Lecture 4 - scribbles

Friday, September 11, 2020 13:09

today: • finish prob.
• frequentist vs. Bayesian

binomial distribution:

model n indep. coin flips

sum of n indep. Bern(θ) R.V.

let $X_i \overset{iid}{\sim} \text{Bern}(\theta)$ — iid: (mutually) independent and identically distributed

implicitly talking ⟶ $X_1, \dots, X_n \overset{iid}{\sim}$

let $X = \sum_{i=1}^{n} X_i$ then we have $X \sim \text{Bin}(n, \theta)$

"binomial with parameter n & θ"

$\Omega_X = \{0, \dots, n\}$

pmf: $p(x; \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$ for $x \in \Omega_X$

$\binom{n}{x}$: # of ways to chose x elements out of n — "n chose x"

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$

$\theta^{\sum_i x_i} (1-\theta)^{\sum_i (1-x_i)}$, where $\sum_i (1-x_i) = n - \sum_i x_i$ and $\sum_i x_i = x$

$= \prod_{i=1}^{n} \text{Bern}(x_i; \theta) = p(x_1, \dots, x_n)$, where $\text{Bern}(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$

mean: $X = \sum_i X_i$

$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n\theta$

similarly, $\text{Var}(\sum_i X_i) \overset{\text{indep.}}{=} \sum_i \text{Var}(X_i) = n\theta(1-\theta)$

other distributions:

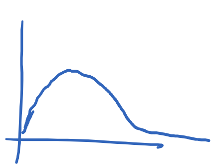
Poisson(λ) (λ: mean) $\Omega_X = \{0, 1, 2, \dots\} = \mathbb{N}$ [count data]

Gaussian in 1D $\mathcal{N}(\mu, \sigma^2)$ (μ: mean, σ^2: variance) $\Omega_X = \mathbb{R}$

gamma $\Gamma(\alpha, \beta)$ $\Omega_X = \mathbb{R}^+$

α: shape; β: inverse scale aka rate

mean $\frac{\alpha}{\beta}$ variance $\frac{\alpha}{\beta^2}$

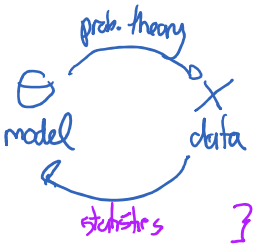


scale mean $\frac{\alpha}{\beta}$ variance $\frac{\alpha}{\beta^2}$

other: Laplace, Cauchy, exponential ↳ $\Gamma(1,\beta)$, beta, Dirichlet ↳ Dirichlet on 2 elements $\Omega_X = [0,1]$

Statiscal concepts

cartoon



} well-defined

} ill defined "inverse problem"

epistemology of science?

example: model n inkep coin flips

prob. theory → prob. k heads in a row

statistics: I have observed k heads, n-k tails, what is θ?

frequentist vs. Bayesian

Semantic of prob.: meaning of a prob?

a) (traditional) frequentist semantic

$P\{X=x\}$ represents the limiting frequency of observing $X=x$ if I could repeat ∞ # of iid experiments

b) Bayesian (subjective) semantic

$P\{X=x\}$ encodes an agent "belief" that $X=x$

laws of prob. characterizes a "rational" way to combine "beliefs" and "evidence" [observations]

[→ has motivation in terms of gambling, utility/decision theory, etc...]

operationally:

Bayesian approach: (*) very simple philosophically:

- treat **all** uncertain quantities as **R.V.**

i.e. encode all knowledge about the system ("beliefs") as a "prior" on probabilistic models and then use law of prob (and Bayes rule) to get updated beliefs and answers.

justification for frequentist semantic:

- for discrete R.V. X, suppose $P\{X=x\} = \theta$
  $\Rightarrow P\{X \neq x\} = 1-\theta$

$B \triangleq \mathbb{1}_{\{X=x\}} \Rightarrow B \sim Bern(\theta)$ R.V.

indicator fct. $\mathbb{1}_A(u) = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{o.w.} \end{cases}$

repeat iid experiments i.e. $B_i \overset{iid}{\sim} Bern(\theta)$

by L.L.N. (law of large numbers) $\frac{1}{n}\sum_{i=1}^{n} B_i \xrightarrow{a.s.} \mathbb{E}[B_i] = \theta$ (limiting frequency)

by CLT $\sqrt{n}\left(\frac{1}{n}Bin(n,\theta) - \theta\right) \xrightarrow{d} N(0, \theta(1-\theta))$

Coin flips – Bayesian approach

biased coin flips

we believe $X \sim Bin(n, \theta)$ — unknown $\Rightarrow$ model it as a R.V.

