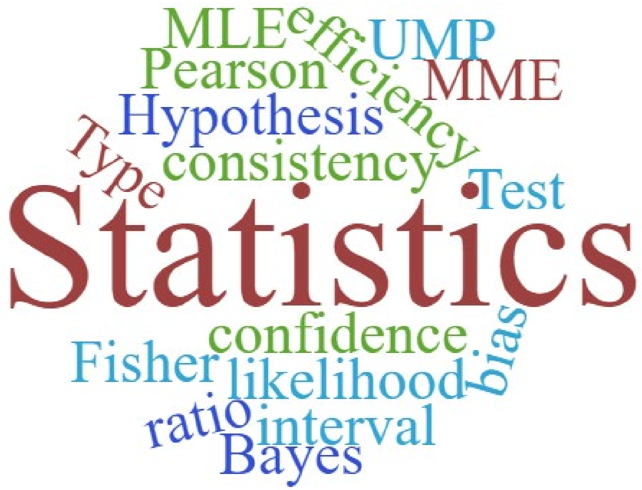
Math 181A: Mathematical Statistics



Learning Goals:

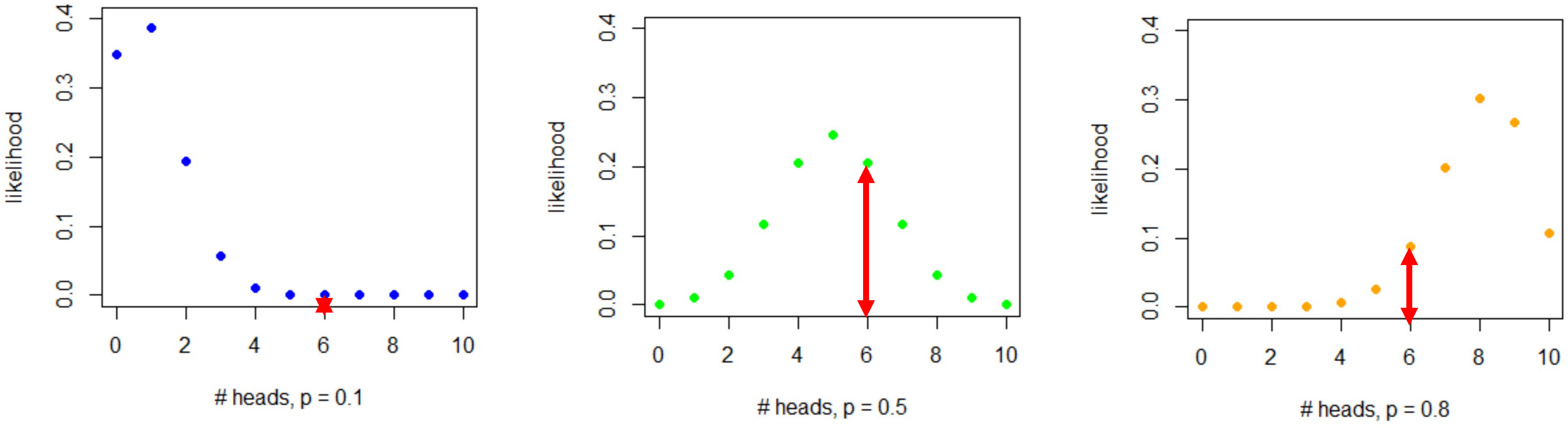
- Get an intuitive sense for the idea of likelihood
- Learn to write the likelihood expression for *n* data points from a given pmf/pdf
- Practice the steps in finding an MLE
- Internalize pro tips that make this complex process easier, quicker, and better displayed on the page

David Quarfoot
Associate Teaching Professor
Email: dquarfoot@ucsd.edu
Office: APM 5220

