWXK MIG LPGZ ISAZFXK FAKTBASZIC CTSX SAZ SXIKCJ GA OKAIE SAK JXZ GA OCINV ISE ZFX GVJ MIG LPGZ I KAM AQ CASY ISYKJ KXE CTSXG ISE EXSGX OCINV CTSXG TSZXKHTRXE AS ZFX XEYX AQ ZFX KTWXK T NAPCE QITSZCJ HIVX APZ ZFX ASCJ ZMA OCINV ZFTSYG TS ICC ZFX UKAGUXNZ ZFIZ GXXHXE ZA OX GZISETSY PUKTYFZ ASX AQ ZFXGX MIG ZFX OXINAS OJ MFTNF ZFX GITCAKG GZXXKXECTVX IS PSFAAUXE NIGV PUAS I UACXIS PYCJ ZFTSY MFXS JAP MXKX SXIK TZ ZFX AZFXK I YTOOXZ MTZF GAHX NFITSG FISYTSY ZA TZ MFTNF FIE ASNX FXCE I UTKIZX ZFX HIS MIG CTHUTSY AS ZAMIKEG ZFTG CIZZXK IG TQ FX MXKX ZFX UTKIZX NAHX ZA CTQX ISE NAHX EAMS ISE YATSY OINV ZA FAAV FTHGXCQ PU IYITS TZ YIWX HX I ZXKKTOCX ZPKS MFXS T ZFAPYFZ GA ISE IG T GIM ZFX NIZZCX CTQZTSY ZFXTK FXIEG ZA YIBX IQZXK FTH T MASEXKXE MFXZFXK ZFXJ ZFAPYFZ GA ZAA T CAAVXE ICC KAPSE QAK ZFX FAKKTOCX JAPSY HIS ISE NAPCE GXX SA GTYSG AQ FTH OPZ SAM T MIG QKTYFZXSXE IYITS ISE KIS FAHX MTZFAPZ GZAUUTSY"

MODELS

SINGLE-CHAIN MCMC

1. Initialize a random mapping of letters
2. Use the transition matrix to calculate the initial log likelihood of the text decoded using the map
3. Create a proposal by sampling two letters to swap in the mapping
4. Calculate the decoded text's log likelihood:
   a. If the log likelihood meets the acceptance criteria, which favors maximizing likelihood with a degree of random rejection, accept the swap.
   b. Otherwise, retain the mapping from the previous step.
5. Run a few hundred iterations and time the full process as well as the time taken in each step.

EVOLUTIONARY MCMC

1. Initialize three different random mappings
2. Calculate their associated likelihoods
3. For each of the 3 chains, make ten proposals by sampling ten sets of two letters to swap
4. Calculate the decoded text's log likelihood and accept the swaps into the population of 30 if they meet the acceptance criteria outlined to the left
5. Cull the population, keeping only the three highest likelihood mappings (a.k.a. survival or recombination)
6. Run a few hundred iterations (generations) and time the full process and the time taken in each step.

1. HJB LEYWJBW VBYB CIWH E NSAU MNEKT JSYZXSAHEN NZAB HJBA EW Z WHSFFBP HS NSST EOHBY JZL EAP HJB YZGBY VEW CIWH EASHJBY JSYZXSAHEN NZAB ASH ABEYNR WS MYSEP ASY RBH WS MNEKT EAP HJB WTR VEW CIWH E YSV SO NSAU EAUYR YBP NZABW EAP PBAWB MNEKT NZABW ZAHBYLZDBP SA HJB BPUB SO HJB YZGBY Z KSINP OEZAHNR LETB SIH HJB SANR HVS MNEKT HJZAUW ZA ENN HJB FYSWFBKH HJEH WBBLBP HS MB WHEAPZAU IFYZUJH SAB SO HJBWB VEW HJB MBEKSA MR VJZKJ HJB WEZNSYW WHBBYBPNZTB EA IAJSSFBP KEWT IFSA E FSNBEA IUNR HJZAU VJBA RSI VBYB ABEY ZH HJB SHJBY E UZMMBH VZHJ WSLB KJEZAW JEAUZAU HS ZH VJZKJ JEP SAKB JBNP E FZYEHB HJB LEA VEW NZLFZAU SA HSVEYPW HJZW NEHHBY EW ZO JB VBYB HJB FZYEHB KSLB HS NZOB EAP KSLB PSVA EAP USZAU MEKT HS JSST JZLWBNO IF EUEZA ZH UEGB LB E HBYYZMNB HIYA VJBA Z HJSIUJH WS EAP EW Z WEV HJB KEHHNB NZOHZAU HJBZY JBEPW HS UEXB EOHBY JZL Z VSAPBYBP VJBHJBY HJBR HJSIUJH WS HSS Z NSSTBP ENN YSIAP OSY HJB JSYYZMNB RSIAU LEA EAP KSINP WBB AS WZUAW SO JZL MIH ASV Z VEW OYZUJHBABP EUEZA EAP YEA JSLB VZHJSIH WHSFFZAU

