Probability

Basics of probability

Example: Independent Events

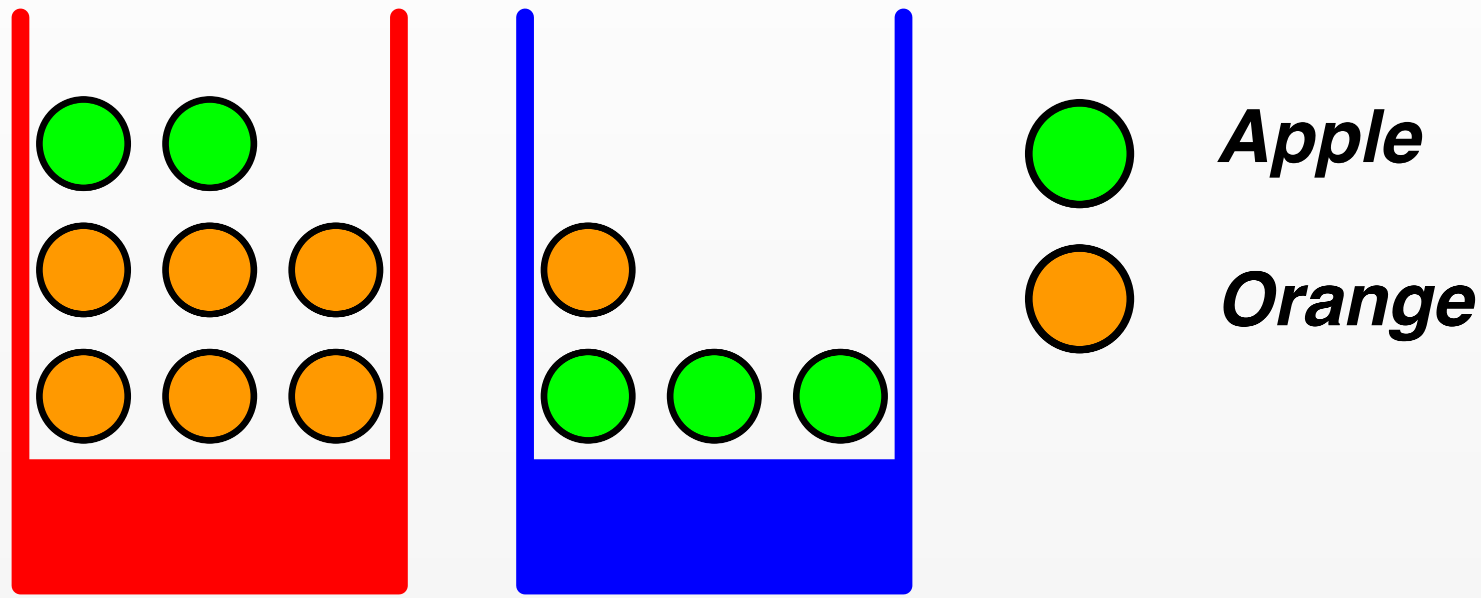
1. What's the probability of getting a sequence of **1,2,3,4,5,6** if we roll a dice six times

   Answer: $(1 / 6)^6$

2. A school survey found that **9** out of **10** students like pizza. If three students are chosen at random, what is the probability that all three students like pizza

   Answer: $(9 / 10)^3$

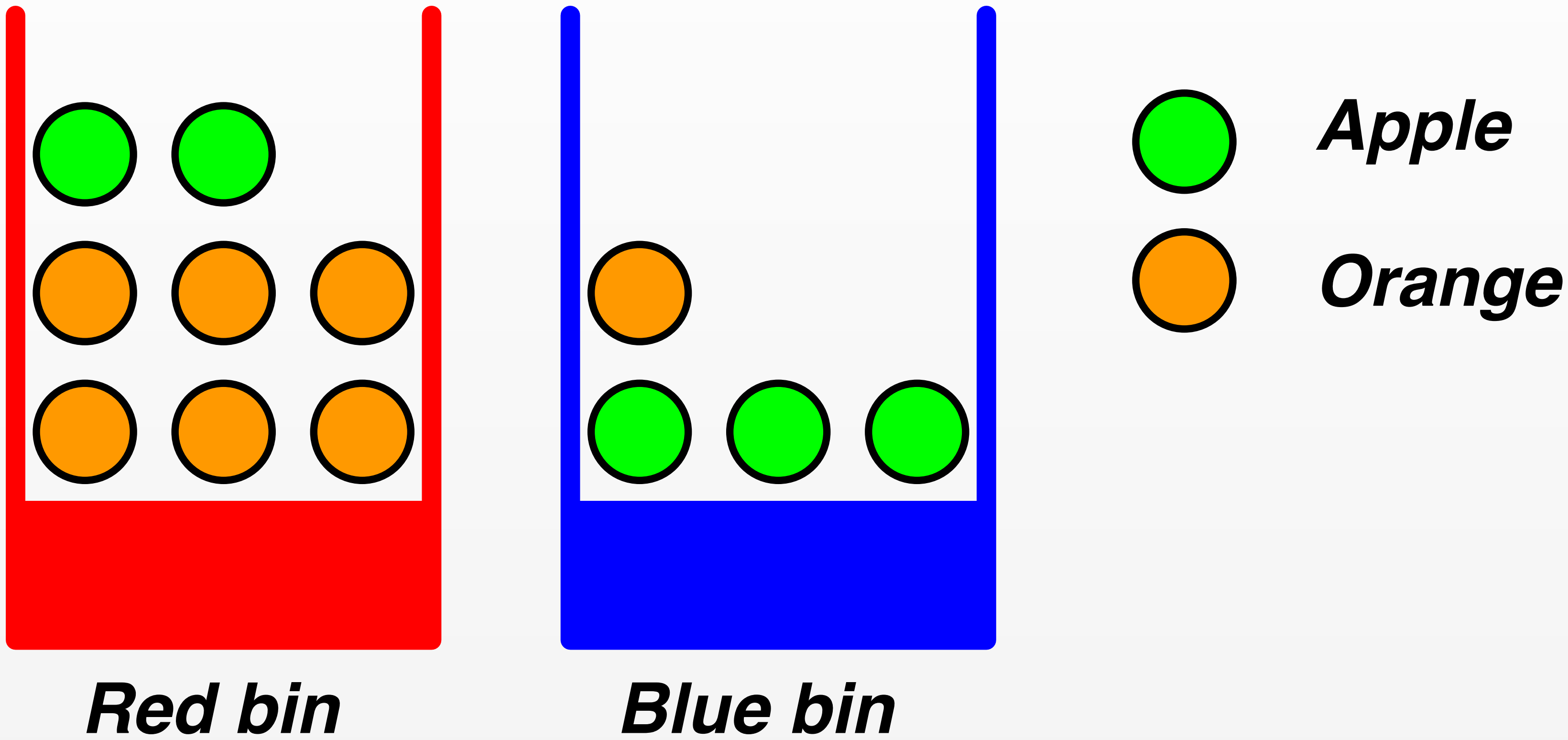
Dependent Events



What is the probability of picking an apple?

