IFT6269: Probabilistic Graphical Models

Michael Noukhovitch

Fall 2017, Universite de Montreal

Notes written from Simone Lacoste-Julien's lectures.

Contents

1 Probability Review

we need probability to model **uncertainty**

- intrinsic (quantum mechanics)
- incomplete information (dice)
- incomplete modelling ("most birds can fly")

1.1 Notation

sample space Ω

realization $x_1 \in \Omega$

random variable a measurable mapping $X : \Omega \to \mathbb{R}$

indicator function $\mathbb{1}_A(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{else} \end{cases}$

probability distribution a mapping $P : 2^\Omega \to [0, 1]$

set of events $E =$ set of all subsets of Ω

event $\{X = x_1\}$ represents both

- the event $\{x\} \in \Omega_x$
- the event $\{w \in \Omega : X(w) = x_1\} \in E$

joint distribution a random vector $P_{x,y}\{X = x, Y = y\}$

marginal distribution distribution on the components of the random vector $P\{X = x\} = \sum_{y=\Omega_y} P\{X = x, Y = y\}$

1.2 Kologomorov Axioms

1. $P(E_i) \geq 0, \forall E_i \in E$
2. $P(\Omega) = 1$
3. $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ when E_is are disjoint

1.3 Random Variable Basics

probability mass function $P_X(x) = P\{X = x\}, x \in \Omega_x$

cumulative distribtion function $F_X(x) = P\{X \leq x\}$

- non-decreasing
- $\lim_{x\to-\infty} F_X(x) = 0$
- $\lim_{x\to\infty} F_X(x) = 1$

probability density function function $p(x)$ s.t. $F_X(x) = \int p(x)dx$

discrete variable Ω_x is countable, definted by its PMF

continuous Ω_x is uncountable, defined by its PDF

1.4 Other Probability Review

expectation/mean

$$\begin{aligned} E[X] &= \sum_{x \in \Omega_x} xp(x) \\ &= \int_{\Omega} xp(x)d(x) \end{aligned}$$

variance $Var[X] = E[(X - E[X])^2]$

independence $X \perp\!\!\!\perp Y$ if $p(x, y) = p(x)p(y)$

conditional $P(A|B) = \frac{P(A \cap B)}{P(B)}$

1.5 Rules

bayes rule $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

product rule $P(A, B) = P(A|B)P(B)$

conditional independence $X \perp\!\!\!\perp Y|Z \iff p(x, y|z) = p(x|z)p(y|z)$

2 Parametric Models

2.0.1 Bernoilli

A coin flip with probability $\theta, X \sim Bern(\theta)$

- $p(x = 1|\theta) = \theta$
- $\Theta = [0, 1]$
- $\Omega_x = \{0, 1\}$
- $E[X] = \theta$
- $Var[X] = \theta(1 - \theta)$

2.0.2 Binomial

N independent coin flips, $X \sim Bin(n, \theta)$

- let $X_i \overset{iid}{\sim} Bern(\theta)$, then $X = \sum^n X_i$
- $p(x; \theta) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}$
- $E[X] = n\theta$
- $Var[X] = n\theta(1 - \theta)$

2.0.3 Other Distributions

- Poisson $\Omega_x = \{0, 1, \ldots\} = \mathbb{N}$
- Gaussian $N(\mu, \sigma^2), \Omega_x = \mathbb{R}$
- Gamma $\Gamma(\alpha, \beta), \Omega_x = \mathbb{R}_+$

3 Probability

3.1 Maximum Likelihood Estimator

Maximize $p(x|\theta)$ for binomial where $p(x|\theta) = \binom{n}{x} + x$
We use log likelihood instead, because if $a < b$ then $\log a < \log b$

$$\log \binom{n}{x} + n \log x + (n-k) \log(1-x)$$

so $f'(\theta) = 0, f''(\theta) = 0$ is necessary condition for local max