Looking at Data and Public Health Problems with Bayesian Glasses

Michael Escobar

Dalla Lana School of Public Health

University of Toronto

m.escobar@utoronto.ca

http://individual.utoronto.ca/Escobar/

Bayesian Statistics. What is it?

(Without going to much into the technicalities.)

Maybe:

- Bayesian Statisticians: Savage, Lindley, DeFinetti, etc.
- Bayesian methods: Posterior distributions, credible regions, shrinkage to the mean.... Also, maybe, Bayesian nets, Bayes factors...
- When speaking about statistical foundations, one often talks about ”The Bayesian paradigm”. What is a a paradigm?

$\{$ Warning: This is my own take...$\}$

What is a Paradigm

- Not discussed much by ”hard” scientist. Mostly by social scientist.
- A paradigm is ”an analytical lens, a say of viewing the world and a framework from which to understand the human experience” [a]

[a] Blackstone, Sociological Inquiry Principles: Qualitative and Quantitative Methods. From the website

Paradigm: Kuhn

From Wikipedia, summarizing Kuhn, 1996, The Structure of Scientific Revolution.

- *What* is to be observed and scrutinized.
- The kind of *questions* that are to be asked and probed for answers in relationship to this subject.
- *How* these questions are to be structured.
- *How* the results of scientific investigations should be interpreted.
- *How* is the experiment to be conducted, and *what* equipment is available to conduct the experiment.

What is the Bayesian paradigm

Simple definition:

- Quantification of uncertainty

Maybe little more accurate:

- Quantification of the uncertainty in the numerical value in the parameter of the model.

Bayesian paradigm

- Uncertainty is expressed as a probability.
- Probability is meant both as a technical mathematical sense and also meant as in common usage. (Bernardo: common definition of probability... ”conditional measure of uncertainty.”)
- Once the “real world” sense of uncertainty is express in terms of mathematics, then one can use mathematical rules to update the belief given new data.
- Then, one can then interpret the mathematical statements back to statements in the real world.

For example consider your belief of the height of the next person to walk through the doorway.

Bayesian paradigm

- Express ones belief in a value before seeing new data. The *prior distribution*.
- Express the likelihood of seeing the data as a function of the parameter. (This is the value of the distribution of seeing this data point given a value of the parameter.) This is the *likelihood function*.
- **Turn the Bayesian Crank:** Use Bayes theory to get the new belief in the value of the parameter. This the *posterior distribution*.
- The main formula:

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

Sometimes ”turning the crank” has some technical issues. However, the *what to do* is obvious.

Simple example

- The prior distribution of μ is a normal distribution with mean μ_0 and variance σ_0^2.
- Observe $X_1, \ldots, X_n$ which are each from a normal distribution with (unknown) mean μ and variance σ_i^2. This information gives us the likelihood function.
- After some math, the posterior distribution of μ given the data is normal with mean

$$\frac{\frac{1}{\sigma_0^2}\mu_0 + \sum_{i=1}^n \frac{1}{\sigma_i^2} X_i}{\frac{1}{\sigma_0^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2}}$$

and variance

$$\left(\frac{1}{\sigma_0^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^{-1}.$$

Note: posterior mean is a weighted average of mean from prior and from the data. The weights are the inverses of the variances. As sample size gets bigger, the weight of the prior disappears.

Micro versus Macro Analyses

Consider what I would call Micro and Macro Analysis

- Micro: One study. A typical paper that one would see in the medical/public health field. Looks at one collection of data from one population with one question and study provides that results of this study.
- Macro: A collection of studies. The idea is to synthesis different results together.

Micro Analysis

Let's consider the simple question of the height of people walking through the doorway.

- Assume that the heights, X_i, have a normal distribution with common mean μ and common variance σ^2.
- In a "typical, non-Bayesian" analysis, one might get the sample mean $\bar{X}$ and the usual estimate S of σ.
- Then, one can estimate μ by $\bar{X}$.
- Also, one can a *95% confidence interval* for μ which would be approximately the interval $[\bar{X} - 1.96S/\sqrt{n}, \bar{X} + 1.96S/\sqrt{n}]$.
- And, of course a confidence interval is..... Well, usually only the students who get an A in the basic statistic course gets this right and no one else really cares what it is or what it means...

Micro Analysis Seen Through a Bayesian Lens

If one adheres to the Bayesian paradigm, then one sees uncertainties as probability distribution and one has a way of incorporating the data to update one's belief (which can be expressed as a probability distributions.)

