# 1. CH.1: Introduction

## 1.1. Theory: Posteriors

The key is the **posterior density -** the probability density of parameters given data.
Apply Bayes Rule that $\Pr[B|A] = \{\Pr[A|B] \times \Pr[B]\}/\Pr[A]$ to the probability of parameter vector $\boldsymbol{\theta}$ (event $B$) given data vector $\mathbf{y}$ (event $A$):

1. Likelihood function (for data given parameters) $p(\mathbf{y}|\boldsymbol{\theta})$
2. Prior density (for parameters) $p(\boldsymbol{\theta})$
3. Posterior density (for parameters given data) $p(\boldsymbol{\theta}|\mathbf{y}) = p(\boldsymbol{\theta}) \times p(\mathbf{y}|\boldsymbol{\theta})/p(\mathbf{y})$.

The user inputs 1. (as in ML) and 2. (special to Bayesian) and gets out 3.
The term $p(\mathbf{y})$ in 3. is a constant that does not depend on $\theta$. For some aspects of Bayesian analysis we can ignore the constant $p(\mathbf{y})$ and simply say

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta}) \times p(\mathbf{y}|\boldsymbol{\theta}).$$

[Aside: In general if density $f(x) = kg(x)$ where $k$ is a constant that does not depend on $x$ then $g(x)$ is called the **kernel** of the density. Much Bayesian analysis works with just the density kernel].

The **posterior density** $p(\boldsymbol{\theta}|\mathbf{y})$ is the starting point for further analysis:

1. Making statements about $\boldsymbol{\theta}$: e.g. the posterior mean $\mathrm{E}[\boldsymbol{\theta}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$
2. Model selection: e.g. posterior probability for model $j$ $p(M_j|\mathbf{y})$
3. Prediction of new observation(s) $\mathbf{y}^*$ given data $\mathbf{y}$ $p(\mathbf{y}^*|\mathbf{y})$

## 1.2. Computation

Computation is used extensively in Bayesian analysis. One reason is that calculations using the posterior usually involve an integral. Koop begins with this.
Suppose we know the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ but not how to calculate a posterior moment $\mathrm{E}[g(\boldsymbol{\theta})]$ such as the posterior mean. To estimate $\mathrm{E}[g(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ make many draws from $p(\boldsymbol{\theta}|\mathbf{y})$ and for each draw of $\boldsymbol{\theta}$ calculate $g(\boldsymbol{\theta})$ and average.
With $S$ draws

$$\widehat{g}^S = \widehat{\mathrm{E}}[g(\boldsymbol{\theta})] = \frac{1}{S}\sum_{s=1}^{S} g(\boldsymbol{\theta}^{(s)}).$$

A second reason for computational methods is when there is no closed from the posterior but analysis is still possible using modern MCMC methods. This is not introduced until chapter 4 but is the basis for the Bayesan revival.

## 1.3. Exercises

Much of Koop (2003) is about computational methods.

Even so we begin with questions 1 and 2 that are analytical learning exercises in becoming familiar with Bayes theorem. They illustrate that getting the posterior analytically involves tricky math even in these “simple” examples.

Question 1 is easier.

Question 2 is the basis for extension to linear regression in chapter 2.

Question 3 involves computation once we know the posterior.

**1-1** Exponential density and exponential prior, where one way to write the exponential density with parameter $\lambda$ is $f(x) = \lambda e^{-\lambda x}$ where the mean $\mathrm{E}[x] = 1/\lambda$.

**(a)** Suppose $y$ is exponential distributed with parameter $\theta$ (so $\mathrm{E}[y] = 1/\theta$) . Let $\mathbf{y} = (y_1, ..., y_N)$ be data from a random sample of size $N$. Show that the likelihood is $p(\mathbf{y}|\theta) = \theta^N e^{-\theta N\bar{y}}$ where $\bar{y}$ is the sample mean.

**(b)** Show that the MLE which minimizes $\ln f(\mathbf{y}|\theta)$ is $\widehat{\theta} = 1/\bar{y}$.

**(c)** Suppose the prior density for $\theta$ is exponential with parameter 1 (so $\mathrm{E}[\theta] = 1$). Show that the posterior density $p(\theta|\mathbf{y}) \propto \theta^N e^{-\theta(N\bar{y}+1)}$ (where we have ignored $p(y) = \int \theta^{-N} e^{-\theta(\bar{y}+1)} d\theta$).

**(d)** One way to write a gamma density with parameters $a$ and $b$ is $f(x) \propto x^{a=1} e^{-x/b}$ where the mean $\mathrm{E}[x] = ab$.

Given this information, state the posterior density in part (c) and show that the posterior mean is $(N + 1)/(N\bar{y} + 1)$.

**1-2** Normal density and normal prior for mean.

**(a)** Suppose $y$ is normally distributed with mean parameter $\theta$ and variance 1. Let $\mathbf{y} = (y_1, ..., y_N)$ be data from a random sample of size $N$. Show that the likelihood $p(\mathbf{y}|\theta) \propto \exp\{-N(\bar{y}-\theta)^2/2\}$ [Hint: $\sum_{i=1}^N (y_i - \theta)^2 = \{\sum_{i=1}^N (y_i - \bar{y})^2\} + N(\bar{y}-\theta)^2$].

**(b)** Show that the MLE which minimizes $\ln f(\mathbf{y}|\theta)$ is $\widehat{\theta} = \bar{y}$.

**(c)** Suppose the prior density for $\theta$ is normal with mean 0 and variance 1. Show that the posterior density $p(\theta|\mathbf{y}) \propto \exp\{-\frac{1}{2}[N(\bar{y} - \theta)^2 + \theta^2]\}$.

**(d)** Show by completing the square that $N(\bar{y} - \theta)^2 + \theta^2 = (N + 1)(\theta - \frac{N}{N+1}\bar{y})^2 - (N + 1)(\frac{N}{N+1}\bar{y})^2 + N\bar{y}^2$.

**(e)** Using the result in part (d) show that the posterior density $p(\theta|\mathbf{y}) \propto \exp\{-\frac{1}{2}[(N+1)(\theta - \frac{N}{N+1}\bar{y})^2]\}$. Since this is the kernel of the normal density, give the posterior mean.
**(f)** Also give the posterior variance. Is it larger or smaller than the variance of the MLE from part (b)?

**1-3** This question uses simulated methods to calculate moments, using $\widehat{\mathrm{E}}[g(\boldsymbol{\theta})] = \frac{1}{S}\sum_{s=1}^{S} g(\boldsymbol{\theta}^{(s)})$. Suppose the posterior density for parameter $\theta$ is $N[0,1]$.
**(a)** Calculate the posterior mean of $\theta$ using $S = 10$, 100, 1000 and $10,000$ draws.
**(b)** Calculate the variance of this simulation estimate when $S = 100$ (see Koop).
**(c)** Calculate $\mathrm{E}[\exp(-\exp(x))]$ using $S = 10$, 100, 1000 and $10,000$ draws.
**(d)** Does $\mathrm{E}[\exp(-\exp(x))]$ actually exist? [Hint: This is tricky and can be skipped.]