

blog post ideas

YOU ARE (BECOMING) AN ML EXPERT

heteroskedastic linear regression with tensorflow probability

bayesian a/b testing with tensorflow

Achieving and Demonstrating Deep Understanding of Bayesian Statistics, ML in production and Business Intelligence

network and pitch this angle and find the perfect job

write blog posts and do projects that are fun and/or demonstrate business intelligence/web development/useful for Basyn

prioritize what is worthwhile to display or has clear strategic benefits

Find relevant materials in course lecture/assignments/notes

1. make notes for self and blog posts
2. low hanging fruit notebooks and projects
 1. assess how much material is available that would be impressive and worth effort to display

Go through Bayesian Programming for Tensorflow Probability

1. Take extensive notes for yourself and blog posts
2. Do low hanging fruit example notebooks and

Go through Bayesian Data Analysis

1. assess topics I want to deep dive into
2. Take extensive notes for yourself and blog posts
3. Do low hanging fruit example notebooks and

Network, supplement, seek advice, put self in situations for compensated competence

1. ...? (attend webinars, watch relevant youtube videos, blogs, ask Elizabeth for pointers)
2. find an expert who does bayesian business intelligence/ml production code get their advice, ask them to review your materials
3. be aggressive and confident in the value you bring in your niche

Logic of Science Notes

precourse

pascal fermat's advice for 'fair odds' for gambling games..what math did it lead to?

- what's the correct decision to make during a gambling game, given you dont know what's gonna happen
- developed into frequentist tools

tools from measure theory (kolmogorov's contributions)

- random variable, expected value, law of large numbers, central limit theorem, markov process, random walk
- advanced - sigma algebra, filtration, brownian motion, stochastic calculus

connection of these fundamental ideas of probability with data questions in the real world

probability is a language of ideas, used all the time, in the news, 'chances are...', 'high probability of...', 'odds are against...'

conflict in ways people have understood probability, word is ambiguous

- correct legal judgments vs fair odds
- plausibility of evidence or belief vs frequency and fair odds (the major conflict in the interpretations of probability)
 - what jaynes hopes to settle in the book

statistics

- first statisticians were measurements of the state like census
- always understood to have a measurement error
- sampling variation
- physicists, astronomy questions, had data to measure, how to combine for accurate
 - techniques developed by laplace
 - linear regression
 - theory of distribution of errors
- statistical inference
 - given we have all these measurements with some error, how do we infer the true state of being of the phenomenon
 - fisher vs pearson, 2 competing schools developed in 20th century, developed orthodox techniques
 - unbiased estimator, confidence interval, p value, type 1 and 2 error, hypothesis test
 - based on pascal and fermats early work in gambling rates
 - bayesian inference
 - how our probability estimates gets updated with new data
 - richard price in honor of bayes developed the theory
- still a tension in use of probability
 - kolmogorov's math doesnt resolve, all it tells us is how do probabilities combine and aggregate, once sample space and probability measure is defined
 - how can these be mapped onto the real world for practical understanding of probability

- common usage of probability in language vs technical orthodox
 - probability of future events for frequentists means parallel universes, but what differences between these universes lead to measured differences?
- subjectivist view
 - probability means a degree of belief a statement is true, common usage in the language
 - tensions for uptake in 'objective' realist science applications

this helps break apart the confusion I felt starting in probability, as people fluidly jump from one to one

- kolmogorov's toolset of marginal and conditional probability favors frequentist
- vs data science applications building on bayes

to help solve this tension, jaynes considers probability as an extension of logic

- the apparent separation between probability and statistics will dissipate
- for situations of reasoning with incomplete information
- axioms of kolmogorov are derivable from more fundamental conditions of consistent logical reasonings
 - define the fundamental qualitative terms of logic we agree on, then derive the probability theory
- proposes to elevate bayes theorem to the same status of Aristotelean logic (his syllogism for example)
 - even bayes rule as a superset of this logic, with syllogism as a special case
 - bayes rule as the test for the statistical 'ad hoc' and measured as how consistent they are
 - a thorough critique of standard statistics (fisher school)
 - the approach of you start with a null hypothesis (no difference in our control vs treatment)
 - statistics asks what probabilities (frequencies) of the data due we observe, how frequently we expect them in a certain range
 - jaynes says no its the other way around: given we have a set of data, what probability we associate to the hypothesis that we actually care about, what we test about
 - ie, instead of hypothesis to data, we go from data to hypothesis
 - how again is this considered more subjective??? priors seem less subjective than hypothesis
 - book amounts to a demolition of the last hundred years of statistics
 - jaynes attempts a better grounding of subjectivist position
 - set of background assumptions/beliefs of probability can be incorrect, people can be deceived
 - broader understanding of probability that includes plausibility of a hypothesis, allows for sound logical statistical inference, when traditional techniques of stats breaks down
 - so much of modern science development depends on p tests, the stats could be holding it back
- 400 years of probability theory resolved, aristotelean logic around for thousands of years extended to new classes of problems for reasoning logically about things that couldnt be reasoned about before, a hundred years of statistics dismantled

