Advanced Statistics - Tutorial 05/08/2021

Part 1: Sum of two independent random variables / convolution

Theorem:

X_1, X_2 real-valued independent random variables with prob. densities $f_1(x), f_2(x)$

$\Rightarrow X_1 + X_2$ has prob. density $g(x) = \int_{-\infty}^{\infty} f_1(y)\, f_2(x-y)\, dy$

$$= \int_{-\infty}^{\infty} f_1(x-y)\, f_2(y)\, dy$$

Part 2: Bayesian statistics

① Classical vs. Bayesian statistics

Determine parameter ϑ / θ

based on observations $x = (x_1, \ldots, x_n)$

Classical statistics

- x is a realisation of a random variable X
- ϑ is an unknown but deterministic parameter
- Mathematical model:

  $f_\vartheta(x)$ likelihood

- Point estimators:
  - Maximum likelihood of $f_\vartheta(x)$
  - ...

Bayesian statistics

- x is a realisation of a random variable X
- ϑ is a realisation of a random variable Θ
- Mathematical model:

  $$f(\vartheta \mid x) \propto f(x \mid \vartheta)\, f(\vartheta)$$

  a posteriori density of Θ — likelihood — a priori density of Θ

  proportional with const

  $\left(\int f(x \mid \vartheta)\, f(\vartheta)\, d\vartheta \right)^{-1}$ continuous case

  $\left(\sum_{\vartheta} f(x \mid \vartheta)\, f(\vartheta) \right)^{-1}$ discrete case

- Point estimators:
  - Mode
  - Median
  - Mean

  $\left.\right\}$ of $f(\vartheta \mid x)$
  - ...

- Interval estimations:

  $\vartheta \in [a(x), b(x)]$ with confidence level $1-\alpha$

  $\Leftrightarrow P_\vartheta\left(a(X) \leq \vartheta \leq b(X)\right) \geq 1-\alpha$ for all ϑ

  random variable

- Interval estimations:

  $\theta \in [a(x), b(x)]$ with probability $1-\alpha$

  $\Leftrightarrow P\left(\theta \in [a(x), b(x)] \mid X = x\right) \geq 1-\alpha$

② Example 1: AIDS test

$f(\text{pos} \mid \text{AIDS}) = 0.999$ (accuracy of the test)

$f(\text{pos} \mid \neg\text{AIDS}) = 0.05$ (false positive rate)

$f(\text{AIDS}) = 0.00035$ (prior / prevalence of AIDS)

$$\Rightarrow f(\text{AIDS} \mid \text{pos}) = \frac{f(\text{pos} \mid \text{AIDS})\, f(\text{AIDS})}{f(\text{pos} \mid \text{AIDS})\, f(\text{AIDS}) + f(\text{pos} \mid \neg\text{AIDS})\, f(\neg\text{AIDS})} \approx 0.069$$

③ Example 2: Toss a coin n times and estimate probability ϑ of "head"

Likelihood: Binomial(n, ϑ) distribution

$$\Rightarrow f(k \mid \vartheta) = \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} \quad \text{for } k \in \{0, 1, \dots, n\}$$

Prior: depends on our beliefs, e.g. Beta(a, b) distribution

$$\Rightarrow f(\vartheta) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}_{= \frac{1}{B(a,b)} \text{ Beta function}} \vartheta^{a-1} (1-\vartheta)^{b-1} \propto \vartheta^{a-1} (1-\vartheta)^{b-1}$$

Posterior:

$$\begin{aligned} \Rightarrow f(\vartheta \mid k) &\propto f(k \mid \vartheta)\, f(\vartheta) \\ &\propto \vartheta^k (1-\vartheta)^{n-k} \cdot \vartheta^{a-1} (1-\vartheta)^{b-1} \\ &= \vartheta^{k+a-1} (1-\vartheta)^{n-k+b-1} \end{aligned}$$

Beta$(k+a, n-k+b)$ distribution

$\Rightarrow$ Prior and posterior distribution are from the same class of distributions ("conjugate prior")