Bayesian Analysis - A First Example

This script works through the example in Hoff (2009), section 1.2.1 We are interested in a single parameter: θ, the fraction of individuals in a city population with with a specific infectious disease.

Contents

We have a small sample of 20 individuals to provide empirical evidence for our estimation of theta. Let y be the observed number of infected individuals in the sample.

```
n=20;
```

Matlab Preliminaries

Set random number seed. This will allow you to EXACTLY replicate your results in re-runs. Use "37" to get my results exactly.

```
rand('state',37); % set arbitrary seed for uniform draws
randn('state',37); % set arbitrary seed for normal draws
```

Sampling Distribution

If we knew theta, than a reasonable sampling model (and likelihood function) for the total number of infected citizens would be the binomial (n, θ) density. Thus:

$$p\left(y|\theta\right) = (\; n \;)\, y\theta^{y} \left(1-\theta\right)^{n-y}, \quad \theta \in \left[0,1\right], \quad y = 1, 2, \ldots, n \tag{1}$$

Let's plot the bin(20,theta) for a few values of theta:

```
n=20; %our assumed sample size

th1=0.05;
th2=0.1;
```

```
th3=0.2;
th4=0.5;

R=10000; %number of draws - your choice

y1=binornd(20,th1,R,1); %this gives us an R by 1 vector of proportions.
y2=binornd(20,th2,R,1);
y3=binornd(20,th3,R,1);
y4=binornd(20,th4,R,1);

% Plot the results in a single figure, but with separate windows
figure(1);
subplot(2,2,1); %call first window of a 2 x 2 grid
seq=0:1:20; %determine x-values over which to plot
[counts,bincenters]=hist(y1,seq);
%pcaptures # of obs. for each x-value in "counts", and captures the integers
%over which we wish to plot in "bincenters" (so bincenters = seq);
bar(bincenters,counts/R); %generates actual plot
title('$\theta=0.05$','interpreter','latex','fontsize',15);
xlabel('# of successes (infected persons)');
ylabel('probability');
set(gca,'Xlim',[-1 20]);
set(gca,'Ylim',[0 0.5]);
set(gca,'XTick',0:2:20);

subplot(2,2,2); %fill second window
seq=0:1:20;
[counts,bincenters]=hist(y2,seq);
bar(bincenters,counts/R);
title('$\theta=0.1$','interpreter','latex','fontsize',15);
xlabel('# of successes (infected persons)');
ylabel('probability');
set(gca,'Xlim',[-1 20]);
set(gca,'Ylim',[0 0.5]);
set(gca,'XTick',0:2:20);

subplot(2,2,3); %fill third window
seq=0:1:20;
[counts,bincenters]=hist(y3,seq);
bar(bincenters,counts/R);
title('$\theta=0.2$','interpreter','latex','fontsize',15);
xlabel('# of successes (infected persons)');
ylabel('probability');
set(gca,'Xlim',[-1 20]);
set(gca,'Ylim',[0 0.5]);
set(gca,'XTick',0:2:20);
```

```
subplot(2,2,4); %fill fourth window
seq=0:1:20;
[counts,bincenters]=hist(y4,seq);
bar(bincenters,counts/R);
title('$\theta=0.5$','interpreter','latex','fontsize',15);
xlabel('# of successes (infected persons)');
ylabel('probability');
set(gca,'Xlim',[-1 20]);
set(gca,'Ylim',[0 0.5]);
set(gca,'XTick',0:2:20);

%Save figure in Matlab format
to = 'c:\klaus\AAEC6984\mlab\figures\BinomExample'; %designate destination
saveas(gcf,to);%"gcf" means "get current figure"

% Export figure so it can be imported into latex
to = 'c:\klaus\AAEC6984\tex\module1\BinomExample'; %designate destination
print('-depsc',to);%this automatically selects the current figure
```

Prior

Infection rates in other cities range from 0.05 to 0.2, with an average prevalence of 0.1. Thus, a reasonable prior distribution must satisfy the following conditions:

- bound between 0 and 1
- continuous between bounds
- mean near 0.1

Even with these conditions, there are many priors to choose from. In such cases, we add a 4th condition: computational convenience. This leads to a beta prior. The beta(a,b) has mean a/(a+b) and mode (a-1)/(a-1+b-1). The density is given as:

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \theta \in [0,1] \tag{2}$$

The $Beta(2,20)$ might be a resonable choice. Let's draw from it and plot it (along with the posterior, below).

```
a0=2;
b0=20;
pr=betarnd(a0,b0,R,1);
```

```
Eth=mean(pr)
stdth=std(pr)
minth=min(pr)
maxth=max(pr)
```

The Posterior

Assume we observe y=0 (i.e. nobody of the 20 individuals is infected). It turns out that a $Beta\left(\alpha,\beta\right)$ prior for θ, combined with a $Bin\left(\theta,n\right)$ sampling distribution yields a $Beta\left(\alpha+y,\beta+n-y\right)$. Knowing this, we can immediately draw from this posterior and compute its moments.

```
y=0;
a1=a0+y;
b1=b0+n-y;
po=betarnd(a1,b1,R,1);
Eth=mean(po)
stdth=std(po)
minth=min(po)
maxth=max(po)
```

Let's plot the prior vs. the posterior

```
[f1,x1]=ksdensity(pr,'kernel','epanechnikov','npoints',200,'support',[0 1]);
[f2,x2]=ksdensity(po,'kernel','epanechnikov','npoints',200,'support',[0 1]);

figure(2)
plot(x1,f1,'-b',x2,f2,'-k','LineWidth',1);
title('prior and posterior distributions');
xlabel('$\theta$','interpreter','latex','fontsize',12);
ylabel('density');
h=legend('$p\left(\theta\right)$','$p\left(\theta|y\right)$');
set(h,'interpreter','latex','fontsize',12);
```

Matlab Finals

save your numerical output, if any

```
save c:\klaus\AAEC6984\mlab\worksp\mod1_1a pr po;
```

Publishing Notes

after clicking on "save and publish" in the menu bar above, open the resulting .tex file in TexnicCenter. Before compiling, add the following commands at the top of the file, before the \begin{document} line:

\usepackage{amsmath} \usepackage{epstopdf}