A Bowl of Oatmeal

- Imagine a huge pot of oatmeal with raisins. Suppose the number of raisins per unit volume in the pot is \( \rho \). We are served a bowl of oatmeal of volume \( S \) and expected number of raisins equal to \( \lambda = \rho S \).
- We eat the oatmeal with a very small spoon. Because the spoon is so small, we will have to take a large number \( n \) of equally sized spoonfuls, so that the expected number of raisins in each spoonful is \( \lambda/n << 1 \).
- The probability that a spoonful contains a raisin is approximately \( \lambda/n \).
- The probability that it does not contain a raisin is approximately \( 1 - \lambda/n \).
- Assume \( n \) is so large that we can ignore the possibility of more than one raisin in a spoonful.

\[
P(k \mid \alpha, n) = C_n^k \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]

For very large \( m \), we have
\[
\frac{C_n^k}{n!} = \frac{1}{k!} + O\left( \frac{1}{n} \right), \quad \left( 1 - \frac{\lambda}{n} \right)^{n-k} = e^{-\lambda} + O\left( \frac{1}{n} \right)
\]

So in the limit as \( n \to \infty \)
\[
P(k \mid \lambda, n) \to P(k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}
\]

A Bowl of Oatmeal

- Note that
\[
\frac{C_n^k}{n^k} = \frac{n(n-1) \cdots (n-k+1)}{k! n^k} = \frac{1}{k!} \left( 1 - \frac{0}{n} \right) \left( 1 - \frac{1}{n} \right) \cdots \left( 1 - \frac{k-1}{n} \right) = \frac{1}{k!} + \frac{k(k-1)}{2n} + O\left( \frac{1}{n^2} \right)
\]

and
\[
\log \left( 1 - \frac{\lambda}{n} \right)^{n-k} = (n-k) \log \left( 1 - \frac{\lambda}{n} \right)
\]

\[
= -(n-k) \left( \frac{\lambda}{n} + O\left( \frac{1}{n} \right) \right) = -\lambda + O\left( \frac{1}{n} \right)
\]

so that
\[
\left( 1 - \frac{\lambda}{n} \right)^{n-k} = e^{-\lambda} + O(n)
\]

Poisson Distribution

- Under these assumptions, the distribution of the number \( k \) of raisins in all of the spoonfuls is binomial:

\[
P(k \mid \alpha, n) = C_n^k \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]

- This distribution,

\[
P(k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}
\]

is the Poisson distribution.

- The Poisson distribution is a distribution on integers, and has a real parameter \( \lambda \).

- It governs “counting” situations such as the random arrival of clients at a serving station, the random decay of atoms in a lump of radioactive material, the arrival of photons at a detector, the occurrence of automobile accidents, deaths of patients in a hospital, etc.

- Each example is characterized by a \textit{rate} (expected occurrences per unit time, volume, or other unit) \( \lambda \), and the \textit{number of occurrences} \( k \).
Heart Transplant Mortality Rates

- In §3.3, Albert considers heart transplant mortality rates.
- He models the data, the number of patients who die in a given period of time after an operation, with the Poisson distribution.
- He chooses as his prior a gamma distribution. The gamma distribution has the general form
  \[ p(\lambda | \alpha, \beta) \propto \lambda^{\alpha-1} \exp(-\beta \lambda) \]
  with shape parameter \( \alpha \) and rate parameter \( \beta \), or alternatively
  \[ p(\lambda | \alpha, \sigma) \propto \lambda^{\alpha-1} \exp(-\lambda / \sigma) \]
  where \( \sigma = 1/\beta \) is the scale parameter.

Heart Transplant Mortality Rates

- Albert wants to estimate the mortality rate from \( e \), the exposure (expected number of deaths case mix), and \( y \), the number of deaths within 30 days. The exposure is calculated by adding up the probability of death for each patient undergoing the surgery (a statistical model is used to estimate this, patient by patient, using data unique to each patient). He models the number of deaths \( y \) as Poisson with mean \( e \lambda \), and he wishes to estimate \( \lambda \), the mortality rate per unit exposure. Thus
- Albert’s statement of the problem on p. 41 contains an error. He identified \( e \) as the expected number of deaths at the hospital. The actual data had an extra factor of 1000, so the \( e \) in the book is 1000 times too large. Therefore we will remove the factor of 1000 in our calculations.

