Decision and Risk
Lecture 5: Nonconjugate Priors

Gordon J. Ross

Today's Material

- Two weeks ago: How to perform risk calculations and reason about the probability of extreme events occurring when observations follow a Gaussian distribution
- Last week: How to perform similar calculations when the observations follow a distribution that looks to be Exponential or Lognormal

This week we will discuss how to do similar calculations in situations where we cannot find conjugate priors

Dow Jones Stock Index

A **stock market index** is used to measure the value of a particular section of the stock market. It is constructed by averaging the stock prices of a group of companies, and is hence an example of a **portfolio** of stocks.

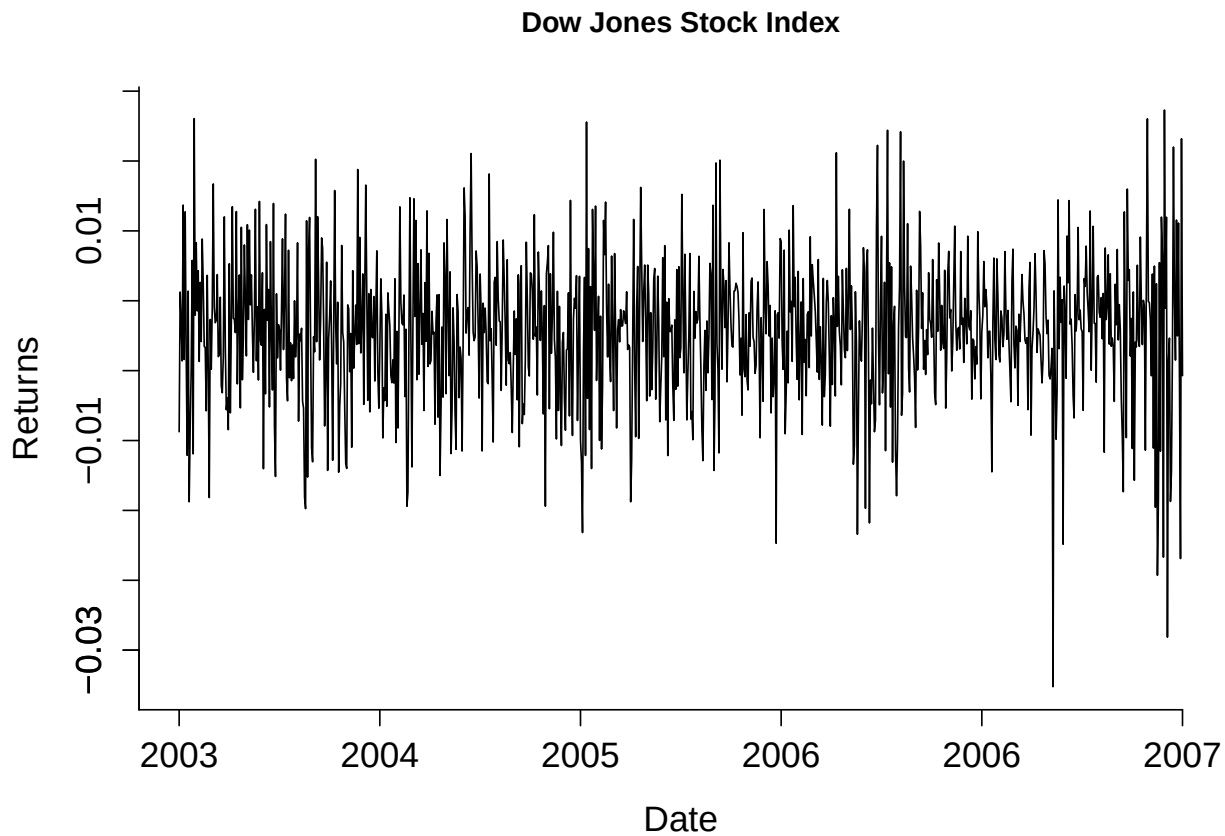
Popular stock indexes include the FTSE 100 which is made up of the 100 largest companies on the London Stock Exchange, and the Dow Jones Industrial Average which is made up of major American companies.

