Bayesian Reasoning Repetition

Prof. Dr. Ulrich Mansmann
mansmann@ibe.med.uni-muenchen.de
IBE – Department of Medical Informatics, Biometry, and Epidemiology
LMU – Ludwig-Maximilians-Universität München
• The Poisson distribution is used in epidemiology to model the number of rare events in large populations and specific time periods.

• It uses the incidence rate $\lambda$ of a specific event and the total time under observation $T_{tot}$ to calculate the expected number of events in the population of interest: $\lambda \cdot T_{tot}$.

• A very specific feature of the Poisson distribution is that its mean ($\lambda \cdot T_{tot}$) equals its variance ($\lambda \cdot T_{tot}$).
The probability to observe exactly $y$ events in a population with a total time of $T$ patient years and incidence rate $\lambda$ is given by

$$f(y; \lambda) = \frac{(\lambda T)^y}{y!} \exp(-\lambda T), \quad y = 0, 1, \ldots$$
In 1983, ten cases of leukemia were observed in a population of 30000 children who lived within a 10 km radius around a nuclear power plant. The population-specific yearly incidence rate is 10 cases per 100000 children.

What is the probability of observing at least 10 cases in the population of interest using the population-specific incidence rate?

\[
P(\text{cases} \geq 10) = 1 - P(\text{cases} \leq 9) = 1 - P(0) - P(1) - \ldots - P(9)
\]

\[
\lambda \cdot T = (10/100000) \cdot 30000 = 3
\]

\[
P(\text{cases} \geq 10) = 1 - P(\text{cases} \leq 9) = 1 - 0.9989 = 0.0011
\]

> 1 - sum(dpois(0:9, 3))

[1] 0.001102488
Poisson distribution – Example

Try to formulate the result in terms of a classical statistical test.

What is $H_0$?

We will use a one-sided null-hypothesis: $\lambda \leq \lambda_{\text{pop}}$

What is $H_1$?

We will use a one-sided alternative: $\lambda > \lambda_{\text{pop}}$

What is the p-value?

$p = 0.0011$
Poisson distribution – Example

How to address this issue as a Bayesian?

What is the prior distribution?

What is the posterior distribution?

How to use the posterior distribution to address the public health problem?
What is the *prior distribution*?

Use the idea of a conjugate family of priors: The posterior is of the same class as the prior.

The gamma distribution is the conjugate family of distributions when data from Poisson distribution is observed.

If the parameter $\lambda$ of the Poisson distribution has a gamma distributed prior, $\Gamma(\alpha,\beta)$, then the posterior based on the $j$ observations $k_1, ..., k_j$ is again a Gamma distribution:

$$p(\lambda | k) = \text{Gamma}(\lambda | \alpha + k_1 + ... + k_j, \beta + j)$$
Gamma distribution

See Table A.5 of Lesaffre & Lawson, Bayesian Biostatistics

\[ \Gamma_{\alpha,\beta}(x), \ x \geq 0, \ \alpha > 0, \ \beta > 0 \]

\[ \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \]

For \( \alpha=1 \) the Gamma distribution is an exponential distribution.

Mean: \( \alpha/\beta \)

Variance: \( \alpha/\beta^2 \)
How certain are we about the population incidence $\lambda_{\text{pop}}$?
What makes the population incidence, $\lambda_{\text{pop}}$, uncertain?

Prior $\alpha=30$, $\beta=10$ [$\alpha=0.6$, $\beta=0.2$]
Technical note: The Gamma distribution has parameters shape = $\alpha$ and scale = $1/\beta$

Prior $\alpha=30 + 10$, $\beta=10 +1$ [$\alpha=10.6$, $\beta=1.2$]

```
xx<-seq(0,20,by=0.1)
par(mfrow=c(1,2))
plot(xx, dgamma(xx,scale=0.1,shape=30),xlab="",ylab="gamma density",type="l")
lines(xx, dgamma(xx,scale=0.0909,shape=40),col="red")
plot(xx, dgamma(xx,scale=5,shape=0.6),xlab="",ylab="gamma density",type="l")
lines(xx, dgamma(xx,scale=1/1.2,shape=10.6),col="red")
par(mfrow=c(1,1))
```
Poisson distribution – Example

