More on Multivariate Probability

Yingzhen Li

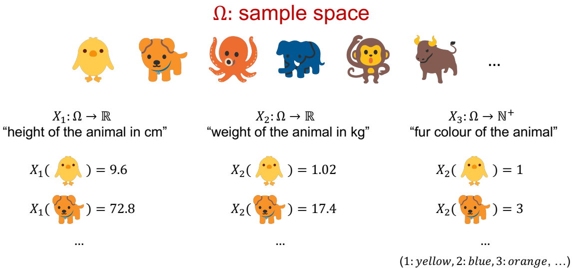
Department of Computing
Imperial College London

@liyzhen2
yingzhen.li@imperial.ac.uk

October 31, 2022

Recap: Multivariate probability

Let's say you are in a zoo that has infinite number of animals:



- Support: $\mathcal{A} = \{(x_1, x_2, x_3) : X_n(\omega) = x_n, \omega \in \Omega\}$
- Set $P(A) := \mathbb{P}(E)$ for the biggest event set $E \subset \Omega$ such that $(X_1, ..., X_N)(E) := \{(X_1(\omega), ..., X_N(\omega)) : \omega \in E\} \subset A$,
- PMF/PDF $p(x_1, x_2, x_3)$ satisfies:

$$\int_{(x_1,x_2,x_3)\in A} p(x_1, x_2, x_3) dx_1 dx_2 dx_3 = P(A), \quad \forall A \subset \mathcal{A}.$$

Conditional probability

Joint probability of events $E_A, E_B \subset \Omega$, $\mathcal{E} = 2^{\Omega}$:

$$\mathbb{P}(E_A \cap E_B)$$

Conditional probability: given that E_A occurs, what is the probability that E_B also occurs?

$$\mathbb{P}(E_B|E_A) := \frac{\mathbb{P}(E_A \cap E_B)}{\mathbb{P}(E_A)}$$

By symmetricity:

$$\mathbb{P}(E_A|E_B) := \frac{\mathbb{P}(E_A \cap E_B)}{\mathbb{P}(E_B)}$$

We can write these expressions in random variable support space: for $A, B \subset \mathcal{A}$,

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

Conditional probability

Conditional probability for $A, B \subset \mathcal{A}$,
$\mathcal{A} = \{(x_1, x_2, ..., x_N) : X_n(\omega) = x_n, \omega \in \Omega\}$,

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

Specifically, we can define $P(X_2 \in A_2 | x_1 \in A_1)$ if we define:

$$A = \{(x_1, x_2, ..., x_N) : X_n(\omega) = x_n, \omega \in \Omega, x_1 \in A_1\}$$

$$B = \{(x_1, x_2, ..., x_N) : X_n(\omega) = x_n, \omega \in \Omega, x_2 \in A_2\}$$

Conditional PMF/PDF $p(X_2 = x_2 | x_1 \in A_1)$ can be defined similar to the joint PMF/PDF case: just need to ensure $\forall A_1 \subset V_{X_1}, A_2 \subset V_{X_2}$

$$\int_{A_2} p(X_2 = x_2 | X_1 \in A_1) dx_2 = P(X_2 \in A_2 | X_1 \in A_1).$$

Conditional probability

Let's say you are in a zoo that has infinite number of animals:

Ω: sample space

...

$X_1 : \Omega \to \mathbb{R}$
"height of the animal in cm"

$X_1($  $) = 9.6$

$X_1($  $) = 72.8$

...

$X_2 : \Omega \to \mathbb{R}$
"weight of the animal in kg"

$X_2($  $) = 1.02$

$X_2($  $) = 17.4$

...

$X_3 : \Omega \to \mathbb{N}^+$
"fur colour of the animal"

$X_2($  $) = 1$

$X_2($  $) = 3$

...