Let X_t denote the price of the index on day t. Stock market participants are usually interested in the log returns of the index, denoted:

$$Y_t = \log \frac{X_t}{X_{t-1}} = \log(X_t) - \log(X_{t-1})$$

Dow Jones Stock Index

Here are the log returns of the Dow Jones between the years 2002 and 2007:



Dow Jones Stock Index

As we done 2 weeks ago, assume that the returns on day Y_t have a $N(\mu, \sigma^2)$ distribution. For VaR analysis, we may want to know the probability of seeing a daily loss greater than 0.01 (i.e. a return less than -0.01). As always, if μ and σ^2 are known, this is given by:

$$p(\tilde{Y} < -0.01) = \int_{-\infty}^{-0.01} p(Y|\mu, \sigma) dY$$

Dow Jones Stock Index

For the Dow Jones, the empirical mean is 0.00042 and standard deviation is 0.0068. If we treat these as the real values (ignoring estimation uncertainty) then we can compute the probabilities in R using:

```
> pnorm(-0.01,0.00042,0.0068)
[1] 0.06271768
```

So the probability of observing a daily return less than 0.01 is around 6.2%.

Dow Jones Stock Index

If μ and σ^2 are not known, they must be estimated. In general, we have seen that if Y has a probability distribution with a vector of unknown parameter θ (here $\theta = (\mu, \sigma^2)$) then the calculation as the same as in the known case except that we must now average over the posterior:

$$p(\tilde{Y} < D) = \int_{-\infty}^{\infty} p(\tilde{Y} < D|\theta)p(\theta|Y)d\theta = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{D} p(\tilde{Y}|\theta)d\tilde{Y}\right] p(\theta|Y)d\theta$$

This means that we need to be able to compute the posterior distribution $p(\theta|Y)$

Dow Jones Stock Index

We saw how to compute posterior distributions $p(\theta|Y)$ when using a conjugate prior for θ, e.g.:

- When Y had a Binomial distribution and the prior was Beta
- When Y had an Exponential or Poisson distribution and the prior was Gamma
- When Y had a Normal distribution, σ^2 was known, and we had a Normal prior on μ.

However in general we will not be able to find conjugate priors....

Gaussian Distribution - Unknown Mean and Variance

Suppose the observations Y_t (which are the return in our case) have a Gaussian distribution where **neither** μ nor σ^2 are known. In this case, both must be learned.

We must specify a prior distribution. If we view our prior beliefs about μ as being independent of our prior beliefs about σ^2 (a reasonable assumption in most cases) we will have a separable prior $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$

In general this **cannot be chosen to be conjugate**

Conjugate Priors

Remember why conjugate priors are important. We have a likelihood function $p(Y|\theta)$ and a prior $p(\theta)$. We want the posterior distribution $p(\theta|Y)$. This is given by:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

In general **we will not be able to solve the integral on the bottom using standard techniques**. However when we use a conjugate prior, we can always solve it using the trick we have used repeatedly where we recognise it has the same form as the prior.

We cannot do this with non-conjugate priors

Integration

Remember some of the standard techniques you will have learned for solving integrals:

- Integration by substiution
- Integrating by parts
- Residue Integration (if you have taken a class in complex analysis

+ several more.

However despite this, many/most integrals that we encounter in applied mathematics cannot be solved using these techniques. Instead, we must use **numeric integration**

Numerical Integration

Numeric integration is a set of techniques for solving **definite integrals** in cases where we cannot find an analytic solution for the indefinite integral. I.e. it lets us do integrals of the form:

$$\int_a^b f(x)\,dx$$

When doing Bayesian analysis on any non-trivial problem, some form of numeric integration will usually be required. Indeed, while most of the theory and mathematics of Bayesian inference was initially worked out during the years 1900-1970, it was only with the widespread availability of fast computers able to quickly perform numeric integration that it became popuar

Numerical Integration

So, to do Bayesian inference in practice, we will often need to compute integrals which can't be solved using standard methods.

In this case, we use numeric integration techniques to solve the integral $p(Y)$ that comes pun in Bayes theorem:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

There are many different types of numeric integration that can be used. The two most popular are Gaussian quadratures, and Markov Chain Monte Carlo. We will focus on the former.

Numerical Integration - Trapezoid Rule

A very simple example of a numerical integration method which you may have seen at A-Level is the Trapezoid rule.

