IFT 6269: Probabilistic Graphical Models **Fall 2018**

Lecture 2 — September 7, 2018

Lecturer: Simon Lacoste-Julien *Scribe: Kanika Madan*

Disclaimer: These notes have only been lightly proofread.

2.1 Probability review

2.1.1 Motivation

Question : Why do we use need to use probability?
Answer : Probability provides a principled framework to model uncertainty.

Question : What are the sources of uncertainty?
Answer : Several sources of uncertainty could be:

1. Intrinsic uncertainty: an uncertainty which is built into the system. For example, in quantum mechanics.
2. Uncertainty due to partial information or partial observations. For example:
   (a) In card games, the information and observations might not be complete.
   (b) In rolling a die, the initial conditions might not be known.
3. Uncertainty due to incomplete modeling of a complex phenomenon. This could also be due to computation issues. For example, a rule such as "most birds can fly" has an advantage of being a simple rule, but it also yields uncertainty.

2.1.2 Notation

A review of some of the commonly used notation:
Random variables will be noted $X_1, X_2, X_3, \dots$ or $X, Y, Z, \dots$ and are usually real-valued.

The realizations of these random variables are often denoted by small letters: $x_1, x_2, x_3, ...$ (or $x, y, z, ...$) and are the values that the respective random variables can take. If X is a random variable, then it represents an uncertain quantity. $X = x$ represents that random variable X takes value x.

Formally

Let us define Ω a sample space of elementary events, $\{\omega_1, \omega_2, \omega_3, \dots\}$[1]. The sample space represents the possible values of a given random variable.

For example, let X be the result of a die throw, then $\Omega = \{1, 2, 3, 4, 5, 6\}$.

There are two types of random variables:

1. Discrete Random Variable: when Ω is countable
2. Continuous Random Variable: when Ω is uncountable

If we assume that Ω is countable, then a random variable X is characterized by a probability mass function pmf, which is defined as:

$$p(x)\ \forall x \in \Omega \triangleq \begin{cases} p(x) \geq 0\ \forall x \\ \Sigma_{x \in \Omega} p(x) = 1 \end{cases}$$

A probability distribution P is a mapping $P : \mathcal{E} \mapsto [0, 1]$ where $\mathcal{E}$ is the set of all subsets of Ω, i.e. the set of events (i.e. 2^Ω, i.e. a σ-field[2]) ; such that:

$$\left.\begin{array}{l} -P(E) \geq 0 \quad \forall E \in \mathcal{E} \\ -P(\Omega) = 1 \\ -P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty}(E_i) \quad \text{when } E_1, E_2, \dots \text{ are disjoint.} \end{array}\right\} \text{Kolmogorov axioms}$$

A continuous random variable is characterized by a probability density function p, pdf such that:

$$\begin{cases} p(x) \geq 0\ \forall x \in \Omega \\ \text{p is integrable and } \displaystyle\int_\Omega p(x) = 1 \end{cases}$$

For continuous random variable, we have: for $\Omega \in \mathbb{R}$: $P([a, b]) = \int_a^b p(x)$

Notation: for a random variable, we write as: $p_X(x)$ to denote random variable X taking value x, and $P_Y(y)$ to denote random variable Y taking value y. Also, $P_Y(y) = P_Y(x)$ when $x = y$

[1]temporarily assumed to be a countable set

[2]the σ-field formalism is necessary when Ω is uncountable, which happens as soon as we consider a continuous random variable.

Recap:

For discrete random variable X, pmf $p(x) \leftrightarrow P\{X = x\} = p(x)$

For continuous random variable X, pdf $p(x) \leftrightarrow P\{X \in x \pm \frac{dx}{2} = p(x)dx$

$$P\{X \in x \pm \frac{a}{2} = \int_{x-\frac{a}{2}}^{x+\frac{a}{2}} p(u)du$$

2.1.3 Other random variable basics

Joint random variable

Let $Z = (X, Y)$ be a joint random variable defined over $\Omega_Z \triangleq \Omega_X \times \Omega_Y$. Then the pmf of Z is the "joint pmf" on X and Y such that $p(x, y) = P\{X = x, Y = y\}$. This is called joint distribution.

For example, in a die roll, let X be the result of the die roll as an even number, and Y be the result of the die roll as an odd number. Then we can represent the elementary events of $\Omega_{(X,Y)}$ as the following table:

	$X = 0$	$X = 1$
$Y = 0$	0	$\frac{1}{2}$
$Y = 1$	$\frac{1}{2}$	0

If X and Y are continuous, then $P\{box\} = \displaystyle\int_{box} p(x, y)dxdy$.

Marginal Distribution

(Defined in the context of joint distribution)

Marginal distribution is a distribution of a component of a random vector.

$$P\{X = x\} = \Sigma_{y \in \Omega_Y} P\{X = x, Y = y\} = \Sigma_{y \in \Omega_Y} P(x, y).$$

This is called marginalizing out Y.

Independence

X is independent of Y $\Leftrightarrow$ p(x, y) = p(x) p(y) $\forall\ (x, y) \in \Omega_X \times \Omega_Y$. Independence is denoted as: $X \perp\!\!\!\perp Y$.

Random variables $x_1, x_2, ...$, are mutually independent $\Leftrightarrow$ p(x$_1, x_2, \ldots, _n$) $= \prod_{i=1}^{n} p(x_i)$, $\forall x_{1:n} \in \times_{i=1}^{n} \Omega_{X_i}$.

2.1.4 Conditioning

For events A and B, suppose that $P(B) \neq 0$. We define the probability of A given B,

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

In terms of sample space, that means we look at the subspace where B happens, and in that space, we look at the subspace where A also happens.

For discrete random variables X and Y, conditional pmf is defined as:

$$p(x|y) \triangleq P(X = x|Y = y) \triangleq \frac{P(X = x, Y = y)}{P(Y = y)}$$

where $P(Y = y) = \sum_x P(X = x, Y = y)$ is a normalization constant and also a marginal pmf, necessary in order to get a real probability distribution. $P(X = x, Y = y)$ is the joint pmf.

For continuous random variables X and Y, conditional pmf is defined as:

$$p(x|y) \triangleq \frac{p(x, y)}{p(y)}$$

where $p(y)$ is the probability density.

It is always true, with the subtle point that $p(x|y)$ is undefined if $p(y) = 0$.