Gaissian Models Problems

ELG 5218 – Uncertainty Evaluation in Engineering Measurements and Machine Learning

Instructor: Miodrag Bolić, University of Ottawa

Date: January 28, 2026

PART A: GAUSSIAN INTUITION

A1. Why precisions (not variances) add

Consider the univariate Gaussian model with known variance σ^2:

$$x_i \mid \mu \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n,$$

and prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

(a) Show algebraically that the posterior precision for μ is

$$\lambda_n = \frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}.$$

A2. Precision-weighted averaging: limiting cases

Using the same model and notation as above, the posterior mean is

$$\mu_n = w\,\bar{x} + (1 - w)\mu_0, \quad w = \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2}.$$

(a) Evaluate w in the limit $n \to \infty$ (with fixed σ^2, σ_0^2). What happens to μ_n?

(b) Evaluate w in the limit $\sigma_0^2 \to 0$ (i.e., prior becomes extremely concentrated at μ_0) with fixed n, σ^2. What happens to μ_n?

(c) Interpret these two limits in words.

PART B : SEQUENTIAL GAUSSIAN UPDATES

B1. Learning rate dynamics in online Gaussian learning

In the online update for known variance,

$$\mu_n = \mu_{n-1} + w_n(x_n - \mu_{n-1}), \quad w_n = \frac{\lambda}{\lambda_{n-1} + \lambda}, \quad \lambda = \frac{1}{\sigma^2},$$

where λ_{n-1} is the posterior precision after $n - 1$ observations.

(a) Show that w_n can be written explicitly in terms of n and σ^2, σ_0^2:

$$w_n = \frac{1/\sigma^2}{1/\sigma_0^2 + n/\sigma^2}.$$

(b) Show that w_n is decreasing in n.

(c) Intuition: Explain in one or two sentences why it makes sense that the "step size" toward each new x_n shrinks as more data are observed.

PART C : UNKNOWN VARIANCE – NORMAL–GAMMA

C1. Interpreting Normal–Gamma hyperparameters

In the Normal–Gamma prior

$$\mu \mid \lambda \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0 \lambda}\right), \quad \lambda \sim \text{Gamma}(\alpha_0, \beta_0),$$

we update to

$$\kappa_n = \kappa_0 + n, \quad \mu_n = \frac{\kappa_0 \mu_0 + n\bar{x}}{\kappa_n}, \quad \alpha_n = \alpha_0 + \frac{n}{2}, \quad \beta_n = \beta_0 + \frac{1}{2}\left[\sum (x_i - \bar{x})^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{x} - \mu_0)^2\right].$$

(a) Explain the role of κ_0 in the prior. Why is it often interpreted as an "effective prior sample size"?

(b) Explain the two components inside the bracket in β_n: $\sum (x_i - \bar{x})^2$ and $\frac{\kappa_0 n}{\kappa_0 + n}(\bar{x} - \mu_0)^2$.

(c) Intuition: How does a large discrepancy between $\bar{x}$ and μ_0 affect the posterior for λ via β_n?

PART D : STUDENT-t MARGINAL AND ROBUSTNESS

D1. Tail behavior: Gaussian vs Student-t

Suppose $\mu \mid$ data is (i) Gaussian $\mathcal{N}(\mu_n, s^2)$ and (ii) Student-t with ν degrees of freedom, location μ_n, and scale s.

(a) For large $|\mu - \mu_n|$, which posterior (Gaussian or Student-t) assigns more mass to extreme values, and why?

PART E : DATA-DRIVEN GAUSSIAN ANALYSIS

E1. Simulated Gaussian data with unknown variance

Suppose we simulate $n = 20$ observations from a true model $x_i \sim \mathcal{N}(\mu^\star, \sigma_\star^2)$ with $\mu^\star = 5$ and $\sigma_\star = 2$. We use a Normal–Gamma prior

$$\mu \mid \lambda \sim \mathcal{N}(0, (1 \cdot \lambda)^{-1}), \quad \lambda \sim \text{Gamma}(\alpha_0 = 2, \beta_0 = 2).$$

The sample mean and variance (from one realization) are:

$$\bar{x} \approx 4.8, \quad s^2 \approx 3.4.$$

An MCMC sampler is run to sample (μ, λ) from the posterior. After burn-in, we obtain 4000 iterations from one chain:

- Posterior summaries (from the chain):

$$\hat{\mu} \approx 4.9, \quad \widehat{\text{sd}}(\mu) \approx 0.5; \quad \hat{\sigma}^2 = 1/\hat{\lambda} \approx 3.6.$$

- Diagnostics for μ:

- Trace: looks stationary, centered near 5, with no trend.
- ACF: $\rho(1) \approx 0.3$, $\rho(5) \approx 0.05$, near zero after lag ≈ 10.
- ESS ≈ 1600.

(a) Based on the summaries, does the posterior mean $\hat{\mu}$ appear consistent with the true value $\mu^\star = 5$? Comment briefly.

(b) Using the diagnostics, assess convergence and mixing of the chain for μ.

