Priors and Likelihood combine to give Posterior

Giles Innocent with input from Sonja Hartnack

COST-Harmony Madrid July 2023

Bayes rule

$$P(A \wedge B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

$$\implies P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

Or, in our case:

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

Or, typically, as $P(Y)$ is constant:

$$P(\theta|Y) \propto P(\theta) \times P(Y|\theta)$$

Distributions as priors

- Typically we are interested in prior *distributions*
- Rather than multiplying individual values we must do this across all possible values of θ the parameter vector
- Typically the probability of the data is difficult to calculate
- We could integrate over the whole of parameter space, but quickly becomes difficult
- As we want to find the Posterior *distribution* MCMC provides a relatively simple solution to finding the posterior

MCMC

- We wish to sample from the posterior
- This sample, if sufficiently large, allows us to approximate the posterior probability distribution
- Hence estimates of posterior mean, median, mode, variance etc.
- Typically this requires a "guess" for the next set of parameters, followed by an accept/reject decision, based on the posterior probability at this set of parameters

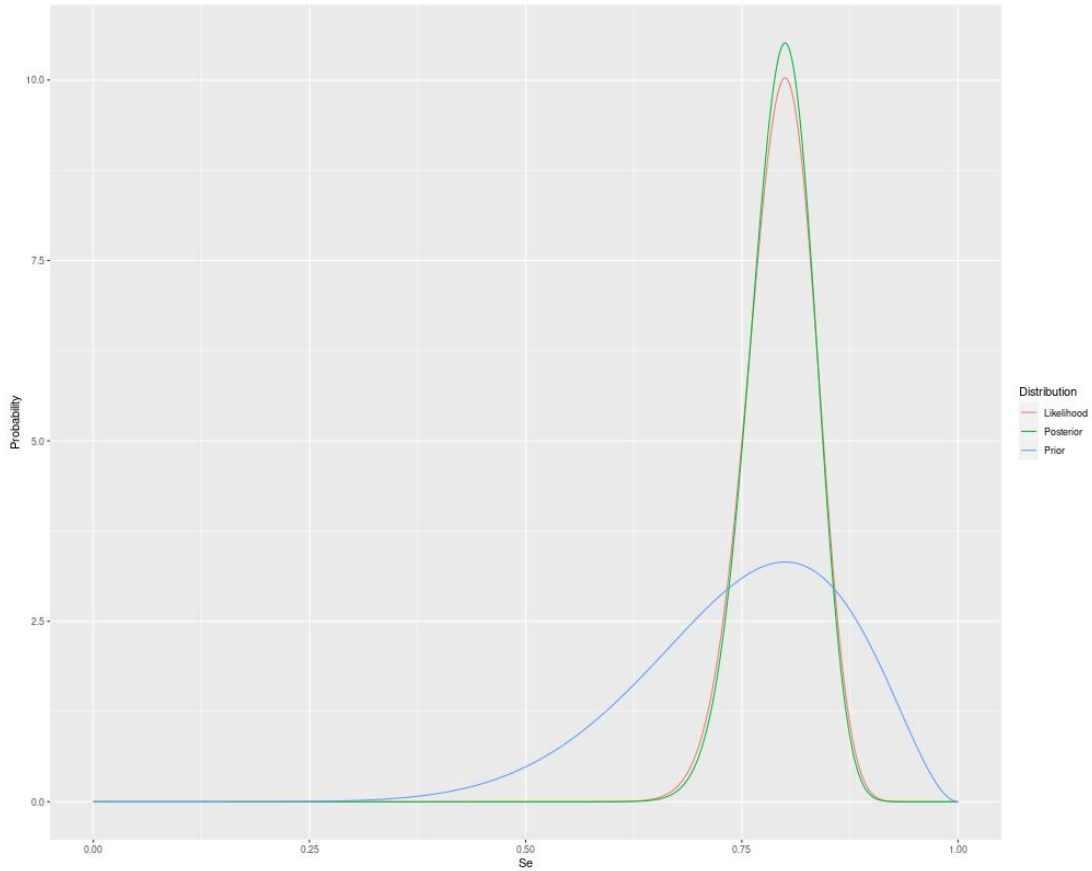
Do priors matter?

- Posterior combines the prior with the (likelihood of) the data
- If there is sufficient data it overwhelms the prior
- The "ghost" of the prior remains
- If the prior only poorly supports the data they can be an issue
- If the prior does not cover the true range of the parameter(s)
  - E.G. you have a strictly positive prior and negative values are supported for the parameter
  - you are "sure" that the Se > 0.5, so you use a prior where the $p(Se < 0.5)$ is vanishingly small, but the data suggest that it may not be.
- No amount of data can fix an incorrect prior

