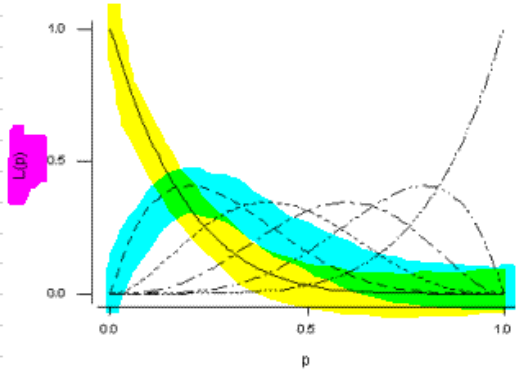
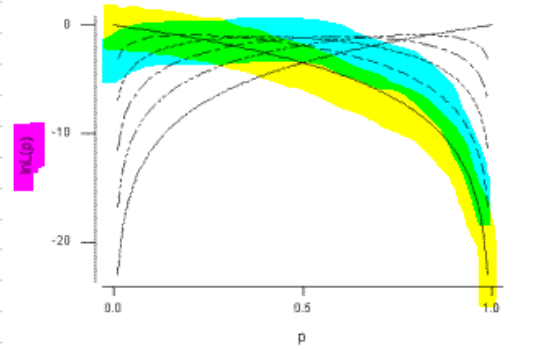
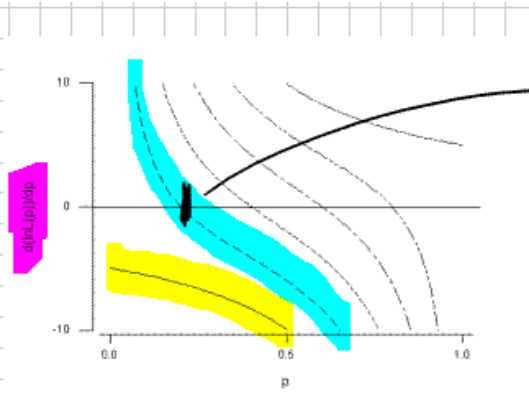
Stat 543 2-9-05

Recall Binomial (5, p)







$\hat{p} = \frac{1}{5}$ 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

<u>Example</u> 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 $\sum p_j = 1$

pmf for X is

$$f(x \mid p) = \prod_{i=1}^{k} p_j^{n_j}$$

for x with each $x_{ij} = 0$ or 1, $\sum_j x_{ij} = 1$, $n_j = 0$ integers with $\sum n_j = n$

where $n_j = \sum_i x_{ij}$

In this problem 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)

$$\log f(x|p) = \sum_{j=1}^{k-1} n_j \log p_j + n_k \log\left(1 - \sum_{j=1}^{k-1} p_j\right)$$

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

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

$$= \frac{n_j}{p_j} - \frac{n_k}{p_k}$$

So the likelihood equations are

$$\frac{n_j}{p_j} = \frac{n_k}{p_k} \qquad j = 1, 2, \ldots, k-1$$

$$p_j = \frac{n_j}{n_k} p_k \qquad j = 1, 2, \ldots, k-1$$

Sum to get

$$1 - p_k = \sum_{j=1}^{k-1} p_j = \frac{n - n_k}{n_k} p_k = \frac{n}{n_k} p_k - p_k$$

$$\therefore \quad p_k = \frac{n_k}{n}$$

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

Trial 1	Information	In terms of X_{ij}'s
1	outcome is 1	$X_{11} = 1$
2	" " 3	$X_{23} = 1$
3	" " 2 or 4	$X_{32} + X_{34} = 1$
4	" " 2	$X_{42} = 1$
5	" " 3	$X_{53} = 1$
6	" " 2 or 3	$X_{62} + X_{63} = 1$
7	" " 1	$X_{71} = 1$
8	" " 1 or 2	$X_{81} + X_{82} = 1$
9	" " 2	$X_{92} = 1$
10	" " 4	$X_{104} = 1$

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

rather I observe

$$Y_3 = (X_{31}, X_{33}, X_{32} + X_{34})$$

---

maximum likelihood for $p = (p_1, p_2, p_3, p_4)$?

the likelihood function based on Y is

$$L_Y(p) = p_1^2 \, p_2^2 \, p_3^2 \, (1 - p_1 - p_2 - p_3)$$
$$\times \underbrace{(1 - p_1 - p_3)}_{p_2 + p_4} (p_2 + p_3)(p_1 + p_2)$$

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 $L_Y(p)$

$$L_x(p) = p_1^{n_1} p_2^{n_2} p_3^{n_3} (1 - p_1 - p_2 - p_3)^{n_4}$$

and $\ell_x(p) = \log L_x(p)$

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

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

$$n_1 = 2 + X_{81}$$
$$n_2 = 2 + X_{32} + X_{62} + X_{82}$$
$$n_3 = 2 + X_{63}$$
$$n_4 = 1 + X_{34}$$