Baysian Inference

Firstly, (statistical) **inference**is the process of deducing properties about a population or probability distribution from data. Maximum Likelihood is a good example.

**Bayesian inference**is therefore just the process of deducing properties about a population or probability distribution from data *using Bayes’ theorem.* That’s it.

Problem with maximum likelihood: It is a fixed value: for example 0.3. But the probability could be 0.25 or 0.4

Therefore we use a probability distribution because this might be more appropriated!

This distribution is known as the **prior distribution.**

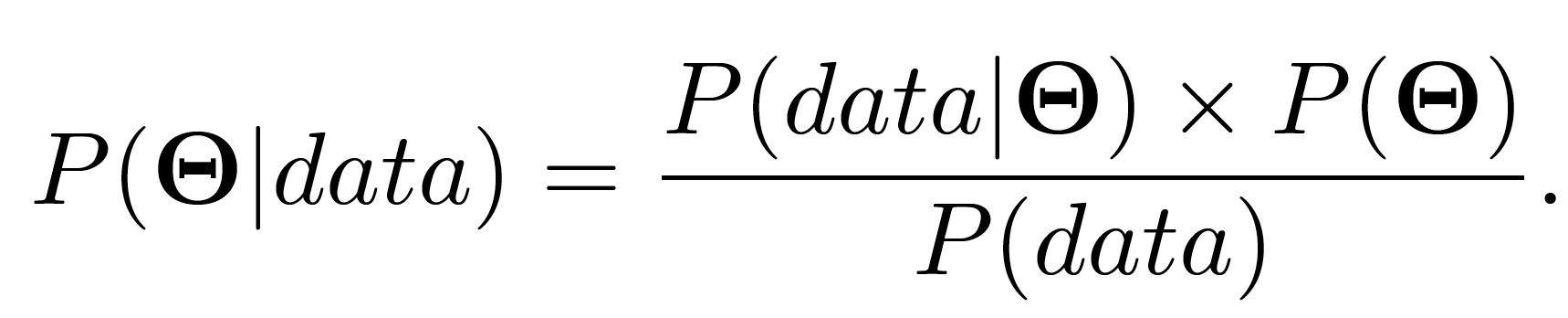
🡪 Two plots: both peak at 0.3

🡪 The higher the peak the more certain

Instead of event A, we’ll typically see Θ, this symbol is called Theta. Theta is what we’re interested in, it represents the set of parameters. So if we’re trying to estimate the parameter values of a Gaussian distribution then Θ represents both the mean, μ and the standard deviation, σ (written mathematically as Θ = {μ, σ}).

Instead of event B, we’ll see *data*or ***y =****{y1, y2, …, yn}.*These represent the data, i.e. the set of observations that we have. I’ll explicitly use *data*in the equation to hopefully make the equation a little less cryptic.

So now Bayes’ theorem in model form is written as:



We‘ve seen that P(Θ) is the prior distribution. It represents our beliefs about the true value of the parameters, just like we had distributions representing our belief about the probability of selling ice cream.

*P(*Θ*|data)*on the left hand side is known as the **posterior distribution.**This is the distribution representing our belief about the parameter values after we have calculated everything on the right hand side taking the observed data into account.

*P(data|*Θ*)*is something we’ve come across before. If you made it to the end of my previous post on maximum likelihood then you’ll remember that we said *L(data; μ, σ)*is the likelihood distribution (for a Gaussian distribution). Well *P(data|*Θ*)*is exactly this, it’s the **likelihood distribution** in disguise. Sometimes it’s written as ℒ*(*Θ; *data)* but it’s the same thing here**.**

Grundsätzliche Schreibweise:



Da P(X) keine Datenabhängigkeit beinhaltet, ist die Form der Posterior-Verteilung nur vom Zähler von Bayes abhängig. 🡪 Likelihood & Prior

Posteroir Peak between Likelihood Peak and Prior Peak.

Amount of Data (x wird größer) wird größer, dann wird Likelihood kleiner P(X|Teta)