# Bayes' Theorem

Let $A$ and $B_1, \ldots, B_k$ be events in a sample space $\Omega$.

*Inversion problem:* given $\mathbb{P}(A|B_j)$ (and $\mathbb{P}(B_j)$) find $\mathbb{P}(B_j|A)$

**Bayes' Theorem:**

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\sum_{i=1}^{k} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}$$

For continuous random variables $X$ and $Y$, Bayes' Theorem is formulated in terms of densities:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\,f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)\,f_X(x)}{\int f_{Y|X}(y|x)\,f_X(x)\,dx}$$

*Application to statistical inference:*

- Probabilistic model: $f(y|\theta)$ - distribution of $Y$ for fixed $\theta$
- Statistical problem: given data $y$ make statements about $\theta$
- Likelihood: $l(\theta|y) = f(y|\theta)$ (reflects inversion problem)

**Bayesian approach:**

A Bayesian statistical (parametric) model consists of

- $f(y|\theta)$, a parametric statistical model (likelihood function), and
- $\pi(\theta)$, a prior distribution on the parameters.

The posterior distribution of the parameter $\theta$ is

$$\pi(\theta|y) = \frac{f_{Y|\theta}(y|\theta)\,\pi(\theta)}{\int_\Theta f_{Y|\theta}(y|\theta)\,\pi(\theta)\,d\theta} \sim f_{Y|\theta}(y|\theta)\,\pi(\theta)$$

The Bayesian modelling approach can be summarized by

posterior $\sim$ likelihood $\times$ prior.

*Bayesian interpretation of probability*

probability = (subjective) uncertainty

# Bayesian Inference

**Example:** Binomial distribution

- Likelihood function

  $$Y|\theta \sim \mathrm{Bin}(n, \theta)$$

- Prior distribution

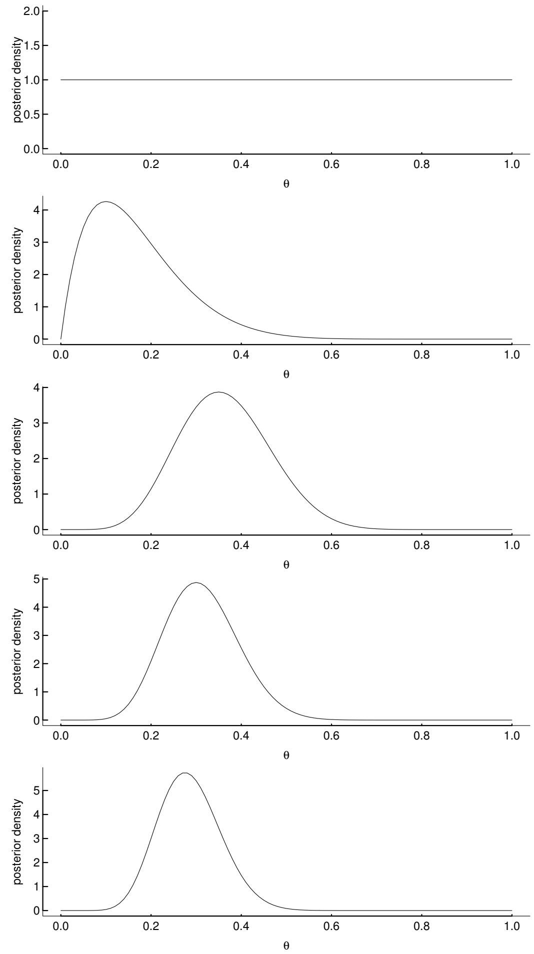
  $$\theta \sim U(0, 1) = \mathrm{Beta}(1, 1)$$

- Posterior distribution

  $$\theta|Y \sim \mathrm{Beta}(1 + Y, 1 + n - Y)$$

Uncertainty about parameter can be updated repeatedly when new data are available:

- take current posterior distribution as prior
- compute new posterior distribution conditional on new data




The *posterior distribution* is used for inference about $\theta$:

- posterior mean

  $$\mathbb{E}(\theta|Y)$$

- posterior variance

  $$\mathrm{var}(\theta|Y) = \mathbb{E}\big((\theta - \mathbb{E}(\theta|Y))^2 \big| Y\big)$$

- posterior confidence interval (credibility interval)

  $$\int_{\theta_l}^{\theta_r} \pi(\theta|Y)\, d\theta = 1 - \alpha$$

# Conjugate Priors

A mathematical convenient choice are conjugate priors: The posterior distribution belongs to the same parametric family as the prior distribution with different parameters:

| **Likelihood** $\boldsymbol{f(y\|\theta)}$ | **Prior** $\boldsymbol{\pi(\theta)}$ | **Posterior** $\boldsymbol{\pi(\theta\|y)}$ |
|---|---|---|
| Normal $\mathcal{N}(\theta, \sigma^2)$ | Normal $\mathcal{N}(\mu, \tau^2)$ | Normal $\mathcal{N}\left(\frac{\sigma^2\mu+\tau^2 y}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$ |
| Poisson $\text{Poisson}(\theta)$ | Gamma $\Gamma(\alpha, \beta)$ | Gamma $\Gamma(\alpha+y, \beta+1)$ |
| Gamma $\Gamma(\nu, \theta)$ | Gamma $\Gamma(\alpha, \beta)$ | Gamma $\Gamma(\alpha+\nu, \beta+y)$ |
| Binomial $\text{Bin}(n, \theta)$ | Beta $\text{Beta}(\alpha, \beta)$ | Beta $\text{Beta}(\alpha+y, \beta+n-y)$ |
| Multinomial $\text{M}_k(\theta_1, \ldots, \theta_k)$ | Dirichlet $\text{D}(\alpha_1, \ldots, \alpha_k)$ | Dirichlet $\text{D}(\alpha_1+y_1, \ldots, \alpha_k+y_k)$ |
| Normal $\mathcal{N}(\mu, 1/\theta)$ | Gamma $\Gamma(\alpha, \beta)$ | Gamma $\Gamma\left(\alpha+\frac{1}{2}, \beta+\frac{1}{2}(\mu-y)^2\right)$ |

**Problems in choice of prior:**

- The conjugate priors might not reflect our uncertainty about $\theta$ correctly.
- In general, for non-conjugate priors the posterior distribution is not available in analytic form.
- It is difficult to describe uncertainty about $\theta$ in form of a particular distribution. In particular, we might be uncertain about the parameters of the prior distribution ($\rightsquigarrow$ hierarchical modelling, empirical Bayesian methods).

# Bayesian Analysis with Missing Data

**Bayesian statistical model:**

- *Data model:*

$$\begin{array}{ll} f(Y|\theta) & \text{complete-data likelihood} \\ f(R|Y,\xi) & \text{missing-data mechanism} \end{array}$$

- *Prior distribution:*

$$\pi(\theta,\xi)$$

The *posterior distribution* of $\theta$ and $\xi$ is

$$\begin{aligned} \pi(\theta,\xi|Y_{\text{obs}},R) &\sim f(Y_{\text{obs}},R|\theta,\xi)\,\pi(\theta,\xi) \\ &= \int f(Y_{\text{obs}},y_{\text{mis}},R|\theta,\xi)\,\pi(\theta,\xi)\,dy_{\text{mis}} \\ &= \int f(Y_{\text{obs}},y_{\text{mis}}|\theta)\,f(R|Y_{\text{obs}},y_{\text{mis}},\xi)\,\pi(\theta,\xi)\,dy_{\text{mis}} \end{aligned}$$

If the data are missing at random (MAR) then

$$\begin{aligned} \pi(\theta,\xi|Y_{\text{obs}},R) &\sim \int f(Y_{\text{obs}},y_{\text{mis}}|\theta)\,f(R|Y_{\text{obs}},\xi)\,\pi(\theta,\xi)\,dy_{\text{mis}} \\ &= \int f(Y_{\text{obs}},y_{\text{mis}}|\theta)\,dy_{\text{mis}}\,f(R|Y_{\text{obs}},\xi)\,\pi(\theta,\xi) \\ &= f(Y_{\text{obs}}|\theta)\,f(R|Y_{\text{obs}},\xi)\,\pi(\theta,\xi) \end{aligned}$$

# Bayesian Analysis with Missing Data

For inference on $\theta$, we consider the marginal posterior distribution of $\theta$

$$\begin{aligned}\pi(\theta|Y_{\text{obs}}, R) &= \int_{\Xi} \pi(\theta, \xi|Y_{\text{obs}}, R)\, d\xi \\ &\sim \int_{\Xi} f(Y_{\text{obs}}|\theta)\, f(R|Y_{\text{obs}}, \xi)\, \pi(\theta, \xi)\, d\xi\end{aligned}$$

If the parameters are distinct in the sense that

$$\pi(\theta, \xi) = \pi(\theta)\, \pi(\xi)$$

then the marginal posterior distribution of $\theta$ satisfies

$$\pi(\theta|Y_{\text{obs}}, R) \sim f(Y_{\text{obs}}|\theta)\, \pi(\theta) \cdot \int_{\Xi} f(R|Y_{\text{obs}}, \xi)\, \pi(\xi)\, d\xi$$

It follows that

$$\pi(\theta|Y_{\text{obs}}, R) = \frac{f(Y_{\text{obs}}|\theta)\, \pi(\theta)}{\int_{\Theta} f(Y_{\text{obs}}|\theta)\, \pi(\theta)\, d\theta}$$

and hence $\pi(\theta|Y_{\text{obs}}, R) = \pi(\theta|Y_{\text{obs}})$.

**Result:**

The missing data mechanism is ignorable for posterior inference about the parameter $\theta$ if

- the data are missing at random (MAR) and
- the parameters $\theta$ and $\xi$ are distinct, that is

$$\pi(\theta, \xi) = \pi(\theta)\, \pi(\xi).$$