- Our model is therefore
  \[ f(y | e, \lambda) \propto \frac{(e \lambda)^y \exp(-e \lambda)}{y!} \]
- Multiplying this by a gamma prior and dropping constants independent of \( \lambda \), we get
  \[ g(\lambda | y, e, \alpha, \beta) \propto \frac{(e \lambda)^y \exp(-e \lambda)}{y!} \lambda^{\alpha - 1} \exp(-\beta \lambda) \]
  \[ \propto \lambda^{y+\alpha-1} \exp(-(\beta + e) \lambda) \]
- So, the prior, which was gamma with shape parameter \( \alpha \) and rate parameter \( \beta \), goes over to a posterior which is gamma with corresponding parameters \((\alpha + y)\) and \((\beta + e)\). It is an example of a gamma-Poisson conjugate family.

- If we had no data, the usual default prior for \( \lambda \) is \( g(\lambda) = 1/\lambda \). However, Albert says that there is data on a group of 10 hospitals that are similar to the one we are interested in, so he proposes using these data to estimate a prior for \( \lambda \).
- In the \( j \)th hospital, the total exposure was \( o_j \), and there were \( z_j \) deaths. The data are independent, so with the above prior, the posterior on \( \lambda \) is given by
  \[ f(\lambda | o, z) = \prod \left[ \frac{(\lambda o_j)^{z_j} \exp(-(\lambda o_j))}{z_j!} \right] \frac{1}{\lambda} \]
  \[ \propto \lambda^{\sum z_j - 1} \exp(-\lambda \sum o_j) \]
Heart Transplant Mortality Rates

- So we see that the prior distribution, taking into account the 10 hospitals (but not our hospital) is gamma, with shape and rate parameters:
  \[ \sum z_j = 16 \]
  \[ \sum o_j = 15.174 \]

- Using this gamma distribution as the prior for the data on our 11th hospital, for which we had one death and total exposure 0.066, we find that the posterior distribution on \( \lambda \) is gamma with shape parameter 16+1=17 and rate parameter 15.174+0.066=15.240

Heart Transplant Mortality Rates

- First, let us compute the posterior distribution on \( \lambda \). We do this by simulation, obtaining a large sample of \( \lambda \)'s. The distribution is gamma with the parameters we've just computed.

  \[
  \begin{align*}
  \alpha &= 16 \\
  \beta &= 15.174 \\
  y_{obs} &= 1 \\
  ex &= 0.066 \\
  \lambda_{AA} &= rgamma(1000, shape=\alpha+y_{obs}, rate=\beta+ex) \\
  \text{hist}(& \lambda_{AA}, 50) \\
  \text{quantile}(\lambda_{AA}, c(0.025, 0.975))
  \end{align*}
  \]

Heart Transplant Mortality Rates

- We can check whether our data are consistent with the model by looking at the prior predictive density of the data. This is our prediction of the data, before we look at the data for the hospital in question. It is gotten by solving Bayes’s theorem for the marginal likelihood (“Student’s Theorem”):

  \[
  f(y) = \frac{f(y|\lambda)g(\lambda)}{g(\lambda|y)}
  \]

- Note that the left-hand side of this equation is independent of \( \lambda \); therefore, so is the right-hand side, and we can evaluate it for any \( \lambda \) that we choose. Albert chooses to evaluate it at the MLE for \( \lambda = \alpha / \beta \).

Heart Transplant Mortality Rates

- Here is Albert’s code:

  \[
  \begin{align*}
  \alpha &= 16; \ \beta = 15.174 \\
  y_{obs} &= 1; \ \text{ex} = 0.066 \\
  y &= 0:10 \\
  \text{lam} &= \alpha / \beta \\
  \text{py} &= \text{dpois}(y, \text{lam}*\text{ex})*\text{dgamma}(\text{lam}, \text{shape}=\alpha, \text{rate}=\beta)/\text{dgamma}(\text{lam}, \text{shape}=\alpha+y, \text{rate}=\beta+\text{ex}) \\
  \text{cbind}(y, \text{round(py, 3)})
  \end{align*}
  \]

