Probability review

Zahin Mohammad

May 2020

1 Elements of probability

$A \cap B$ can be read as A and B (the common area in a ven diagram). $A \cap B = 0$ signifies that there are no common areas between A and B, they are disjoint (mutually exclusive).

Independence is defined as $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$. Therefore disjoint events are not independent.

Sample space Ω is the set of all outcomes in the experiment.

Event space F is a subset of Ω of possible values of the experiment.

Probability measure P is a function that takes event(s) and returns their probability of being true. If all events are disjoint, the probability of the events is the summation of the probability of individual events.

Properties:

- $A \subseteq B \Rightarrow P(A) \leq P(B)$

- $P(A \cap B) \leq min(P(A), P(B))$

- $P(A \cup B) \leq P(A) + P(B)$

- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$ This is conditional probability

- $P(s_1 \cap s_2 ... \cap s_n) = P(s_1)P(s_2|s_1)P(s_3|s_1 \cap s_2)...P(s_n|s_{n-1} \cap s_{n-2}... \cap s_1)$

  This is chain rule

2 Random variables

Discrete random variable: The random variable can only take finite values. If X is a random variable denoting 10 coin tosses, it is discrete. $Val(x)$ denotes all possible events of random variable X. $Val(DieThrow) = 1, 2, ...6$.

Continuous random variable: The random variable can take on infinite values. Cannot find probability of a single point, but instead a range. If X is a random variable denoting the height of a randomly selected tree (can keep dividing the height of the last unit into smaller units).

2.1 CDF: Cumulative distribution function

Basically the sum of all probabilities up until an event. Usually used for continuous random variables.

$$F_X(x) = P(X \leq x) \tag{1}$$

2.2 PMF: Probability Mass function

Probability of a single event. Can only be used for discrete random variables.

$$p_X(x) = P(X = x) \tag{2}$$

2.3 PDF: Probability density functions

Defined only if CDF is differentiable everywhere. PDF is just the derivative of the CDF.

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{3}$$

Properties:

- $f_X(x) \geq 0$

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$ Intuitively this makes sense because $f_X(x) \approx p_X(x)$

2.4 Expectation

Expectation is a weighted average of a random variable. Let $g(X)$ be random variable, that takes input X another random variable.

Discrete:

$$E[g(X)] = \sum_{x \in Val(X)} g(x)p_X(x) \tag{4}$$

Continuous:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \tag{5}$$

Note: $E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$

Properties:

- $E[a] = a$

- $E[aX] = aE[X]$

- $E[X + Y] = E[X] + E[Y]$

2.5 Variance

Variance is how concentrated a distribution is around the mean.

$$Var[X] = E[X - E[X]]^2 \tag{6}$$

$$= E[X^2] - (E[X])^2 \tag{7}$$

Properties:

- $Var[a] = 0$

- $Var[aX] = a^2 Var[X]$

2.6   Common random variables

Discrete:

- Bernouli

- Binomial

- Geometric

- Poisson

Continuous:

- Uniform

- Exponential

- Normal

Check the course notes for their properties.

3   Two random variables

An example is X representing number of heads in a coin flip and Y representing the longest streak of heads.

3.1   Joint and marginal distributions

$$F_{XY}(x, y) = P(X < x, Y < y) \tag{8}$$

The marginal cumulative distribution functions are defined as

$$F_Y(y) = \lim x \to \infty F_{XY}(x, y) \tag{9}$$

$$F_X(x) = \lim y \to \infty F_{XY}(x, y). \tag{10}$$

3.2 Joint and marginal probability mass functions

$$p_{XY}(x, y) = P(X = x, Y = y) \tag{11}$$

The marginal probability mass functions are defined as

$$p_X(x) = \sum_y P_{XY}(x, y) \tag{12}$$

$$p_Y(y) = \sum_x P_{XY}(x, y) \tag{13}$$

3.3 Joint and marginal probability density functions

Let X and Y be continuous random variables and their joint distribution function is differentiable in x and y.

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \tag{14}$$

$$\int \int_{(x,y) \in A} f_{XY}(x, y) dx dy = P((X, Y) \in A) \tag{15}$$

Note: The integral over infinity is 1, even though the joint probability density function is always non-negative and potentially above 1 (think dirac delta).

$$f_X(x) = \int_{-infty}^{\infty} f_{XY}(x, y) dy \tag{16}$$

$$f_Y(y) = \int_{-infty}^{\infty} f_{XY}(x, y) dx \tag{17}$$

3.4 Conditional distributions

What is the probability distribution function of Y when $X = x$.

Discrete Case:

$$p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)} \tag{18}$$

Continuous Case:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{19}$$

3.5 Chain Rule

$$p_{X_1,...X_n}(x_1,...x_n) = p_{X_1}(x_1)p_{X_2|x_1}(x_2|x_1)...p_{X_n|X_1...X_n-1}(x_n|x_1,...x_n) \tag{20}$$

3.6 Bayes Rule

Don't understand the y' in this!

3.7 Independence

- $F_{XY}(x,y) = F_X(x)F_Y(y)$

- Discrete: $p_{XY}(x,y) = p_X(x)p_Y(y)$

- Discrete: $p_{Y|X}(y|x) = p_Y(y)$

- Continuous: $f_{XY}(x,y) = f_X(x)p_Y(y)$

- Continuous: $f_{Y|X}(y|x) = f_Y(y)$

3.8 Expectation and covariance

- $E[g(X,Y)] = \sum_{x \in Val(X)} \sum_{y \in Val(Y)} g(x,y)p_{XY}(x,y)$

- $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dx dy$

- $[X,Y] = E[(XE[X])(YE[Y])]$