(c) Intuition: Why is the posterior variance for μ relatively small (around 0.5^2) even though σ^2 is unknown?

OTHER QUESTIONS

Problem 1 – Robustness: Known vs Unknown Variance

Consider the internet speed example from the lecture: we observe data x_i in Mbps and model

$$x_i \mid \mu, \sigma^2 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

We revisit two cases (A and B), but now with a robustness focus.

Case A (known variance). Assume $\sigma^2 = 25$ is known. Prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ with $\mu_0 = 20, \sigma_0^2 = 25$.

Case B (unknown variance). Assume a Normal–Gamma prior:

$$\mu \mid \lambda \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0 \lambda}\right), \quad \lambda \sim \text{Gamma}(\alpha_0, \beta_0),$$

with $\mu_0 = 20, \kappa_0 = 1, \alpha_0 = 2.5, \beta_0 = 12.5$. Here $\lambda = 1/\sigma^2$.

Suppose the observed dataset of size $n = 5$ is

$$x = \{15.77,\ 20.5,\ 8.26,\ 14.37,\ 21.09\}, \quad \bar{x} \approx 16.0, \quad \sum_i (x_i - \bar{x})^2 = 67.6.$$

(a) Derive the posterior for μ in Case A and give the posterior mean and variance in closed form. Then plug in the numbers above to compute μ_n and σ_n^2.

(b) Derive the posterior Normal–Gamma parameters $(\mu_n, \kappa_n, \alpha_n, \beta_n)$ in Case B, and write down the marginal posterior for $\mu \mid x$ (its Student-t form, with ν_n, μ_n, σ_n^2). Compute the numerical values of $\kappa_n, \mu_n, \alpha_n, \beta_n, \nu_n, \sigma_n^2$.

(c) Intuition: Explain in detail why the marginal posterior for μ in Case B has heavier tails than the Gaussian posterior in Case A. In your answer, explicitly connect: (i) uncertainty in σ^2, (ii) the mixture interpretation over σ^2, and (iii) robustness to outliers.

(d) Suppose we add a single extreme outlier, $x_6 = 60$ Mbps, so $n = 6$. We re-run Case A and Case B. Without recomputing full algebra, argue qualitatively (but precisely) how the posterior mean and 95% credible interval for μ will change in each case. Which model (A or B) is more robust to this outlier, and why?

Problem 2 – Sequential Bayesian Filtering vs Exponential Moving Average

Consider the streaming temperature example: $x_t \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$ with known σ^2, and a Gaussian prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. We process data sequentially.

(a) Derive the exact sequential update for the posterior mean μ_t and precision λ_t given μ_{t-1}, λ_{t-1} and new observation x_t. Show that

$$\mu_t = \mu_{t-1} + w_t(x_t - \mu_{t-1}), \quad w_t = \frac{\lambda}{\lambda_{t-1} + \lambda},$$

where $\lambda = 1/\sigma^2$.

(b) Compare the above update with an exponential moving average (EMA) of the form

$$m_t = (1 - \alpha)m_{t-1} + \alpha x_t,$$

where α is fixed. For large t, how does w_t behave? In what sense is Bayesian updating a *data-adaptive* EMA?

(c) Suppose you stream $n = 10^6$ observations and you can only run the computation once (no revisiting old data). Argue why the Bayesian sequential filter has an advantage over the fixed-α EMA for quantifying uncertainty about μ. Be explicit about what you gain (and what you lose) if you only track the EMA.

Problem 3 – Multivariate Gaussian Fusion and Geometry

We consider a 2D sensor fusion scenario. The true latent state is $\mu \in \mathbb{R}^2$, and we observe noisy measurements from two sensors:

$$y_{1,i} \mid \mu \sim \mathcal{N}(\mu, \Sigma_1), \quad y_{2,i} \mid \mu \sim \mathcal{N}(\mu, \Sigma_2),$$

where Σ_1, Σ_2 are known positive-definite covariance matrices, and all observations are conditionally independent given μ.

Assume a Gaussian prior $\mu \sim \mathcal{N}(\mu_0, \Sigma_0)$.

Let n_1, n_2 be the number of readings from each sensor, with empirical means $\bar{y}_1, \bar{y}_2$.

(a) Show that the posterior is $\mu \mid y \sim \mathcal{N}(\mu_n, \Sigma_n)$ with

$$\Sigma_n^{-1} = \Sigma_0^{-1} + n_1\Sigma_1^{-1} + n_2\Sigma_2^{-1}, \quad \mu_n = \Sigma_n \left(\Sigma_0^{-1}\mu_0 + n_1\Sigma_1^{-1}\bar{y}_1 + n_2\Sigma_2^{-1}\bar{y}_2\right).$$

(b) Geometric intuition: Let C_0, C_1, C_2 be the 95% credible ellipses of the prior, sensor-1 likelihood, and sensor-2 likelihood respectively. Describe qualitatively how the shape and orientation of the posterior ellipse C_n will change as:

- n_1 increases with n_2 fixed.
- The correlation in Σ_1 becomes very strong and aligned with one axis, while Σ_2 is nearly spherical.