$(1: yellow, 2: blue, 3: orange, \dots)$

Joint probability $P(1.0 \leqslant X_2 \leqslant 10.0, 10.0 \leqslant X_1 \leqslant 50.0)$:

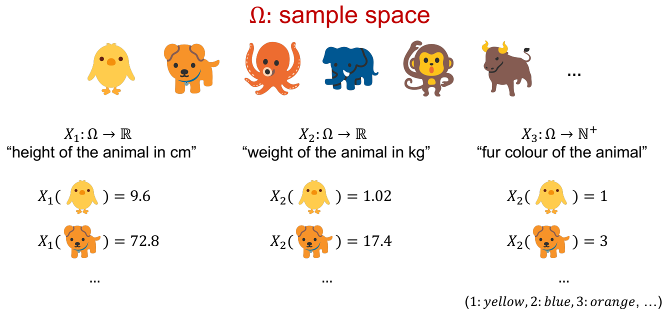
- Figure out the event sets

$$E_1 = \{\omega \in \Omega | 10.0 \leqslant X_1(\omega) \leqslant 50.0\}, \quad E_2 = \{\omega \in \Omega | 1.0 \leqslant X_2(\omega) \leqslant 10.0\}$$

$$\Rightarrow \quad E_1 \cap E_2 = \{\omega \in \Omega | 10.0 \leqslant X_1(\omega) \leqslant 50.0, 1.0 \leqslant X_2(\omega) \leqslant 10.0\}$$

Conditional probability

Let's say you are in a zoo that has infinite number of animals:



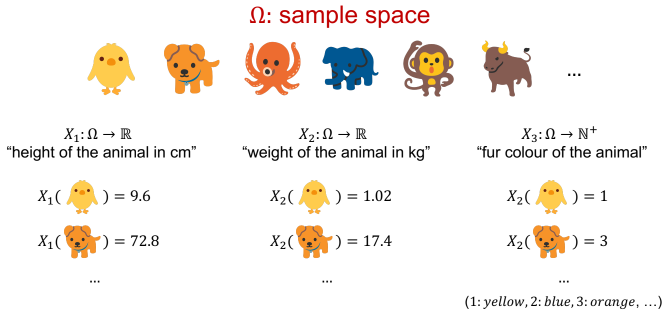
Joint probability $P(1.0 \leqslant X_2 \leqslant 10.0, 10.0 \leqslant X_1 \leqslant 50.0)$:

- Compute $P(1.0 \leqslant X_2 \leqslant 10.0, 10.0 \leqslant X_1 \leqslant 50.0)$ as

$$P(1.0 \leqslant X_2 \leqslant 10.0, 10.0 \leqslant X_1 \leqslant 50.0) = \mathbb{P}(E_1 \cap E_2)$$

Conditional probability

Let's say you are in a zoo that has infinite number of animals:



Marginal probability $P(10.0 \leqslant X_1 \leqslant 50.0)$:

- Figure out the event $E_1 = \{\omega \in \Omega | 10.0 \leqslant X_1(\omega) \leqslant 50.0\}$, then

$$P(10.0 \leqslant X_1 \leqslant 50.0) = \mathbb{P}(E_1)$$

Conditional probability

Let's say you are in a zoo that has infinite number of animals:

Ω: sample space

...

$X_1: \Omega \to \mathbb{R}$
"height of the animal in cm"

$X_1($  $) = 9.6$

$X_1($  $) = 72.8$

...

$X_2: \Omega \to \mathbb{R}$
"weight of the animal in kg"

$X_2($  $) = 1.02$

$X_2($  $) = 17.4$

...

$X_3: \Omega \to \mathbb{N}^+$
"fur colour of the animal"

$X_2($  $) = 1$

$X_2($  $) = 3$

...

$(1: yellow, 2: blue, 3: orange, \ldots)$

Conditional probability $P(1.0 \leqslant X_2 \leqslant 10.0 | 10.0 \leqslant X_1 \leqslant 50.0)$:

- Compute $P(1.0 \leqslant X_2 \leqslant 10.0 | 10.0 \leqslant X_1 \leqslant 50.0)$ as

$$P(1.0 \leqslant X_2 \leqslant 10.0 | 10.0 \leqslant X_1 \leqslant 50.0) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)}$$

Sum rule and product rule

By definition of the **conditional probability**:

$$P(X_2 \in A_2 | X_1 \in A_1) := \frac{P(X_2 \in A_2, X_1 \in A_1)}{P(X_1 \in A_1)}$$

Product rule:

$$P(X_2 \in A_2, X_1 \in A_1) = P(X_2 \in A_2 | X_1 \in A_1) \times P(X_1 \in A_1)$$

"Joint dist. = conditional dist. $\times$ marginal dist."