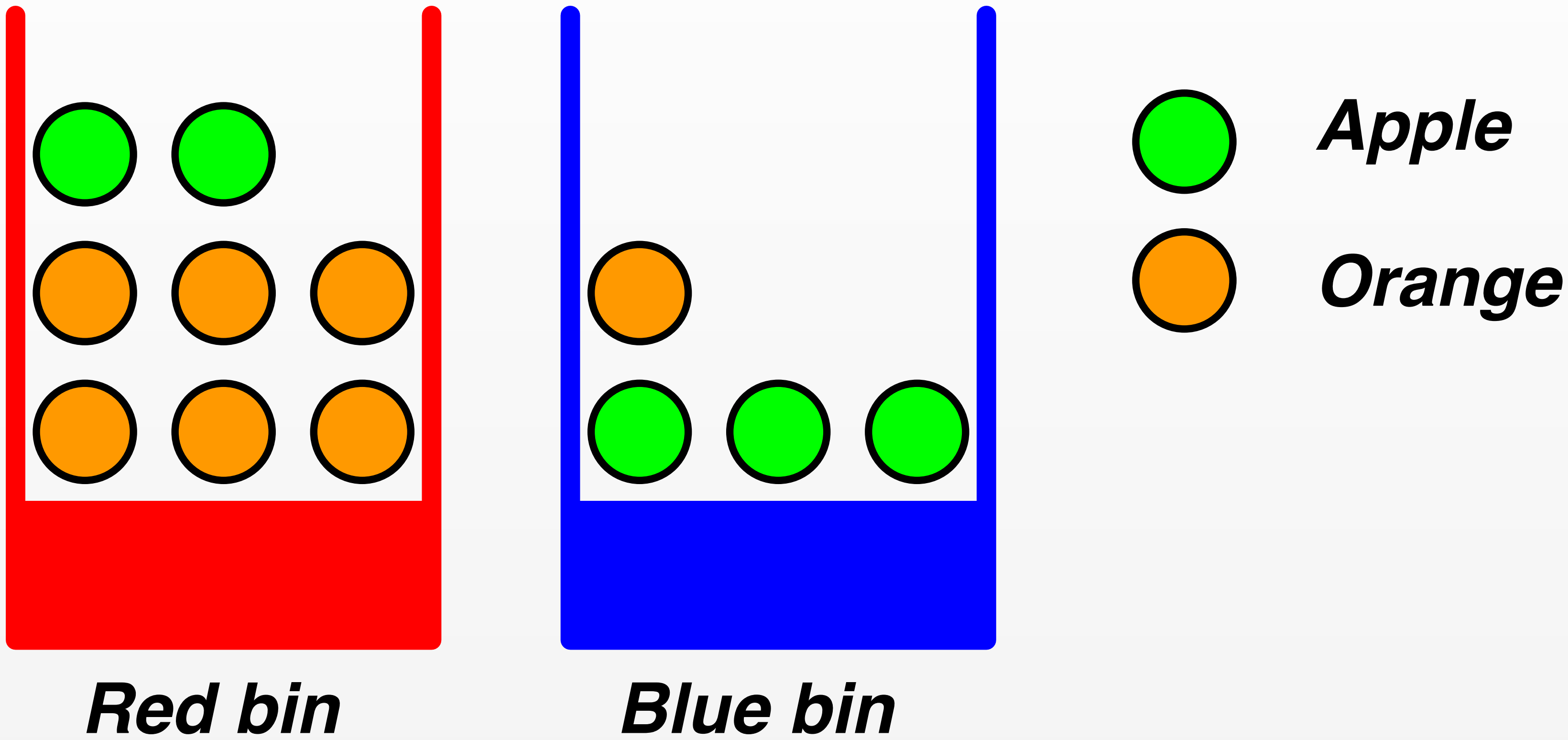
P(fruit = apple)= 5/12

Dependent Events



*If I randomly pick a fruit from the **red** bin,*
*what is the probability that I get an **apple**?*

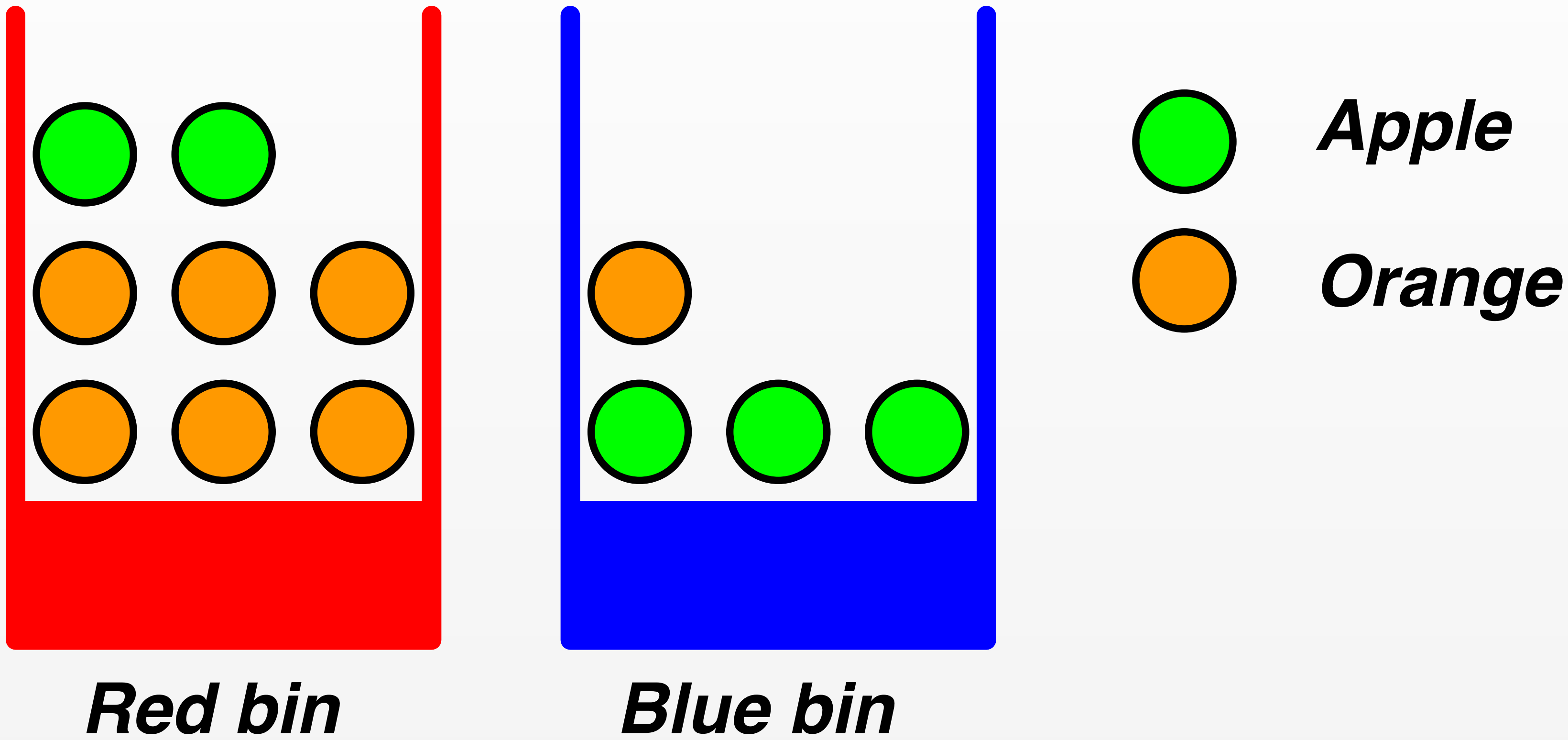
Dependent Events



Conditional Probability

P(fruit = apple | bin = red) = 2 / 8

Dependent Events



Conditional Probability

P(fruit = apple | bin = **blue**) = 3 / 4

Dependent Events

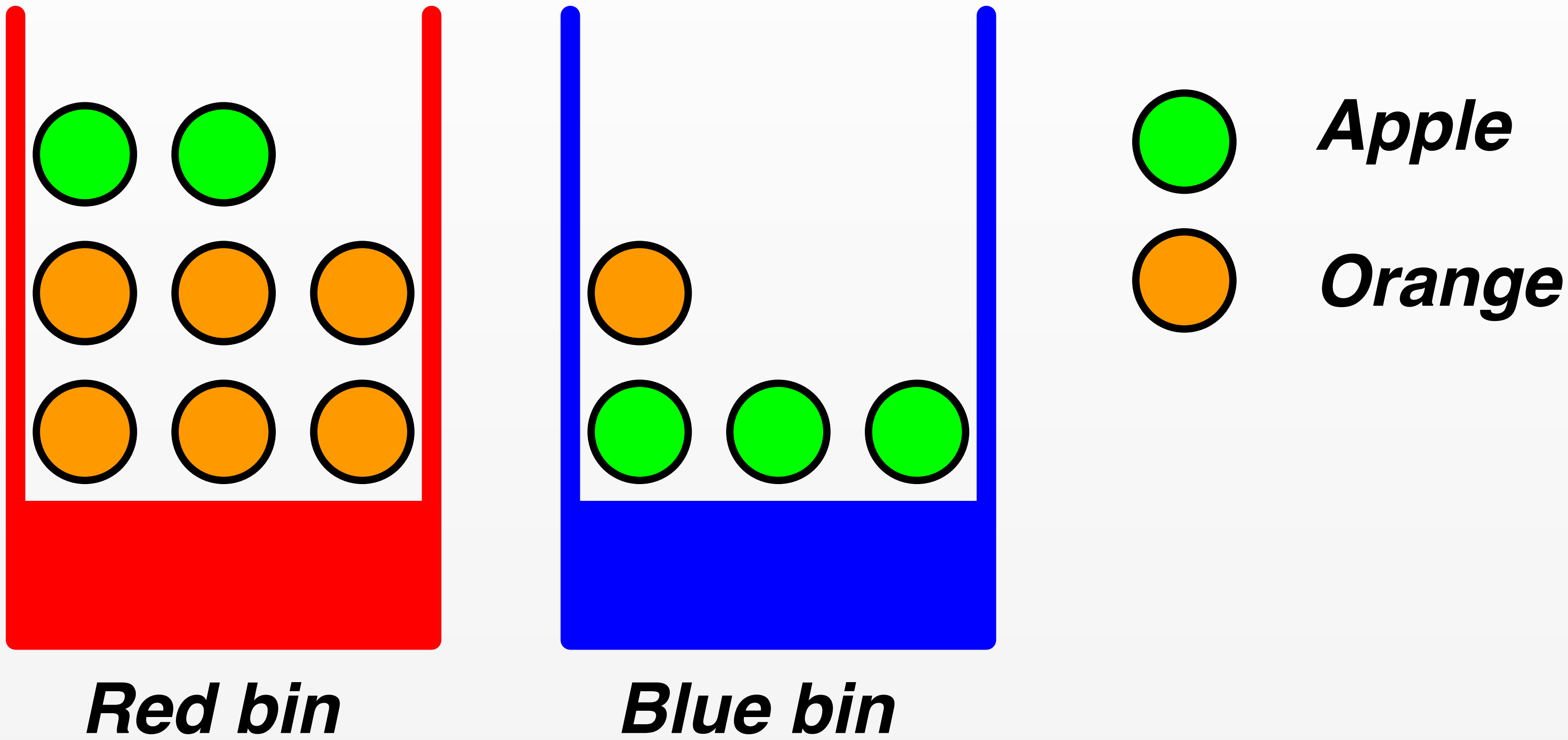
The probability of the event fruit = apple changes if we know which bin was selected.

