Winter 2019 Math 106
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 6: Bayesian Inference
6.1 Bayesian inference

In the estimation problem of a variable $\theta$ of interest using a sample of a sample $\{X_i\}$, the conditional probability density of $\theta$ for the given sample $\{X_i\}$ is given by

$$p(\theta|\{X_i\}) = \frac{p(\theta)p(\{X_i\}|	heta)}{\int p(\theta)p(\{X_i\}|	heta)d\theta}$$

from the Bayes’ theorem.

- $p(\theta)$ is a prior density of $\theta$.
- $p(\{X_i\}|	heta)$ is the likelihood of $\{X_i\}$.
- The denominator is a normalization constant.
6.1 Bayesian inference

In Lecture 4, we discussed a parametric inference problem using a parameter $\theta$ and a sample $\{X_i\}$.

- Likelihood $\mathcal{L}_n(\theta) = \prod_i^np(X_i; \theta)$.
- The likelihood is not a probability density of $\theta$.
- $\theta$ is a fixed value and we make probability statements only for the random variables related to the sample for an increasing sample size.

In Bayesian inference,

- We make probability statements about $\theta$, that is, $\theta$ is a random variable.
- The probability describes degree of belief.
- For example, "the probability that it will rain tomorrow is .35"
6.1 Bayesian inference

What do we do with the posterior density?

▶ For a point estimate, we can use the mean of mode of the posterior

▶ We can also obtain a Bayesian interval estimate $C$

\[
\mu(\theta \in C|\{X_i\}) = \int_C p(\theta|\{X_i\})d\theta = 1 - \alpha.
\]

Here, we assume that $\theta$ is a random variable and $\{X_i\}$ is fixed.
6.2 Priors

▶ If we assume a constant for the prior, that is, a uniform density, the mode of the posterior is equal to the maximum likelihood estimator (MLE) because

\[ p(\theta|\{X_i\}) \approx p(\{X_i\}|\theta). \]

Thus, MLE is related to the Bayesian estimator.

▶ However, this does not always hold; if \( \theta \in \mathbb{R} \), there is no uniform density on \( \mathbb{R} \) because

\[ \int_{\mathbb{R}} c dx = \infty. \]

for any constant \( c > 0 \).
6.2 Priors

▶ A constant prior is **not** transformation invariant.
Let’s assume a uniform prior density for $\theta \in (0, 1)$ because of lack of any prior information. For a transformation of $\theta$, $\psi = \ln(\theta/(1 - \theta))$, we also have no prior information and we may assume a uniform prior density for $\psi$.
It is a straightforward exercise to check that the density of $\psi$ is

$$p(\psi) = \frac{e^\psi}{(1 + e^\psi)^2}$$

if we assume a uniform density for $\theta$. 
**Exercise.** Let $X_1, X_2, \ldots, X_n$ be IID of $N(\theta, \sigma^2)$ where $\theta$ is unknown and $\sigma$ is known. Suppose we take as a prior $\theta$ is $N(a_{\text{prior}}, b^2)$ where $a_{\text{prior}}$ and $b$ are known constants.

- The posterior is Gaussian, that is,
  
  \[ p(\theta|\{X_i\}) = \phi(x; a_{\text{post}}, b_{\text{post}}^2) \]
  where $\phi$ is a Gaussian density.

- The posterior mean and variance are

  \[
  a_{\text{post}} = k \left( \frac{1}{n} \sum_i X_i \right) + (1-k) a_{\text{prior}} = a_{\text{prior}} + k \left( \frac{1}{n} \sum_i X_i - a_{\text{prior}} \right)
  \]

  where

  \[
  k = \frac{n}{\sigma^2 + \frac{1}{b^2}}
  \]

  and

  \[
  b_{\text{post}}^2 = \frac{b^2 \sigma^2 / n}{b^2 + \sigma^2 / n}
  \]
6.3 Kalman Filtering

- Kalman filter was co-invented and developed by R.E. Kalman (National Medal of Science 2009).
- Kalman filter is also known as linear quadratic estimation (LQE).
- Kalman filter uses a series of measurements observed over time to estimate unknown variables.
- Kalman filter estimate the conditional density of unknown variables at each time when measurements are available.
### 6.3 Kalman Filtering

1. **Forecast (prediction)**
   - $u_{m+1,\text{post}}$: posterior mean at the $m + 1$-th step.
   - $u_{m+1,\text{prior}}$: prior mean at the $m + 1$-th step.
   - $v_{m+1}$: observation at the $m + 1$-th step.

   The forecast step is represented by the true signal, which is not observed.

2. **Analysis (correction)**
   - $u_{m+1,\text{post}} = u_{m+1,\text{prior}} + K(v_{m+1} - u_{m+1,\text{prior}})$

   where $K$ is the Kalman gain.

   

$$K = \frac{\sigma^2_{m+1,\text{prior}}}{\sigma^2_{\text{obs}} + \sigma^2_{m+1,\text{prior}}}$$