Good knowledge on $\lambda_{\text{pop}}$

Uncertainty on $\lambda_{\text{pop}}$
How to report the results on $\lambda_{\text{pop}}$: 

\[ P(\lambda|k) = \text{Gamma}(\lambda | \alpha + k_1 + \ldots + k_j, \beta + j) \]

Prior $\alpha=30$, $\beta=10$ [$\alpha=0.6$, $\beta=0.2$]

Prior $\alpha=30 + 10$, $\beta=10 + 1$ [$\alpha=10.6$, $\beta=1.2$]

1-$\text{pgamma}(3,\text{scale}=0.0909,\text{shape}=40)$
1-$\text{pgamma}(3,\text{scale}=1/1.2,\text{shape}=10.6)$

Above we calculated the posterior evidence that $\lambda_{\text{pop}}$ is above 3.

Situation with certainty: $\alpha=30$, $\beta=10 : 0.8696496$
Situation with uncertainty: $\alpha=0.6$, $\beta=0.2 : 0.9979639$

Why does a kind of contradiction exist when comparing the Bayesian results with the frequentist approach?
Using alternative (non-conjugated) priors:

```r
model
{
  lambda~dunif(lower,upper)
k~dpois(lambda)
lambda.mod<-lambda-3
count.above<-step(lambda.mod)
}
data
list(k=10,lower=2,upper=4)

init
list(lambda=3)
```

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>count.above</td>
<td>0.8684</td>
<td>0.3381</td>
<td>0.003203</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1001</td>
</tr>
<tr>
<td>lambda</td>
<td>3.502</td>
<td>0.4113</td>
<td>0.003961</td>
<td>2.463</td>
<td>3.613</td>
<td>3.983</td>
<td>10001</td>
<td>10000</td>
</tr>
</tbody>
</table>
Predictive distributions, predictive power

Conditional power versus predictive power

Halperin et al. [CCT, 1982]:

*Conditional power:* Probability to reject the null-hypothesis if study is extended (new patients) conditional to:
   1. *Null-hypothesis is true*
   2. *Alternative is true*

Spiegelhalter et al. [CCT, 1986]:

*Predictive Power:* Non-conditional prediction if study is extended (new patients included): *With the data so far, what is the chance that the trial will end up showing a conclusive result?* Consider all hypotheses which are plausible given the observed data.
Predictive distributions, predictive power

Concept of predictive power:

• $X_f$ - future data when study is extended (not observed – random variable)
• $x_0$ - data observed so far
• $X_f$ and $x_0$ are samples from a population with unknown parameter $\theta$
• $X$ is the overall sample after extension: $X = (x_0, X_f)$
• $R$ - rejection region for the frequentist test: Test statistics $T(X)$ is in $R$, then we reject the null-hypothesis, this is a conclusive result of the study
• Conditional Power: $P[T(X) \in R \mid \theta]$
• Posteriori of $\theta$: $P[\theta \mid x_0]$ using a non-informative prior $P[\theta]$
• Predictive Power: $P[T(X) \in R \mid x_0] = \int P[T(X) \in R \mid \theta]P[\theta \mid x_0]$ 
• Predictive distribution of future data: $P[X_f \mid x_0] = \int P[X_f \mid \theta]P[\theta \mid x_0]$

Bayesian Reasoning
Hyperthermia Study
Projects A1, B1 of SFB 273 – Hyperthermia, interims review in fall 2000

123 patients recruited between June 1994 and June 2000 (~ 20 patients per year)
Data of 113 patients can be analyzed

