

Lecture 3 - scribbles

Tuesday, September 11, 2018 14:30

today: • finish prob. review
• parametric models

Bayes rule:

→ invert the conditioning

by definition: $p(y|x) = \frac{p(x,y)}{p(x)}$

Bayes rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

⇒ $p(x,y) = p(y|x)p(x)$
"product rule"

chain rule → successive application of product rule

$$p(x_1, \dots, x_n) = p(x_{1:n-1}) p(x_n | x_{1:n-1})$$

= ...

$$\rightarrow = \prod_{i=1}^n p(x_i | x_{1:i-1})$$

always true:

can be simplified using conditional indep. assumptions
(^{from} graphical model)

conditional independence:

X is cond. indep of Y given Z notation: $X \perp\!\!\!\perp Y \mid Z$

$$\Leftrightarrow p(x, y \mid z) = p(x \mid z) p(y \mid z) \quad \forall x \in \Omega_x$$

$$\forall y \in \Omega_y$$

$$\forall z \in \Omega_z$$

$$\text{st. } p(z) \neq 0$$

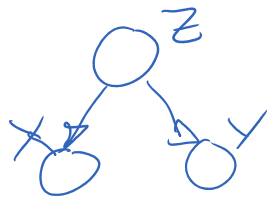
example: Z = indicator whether mother carries a genetic disease

X = son 1 has disease
 Y = son 2 " "

$$X \perp\!\!\!\perp Y \mid Z$$

here, X is not "marginally indep." of Y (ie. $X \not\perp\!\!\!\perp Y$)
 in this phenomenon

$$\exists x, y \text{ st. } p(x, y) \neq p(x)p(y)$$



exercise to the reader: prove that if $X \perp\!\!\!\perp Y \mid Z$

$$\text{then } p(x \mid y, z) = p(x \mid z)$$

$$\left[\text{conditional analog to: } X \perp\!\!\!\perp Y \Rightarrow p(x \mid y) = p(x) \right]$$

example of pairwise independent R.V. that are not mutually independent:

X = coin flip as $\{0, 1\}$
 Y = coin flip
 Z = coin flip

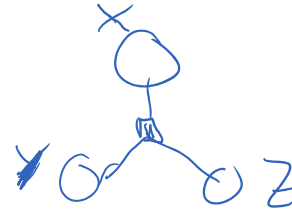


$X = \text{coin flip as } \{0,1\}$
 $Y = \text{coin flip}$
 and $X \perp\!\!\!\perp Y$

define $Z \triangleq X \text{ xor } Y$

we still have $X \perp\!\!\!\perp Z$
 $Y \perp\!\!\!\perp Z$

$Z \not\perp\!\!\!\perp (X, Y)$



other notation:

expectation/mean of a R.V.

$\mathbb{E}[X] \triangleq \sum_{x \in \Omega_X} x p(x)$ (for discrete R.V.)

"expectation of R.V. X"

$\int_{\Omega} x p(x) dx$ ("cts. R.V.")

$\mathbb{E}[\cdot]$ is a linear operator $\Rightarrow \mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$
 (where a is a scalar and X, Y are R.V.)

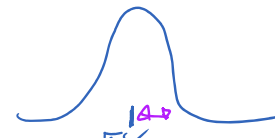
(note: $\mathbb{E}[X]$ does not always exist \rightarrow e.g. Cauchy distribution has integral diverging)

variance:

$$\text{Var}[X] \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2]$$

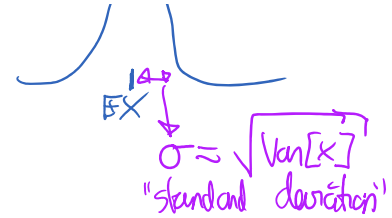
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

"measure of dispersion"



$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}X)^2] \\ = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

← "measure of dispersion"



parametric models:

→ family of distributions

set of valid parameters

$$\mathcal{P}_{\Theta} = \{ p(\cdot; \theta) \mid \theta \in \Theta \}$$

possible pmf/pdf depending on parameter θ

from context: "support of distribution" i.e. Ω_X is usually fixed for all θ

examples: $\Omega_X = \{0, 1\}$ for a coin flip

$\Omega_X = [0, +\infty[$ for Gamma distribution

$[a, b]$
 $[a, b[$

open interval in French !!

abuse of notation $\ni \{ p(x; \theta) \mid \theta \in \Theta \}$

"better" notation $\ni p(X; \theta) \mid \theta \in \Theta$

notation: $X \sim \text{Bern}(\theta)$

↑ "R.V. X is distributed as a Bern(θ) distribution"

$$p(x; \theta) = \text{Bern}(x; \theta)$$

↑ variable for pmf

↑ parameter of dist.

another example:

$$p(x; (\mu, \sigma^2)) = N(x \mid \mu, \sigma^2) \\ \text{or} \\ N(x; \mu, \sigma^2)$$

Bernoulli: prob. of a coin flip $\Omega_X = \{0, 1\}$

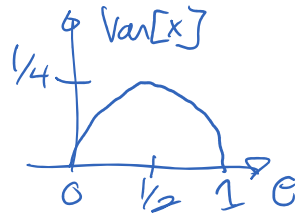
$$P\{X=1 \mid \theta\} = \theta \Rightarrow \Theta = [0, 1] \\ P\{X=0\} \text{ or } 1 - \theta$$

$$Y \sim \text{Bernoulli}(\theta) = \{0, 1\} \quad Y \sim \text{Bernoulli}(\theta)$$

$$P\{X=1\} = \theta$$

$$E[X] = \theta$$

$$\text{Var}[X] = \theta(1-\theta)$$



15h26

binomial distribution: models n indep. coin flips

sum of n indep. $\text{Bern}(\theta)$ R.V.