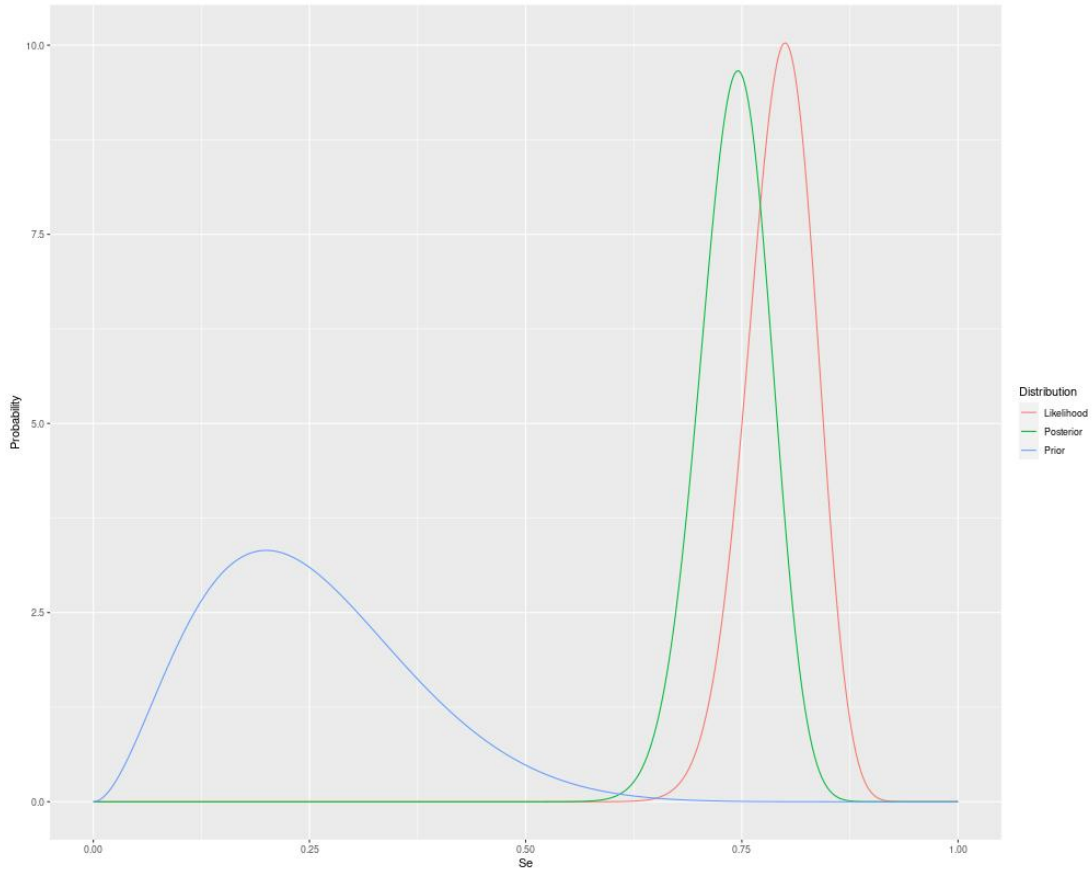
How can we test priors?

- Plot posterior
  - does it all stack up at one end?
  - does this coincide with a limit in the prior?
- Plot prior and posterior together
  - good
  - OK
  - poor
  - see next slides for details

