Pre-Class 03: Mathematical Background III

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

AI Department, Dongguk University

Basics: probability

Axioms of probability

Axioms for events

1. Ω is an event.
2. For every sequence of events $A_1, A_2, \ldots$, the union $\bigcup_{n=1}^{\infty} A_n$ is an event.
3. For every event A, the complement A^c is an event.

Axioms for probability

1. $\Pr\{\Omega\} = 1$.
2. For every event A, $\Pr\{A\} \geq 0$.
3. The probability of the union of any sequence $A_1, A_2, \ldots$ of disjoint events is given by

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \Pr\{A_n\}$$

Axioms of probability theory (cont'd)

Corollaries

$$
\begin{aligned}
\Pr\{\emptyset\} &= 0 && \\
\Pr\left\{\bigcup_{n=1}^{m} A_n\right\} &= \sum_{n=1}^{m} \Pr\{A_n\} && \text{for } A_1, \ldots, A_m \text{ disjoint} \\
\Pr\{A^c\} &= 1 - \Pr\{A\} && \text{for all } A \\
\Pr\{A\} &\le \Pr\{B\} && \text{for all } A \subseteq B \\
\sum_n \Pr\{A_n\} &\le 1 && \text{for all } A_1, \ldots \text{ disjoint}
\end{aligned}
$$

Conditional probability

Definition For any two events A and B (with $\Pr\{B\} > 0$), the conditional probability of A, conditional on B, is defined by

$$\Pr\{A \mid B\} = \Pr\{A \cap B\}/\Pr\{B\}$$

Definition Two events, A and B, are statistically independent if

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$$

Random variables

- A **random variable** X takes on a defined set of values with different probabilities
  - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability $1/6$
  - For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition A" is a random variable (the percentage will be slightly different every time you poll)
- Roughly, **probability** is how frequently we expect different outcomes to occur if we repeat the experiment over and over

Discrete and continuous random variable

- **Discrete** random variables have a countable number of outcomes
  - Dead/live, dice, counts, etc.
  - Probability mass function (pmf, p.m.f.)
- **Continuous** random variables have an infinite continuum of possible values
  - Blood pressure, weight, real number, etc.
  - Probability density function (pdf, p.d.f.)
- **Probability distribution** $F_X(x) = \Pr(X \leq x)$
- **Probability mass function** $p_X(x) = \Pr(X = x)$
- **Probability density function** $f_X(x) = \frac{d}{dx}F_X(x)$

Conditional probability

Two rv's, say X and Y, are *statistically independent* if

$$F_{XY}(x, y) = F_X(x)F_Y(y) \text{ for each value } x_i \text{ of } X \text{ and } y_i \text{ of } Y$$

Discrete rv's

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$$
$$p_{X|Y}(x_i \mid y_j) = p_X(x_i)$$

Continuous rv's

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$
$$f_{X|Y}(x \mid y) = f_X(x)$$

Expectation

The expected value $\mathrm{E}[X]$ of a random variable X is also called the expectation or the mean and is frequently denoted as $\bar{X}$. Considering nonnegative discrete rv's, the expected value is then given by

$$\mathrm{E}[X] = \sum_x x p_X(x).$$

- Discrete case: $\mathrm{E}(X) = \sum_x x p_X(x)$
- Continuous case: $\mathrm{E}(X) = \int_x x f_X(x) dx$

Variance and standard deviation

The *variance* is denoted by σ_X^2 or $\mathrm{Var}[X]$. It is given by

$$\sigma_X^2 = \mathrm{E}[(X - \bar{X})^2] = \mathrm{E}[X^2] - \bar{X}^2$$

The *standard deviation* σ_X of X is the square root of the variance and provides a measure of dispersion of the rv around the mean. Thus the mean is a rough measure of typical values for the outcome of the rv, and σ_X is a measure of the typical difference between X and $\bar{X}$.

Example: expectation and variance

Table 1: Example: discrete random variable

X	10	11	12	13	14
$p_X(x)$	0.4	0.2	0.2	0.1	0.1

- Expectation

$$\begin{aligned} \mathrm{E}(X) &= \sum_x x p_X(x) && (1) \\ &= 10 \cdot 0.4 + 11 \cdot 0.2 + 12 \cdot 0.2 + 13 \cdot 0.1 + 14 \cdot 0.1 && (2) \\ &= 11.3 && (3) \end{aligned}$$

Example: expectation and variance

Table 2: Example: discrete random variable

X	10	11	12	13	14
$p_X(x)$	0.4	0.2	0.2	0.1	0.1

- Variance

$$
\begin{aligned}
\mathrm{E}\left[(X - \mathrm{E}(X))^2\right] &= \sum_x (x - 11.3)^2 p_X(x) && (4)\\
&= (10 - 11.3)^2 \cdot 0.4 + (11 - 11.3)^2 \cdot 0.2 && (5)\\
&\quad + (12 - 11.3)^2 \cdot 0.2 + (13 - 11.3)^2 \cdot 0.1 && (6)\\
&\quad + (14 - 11.3)^2 \cdot 0.1 && (7)\\
&= 1.81 && (8)
\end{aligned}
$$

Example: expectation and variance

Table 3: Example: discrete random variable

X	10	11	12	13	14
$p_X(x)$	0.4	0.2	0.2	0.1	0.1

- Variance

$$\begin{aligned}
\mathrm{E}\left[X^2\right] - \{\mathrm{E}(X)\}^2 &= \sum_x x^2 p_X(x) - 11.3^2 && (9)\\
&= 10^2 \cdot 0.4 + 11^2 \cdot 0.2 + 12^2 \cdot 0.2 + 13^2 \cdot 0.1 && (10)\\
&\quad + 14^2 \cdot 0.1 - 11.3^2 && (11)\\
&= 40 + 24.2 + 28.8 + 16.9 + 19.6 - 127.69 && (12)\\
&= 1.81 && (13)
\end{aligned}$$