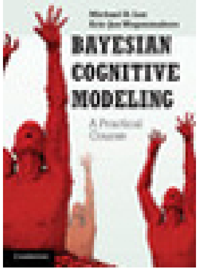


Bayesian Cognitive Modeling

A Practical Course

Michael D. Lee, Eric-Jan Wagenmakers

Book DOI: http://dx.doi.org/10.1017/CBO9781139087759

Online ISBN: 9781139087759

Hardback ISBN: 9781107018457

Paperback ISBN: 9781107603578

Chapter

4 - Inferences with Gaussians pp. 54-59

Chapter DOI: http://dx.doi.org/10.1017/CBO9781139087759.006

Cambridge University Press

4 Inferences with Gaussians

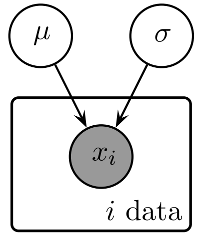
4.1 Inferring a mean and standard deviation

One of the most common inference problems involves assuming data following a Gaussian (also known as a Normal, Central, or Maxwellian) distribution, and inferring the mean and standard deviation of this distribution from a sample of observed independent data.

The graphical model representation for this problem is shown in Figure 4.1. The data are the n observations $x_1, \ldots, x_n$. The mean of the Gaussian is μ and the standard deviation is σ. WinBUGS parameterizes the Gaussian distribution in terms of the mean and precision, not the mean and variance or the mean and standard deviation. These are all simply related, with the variance being σ^2 and the precision being $\lambda = 1/\sigma^2$.

Here the prior used for μ is intended to be only weakly informative. That is, it is a prior intended to convey little information about the mean, so that inference will be primarily dependent upon relevant data. It is a Gaussian centered on zero, but with very low precision (i.e., very large variance), and gives prior probability to a wide range of possible means for the data. When the goal is to estimate parameters, this sort of approach is relatively non-controversial.

Setting priors for standard deviations (or variances, or precisions) is trickier, and certainly more controversial. If there is any relevant information that helps put the data on scale, so that bounds can be set on reasonable possibilities for the standard deviation, then setting a uniform over that range is advocated by Gelman (2006). In this first example, we assume the data are all small enough that setting an upper bound of 10 on the standard deviation covers all the possibilities.



$$\mu \sim \text{Gaussian}(0, 0.001)$$
$$\sigma \sim \text{Uniform}(0, 10)$$
$$x_i \sim \text{Gaussian}(\mu, \tfrac{1}{\sigma^2})$$

Fig. 4.1 Graphical model for inferring the mean and standard deviation of data generated by a Gaussian distribution.

The script Gaussian.txt implements the graphical model in WinBUGS. Note the conversion of the standard deviation sigma into the precision parameter lambda used to sample from a Gaussian:

```
# Inferring the Mean and Standard Deviation of a Gaussian
model{
  # Data Come From A Gaussian
  for (i in 1:n){
    x[i] ~ dnorm(mu,lambda)
  }
  # Priors
  mu ~ dnorm(0,.001)
  sigma ~ dunif(0,10)
  lambda <- 1/pow(sigma,2)
}
```

The code Gaussian.m or Gaussian.R creates some artificial data, and applies the graphical model to make inferences from data. The code does not produce a graph, or any other output. But all of the information you need to analyze the results is in the returned variables samples and stats.

Exercises

Exercise 4.1.1 Try a few data sets, varying what you expect the mean and standard deviation to be, and how many data you observe.

Exercise 4.1.2 Plot the *joint* posterior of μ and σ. That is, plot the samples from μ against those of σ. Interpret the shape of the joint posterior.

Exercise 4.1.3 Suppose you knew the standard deviation of the Gaussian was 1.0, but still wanted to infer the mean from data. This is a realistic question: For example, knowing the standard deviation might amount to knowing the noise associated with measuring some psychological trait using a test instrument. The x_i values could then be repeated measures for the same person, and their mean the trait value you are trying to infer. Modify the WinBUGS script and Matlab or R code to do this. What does the revised graphical model look like?