$\Rightarrow$ need a $p(\theta)$ "prior distribution"

$\Omega_\Theta = [0,1]$

Suppose we observe $X=x$ (result of n flips)

then we can "update" our belief about Θ using **Bayes rule**

$$\underbrace{p(\Theta=\theta \mid X=x)}_{\text{posterior belief}} = \frac{\overbrace{p(X=x \mid \theta)}^{\text{observation model}}\ \overbrace{p(\Theta=\theta)}^{\text{prior belief}}}{\underbrace{p(x)}_{\text{normalization "marginal likelihood"}}}$$

Likelihood

[note: $p(x|\theta) \to$ pmf $p(\theta) \to$ pdf $p(x,\theta)$ is a "mixed distribution"]

Example:

Suppose $p(\theta)$ is uniform on $[0,1]$ "no specific preference"

$$p(\theta|x) \propto \underbrace{\theta^x(1-\theta)^{n-x}}_{p(x|\theta) \text{ up to a constant}} \underbrace{\mathbb{1}_{[0,1]}(\theta)}_{p(\theta)}$$

("proportional to")

Scaling: $\int_0^1 \theta^x(1-\theta)^{n-x}\,d\theta = B(x+1,(n-x)+1)$

normalization constant $\int_\theta p(\theta|x)\,d\theta = 1$

$B(a,b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ (Beta fct.) (gamma fct.)

$\Gamma(a) = \int_0^\infty u^{a-1}e^{-u}\,du$

here $p(\theta|x)$ is called a "beta distribution"

$$B(\theta|\alpha,\beta) \triangleq \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{1}_{[0,1]}(\theta)}{B(\alpha,\beta)}$$

(α,β: parameters)

- uniform distribution: $B(\theta|1,1)$
- posterior $B(\theta|x+1, n-x+1)$

exercise to the reader: if use $B(\alpha_0,\beta_0)$ as prior

posterior will be $B(x+\alpha_0, n-x+\beta_0)$

(*) posterior $p(\theta|\underbrace{X=x}_{\text{observation}})$ contains all the info from data x that we need to answer new queries about θ

e.g. Question: what is prob. of head ($F=1$) on the next flip (flip outcome)

as a frequentist $P(F=1|\text{data}) = \hat{\theta}$ ← notation to mean "estimate"

as a Bayesian $p(F=1|X=x) = \int_\theta p(F=1,\theta|X=x)\,d\theta$ (product rule)

$$= \int_{\Theta} \underbrace{p(F=1 \mid \theta, X=x)}_{=\theta \text{ (by our model)}} \underbrace{p(\theta \mid X=x)}_{\text{posterior}} d\theta \quad \text{(product rule)}$$

$$= \int_{\Theta} \theta\, p(\theta \mid X=x)\, d\theta = \mathbb{E}[\theta \mid X=x]$$

"posterior mean of θ"

* a meaningful "Bayesian" estimator of θ

$$\hat{\theta}_{Bayes}(x) \triangleq \mathbb{E}[\theta \mid X=x] \quad \text{(posterior mean)}$$

notation: $\hat{\theta}$: observation $\rightarrow \Theta$

our coin example : $p(\theta \mid x) = \text{Beta}(\theta \mid \alpha = x+1, \beta = n-x+1)$

mean of a beta R.V. $\frac{\alpha}{\alpha+\beta}$

thus $\boxed{\hat{\theta}_{Bayes}(x) = \mathbb{E}[\theta \mid x] = \frac{x+1}{n+2}}$

here, biased estimator $\mathbb{E}_X[\hat{\theta}(X)] \neq \theta$

$$= \mathbb{E}\left[\frac{X+1}{n+2}\right] = \frac{\mathbb{E}X+1}{n+2} = \frac{n\theta+1}{n+2}$$

but asymptotically unbiased $\xrightarrow{n \to \infty} \theta$

compare { contrast with $\hat{\theta}_{MLE}(x) = \frac{x}{n}$ [unbiased $\mathbb{E}\left[\frac{X}{n}\right] = \frac{n\theta}{n} = \theta$]

to summarize :

- as a Bayesian: get a posterior + use laws of probabilities
- in "frequentist statistics"
  consider multiple estimators
  - MLE
  - moment matching
  - Bayesian posterior mean
  - MAP
  - regularized MLE

  and then analyze their statistical properties

- biased? ?
- Variance ?
- consistent ?

Maximum likelihood principle

setup: given a parametric family $p(x;\theta)$ for $\theta \in \Theta$

we want to estimate/learn θ from x