# Coursework 8: STAT 570

Philip Pham

November 26, 2018

1. Consider n experiments with $Z_{ij}$ , $j = 1, 2, \ldots, N_i$, the binary outcomes within cluster (experiment) $i$ with $Y_i = \sum_{j=1}^{N} Z_{ij}$, $i = 1, \ldots, n$.

   (a) By writing
   $$\operatorname{var}(Y_i) = \sum_{j=1}^{N_i} \operatorname{var}(Z_{ij}) + \sum_{j=1}^{N_i} \sum_{j \neq k} \operatorname{cov}(Z_{ij}, Z_{ik}), \tag{1}$$
   show that
   $$\operatorname{var}(Y_i) = N_i p_i (1 - p_i) \times \left[1 + (N_i - 1)\tau_i^2\right], \tag{2}$$
   where $p_i = \mathbb{E}[Z_{ij}]$ and $\tau_i^2$ is the correlation of outcomes within cluster $i$.

   **Solution:** Using the variance for a Bernoulli random variable and the definition of the correlation coefficient, we have that
   $$\begin{aligned} \operatorname{var}(Z_{ij}) &= p_i (1 - p_i) \\ \operatorname{cov}(Z_{ij}, Z_{ik}) &= \tau_i^2 p_i (1 - p_i) \text{ for } j \neq k. \end{aligned} \tag{3}$$
   Since $Z_{ij}$ are identically distributed for different $j$, we can rewrite Equation 1 as
   $$\operatorname{var}(Y_i) = N_i \operatorname{var}(Z_{i1}) + N_i (N_i - 1) \operatorname{cov}(Z_{i1}, Z_{i2}). \tag{4}$$
   Applying Equation 3 to Equation 4, we have the result
   $$\begin{aligned} \operatorname{var}(Y_i) &= N_i p_i (1 - p_i) + N_i (N_i - 1) \tau_i^2 p_i (1 - p_i) \\ &= N_i p_i (1 - p_i) \times \left[1 + (N_i - 1)\tau_i^2\right] \end{aligned}$$
   as desired.

   (b) Consider the model
   $$\begin{aligned} Y_i \mid q_i &\sim \operatorname{Binomial}(N_i, q_i) && (5) \\ q_i &\sim \operatorname{Beta}(a_i, b_i), && (6) \end{aligned}$$
   where we can parameterize as $a_i = dp_i$, $b_i = d(1 - p_i)$, so that
   $$\begin{aligned} \mathbb{E}[q_i] = p_i &= \frac{a_i}{d} && (7) \\ \operatorname{var}(q_i) &= \frac{p_1 (1 - p_i)}{d + 1}. && (8) \end{aligned}$$
   Obtain the marginal moments and show that the variance is of the form in Equation 2, and identify $\tau_i^2$.

**Solution:** We have that

$$\begin{aligned}
\mathbb{P}\left(Y_i = y\right) &= \int_0^1 \mathbb{P}\left(Y_i \mid q_i\right) p\left(q_i\right) \, \mathrm{d}q_i \\
&= \int_0^1 \left(\binom{N_i}{y} q_i^y \left(1 - q_i\right)^{N_i - y}\right) \left(\frac{\Gamma\left(a_i + b_i\right)}{\Gamma\left(a_i\right)\Gamma\left(b_i\right)} q_i^{a_i - 1} \left(1 - q_i\right)^{b_i - 1}\right) \mathrm{d}q_i \\
&= \binom{N_i}{y} \frac{\Gamma\left(a_i + b_i\right)}{\Gamma\left(a_i\right)\Gamma\left(b_i\right)} \int_0^1 q_i^{a_i + y - 1} \left(1 - q_i\right)^{b_i + N_i - y - 1} \, \mathrm{d}q_i \\
&= \binom{N_i}{y} \left(\frac{\Gamma\left(a_i + b_i\right)}{\Gamma\left(a_i\right)\Gamma\left(b_i\right)}\right) \left(\frac{\Gamma\left(y + a_i\right)\Gamma\left(N_i - y + b_i\right)}{\Gamma\left(N_i + a_i + b_i\right)}\right), \qquad (9)
\end{aligned}$$

so $Y_i \sim \text{BetaBinomial}\left(N_i, a_i, b_i\right)$.
Using Equation 9, the expectation of $Y_i$ is

$$\mathbb{E}\left[Y_i\right] = \sum_{y=0}^{N_i} y \mathbb{P}\left(Y_i = y\right) = \sum_{y=1}^{N_i} y \mathbb{P}\left(Y_i = y\right). \qquad (10)$$

Note that when $N_i = 1$, Equation 10 trivially becomes $a_i / \left(a_i + b_i\right)$. In general, we can show that $\mathbb{E}\left[Y_i\right] = N_i \dfrac{a_i}{a_i + b_i}$. With the $N_i = 1$ base case established, we now have

$$\begin{aligned}
\mathbb{E}\left[Y_i\right] &= \sum_{y=1}^{N_i} y \mathbb{P}\left(Y_i = y\right) \\
&= \sum_{y=1}^{N_i} y \binom{N_i}{y} \left(\frac{\Gamma\left(a_i + b_i\right)}{\Gamma\left(a_i\right)\Gamma\left(b_i\right)}\right) \left(\frac{\Gamma\left(y + a_i\right)\Gamma\left(N_i - y + b_i\right)}{\Gamma\left(N_i + a_i + b_i\right)}\right) \\
&= \frac{N_i}{N_i - 1 + a_i + b_i} \sum_{y=1}^{N_i} \left(y - 1 + a_i\right) \text{BetaBinomial}_{N_i - 1, a_i, b_i}\left(y - 1\right) \\
&= \frac{N_i}{N_i - 1 + a_i + b_i} \left(\frac{\left(N_i - 1\right) a_i}{a_i + b_i} + a_i\right) = N_i \frac{a_i}{a_i + b_i} \qquad (11)
\end{aligned}$$

as expected. Substituting, we have that

$$\mathbb{E}\left[Y_i\right] = N_i \frac{dp_i}{dp_i + d\left(1 - p_i\right)} = N_i p_i. \qquad (12)$$

Using a similar strategy as in Equation 11, we have that

$$\mathbb{E}\left[Y_i\left(Y_i - 1\right)\right] == \mathbb{E}\left[Y_i^2\right] - \mathbb{E}\left[Y_i\right]. \qquad (13)$$