$$P(\text{fruit} = \text{apple}) = 5/12$$

$$P(\text{fruit} = \text{apple} \mid \text{bin} = \text{red}) = 2 / 8$$

$$P(\text{fruit} = \text{apple} \mid \text{bin} = \text{blue}) = 3 / 4$$

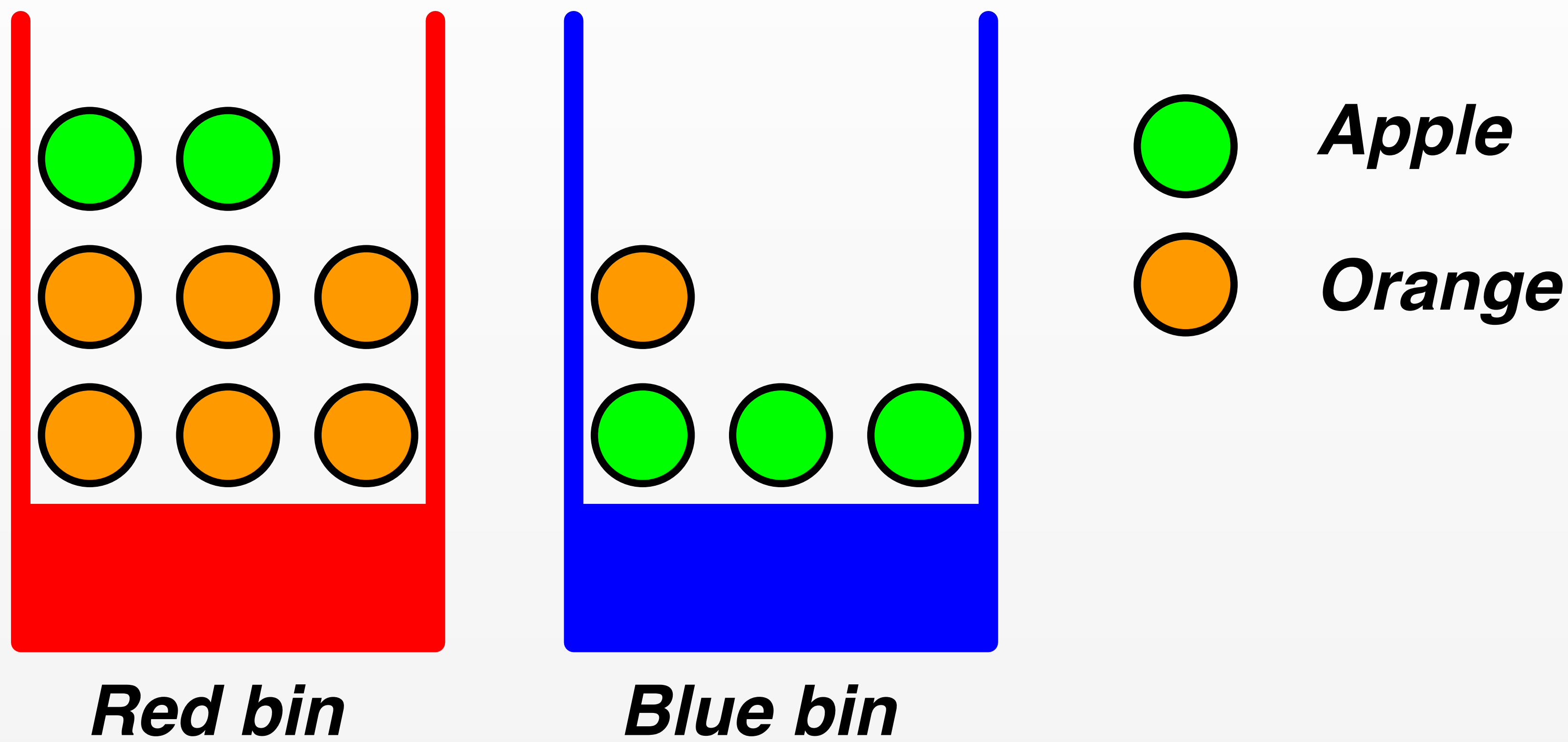
Dependent Events



Joint Probability

P(fruit = apple , bin = red) = 2 / 12

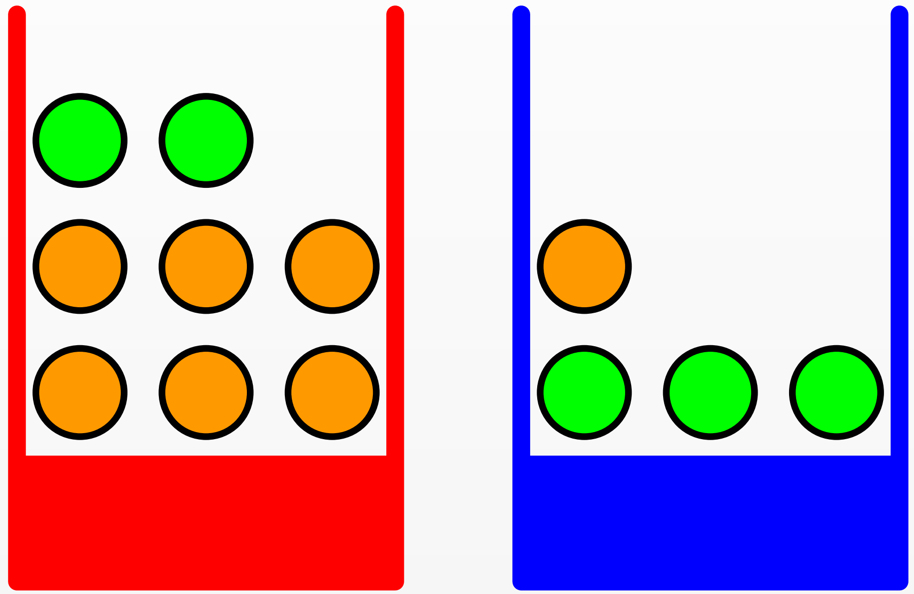
Dependent Events



Joint Probability

P(fruit = apple , bin = blue) = 3 / 12

Two rules of Probability

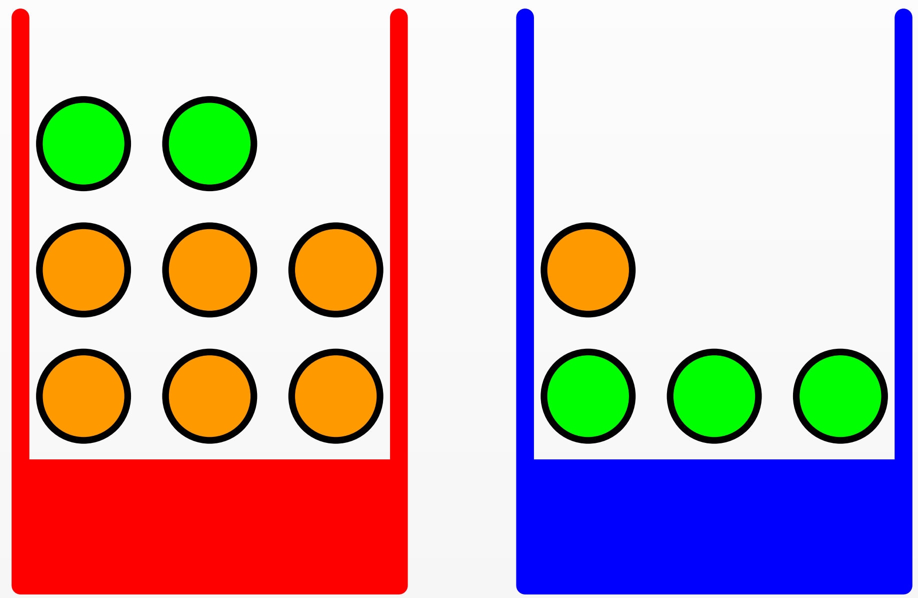


1. Sum Rule (Marginal Probabilities)

$$P(\text{fruit} = \text{apple}) = P(\text{fruit} = \text{apple}, \text{bin} = \text{blue}) + P(\text{fruit} = \text{apple}, \text{bin} = \text{red})$$