**Primary endpoint:** Tumor free survival
**Population:** 113 patients with a locally advanced rectum carcinoma
**Treatment:** Preoperative Radio-Chemotherapy (N=50) versus combination of preoperative Hyperthermia and Radio-Chemotherapy (N=63)
**Stratified Randomisation** (using uT3 and uT4)
**Stratified Log-Rank Test:** p = 0.796
Predictive distributions, predictive power

![Tumour free survival graph]

Bayesian Reasoning
Predictive distributions, predictive power
Predictive distributions, predictive power

Parametric survival models

Parametric regression
\[ \log(T) = \theta^T X + \sigma \cdot \log[E] \quad E \sim \text{Exp}[1] \]

Exponential model
\[ \log(T) = -\log(\lambda) + \log[E] \]

Weibull Model
\[ \log(T) = -\log(\lambda) + \frac{1}{\alpha} \cdot \log[E] \quad S(t) = \exp\{-(\lambda \cdot t)^{\alpha}\} \]

Linear Predictor
\[ \theta^T X = \theta_0 + \theta_1 \cdot 1_{\text{HRCT}} \]

<table>
<thead>
<tr>
<th></th>
<th>Exponential</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.9289 (0.707)</td>
<td>4.5413 (0.498)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.0929 (0.447)</td>
<td>0.0422 (0.303)</td>
</tr>
<tr>
<td>Log(1/\alpha)</td>
<td></td>
<td>-0.3906 (0.183)</td>
</tr>
<tr>
<td>\alpha</td>
<td></td>
<td>1.4779</td>
</tr>
</tbody>
</table>
Predictive distributions, predictive power
Predictive distributions, predictive power

**Exponential**

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>deviance</td>
<td>1383.0</td>
<td>47.92</td>
<td>1.484</td>
<td>1297.0</td>
<td>1381.0</td>
<td>1483.0</td>
<td>10001</td>
<td>10000</td>
</tr>
<tr>
<td>teta.0</td>
<td>-5.067</td>
<td>0.3322</td>
<td>0.0154</td>
<td>-5.794</td>
<td>-5.046</td>
<td>-4.463</td>
<td>10001</td>
<td>10000</td>
</tr>
<tr>
<td>teta.1</td>
<td>-0.096</td>
<td>0.4666</td>
<td>0.0221</td>
<td>-1.0</td>
<td>-0.094</td>
<td>0.8364</td>
<td>10001</td>
<td>10000</td>
</tr>
</tbody>
</table>

Deviance at mean(θ): 241.7723

$\text{DIC: } 2 \cdot 1383.0 - 241.7723 = 2524.228$

**Weibull**

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>deviance</td>
<td>241.9</td>
<td>2.41</td>
<td>0.1366</td>
<td>239.1</td>
<td>241.3</td>
<td>247.9</td>
<td>10001</td>
<td>10000</td>
</tr>
<tr>
<td>r</td>
<td>1.495</td>
<td>0.2062</td>
<td>0.0266</td>
<td>0.9708</td>
<td>1.469</td>
<td>2.092</td>
<td>10001</td>
<td>10000</td>
</tr>
<tr>
<td>teta.0</td>
<td>-6.869</td>
<td>1.135</td>
<td>0.1046</td>
<td>-9.234</td>
<td>-6.757</td>
<td>-4.912</td>
<td>10001</td>
<td>10000</td>
</tr>
<tr>
<td>teta.1</td>
<td>-0.099</td>
<td>0.445</td>
<td>0.0192</td>
<td>-0.971</td>
<td>-0.09</td>
<td>0.764</td>
<td>10001</td>
<td>10000</td>
</tr>
</tbody>
</table>

$6.869/1.495 = 4.595\quad 0.099/1.495 = 0.0662$

Deviance at mean(θ): 238.94

$\text{DIC: } 2 \cdot 241.9 - 238.94 = 244.86$
Andersen’s CP Exponential Model

- \( N_i \): Number of events in group i,
- \( X_i \): Total time observed in group i
- \( \lambda_i \): Hazard in group i, estimated by \( \frac{N_i}{X_i} \), Hazard ratio (HR) \( \theta = \frac{\lambda_1}{\lambda_2} \)

