Eric Rupert
Bayesian Inference
Dec 18 2017

1. Introduction

Data analysis generally requires use of statistical techniques to forecast, generalize ("nowcast"), and determine level of confidence in a process or result. A major tool for these purposes (and others) is Bayesian Inference. Bayesian Inference allows for assumptions to be used to predict outcomes, based on empirically collected data and is particularly useful when little data can be collected on the desired process (preventing direct frequentist analysis of the process).

The traditional approach to statistical analysis is the Frequentist approach. This view sees statistics as the outcomes due to a large number of trials. For example, a coin is considered fair if it has been flipped a million times and came up heads approximately five hundred thousand of those times. This approach assumes the parameter set (typically denoted Θ) is fixed.

Conversely the Bayesian approach sees not the generative function as fixed, but rather the collected data (D). This way a search can be made over D to find Θ. This method, then allows for the model to be updated as more data is collected. It also allows us to test our beliefs in fairness of systems. For example, if we believe an urn is fair, then half of the time a red ball will be drawn and half of the time a black ball will be drawn, but we can also assume some unfairness (it has 2 black balls for every red ball, for example), and find a degree of certainty of how it is unfair.

Bayesian Inference has gained acceptance recently, and is used in a broad variety of tasks from understanding a test result to setting the speed of a self driving car. A mathematical background will be presented in section two and examples (with MATLAB) will be presented in section three.

2. Background on Bayesian Inference

Bayesian Inference comes from Bayes' Rule, which is derived from the relationship between joint, marginal, and conditional probabilities. A marginal probability is the probability of an event happening, $P_A(a)$, where A denotes the parameter space, and a denotes the value within the space. This marginal probability is independent of other events. A joint probability is the probability of two events a and b occurring simultaneously $P_{A,B}(a,b)$. A conditional probability is the probability that event b will occur, given that event a has already occurred $P_{A,B}(a|b)$.

$$P_{A,B}(a|b)=\frac{P_{A,B}(a,b)}{P_B(b)} \tag{1}$$

likewise,

$$P_{A,B}(b|a)=\frac{P_{A,B}(a,b)}{P_A(a)} \tag{2}$$

Rearranging equation 1 and substituting into equation 2, we have

$$P_{A,B}(b|a)=\frac{P_{A,B}(a|b)*P_B(b)}{P_A(a)} \quad (3)$$

Which is Bayes' Rule, and can be written in marginal form (which is to say partitioning $P_A(a)$ into its dependencies on all possible B outcomes), and finally we have

$$P_{A,B}(b_k|a)=\frac{P_{A,B}(a|b_k)P_B(b_k)}{\sum_i P_{A,B}(a|b_i)P_B(b_i)} \quad (4)$$

Where we are solving for a particular $P_{A,B}(b|a)$ posterior (dependent result), given $P_{A,B}(a|b)$ likelihood and total probability $P_B(b)$ (known as a prior). The summation in the denominator, hopefully, is recognized to be a normalization factor.

While the mathematics behind Bayes' Rule is correct, there has been controversy around this interpretation of probability. This is due to the selection of priors, which may be based on incomplete or flawed knowledge. Take for example an urn, where red and black balls are drawn with replacement. If 10 black balls are drawn in a row, and we wish to test for fairness, we need to have prior knowledge of the probability that we have a fair urn (where the probability of retrieving a black ball is .5 on any draw).

$$P(\text{fair urn}|a=10\text{ black balls in a row})=\frac{P(a|\text{fair urn})P(\text{fair urn})}{P(a|\text{fair urn})P(\text{fair urn})+P(a|\text{unfair urn})P(\text{unfair urn})} \quad (5)$$

Here we can see another problem. If we have experience with the urn manufacturer, giving us the prior P(fair urn), do we know the likelihood P(a|unfair urn)? Because this likelihood could lie anywhere on a continuum, we may have to guess, calling the entire result into question.

3. Examples

MATLAB will be used to explore two examples. The first example will deal with probability mass functions encountered in medicine, and the second will take a general problem in signal processing, where a continuum is approximated.

In medicine it is common to test for various illnesses, although often tests are reserved for those who exhibit some symptoms, and this example will illustrate why. If a certain cancer B has a rate of occurrence of P(b)=.00001 in the general population, and a test for that cancer returns positive in the presence of the cancer with a rate P(a|b)=.995, and returns a false positive P(a|~b)=.1, then, in the general populace, the probability of cancer, given a positive test is

$$P(cancer|positive\,test) = \frac{P(positive\,test|cancer)\,P(cancer)}{P(positive\,test|cancer)\,P(cancer)+P(positive\,test|no\,cancer)\,P(no\,cancer)}$$
$$= \frac{.995*.00001}{.995*.00001+.1*.99999}$$
$$= 9.95\text{x}10^{-5} \quad (6)$$

Here we can see that a positive test for cancer gives no cause for concern, due to the false positive rate and the comparative enormity of the population without cancer. This also serves to illustrate the importance of choosing priors. If a patient exhibits symptoms which increases their prior probability, the false positives may no longer make the test inconsequential. Also, if insufficient studies have been performed to determine the prior, there can be no way of knowing if a positive result is cause for further action or not.