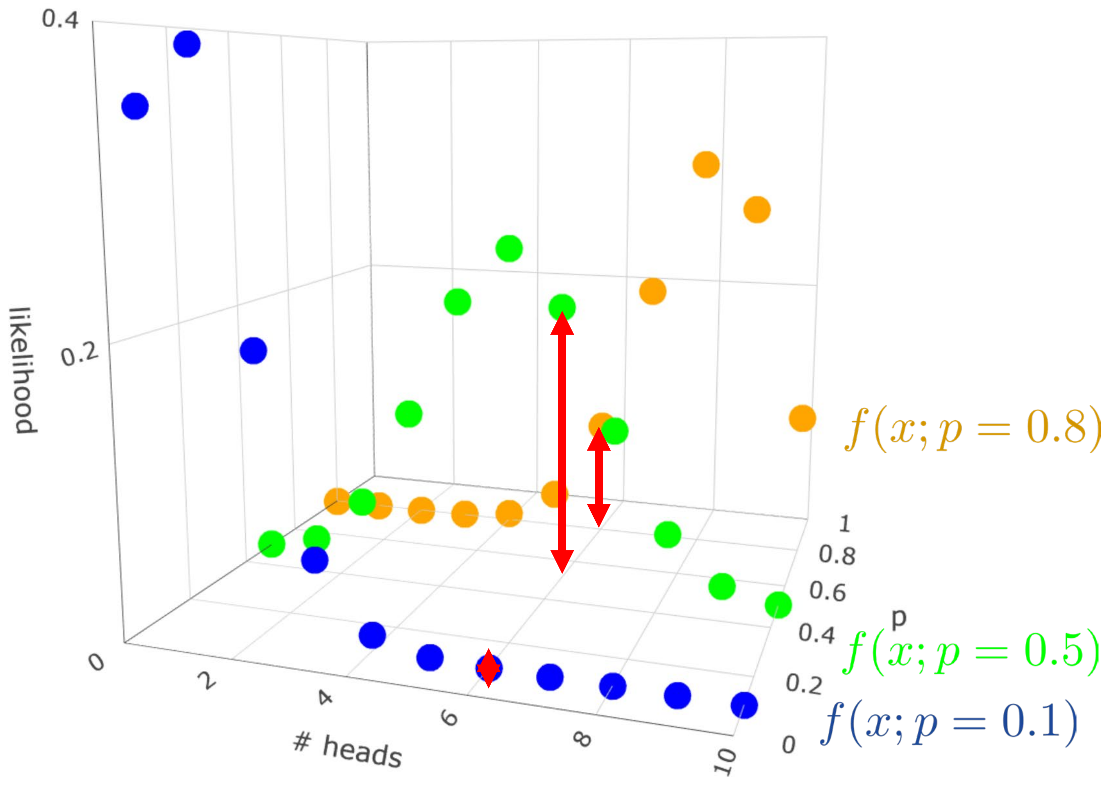
Maximum Likelihood Estimators/Estimates (MLEs)

Quick Q: A coin with unknown $p = p_{heads}$ is flipped 10 times and comes up heads 6 times. Which *of these* values for p feels most likely and why: 0.1, 0.5, or 0.8? Of *all possible* choices for p, what seems most reasonable to you?

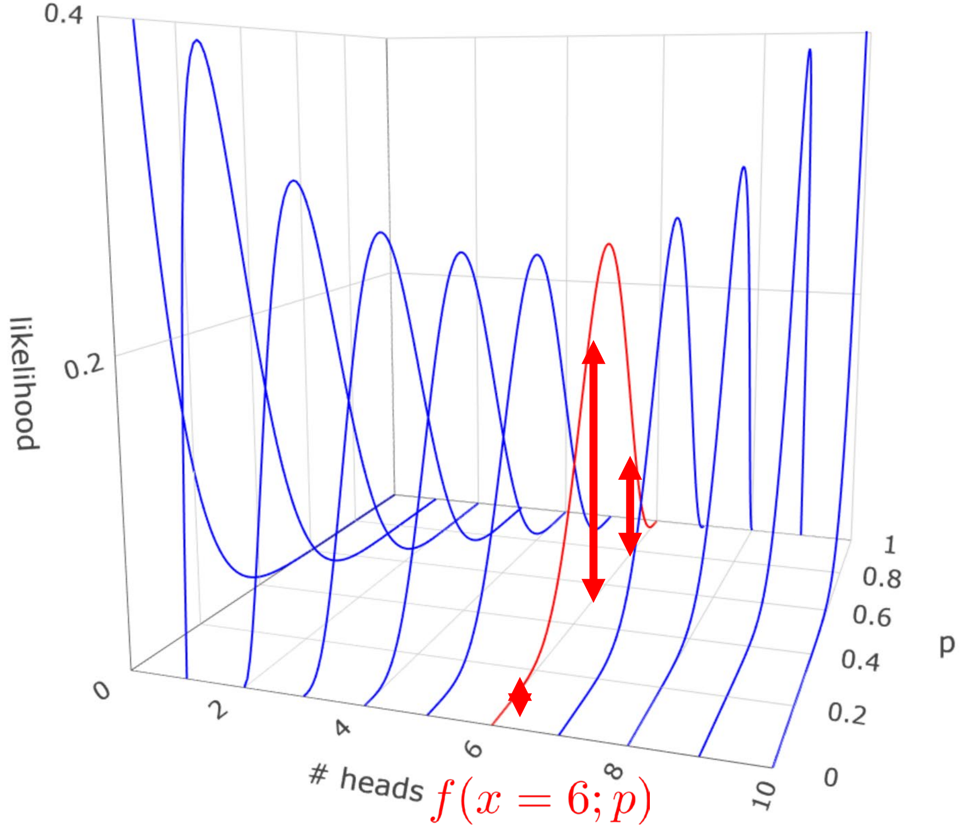
Estimation framing: $X \sim Binom(n = 10, p = ?)$. If $X = 6$, what is p?