50. STE MADNTEN HEDE FKNS A CORL PCAIG TODUVORSAC CURE STER AN U NSOBBEY SO COOG AZSED TUM ARY STE DUJED HAN FKNS AROSTED TODUVORSAC CURE ROS READCW NO PDOAY ROD WES NO PCAIG ARY STE NGW HAN FKNS A DOH OZ CORL ARLDW DEY CUREN ARY YERNE PCAIG CUREN URSEDMUXEY OR STE EYLE OZ STE DUJED U IOKCY ZAURSCW MAGE OKS STE ORCW SHO PCAIG STURLN UR ACC STE BDONBEIS STAS NEEMEY SO PE NSARYURL KBDULTS ORE OZ STENE HAN STE PEAIOR PW HTUIT STE NAUCODN NSEEDEYCUGE AR KRTOOBEY IANG KBOR A BOCEAR KLCW STURL HTER WOK HEDE READ US STE OSTED A LUPPES HUST NOME ITAURN TARLURL SO US HTUIT TAY ORIE TECY A BUDASE STE MAR HAN CUMBURL OR SOHADYN STUN CASSED AN UZ TE HEDE STE BUDASE IOME SO CUZE ARY IOME YOHR ARY LOURL PAIG SO TOOG TUMNECZ KB ALAUR US LAJE ME A SEDDUPCE SKDR HTER U STOKLTS NO ARY AN U NAH STE IASSCE CUZSURL STEUD TEAYN SO LAVE AZSED TUM U HORYEDEY HTESTED STEW STOKLTS NO SOO U COOGEY ACC DOKRY ZOD STE TODDUPCE WOKRL MAR ARY IOKCY NEE RO NULRN OZ TUM PKS ROH U HAN ZDULTSEREY ALAUR ARY DAR TOME HUSTOKS NSOBBURL

75. SHE MADRHER PEDE JURS A LONT BLAZK HODICONSAL LINE SHEN AR I RSOFFEY SO LOOK AGSED HIM ANY SHE DIVED PAR JURS ANOSHED HODICONSAL LINE NOS NEADLW RO BDOAY NOD WES RO BLAZK ANY SHE RKW PAR JURS A DOP OG LONT ANTDW DEY LINER ANY YENRE BLAZK LINER INSEDMIXEY ON SHE EYTE OG SHE DIVED I ZOULY GAINSLW MAKE OUS SHE ONLW SPO BLAZK SHINTR IN ALL SHE FDORFEZS SHAS REEMEY SO BE RSANYINT UFDITHS ONE OG SHERE PAR SHE BEAZON BW PHIZH SHE RAILODR RSEEDEYLIKE AN UNHOOFEY ZARK UFON A FOLEAN UTLW SHINT PHEN WOU PEDE NEAD IS SHE OSHED A TIBBES PISH ROME ZHAINR HANTINT SO IS PHIZH HAY ONZE HELY A FIDASE SHE MAN PAR LIMFINT ON SOPADYR SHIR LASSED AR IG HE PEDE SHE FIDASE ZOME SO LIGE ANY ZOME YOPN ANY TOINT BAZK SO HOOK HIMRELG UF ATAIN IS TAVE ME A SEDDIBLE SUDN PHEN I SHOUTHS RO ANY AR I RAP SHE ZASSLE LIGSINT SHEID HEAYR SO TACE AGSED HIM I PONYEDEY PHESHED SHEW SHOUTHS RO SOO I LOOKEY ALL DOUNY GOD SHE HODDIBLE WOUNT MAN ANY ZOULY REE NO RITNR OG HIM BUS NOP I PAR GDITHSENEY ATAIN ANY DAN HOME PISHOUS RSOFFINT

