Ex: Coin with unknown bias $\theta = P(\text{heads})$.
Flip it independently, getting outcomes
$x_1, x_2, \ldots, x_n$, where $x_i \in \{H, T\}$.
Let n_0 be the number of tails and n_1 the
number of heads, ~~n_0~~ where $n_0 + n_1 = n$.

$$L(x_1, \ldots, x_n \mid \theta) = (1-\theta)^{n_0}\,\theta^{n_1} \qquad \text{factor } \binom{n}{n_1}?$$

$$\ln L(x_1, \ldots, x_n \mid \theta) = n_0 \ln(1-\theta) + n_1 \ln \theta$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, \ldots, x_n \mid \theta) = -\frac{n_0}{1-\theta} + \frac{n_1}{\theta} = 0$$

$$\frac{n_1}{\hat\theta} = \frac{n_0}{1-\hat\theta}$$

$$n_1(1-\hat\theta) = n_0\,\hat\theta$$

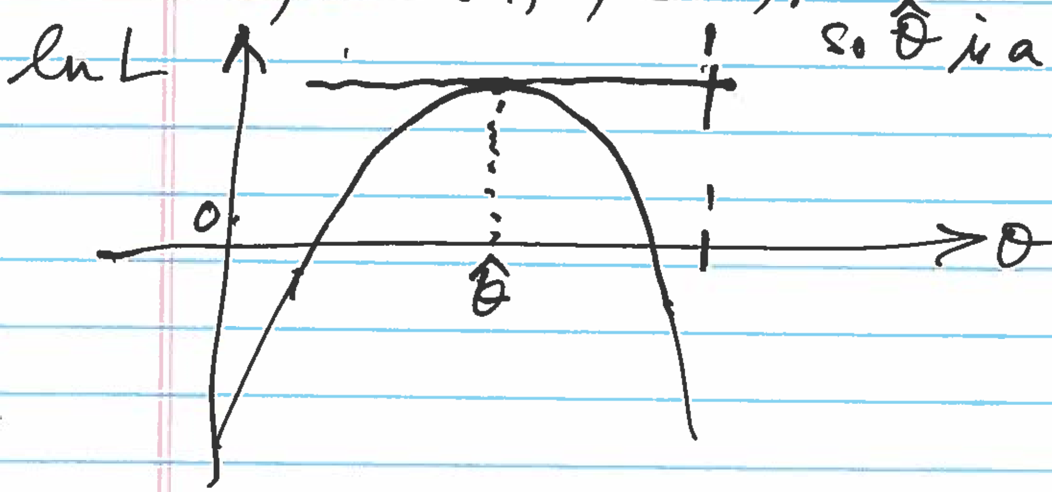
$$n_1 = n_0\hat\theta + n_1\hat\theta = (n_0+n_1)\hat\theta = n\hat\theta$$

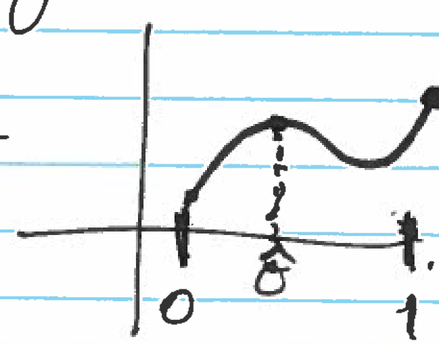
$$\hat\theta = \frac{n_1}{n}$$, the fraction of samples that were H

Is $\hat\theta$ a max or min or inflection point?

$$\frac{\partial^2}{\partial \theta^2} \ln L(x_1, \ldots, x_n \mid \theta) = -\frac{n_0}{(1-\theta)^2} - \frac{n_1}{\theta^2} < 0$$

Because this is negative for all values of
θ, $\ln L(x_1, \ldots, x_n \mid \theta)$ is concave downward.
So $\hat\theta$ is a global max.





Ex: Suppose $x_1, x_2, \dots x_n$ are independent samples from $N(\mu, \sigma^2)$, where both $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ are unknown

$$L(x_1, \dots, x_n | \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \cdot e^{-(x_i - \theta_1)^2 / 2\theta_2}$$

$$\ln L(x_1, \dots, x_n | \theta_1, \theta_2) = \sum_{i=1}^{n} \left(-\frac{1}{2} \ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \right)$$

$$\frac{\partial}{\partial\theta_1} \ln L(x_1, \dots, x_n | \theta_1, \theta_2) = \sum_{i=1}^{n} + \frac{\cancel{2}(x_i - \theta_1)}{\cancel{2}\theta_2} = 0$$

$$\sum_{i=1}^{n} (x_i - \hat{\theta}_1) = 0$$

$$n\hat{\theta}_1 = \sum_{i=1}^{n} x_i$$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i, \text{ the sample mean}$$

$$\frac{\partial^2}{\partial\theta_1^2} \ln L(x_1, \dots, x_n | \theta_1, \theta_2) = \sum_{i=1}^{n} \frac{-1}{\theta_2} < 0$$

so $\hat{\theta}_1$ ~~is a max~~ maximizes $\ln L(--)$

Why was it ok to multiply densities as though they were probabilities?

$$P\left(x_i - \frac{\delta}{2} \le X \le x_i + \frac{\delta}{2}\right) \approx f(x_i) \cdot \delta$$

δ disappears when $\frac{\partial}{\partial\theta} \ln L \dots$