More Math Stuff

Collapsing tables

In lecture 30, we gave conditions that imply that the marginal and conditional OR's are the same. We show this for a 2x2x2 table below, with factors A, B, C.

Let μ_{ijk} be the mean count for the ijk cell. Since we are dealing with a 2x2x2 table, we don't need subscripts on the main effects and interactions (we can write $\alpha_2 = \alpha$, $\beta_2 = \beta$ and so on). Recall that any main effect or interaction with a 1 subscript is zero. Thus

$$\begin{aligned}
\log(\mu_{111}) &= (Int),\\
\log(\mu_{211}) &= (Int) + \alpha,\\
\log(\mu_{121}) &= (Int) + \beta,\\
\log(\mu_{221}) &= (Int) + \alpha + \beta + \alpha\beta,\\
\log(\mu_{112}) &= (Int) + \gamma,\\
\log(\mu_{212}) &= (Int) + \alpha + \gamma + \alpha\gamma,\\
\log(\mu_{122}) &= (Int) + \beta + \gamma + \beta\gamma,\\
\log(\mu_{222}) &= (Int) + \alpha + \beta + \gamma + \alpha\beta + \beta\gamma + \alpha\gamma + \alpha\beta\gamma.
\end{aligned}$$

Note also that e.g. $\alpha\beta$ in this context does not mean α multiplied by β, rather it is a single number representing the interaction between α and β.

The marginal table of A and B (collapsing over C) has probabilities

$$\begin{aligned}
P(A = i, B = j) &= \pi_{ij+}\\
&= \pi_{ij1} + \pi_{ij2}\\
&= \frac{\mu_{ij1} + \mu_{ij2}}{\mu_{111} + \cdots + \mu_{222}}
\end{aligned}$$

so the odds ratio between A and B in the collapsed table is

$$\begin{aligned}
&\frac{(\pi_{221} + \pi_{222})(\pi_{111} + \pi_{112})}{(\pi_{121} + \pi_{122})(\pi_{211} + \pi_{212})}\\
&= \frac{(\mu_{221} + \mu_{222})(\mu_{111} + \mu_{112})}{(\mu_{121} + \mu_{122})(\mu_{211} + \mu_{212})}\\
&= \frac{(e^{(Int)+\alpha+\beta+\alpha\beta} + e^{(Int)+\alpha+\beta+\gamma+\alpha\beta+\beta\gamma+\alpha\gamma+\alpha\beta\gamma}) \times (e^{(Int)} + e^{(Int+\gamma)})}{[e^{(Int)+\beta} + e^{(Int)+\beta+\gamma+\beta\gamma}) \times (e^{(Int)+\alpha} + e^{(Int)+\alpha+\gamma+\alpha\gamma})}\\
&= \frac{e^{(Int)+\alpha+\beta+\alpha\beta} e^{(Int)} [1 + e^{(\text{Int})+\gamma+\beta\gamma+\alpha\gamma+\alpha\beta\gamma}][1 + e^{(\text{Int})+\gamma}]}{e^{(Int)+\beta} e^{(Int+\alpha)} [1 + e^{(\text{Int})+\gamma+\beta\gamma}][1 + e^{(\text{Int})+\gamma+\alpha\gamma}]}\\
&= \frac{e^{\alpha\beta} [1 + e^{(\text{Int})+\gamma+\beta\gamma+\alpha\gamma+\alpha\beta\gamma}][1 + e^{(\text{Int})+\gamma}]}{[1 + e^{(\text{Int})+\gamma+\beta\gamma}][1 + e^{(\text{Int})+\gamma+\alpha\gamma}]}
\end{aligned}$$

Suppose that C is conditionally independent of A, given B. Then $\alpha\beta\gamma = 0$ and $\alpha\gamma = 0$. Substituting this in the expression above gives

$$\frac{e^{\alpha\beta}[1+e^{(\text{Int})+\gamma+\beta\gamma}][1+e^{(\text{Int})+\gamma}]}{[1+e^{(\text{Int})+\gamma+\beta\gamma}][1+e^{(\text{Int})+\gamma}]} = e^{\alpha\beta}.$$

When $\alpha\beta\gamma = 0$, this is just the conditional odds ratio between A and B given C (see the formula in the last Math Stuff handout).

A similar argument works when C is conditionally independent of B, given A.

Maximising the Poisson log-likelihood in the flying bomb example

Let y_i be the number of squares hit by i flying bombs, i = 0, 1, 2, 3, 4, 5. The log-likelihood under the model is $\sum_{i=0}^{5} y_i \log \pi_i(\mu)$ where $\pi_i(\mu)$ is the probability that a Poisson variable Y with mean μ has value $Y = i$, namely

$$\pi_i(\mu) = \frac{e^{-\mu}\mu^i}{i!}.$$

We need to maximize the log-likelihood as a function of μ. The log likelihood is

$$\sum_{i=0}^{5} y_i \log \pi_i(\mu) = \sum_{i=0}^{5} y_i (i \log(\mu) - \mu).$$

Differentiating with respect to μ and equating the derivative to zero gives $\sum_{i=0}^{5} y_i (\frac{i}{\mu} - 1) = 0$ which has solution $\mu = \frac{\sum_{i=0}^{5} i y_i}{\sum_{i=0}^{5} y_i}$, i.e. the mean number of hits per square. This is the value that maximizes the log-likelihood since the second derivative is negative.