The Bayesian Paradigm-continued

$$P(H|data) = \frac{P(data|H)P(H)}{P(data)}$$

$$P(H|data) \propto P(data|H)P(H)$$

15.3.1 Calibrating degrees of belief

Suppose I wanted to discover "Your" probability that average adult male emperor penguins weigh more than 50 lbs? We will go through comparison experiments:

1. Would you rather bet on getting one green chip out of 1 R 1G or bet on A true?

   Suppose you prefer the latter.

2. Would you rather bet on getting a green chip out of 3G and 1 R ?

....etc... This allows for statements that enable us to bound probabilites.

Another type of thought experiment could be used to build $P[\Theta leq \theta]$ for an increasing sequence of θ's.

This is not usually how priors are built though because it seems quite an exhaustive process to build up a whole density prior, instead we are going to use families of priors who have easy updating processes with regards to the specific likelihoods at hand.

15.4 Conjugate Priors

This is true for exponential families for which we have the multiplications property that enables to have the conjugate families. We saw the examples of *Binomial-Beta* and *Normal-Normal.*

We saw that the $Beta(1, 1)$ was the uniform, and that as a and b increase the strength the prior intensifies, in fact $Beta(a, a + b)$ corresponds to experimental evidence of having seen a out of $a + b$ successes.

15.4.1 Examples

We want to establish the exact weight of a nominal weight of 10 grams, we have a the actual weight as 10+μ micrograms, we have an unbiased precision balance and make 5 weighings, we suppose that the measurements are random with sd 6 micrograms, $X \sim \mathcal{N}\ (\mu, 36)$ and suppose our prior for μ is $\mathcal{N}\ (0, 16)$:

$$g(\mu) \propto e^{-\frac{1}{32}\mu^2}$$

It turns out $\bar{X}\ = 7.0$. Since $\frac{\sigma^2}{5} = \frac{36}{5}$, the likelihood for $\bar{X}$ is

$$L(\mu) = e^{-\frac{5}{72}(\mu-7)^2}$$

The posterior density is thus

$$H(\mu|\bar{X}) \propto e^{-\frac{5}{72}(\mu-7)^2} \times e^{-\frac{1}{32}\mu^2} = e^{-\frac{29}{288}(\mu-140/29)^2}$$

This is a normal pdf with mean 140/29=4.83 and with sd $\sqrt{144/29} = 2.23$, the data has shifted the distribution of X from 0 to 4.8 and narrowed it somewhat from a sd of 4 to a sd of 2.23.

Bayesian predictive distribution of a new observation is the average of the pdf's with regards to θ , here's an example:

Broken windows.

Twenty windows in a high rise break in the first year it was built. Was this caused by the defect D in the glass? If caused by D, the glass manufacturer has to replace all the glass in the building. If not caused by D, then the company that installed the glass is liable.

Only 4 of the 20 are available for analysis, we will assume this is a random sample of the 20. All four turned out to have defect D , the question the court asked was how many more of the 16 other windows had default D?

The windows from the building were all from the same lot, the lot-proportion p of defect D among all windows with defects varies with a distribution with shape:

$$g(p) \propto p^{-3/4}(1-p)^{-1/4}, \qquad 0 < p < 1$$

This is a beta density with mean $1/4$.

We consider 20 indpt obervations of $Ber(p)$, we observe $U = X_1 + X_2 + X_3 + X_4 = 4$. The likelihood of such an observation is p^4, the posterior for p is:

$$h(p|U = 4) \propto p^4 p^{-3/4}(1-p)^{-1/4}, \qquad 0 < p < 1$$

which is Beta(4.25,.75) with mean 4.25/5.

We want the predictive distribution for $Z = Y_1 + \cdots Y_{16}$, given p, we have

$$f^*(z|p) = \binom{16}{z} p^z (1-p)^{16-z}, z = 0, 1, \ldots, 16$$

which is the joint pdf of the new observations, the predictive distribution is the expectation of this with regards to the distribution of p:

$$f(z) = E[f^*(z|p)] = \binom{16}{z} \frac{\int_0^1 p^{z+3.25}(1-p)^{15.75-z} dp}{\int_0^1 p^{3.25}(1-p)^{-.25} dp}$$

and

$$E(z) = E(E(Z|p)) = 16E(p) = 13.6$$

The expected number of broken windows to have D, given that the first 4 had D, would be 17.6.

Mixtures of Betas

If we are interested in a parameter that is in the compact interval $[0, 1]$ we are actually always in good shape to be able to use a mixture of betas, for which the updating is always easy.