Next let's look at a case involving signal processing. If we have a sensor (the type is irrelevant) and it gives 3 readings (in this example they are 4.11, 8.79, and 7.98) of the same thing (E.G. temperature measurements across the surface of a pan heating on a stove, the weight of an excited dog on a scale, etc.). We could take the mean and standard deviation of them and assume that the measurement should be 6.96 with a standard deviation of 2.5. This is the same as having a flat prior: that is, we assume all readings are equally likely (if we take the temperature of a frying pan, can we assume that 5000C is as likely as 350C?). When we solve Bayes' rule this way:

$$P(B|A=[4.11,8.79,7.98])=\frac{P(A|B)P(B)}{P(A)} \quad (7)$$

since P(B) is a uniform distribution, it's probability distribution is constant over all values, and so P(B)/P(A)=1, thus:

$$P(B|A=[4.11,8.79,7.98])=P(A=4.11|B)*P(A=8.79|B)*P(A=7.98|B) \quad (8)$$

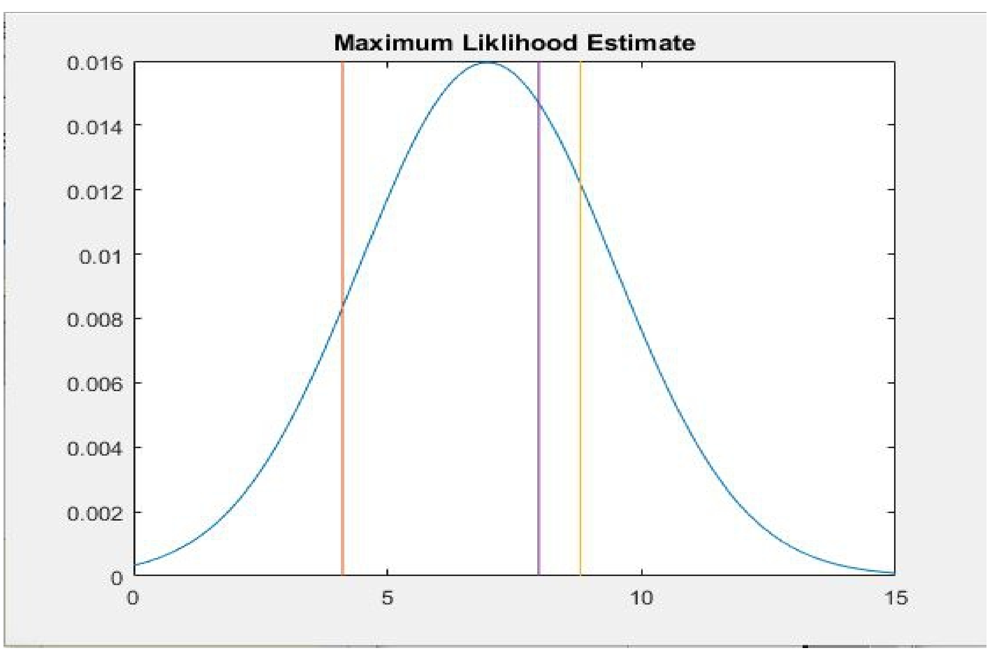

Illustration 1: Maximum Likelihood Estimate. Mean=6.96, Std Dev=2.5

Which has a maxima at B=6.96, exactly the same as taking the mean. If instead, we know something about the problem (last time my dog weighed 7.5 lbs and I think I can detect +/- 1.5 lbs change (σ), or the oil I use just starts to smoke at some known temperature, with a small variation due to atmospheric pressure changes), then we can assign a prior. In this case, we'll let P(B) ~ $\mathcal{N}(\mu=7.5, \sigma^2=2.25)$, now

