Maximum Likelihood Estimation

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SEPTEMBER 29, 2016

Roadmap

- Going from data and distributions to parameters
- Mathematical aside: optimization with constraints
- Poisson MLE
- Gaussian MLE
- Multinomial MLE

Why MLE?

- Before: Distribution + Parameter $\longrightarrow x$
- Now: x + Distribution $\longrightarrow$ Parameter
- (Much more realistic)
- NB: Says nothing about how good a fit a distribution is

Likelihood

- Likelihood is $p(x; \theta)$
- We want estimate of θ that best explains data we seen
- I.e., Maximum Likelihood Estimate (MLE)

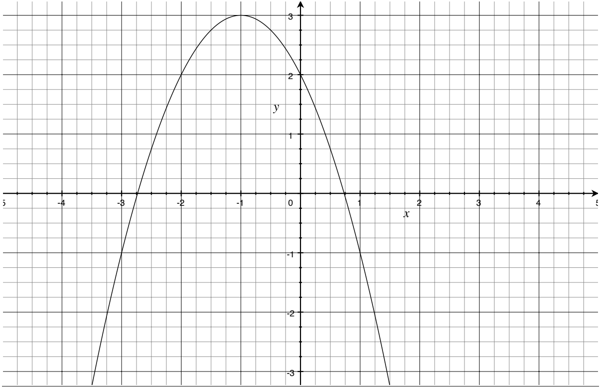
Likelihood

- The *likelihood function* refers to the PMF (discrete) or PDF (continuous).
- For discrete distributions, the likelihood of x is $P(X = x)$.
- For continuous distributions, the likelihood of x is the density $f(x)$.
- We will often refer to likelihood rather than probability/mass/density so that the term applies to either scenario.

Optimizing Unconstrained Functions

Suppose we wanted to optimize

$$\ell = x^2 - 2x + 2 \qquad (1)$$

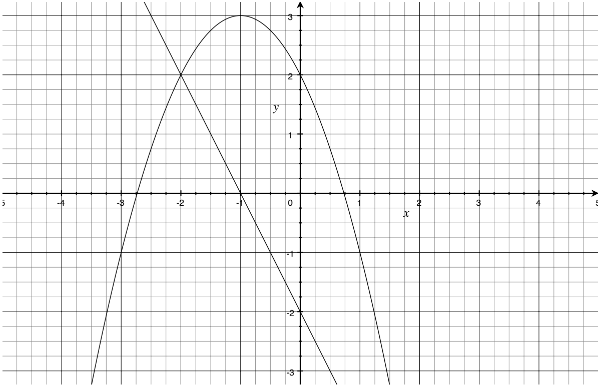


Optimizing Unconstrained Functions

Suppose we wanted to optimize

$$\ell = x^2 - 2x + 2 \qquad (1)$$

$$\frac{\partial \ell}{\partial x} = -2x - 2 \qquad (2)$$



Optimizing Unconstrained Functions

$$\frac{\partial \ell}{\partial x} = 0 \tag{3}$$

$$-2x - 2 = 0 \tag{4}$$

$$x = -1 \tag{5}$$

(Should also check that second derivative is negative)

Optimizing Constrained Functions

Theorem: Lagrange Multiplier Method

Given functions $f(x_1, \ldots x_n)$ and $g(x_1, \ldots x_n)$, the critical points of f restricted to the set $g = 0$ are solutions to equations:

$$\frac{\partial f}{\partial x_i}(x_1, \ldots x_n) = \lambda \frac{\partial g}{\partial x_i}(x_1, \ldots x_n) \quad \forall i$$

$$g(x_1, \ldots x_n) = 0$$

This is $n+1$ equations in the $n+1$ variables $x_1, \ldots x_n, \lambda$.

Lagrange Example

Maximize $\ell(x, y) = \sqrt{xy}$ subject to the constraint $20x + 10y = 200$.

- Compute derivatives

Lagrange Example

Maximize $\ell(x,y) = \sqrt{xy}$ subject to the constraint $20x + 10y = 200$.

- Compute derivatives

$$\frac{\partial \ell}{\partial x} = \frac{1}{2}\sqrt{\frac{y}{x}} \quad \frac{\partial g}{\partial x} = 20$$
$$\frac{\partial \ell}{\partial y} = \frac{1}{2}\sqrt{\frac{x}{y}} \quad \frac{\partial g}{\partial y} = 10$$

Lagrange Example

Maximize $\ell(x,y) = \sqrt{xy}$ subject to the constraint $20x + 10y = 200$.

- Compute derivatives

$$\frac{\partial \ell}{\partial x} = \frac{1}{2}\sqrt{\frac{y}{x}} \quad \frac{\partial g}{\partial x} = 20$$
$$\frac{\partial \ell}{\partial y} = \frac{1}{2}\sqrt{\frac{x}{y}} \quad \frac{\partial g}{\partial y} = 10$$

- Create new systems of equations

Lagrange Example

Maximize $\ell(x,y)=\sqrt{xy}$ subject to the constraint $20x+10y=200$.

- Compute derivatives

$$\frac{\partial \ell}{\partial x}=\frac{1}{2}\sqrt{\frac{y}{x}} \quad \frac{\partial g}{\partial x}=20$$

$$\frac{\partial \ell}{\partial y}=\frac{1}{2}\sqrt{\frac{x}{y}} \quad \frac{\partial g}{\partial y}=10$$

- Create new systems of equations

$$\frac{1}{2}\sqrt{\frac{y}{x}}=20\lambda$$

$$\frac{1}{2}\sqrt{\frac{x}{y}}=10\lambda$$

$$20x+10y=200$$

Lagrange Example

- Dividing the first equation by the second gives us

$$\frac{y}{x} = 2 \tag{6}$$

- which means $y = 2x$, plugging this into the constraint equation gives:

$$\begin{aligned} 20x + 10(2x) &= 200 \\ x = 5 \Rightarrow y &= 10 \end{aligned}$$