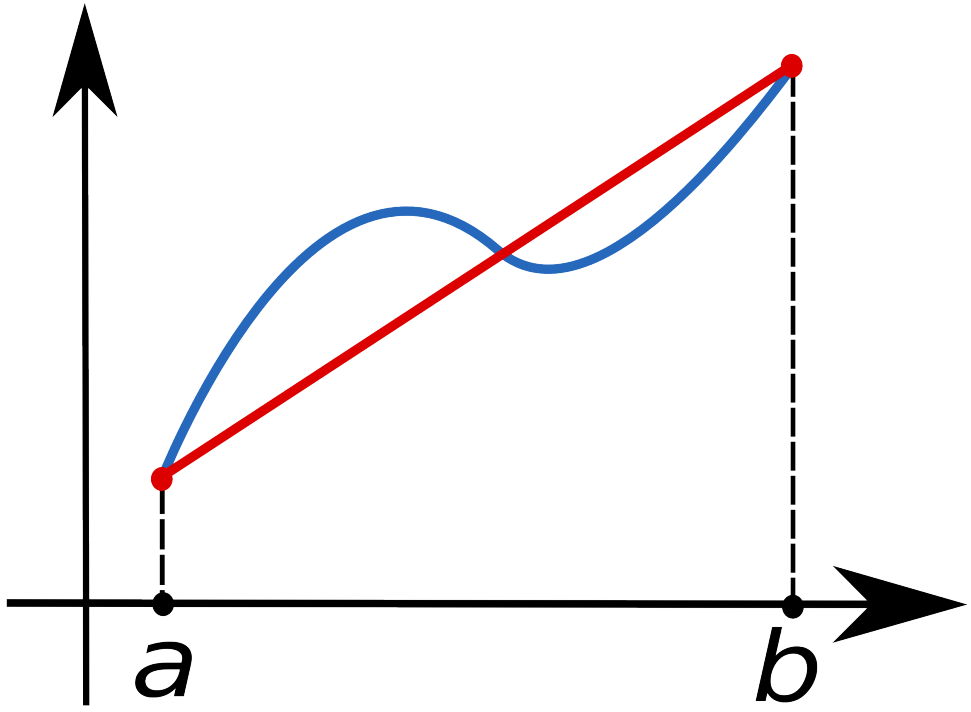
The Trapezoid rule says that we approximate an integral by:

$$\int_a^b f(x)dx \approx (b-a)\frac{f(a)+f(b)}{2}$$

Numerical Integration - Trapezoid Rule

Why does this work? Remember that the integral of a function $f(x)$ over $[a, b]$ is equal to the area under the curve of the function between a and b.

The Trapezoid rule approximates this area with the best fitting rectangle



Numerical Integration - Trapezoid Rule - Example

Suppose we want to do:

$$\int_1^2 x^2 dx$$

We all know that this is just

$$\int_1^2 x^2 dx = \left[\frac{x^3}{3}\right]_1^2 = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3} = 2.333$$

But suppose we had forgot how to do this.

Numerical Integration - Trapezoid Rule - Example

Using the Trapezoid rule;;;

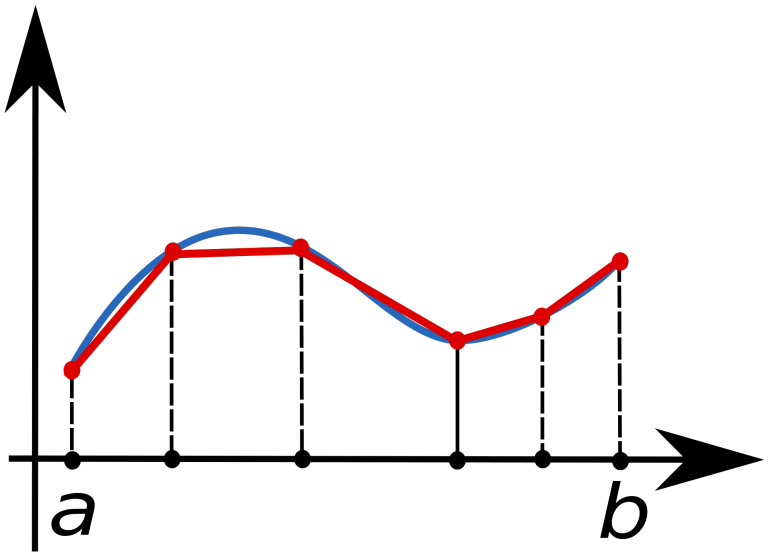
$$\int_1^2 x^2 dx \approx (b-a)\frac{f(a)+f(b)}{2} = (2-1)\frac{2^2+1^2}{2} = 2.5$$

The true value was 2.33, so this isn't too bad!