85. SHE MARTHET WERE QUTS A LONG BLACK HORIYONSAL LINE SHEN AT I TSOPPED SO LOOK AFSER HIM AND SHE RIVER WAT QUTS ANOSHER HORIYONSAL LINE NOS NEARLZ TO BROAD NOR ZES TO BLACK AND SHE TKZ WAT QUTS A ROW OF LONG ANGRZ RED LINET AND DENTE BLACK LINET INSERMIXED ON SHE EDGE OF SHE RIVER I COULD FAINSLZ MAKE OUS SHE ONLZ SWO BLACK SHINGT IN ALL SHE PROTPECS SHAS TEEMED SO BE TSANDING UPRIGHS ONE OF SHETE WAT SHE BEACON BZ WHICH SHE TAILORT TSEEREDLIKE AN UNHOOPED CATK UPON A POLEAN UGLZ SHING WHEN ZOU WERE NEAR IS SHE OSHER A GIBBES WISH TOME CHAINT HANGING SO IS WHICH HAD ONCE HELD A PIRASE SHE MAN WAT LIMPING ON SOWARDT SHIT LASSER AT IF HE WERE SHE PIRASE COME SO LIFE AND COME DOWN AND GOING BACK SO HOOK HIMTELF UP AGAIN IS GAVE ME A SERRIBLE SURN WHEN I SHOUGHS TO AND AT I TAW SHE CASSLE LIFSING SHEIR HEADT SO GAYE AFSER HIM I WONDERED WHESHER SHEZ SHOUGHS TO SOO I LOOKED ALL ROUND FOR SHE HORRIBLE ZOUNG MAN AND COULD TEE NO TIGNT OF HIM BUS NOW I WAT FRIGHSENED AGAIN AND RAN HOME WISHOUS TSOPPING

90. THE MARSHES WERE QUST A LONG BLACK HORIVONTAL LINE THEN AS I STOPPED TO LOOK AFTER HIM AND THE RIZER WAS QUST ANOTHER HORIVONTAL LINE NOT NEARLY SO BROAD NOR YET SO BLACK AND THE SKY WAS QUST A ROW OF LONG ANGRY RED LINES AND DENSE BLACK LINES INTERMIXED ON THE EDGE OF THE RIZER I COULD FAINTLY MAKE OUT THE ONLY TWO BLACK THINGS IN ALL THE PROSPECT THAT SEEMED TO BE STANDING UPRIGHT ONE OF THESE WAS THE BEACON BY WHICH THE SAILORS STEEREDLIKE AN UNHOOPED CASK UPON A POLEAN UGLY THING WHEN YOU WERE NEAR IT THE OTHER A GIBBET WITH SOME CHAINS HANGING TO IT WHICH HAD ONCE HELD A PIRATE THE MAN WAS LIMPING ON TOWARDS THIS LATTER AS IF HE WERE THE PIRATE COME TO LIFE AND COME DOWN AND GOING BACK TO HOOK HIMSELF UP AGAIN IT GAZE ME A TERRIBLE TURN WHEN I THOUGHT SO AND AS I SAW THE CATTLE LIFTING THEIR HEADS TO GAVE AFTER HIM I WONDERED WHETHER THEY THOUGHT SO TOO I LOOKED ALL ROUND FOR THE HORRIBLE YOUNG MAN AND COULD SEE NO SIGNS OF HIM BUT NOW I WAS FRIGHTENED AGAIN AND RAN HOME WITHOUT STOPPING

92. THE MARSHES WERE JUST A LONG BLACK HORIZONTAL LINE THEN AS I STOPPED TO LOOK AFTER HIM AND THE RIVER WAS JUST ANOTHER HORIZONTAL LINE NOT NEARLY SO BROAD NOR YET SO BLACK AND THE SKY WAS JUST A ROW OF LONG ANGRY RED LINES AND DENSE BLACK LINES INTERMIXED ON THE EDGE OF THE RIVER I COULD FAINTLY MAKE OUT THE ONLY TWO BLACK THINGS IN ALL THE PROSPECT THAT SEEMED TO BE STANDING UPRIGHT ONE OF THESE WAS THE BEACON BY WHICH THE SAILORS STEEREDLIKE AN UNHOOPED CASK UPON A POLEAN UGLY THING WHEN YOU WERE NEAR IT THE OTHER A GIBBET WITH SOME CHAINS HANGING TO IT WHICH HAD ONCE HELD A PIRATE THE MAN WAS LIMPING ON TOWARDS THIS LATTER AS IF HE WERE THE PIRATE COME TO LIFE AND COME DOWN AND GOING BACK TO HOOK HIMSELF UP AGAIN IT GAVE ME A TERRIBLE TURN WHEN I THOUGHT SO AND AS I SAW THE CATTLE LIFTING THEIR HEADS TO GAZE AFTER HIM I WONDERED WHETHER THEY THOUGHT SO TOO I LOOKED ALL ROUND FOR THE HORRIBLE YOUNG MAN AND COULD SEE NO SIGNS OF HIM BUT NOW I WAS FRIGHTENED AGAIN AND RAN HOME WITHOUT STOPPING

