Statistics primer & exponential families

“Algebraic & Geometric Methods in Statistics”

Sonja Petrović
Created for Math/Stat 561

Jan 30, 2023.

Objective

- Review some statistics fundamentals;
- understand the setup of exponential families.

Material:

Sourced from chapters 5 (“Statistics primer”) and 6 (“exponential families”) of the textbook. Other resources provided in subsequent links.

Parametric models: self-review

- What is a parametric statistical model?
- What is an implicit statistical model?
- What does it mean for random variables $X_1, \ldots, X_n$ to be *iid* (independent and identically distributed)?
- What does it mean for random variables to be exchangable?
- What is an *iid* sample?

Task:

Look up, write down, and adopt these definitions.
See Sections 5.1. and 5.2. of the textbook (handout).

The running example

Example 5.3.2.: Binomial random variable: $r+1$ states, $0, \ldots, r$.

- The model consists of all distributions of the form

$$\{\Big(\pi^r, r\pi^{r-1}(1-\pi), \ldots, (1-\pi)^r\Big) : \pi \in [0,1]\}.$$

In other words, the model is the set $\{p_\pi\} \subset \mathbb{R}^{r+1}$ where each p_π has the above form.

- Data collected from an *iid* sample of size n: $X^{(1)}, \ldots, X^{(n)}$, from an underlying distribution p_{π_0}.
- π_0 is the unknown but fixed parameter we would like to estimate using the data.

Statistics vs. parameters

Parameter [Definition 5.3.1.]

Let $\mathcal{M}_\theta$ be a parametric statistical model with parameter space Θ. A parameter of a statistical model is a function $s : \Theta \to \mathbb{R}$.

Statistic [Definition 5.1.5.]

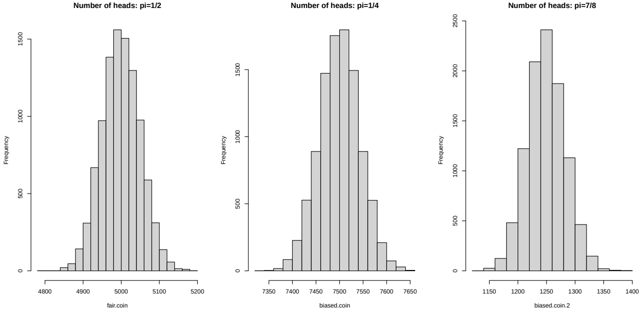
A statistic is a function from the state space to some other set.

- A statistic $T(X)$ is sufficient for the model if $P(X = x | T(X) = t, \theta) = P(X = x | T(X) = t)$.

Estimator [Definition 5.3.1.]

An estimator $\hat{\theta}$ is a function from the data space D to $\mathbb{R}$.

- An estimator is consistent if $\hat{\theta} \to_P \theta$. $\leftarrow$ it converges to the true parameter as the sample size $\to \infty$.
  - There are *many* ways to compute an estimator.



Three simulations, **one** parametric model, **one** unknown parameter

$\rightarrow$ The parametric model: $Bin(10000, \pi)$

- Histogram 1: data simulated with $\pi = 1/2$.
- Histogram 2: data simulated with $\pi = 1/4$.
- Histogram 3: data simulated with $\pi = 7/8$.

What is the parameter estimation problem on this example?

Write it out.

(What is X, θ or π, r, n, a statistic, an estimator of a parameter?)

The parameter estimation problem

- There are many ways to compute estimatores.
  - See Math 563, for starters; Method of moments, for example
  - **READ** Examples in the book re: binomial r.v.: 5.3.2, 5.3.4, 5.3.6.

Maximum likelihood estimation [Defn. 5.3.5.]

Let D be data from some model w/ parameter space Θ. Likelihood function:

$$L(\theta|D) := p_\theta(D) \text{ or } L(\theta|D) := f_\theta(D).$$

- L is a function of the parameter(s)! Data is *fixed* in the likelihood function.

The maximum likelihood estimate (MLE) $\hat{\theta}$ is the maximizer of the likelihood function:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta|D).$$

- MLE = the particular value $\hat{\theta}$ of the parameter that makes D most likely to have been observed udner the model.

Let's look at the Binomial example again.

- The data $D = X^{(1)}, \ldots, X^{(n)}$ is summarized by a vector of counts $u = (u_0, ..., u_r)$, where $u_i = |\{j : X(j) = i\}|$.
- In the case of discrete data, this likelihood function is thus only a function of the vector of counts u:

$$L(\theta|D) = \prod_j p_\theta(j)^{u_j}.$$

* It is common to study the log-likelihood function $\ell(\theta|D) = \log L(\theta|D)$.

Binomial likelihood

Go over **Example 5.3.6**.

- What is the likelihood function?
- What is the MLE?

MLE for $\mathcal{M}_{1 \perp\!\!\!\perp 2}$.

Go over Proposition 5.3.8. and proof.

Exponential families

- definition [book. example w/ A and h.]
- binomial distribution as an expo family
- discrete. examples 6.2.6 and 6.2.7.
- log-linear models.

under construction

Unfortunately, I didnot have time to re-type these notes, so for now I will switch to lecture notes by Prof. Kaie Kubjas, who is also using the same textbook!

go to pdf

Other resources

- Check out this lovely tutorial on MLE by Prof. Andrew Moore.
- Larry Wasserman's intermediate statistics notes on likelihood and sufficiency: read this and this.
- Eliana Duarte's summer school lectures include these slides on [exponential families: an algebraic statistics perspective], see page 13-18. **link will be provided ASAP.**
- Michael I. Jordan's chapter on exponential families provides another resource equivalent to the background in Chapter 6.

License