ADVANCED BAYESIAN MODELING

Motivation: Rat Tumor Data

Rat Tumor Example

Rats have been used in many experiments to evaluate the effect of drugs on development of tumors.

Tumors may develop even if no drug is administered (zero dose).

What is the baseline probability of a tumor?

A survey of 71 different experiments is conducted, recording

- Number of rats in control (zero-dose) group
- Number of control group rats developing a tumor

(More: BDA3, Sec. 5.1, Tarone, 1982)

Let

$$
\begin{aligned}
n_j &= \text{total number of rats in control group of experiment } j \\
y_j &= \text{number in control group of experiment } j \text{ that develop a tumor} \\
j &= 1, \ldots, 71
\end{aligned}
$$

Data pairs (y_j, n_j):

$$(0, 20), \quad (0, 20), \quad (0, 20), \quad \ldots, \quad (9, 24), \quad (4, 14)$$

Assume n_j fixed, but y_j random.

If control group rats in experiment j develop tumors independently, and with the same probability,

$$y_j \sim \text{Bin}(n_j, \, ?\,)$$

If

$$\theta_j = \text{control-group tumor probability in experiment } j$$

then

$$y_j \mid \theta_j \sim \text{Bin}(n_j, \theta_j)$$

If there was only one experiment, could use methods similar to earlier example (dog-walking service)

What can we do with data from *several* experiments?
What assumptions should we make?

Option 1: Same probability for all experiments ($\theta_j = \theta$)

Option 2: Completely unrelated probabilities for different experiments

Option 3: Compromise: Distinct, but related, probabilities

Option 1: Common Probability θ

Just pool data into one binomial sample:

$$y_1 + \cdots + y_{71} \mid \theta \quad \sim \quad \text{Bin}(n_1 + \cdots + n_{71}, \theta)$$

Give a prior to θ (e.g., beta distribution), and compute posterior.

Drawbacks:

- Contradicted by data (results not shown here)
- No way to assess variability among experiments

Option 2: Unrelated Probabilities θ_j

Analyze each experiment separately (with its own prior).

Drawbacks:

- Limited by precision of each individual experiment
- No Bayesian way to answer questions about *overall* probability
- No way to predict result of a *new* experiment

Option 3: Compromise

Allow separate tumor probabilities

$$\theta_1, \ \ldots, \ \theta_{71}$$

but assume they have something in common.

Idea: Regard the experiments as if sampled from a population of possible experiments.

Then the θ_js are independently random from a common (population) distribution.

Advantages:

- Can define "overall" probability of tumor
- Can assess variability among experiments
- Can possibly improve estimation of individual θ_js through a more informative prior (by pooling information across experiments – later)

But distribution of θ_js is unknown.

Need a higher-level model to make inference based on the data.

Leads to idea of *hierarchical models* ...