1. Replacing gradient with stochastic gradient will introduce noise. We try to mitigate noise with various methods (eg: friction)
   1. What if we try using nn models to predict noise and it’s covariance and then use that as friction to cancel noise from stochastic gradients.
2. Why do we random sample momentum?
3. What is testing hypothesis, T & Z statistic?
4. Use chunked hyper networks to generate MCMC iterates (predict iterates instead of simulating)
5. Need n2 random steps to move n steps in a given direction (from law of large numbers theorem ? or something similar)
6. Study feature space sim/diff for models trained with MLP/MAP vs SG-MCMC iterates
7. Should start MCMC at MAP for faster sampling of the target distribution
8. Need to mention what mixing rates are, coverage of target distribution, detailed balance, correlated samples.
9. What if we distill MCMC iterates in 1 model?
10. SG-MCMC gives samples from target dist which are not necessarily at the mode. In contrast, ensemble and SWA give samples from the model.
    1. Compare performance
11. SGD updates weights by multiplying gradient of weights with learning rate VS SG-MCMC simulates SDE to update weights
12. Metropolis-Hastings method does not need conditional distributions VS Gibbs sampling which needs conditional distributions which is not always possible
13. Bayes is better than Frequentist method
    1. One reason no complexity issue, in Bayes, use maximum complexity model and expectation of prediction will handle overfitting
14. Hamiltonian Monte Carlo is faster than Metropolis-Hastings, since there is no random sampling.
15. Benchmark PSGLS with all quasi Newtonian methods (Adam, Nadam, ...) and compare with Riemannian geometry methods
    1. mass matrix acts as preconditioning or Riemannian metric
16. Covariance matrix has to be positive semidefinite

**Timeline**

1949

* [The Monte Carlo Method](https://www.semanticscholar.org/paper/The-Monte-Carlo-Method-Metropolis-Ulam/b9ff7f1bf51982dab33ff1e918f9fac9e7fe21dd)
* Stanisław Marcin Ulam came up with monte carlo when on the manhattan project
* Worked with metropolis
* Having fast computers played a big role

1953

* [Equation of state calculations by fast computing machines](https://www.semanticscholar.org/paper/Equation-of-state-calculations-by-fast-computing-Metropolis-Rosenbluth/f6a13f116e270dde9d67848495f801cdb8efa25d)
* Metropolis introduced the metropolis algorithm, first mcmc method
* Not sure who did the work

1978

* [Brownian dynamics as smart Monte Carlo simulation](https://pdfs.semanticscholar.org/9873/12496fb243bb56518fb9ecfa051e9df8f35d.pdf?_ga=2.257652872.2011050715.1592016218-642035832.1590635133)
* Introduced dynamics based sampling instead of random sampling in original method
* did not think of switching between vanilla and brownian to make the hybrid monte carlo method

1987

* [Hybrid Monte Carlo](https://www.sciencedirect.com/science/article/abs/pii/0920563288901570?via%3Dihub)

2011

* Bayesian Learning via Stochastic Gradient Langevin Dynamics
* Allows use of stochastic gradients to in Langevin Dynamics
* Langevin dynamics don’t use hamiltonian dynamics equation