UNIVERSITY OF CAMBRIDGE

Theoretical Statistics

May 21, 2021

Overview

Statistics 101

Bayesian statistics

Setup

Definition

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and $\mathcal{P} = \{P_\theta \mid \theta \in \Omega\}$ a set of probability measures on $(\mathcal{X}, \mathcal{F})$. We call $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ a *statistical model*.

Definition

A function $T : \mathcal{X} \to \mathbb{R}^n$ is called a *statistic*.

Definition

Given a sample X, we want to find a statistic δ such that $\delta(X)$ would be close to $g(\theta)$. In this context δ is called an *estimator*.

Loss and Risk

Definition

To be able to compare different estimators we define the *loss* function $L(\theta, d)$ as the cost of estimating $g(\theta)$ with d.

We can then define the *risk* of a particular estimator δ as

$$R(\theta, \delta) = \mathbb{E}_{X \sim P_\theta}\left[L(\theta, \delta(X))\right]$$

Loss and Risk

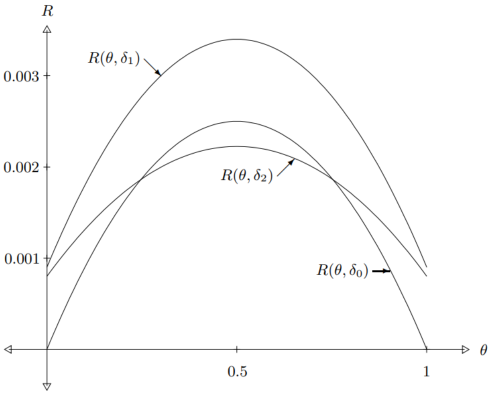

Fig. 3.2. Risks for δ_0, δ_1, and δ_2.

UNIVERSITY OF CAMBRIDGE

Loss and Risk (Decision Theory)

Definition

Alternatively we could consider the *action space* $\mathcal{A}$, define the loss function and the *nonrandomized* decision rule as

$$L : \Omega \times \mathcal{A} \to \mathbb{R}$$
$$\delta : \mathcal{X} \to \mathcal{A}$$

Then the risk is defined in the same way

$$R(\theta, \delta) = \mathbb{E}_\theta\big[L(\theta, \delta(X))\big]$$

Hypothesis Testing

Hypothesis testing as a Decision Problem

Given hypotheses $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega_1$ we can associate

1. an action space $\mathcal{A} = \{0, 1\}$.
2. a loss function

$$L(\theta, a) = I\{a = 1, \theta \in \Omega_0\} + I\{a = 0, \theta \in \Omega_1\}$$

Sufficient statistics

Definition

Given a statistical model $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, a statistic T is called a *sufficient* if $X|T = t$ under P_θ does not depend on θ.

Equivalently, if $I(\Theta, X) = I(\Theta, T(X))$ for any distribution on Θ.

Simulating samples using sufficient statistics

If T is a sufficient statistic, then define

$$Q_t(B) = \mathbb{P}(X \in B \mid T = t)$$

We can now sample $X \sim P_\theta$, $T = T(X)$ and $\widetilde{X} \sim Q_T$.

Neyman's Factorization Theorem

Theorem

Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be a statistical model where each P_θ is dominated by μ and has density p_θ. $T(X)$ is a sufficient statistic if and only if there exist functions $g_\theta(x)$ and $h(x)$ such that for all θ

$$p_\theta(x) = g_\theta(T(x))h(x)$$

UNIVERSITY OF CAMBRIDGE

Rao-Blackwell theorem

Theorem

Let T be a sufficient statistic for $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ and δ and estimator of $g(\theta)$ and define

$$\nu(T) = \mathbb{E}\big[\delta(X) \mid T\big]$$

If $L(\theta, \cdot)$ is convex and $R(\theta, \delta) < \infty$, then

$$R(\theta, \nu) \leq R(\theta, \delta)$$

Bayesian statistics

Definition

In Bayesian statistics, one assumes an additional prior distribution $\Lambda(\cdot)$ on Ω. Then we might want to optimize

$$\int_\Omega R(\theta, \delta)\Lambda(\mathrm{d}\theta)$$

Estimators δ which optimize this criterion are called Bayes estimators.

Theorem

Suppose that $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ is a statistical model together with a prior distribution Λ on Ω and that $L(\theta, d) \geq 0$ for all relevant arguments. If

1. $\mathbb{E}\big[L(\Theta, \delta_0)\big] < \infty$ for some δ_0
2. for a.e. x there exists a value $\delta_\Lambda(x)$ that minimizes

$$\mathbb{E}\big[L(\Theta, d) | X = x\big]$$

   over d

then δ_Λ is a Bayes estimator.

Bayesian sufficient statistics

Definition

If $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ is a statistical model and $\Lambda(\cdot)$ is a Bayesian prior, then T is a *Bayesian sufficient statistic* if for almost every x

$$P(\theta|X = x) = P(\theta|T(X) = t(x))$$

Definition

If $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ is a statistical model and $\Lambda(\cdot)$ is a Bayesian prior, then T is a *predictive sufficient statistic* if for almost every x

$$P(X^* = x^*|X = x) = P(X^* = x^*|T(X) = t(x))$$

Admissibility

Definition

Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be a statistical model and let δ be an estimator. If there is a δ_* such that

$$R(\theta, \delta_*) \leq R(\theta, \delta)$$

with strict inequality for at least one θ, then δ is called inadmissible. Otherwise it is admissible.

Theorem

If δ_Λ is an essentially unique Bayes estimator, it is admissible.

Complete Class Theorems

Definition

A family of estimators is called a complete class if any δ **not** in the class is inadmissible.

Theorem

Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be a statistical model and let $\mathcal{B}_0$ be the class of all Bayes estimators w.r.t. different priors $\Lambda(\cdot)$ on **finite** subsets of Ω. Suppose that

- $\mathcal{P}$ is dominated with densities $p_\theta(x) > 0$ everywhere
- $L(\theta, \cdot)$ is nonnegative and strictly convex
- $L(\theta, a) \to \infty$ as $\|a\| \to \infty$

Then the set of pointwise limits of $\mathcal{B}_0$ is a complete class.

Bayesian vs Frequentist

$$l(\theta, \delta(x))$$

- Frequentist - $R(\theta, \delta) = \mathbb{E}_\theta\big[l(\theta, \delta(X))\big]$
- Bayesian - $\rho(x, \delta) = \mathbb{E}\big[l(\Theta, \delta(X))|X = x\big]$
- Coherency – decisions/beliefs are consistent with Bayes theorem. Frequentist statistics are incoherent – Dutch Book argument.

Objective vs Subjective Bayes

- Subjective Bayesian – allows any priors
- Objective Bayesian – restricts to diffuse priors
- Empirical Bayes – choose prior from data (James-Stein estimator)
- `https://stats.stackexchange.com/questions/381825/objective-vs-subjective-bayesian-paradigms`

James-Stein estimator

S

uppose that $X_i \sim N(\theta_i, 1)$ and consider the loss function $l(\theta, \delta(X)) = \|\theta - \delta(X)\|^2$. Then the estimator

$$\delta(x) = x$$

is admissible for $n = 1, 2$, but inadmissible for $n > 2$.
It is dominated by the James-Stein estimator:

$$\delta_{\mathrm{JS}}(x) = \left(1 - \frac{n-2}{\|x\|^2}\right)x$$

References I

- José M Bernardo and Adrian FM Smith, *Bayesian theory*, vol. 405, John Wiley & Sons, 2009.
- Robert W Keener, *Theoretical statistics: Topics for a core course*, Springer Science & Business Media, 2010.