$$=$$

Two rules of Probability



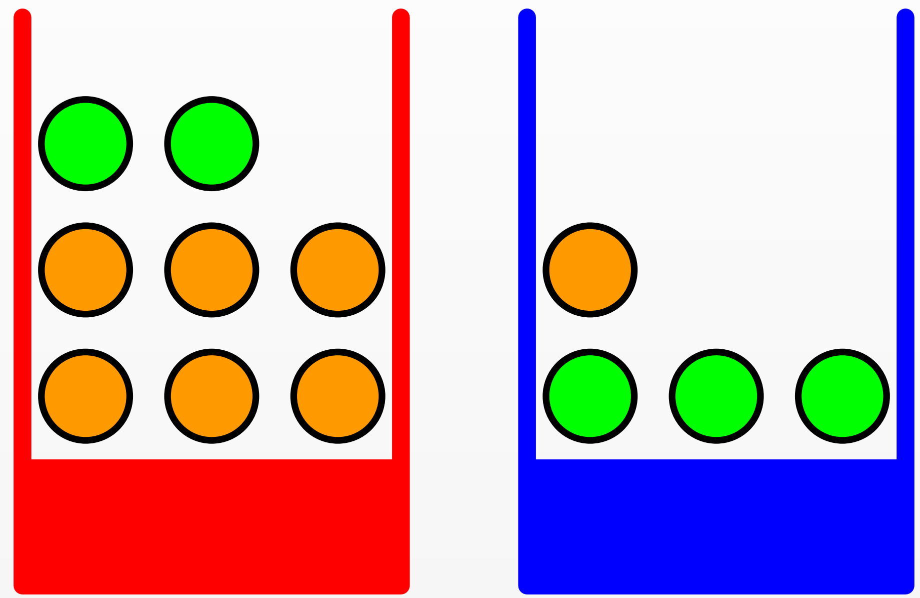
1. Sum Rule (Marginal Probabilities)

P(fruit = apple) = P(fruit = apple , bin = blue)

+ P(fruit = apple , bin = red)

= 3 / 12 + 2 / 12 = 5 / 12

Two rules of Probability



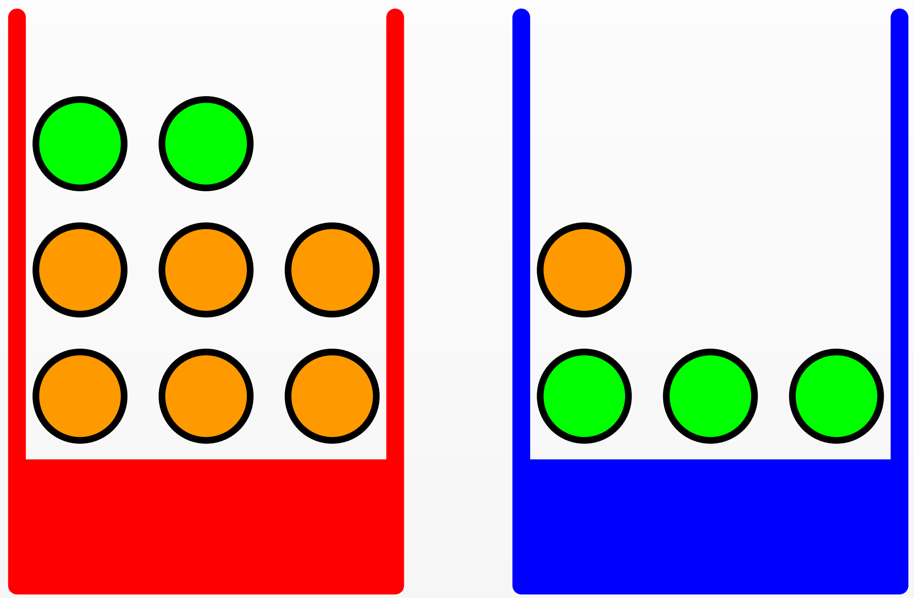
1. Sum Rule (Marginal Probabilities)

P(bin = red) = P(fruit = apple , bin = red)

+ P(fruit = orange , bin = red)

= 2 / 12 + 6 / 12 = 8 / 12

Two rules of Probability



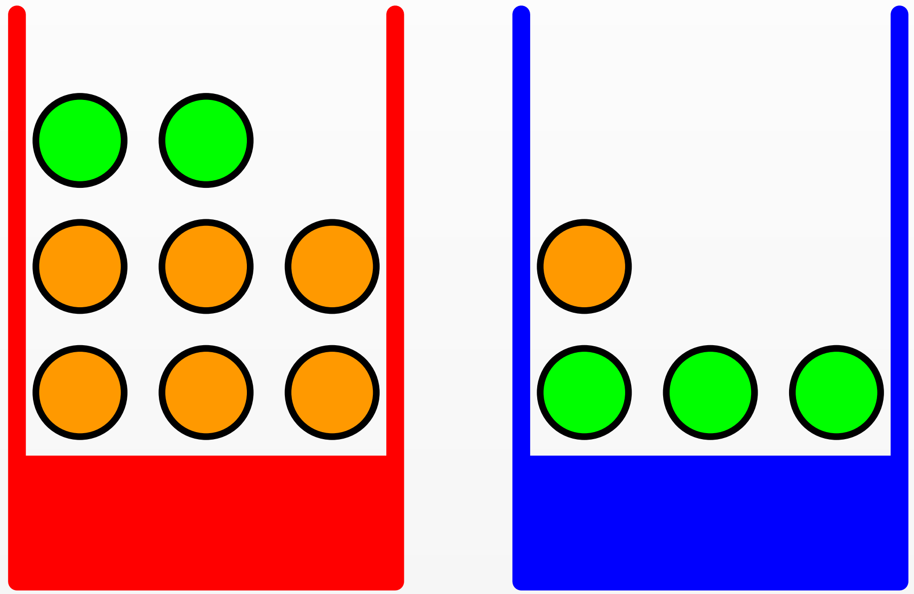
2. Product Rule

$P(\text{fruit} = \text{apple}, \text{bin} = \text{red}) =$

$P(\text{fruit} = \text{apple} \mid \text{bin} = \text{red})\, p(\text{bin} = \text{red})$

$=$

Two rules of Probability



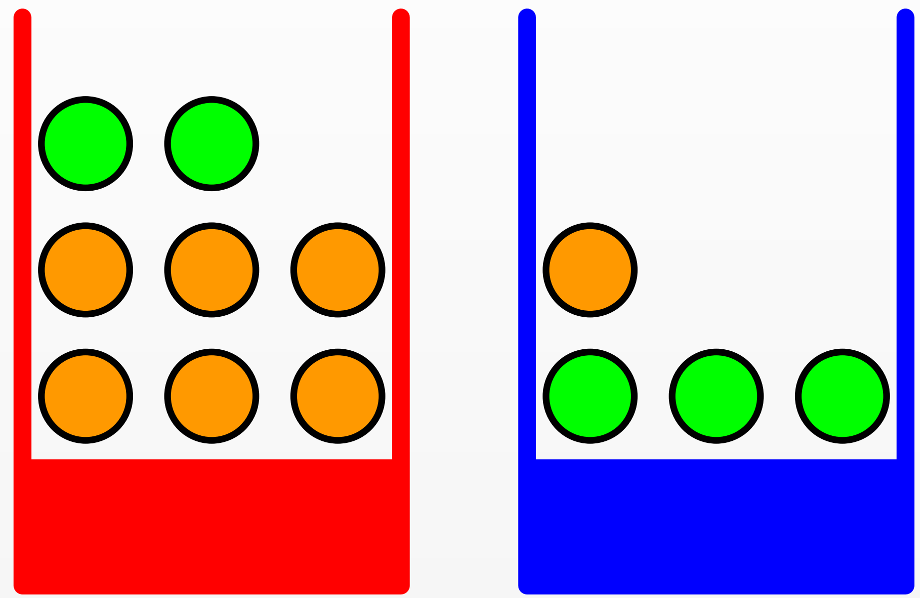
2. Product Rule

P(fruit = apple , bin = red) =

P(fruit = apple | bin = red) p(bin = red)