- The data (1 death on exposure of 0.066) is consistent with this prediction
Heart Transplant Mortality Rates

- He then considers the prior predictive for a larger hospital:

\[ y_{obs} = 4; \ ex = 1.767 \]

\[ y = 0:10 \]

\[ \text{lam} = \alpha/\beta \]

\[ \text{py} = \text{dpois}(y, \text{lam} * \text{ex})*\text{dgamma}(\text{lam}, \text{shape} = \alpha, \ \text{rate} = \beta)/\text{dgamma}(\text{lam}, \text{shape} = \alpha + y, \ \text{rate} = \beta + \text{ex}) \]

\[ \text{cbind}(y, \text{round(py, 3)}) \]

- The data (4 deaths with exposure 1.767) is also consistent with this prior predictive distribution.

Heart Transplant Mortality Rates

- Finally he compares prior and posterior for the two hospitals:

\[ \lambda = \text{seq}(0, \max(\text{c}($\lambda_A$, $\lambda_B$))), \ \text{length}=500) \]

\[ \text{par(mfrow=c(2,1))} \]

\[ \text{hist}(\lambda_A, \text{freq=F, main="", breaks=15, xlim=c(0,3), ylim=c(0,2))} \]

\[ \text{lines}(\lambda, \text{dgamma}(\lambda, \text{shape}=\alpha, \ \text{rate}=\beta)) \]

\[ \text{hist}(\lambda_B, \text{freq=F, main="", breaks=15, xlim=c(0,3), ylim=c(0,2))} \]

\[ \text{lines}(\lambda, \text{dgamma}(\lambda, \text{shape}=\alpha, \ \text{rate}=\beta)) \]

Heart Transplant Mortality Rates

- We see that for the smaller hospital with a low exposure due to few surgeries, the posterior and the prior are very similar; however, there is enough data for the large hospital that the difference between the prior and posterior is easily noticeable.

- This example is rather artificial, since we’ve estimated one hospital’s mortality rate using a prior generated by ten other hospitals. Ideally, we should estimate the mortality rates for all hospitals together, but using data from all hospitals in a “pooled” but consistent way so that each hospital contributes to the prior on all the others.

- We will show how to do this in a later lecture.
Normal Data with Both Parameters Unknown

- §4.2 of the book
- Assume \( \{y_1, y_2, ..., y_n\} \) are independent and identically distributed samples from some normal distribution ("iid \( N(\mu, \sigma^2) \)"
- The likelihood function for a single observation is
  \[
  N(y_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)
  \]

Normal Data with Both Parameters Unknown

- The most common case unknown mean and variance
- Take flat prior for \( \mu \) and Jeffreys prior for \( \sigma^2 \propto 1/\sigma^2 \)
- Then posterior \( \propto \) prior \( \times \) likelihood
  \[
  pd = g(\mu)g(\sigma^2 | \mu) f(y | \sigma^2, \mu)
  \]
  \[
  \propto \frac{1}{\sigma^2} \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(S + n(\mu - \bar{y})^2)\right]
  \]
  \[
  = \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(S + n(\mu - \bar{y})^2)\right]
  \]
  \[
  = \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{S}{2\sigma^2}\right] \exp\left[-\frac{n(\mu - \bar{y})^2}{2\sigma^2}\right]
  \]
- The posterior is symmetric in \( \mu \) but asymmetric in \( \sigma \)

Normal Data with Both Parameters Unknown

- The likelihood function for all the observations is
  \[
  f(y_1, y_2, ..., y_n, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum (y_i - \mu)^2\right)
  \]
- To simplify this we complete the square:
  \[
  \sum (y_i - \mu)^2 = \sum (y_i - \bar{y} + \bar{y} - \mu)^2
  \]
  \[
  = \sum (y_i - \bar{y})^2 + n(\mu - \bar{y})^2
  \]
  \[
  = S + n(\mu - \bar{y})^2
  \]
  \[
  \bar{y} = \frac{\sum y_i}{n}
  \]
  \[
  S = \sum (y_i - \bar{y})^2
  \]
  Sufficient statistics