RESULTS

- I encrypted the text 5 times and decrypted it using each method to determine its average efficacy. The results of each run in number of steps taken, total time taken, and average step time are displayed in the following table:

Run	MCMC steps	eMCMC steps	MCMC total time (secs)	eMCMC total time (secs)	MCMC avg step time (secs)	eMCMC avg step time (secs)
1	114	93	145.96	96.38	1.29	1.02
2	53	103	60.71	109.69	1.15	1.05
3	98	108	106.19	113.23	1.09	1.02
4	75	86	65.73	88.26	0.88	1.03
5	79	73	95.70	93.87	1.22	1.28
Avg	**83.8**	**92.6**	**94.86**	**100.29**	**1.12**	**1.08**

- In some cases, the algorithm did not converge - this occurred three times with the single-chain Markov Chain Monte Carlo method, and once with the evolutionary Markov Chain Monte Carlo method. These runs were not included above but I think they should be considered in analyzing the results.

ANALYSIS & CONCLUSION

- Counterintuitive results from steps, step times, and total time
- Generally, the evolutionary MCMC method performed similarly to the traditional single chain MCMC method, likely due to:
  - Computing resources
  - Heuristic parameter optimization
  - Convergence frequently tied to initial mapping proposal
- With a limited sample there are no clear efficiencies in using an evolutionary Markov Chain Monte Carlo approach relative to a singe chain Markov Chain Monte Carlo approach
- Still, Evolutionary MCMC shows promise in reducing likelihood of non-convergence

NEXT STEPS AND FUTURE RESEARCH

Computing in the cloud:

- More runs
- Controlled computing conditions
- Test a wider range of eMCMC parameter values:
    - Number of chains
    - Number of proposals
- Parallelization

REFERENCES

- Agarwal, Rahul. (2019). Applications of MCMC for Cryptography and Optimization. (https://towardsdatascience.com/applications-of-mcmc-for-cryptography-and-optimization- 1f99222b7132)
- Alsharafat, Wafa. (2015). Evolutionary genetic algorithm for encryption. 10.1109/ICCIC.2014.7238559. (https://www.researchgate.net/publication/285612552_Evolutionary_genetic_algorithm_for_encryptio n)
- Byoung-Tak Zhang, Dong-Yeon Cho, System identification using evolutionary Markov chain Monte Carlo, Journal of Systems Architecture, Volume 47, Issue 7, 2001, Pages 587-599, ISSN 1383-7621, https://doi.org/10.1016/S1383-7621(01)00017-0. (http://www.sciencedirect.com/science/article/pii/S1383762101000170)
- Chen, Jian & Rosenthal, Jeffrey. (2012). Decrypting classical cipher text using Markov chain Monte Carlo. Statistics and Computing. 22. 397-413. 10.1007/s11222-011-9232-5. (http://probability.ca/jeff/ftpdir/decipherart.pdf)
- Diaconis, Persi. (2009). The Markov Chain Monte Carlo Revolution. Bulletin of the American Mathematical Society. 46. 179-205. 10.1090/S0273-0979-08-01238-X.
- (https://www.statweb.stanford.edu/~cgates/PERSI/papers/MCMCRev.pdf)
- Dickens, C., & Mitchell, C. (2003). *Great Expectations*. (https://www.gutenberg.org/files/1400/1400-h/1400-h.htm)
- Drugan, Madalina & Thierens, Dirk. (2003). Evolutionary Markov Chain Monte Carlo. 63-76. (https://dspace.library.uu.nl/bitstream/handle/1874/24000/drugan_03_evolutionarymarkov.pdf?seque nce=1)
- Garg, Poonam. (2010). Evolutionary Computation Algorithms for Cryptanalysis: A Study. International Journal of Computer Science and Information Security, Vol. 7, No. 1. (https://arxiv.org/pdf/1006.5745)
- Landgraf, Andrew J. (2013). Text Decryption Using MCMC. (https://alandgraf.blogspot.com/2013/01/text-decryption-using-mcmc.html)
- Tolstoy, L., Pevear, R., & Volokhonsky, L. (2007). *War and Peace*. (http://www.gutenberg.org/files/2600/2600-h/2600-h.htm)