= 2 / 8 * 8 / 12 = 2 / 12

Two rules of Probability



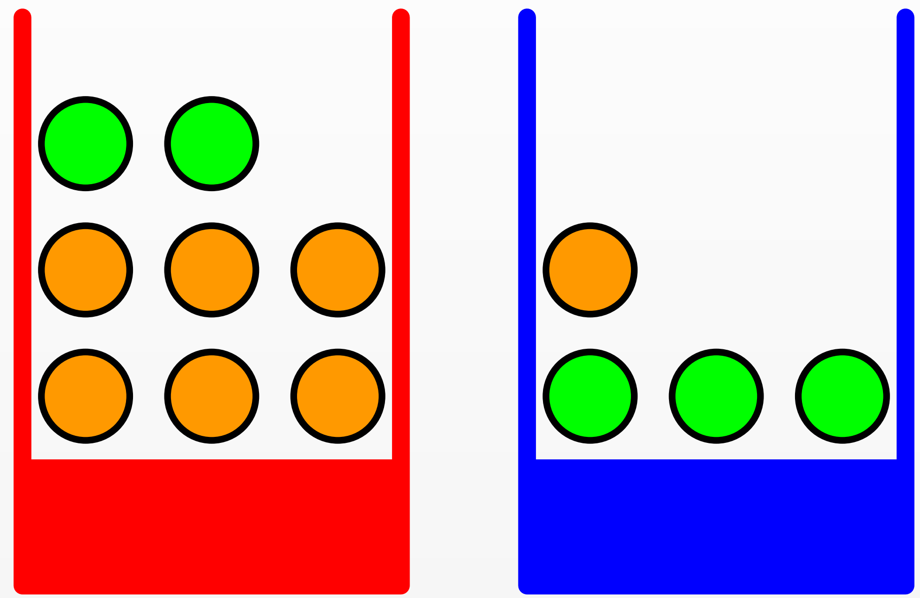
2. Product Rule (reversed)

P(fruit = apple , bin = red) =

P(bin = red | fruit = apple) p(fruit = apple)

=

Two rules of Probability



2. Product Rule (reversed)

P(fruit = apple , bin = red) =

P(bin = red | fruit = apple) p(fruit = apple)

= 2 / 5 * 5 / 12 = 2 / 12

Probability

Random variables Measures and Densities

Probability Measures

Definition: A probability space $(\Omega, \mathcal{F}, P)$ consists of

- A sample space Ω (i.e. the set of *outcomes*)
- A set of events $\mathcal{F}$ (i.e. the set possible sets)
- A probability measure P (maps events to probabilities)

Axioms of Probability

$P : \mathcal{F} \rightarrow [0,1] \qquad P(E) \geq 0 \ \forall E \in \mathcal{F} \qquad P(\Omega) = 1$

$$P(E_1 \cup E_2) = p(E_1) + P(E_2) \ when \ E_1 \cap E_2 = \emptyset$$

Axiom that can be derived from the basic axioms

Complement of E

$$P(E) = 1 - P(\neg E)$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E$$

Joint and Conditional Probability

- **Definition:** Joint Probability

Outcomes in both *A* and *B*

$$P(A,B) = P(A \cap B)$$

Events (i.e. sets of outcomes)

- **Definition:** Conditional Probability

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

The product rule can be derived from this formulae

Random Variables

- *G*ives a language to talk about probability of events.
- F*ormalizes the relationship between two probability spaces.*
- *Allows expressing joint and conditional probabilities of events in different probability spaces.*

Random Variables

- ***Random Variable:*** A variable that takes value in the sample space $\Omega = \{1,2,3,4,5,6\}$

  $X \in \Omega$

  - ***Event:*** A set of *outcomes*

    $X >= 3$ $\{3, 4, 5, 6\}$

    $X = 5$ $\{5\}$

- ***Probability:*** The chance that a randomly selected *outcome* is part of an *event*

  $P(X >= 3) = 4 / 6$

Random Variables

- ***Define random variable***

$$Z : \Omega \to \{0,1\}$$

$$Z = \begin{cases} 1 & X \text{ } is \text{ } even \\ 0 & X \text{ } is \text{ } odd \end{cases}$$

- ***Sample Space:***

$$\Omega_Z = \{0,1\}$$

- ***Probability:*** The chance that a randomly selected *outcome* is part of an *event*

$$P(Z = 1) = P(X \in \{2,4,6\}) = 3/6$$

Dependence of random variables

- *Joint probabilities of events in different probability spaces*

$$
\begin{aligned}
P(X = 2, Z = 1) &= P(X \in \{2\}, X \in \{2,4,6\}) \\
&= P(X \in \{2\} \cap \{2,4,6\}) \\
&= 1/6
\end{aligned}
$$

- *Conditional probabilities of events in different probability spaces*

$$
\begin{aligned}
P(X = 2 \,|\, Z = 1) &= P(X \in \{2\} \,|\, X \in \{2,4,6\}) \\
&= P(X \in \{2\} \cap \{2,4,6\}) / P(X \in \{2,4,6\}) \\
&= (1/6)/(3/6) \\
&= 1/3
\end{aligned}
$$

Distribution

- Commonly used (or abused) shorthands:

Full Expression	Shorthand 1	Shorthand 2	Shorthand 3
$P(X \in E)$		$P_X(E)$	$P(E)$
$P(X \in \{x\})$	$P(X = x)$	$P_X(x)$	$P(x)$
$P(X \in [x, \infty])$	$P(X \geq x)$		
$P(X \in [x_1, x_2])$	$P(x_1 \leq X \leq x_2)$		

Bayes' Rule

- Let X and Z be two random variables.
- Let Ω_X be the sample space of X.
- Let Ω_Z be the sample space of Z.

Rules for singleton events

Sum rule

$$P(x) = \sum_{z \in \Omega_Z} P(x, z)$$

$$P(z) = \sum_{x \in \Omega_X} P(x, z)$$

Product rule

$$P(x, z) = P(x \mid z) P(z) = P(z \mid x) P(x)$$

Bayes' rule

$$P(z \mid x) = \frac{P(x \mid z) P(z)}{P(x)}$$

Bayes' Rule

Likelihood Prior

$$P(z \mid x) = \frac{P(x \mid z)P(z)}{P(x)}$$

Posterior

$$\Omega_X = \{d, \neg d\}$$
$$\Omega_Z = \{t, \neg t\}$$

$P(d)$	*Probability of rare disease:* 0.005
$P(t \mid d)$	*Probability of detection:* 0.98
$P(t \mid \neg d)$	*Probability of false positive:* 0.05
$P(d \mid t)$	*Probability of disease when test positive?*

Bayes' Rule

$$\begin{aligned} P(d,t) &= P(t \mid d)P(d) \\ &= 0.98 \times 0.005 \\ &\approx 0.005 \end{aligned}$$

$$\begin{aligned} P(\neg d,t) &= P(t \mid \neg d)P(\neg d) \\ &= 0.05 \times 0.995 \\ &\approx 0.05 \end{aligned}$$

Product rule

$$P(d) = 0.005$$
$$P(t \mid d) = 0.98$$
$$P(t \mid \neg d) = 0.05$$

$$\begin{aligned} P(t) &= P(t,d) + P(t,\neg d) \\ &= 0.98 \times 0.005 \\ &\approx 0.055 \end{aligned}$$

Sum rule

$$\begin{aligned} P(d \mid t) &= P(t \mid d)P(d)/P(t) \\ &\approx 0.005/0.055 \\ &= 0.09 \end{aligned}$$