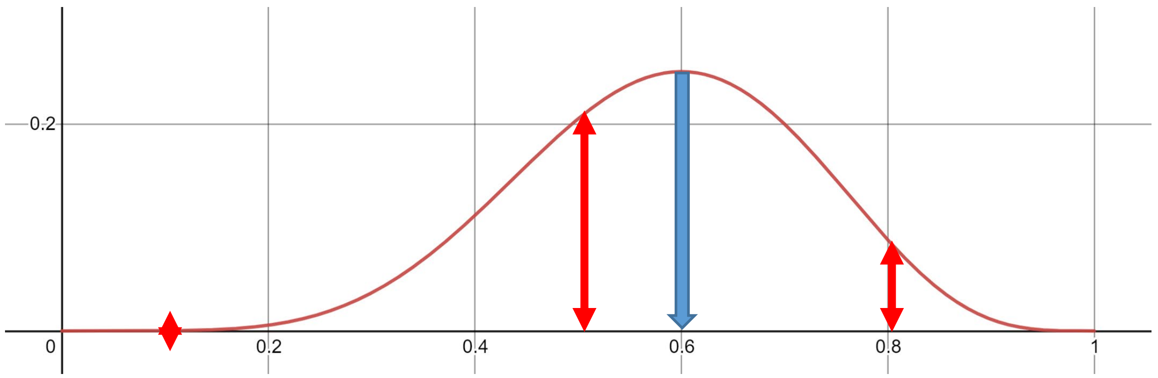
Perhaps there is a way to combine all these images into a single one...



What if we consider values for p other than 0.1, 0.5, and 0.8?



$$\text{Recall } f(x;p) = \binom{10}{x} p^x (1-p)^{10-x} \text{ is the pmf for } X \sim Binom(10, p)$$



$f(x = 6, p)$

A **likelihood function** is just a pmf or pdf where x is viewed as given and the parameter(s) is viewed as the unknown.

Maximum Likelihood Estimation (MLE) finds the best guess for a parameter by finding what parameter value is linked to the maximum of the likelihood function.

$$L(p) = f(x = 6, p) = \binom{10}{6} p^6 (1-p)^4$$

$$L'(p) = \binom{10}{6} \left[p^6 \cdot 4(1-p)^3(-1) + (1-p)^4 \cdot 6p^5\right]$$

$$0 = -4p^6(1-p)^3 + 6p^5(1-p)^4$$

$$0 = -4p + 6(1-p)$$

$$10p = 6$$

$$\boxed{\widehat{p}_{\text{MLE}} = 0.6}$$

This confirms our intuition!