$$P(B|A)=\frac{P(A|B)P(B)}{\int_{-inf}^{inf} P(A|B)P(B)} \tag{9}$$

Which we can solve in MATLAB and find

$$\mu_{B|A}=7.35$$
$$\sigma^2_{B|A}=1.65 \tag{10}$$

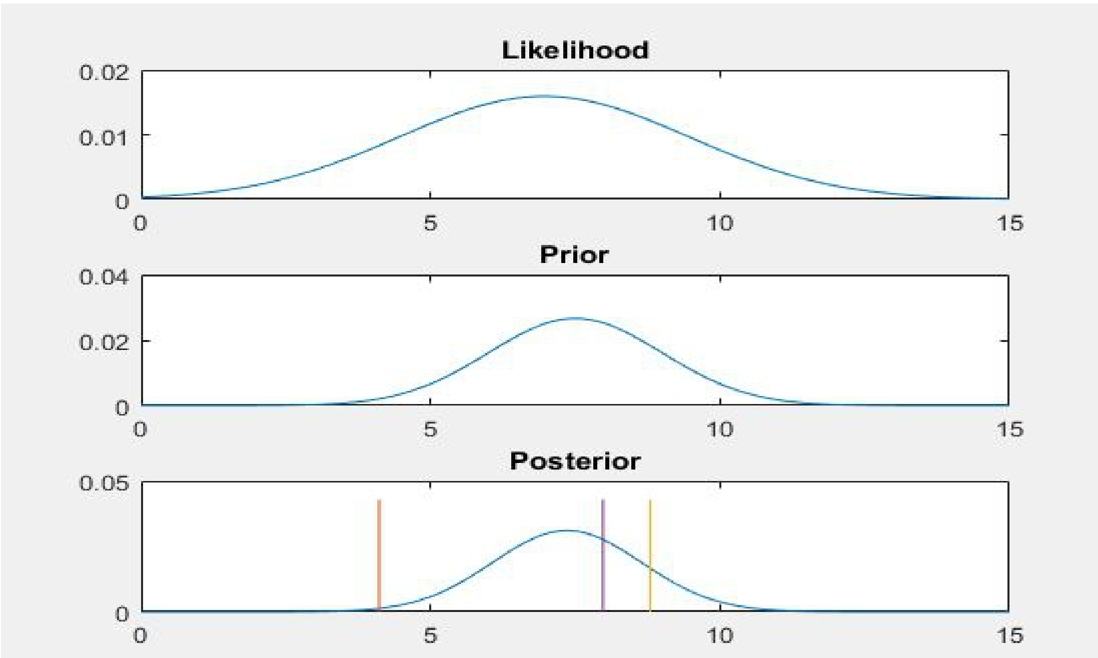

Illustration 2: Likelihood, Prior, and Posterior probabilities

This allows one value to nearly be discarded, based on prior belief (which, if done correctly, is based on accumulated data). In this case, we are updating our parameter set, Θ, based on the data, D.

4. MATLAB code

```
%% Probability of Cancer in general population given a positive test

close all; clear; clc;

% Probability of cancer B in total population
Pb=.001;
```

```
% Probability of positive test, given cancer is present
Pab=.995;

% Probability of positive test, given cancer is not present
Panb=.1;

% Probability of cancer B, given a positive test
Pba=Pab*Pb/(Pab*Pb+(1-Pb)*Panb);                %Pba=9.9491e-5


%% Estimation given three values with and with out prior beliefs

close all; clear; clc;
rng('default')

A=randn(2,1);                          %throw away values
sigma=1.5;
X=(randn(3,1)+5)*sigma;

Xhat=mean(X);
SigHat=std(X);

dXrng=.1;
Xrng=0:dXrng:15;
Likelihood=(1/(SigHat*sqrt(2*pi)))*exp(-0.5*(Xrng-Xhat).^2/SigHat^2)*dXrng;

figure
plot(Xrng, Likelihood, [X(1), X(1)], [0, 0.016], [X(2), X(2)], [0, 0.016], [X(3),
X(3)], [0, 0.016])
title('Maximum Liklihood Estimate')

SigP=1.5;
MeanP=7.5;
Prior=(1/(SigP*sqrt(2*pi)))*exp(-0.5*(Xrng-MeanP).^2/SigP^2)*dXrng;

Posterior=Likelihood.*Prior/sum(Likelihood.*Prior);
MuPostHat=sum(Posterior.*Xrng);
VarPostHat=sum(Posterior.*Xrng.^2)-MuPostHat^2;

figure
subplot(3,1,1)
plot(Xrng, Likelihood)
title('Likelihood')
subplot(3,1,2)
plot(Xrng, Prior)
title('Prior')
subplot(3,1,3)
plot(Xrng, Posterior, [X(1), X(1)], [0, 0.0429], [X(2), X(2)], [0, 0.0429], [X(3),
X(3)], [0, 0.0429])
title('Posterior')
```

References

JOYCE, JAMES. “Bayes' Theorem (Stanford Encyclopedia of Philosophy).” 28 Jun. 2008, https://plato.stanford.edu/entries/bayes-theorem/

KAY, STEVEN. “Conditional Probability.” INTUITIVE PROBABILITY AND RANDOM PROCESSES USING MATLAB, SPRINGER, 2016, pp. 86–89.

XIANG, NING and FACKLER, CAMERON. “Objective Bayesian Analysis in Acoustics.” Objective-Bayesian-Analysis-in-Acoustics-Ning-Xiang-and-Cameron-Fackler.pdf, 16 May 2016, 2:57:41PM, http://acousticstoday.org/wp-content/uploads/2015/06/Objective-Bayesian-Analysis-in-Acoustics-Ning-Xiang-and-Cameron-Fackler.pdf