Sum rule and product rule

By definition of the **conditional probability**:

$$P(X_2 \in A_2 | X_1 \in A_1) := \frac{P(X_2 \in A_2, X_1 \in A_1)}{P(X_1 \in A_1)}$$

Product rule:

$$P(X_2 \in A_2, X_1 \in A_1) = P(X_2 \in A_2 | X_1 \in A_1) \times P(X_1 \in A_1)$$

"Joint dist. = conditional dist. × marginal dist."

Sum rule:

$$P(X_1 \in A_1) = \int_{V_{X_2}} p(X_1 \in A_1, X_2 = x_2) dx_2$$

"Marginal dist. = sum/integral of joint dist."

Sum rule and product rule

By definition of the **conditional probability**:

$$P(X_2 \in A_2 | X_1 \in A_1) := \frac{P(X_2 \in A_2, X_1 \in A_1)}{P(X_1 \in A_1)}$$

Product rule:

$$P(X_2 \in A_2, X_1 \in A_1) = P(X_2 \in A_2 | X_1 \in A_1) \times P(X_1 \in A_1)$$

"Joint dist. = conditional dist. × marginal dist."

Sum rule:

$$P(X_1 \in A_1) = \int_{V_{X_2}} p(X_1 \in A_1, X_2 = x_2) dx_2$$

"Marginal dist. = sum/integral of joint dist."

Combining both:

$$P(X_1 \in A_1) = \int_{V_{X_2}} P(X_2 = x_2 | X_1 \in A_1) \times P(X_1 \in A_1) dx_2$$

Conditional independence

Independence of two random variables X_1, X_2:

$$X_1 \perp\!\!\!\perp X_2 \quad \Leftrightarrow \quad p(X_1, X_2) = p(X_1)p(X_2)$$

Equivalently, using product rule:

$$X_1 \perp\!\!\!\perp X_2 \quad \Leftrightarrow \quad p(X_1|X_2) = p(X_1), p(X_2|X_1) = p(X_2)$$

Conditional independence

Independence of two random variables X_1, X_2:

$$X_1 \perp\!\!\!\perp X_2 \quad \Leftrightarrow \quad p(X_1, X_2) = p(X_1)p(X_2)$$

Equivalently, using product rule:

$$X_1 \perp\!\!\!\perp X_2 \quad \Leftrightarrow \quad p(X_1|X_2) = p(X_1), p(X_2|X_1) = p(X_2)$$

Conditional Independence of two random variables X_1, X_2 given X_3:

$$X_1 \perp\!\!\!\perp X_2|X_3 \quad \Leftrightarrow \quad p(X_1, X_2|X_3) = p(X_1|X_3)p(X_2|X_3)$$

Equivalently, using product rule:

$$X_1 \perp\!\!\!\perp X_2|X_3 \quad \Leftrightarrow \quad p(X_1|X_2, X_3) = p(X_1|X_3), p(X_2|X_1, X_3) = p(X_2|X_3)$$

Conditional independence

Example: drawing 5 cards from a standard 52-card poker deck

Define the following random variables:

- X_1: number of hearts $\heartsuit$
- X_2: number of diamonds $\diamondsuit$
- X_3: number of clubs $\clubsuit$
- X_4: number of spades $\spadesuit$

What is the joint distribution $p(X_1, X_2, X_3, X_4)$ for the card draws –

- with replacement?
- without replacement?

Vector mean & covariance

Univariate case:

$$\text{mean:} \quad \mathbb{E}[X] = \int x p_X(x) dx$$

$$\text{variance:} \mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])] = \int (x - \mathbb{E}[X]) p_X(x) dx$$

Multivariate case: write $X = (X_1, ..., X_N)^\top$

$$\text{mean:} \quad \mathbb{E}[X] = (\mathbb{E}[X_1], ..., \mathbb{E}[X_N])^\top$$

$$\text{covariance:} \quad \mathbb{V}[X] = \Sigma, \quad \Sigma_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$$

Using **sum rule**: only need marginals

$$\mathbb{E}[X_i] = \int x_i p_X(\boldsymbol{x}) d\boldsymbol{x} = \int x_i p_X(X_i = x_i) dx_i$$

$$\Sigma_{ij} = \int (x_i - \mathbb{E}[X_i])(x_j - \mathbb{E}[X_j]) p_X(\boldsymbol{x}) d\boldsymbol{x}$$

$$= \int (x_i - \mathbb{E}[X_i])(x_j - \mathbb{E}[X_j]) p_X(X_i = x_i, X_j = x_j) dx_i dx_j$$