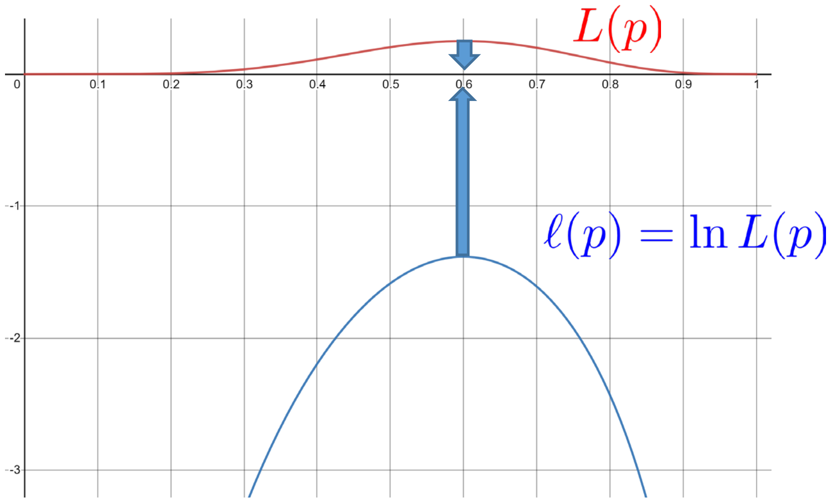
The MLE Procedure (Improved!)

1) Using the pmf/pdf, write the likelihood function. Notation: $L(p)$ or $L(\text{parameter})$

2) Take the ln of $L(p)$ to create the **log-likelihood** function. Notation: $\ell(p)$

Helpful fact: $g(x)$ and $\ln g(x)$ attain a max at the same x location.

Since $\ln(a \cdot b) = \ln a + \ln b$ and $\ln\left(\frac{a}{b}\right) = \ln a - \ln b$, the expression for $\ell(p)$ is often MUCH easier to differentiate than $L(p)$.



3) Take the derivative and set it equal to 0.

4) Solve to find the MLE.

5) Plug the MLE into the second derivative to prove you have a max (not min).

Pro Tips (PTs) show how to make the MLE steps more efficient and reliable. Use them!

Your Turn: Improved You!

Follow the improved MLE steps for $L(p) = \binom{10}{6} p^6 (1-p)^4$.

PT 1: Skip this step

$$\ell(p) = \ln L(p) = \ln\left[\binom{10}{6} \cdot p^6 \cdot (1-p)^4\right] = \ln\binom{10}{6} + \ln p^6 + \ln(1-p)^4$$

So, $\ell(p) = C + 6\ln p + 4\ln(1-p)$

PT 2: Anything without a p can be hidden behind a C.

$\ell'(p) = 0 + \frac{6}{p} - \frac{4}{1-p}$

$\frac{4}{1-\widehat{p}} = \frac{6}{\widehat{p}}$

PT 4: Put hats on the parameter once you set $\ell' = 0$.

So, $4\widehat{p} = 6(1-\widehat{p})$, or $\widehat{p}_{\text{MLE}} = 0.6$.

$\ell''(p) = -\frac{6}{p^2} - \frac{4}{(1-p)^2}$

Since $\ell''(p) < 0$ for all p, we must have $\ell''(\widehat{p}_{\text{MLE}}) < 0$, and hence, a maximum by the Second Derivative Test.

You MUST check this to earn full credit.

PT 3: Divide the page in half to remember to find $\ell'(p)$ AND $\ell''(p)$.

Beyond One Piece of Data

Usually, we have iid data $(x_1, \ldots, x_n)$ from a RV $X \sim f_X(x;\theta)$.

Here, $L(\theta) = f_{(X_1,\ldots,X_n)}(x_1 \text{ and } x_2 \text{ and } \cdots \text{ and } x_n;\theta)$

$$= f_{X_1}(x_1;\theta) \cdot f_{X_2}(x_2;\theta) \cdots f_{X_n}(x_n;\theta)$$ (by independence)

So, $\boxed{L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)}$. (by identically distributed)

Thus, $\ell(\theta) = \ln \prod_{i=1}^{n} f(x_i;\theta)$. Therefore, $\boxed{\ell(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)}$.

Differentiating sums is MUCH easier than products!

Your Turn: Likely Story

Write the likelihood and simplified log-likelihood function for each setup.

1. $X \sim Poisson(\lambda)$ and you have data $x_1, \ldots, x_n$.

$$L(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \boxed{\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}}$$

$$\ell(\lambda) = \sum_{i=1}^{n} \ln f(x_i; \lambda) = \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) = \sum_{i=1}^{n} \left[\ln e^{-\lambda} + \ln \lambda^{x_i} - \ln(x_i!)\right]$$

$$= \boxed{\sum_{i=1}^{n} [-\lambda + x_i \ln \lambda - \ln(x_i!)]}$$

This is too many steps to show when finding $\ell(\lambda)$ (see next slide)!

