Introduction in nutshell

Bayesian interpretation:

- Probability = degree of uncertainty (not frequency).
- Uncertainty: can be thought as epistemic uncertainty and aleatory uncertainty.
- Epistemic uncertainty: e.g. not knowing the exact disease percentage in a population. Could be reduced in principle.
- Aleatory uncertainty: e.g. not knowing the outcome of a sample. Could not be ***easily*** reduced (Jaynes: in principle, this too is a matter of not knowing exact initial conditions. Therefore, also a form of uncertainty. No ”true randomness” need to be assumed).
- ***”The essence of the present theory is that <u>no</u> probability, direct, prior, or posterior, is simply a frequency.”***

–Harold Jeffreys 1939.

Introduction in nutshell

Procedure of bayesian inference:

- Specify the unknown parameter of interest θ.
- Specify your prior distribution $P(\theta)$.
- Specify the conditional probability $P(X|\theta)$ of observations X, given θ.
- Update the prior by the information of data X, i.e. compute posterior probability distribution $P(\theta \mid X)$.

Introduction in nutshell

Procedure of bayesian inference:

- If the prior distribution was conjugate, the posterior distribution will take the form of a standard distribution. Determine the parameters for this.
- If the prior was not conjugate, we need numerical methods (e.g. WinBUGS) to compute something from the posterior distribution.
- Make a summary of the posterior results.

Introduction in nutshell

Different approaches for prior distribution:

- **Uninformative priors:** aim to be standard for representing ’no prior knowledge’.
- Not exactly possible to define universally uninformative prior with respect to all conceivable aspects.
- Usually good (and easy) choice if there will be enough data that will dominate the posterior distribution, even improper flat priors may work.

Introduction in nutshell

Different approaches for prior distribution:

- **Informative priors:** aim to represent prior knowledge that is thought relevant.
- Could be based on earlier similar experiments (as ’prior data’), earlier published estimates, or expert knowledge elicitation, or logical judgements of substance.
- Can be useful when prior information is thought important and justified, and if the data will contain only limited information.

Introduction in nutshell

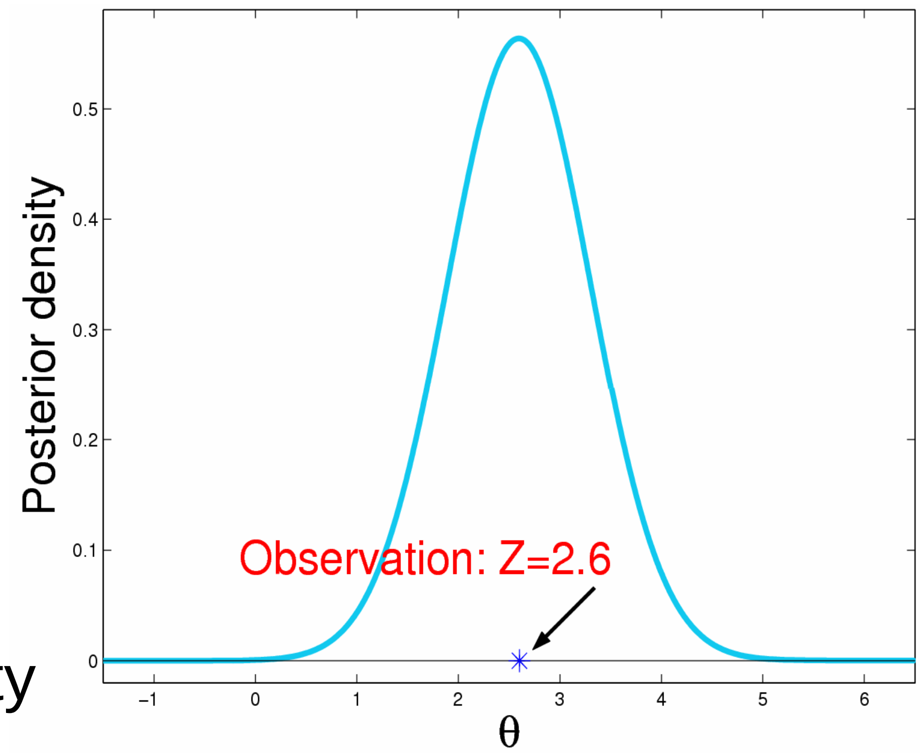
Example of unknown mean θ:

- Assumed model for observable variables X and Y, given θ:

  X ~ N(θ,1) and Y ~ N(θ,1)

- The model for Z = (X+Y)/2 is then: Z ~ N(θ,0.5)
- Assume that we can observe Z.
- Assume a prior for θ. Let's say uninformative, flat, prior over entire real axis: P(θ)=1, for all θ.
- Compute the posterior probability distribution **for** θ, **given Z**:

P(θ | Z)=P(Z | θ)*P(θ)/const =N(Z,0.5).



Introduction in nutshell

Some classical examples of conjugate priors:

- $p \sim \text{Beta}(\alpha,\beta)$ with $x \sim \text{Bin}(n,p)$
  → $p \sim \text{Beta}(x+\alpha, n-x+\beta)$.

- $\lambda \sim \text{Gamma}(\alpha,\beta)$ with $x_i \sim \text{Poisson}(\lambda)$ when $i=1,\ldots,n$
  → $\lambda \sim \text{Gamma}(\Sigma x_i+\alpha, n+\beta)$

- $\mu \sim N(\mu_0,\sigma_0^2)$ with $x_i \sim N(\mu,\sigma^2)$ when $i=1,\ldots,n$
  → $\mu \sim N(\ (n_0\mu_0+ \Sigma x_i/n)/(n_0+1)\ ,\ (\sigma^2/n)/(n_0+1)\ )$
  where $n_0 = \sigma^2/(n\sigma_0^2)$

Introduction in nutshell

Conjugate priors:

- Advantages of conjugacy:
  - – posterior is a known standard distribution
  - – posterior mean can be seen as a weighted average of data and ’prior data’.
- Disadvantages of conjugacy:
  - – very restricted use, only some solutions exist, but may serve as a first approach and ’benchmark’.

Introduction in nutshell

Exchangeability:

- If a series of random (i.e. uncertain) variables $X_1, X_2, X_3, \ldots$ is such that for us, they appear to be infinitely exchangeable, then it logically follows that our probability can be expressed as if
  - $P(X_i \mid \theta)$ (each of them are conditionally independent of others, given θ)
  - $P(\theta)$ represents a prior for θ.

Introduction in nutshell

Exchangeability:

- Then, we obtain
  - prior predictive distribution, integrating over prior uncertainty:

    $$P(X_i) = \int P(X_i \mid \theta)P(\theta)\, d\theta$$

  - posterior predictive distribution, integrating over posterior uncertainty:

    $$P(x_{n+1} \mid x_1,\ldots,x_n) = \int P(x_{n+1} \mid \theta)P(\theta \mid x_1,\ldots,x_n)\, d\theta$$