Parametric Statistics

Statistical Inference

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

October 30, 2023

Lecture Summary

- 7.1 Statistical Inference
- 7.2 Prior and Posterior Distributions
- 7.3 Conjugate Prior Distributions
- 5.7 The Beta Distributions
- 5.8 The Gamma Distributions

Statistical Inference

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

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

For example:

- Life time of a Christmas light series follows the Expo(θ).
- The average volume of 63 drinks is approximately normal with mean θ.
- The number of people that have a disease out of a group of N people follows the Binomial(N, θ) 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 Expo(θ), 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?

Statistic

I want to use the 100 patients to make the statistical inferences above. Do I need to keep all 100 values, or can I use a summary?

Definition

Suppose that the observable random variables of interest are $X_1, \ldots, X_n$. Let r be an arbitrary real-valued function of n real variables. Then the random variable $T = r(X_1, \ldots, X_n)$ is called a statistic.

Examples of Statistics.

- The sample mean $\overline{X}_n$.
- The maximum Y_n of the values of $X_1, \ldots, X_n$.
- The function $r(X_1, \ldots, X_n)$, which has the constant value 3 for all values of $X_1, \ldots, X_n$.

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 pf of $\mathbf{X}$ given θ or
- do we think about $f(\mathbf{x} \mid \theta)$ as a 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

Calculating the posterior

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

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

where

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

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 pf: $\xi(\theta)/f(\theta)$

Likelihood

When the joint 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

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

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

Example: Bernoulli Likelihood and a Beta Prior

A Clinical Trial. Suppose that 40 patients are going to be given a treatment for a condition and that we will observe for each patient whether or not they recover from the condition. We are most likely also interested in a large collection of additional patients besides the 40 to be observed.

- For each patient $i = 1, 2, \ldots$, let $X_i = 1$ if patient i recovers, and let $X_i = 0$ if not.
- $X_i \sim Bernoulli(p), 0 \leq p \leq 1$.
- WLLN: The proportion of the first n patients who recover $\overline{X}_n \xrightarrow{p} p$ as n goes to infinity.
- I need a prior defined on the parameter space $[0, 1]$

Beta Distributions distributions

Definition (Beta distribution)

A RV 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 RVs 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.

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., $Beta(2, 2)$:

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

- Compute the likelihood:

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., $Beta(2,2)$:

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

- 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:

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., $Beta(2,2)$:

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

- 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:

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

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

- 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 $\xi(\theta) = Beta(\alpha, \beta)$, 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.

Prior and Posterior

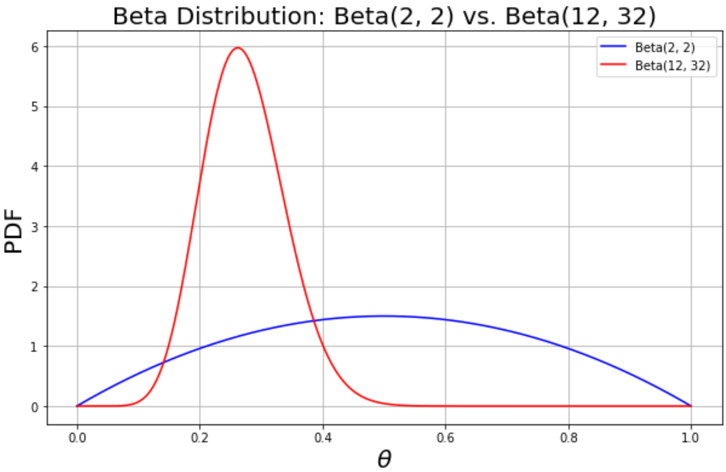


Figure: Prior and posterior distributions for parameter θ

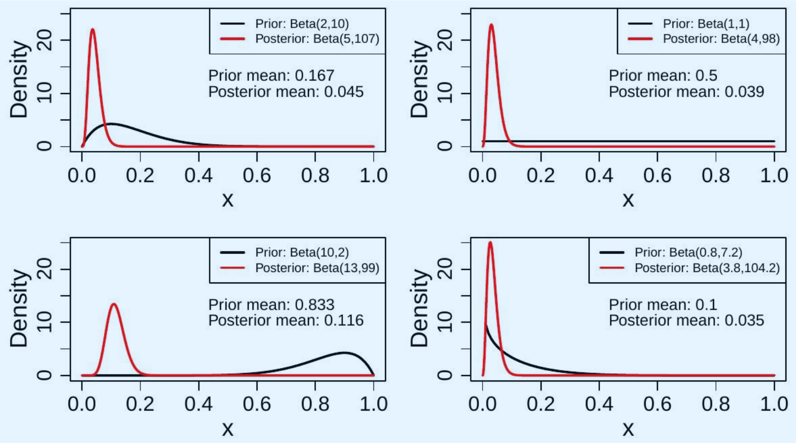
Prior distributions

- The prior distribution should reflect what we know apriori about θ.
- For example: Beta$(2, 10)$ puts almost all of the density below 0.5 and has a mean $2/(2+10) = 0.167$, saying that a prevalence of more then 50% is very unlikely.
- Using Beta $(1, 1)$, i.e. the Uniform $(0, 1)$ indicates that a priori all values between 0 and 1 are equally likely.

Choosing a prior

- Deciding what prior distribution to use can be very difficult.
- We need a distribution (e.g. Beta) and its hyperparameters (e.g. α, β).
- When hyperparameters are difficult to interpret we can sometimes set a mean and a variance and solve for parameters E.g: What Beta prior has mean 0.1 and variance 0.1^2 ?
- If more than one option seems sensible, we perform sensitivity analysis.

Sensitivity Analysis



We compare the posteriors we get when using the different priors.

Sensitivity Analysis

The posterior is influenced both by sample size and the prior variance

- Larger sample size $\Rightarrow$ less the prior influences the posterior
- Larger prior variance $\Rightarrow$ the less the prior influences the posterior Prior variance: 0.011

Steps To Bayesian Estimation

- Pick a prior distribution.
- Compute the likelihood.
- Use Bayes' theorem to compute the posterior distribution:

$$\text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution}$$

- Perform Sensitivity Analysis.
- Summarize the posterior distribution.

Gamma Distributions

Definition (Gamma distribution)

A RV 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 RVs 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 $t < \beta$

Steps To Bayesian Estimation

- Pick a prior distribution.
- Compute the likelihood.
- Use Bayes' theorem to compute the posterior distribution:

$$\text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution}$$

Another Example: Exponential Distribution

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

Reminder

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

$$Expo(\beta) : f(x \mid \beta) = \begin{cases} \beta e^{-\beta x} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0. \end{cases}$$

Recap: Steps To Bayesian Estimation

- Pick a prior distribution.
- Compute the likelihood.
- Use Bayes' theorem to compute the posterior distribution:

  $$\text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution}$$

- Perform Sensitivity Analysis.
- Summarize the posterior distribution.