Under \( H_0 \): \( \theta=1 \) and the test statistics \( W \) is standard normal distributed \( N(0,1) \).

\[
W = \log\left( \frac{\hat{\theta}}{\sqrt{1/N_1 + 1/N_2}} \right)
\]

Under \( H_A \): \( \theta = \theta_A \) the test statistics \( W \sim N(\mu_A, 1) \)

\[
\mu_A = \log\left( \frac{\theta_A}{\sqrt{1/E_A[N_1] + 1/E_A[N_2]}} \right)
\]

The power function is given by \( \gamma(\theta_A) = \Phi(Z_{\alpha/2}-\mu_A) + 1 - \Phi(Z_{1-\alpha/2}-\mu_A) \).
Andersen’s CP Exponential Model

Study will be continued for T additional time units

\( D_i \): Number of additional events in group \( i \)

\( s_i \): expected additional observation time \( s_i = s_i^{old} + s_i^{new} \)

\[
\begin{align*}
    s_i^{old} &= y_i \cdot \left[ 1 - \exp\left\{ -q_i \cdot T \right\} \right] / q_i \\
    s_i^{new} &= r_i \cdot T / q_i - r_i \cdot \left[ 1 - \exp\left\{ -q_i \cdot T \right\} \right] / (q_i^2)
\end{align*}
\]

\( y_i \) is the number of patients in group \( i \), who will enter the new phase of the study,
\( q_i \) is the combined death and drop-out rate,
\( r_i \) is the number of patients who will be recruited per time unit.

Old: Patients who enter from the first phase into the extension

For fixed \( s_i \) \( D_i \sim N(\lambda_i \cdot s_i, \lambda_i \cdot s_i) \).
Andersen’s CP Exponential Model

Under $H_A: \theta = \theta_A$ the test statistic

$$W_T = \log\left(\frac{N_2 + D_2}{X_2 + s_2}\right) - \log\left(\frac{N_1 + D_1}{X_1 + s_1}\right)$$

is normally distributed with mean

$$\mu_{CP} = \log\left(\frac{\theta_A \cdot \lambda \cdot s_2 + n_2}{x_2 + s_2}\right) - \log\left(\frac{\lambda \cdot s_1 + n_1}{x_1 + s_1}\right)$$

and variance

$$\nu_{CP} = \frac{\theta_A \cdot \lambda \cdot s_2}{(\theta_A \cdot \lambda \cdot s_2 + n_2)^2} + \frac{\lambda \cdot s_1}{(\lambda \cdot s_1 + n_1)^2}$$
Andersen’s CP Exponential Model

The conditional power is given by

\[ \gamma_{CP}(\theta_A) = \Phi\left(\frac{-k_\alpha - \mu_{CP}}{\sqrt{\nu_{CP}}}\right) + 1 - \Phi\left(\frac{k_\alpha - \mu_{CP}}{\sqrt{\nu_{CP}}}\right) \]

with

\[ k_\alpha = Z_{1-\alpha/2} \cdot \text{Var}[W_T | \theta = 1] \approx Z_{1-\alpha/2} \cdot \sqrt{(1/[x_1 + s_1] + 1/[x_2 + s_2]) / \hat{\lambda}} \]

\[ \hat{\lambda} = (\lambda_1 + \lambda_2) / 2 \]
Andersen’s CP Exponential Model

X.1  10
N.1  1516.813
X.2  10
N.2  1664.427
Drop-out 2/12
r  10/12
T  60
Time unit: months
Extension of Hyperthermia Study

Szenario:


05.06.2000

1304 Tage

31.12.2005

Ende des Follow-Up

It is planned to recruit 100 new patients, about 30 per year. The patients enter continuously.
Integrating the conditional power using the posterior of the treatment effect

Integration over the posterior distribution of $\theta$ results in a predictive power of 0.299 (~30%)
Predictive distributions, predictive power

