Derivations of Equations 9 and 10

To derive equation 9 note that:

\[
P(#m|\#B, \#U, \theta_{B,f}) = \int_0^1 \sum_{i=0}^{#m} [\text{bin}(\#B, i, \theta_{B,f}) \times \text{bin}(\#U, #m - i, \theta_{U,f})] \ d\theta_{U,f} \tag{1}
\]

\[
= \sum_{i=0}^{#m} [\text{bin}(\#B, i, \theta_{B,f}) \times \int_0^1 \text{bin}(\#U, #m - i, \theta_{U,f}) \ d\theta_{U,f}] \tag{2}
\]


\[
\int_0^1 \text{bin}(\#U, #m - i, \theta_{U,f}) \ d\theta_{U,f} \tag{3}
\]

\[
= \frac{\Gamma(#m - i + 1)\Gamma(#U - (#m - i) + 1)}{\Gamma(#U + 2)} \tag{4}
\]

\[
= C(#U, #m - i) \frac{\Gamma(#m - i + 1)\Gamma(#U - (#m - i) + 1)}{\Gamma(#U + 2)} \tag{5}
\]

\[
\frac{1}{#U + 1} \tag{6}
\]

Here, \(C(a, b)\) is the number of combinations of \(a\) items taken \(b\) at a time, and \(\Gamma(x)\) is Euler’s gamma function which equals \((x - 1)!\) for positive integers. Putting this together (and moving the constant out of the summation) gives

\[
P(#m|\#B, \#U, \theta_{B,f}) = \frac{1}{#U + 1} \sum_{i=0}^{#m} \text{bin}(\#B, i, \theta_{B,f}) \tag{9}
\]

Equation 10 assumed that the counts of the findings, \(\{#f = c_f\}\), are conditionally independent. To show this, suppose there are findings \(f_1, \ldots, f_n\). We want to show that:

\[
P(#f_1 = c_1, \ldots, #f_n = c_n|\#B, \#U) = \prod_{i=1}^{n} P(#f_i = c_i|\#B, \#U) \tag{10}
\]

We will show this by induction on \(n\). It is trivial for \(n = 1\). Suppose it is true for \(n - 1\). We will suppress the conditions \(\#B, \#U\) which are common to all formulas, and assume they are part of the knowledge base.
We can express $P(\#f_1 = c_1, \ldots, \#f_n = c_n)$ as a sum of probabilities of truth-assignments to the $\{f_{i,j}\}$ that obey the constraints $\#f_1 = c_1, \ldots, \#f_n = c_n$. We can also separate each of these probabilities into a product of two terms where the left term is the probability of a truth-assignment to the $\{f_{1,j}\}$ and the right term is the probability of a truth-assignment to the $\{f_{i>1,j}\}$.

Let $\{\alpha_p\}$ be the set of all truth assignments to $\{f_{1,j}\}$ that satisfy the constraints, and let $\{\beta_q\}$ be the set of all truth assignments to $\{f_{i>1,j}\}$ that satisfy the constraints. Then:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \sum_p \sum_q P(\alpha_p \cup \beta_q)$$  \hspace{1cm} (11)

By the independence assumptions:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \sum_p \sum_q P(\alpha_p)P(\beta_q)$$  \hspace{1cm} (12)

Undistributing the $\{\alpha_p\}$:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \sum_p [P(\alpha_p) \sum_q P(\beta_q)]$$  \hspace{1cm} (13)

The inner sum is over the set of probabilities of truth assignments that satisfy $\#f_2 = c_1 \ldots, \#f_n = c_n$, so:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \sum_p [P(\alpha_p)[P(\#f_2 = c_2 \ldots, \#f_n = c_n)]]$$  \hspace{1cm} (14)

Undistibute again to get:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \left[\sum_p P(\alpha_p)\right][P(\#f_2 = c_2 \ldots, \#f_n = c_n)]$$  \hspace{1cm} (15)

Now, the sum is over the set of probabilities of truth assignments that satisfy $\#f_1 = c_1$, so:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = P(\#f_1 = c_1) P(\#f_2 = c_2 \ldots, \#f_n = c_n)$$  \hspace{1cm} (16)

Finally, applying the inductive hypothesis gives:

$$P(\#f_1 = c_1, \ldots, \#f_n = c_n) = \prod_{i=1}^{n} P(\#f_i = c_i)$$  \hspace{1cm} (17)