Normal Data with Both Parameters Unknown

- We see that the posterior distribution can be factored as
  the product of a piece independent of \( \mu \) times a piece
  depending on both \( \mu \) and \( \sigma^2 \).

  \[
  pd = g(\mu)g(\sigma^2 | \mu) f(y | \sigma^2, \mu)
  \]
  \[
  \propto \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{S}{2\sigma^2}\right] \exp\left[-\frac{n(\mu - \bar{y})^2}{2\sigma^2}\right]
  \]
- To sample from this distribution, we can first integrate
  over \( \mu \) to get the marginal distribution of \( \sigma^2 \). We can
  sample from this distribution to obtain a sample of \( \sigma^2 \), and
  then sample from the conditional distribution of \( \mu \) given
  the value of \( \sigma^2 \) we just sampled.
- Repeating this we can get a sample as large as we want
Normal Data with Both Parameters Unknown

- The integral over \( \mu \) produces an extra factor of \( \sigma \) in the denominator, resulting in:

\[
pd \propto \frac{\sigma}{(\sigma^2)^{n/2+1/2}} \exp \left[ -\frac{S}{2\sigma^2} \right] = \frac{1}{(\sigma^2)^{n/2+1/2}} \exp \left[ -\frac{S}{2\sigma^2} \right]
\]

Normal Data with Both Parameters Unknown

- We have to transform from \( \sigma^2 \) to \( \chi^2 = S/\sigma^2 \). This means multiplying the distribution on \( \sigma^2 \) by the Jacobian, which is

\[
\frac{d\sigma^2}{d\chi^2} \propto \frac{1}{(\chi^2)^{2}} \propto (\sigma^2)^{2}
\]

- Then, mechanically substitute for \( \sigma^2 \). The result is

\[
g(\chi^2 | \mu, y) \propto (\sigma^2)^{2} (\sigma^2)^{n/2+1/2} \exp \left[ -\frac{S + n(\mu - y)}{2\sigma^2} \right] = (\sigma^2)^{n/2-3/2} \exp \left[ -\frac{\chi^2}{2} \right] = (\chi^2)^{n-1/2-1} \exp \left[ -\frac{\chi^2}{2} \right]
\]

- This is a \( \chi^2 \) distribution with \( n-1 \) degrees of freedom

Normal Data with Both Parameters Unknown

- The Jacobian has to be included whenever we change variables, since when we are working with probability densities, integration is always implied. So, suppose we have a distribution on a variable \( \theta \), and wish to convert to a variable \( \phi \). If we were integrating, we would do it by substituting for \( d\theta \) in the integral using the formula

\[
d\theta = \frac{d\theta}{d\phi} d\phi
\]

In this expression, the quantity \( d\theta/d\phi \) is the Jacobian. In the case of more than one variable, a more complicated factor is required.

Normal Data with Both Parameters Unknown

- So, to sample on \( \sigma^2 \), first draw a sample from a standard \( \chi^2 \) distribution on \( n-1 \) degrees of freedom; then calculate \( \hat{\sigma}^2 = S/\chi^2 \).

- We’ve sampled from an inverse \( \chi^2 \) distribution

- Once we have a sample on \( \sigma^2 \), we can draw a sample from the conditional distribution on \( \mu \)

\[
p(\mu | \sigma^2, y) \propto \exp \left[ -\frac{n(\mu - \bar{y})^2}{2\sigma^2} \right] = \exp \left[ -\frac{(\mu - \bar{y})^2}{2(\sigma^2/n)} \right]
\]

- By inspection we see that \( \mu \) is distributed normally with mean \( \bar{y} \) and variance \( \sigma^2/n \)
Normal Data with Both Parameters Unknown

• To illustrate this, we duplicate Albert’s code:

```r
data(marathontimes)
attach(marathontimes)
d=mycontour(normchi2post,c(220,330,500,9000),time)
title(xlab="mean",ylab="variance")
S=sum((time-mean(time))^2)
n=length(time)
sigma2=S/rchisq(1000,n-1)
u=runif(1000,mean=mean(time),sd=sqrt(sigma2)/sqrt(n))
points(mu,sigma2)
```