The Bayesian approach does not use subjective apriori assumptions

*What given the data so far, are the chances of getting a conclusive result if the extension of the trial is completed.*

The CP approach uses hypotheses without evaluating their validity based on the data observed so far. Are these hypotheses plausible?
Predictive distributions, predictive power

Halperin M, Gordon KK, Ware JH, Johnson NJ, DeMets DL (1982) *An Aid to Data Monitoring in Long-Term Clinical Trials*, Controlled Clinical Trials, 3, 311-323

Spiegelhalter DJ, Freedman LS, Blackburn PR (1986) *Monitoring Clinical Trials: Conditional or Predictive Power*, Controlled Clinical Trials, 7, 8-17

Andersen PK (1987) *Conditional Power Calculations as an Aid in the Decision Whether to Continue a Clinical Trial*, Controlled Clinical Trials, 8, 67-74
Missing data

Understanding the reasons why data are missing can help with analyzing the remaining data. If values are missing at random, the data sample may still be representative of the population. But if the values are missing systematically, analysis may be harder.

For example, in a study of the relation between IQ and income, participants with an above-average IQ might tend to skip the question ‘What is your salary?’ Analysis may falsely show no association between IQ and salary, while in fact there may be a relationship.

Because of these problems, methodologists routinely advise researchers to design studies to minimize the incidence of missing values.
Missed data

**Missing completely at random**
Values in a data set are **missing completely at random (MCAR)** if the events that lead to any particular data-item being missing are independent both of observable variables and of unobservable parameters of interest, and occur entirely at random. When data are MCAR, the analyses performed on the data are unbiased; however, data are rarely MCAR.

**Missing at random (MAR)** is an alternative, and occurs when the missingness is related to a particular variable, but it is not related to the value of the variable that has missing data. An example is that males are less likely to fill in a depression survey but this has nothing to do with their level of depression.
Missing data

**Missing not at random (MNAR)** is data that is missing for a specific reason (i.e. the value of the variable that's missing is related to the reason it's missing).

An example is if men failed to fill in a depression survey because of their level of depression.
Imputing unobserved data

Rats example
#### Ranking institutions

This example considers mortality rates in 12 hospitals performing cardiac surgery in babies.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>No of ops</th>
<th>No of deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>148</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>119</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>810</td>
<td>46</td>
</tr>
<tr>
<td>E</td>
<td>211</td>
<td>8</td>
</tr>
<tr>
<td>F</td>
<td>196</td>
<td>13</td>
</tr>
<tr>
<td>G</td>
<td>148</td>
<td>9</td>
</tr>
<tr>
<td>H</td>
<td>215</td>
<td>31</td>
</tr>
<tr>
<td>I</td>
<td>207</td>
<td>14</td>
</tr>
<tr>
<td>J</td>
<td>97</td>
<td>8</td>
</tr>
<tr>
<td>K</td>
<td>256</td>
<td>29</td>
</tr>
<tr>
<td>L</td>
<td>360</td>
<td>24</td>
</tr>
</tbody>
</table>

The number of deaths $r_i$ for hospital $i$ are modelled as a binary response variable with *true* failure probability $p_i$:

$$r_i \sim \text{Binomial}(p_i, n_i)$$

We first assume that the *true* failure probabilities are independent (i.e. fixed effects) for each hospital. This is equivalent to assuming a standard non-informative prior distribution for the $p_i$'s, namely:

$$p_i \sim \text{Beta}(1.0, 1.0)$$
Modelling the performance of institutions

model
{
  for ( i in 1 : N ) {
    p[i] ~ dbeta(1.0, 1.0)
    r[i] ~ dbin(p[i], n[i])
  }
}

Data

list(n = c(47, 148, 119, 810, 211, 196, 148, 215, 207, 97, 256, 360),
     r = c(0, 18, 8, 46, 8, 13, 9, 31, 14, 8, 29, 24),
     N = 12)

Inits

list(p = c(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1))
Ranking institutions

Use and initialize rank monitor