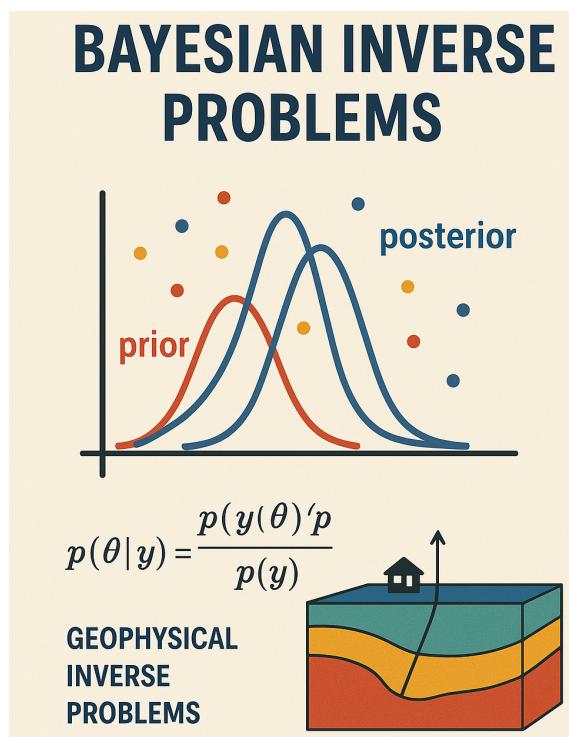


# Bayesian Inverse Problems

## PRACTICAL

---

Mark Asch - MAKUTU/2025



BIP - M. Asch - Practical for Lecture 02

## Ex.1: Bayes - Priors and Posteriors

Suppose that the random variable  $X \sim \mathcal{U}[0, 1]$  and suppose that we know the conditional likelihood

$$Y|X = x \sim \mathcal{G}(x),$$

where the geometric law,  $\mathcal{G}(x)$ , implies that

$$P_{Y|X}(y|x) = x(1-x)^{y-1}$$

for  $y = 1, 2, \dots$

1. What event does the geometric law model?
2. Find the posterior density of  $X$  given  $Y = 2$ .
3. Plot the prior, likelihood and posterior.
4. Conclusions.

# Theory: MAP estimator

[Wikipedia]

- Assume that we want to estimate an unobserved population parameter  $\theta$  from observations  $y$
- Let  $f$  be the distribution of  $y$  so that  $f(y | \theta)$  is the probability of  $y$  when the underlying population parameter is  $\theta$
- Then the function  $\theta \mapsto f(y | \theta)$  is known as the likelihood function and the estimate

$$\hat{\theta}_{\text{MLE}}(y) = \arg \max_{\theta} f(y | \theta)$$

is the **maximum likelihood estimate** (MLE) of  $\theta$ .

- Now assume that a prior distribution  $g(\theta)$  over  $\theta$  exists, which allows us to treat  $\theta$  as a random variable

- According to Bayes' Law, we can calculate the posterior density of  $\theta$

$$\theta \mapsto f(\theta | y) = \frac{f(y | \theta) g(\theta)}{\int_{\Theta} f(y | \vartheta) g(\vartheta) d\vartheta}.$$

- The method of maximum a posteriori (MAP) estimation then estimates  $\theta$  as the mode of the posterior density of this random variable

$$\begin{aligned}\hat{\theta}_{\text{MAP}}(x) &= \arg \max_{\theta} f(\theta | y) \\ &= \arg \max_{\theta} \frac{f(y | \theta) g(\theta)}{\int_{\Theta} f(y | \vartheta) g(\vartheta) d\vartheta} \\ &= \arg \max_{\theta} f(y | \theta) g(\theta).\end{aligned}$$

- Notes:

$\Rightarrow$  The denominator of the posterior density (the marginal likelihood of the model) is always pos-

itive and does not depend on  $\theta$  and therefore plays no role in the optimization.

⇒ The MAP estimate of  $\theta$  coincides with the ML estimate when the prior  $g$  is uniform (i.e.,  $g$  is a constant function).

- MAP estimates can be **computed** in several ways:
  - ⇒ Analytically, when the mode(s) of the posterior density can be given in closed form. This is the case when conjugate priors are used.
  - ⇒ By numerical optimization such as the conjugate gradient method or some quasi-Newton method. This usually requires first or second derivatives, which have to be evaluated analytically or numerically.
  - ⇒ By a modification of an expectation-maximization (EM) algorithm. This does not require derivatives of the posterior density.
  - ⇒ By a Monte Carlo method.

## Ex. 2: MAP estimator

Let  $X$  be a continuous random variable with PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

and suppose that

$$Y | X = x \sim \mathcal{G}(x)$$

is a geometric random variable.

1. Find the MAP estimate of  $X$  given  $Y = 3$ .
2. Plot the prior, likelihood and posterior, as well as the MAP estimate.

## Ex. 3: MAP estimator of mean of IID random variables

Suppose that we are given a sequence  $(x_1, \dots, x_n)$  of IID  $\mathcal{N}(\mu, \sigma_v^2)$  random variables and a prior distribution of the mean  $g(\mu)$  is given by  $\mathcal{N}(\mu_0, \sigma_m^2)$ .

1. Find the MAP estimate of  $\mu$ . (long...)
2. Plot the prior, likelihood and posterior, as well as the MAP estimate.

## Ex. 4: MAP estimator of a noisy signal

Suppose that the signal  $X \sim \mathcal{N}(0, \sigma_X^2)$  is transmitted over a communication channel. Assume that the received signal is given by

$$Y = X + \eta$$

where  $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$  is independent of  $X$ .