Bayes' rule

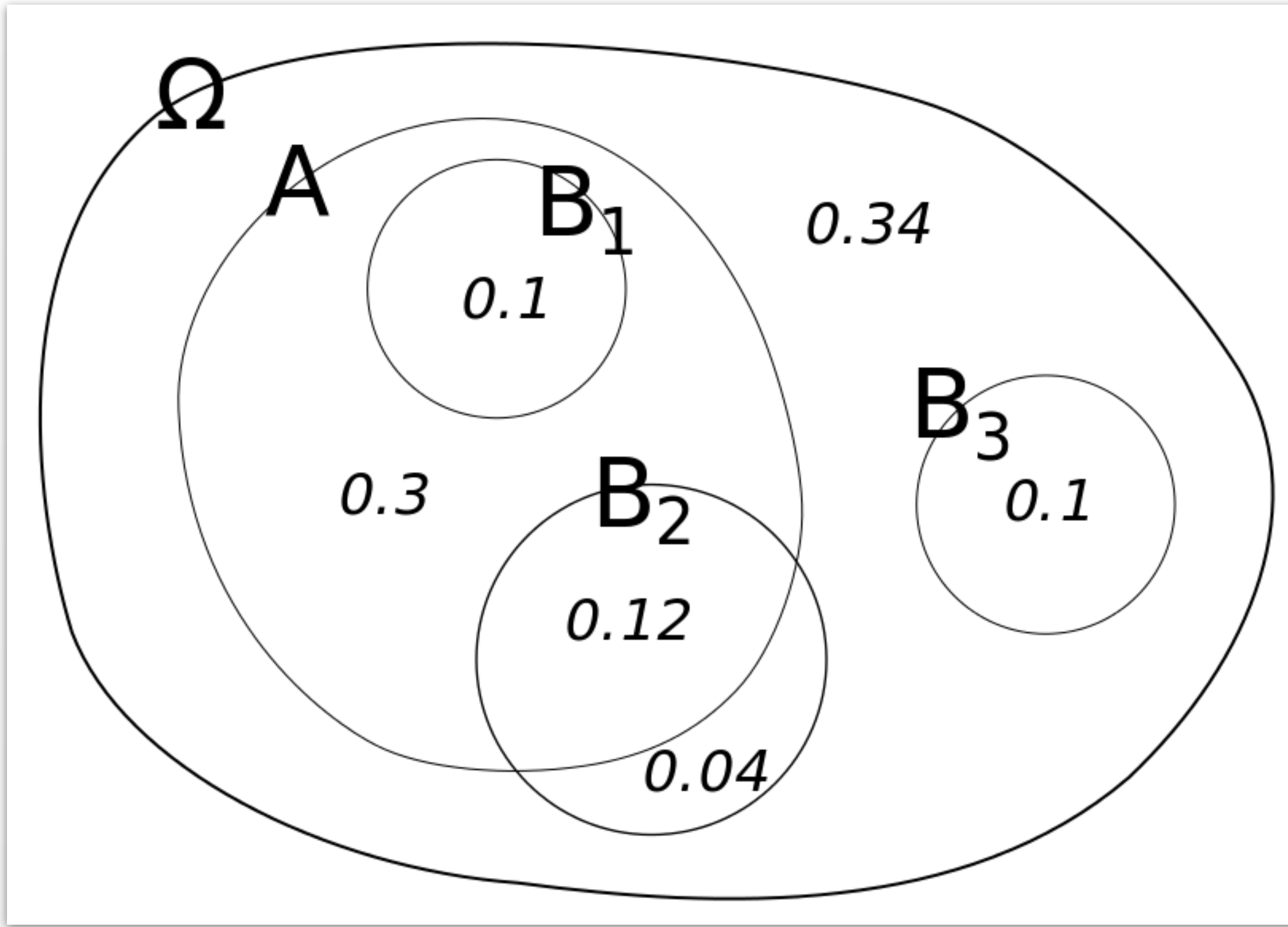
$$P(z \mid x) = \frac{P(x \mid z)P(z)}{P(x)}$$

Probability Mass Function (pmf)

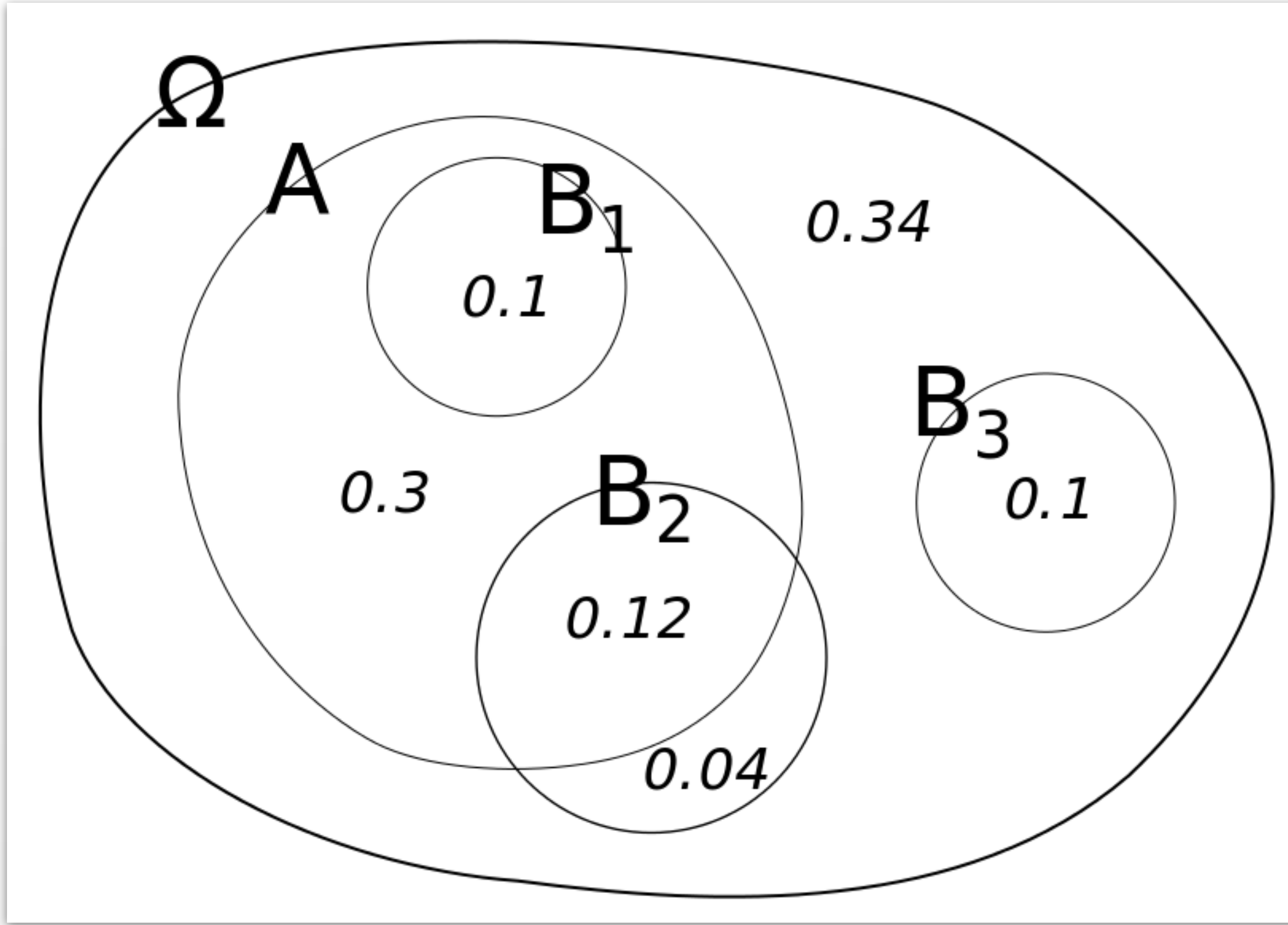
Assigns probability to each outcome of a discrete sample space.

