STAT 3006: TUTORIAL 1

1. Bayesian Estimator.
2. Shortest Confidence Interval.

BAYESIAN STATISTICS

- Three features of Bayesian statistics:
- Likelihood function $\mathrm{L}(\Theta \mid X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots, X_n \sim p(x|\Theta)$.
- Prior distribution $\boldsymbol{\pi}(\Theta)$ for Θ (our initial knowledge about Θ).
  - Expert advice
  - Previous study
- Posterior distribution $\boldsymbol{\pi}(\Theta|X_1, X_2, \ldots, X_n)$ for Θ.
  - Bayes rule
- Remark:
  - M.L.E. = argmax $\mathrm{L}(\Theta \mid X_1, X_2, \ldots, X_n)$ (frequentist).
  - Posterior mean = the mean of the posterior distribution $\boldsymbol{\pi}(\Theta|X_1, X_2, \ldots, X_n)$ (Bayesian).
  - Why use posterior mean?
  - Posterior mean minimizes the Bayesian risk $\int\int (\boldsymbol{\delta} - \Theta)^2 p(X|\Theta)\boldsymbol{\pi}(\Theta)\, dX d\Theta$.

TWO IMPORTANT NOTES

- Prior and posterior are about Θ (parameters of interest)

- Likelihood function is about the sample $X_1, X_2, \ldots, X_n$.

CONJUGATE PRIOR

- (*) Prior and posterior belong to the same distribution family.
- When condition (*) holds, we always refer to $\boldsymbol{\pi}(\Theta)$ as a conjugate prior for $p(x|\Theta)$.
- Examples:
  - Normal distribution is conjugate for normal distribution.
  - Gamma distribution is conjugate for exponential distribution.
  - Gamma distribution is conjugate for poisson distribution.
  - Beta distribution is conjugate for Bernoulli distribution.
- Remark:
  - When we calculate $\boldsymbol{\pi}(\Theta|X_1,X_2,\ldots,X_n)$, we only focus on the part that involves Θ. The part is called kernel.
  - Once we have derived the kernel, we can determine which distribution the posterior distribution is.

NORMAL IS CONJUGATE FOR NORMAL

- $X_1, X_2, \ldots, X_n \sim N(\boldsymbol{\mu}, 1)$
- Prior $\boldsymbol{\mu} \sim N(a, b^2)$
  - The parameters in the prior (a and b) are called *hyper-parameters.*
  - Hyper-parameters are selected based on our experiences.
- Posterior $\boldsymbol{\pi}(\boldsymbol{\mu}|X_1, X_2, \ldots, X_n) \propto e^{-\frac{(\boldsymbol{\mu}-\eta)^2}{2\tau^2}}$
  - $\eta = \frac{\frac{a}{b^2}+n\,\bar{X}}{\frac{1}{b^2}+n}, \tau^2 = \frac{1}{n+\frac{1}{b^2}}$
  - The kernel is a normal kernel
  - The posterior distribution of $\boldsymbol{\mu}$ is a normal distribution with mean η and variance $\boldsymbol{\tau}^2$.

GAMMA IS CONJUGATE FOR EXPONENTIAL

- $X_1, X_2, \ldots, X_n \sim Exp(\boldsymbol{\lambda})$
- $\boldsymbol{\lambda} \sim \text{Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta})$; $\boldsymbol{\pi}(\boldsymbol{\lambda}) \propto \boldsymbol{\lambda}^{\boldsymbol{\alpha}-1} \mathrm{e}^{-\boldsymbol{\beta}\boldsymbol{\lambda}}$.
- Posterior distribution is $\text{Gamma}(\mathrm{n} + \boldsymbol{\alpha}, \sum \mathrm{X_i} + \boldsymbol{\beta})$.

GAMMA IS CONJUGATE FOR POISSON

- $X_1, X_2, \ldots, X_n \sim Poi(\boldsymbol{\lambda})$
- $\boldsymbol{\lambda} \sim \mathrm{Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta})$; $\boldsymbol{\pi}(\boldsymbol{\lambda}) \propto \boldsymbol{\lambda}^{\boldsymbol{\alpha}-1} \mathrm{e}^{-\boldsymbol{\beta}\boldsymbol{\lambda}}$.
- Posterior distribution is $\mathrm{Gamma}(\sum \mathrm{X_i} + \boldsymbol{\alpha}, \mathrm{n} + \boldsymbol{\beta})$.

BETA IS CONJUGATE FOR BERNOULLI

- $X_1, X_2, \ldots, X_n \sim Ber(\boldsymbol{\lambda})$
- $\boldsymbol{\lambda} \sim Beta(\mathrm{a}, b)$; $\boldsymbol{\pi}(\boldsymbol{\lambda}) \propto \boldsymbol{\lambda}^{\mathrm{a}-1}(1 - \boldsymbol{\lambda})^{b-1}$.
- Posterior distribution is Beta$(\sum \mathrm{X_i} + \mathrm{a}, \mathrm{n} - \sum \mathrm{X_i} + b)$.

- From the four examples, can you see the relationship between Bayesian estimator and MLE?

BISECTION METHOD

- Find a zero point for a univariate and continuous function.

Algorithm:

INPUT: continuous and univariate function f and interval $[a, b]$ with $f(a)f(b) < 0$.

INITIALIZE: $a^{(0)} \leftarrow a$ and $b^{(0)} \leftarrow b$, and $t \leftarrow 0$.

Repeat

calculate $c^{(t)} \leftarrow \frac{a^{(t)}+b^{(t)}}{2}$;

If $f(c^{(t)}) \cdot f(a^{(t)}) < 0$, let $a^{(t+1)} \leftarrow a^{(t)}$ and $b^{(t+1)} \leftarrow c^{(t)}$;

else if $f(c^{(t)}) \cdot f(b^{(t)}) < 0$, let $a^{(t+1)} \leftarrow c^{(t)}$ and $b^{(t+1)} \leftarrow b^{(t)}$;

else break;

$t \leftarrow t + 1$;

Until $|a^{(t)} - b^{(t)}| < \epsilon$.

OUTPUT: $a^{(t)}$, $b^{(t)}$ in the last iteration. $c^{(t)} \leftarrow \frac{a^{(t)}+b^{(t)}}{2}$ is the final answer.

THE SHORTEST CONFIDENCE INTERVAL

- Problem: we have a density function f(y), and we want to find the interval [a, b] with the smallest length satisfying $\int_a^b f(y)dy = \boldsymbol{\alpha}$. ($\boldsymbol{\alpha}$ is large, e.g. 0.95.)
- Details to calculate the [a, b] can be found in the lecture note, but why the shortest confidence interval must satisfy $f(a) = f(b) = some\ \boldsymbol{\lambda}$?
- Assume there exists a number m s.t. f(x) is strictly increasing when x < m, and f(x) is strictly decreasing when x > m.
- If one interval $[a^*, b^*]$ satisfy $\int_{a^*}^{b^*} f(y)dy = \boldsymbol{\alpha}$ and $f(a^*) = f(b^*)$, then for any other interval [a, b] with $\int_a^b f(y)dy = \boldsymbol{\alpha}$, we always have $b - a > b^* - a^*$.
- Proof:…