Vector mean & covariance

Connecting univariate & multivariate cases:

$$\begin{aligned}\mathbb{E}[\mathbf{a}^\top X] &= \int \mathbf{a}^\top \boldsymbol{x} p_X(\boldsymbol{x}) d\boldsymbol{x} = \int \sum_i a_i x_i p(\boldsymbol{x}) d\boldsymbol{x} = \sum_i a_i \int x_i p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \sum_i a_i \int x_i p(x_i) p(\{x_j\}_{j\neq i} | x_i) d\boldsymbol{x} \\ &= \sum_i a_i \int x_i p(x_i) dx_i \int p(\{x_j\}_{j\neq i} | x_i) d\boldsymbol{x}_{-i} \\ &= \mathbf{a}^\top (\mathbb{E}[X_1], ..., \mathbb{E}[X_N])^\top = \mathbf{a}^\top \mathbb{E}[X]\end{aligned}$$

$\implies$ mean vector is the mean of each marginal!

Vector mean & coariance

Connecting univariate & multivariate cases:
Write $\bar{\boldsymbol{x}} := \mathbb{E}[X]$ and use sum rule

$$
\begin{aligned}
\mathbb{V}[\mathbf{a}^\top X] &= \int (\mathbf{a}^\top \boldsymbol{x} - \mathbf{a}^\top \bar{\boldsymbol{x}})^2 p_X(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \Big(\sum_i a_i x_i - a_i \bar{x}_i\Big)\Big(\sum_j a_j x_j - a_j \bar{x}_j\Big) p_X(\boldsymbol{x}) d\boldsymbol{x} \\
&= \sum_i \sum_j a_i a_j \int (x_i - \bar{x}_i)(x_j - \bar{x}_j) {\color{red} p_X(X_i = x_i, X_j = x_j) dx_i dx_j} \\
&= \sum_i \sum_j a_i a_j \Sigma_{ij} = \mathbf{a}^\top \Sigma \mathbf{a}
\end{aligned}
$$

The covariance Σ allows us to find the scalar variance in any direction.

Conditional expectations

Expectation of a function $f(X)$ under distribution $p(X)$:

$$\mathbb{E}_{p(X)}[f(X)] = \int f(x)p(X = x)dx$$

Conditional expectation of a function $g(Y)$ given $X = x$ under conditional distribution $p(Y|X = x)$:

$$\mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

An equivalent notation: $\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)]$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\mathbb{E}_{p(Y)}[Y] = \mathbb{E}_{p(X)}[\mathbb{E}_{p(Y|X)}[Y]]$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\begin{aligned}\mathbb{E}_{p(Y)}[Y] &= \mathbb{E}_{p(X)}[\mathbb{E}_{p(Y|X)}[Y]] \\ &= \int p(X = x)\mathbb{E}_{p(Y|X=x)}[Y]dx\end{aligned}$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\begin{aligned}\mathbb{E}_{p(Y)}[Y] &= \mathbb{E}_{p(X)}[\mathbb{E}_{p(Y|X)}[Y]] \\ &= \int p(X = x)p(Y = y|X = x)ydydx\end{aligned}$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\begin{aligned}\mathbb{E}_{p(Y)}[Y] &= \mathbb{E}_{p(X)}[\mathbb{E}_{p(Y|X)}[Y]] \\ &= \int p(Y = y, X = x)ydydx \quad \text{(product rule)}\end{aligned}$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:

$$\begin{aligned} \mathbb{E}_{p(Y)}[Y] &= \mathbb{E}_{p(X)}[\mathbb{E}_{p(Y|X)}[Y]] \\ &= \int p(Y = y)ydy \quad \text{(sum rule)} \end{aligned}$$

Conditional expectations

Conditional expectation of a function $g(Y)$ given $X = x$:

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}_{p(Y|X=x)}[g(Y)] = \int g(y)p(Y = y|X = x)dy$$

- $\mathbb{E}[g(Y)|X = x]$ is a function of x
- As X is a random variable, $\mathbb{E}[g(Y)|X]$ is also a random variable