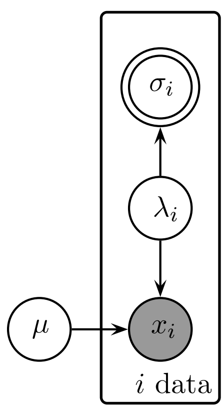
Exercise 4.1.4 Suppose you knew the mean of the Gaussian was zero, but wanted to infer the standard deviation from data. This is also a realistic question: Suppose you know that the error associated with a measurement is unbiased, so its average or mean is zero, but you are unsure how much noise there is in the instrument. Inferring the standard deviation is then a sensible way to infer the noisiness of the instrument. Once again, modify the WinBUGS script and Matlab or R code to do this. Once again, what does the revised graphical model look like?

4.2 The seven scientists

This problem is from MacKay (2003, p. 309) where it is, among other things, treated to a Bayesian solution, but not quite using a graphical modeling approach, nor relying on computational sampling methods.

Seven scientists with wildly-differing experimental skills all make a measurement of the same quantity. They get the answers $x = \{-27.020, 3.570, 8.191, 9.898, 9.603, 9.945, 10.056\}$. Intuitively, it seems clear that the first two scientists are pretty inept measurers, and that the true value of the quantity is probably just a bit below 10. The main problem is to find the posterior distribution over the measured quantity, telling us what we can infer from the measurement. A secondary problem is to infer something about the measurement skills of the seven scientists.

The graphical model for one way of solving this problem is shown in Figure 4.2. The assumption is that all the scientists have measurements that follow a Gaussian distribution, but with different standard deviations. However, because they are all measuring the same quantity, each Gaussian has the same mean, and it is just the standard deviation that differs.



$$\mu \sim \text{Gaussian}(0, 0.001)$$
$$\lambda_i \sim \text{Gamma}(0.001, 0.001)$$
$$\sigma_i \leftarrow 1/\sqrt{\lambda_i}$$
$$x_i \sim \text{Gaussian}(\mu, \lambda_i)$$

Fig. 4.2 Graphical model for the seven scientists problem.

Notice that we have used a different approach to assign priors to the standard deviations. The previous example, as shown in Figure 4.1, used a uniform distribution. The current example, shown in Figure 4.2, uses a gamma distribution for the priors on the precisions. This is another standard approach, which has some attractive theoretical motivations, but is critically discussed by Gelman (2006).

The script `SevenScientists.txt` implements the graphical model in Figure 4.2 in WinBUGS:

```
# The Seven Scientists
model{
  # Data Come From Gaussians With Common Mean But Different Precisions
  for (i in 1:n){
    x[i] ~ dnorm(mu,lambda[i])
```

Box 4.1 Priors on precisions

The practice of assigning $\mathrm{Gamma}(0.001, 0.001)$ priors on precision parameters is theoretically motivated by scale invariance arguments, meaning that priors are chosen so that changing the measurement scale of the data does not affect inference. The invariant prior on precision λ corresponds to a uniform distribution on $\log \sigma$, that is, $p(\sigma^2) \propto 1/\sigma^2$, or a $\mathrm{Gamma}(a \to 0, b \to 0)$ distribution. This invariant prior distribution, however, is *improper* (i.e., the area under the curve is unbounded), which means it is not really a distribution, but the limit of a sequence of distributions (see Jaynes, 2003). WinBUGS requires the use of proper distributions, and the $\mathrm{Gamma}(0.001, 0.001)$ prior is intended as a proper approximation to the theoretically motivated improper prior. This raises the issue of whether inference is sensitive to the essentially arbitrary value 0.001, and it is sometimes the case that using other small values such as 0.01 or 0.1 leads to more stable sampling in WinBUGS.

```
  }
  # Priors
  mu ~ dnorm(0,.001)
  for (i in 1:n){
    lambda[i] ~ dgamma(.001,.001)
    sigma[i] <- 1/sqrt(lambda[i])
  }
}
```

Notice that the graphical model implements the prior on the precisions, but also re-parameterizes to the standard deviation scale, which is often more easily interpretable.