often denoted p

Let $X_i \sim \text{Bern}(\theta)$ independent & identically distributed

implicitly talking about X_1, \dots, X_n

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

"binomial with parameter n, θ "

$$\Omega_X = \{0, 1, \dots, n\}$$

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

of ways to get x successes out of n trials

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

71 of ways to get x successes out of n trials

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

$$\prod_{i=1}^n \epsilon^{x_i} (1-\epsilon)^{n-x_i}$$

$$= \prod_{i=1}^n \text{Bern}(x_i; \epsilon)$$

mean: $X = \sum_{i=1}^n x_i$

$$E[X] = \sum_{i=1}^n E[x_i] = n\epsilon$$

similarly $\text{Var}[X] = n\epsilon(1-\epsilon)$

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

mean

mean variance

Gaussian in 1D $N(\mu, \sigma^2)$ $\Omega_X = \mathbb{R}$

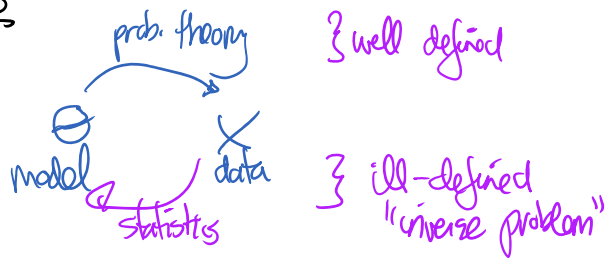
gamma $\Gamma(\alpha, \beta)$ $\Omega_X = \mathbb{R}^+$

shape scale

Other: Laplace, Cauchy, exponential, beta, Dirichlet, etc.

Statistical concepts

cartoon



epistemology of science?

Statistics - "inverse problem"

example: model n indep coin-flips

prob theory \rightarrow prob k heads in a row

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

frequentist vs Bayesian:

philosophy: meaning of a probability?

a) (traditional) frequentist semantic

$P\{X=x\}$ represents the limiting frequency of observing $X=x$

if I could repeat ∞ # of i.i.d. 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]

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

operationally:

(frequentist justification)

• for a discrete R.V., suppose that $P\{X=x\} = \theta$

$$\Rightarrow P\{X \neq x\} = 1 - \theta$$

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

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

repeat i.i.d. experiments i.e. $B_i \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(\theta)$

by L.L.N.
(law of large numbers) $\frac{1}{n} \sum_i B_i \xrightarrow{\text{a.s.}} E[B_1] = \theta$

\uparrow limiting frequency

[aside: $\sum_{i=1}^n B_i \sim \text{Bin}(n, \theta)$ $\frac{1}{n} E[\sum_i B_i] = \frac{n\theta}{n} = \theta$]

$$\text{Var}\left[\frac{1}{n} \sum_i B_i\right] = \frac{1}{n^2} \text{Var}[\text{Bin}(n, \theta)] = \frac{1}{n^2} n\theta(1-\theta) = \frac{1}{n} \theta(1-\theta)$$

$$\text{Var}[ax] = a^2 \text{Var}[x]$$

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

CLT
Central Limit Theorem

\rightarrow all works

Bayesian approach:

⊛ very simple philosophically: treat all uncertain quantities as R.V.

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

simplest example: biased coin

we believe $X \sim \text{Bin}(n, \epsilon)$

↑ unknown ⇒ model it as a R.V.

⇒ we need a $p(\epsilon)$ "prior distribution"

$$\Omega_{\epsilon} = [0, 1]$$

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

then we can "update" our belief about ϵ using Bayes rule

$$\underbrace{p(\epsilon | X=x)}_{\text{posterior belief}} = \underbrace{p(X=x | \epsilon)}_{\substack{\text{observation model} \\ \text{likelihood}}} \underbrace{p(\epsilon)}_{\substack{\text{prior belief} \\ \text{normalization} \\ \text{"marginal likelihood"}}$$

[note: $p(x|\epsilon)$ → is pmf $p(x, \epsilon) \propto$ is a "mixed" dist.]
 $p(\epsilon)$ → pdf

example:

→ ie. $p(\epsilon) = \begin{cases} 1 \\ 0 \end{cases} \int_{0.5}^1 \delta_{\epsilon} \text{ for } \epsilon \in [0.5, 1]$ ie. $p(\epsilon) = \mathbb{1}_{[0.5, 1]}(\epsilon)$

Example: suppose $p(\epsilon)$ is uniform on $[0,1]$ "no specific preference"
 i.e. $p(\epsilon) = \sum_0^1 \delta_{\epsilon, \omega}$ for $\omega \in [0,1]$ i.e. $p(\epsilon) = \mathbb{1}_{[0,1]}(\epsilon)$

$$p(\epsilon|x) \propto \underbrace{\theta^x (1-\theta)^{n-x}}_{p(x|\theta) \text{ up to scaling}} \underbrace{\mathbb{1}_{[0,1]}(\theta)}_{p(\theta)} \quad (x \in \{0, 1, \dots, n\})$$

"proportional" to

scaling: $\int_0^1 \theta^x (1-\theta)^{n-x} d\theta = B(x+1, n-x+1)$
 normalization constant
 so that $\int p(\theta|x) d\theta = 1$

$$B(a,b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad \Gamma(a) \triangleq \int_0^\infty u^{a-1} e^{-u} du$$

Beta function

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

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

parameters support

$$\mathbb{1}_A: \mathbb{R} \rightarrow \mathbb{R}$$

set s.t. $\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{o.w.} \end{cases}$