Normal Data with Both Parameters Unknown

• Albert also computes quantiles:

```r
quantile(mu,c(0.025,0.975))
quantile(sqrt(sigma2),c(0.025,0.975))
```

Bioassay Experiment

• In §4.4, Albert describes the following problem
• Animals are dosed with a compound, with different doses to each group
• The number of animals that die in each group is recorded
• We wish to predict the probability that an animal will die, as a function of the dose

<table>
<thead>
<tr>
<th>Dose (log g/ml)</th>
<th>Deaths</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.86</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>-0.30</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>-0.05</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.73</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Bioassay Experiment

• Comment 1: Notice that Albert has transformed the covariate (dose) by using the log. This changes the range of the covariate from \((0, \infty)\) to \((-\infty, \infty)\), which is generally thought to be a good idea
• Comment 2: Similarly, we are trying to estimate probabilities, which are numbers between 0 and 1. A similar analysis would suggest transforming the probabilities into the range \((-\infty, \infty)\). A standard transformation for this is the logit function:

\[
q = \text{logit}(p) = \log\left(\frac{p}{1-p}\right), \text{ with inverse } p = \frac{e^q}{1+e^q}
\]

• A similar effect can be obtained with the probit function, see §10.3 for an example
Bioassay Experiment

- Therefore, Albert writes this as a logit regression model with a constant term and one linear in the (transformed) dose:
  \[ \text{logit}(p_i) = \beta_0 + \beta_1 x_i \]

- Using the inverse of the logit function, this can be rewritten
  \[ p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \]

Bioassay Experiment

- The corresponding likelihood function, the probability of the data given the parameters (but considered as a function of the parameters) is
  \[ L(\beta_0, \beta_1) = \prod_i p_i^y (1 - p_i)^{n - y_i} \]

- The usual prior on the parameters is flat, so the posterior is proportional to the likelihood

Bioassay Experiment

- Albert illustrates two ways of solving this problem
  - First he does a MLE calculation using the R function `glm` (generalized linear model). This is a model that links the probabilities (needed for the likelihood) to the covariates via a link function (here, the logit function). `glm` is a standard function that allows one to solve such problems with relatively little fuss. The result is point estimates for each parameter and other information such as standard deviations
  - Then he solves the problem in a fully Bayesian way, both by plotting the contours of the posterior distribution and by drawing a sample from the posterior distribution from which means, medians, and quantiles can be inferred

```r
x=c(-0.86,-0.3,-0.05,0.73)
n=c(5,5,5,5)
y=c(0,1,3,5)
data=cbind(x,n,y)
response=cbind(y,n-y)
results=glm(response~x,
            family=binomial)
summary(results)
```
To solve using a fully Bayesian approach, Albert has written a special function, `logisticpost`, to compute the posterior distribution for this two-parameter logistic problem. He plots a contour plot of the posterior using another function he wrote, `mycontour`:

```r
?logisticpost
?mycontour
```

Here’s the code to generate a plot of the posterior (parameters are chosen to make the plot “pretty”):

```r
mycontour(logisticpost,
          c(-4,8,-5,39),data)
title(xlab="beta0",ylab="beta1")
```

Next, he generates a sample from the posterior distribution using another function, `simcontour`, that restricts the points to the range already used, and plots them on the contour plot that we already have. (Note the similarity to the call to `mycontour`)

```r
s=simcontour(logisticpost,
             c(-4,8,-5,39),data,1000)
points(s$x,s$y)
```

From the sample `s`, we can now compute some quantities of interest:
- A glance at the marginal distribution of the slope of the logistic regression, $\beta_1$
  ```r
  plot(density(s$y),xlab="beta1")
  quantile(s$y,c(0.025,0.975))
  ```
Bioassay Experiment

- One parameter of interest is the LD-50, the dose \( x \) such that the probability of death is 50%. The LD-50 equals \( \theta = \frac{\beta_0}{\beta_1} \).
- We can deal with this by generating a sample of \( \theta \). This is trivial to do, from the samples on \( \beta_0 \) and \( \beta_1 \). Just divide!
- From this sample, we can estimate the quantiles for LD-50 by using the R function quantile:

```r
data <- s$x/s$y
hist(data, xlab = "LD-50")
quantile(data, c(0.025, 0.975))
```