- ambitious book, broad range of concepts, monumentally impressive

ch1 Plausible Reasoning

main influence of the book

- r.t cox
 - consistency theorem
- polya

in this chapter, we consider the qualitative criteria we look for when reasoning with incomplete information, next ch will be quantitative terms

most reasoning is not deductive, we see clouds and think it may rain, we feel a pain and we go to a doctor for treatment

deductive reasoning is a strict usage similar to Aristotle's syllogism, a premise '

$$\begin{array}{l} (P \rightarrow Q) \\ P \\ \therefore Q \end{array}$$

{

if p is true then q is true,

an observation (minor-premise) 'p is true',

conclusion 'q is true'

}

Logically sound explanations of phenomena can't be discounted just because they're highly unlikely

deductive reasoning vs plausible inference

plausible inference could have a similar premise

$$\begin{array}{l} (P \rightarrow Q) \\ Q \\ \therefore P \text{ is more plausible} \end{array}$$

aka if p then q, q, p is more plausible',

personal observation:

- peirce isn't considered a bayesian, but this resonates with pragmatist notion of truth of inexhaustible lists of consequences of an organizing concept or idea, where new observations/experiences update the idea of the concept and whether or not the observed phenomena is described by the concept
- e.g. hardness means you can rub a rock on it and it won't scratch it, takes a rock and rubs it on a diamond, a diamond is hard

observation: not that p is true but 'q is true',

conclusion, p is 'more plausible'

whether evidence makes something more likely has to do with other hypotheses that may explain it better, and what prior probabilities we put on those hypotheses

in next chapter, finding

$$\frac{P(A + B)}{P(AB)}$$

grounds all the truth table of basic logic

Thought Experiment to Introduce Paradigm shift

How do we normally talk about probability (there's a 40% chance it's gonna rain)

A measure of belief

How frequentists view probability, a rate/frequency of an event occurring

this requires multiple universes to talk about forecasted events

that's absurd, but actually what classical statistics of the last 100 years and therefore most mainstream scientific studies require (think p value)

this is problematic bc if you get new information/data you have to start over, rather than update your model adaptively as new data comes in (this is why practically bayesian statistics is already being used far and wide in data science and business/operations studies, meanwhile scientific journals are beholden to p values, and arguably slow down the march of scientific progress where only 2 states can be compared all the time, and it favors the status quo heavily, and new data gets thrown out all the time because of the statistical model rather than throwing out the statistical model)

the solution is deductive logic is actually a subset of plausible reasoning/inference

instead of aristotles syllogism as the prime case of logic

a bird has wings

jerry the seagul has wings

jerry the seagul is a bird

it's actually a subset of plausible reasoning, where we

a bird has wings

we see jerry the seagul has wings

we believe jerry could be a bird, there's a small chance he's something else, if we get new information that he has a beak and lays eggs, we would be more sure he's a bird

paradigm shifts in science historically happen, not super quickly, but they gain a lot of traction as new beliefs solve old outstanding problems, as is happening in physics and neuroscience with bayesian methods, and in this day in age as many scholars and academics take to other channels for their teaching and research, a paradigm shift may not be so slow as we wait for traditional gatekeepers of knowledge to adjust policies (looking at biology and p values)

thoughts?

Topics of Note

Fundamental

frequentist/classical tools boil down to one framework [Alleny Downey post](#)

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

explain bayes theorem in terms of conditional and marginal probability

mean, standard deviation/variance

- standard deviation is the square root of variance

Moving between model and data

- "This is related to the serious scientific challenge posed at the beginning here: if you know the underlying parameters/model, you can figure out the distribution and the result, but often we have only the experimental result and we're trying to figure out the most appropriate model and parameters.