2. $X \sim Geom(p)$ and you find $x_1 = 3$ and $x_2 = 2$.

Recall that $f_X(x;p) = (1-p)^{x-1}p$.

So $L(p) = \prod_{i=1}^{2}(1-p)^{x_i-1}p = (1-p)^{3-1}p \cdot (1-p)^{2-1}p = \boxed{(1-p)^3p^2}$.

$\ell(p) = \ln(1-p)^3 + \ln p^2 = \boxed{3\ln(1-p) + 2\ln p}$

PT 5: Train yourself to get from L to simplified ℓ in one step!

You must think carefully about the efficiency of everything you do. If you build too much inefficiency into what you do, you will not be able to move forward to more advanced settings (in math, music performance, sports, video games, etc.)

Additional Practice!

Practice these before beginning the homework. More Pro Tips await!

Growing Exponentially More Complicated

Find the MLE for a random sample of size n for the exponential distribution.

PT 6: If the data aren't named, create them: $x_1, x_2, \ldots, x_n$.

PT 7: Having all the pmfs/pdfs memorized is VERY helpful.

We know that $f_X(x;\lambda) = \lambda e^{-\lambda x}$ where $x > 0, \lambda > 0$

Thus, $L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$

PT 8: Make sure to write L(actual parameter in problem).

PT 9: Put a product in front; change all xs to x_is.

$$\text{So } \ell(\lambda) = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n} (\ln\lambda - \lambda x_i) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i$$

PT 10: Any term without an i appears n times, any term with an i will need to keep the summation.

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i$$

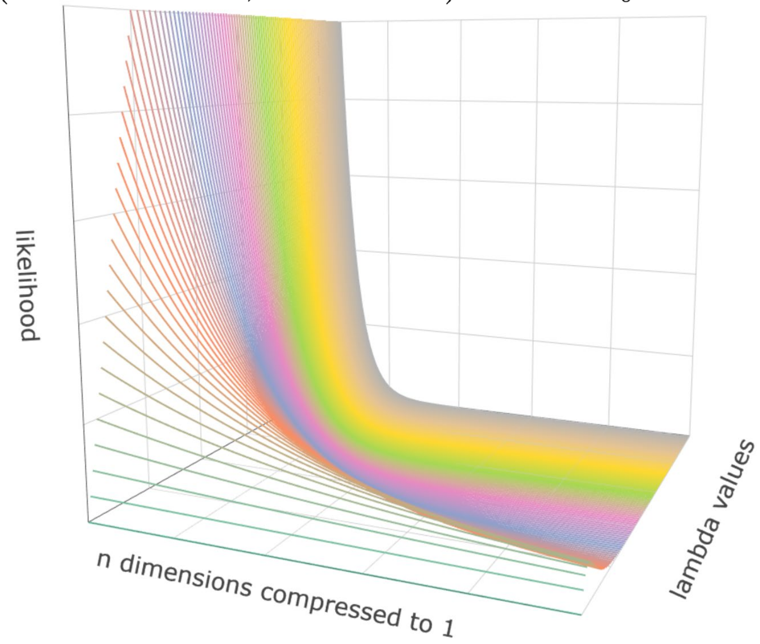
$$\ell''(\lambda) = -\frac{n}{\lambda^2} < 0 \text{ for all } \lambda > 0, \text{ so } \ell''(\widehat{\lambda}_{\text{MLE}}) < 0.$$

PT 11: Feel free to leave the bounds off the $\sum$ after the initial steps.