Numerical Integration - Trapezoid Rule - Refinement

So the Trapezoid rule gives a rough approximation of the integral, but it isn't too accurate. Approximating the function by a rectangle is a bit crude.

We could instead break up the integral and approximate it by several rectangles:



Numerical Integration - Trapezoid Rule - Refinement

For example, if we split up the range into $[1, 1.5]$ and $[1.5, 2]$ we have:

$$\int_1^2 x^2 dx = \int_1^{1.5} x^2 dx + \int_{1.5}^2 x^2 dx$$

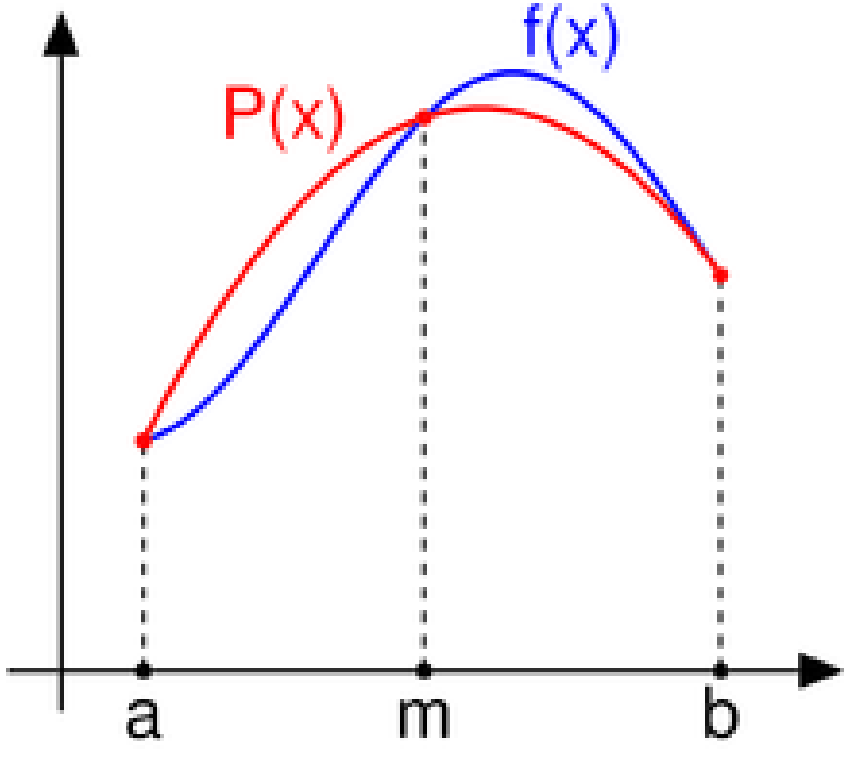
Applying the trapezoid rule to both integrals separately gives:

$$(1.5 - 1)\frac{1^2 + 1.5^2}{2} + (2 - 1.5)\frac{1.5^2 + 2^2}{2} = 0.8125 + 1.5625 = 2.375$$

which is closer to the true value (2.33)

Numerical Integration - Simpson's Rule

Approximating a function by a rectangle is still a bit crude. A slightly more sophisticated approach is to approximate it by a quadratic. This leads to Simpson's Rule.



Numerical Integration - Simpson's Rule

It can be shown that this leads to the approximation

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]$$

Applying this to our example:

$$\int_1^2 x^2 dx \approx \frac{2-1}{6}(1^2 + 4 \times 1.5^2 + 2^2) = 2.33$$

In general this leads to a more exact approximation than the Trapezoid rule (in our case it is exact because f(x) is a quadratic, so approximating it by a quadratic gives a perfect fit)

Numerical Integration - Gaussian Quadratures

The Trapezoid rule and Simpsons rule are very general - we can use them to approximate essentially any integral we like.

Although they are both quite simplistic, they are examples of a general idea which is very powerful. In general, we can approximate the integral of $f(x)$ over $[a, b]$ by a weighted series of n evaluations of $f(x)$ at the evaluation points $x_1, \ldots, x_n$.

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

The Trapezoid Rule and Simpson's rule correspond to different choices of w_i and x_i. More sophisticated methods for numerical integration choose the weights and evaluation points in more sophisticated ways.