It is Bayes' Theorem that lets us move between these." Eric Ma

Basic Bayesian Workflow:

- Statistical Computing Folk Theorem: "If the model takes a long time to fit, your model is poorly specified"
1. Build a model
 2. Diagnose the model by looking for 'divergences'
 3. Reparameterize the Model

distributions, their stories & shapes [1](#) [2](#)

tools from measure theory (kolmogorov's contributions)

- random variable, expected value, law of large numbers, central limit theorem, markov process, random walk

[cross entropy loss](#)

[rmsprop mini-batch optimizer](#)

known knowns

uncertainty quantification (UQ) [wiki](#)

- what are the things we cant measure precisely or measure at all (where probabilistic models comes in)
- how likely certain outcomes are if some aspects of the system are not exactly known
 - e.g. calculating the impact of a car crash on a person depends on a wide range of factors including manufacturing specifics of each vehicle, even how tightly the screws were tightened, that impact the resulting action
- sources of uncertainty:
 - parameter uncertainty (unknown inputs to model),
 - parametric variability (variability of input variables of the model),
 - structural uncertainty (ie model inadequacy. bias, discrepancy, from lack of knowledge of underlying physics),
 - algorithmic uncertainty (ie numerical, discrete: from numerical errors and approximations in impl. model),

- experimental uncertainty (observation error, variability of experimental measurements, even in repeated in same setting for all inputs/variables)
- interpolation uncertainty (lack of available data from model and/or experimental measurements)
- 2 classifications of uncertainty, uncertainty quantification aims at reducing epistemic uncertainties to aleatoric uncertainties [in ml \(tutorial\)](#)
- **the wiki seems to be written by people with different definitions and its still unclear to me, best see how people use this in practice**
 - aleatoric
 - from latin alea (dice, eg game of chance)
 - aka statistical/irreducible uncertainty, commonly referred to as known unknowns
 - a coin flip, no how many seen in the past, we can't predict what the flip will be in the future
 - statistical uncertainty, reps unknowns that are different each time running the experiment,
 - eg firing the same arrow, altitude, direction and velocity will not hit the same point due to random and complicated vibrations of the arrow shaft, and cant be known sufficiently
 - just because its currently unmeasurable doesnt mean it isnt posible with better measurement (if it was unmeasurable, it would be considered epistemic)
 - epistemic uncertainty
 - greek word episteme (knowledge)
 - systemic/reducible uncertainty, unknown unknowns
 - could be known, but not known in practice, may be
 - need another source on this besides wiki

standard linear regression assumes homoscedasticity (consistent variation of error ie distance from regression line - the means of y - to y itself), most stastical processes underestimate the variation in error in order to preserve the function of hypothesis testing [more](#)

- traditionally, the tools werent great one might take a log transform, maybe a square root, but you want is to tell the computer fit the mean of y as an increasing function of x, but also fit the variance of y as an increasing function of x (but theres still not a lot of clean ways to do that), tf probability helps solve this
 - most regression analyses omit a qunatiification of the uncertainty in the predictions (its complex)
 - a solution is to write the regression model as $P(y | x, w)$
 - y, the probability distribution of labels given
 - x, the inputs and
 - w, some parameters
 - fit this model to data by maximizing the probability of the labels, or equivalently, minimizing the negative log likelihood loss: $-\log P(y|x)$
- in python: `negloglik = lambda y, p_y: -p_y.log_prob(y)`
- this loss function can be used with DistributionLambda, bc keras passes the output of the final layer of the model into the loss function, and the models in this example all return distributions
 - we can use a variety of standard continuous and categorical and loss functions with this model of regression

- e.g. mean squared error loss for continuous labels means $P(y | x, w)$ is a normal distribution w/ a fixed scale (standard deviation)
- e.g. cross entropy loss for classification means $P(y | x, w)$ is the categorical distribution

```

• import tensorflow as tf
import tensorflow_probability as tfp
tfd = tfp.distributions
import numpy as np

# Basic TF Linear Regression with no uncertainty
#Build model
model = tf.keras.Sequential([
    tf.keras.layers.Dense(1),
    tfp.layers.DistributionLambda(lambda t: tfd.Normal(loc=t, scale=1)),
])

# Do Inference
model.compile(optimizer=tf.optimizers.Adam(learning_rate=0.01),
loss=negloglik)
model.fit(x, y, epochs=1000, verbose=False);

# Predict
[print(np.squeeze(w.numpy())) for w in model.weights];
yhat = model(x_tst)
asser isinstance(yhat, tfd.Distribution)

