Recall Binomial $\binom{5}{1} p$

\[ \hat{p} = \frac{1}{5}, \quad \text{for } x = 1 \]
Solving the likelihood equation(s)

\[ \nabla \ell(\theta) = 0 \]

is "just" a numerical analysis problem — various search algorithms exist (bisection, coordinate ascent, Newton-Raphson, etc.) — one algorithm that has statistical motivation is the E-M algorithm — we'll introduce it in class.

**Example:** \( k \) possible outcomes, \( n \) independent identical trials, \( P_j \) = probability that any trial produces outcome \( j \)

\[ X_{ij} = I[ \text{trial } i \text{ produces outcome } j ] \]
Suppose $p$ has each $p_j \geq 0$ with $\sum p_j = 1$.

The pmf for $X$ is

$$f(x | p) = \sum_{i=1}^{k} p_{ij}$$

for $x$ with each $x_{ij} = 0$ or $1$

$$\sum x_{ij} = 1$$

where $n_j = \sum x_{ij}$

$n_j = 0$ integers with $\sum n_j = n$

In this problem, the MLE of $p = (p_1, p_2, \ldots, p_k)$ turns out to be

$$\hat{p} = \left( \frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n} \right)$$

To make this plausible, consider a case where each $n_j > 0$ (other cases need to be argued separately... not here.)
\[ \log f(x \mid p) = \sum_{j=1}^{k-1} n_j \log p_j + n_k \log \left(1 - \sum_{j=1}^{k-1} p_j \right) \]

So if (for \( j = 1, 2, \ldots, k-1 \))

\[ \frac{\partial}{\partial p_j} \log f(x \mid p) = \frac{n_j}{p_j} - \frac{n_k}{1 - \sum_{j=1}^{k-1} p_j} \]

\[ = \frac{n_j}{p_j} - \frac{n_k}{P_k} \]

So the likelihood equations are

\[ \frac{n_j}{p_j} = \frac{n_k}{P_k} \quad j = 1, 2, \ldots, k-1 \]

\[ p_j = \frac{n_j}{n_k} P_k \quad j = 1, 2, \ldots, k-1 \]
\[ 1 - P_k = \sum_{j=1}^{k-1} P_j = \frac{n-n_k}{n} \quad P_k = \frac{n}{n_k} P_k - P_k \]

\[ P_k = \frac{n_k}{n} \]

and then \[ P_j = \frac{n_j}{n} \]

But now what if, e.g., \( k = 4 \) and I don't get see all of \( X \) ... instead the information I get is:
For 10 trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Information</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>outcome is 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2 or 4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 or 3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 or 2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

I don't have $X_3 = (X_{31}, X_{32}, X_{33}, X_{34})$
rather I observe

\[ Y_3 = (X_{31}, X_{32}, X_{32} + X_{34}) \]

maximum likelihood for \( p = (P_1, P_2, P_3, P_4) \)?

the likelihood function based on \( Y \) is

\[ L_y(p) = P_1^2 P_2^2 P_3^2 (1 - P_1 - P_2 - P_3) \]
\[ \times (1 - P_1 - P_3) (P_2 + P_3) (P_1 + P_2) \]

\[ P_2 + P_4 \]

I could do numerical analysis here looking for \( P_1, P_2, P_3 \) to optimize \( L_y(p) \)

instead try to do something that makes use of \( L_x(p) \) instead of \( L_y(p) \)
\[ L(x) = p_1^{n_1} p_2^{n_2} p_3^{n_3} (1-p_1-p_2-p_3)^{n_4} \]

and \[ L_x(p) = \log L(x) \]

\[ = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n_4 \log (1-p_1-p_2-p_3) \]

This is easy to maximize ... how to use? I know for the data in hand

\[
\begin{align*}
  n_1 &= 2 + x_{81} \\
  n_2 &= 2 + x_{32} + x_{42} + x_{82} \\
  n_3 &= 2 + x_{43} \\
  n_4 &= 1 + x_{34}
\end{align*}
\]