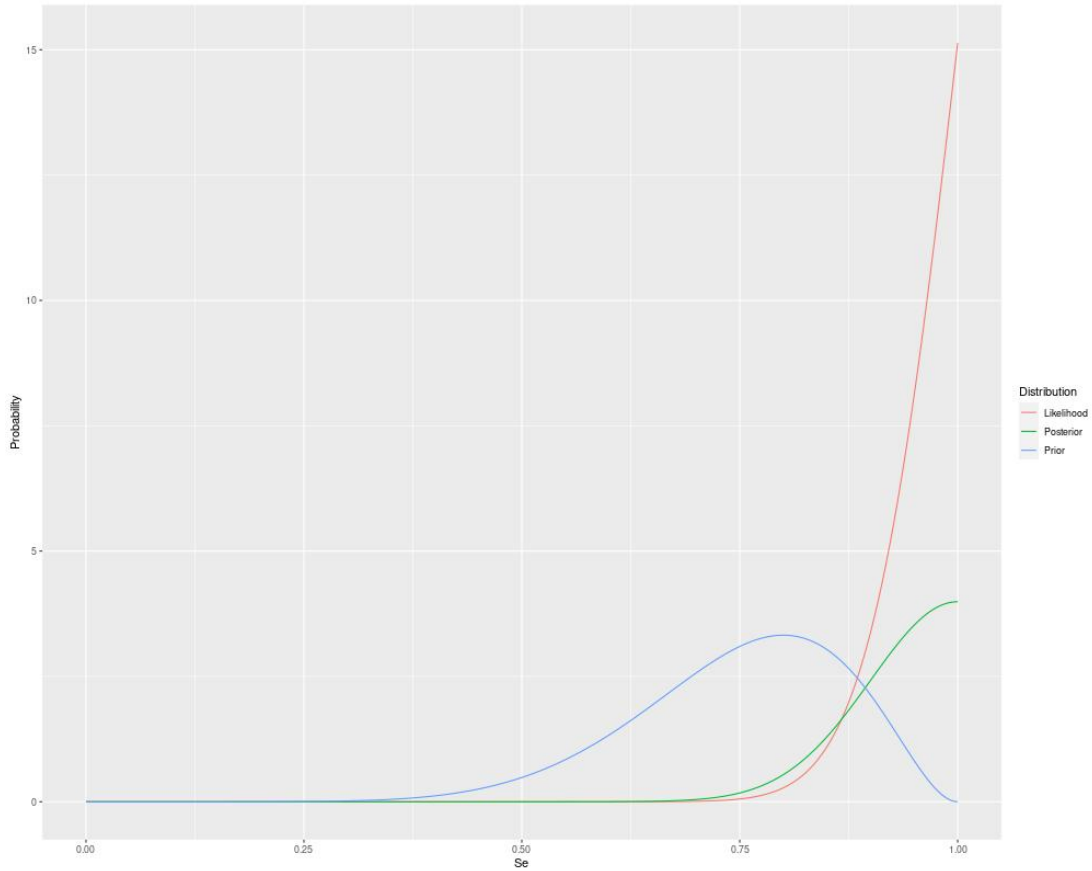
Prior and posterior agree



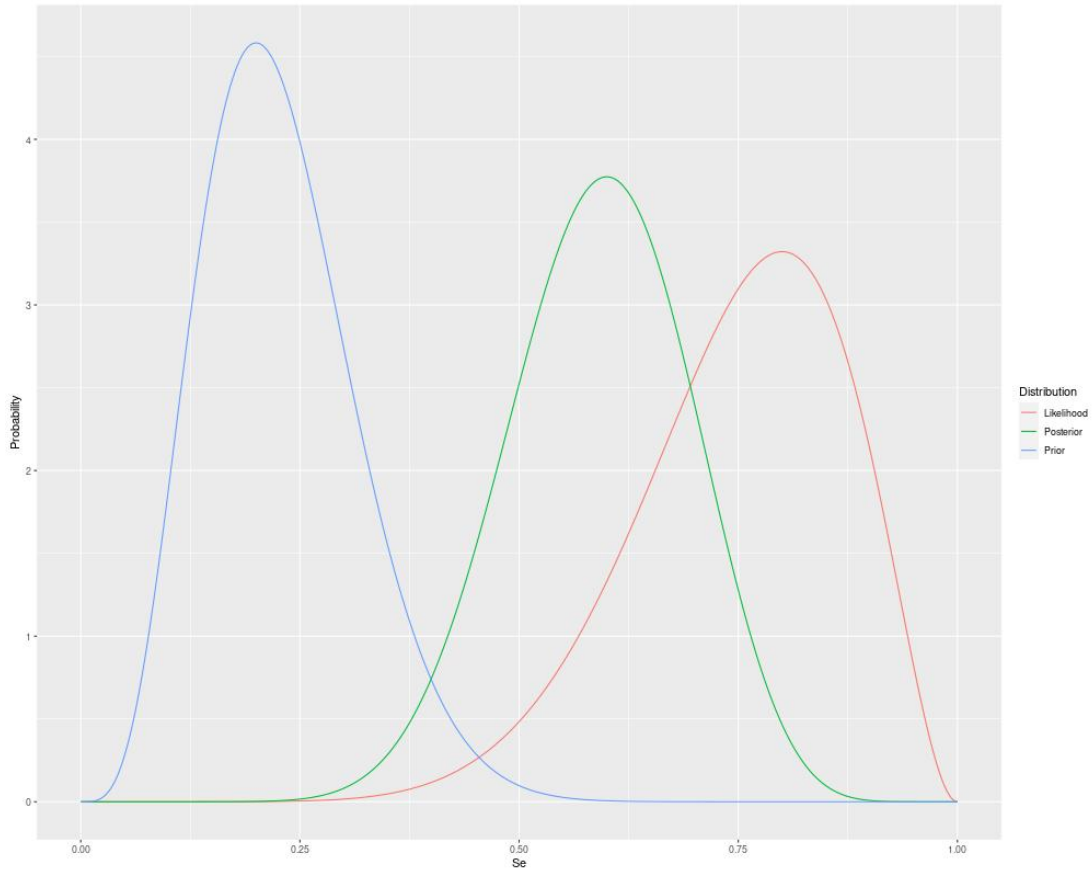
Prior and posterior disagree - prior swamped by data



Poor prior - posterior truncated



Poor prior - posterior distorted



Giles' hierarchy of priors

- Previous data (posterior distribution)
- Prior information; possibly vague
- True expert opinion
- Relatively uninformative
- "Expert" opinion

Frequentists don't use priors do they?

- Frequentists often cite use of priors as an issue with Bayesian methods
  - complain about lack of significance testing
  - complain about lack of power calculations/design
- Frequentists use prior information/knowledge/guess to calculate power/sample size which is only relevant at a specific p-value
- Use of prior information is then "forgotten" during the analysis