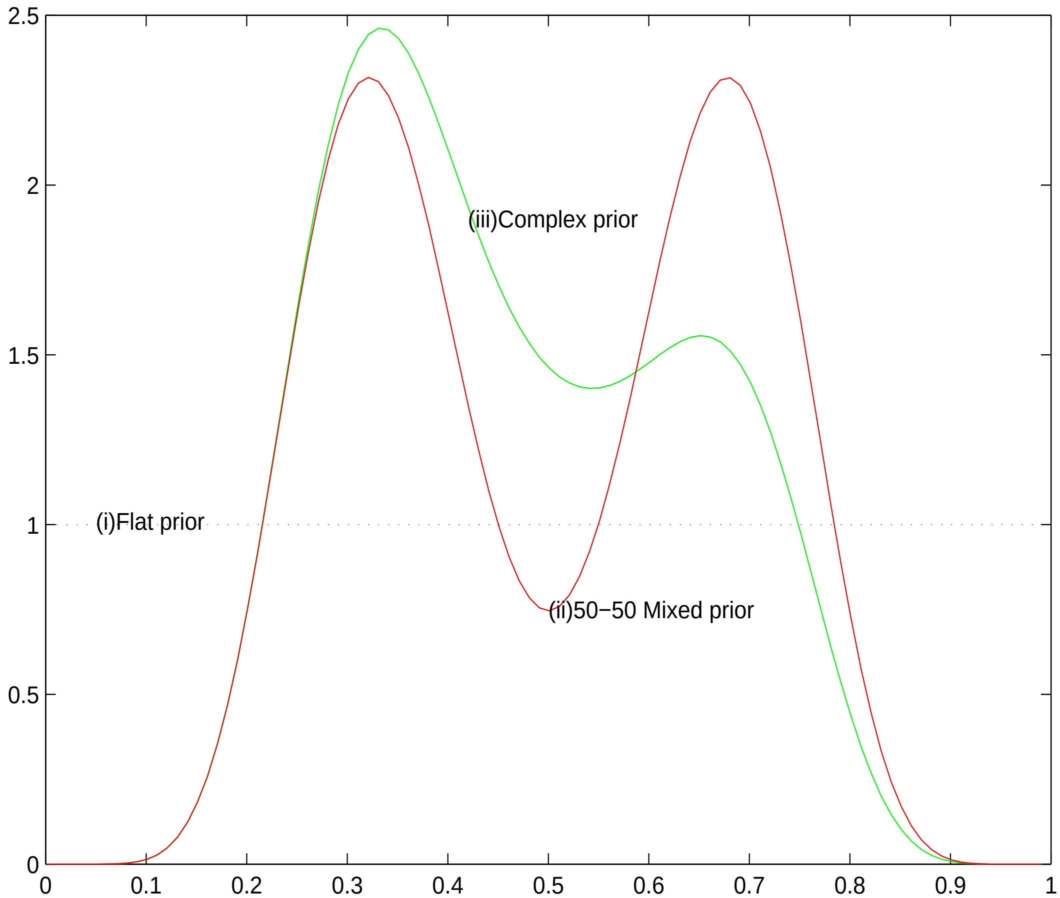
The spinning penny can be modeled by either

$$(i) \qquad B(1,1) = U(0,1)$$

$$(ii) \qquad \frac{1}{2}\beta(10,20) + \frac{1}{2}\beta(10,20)$$

or if we knew a little about the coin that biased towards heads:

$$(iii)\frac{1}{2}\beta(10,20) + \frac{1}{5}\beta(15,15)0.3\beta(20,10)$$



If we observe 3 heads out of n=10 throws, we update this to the following posterior distributions:

$$(i) \qquad \beta(4,8)$$

$$(ii) \qquad 0.84\beta(13,27) + 0.16\beta(23,17)$$

$$(i) \qquad 0.77\beta(13,27) + 0.16\beta(18,22) + 0.07\beta(23,17)$$

Here are the matlab commands to generate these priors:

```
 ps=(0.001:0.01:1);
 pr2=0.5*betapdf(ps,10,20)+0.5*betapdf(ps,20,10);
 pr3=0.5*betapdf(ps,10,20)+0.2*betapdf(ps,15,15)+0.3*betapdf(ps,20,10);
 plot(ps,pr3,'g-')
 hold on;
 plot(ps,pr2,'r-')
 plot(ps,1,'b-')
 text(0.05,1.01,'(i)Flat prior')
 text(0.5,.75,'(ii)50-50 Mixed prior')
 text(0.42,1.9,'(iii)Complex prior')
po1=betapdf(ps,4,8);
po2=0.84*betapdf(ps,13,27)+0.16*betapdf(ps,23,17);
po3=0.77*betapdf(ps,13,27)+0.16*betapdf(ps,18,22)+0.07*betapdf(ps,23,17);
```