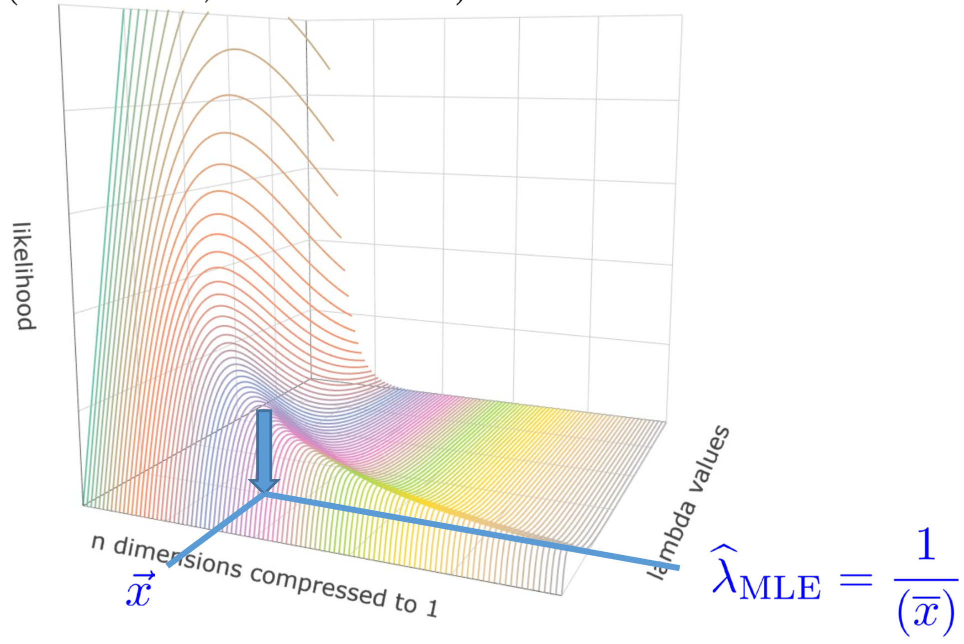
PT 12: Strive for a simple answer that is intuitive.

$$0 = \frac{n}{\widehat{\lambda}} - \sum x_i \implies \sum x_i = \frac{n}{\widehat{\lambda}} \implies \boxed{\widehat{\lambda}_{\text{MLE}} = \frac{n}{\sum x_i} = \frac{1}{\frac{\sum x_i}{n}} = \frac{1}{(\overline{x})} = \frac{1}{(\overline{X})}}$$

$f_{\vec{X}}(\mathbf{x} \text{ unknown}, \lambda \text{ known})$: density outlook



$f_{\vec{X}}(\mathbf{x} \text{ known}, \lambda \text{ unknown})$: likelihood outlook



Your Turn: Don't Be Negative!

Let $X_1, X_2, \ldots, X_n$ be iid based on $X \sim NegBinom(r, p)$ where r is known (fixed). Find the MLE for p. (Follow the PTs!)

Recall that $f_X(x; r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$

$$L(p) = \prod_{i=1}^{n} \binom{x_i - 1}{r - 1} p^r (1-p)^{x_i - r}$$

Remember that this is really $L(p|x_1, \ldots, x_n, r)$. Leaving off the $x_1, \ldots, x_n, r$ focuses you on p.

$$\ell(p) = \sum_{i=1}^{n} \left[\ln \binom{x_i - 1}{r - 1} + r \ln p + (x_i - r) \ln(1-p) \right]$$

Internalize the steps so you can do several at once!

$$= C + nr \ln p + \ln(1-p) \sum (x_i - r)$$

$$\ell'(p) = \frac{nr}{p} + \frac{\sum(x_i - r)}{1-p}(-1) \quad \Big| \quad \ell''(p) = -\frac{nr}{p^2} - \frac{\sum(x_i - r)}{(1-p)^2} < 0 \text{ for all } p$$

$$0 = \frac{nr}{\widehat{p}} - \frac{\sum(x_i - r)}{1-\widehat{p}}$$

Multiply both sides by $\widehat{p}(1-\widehat{p})$

$$0 = nr(1-\widehat{p}) - \widehat{p}\sum(x_i - r) = nr - \cancel{nr\widehat{p}} - \widehat{p}\sum(x_i) + \cancel{\widehat{p}nr}$$

Note: $\sum_{i=1}^{n} r = nr$

So $\widehat{p}\sum x_i = nr$, or $\widehat{p} = \frac{nr}{\sum x_i} = \frac{r}{\frac{\sum x_i}{n}} = \frac{r}{(\overline{x})}$.

Note: Different notations for the MLE:

$\widehat{\theta} = \theta_e = \theta_{\text{MLE}} = \widehat{\theta}_{\text{MLE}}$

In the last two problems, you can show that $\widehat{\theta}_{\text{MME}} = \widehat{\theta}_{\text{MLE}}$. This often occurs, particularly for the common distributions.

History: Gauss, Laplace, Daniel Bernoulli (the idea); Ronald Fisher, 1912 (formalization, theory development)