Law of Total Expectation:
Extention to functions of Y:

$$\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y)|X]]$$

Conditional expectations

Conditional variance of Y given $X = x$:

$$\begin{aligned}\mathbb{V}[Y|X=x] &:= \mathbb{V}_{p(Y|X=x)}[Y] \\ &= \mathbb{E}_{p(Y|X=x)}[(Y - \mathbb{E}_{p(Y|X=x)}[Y])^2] \\ &= \mathbb{E}[(Y - \mathbb{E}[Y|X=x])^2|X=x]\end{aligned}$$

- $\mathbb{V}[Y|X = x]$ is a function of x
- As X is a random variable, $\mathbb{V}[Y|X]$ is also a random variable

Law of Total Variance:

$$\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y|X]] + \mathbb{V}[\mathbb{E}[Y|X]]$$

Summary

Topics we've covered about multivariate probability:

- Definitions and some examples
- Joint, marginal, and conditional distributions
- Sum rule and product rule
- Change-of-variables rule
- Computing mean/variance/expectations

Next lecture: Model selection via cross-validation

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a "split" of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^{K} A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

$$\begin{aligned} \Rightarrow \quad \cup_{k=1}^{K}(\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) &= \mathcal{A}(A_1) \cap (\cup_{k=1}^{K} \mathcal{A}(A_2^k)) \\ &= \mathcal{A}(A_1) \cap \mathcal{A}(V_{X_2}) \\ &= \mathcal{A}(A_1) \cap \mathcal{A} = \mathcal{A}(A_1) \end{aligned}$$

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a "split" of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^{K} A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

$$\begin{aligned} \Rightarrow \quad P(\cup_{k=1}^{K}(\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k))) &= P(\mathcal{A}(A_1) \cap (\cup_{k=1}^{K}\mathcal{A}(A_2^k))) \\ &= P(\mathcal{A}(A_1) \cap \mathcal{A}(V_{X_2})) \\ &= P(\mathcal{A}(A_1) \cap \mathcal{A}) = P(\mathcal{A}(A_1)) \end{aligned}$$

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a "split" of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^{K} A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

As $A_2^i \cap A_2^j = \emptyset \quad \Rightarrow \quad (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) \cap (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) = \emptyset$

$$\Rightarrow \quad P(\mathcal{A}(A_1)) = P(\cup_{k=1}^{K}(\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k))) = \sum_{k=1}^{K} P(\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k))$$

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a "split" of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^{K} A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

As $A_2^i \cap A_2^j = \emptyset \quad \Rightarrow \quad (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) \cap (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) = \emptyset$

$$\Rightarrow \quad P(X_1 \in A_1) = \sum_{k=1}^{K} P(X_1 \in A_1, X_2 \in A_2^k)$$

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a "split" of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^K A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

As $A_2^i \cap A_2^j = \emptyset \quad \Rightarrow \quad (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) \cap (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) = \emptyset$

$$\Rightarrow \quad P(X_1 \in A_1) = \sum_{k=1}^K P(X_1 \in A_1, X_2 \in A_2^k)$$

$$= \sum_{k=1}^K \int_{A_2^k} p(X_1 \in A_1, X_2 = x_2) dx_2$$

Appendix: math formula for deriving sum rule

Deriving **sum rule** using $\mathbb{P}$ defined on sets:
Define for any $A_1 \subset V_{X_1}$, $A_2 \subset V_{X_2}$

$$\mathcal{A}(A_1) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_1(\omega) \in A_1\}$$

$$\mathcal{A}(A_2) = \{(X_1(\omega), X_2(\omega), ..., X_N(\omega)) : X_2(\omega) \in A_2\}$$

Now we can define a “split” of X_2 value space V_{X_2}:

$$V_{X_2} = \cup_{k=1}^{K} A_2^k, \quad A_2^k \neq \emptyset, \quad A_2^i \cap A_2^j = \emptyset, \forall i \neq j$$

As $A_2^i \cap A_2^j = \emptyset \quad \Rightarrow \quad (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) \cap (\mathcal{A}(A_1) \cap \mathcal{A}(A_2^k)) = \emptyset$

$$\Rightarrow \quad P(X_1 \in A_1) = \sum_{k=1}^{K} P(X_1 \in A_1, X_2 \in A_2^k)$$

$$= \int_{V_{X_2}} p(X_1 \in A_1, X_2 = x_2) dx_2 \quad \text{(sum rule)}$$