https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers

**Chapter 1. Introduction**

Points:

1. Bayesian Inference vs Frequentist
2. Discrete Distribution (*Poisson-distributed)*
3. Continuous Distribution(*Exponential-distributed*)
4. PyMC3 (*To sample the posterior samples*)

Examples:

1. *Mandatory coin-flip example:* Toss coins (beta distribution as prior and Bernoulli as likelihood)
2. *Bug, or just sweet, unintended feature?*
3. *Inferring behaviour from text-message data:* User's text-messaging habits have changed over time, either gradually or suddenly, the point of modeling is to judge whether the parameter of Poisson distributed (lambda ) has changed or not after time tau.

References: .

**Chapter 2. A little more on PyMC3**

Points:

1. Model Context and PyMC3 Variables *(e.g. Stochastic Variables & Deterministic Variables)*
2. Something about Theano *(Since PyMC3 is based on Theano)*
3. Including Observations in The Model *(e.g. observed = data)*
4. Artificial Dataset to Make Predictions and Test the Appropriateness
5. An algorithm for human deceit (*Binomial-distributed)*
6. More PyMC3 Tricks (*Arrays of PyMC3 variables*)
7. Normal Distributions (*To sample the posterior samples*)
8. Separation Plots (*A novel data-viz approach to logistic regression*)

Examples:

1. *Bayesian A/B testing:* Effectiveness of drug A vs drug B (True *p* is Uniform distributed, N trials with each trial is a Bernoulli distribution)
2. *Cheating among students:* **Privacy Algorithm**. Use the binomial distribution to determine the frequency of students cheating during an exam. *(To find the proportion p of the cheaters)*

Two methods (samples & deterministic function).

1. *Challenger Space Shuttle Disaster:* 1) At temperature *t*, what is the probability of a damage incident?(*Logistic Function* is used to describe the probability & *Normal Distribution* to describe hyper parameter *alpha* and *beta* & *Bernoulli* *Distribution* to connect the probability *p* and the *(0,1)* result) .2) Is our model appropriate? To generate artificial data set to compare with observed dataset.(Sample from the *posterior distribution* with *stochastic variable*) ( *Draw graph* or use *Bayesian p-values* to assess how good the model fits)

References: .

**Chapter 3. Opening the black box of MCMC**

Points:

1. The Bayesian landscape (Basically how to draw *contour plot* and *3-D(plt.imshow)*).
2. Exploring the landscape using the MCMC *(MCMC returns samples from the posterior distribution, not the distribution itself) .*
3. Algorithms to perform MCMC (1.*Start at current position.2.Propose moving to a new position (investigate a pebble near you).3.Accept/Reject the new position based on the position's adherence to the data and prior distributions (ask if the pebble likely came from the mountain).4.a) If you accept: Move to the new position. Return to Step 1.b) Else: Do not move to new position. Return to Step 1. 5After a large number of iterations, return all accepted positions).*
4. Using MAP to improve convergence./Prediction
5. Speaking of the burn-in period (*As one does not know when the chain has fully converged, a good rule of thumb is to discard the first half of your samples, sometimes up to 90% of the samples for longer runs*).
6. Diagnosing Convergence (*How to Calculate* *Autocorrelation*) / *(This does not imply that a converged MCMC has low autocorrelation. Hence low autocorrelation is not necessary for convergence)*
7. Thinning.( Only returning to the user every *nth* sample)
8. PyMC3 Plot(*traceplot / plot\_posterior/ autocorrplot*)
9. Useful tips for MCMC*(1.Intelligent starting values. 2.Carefully choose priors)*

Examples:

1. *Unsupervised Clustering using a Mixture Model:* 1.For each data point cluster 1 with probability *p*, else choose cluster 2. 2. Draw a random variate from *Normal Distribution* with parameter *$\mu\_i$* and *$\sigma\_i$* where *I* was chosen step 1. 3 Repeat. (*PyMC3 Categorical stochastic variable*).