```

- aleatoric uncertainty (known unknowns)
 - e.g. a coin flip's outcomes
 - assume the variability of the data has a known functional relationship to the value of x
- epistemic uncertainty (unknown unknowns)
 - could be reduced with more data
 - to capture this uncertainty, replace the `tf.keras.layers.Dense` w/ a `tfp.layers.DenseVariational` layer
 - represents the variational posterior $Q(w)$ over the weights to repr the uncertainty in their values
 - regularizes $Q(w)$ to be close to the prior distr $P(w)$, which models uncertainty in the weights before looking into the data
 - using both $P(w)$ and $Q(w)$, Empirical Bayes or Type-II Maximum Likelihood,
 - this method means dont need to specify the location of the prior for the slope and intercept parameters, which is tough w/o prior knowledge, prevents a bad prior from messing up posterior, however you dont get all the regularization benefits over the weights, with some prior info or sophisticated prior, a non trainable prior is possible
 - to model $Q(w)$, we use a multivariate normal distribution for the variational posterior with a trainable diagonal covariance matrix centered on a trainable location
 - to model $P(w)$, we use a standard multivariate normal distribution for the prior with a trainable location and fixed scale
 - `DenseVariational` defines an ensemble of models, the model predicts a different value each time
- the rest of the code is found in this [notebook](#)

log likelihood

- negative log-likelihood loss
 - $-\log P(y|x)$
 - relation to inverse log relation of absorbance and transmittance?

linear mixed effect model

Project:

Simulate different distributions under different conditions

Bayesian Craps Advisor

- pascal fermat's advice for 'fair odds' for gambling games..what math did it lead to?

Blog Post-ify this [video](#)

Resources

bayesian methods for hackers w/ tensor flow probabiilty

Strategic

adding cost/loss functions to a bayesian posterior ('hybrid of statistic and mechanistic model of the world' Eric Ma), useful for modeling the positive events but not so much the cost of implementing..?

Multilevel/hierarchical modeling

- https://docs.pymc.io/notebooks/multilevel_modeling.html

mixture model

shrinkage

- wild estimates being shrunk to estimation means
- modeling an underlying distribution that influences an array of dists

Bayesian T-Test

- Bayesian Estimation Supersedes the t-test (BEST), paper by Kruschke
- freq vs bayesian
- approximating a probability to a normal dist is an approximation of an approximation (scipy 19 lecture), p is bound, normal is -inf to +inf

adjusting your prior according to prior knowledge

- baseball players have a low batting average, not a uniform beta distribution but between 1 and 4, or 1 and 40 for much narrower, 100 to 400, the larger numbers more narrow

Bootstrapping vs Bayes

- [stackoverflow article](#)

[categorical distribution](#)

- often called (inaccurately) multinomial

variational inference vs mcmc

Advanced Topics

Bayesian Neural Net/probabilistic deep learning

- eric ma pydata 2017 [video](#)
 - [notebook](#)
 - deep learning is nothing more than compositions of functions on matrices and their operations
 - bayesian deep learning is grounded on learning a probability distribution for each parameter, rather than point estimates
 - really good illustrations explaining, linear reg, logistic reg, neural net, and bayesian neural net
 - you must get the likelihood function right
- Radford kneel
 - you dont need train test split for bayesian nn

[TF Blog](#)

[Data-Driven Dynamical Systems w/ ML](#)

- [book](#)

Probability Theory As Extended Logic ET Jaynes

- [video series](#)
- book

bayesian decision theory

David MacKay

Yarin Gal

Resources Digested

[Tensorflow Probability Deep Dive](#) Viewed on 7-24 (worth revisiting)

[Bayesian Data Science: Probabilistic Programming SciPy Evan Ma](#) Viewed on 7-25-20

[Arun Subramaniyan](#) Viewed on 7-25-20

- *"probability means solving problems in a fundamentally different way..encourage the practical and philosophical"* 6:55

Intro to TensorFlow Coursera Course 1 Viewed on 7-25-2020

UW ML Research [page](#)

Dynamic Bayesian Network [wiki talk](#)

General Resource

[latex guide](#)

[edward2 github](#)

MIT OCW Intro to Prob and Stat (w Bayes focus) [materials](#)

TODO

Blog Post on Melinda Thielbar's [notebook](#)

Modern Bayesian Workflow PyMC [notebook](#)

Reach out to Jen Bricker jen@datacamp.com for a course proposal Bayesian Data Analysis with Python course

- what are the pre requisites?
- what libraries should be included (tensorflow probability is my preferred and would have your platform stand out)
- how does the lesson content process work
- what budget do you have for this project?