18.650
Statistics for Applications

Chapter 2: Parametric Inference

The rationale behind statistical modeling

- Let $X_1, \ldots, X_n$ be n independent copies of X.
- The goal of statistics is to learn the distribution of X.
- If $X \in \{0, 1\}$, easy! It's $\text{Ber}(p)$ and we only have to learn the parameter p of the Bernoulli distribution.
- Can be more complicated. For example, here is a (partial) dataset with number of siblings (including self) that were collected from college students a few years back: 2, 3, 2, 4, 1, 3, 1, 1, 1, 1, 1, 2, 2, 3, 2, 2, 2, 3, 2, 1, 3, 1, 2, 3, . . .
- We could make no assumption and try to learn the pmf:

x	1	2	3	4	5	6	≥ 7
$\mathbb{P}(X = x)$	p_1	p_2	p_3	p_4	p_5	p_6	$\sum_{i \geq 7} p_i$

  That's 7 parameters to learn.
- Or we could assume that $X \sim \text{Poiss}(\lambda)$. That's 1 parameter to learn!

Statistical model (1)

Formal definition

Let the observed outcome of a statistical experiment be a sample $X_1, \dots, X_n$ of n i.i.d. random variables in some measurable space E (usually $E \subseteq \mathbb{R}$) and denote by $\mathbb{P}$ their common distribution. A *statistical model* associated to that statistical experiment is a pair

$$\left(E, (\mathbb{P}_\theta)_{\theta \in \Theta}\right),$$

where:

- E is *sample space*;
- $(\mathbb{P}_\theta)_{\theta \in \Theta}$ is a family of probability measures on E;
- Θ is any set, called *parameter set*.

Statistical model (2)

- Usually, we will assume that the statistical model is *well specified*, i.e., defined such that $\mathbb{P} = \mathbb{P}_\theta$, for some $\theta \in \Theta$.

- This particular θ is called the true parameter, and is unknown: The aim of the statistical experiment is to *estimate* θ, or check it's properties when they have a special meaning ($\theta > 2$?, $\theta \neq 1/2$?, ...)

- For now, we will always assume that $\Theta \subseteq \mathbb{R}^d$ for some $d \geq 1$: The model is called *parametric*.

Statistical model (3)

Examples

1. For n Bernoulli trials:
$$\left(\{0,1\}, (\mathsf{Ber}(p))_{p\in(0,1)}\right).$$

2. If $X_1, \ldots, X_n \overset{iid}{\sim} \mathsf{Exp}(\lambda)$, for some unknown $\lambda > 0$:
$$\left(\mathbb{R}_+^*, (\mathsf{Exp}(\lambda))_{\lambda>0}\right).$$

3. If $X_1, \ldots, X_n \overset{iid}{\sim} \mathsf{Poiss}(\lambda)$, for some unknown $\lambda > 0$:
$$\left(\mathbb{N}, (\mathsf{Poiss}(\lambda))_{\lambda>0}\right).$$

4. If $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$:
$$\left(\mathbb{R}, \left(\mathcal{N}(\mu, \sigma^2)\right)_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+^*}\right).$$

Identification

The parameter θ is called *identified* iff the map $\theta \in \Theta \mapsto \mathbb{P}_\theta$ is injective, i.e.,

$$\theta = \theta' \Rightarrow \mathbb{P}_\theta = \mathbb{P}_{\theta'}.$$

Examples

1. In all four previous examples, the parameter was identified.

2. If $X_i = \mathbb{1}_{Y_i \geq 0}$, where $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, are unobserved: μ and σ^2 are not identified (but $\theta = \mu/\sigma$ is).

Parameter estimation (1)

Idea: Given an observed sample $X_1, \dots, X_n$ and a statistical model $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$, one wants to *estimate* the parameter θ.

Definitions

- *Statistic*: Any measurable[1] function of the sample, e.g., $\bar{X}_n, \max_i X_i$, $X_1 + \log(1 + |X_n|)$, sample variance, etc...
- *Estimator* of θ: Any statistic whose expression does not depend on θ.
- An estimator $\hat{\theta}_n$ of θ is *weakly* (resp. *strongly*) *consistent* iff

$$\hat{\theta}_n \xrightarrow[n \to \infty]{\mathbb{P} \text{ (resp. } a.s.)} \theta \quad \text{(w.r.t. } \mathbb{P}_\theta).$$

[1] Rule of thumb: if you can compute it exactly once given data, it is measurable. You may have some issues with things that are implicitly defined such as $\sup$ or $\inf$ but not in this class

Parameter estimation (2)

- *Bias* of an estimator $\hat{\theta}_n$ of θ:

$$\mathbb{E}\left[\hat{\theta}_n\right] - \theta.$$

- *Risk* (or *quadratic risk*) of an estimator $\hat{\theta}_n$:

$$\mathbb{E}\left[|\hat{\theta}_n - \theta|^2\right].$$

Remark: If $\Theta \subseteq \mathbb{R}$,

"Quadratic risk $=$ bias^2 $+$ variance".

Confidence intervals (1)

Let $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$ be a statistical model based on observations $X_1, \ldots, X_n$, and assume $\Theta \subseteq \mathbb{R}$.

Definition

Let $\alpha \in (0, 1)$.

- *Confidence interval (C.I.) of level* $1 - \alpha$ for θ: Any random (i.e., depending on $X_1, \ldots, X_n$) interval $\mathcal{I}$ whose boundaries do not depend on θ and such that:

$$\mathbb{P}_\theta \left[\mathcal{I} \ni \theta \right] \geq 1 - \alpha, \qquad \forall \theta \in \Theta.$$

- *C.I. of asymptotic level* $1 - \alpha$ for θ: Any random interval $\mathcal{I}$ whose boundaries do not depend on θ and such that:

$$\lim_{n \to \infty} \mathbb{P}_\theta \left[\mathcal{I} \ni \theta \right] \geq 1 - \alpha, \qquad \forall \theta \in \Theta.$$

Confidence intervals (2)

Example: Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathsf{Ber}(p)$, for some unknown $p \in (0,1)$.

- LLN: The sample average $\bar{X}_n$ is a strongly consistent estimator of p.

- Let $q_{\alpha/2}$ be the $(1 - \frac{\alpha}{2})$-quantile of $\mathcal{N}(0,1)$ and

$$\mathcal{I} = \left[\bar{X}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \bar{X}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} \right].$$

- CLT: $\displaystyle\lim_{n \to \infty} \mathbb{P}_p \left[\mathcal{I} \ni p \right] = 1 - \alpha, \quad \forall p \in (0,1).$

- Problem: $\mathcal{I}$ depends on p !

Confidence intervals (3)

Two solutions:

- Replace $p(1-p)$ with $1/4$ in $\mathcal{I}$ (since $p(1-p) \leq 1/4$).
- Replace p with $\bar{X}_n$ in $\mathcal{I}$ and use Slutsky's theorem.