- Considering the normal model mentioned earlier, then one has an expression for the prior and the likelihood. Also, from that model, one can easily get the posterior.
- One might assume a) that S^2 is a good approximation to σ^2 and b) that the prior variance is much larger than S^2/n. (Note: as n gets large, then both assumptions are very reasonable.)
- Then, through the Bayesian lens, one's new belief in the value of μ is a normal distribution with mean approximately $\bar{X}$ and variance approximately S^2.
- Furthermore, one can say that after seeing the data one believes that there is a 95% probability that the mean is in the interval $[\bar{X} - 1.96S/\sqrt{n}, \bar{X} + 1.96S/\sqrt{n}]$. (This interval is called the credible region.)

data

Observe $\bar{X}$ = 171.5 cm, S = 9.1 cm, n = 71.

Interval for μ: [169.4, 173.6].

Interval for population: [153.7, 189.3].

Micro Analysis Seen Through a Bayesian Lens

Note:

- The "B" students in the basic statistic course usually think that the confidence interval means that there is a 95% probability that μ is in that interval.
- Most of the statistical methods used in the biomedical/public health literature is based on some normal approximation and a similar argument can be made.
- For the non-Bayesian, one starts with sampling distribution and then ends up with estimates and confidence intervals and such. However, the Bayesian starts with beliefs expressed as distributions and ends with updated beliefs which are still expressed as distribution. Therefore, one can take these updated beliefs and feed them into other analyses.

Macro Analysis

The micro analyses are the basic bricks of science. How does one put these bricks together.

- Use one's "wet-wear". People can be very clever.
- Get estimated from one model and plug them into the higher model. However, when one does this, one has trouble propagating the errors. That is, the mean might be good, but the standard error for the model could be very hard to get a handle on.
- With some models, the non-Bayesian methods have been worked out. Example: meta-analysis methods.
- Be Bayesian. The micro-analysis results in our beliefs already converted in probabilities. This means that the beliefs have already been converted into mathematical structures where they can be manipulated mathematically.

Consider the meta-analysis in the next slide.