The code `SevenScientists.m` or `SevenScientists.R` applies the seven scientist data to the graphical model.

Exercises

Exercise 4.2.1 Draw posterior samples using the Matlab or R code, and reach conclusions about the value of the measured quantity, and about the accuracies of the seven scientists.

Exercise 4.2.2 Change the graphical model in Figure 4.2 to use a uniform prior over the standard deviations, as was done in Figure 4.1. Experiment with the effect the upper bound of this uniform prior has on inference.

Box 4.2 Ill-posed problems

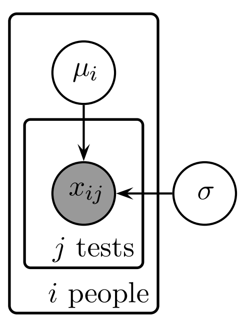
"If one fails to specify the prior information, a problem of inference is just as ill-posed as if one had failed to specify the data." (Jaynes, 2003, p. 373).

4.3 Repeated measurement of IQ

In this example, we consider how to estimate the IQ of a set of people, each of whom have done multiple IQ tests. The data are the measures x_{ij} for the $i = 1, \ldots, n$ people and their $j = 1, \ldots, m$ repeated test scores.

We assume that the differences in repeated test scores are distributed as Gaussian error terms with zero mean and unknown precision. The mean of the Gaussian of a person's test scores corresponds to their latent true IQ. This will be different for each person. The standard deviation of the Gaussians corresponds to the accuracy of the testing instruments in measuring the one underlying IQ value. We assume this is the same for every person, since it is conceived as a property of the tests themselves.

The graphical model for this problem is shown in Figure 4.3. Because we know quite a bit about the IQ scale, it makes sense to set priors for the mean and standard deviation using this knowledge. Our first attempt to set priors (these are revisited in the exercises) simply assume the actual IQ values are equally likely to be anywhere between 0 and 300, and standard deviations are anywhere between 0 and 100.



$$\mu_i \sim \text{Uniform}(0, 300)$$
$$\sigma \sim \text{Uniform}(0, 100)$$
$$x_{ij} \sim \text{Gaussian}(\mu_i, \tfrac{1}{\sigma^2})$$

Fig. 4.3 Graphical model for inferring the IQ from repeated measures.

The script `IQ.txt` implements the graphical model in WinBUGS:

```
# Repeated Measures of IQ
model{
  # Data Come From Gaussians With Different Means But Common Precision
  for (i in 1:n){
    for (j in 1:m){
      x[i,j] ~ dnorm(mu[i],lambda)
    }
  }
```

```
  # Priors
  sigma ~ dunif(0,100)
  lambda <- 1/pow(sigma,2)
  for (i in 1:n){
    mu[i] ~ dunif(0,300)
  }
}
```

The code `IQ.m` or `IQ.R` creates a data set corresponding to there being three people, with test scores of $(90, 95, 100)$, $(105, 110, 115)$, and $(150, 155, 160)$, and applies the graphical model.

Exercises

Exercise 4.3.1 Use the posterior distribution for each person's μ_i to estimate their IQ. What can we say about the precision of the IQ test?

Exercise 4.3.2 Now, use a more realistic prior assumption for the μ_i means. Theoretically, IQ distributions should have a mean of 100, and a standard deviation of 15. This corresponds to having a prior of `mu[i]` $\sim$ `dnorm(100,.0044)`, instead of `mu[i]` $\sim$ `dunif(0,300)`, because $1/15^2 = 0.0044$. Make this change in the WinBUGS script, and re-run the inference. How do the estimates of IQ given by the means change? Why?

Exercise 4.3.3 Repeat both of the above stages (i.e., using both priors on μ_i) with a new, but closely related, data set that has scores of $(94, 95, 96)$, $(109, 110, 111)$, and $(154, 155, 156)$. How do the different prior assumptions affect IQ estimation for these data. Why does it not follow the same pattern as the previous data?