**2 ways to plot (*plt.imshow* & *mpl.colors.ListedColormap (colors)***

**Important: Don't mix posterior samples**

References: .

**Chapter 4.** **The Greatest Theorem Never Told**

Points:

1. The Law of Large Numbers.
2. The Disorder of Small Numbers.
3. How to Compute the Lower bound for a Beta Prior and Binomial Likelihood(*e.g. example 4*).
4. Plot (*plt.errorbar*()*).*
5. Extension to Starred Rating Systems(*1-5 stars to rate instead of upvote/downvote*).

Examples:

1. *A diagram of the Law of Large numbers in action for three different sequences of Poisson random variables & the rate of convergence to expectation.*
2. *Aggregated geographic data(Average height vs. County Population)*
3. *Kaggle's U.S. Census Return Rate Challenge(*predict the census letter mail-back rate of a group block, measured between 0 and 100, using census variables (median income, number of females in the block-group, number of trailer parks, average number of children etc.*)*
4. *How to order Reddit submissions(estimate of the true upvote ratio. Prior ~ Uniform(0,1), then use Binomial to connect observed data with true p, then sort with 5% quantile)*

References: .

**Chapter 5. Loss Functions**

Points:

1. Loss Functions (*some popular loss functions such as squared-error, absolute-loss, zero-one loss and log-loss etc.*)
2. Loss Functions in the Real World*(Unknown true parameter, calculate expectation of loss function over a large number of posterior samples)*
3. Shortcuts*(some useful start values)*
4. Machine Learning via Bayesian Methods*（Frequentist Method: best precision about all possible parameter. Machine Learning Method: best prediction among all possible parameters）*

Examples:

1. *Optimizing for the Showcase on The Price is Right (The Auction Game: very good example for to introduce the loss function) /Minimizing the losses(scipy.optimize.fmin())*
2. *Financial prediction (Prediction about the daily return of given stock) /Least-squares prediction vs. Bayes action prediction. (OLS vs Bayesian Linear regression)(****The minimum of the expected loss is called the Bayes action.****)*
3. ***Kaggle contest on Observing Dark World(****Super great example, better read it with heart)*

References: .

**Chapter 6. Priorities**

Points:

1. Subjective vs Objective Priors *(objective priors: allow the data to influence the posterior the most/ subjective priors:* *allow the practitioner to express his or her views into the prior)*
2. Empirical Bayes *(Empirical Bayes is a trick that combines frequentist and Bayesian inference. However, the author holds the opinion that empirical bayes should not be used casually since it has the risk of using same data twice).*
3. Useful priors to know about *(The Gamma distribution,* *The Wishart distribution, The Beta distribution).*
4. Eliciting expert prior *(Aids speeds of MCMC convergence/More accurate inference/Express our uncertainty better).*
5. Trial roulette method*(A method to find good posterior)*
6. Protips for the Wishart distribution *(The number of unknowns is too large. For a N\*N matrix, 0.5\*N\*(N-1) unknowns need to be estimated). /Several methods of supremacy are given.*
7. Conjugate Priors *(Issues: 1.* *The conjugate prior is not objective.2.* *For larger problems, involving more complicated structures, hope is lost to find a conjugate prior.)*
8. Jefferys Priors *(Google it!)*
9. Bayesian perspective of Penalized Linear Regressions*(Ridge regression/Lasso regression vs. MAP, when use MAP the posterior should be have same form with prior(See reference) )*

Examples:

1. *Bayesian Multi-Armed Bandits (N slot, select the slot with highest probability to obtain the prize)/Very nice example (Beta as prior and binomial as likelihood, combine with loss function).*
2. *Stock Returns (4 stocks(AAPL,TSLA,AMZN,GOOG) with different priors/ Normal distribution to describe* ***mean*** *and* ***std*** *, WishartBaetlett to describe the covariance matrix)/* ***easy way to convert covariance matrix to correlation matrix.***

References: .