- $P : \Omega \to [0,1]$
- Technically, it is different from the probability measure which assigns probability to a subset of Ω, instead of an element.
- However, it is same as the probability measure on a singleton set.
- We use the same notation for pmf and the corresponding probability measure.
- Dice roll: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Conditional Probability: Continuous distributions

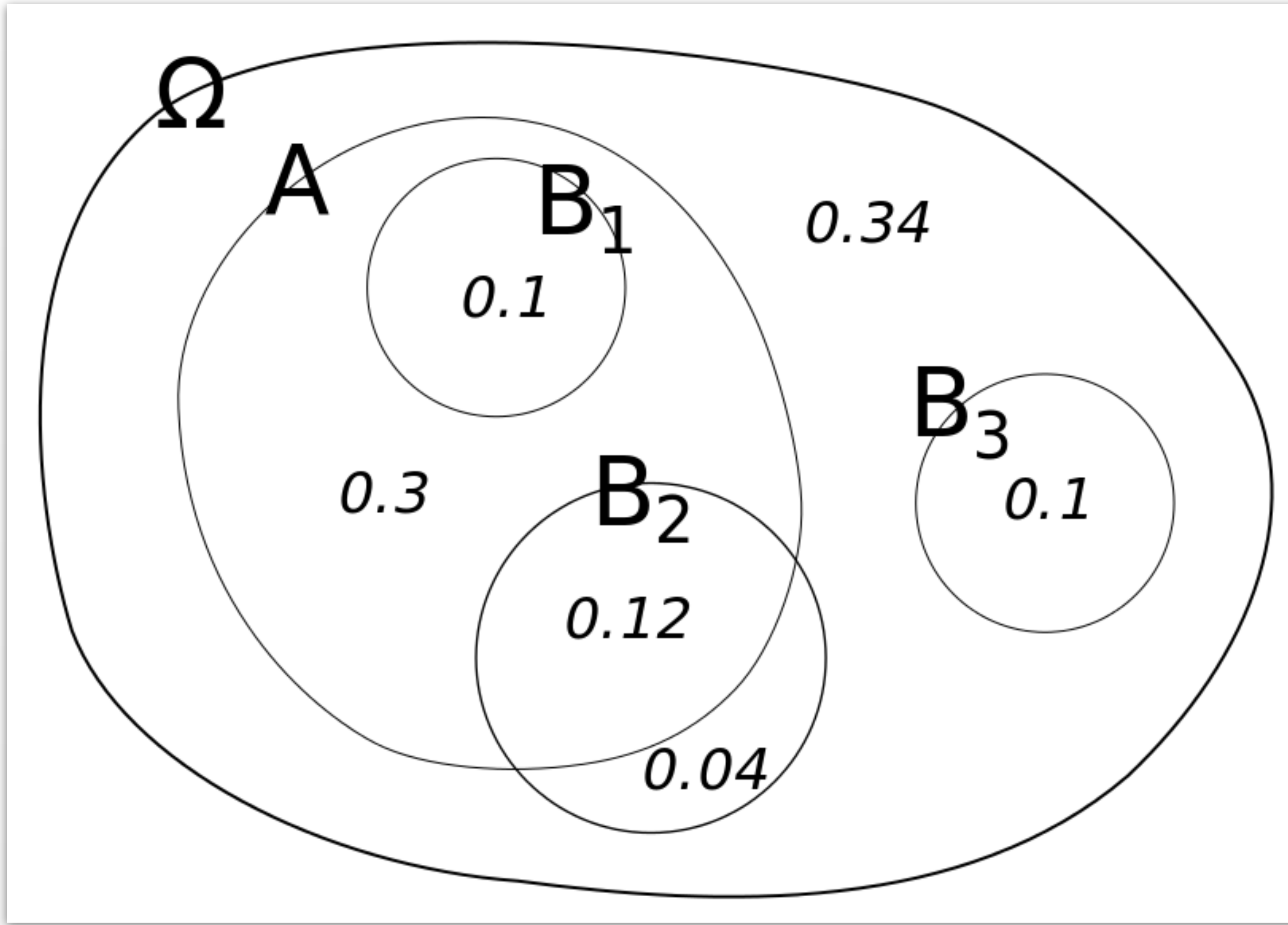


Conditional Probability: Continuous distributions



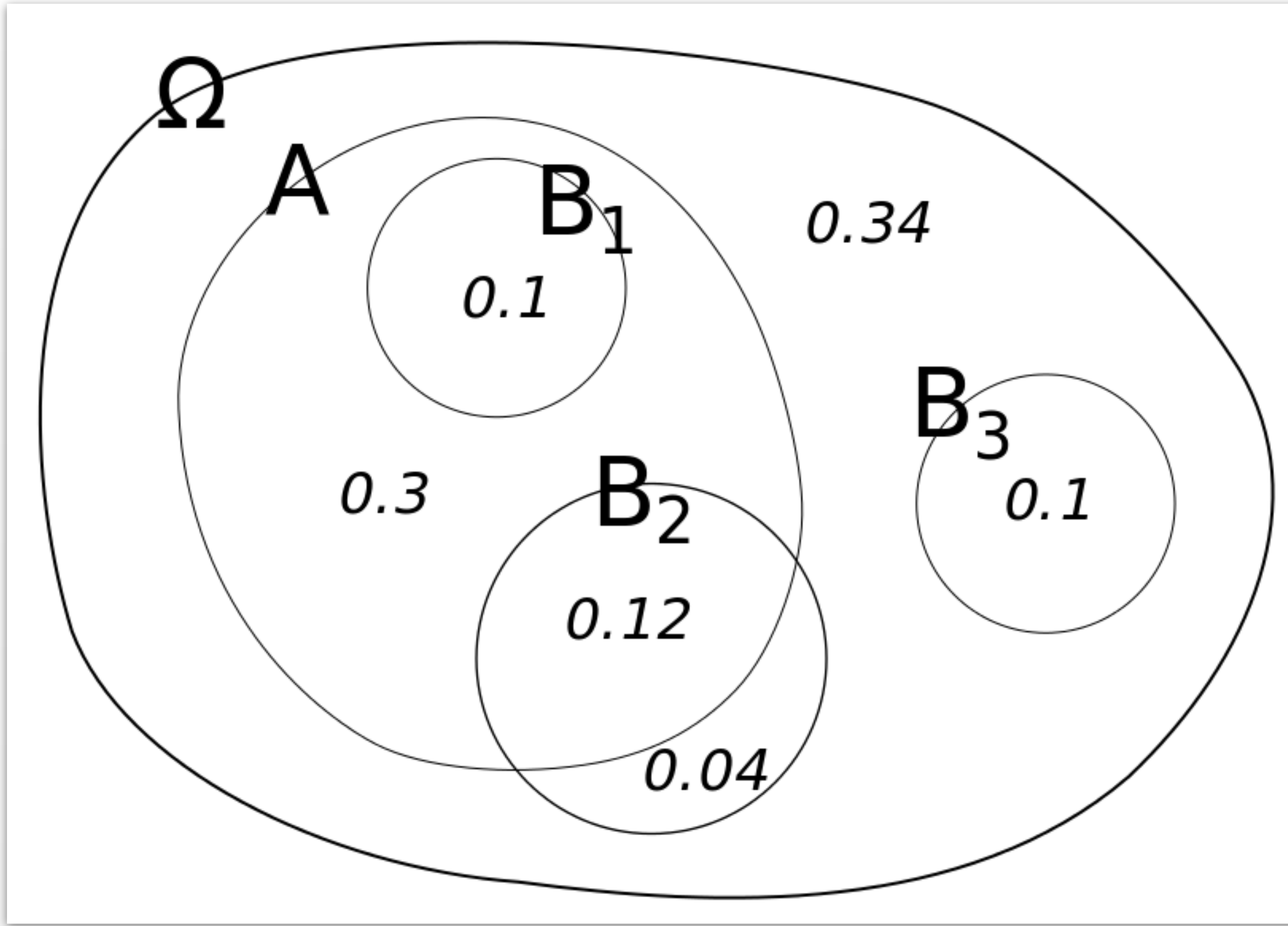
What is the probability $P(B_3)$? 0.1 / 0.34

Conditional Probability: Continuous distributions



What is the probability $P(B_2 \mid A)$? 0.12 / 0.3

Conditional Probability: Continuous distributions



What is the probability $P(B_1 \mid B_3)$? 0.0 / 0.1

Probability Density Functions

- **Problem:** If *X* is a *continuous* variable, then $P(X{=}x)$ is **0** for any outcome ***x***

