Parametric Statistics

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

November 23, 2022

Lecture Summary

- 7.1 Statistical Inference

Bayesian Methods:

- 7.2 Prior and Posterior Distributions
- 7.3 Conjugate Prior Distributions

Statistical Inference

We have seen statistical models in the form of probability distributions:

$$f(x \mid \theta)$$

- In this section the general notation for any parameter will be θ
- The parameter space will be denoted by Ω

For example:

- Life time of a Christmas light series follows the $\mathrm{Expo}(\theta)$
- The average of 63 poured drinks is approximately normal with mean θ
- The number of people that have a disease out of a group of N people follows the $\mathrm{Binomial}(N, \theta)$ distribution.

In practice the value of the parameter θ is unknown.

Statistical Inference

Statistical Inference: Given the data we have observed what can we say about θ ?

- i.e., we observe random variables $X_1, \ldots, X_n$ that we assume follow our statistical model and then we want to draw probabilistic conclusions about the parameter θ.

For example:

- I tested 5 Christmas light series from the same manufacturer and they lasted for $21, 103, 76, 88$ and 96 days.
- Assuming that the life times are independent and follow $\mathrm{Expo}(\theta)$, what does this data set tell me about the failure rate θ?

Statistical Inference - Types of Inference

Say I take a random sample of 100 people and test them all for a disease. If 3 of them have the disease, what can I say about $\theta =$ the prevalence of the disease in the population?

Estimation

Say I estimate θ as $\hat{\theta} = 3/100 = 3\%$.

Confidence intervals

How sure am I about this number? I want uncertainty bounds on my estimate.

Testing hypotheses

Can I be confident that the prevalence of the disease is higher than 2%?

Prediction

If we test 40 more people for the disease, how many people do we predict have the disease?

Bayesian vs. Frequentist Inference

Should a parameter be treated as a random variable?

- Do we think about $f(\mathbf{x} \mid \theta)$ as the conditional pdf/pf of $\mathbf{X}$ given θ or
- do we think about $f(\mathbf{x} \mid \theta)$ as a pdf/pf indexed by θ that is unknown?

Bayesians vs Frequentists:

Consider the prevalence of a disease.

Frequentists

The proportion q of the population that has the disease, is not a random phenomenon but a fixed number that is simply unknown

Bayesians:

The proportion Q of the population that has the disease is unknown and the distribution of Q is a subjective probability distribution that expresses the experimenters (prior) beliefs about Q

Bayesian Inference

Theorem (7.2.1: Calculating the posterior)

Let $X_1, \ldots, X_n$ be a random sample with pdf/pf $f(x \mid \theta)$ and let $f(\theta)$ be the prior pdf/pf *of θ. The the posterior* pdf/pf *is*

$$f(\theta \mid \mathbf{x}) = \frac{f(x_1 \mid \theta) \times \cdots \times f(x_n \mid \theta) f(\theta)}{f(\mathbf{x})}$$

where

$$f(\mathbf{x}) = \int_\Omega f(\mathbf{x} \mid \theta) p(\theta) d\theta$$

is the marginal likelihood of $X_1, \ldots, X_n$

Bayesian Inference

Prior distribution

The distribution we assign to parameters before observing the random variables. Notation for the prior pdf/pf: We will use $p(\theta)$, the book uses $\xi(\theta)$

Likelihood

When the joint $\mathrm{pdf}/\mathrm{pf}\, f(\mathbf{x} \mid \theta)$ is regarded as a function of θ for given observations $x_1, \ldots, x_n$ it is called the likelihood function.

Posterior distribution

Posterior distribution: The conditional distribution of the parameters θ given the observed random variables $X_1, \ldots, X_n$. Notation for the posterior pdf/pf : We will use

$$f\left(\theta \mid x_1, \ldots, x_n\right) = p(\theta \mid \mathbf{x})$$

Example: Bernoulli Likelihood and a Beta prior

I take a random sample of 40 people and test them all for a disease. Let X be the r.v. denoting if a person has the disease or not.

- $X \mid \theta \sim \text{Bernoulli}(\theta)$,

where X denotes the number of people with the disease.

- I need a prior defined in [0, 1]

Beta Distributions distributions

Definition (Beta distribution)

A r.v.X has the Beta distribution with parameters $\alpha, \beta > 0$ if

$$f(x|\alpha,\beta) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & x \in [0,1] \\ 0 & otherwise \end{cases}$$

- Suitable for r.v.s in $[0,1]$
- Parameter space: $\alpha, \beta > 0$.
- $E(X) = \frac{\alpha}{\alpha+\beta}, Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
- MGF: $1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$.

Example: Beta-Bernoulli distribution

- I observe $X_1, \ldots, X_{40}$ with $\sum_{i=1}^{40} x_i = 10$, and I want to find the posterior distribution of θ.
- Pick a prior, e.g., $Beta(2, 2)$:

$$f(\theta) = \frac{1}{B(2,2)}\theta(1-\theta)$$

- Compute the likelihood:

$$f(x_1, \ldots, x_{40}|\theta) = \prod_{i=1}^{40} f(x_i|\theta) = \theta^{10}(1-\theta)^{30}$$

- Compute the posterior up to a constant:

$$f(\theta|x_1, \ldots, x_{40}) = \frac{1}{f(x_1, \ldots, x_{40})} f(\theta) f(x_1, \ldots, x_{40}|\theta) =$$

$$C\theta^{10+1}(1-\theta)^{30+1}$$

- C is a constant, $f(\theta|x_1, \ldots, x_{40})$ is a $Beta(12, 32))$ distribution.

Example:Beta-Bernoulli

- In the general case:
- If the prior is $f(\theta) = Beta(\alpha, \beta)$, the posterior is $f(\theta|x_1, \ldots, x_n) = Beta(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$.
- When the prior and the posterior belong to the same family of distributions, we say the distribution is a conjugate prior for the distribution of the likelihood.
- For example, Beta is a conjugate prior for the Bernoulli distribution.

Gamma Distributions distributions

Definition (Gamma distribution)

A r.v.X has the Gamma distribution with parameters $\alpha, \beta > 0$ if

$$f(x|\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x > 0 \\ 0 & otherwise \end{cases}$$

- Suitable for r.v.s in $(0, \infty)$
- Parameter space: $\alpha, \beta > 0$.
- $E(X) = \frac{\alpha}{\beta}, Var(X) = \frac{\alpha}{\beta^2}$.
- MGF:$\left(1 - \frac{t}{\beta}\right)^{-\alpha}$ for $t < \beta$

Another Example: Exponential Distribution

- I observe $X_1, \ldots, X_n$ where $X_i \sim Expo(\lambda)$
- Pick a prior for λ, e.g., $\lambda \sim Gamma(\alpha, \beta)$
- Compute the posterior up to a constant:

$$f(\lambda | x_1, \ldots, x_n) = C\lambda^{\alpha+n} e^{-\lambda(\beta + \sum_{i=1}^{n} x_i)}$$

- $\lambda \sim Gamma(\alpha + n, \beta + \sum_{i=1}^{n} x_i)$