Exercise 3
MA515 - Foundations of Data Science

IIT Ropar

August 21, 2025

---

1. **Email Spam Classification (Bernoulli Model).** In a spam filter, the presence of a certain keyword is modeled as a Bernoulli random variable with probability p.

   (a) Derive the MLE of p from training data.

   (b) Suppose a Beta(α, β) prior is used to model prior belief about p. Derive the MAP estimate.

   (c) Out of 20 spam emails, the keyword "lottery" appeared in 14. Compute the MLE of p. If the prior is Beta$(2, 3)$, compute the MAP estimate.

2. **Call Center Service Times (Exponential Distribution).** The time between customer calls in a call center is modeled as Exponential(λ).

   (a) Derive the MLE for λ.

   (b) Service times (in minutes) observed are

   $$3.1,\ 2.7,\ 1.8,\ 4.0,\ 2.3.$$

   Compute the MLE of λ.

   (c) If historical data suggests $\lambda \sim \text{Gamma}(\alpha = 3, \beta = 2)$, compute the MAP estimate.

3. **Medical Diagnosis (Disease Prevalence).** A diagnostic test detects a rare disease with probability p. Each patient test result is modeled as Bernoulli(p).

   (a) Derive the MLE of p.

   (b) In a clinical trial, 12 out of 200 patients test positive. Compute the MLE of p.

   (c) If doctors use a prior belief Beta$(5, 95)$ (based on historical prevalence 5%), compute the MAP estimate.

4. **Traffic Flow Modeling (Poisson Distribution).** The number of cars passing a checkpoint in one minute is modeled as Poisson(λ).

   (a) Derive the MLE for λ.

(b) Observed car counts per minute are

$$12,\ 9,\ 10,\ 11,\ 8,\ 13.$$

Compute the MLE of λ.

(c) If historical data suggests $\lambda \sim \text{Gamma}(10, 1)$, compute the MAP estimate.

5. **Stock Return Volatility (Normal Distribution).** Daily returns of a stock are modeled as $\mathcal{N}(0, \sigma^2)$.

   (a) Derive the MLE for σ^2.

   (b) Observed daily returns (in %) are

   $$1.2,\ -0.8,\ 0.5,\ -1.1,\ 1.4.$$

   Compute the MLE of σ^2.

   (c) If analysts place an inverse-Gamma prior on σ^2 with parameters $(\alpha = 3, \beta = 2)$, compute the MAP estimate.

Notes / conventions.

- $\text{Beta}(\alpha, \beta)$ has density proportional to $p^{\alpha-1}(1-p)^{\beta-1}$ on $0 < p < 1$.
- $\text{Gamma}(\alpha, \beta)$ is used as the *rate*–parameter form: density $f(\lambda) \propto \beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}$, so $\mathbb{E}[\lambda] = \alpha/\beta$.
- $\text{Inverse-Gamma}(\alpha, \beta)$ has density proportional to $\beta^\alpha/\Gamma(\alpha)\ x^{-(\alpha+1)} e^{-\beta/x}$ for $x > 0$, and mode $= \beta/(\alpha+1)$ (for $\alpha > 1$).