Notation comparison

$\mathbb{P}(X|Y)$ probability of X, conditioned on the event Y having happened (Y is a subset of the sample space).

$\mathbb{P}(X;\theta)$ probability of X, where to properly define our probability space we need to know the extra piece of information θ. Since θ isn't an event, this is not conditioning.

$\mathcal{L}(X;\theta)$ the likelihood of event X, given that an experiment was run with parameter θ. Likelihoods don't have all the properties we associate with probabilities (e.g. they don't all sum up to 1) and this isn't conditioning on an event (θ is a parameter/rule of how the event could be generated).

Maximizing a Function

CLOSED INTERVALS

Set derivative equal to 0 and solve.

Evaluate likelihood at endpoints and any critical points.

Maximum value must be maximum on that interval.

SECOND DERIVATIVE TEST

Set derivative equal to 0 and solve.

Take the second derivative. If negative everywhere, then the critical point is the maximizer.

Continuous Example

Suppose you get values $x_1, x_2, \dots x_n$ from independent draws of a normal random variable $\mathcal{N}(\mu, 1)$ (for μ unkown)

We'll also call these "realizations" of the random variable.

$$\mathcal{L}(x_i; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

$$\ln(\mathcal{L}(x_i; \mu)) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(x_i - \mu)^2$$

Summary

Given: an event E (usually n i.i.d. samples from a distribution with unknown parameter θ).

1. Find likelihood $\mathcal{L}(E; \theta)$

Usually $\prod \mathbb{P}(x_i; \theta)$ for discrete and $\prod f(x_i; \theta)$ for continuous

2. Maximize the likelihood. Usually:

A. Take the log (if it will make the math easier)

B. Set the derivative to 0 and solve

C. Use the second derivative test to confirm you have a maximizer