1. Find the ML estimate of  $X$  given the observation  $Y = y$ .
2. Find the MAP estimate of  $X$  given the observation  $Y = y$ .
3. Plot the prior, likelihood and posterior, as well as the ML and MAP estimates.

# Theory: CM estimator

- Recall that the posterior distribution,  $f_{X|Y}(x|y)$ , contains all the knowledge that we have about the unknown quantity  $X$
- Therefore, to find a point estimate of  $X$  we can just choose any **summary statistic** of the posterior such as its mean, median, or mode.  
⇒ If we choose the **mode**, the value of  $x$  that maximizes  $f_{X|Y}(x|y)$ , we obtain the MAP estimate of  $X$ .
- Another option would be to choose the **posterior (conditional) mean**,

$$\hat{x} = \mathbb{E}[X|Y = y].$$

- We can show that  $\mathbb{E}[X|Y]$  will give the best estimate of  $X$  in terms of the mean squared error.

For this reason, the conditional expectation is also called the **minimum mean squared error** (MMSE) estimate of  $X$ . It is also called the least mean squares (LMS) estimate or simply the Bayes' estimate of  $X$ .

## Ex. 5: CM (MMSE) estimator of a random variable

Let  $X$  be a continuous random variable with the following prior PDF

$$f_X(x) = 2x, \quad x \in [0, 1].$$

Suppose also that the likelihood is known,

$$f_{Y|X}(y|x) = \begin{cases} 2xy - x + 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the CM estimate of  $X$  given the observation  $Y = y$ .
2. Plot the prior, likelihood and posterior, as well as the estimate.

# Theory: Bayesian Interval Estimation

- Interval estimation has a very natural interpretation in Bayesian inference.
- Suppose that we would like to estimate the value of an unobserved random variable  $X$  given that we have observed  $Y = y$ .
- After calculating the **posterior density**  $f_{X|Y}(x|y)$  we can simply find an interval  $[a, b]$  for which we have

$$P(a \leq X \leq b | Y = y) = 1 - \alpha.$$

- The interval is said to be a  $(1 - \alpha)100\%$  **Bayesian credible interval** (BCI) for  $X$ .

## Ex. 6: BCI for jointly normal RVs

Let  $X$  and  $Y$  be jointly normal random variables with  $X \sim \mathcal{N}(0, 1)$ ,  $Y \sim \mathcal{N}(0, 4)$  and covariance  $\rho(X, Y) = 1/2$ .

1. Find a 95% credible interval for  $X$  given that  $Y = 2$  is observed.
2. Plot the prior, likelihood and posterior, as well as the interval estimate.

**HINT:** We can show, by completing the square, that if  $X$  and  $Y$  are jointly normal random variables with parameters  $\mu_X, \sigma_X^2$ ,  $\mu_Y, \sigma_Y^2$ , and  $\rho$ , then, given  $Y = y$  the variable  $X$  is normally distributed with posterior conditionals

$$E[X|Y = y] = \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}$$

$$\text{Var}(X|Y = y) = (1 - \rho^2)\sigma_X^2$$

## Ex. 7: BIP Theory

### Finite-Dimensional Examples: Scalar case

- Let  $u \in \mathbb{R}$  be a scalar unknown, and suppose that we have measurements  $y \in \mathbb{R}^k$ , with  $k \geq 1$ . This is the overdetermined case.
- The measurement is given by the linear relation

$$y = Au + \eta,$$

where the vector  $A \in \mathbb{R}^{k \times 1}$  and the noise  $\eta \sim \mathcal{N}(0, \delta^2 I)$  is zero-mean Gaussian with variance  $\delta^2$ .

- We model the unknown prior of  $u$  as a standard Gaussian measure  $\mathcal{N}(0, 1)$  with zero mean and unit variance.
1. Use Bayes' Theorem to write the posterior in proportional form (without the denominator).

2. The posterior is Gaussian. Derive expressions for its mean and covariance.
  3. What are the asymptotic values of the mean and covariance In the zero-noise limit? Conclusions?
- 

- Consider now the case where the unknown  $u$  is vector-valued and we have only a scalar measurement. That is,  $u \in \mathbb{R}^n$  with  $n \geq 2$ , and  $y \in \mathbb{R}$ , a scalar.
- Suppose again that the measurement is given by the linear relation

$$y = A^\top u + \eta,$$

where the vector  $A \in \mathbb{R}^n \setminus \{0\}$ .

- For the noise, assume  $\eta \sim \mathcal{N}(0, \delta^2)$  and for the prior  $u \sim \mathcal{N}(0, \Sigma_0)$ .

1. Use Bayes' Theorem to write the posterior in proportional form (without the denominator).
2. The posterior is Gaussian. Derive expressions for its mean and covariance.
3. What are the asymptotic values of the mean and covariance In the zero-noise limit? Conclusions?

## Ex. 7: BIP Theory

Finite-Dimensional Examples: Scalar and vector  
cases -> Hints for solution

1. From the BIP theorem, we (see p.57 of 20\_BIP.pdf)

$$\pi^y(u) \propto \exp\left(-\frac{1}{2\delta^2} \|y - Au\|^2 - \frac{1}{2} |u|^2\right).$$

2. Try *completing the square* to show that the posterior mean and covariance are

$$\mu_\delta = \frac{A^\top y}{\delta^2 + \|A\|^2}, \quad \sigma_\delta^2 = \frac{\delta^2}{\delta^2 + \|A\|^2}.$$

3. Asymptotics - see p. 58

The vector-valued case (see pp. 59–61) is the matrix generalization of the above results. However, the asymptotics are more subtle...