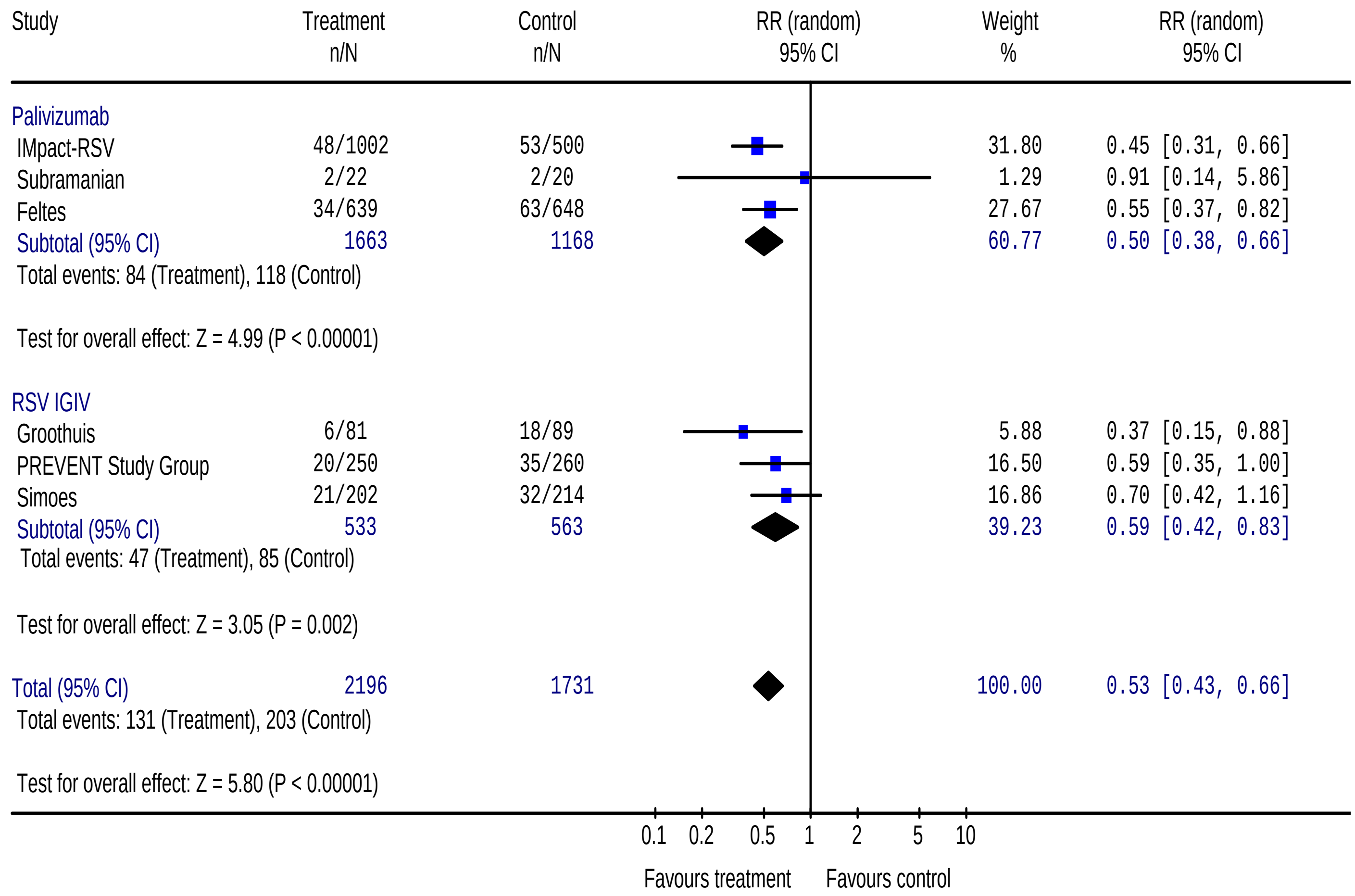

Figure 1: From: Morris, Dzolganovski, Beyene, and Sung, (2009), “A meta-analysis of the effect of antibody therapy for prevention of severe respiratory syncytial virus infection”, BMC Infectious Disease. Downloaded from: http://www.biomedcentral.com/1471-2334/9/106.

Comments on the Meta-Analysis

- Aside: Relative risk are shown, one wants to look at the log(relative risk).
- For the non-Bayesian, the confidence are somewhat interesting to look at, but really are useless to see what the "average effect" is. One needs to get into the parameters of the model. There has been a fair amount of development in the literature to figure out how to do this work.
- For the Bayesian, the log(relative risk)'s for each study are approximately normal. So, this is basically the same problem as before. If one interprets the confidence intervals as Bayesian credible regions, then these intervals give you the belief distribution of the treatment effect in the different study. For the overall effect, one takes a weighted average with the weights equal to one over the variances for the study.
- For the Bayesian, what needs to be done is straight forward.

Technical aside

For the previous meta-analysis, they looked at a "random effect" model. That is, they had the following model:

$$\begin{aligned} (X_i|\mu_i, \sigma_i^2) &\sim N(\mu_i, \sigma_i^2) \\ (\mu_i|\mu, \tau^2) &\sim N(\mu, \tau^2) \end{aligned}$$

So, $(X|\mu, \tau^2, \sigma_i^2) \sim N(\mu, \sigma_i^2 + \tau^2)$. Therefore, the weight for each studies is proportional to $(\sigma_i^2 + \tau^2)^{-1}$.

More Macro Analyses

Besides Meta-Analyses, one can use Bayesian methods to combine other types of information.

For example, one can take output parameters that are estimated in one type of model and then combine them as inputs to for a further model up-stream. This allows one to formally model the uncertainty and variability in the belief in the values of these parameter values.

Some Caveats

“Type III error: the right answer to the wrong question” – John Tukey.

”Computers are useless. They can only give you answers.” – Pablo Picasso(?) (`http://goo.gl/tsu0ps`)

Some Caveats

- Quantification of the uncertainty in the numerical value in the parameter of the model.
- Usually, in public health one has data which are measures of underlying theoretical constructs.

Theoretical Construct

Article on SES and obesity[a]. This paper summarizes several papers looking at the links between SES and obesity.

Different measures of SES used in the papers:

- income and related factors
- education (including schooling and literacy)
- occupation (prestige or status, employment grade or ordered job type.)
- composite indicators
- area level indicators (example: deprivation measures at neighborhood level or regional level instead of individual level.)
- assests and material belongings. (car ownership, renting or owning house.)
- other

[a] McLaren, (2007) “Socioeconomic Status and Obesity”, Epi Review, 29:29:48.

- One usually collects several measures of an underlying construct. (In typical, “old fashion” epidemiology study, one might have an interview of several hours. Since they are measuring the same underlying construct, these variables are usually highly correlated.
- Since there are different possible measures which might be reported in different nodes of the macro model, combining them can be tricky.
- So, one needs to closely collaborate with experts in the field. I would caution against simply data-mining peta-bytes of data.