Numerical Integration - Gaussian Quadratures

Making a 'good' choice for the $w_i's$ and $x_i's$ involves a lot of quite detailed engineering work – typically these points are chosen adaptively by first studying the function, and choosing based on this. The most widely used method for this is called Gaussian Quadratures

We will not explore this in detail since the main goal of this course is not to teach numerical methods. But fortunately almost all computer programming languages will have a numerical integration routine which implements a sophisticated version of this approach.

In R, this is done through the integrate() function

Numerical Integration in R

The integrate() function in R has the form:

$$integrate(f, lower, upper)$$

The first argument is the function to be integrated, the second argument is the lower bound of integration ('a") and the third argument is the upper bound of integration ('b').
So to do:

$$\int_a^b f(x)dx$$

We would call integrate(f,a,b)

Numerical Integration in R - Example

Lets go back to our first example:

$$\int_1^2 x^2 dx$$

which we know is equal to 2.33.
We start by defining the function we want to integrate. In this case we are integrating with respect to x so we write it as a function that takes x as an argument. We call the function 'f' (you can call it anything) and deifne it to return the square of the argument, i.e. x^2

```
f <- function(x) {
  return(x^2)
}
```

Numerical Integration in R - Example

we then pass this function we have created to the built-in integrate function in R. The arguments 'lower' and 'upper' specify the lower and upper bounds of the definite integral (1 and 2 in this case)

```
> integrate(f,lower=1,upper=2)
2.333333 with absolute error < 2.6e-14
```

So the integral is 2.333 (ignore the absolute error part)

Numerical Integration - Example

Say we wanted to do:

$$\int_2^8 e^x sin(x) + x^3 \log(x) dx$$

We do this in R by:

```
f <- function(x) {
  return(exp(x) * sin(x) + x^3 * log(x))
}
> integrate(f, lower=2,upper=8)
3558.161 with absolute error < 4e-11
```

so the integral evaluates to 3558.161

We can use this to do essentially any definite integral.

Bayesian Example

Lets illustrate how this works in practice using the example from the first lecture.

Suppose we are given a coin and told that it could be biased, so the probability of landing heads is not necessarily 0.5. Let θ denote the probability of it landing heads. We wish to learn about θ.

We toss the coin N times and obtain Y heads. This gives a Binomial likelihood

$$p(Y|\theta) = \binom{N}{Y} \theta^Y (1-\theta)^{N-Y}$$

Previously we chose a conjugate Beta prior for θ. But suppose we feel that our beliefs cannot be adequately represented by a Beta distribution and need to choose a non conjugate prior

Bayesian Example

We will assume that our prior for θ is now Gaussian. Lets say our prior beliefs about θ can be represented by $p(\theta) = N(0.5, 0.1^2)$, i.e a mean of 0.5 with a standard deviation of 0.1. We now need to compute the posterior:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

The numerator is:

$$p(\theta|Y) = \binom{N}{Y}\theta^Y(1-\theta)^{N-Y}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\left(\frac{(\theta-\mu)^2}{2\sigma^2}\right)}$$

where $\mu = 0.5$, $\sigma = 0.1$

Bayesian Example

The denominator is:

$$p(Y) = \binom{N}{Y} \int_0^1 \theta^Y (1-\theta)^{N-Y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{(\theta-\mu)^2}{2\sigma^2}\right)} d\theta$$

This cannot be done by hand, however doing it in R using quadratures (via the integrate function) is very easy.

Bayesian Example

Again, we first define the function to be integrated. We are integrating with respect to θ so this should be the argument, with N, Y, μ, σ^2 defined outside the function:

```
f <- function(theta) {
  theta^Y * (1-theta)^(N-Y) * 1/sqrt(2*pi*sigma^2) *
    exp( (theta-mu)^2 / (2*sigma^2))
}
N <- 100 #100 tosses for example
Y <- 48 #48 heads observed
mu <- 0.5
sigma <- 0.1
integrate(f,lower=0,upper=1)
[1] 6.069468e-30 with absolute error < 1e-31
```

Summary

This lets us evaluate the denominator of Bayes theorem. Putting it all together, we can hence compute $p(\theta|Y)$ for any value of θ we like

This essentially lets us apply Bayesian inference to any statistical problem; we no longer need to worry about conjugate priors (although choosing conjugate priors is still useful in practice to keep things simple)

To understand how the numeric integration approach works in practice you will need to work through some examples yourself. In this week's computer workshop we will cover how to use numerical integration for Bayesian inference