Single Outcome

$$X \sim \mathrm{Normal}(0,1)$$

$$P(X = \pi) = 0$$

$$P(3.1 \leq X \leq 3.2) \neq 0$$

Event

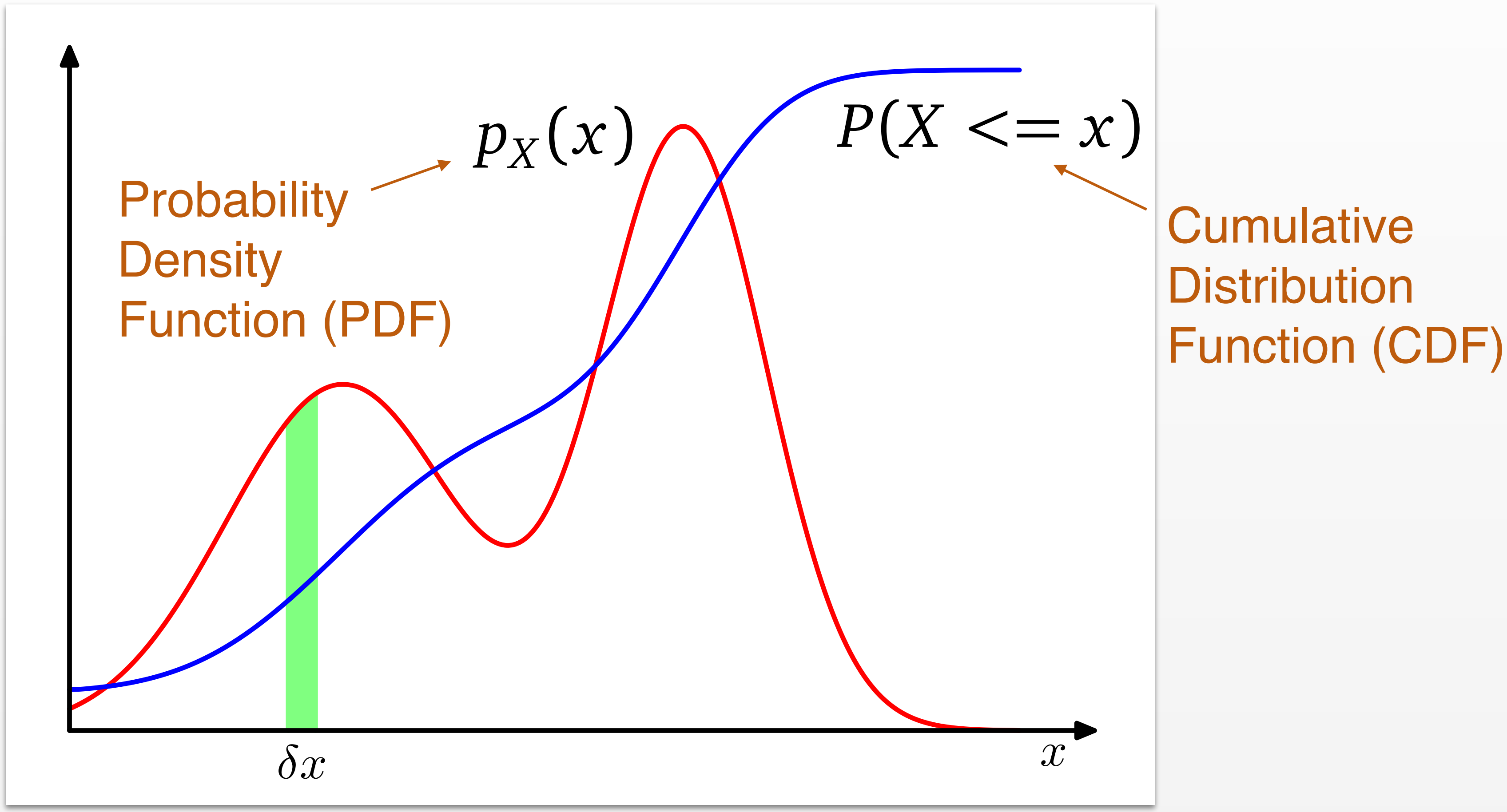
- **Solution:** Define a density function as a derivative

Capital P for probability

$$p_X(x) = \lim_{\delta \to 0} \frac{P(x - \delta < X < x + \delta)}{2\delta}$$

Small p for density

Probability Density Functions



$$p_X(x) = \lim_{\delta \to 0} \frac{P(x - \delta < X < x + \delta)}{2\delta}$$

Discrete vs. Continuous R.V.

		Discrete	Continuous
Sample Space		Any Finite or countably infinite set . e.g.,$\{0,1\}$, set of all integers or rational numbers	Any uncountable set. e.g., real line, interval [0,1].
pmf	$P : \Omega \rightarrow [0,1]$	$P(x)$	Does not exist
pdf	$p : \Omega \rightarrow \mathbf{R}^+$	Does not exist.	$p(x)\ or\ f(x) = \frac{d}{dx}F(x)$
cdf	$F : \Omega \rightarrow [0,1]$	$F(x) = \sum_{t<=x} P(t)$	$F(x) = \int_{-\infty}^{x} p(x)dx$
Probability measure	$P : \mathscr{F} \rightarrow [0,1]$	$P(A) = \sum_{x \in A} P(x)$	$P(A) = \int_{x \in A} p(x)dx$

Mixed R.V.

- There are distributions that are partly continuous and partly discrete.
  - Some points in the sample space have non-zero probabilities; e.g. zero inflated models: distribution of alcohol consumption (large fraction of individuals have 0 alcohol consumption).
  - Won’t talk about this case further in this course.

Notation

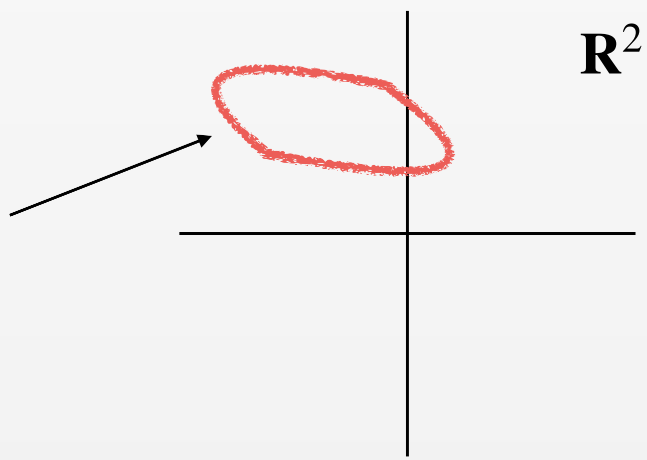
The distribution of a random variable X can be specified in terms of a

- Probability measure: $X \sim P$
- Probability mass function: $X \sim P$
- Probability density function: $X \sim p$
- Cumulative distribution function: $X \sim F$
- Or by name of a well-known probability distribution: $X \sim \text{Normal}(\mu, \sigma)$

Multivariate Distributions/Random Vectors

- **Joint probability measure**: A distribution of a 2 dimensional random vector $\mathbf{X} = [X_1, X_2]$ is specified by a joint probability measure $(\Omega = \Omega_1 \times \Omega_2, \mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2, P)$, where Ω_1 and Ω_2 are the outcome spaces for X_1 and X_2, respectively and $\mathscr{F}_1$ and $\mathscr{F}_2$ is the event space containing subsets of Ω_1 and Ω_2, respectively.
  - If $\Omega_i = \mathbf{R}$ then $\Omega = \mathbf{R}^2$ and $\mathscr{F}$ contains subsets of $\mathbf{R}^2$
  - $P : \mathscr{F} \to [0,1]$ assigns probabilities to subsets of $\mathbf{R}^2$.

P assigns probabilities to subsets of $\mathbf{R}^2$



Multivariate Distributions/Random Vectors

- If Ω_1 and Ω_2 are discrete a joint pmf can be defined $P : \Omega_1 \times \Omega_2 \to [0,1]$
- If Ω_1 and Ω_2 are continuous a joint pdf can be defined $p : \Omega_1 \times \Omega_2 \to \mathbf{R}^+$
- What if Ω_1 is continuous, but Ω_2 is discrete?
  - We can define a function that acts like pdf in the first dimension and a pmf in the second dimension. We use p to denote this function: $p : \Omega_1 \times \Omega_2 \to \mathbf{R}^+$

Marginals

Marginal pdf $\longrightarrow p_{X_1}(x_1) = \sum_{x_2 \in \Omega_2} p(x_1, x_2)$

Marginal pmf $\longrightarrow P_{X_2}(x_2) = \int_{x_1 \in \Omega_1} p(x_1, x_2)\ dx_1$

Joint decomposition

$$p(x_1, x_2) = \underbrace{p_{X_1|X_2}(x_1 \mid x_2)}_{\text{Conditional pdf}} P_{X_2}(x_2) \longrightarrow \text{Marginal pmf}$$

$$= \underbrace{P_{X_2|X_1}(x_2 \mid x_1)}_{\text{Conditional pmf}} p_{X_1}(x_1) \longrightarrow \text{Marginal pdf}$$

Multivariate Distributions/Random Vectors

- **Joint probability measure**: A distribution of a D dimensional random vector $\mathbf{X} = [X_1, X_2 \ldots X_D]$ is specified by a joint probability measure $(\boldsymbol{\Omega} = \prod_{i=1}^{D} \boldsymbol{\Omega}_i, \mathscr{F} = \bigotimes_{i=1}^{D} \mathscr{F}_i, P)$, where $\boldsymbol{\Omega}_i$ is the outcome space for X_i and $\mathscr{F}_i$ is the event space containing subsets of $\boldsymbol{\Omega}_i$.
  - If $\boldsymbol{\Omega}_i = \mathbf{R}$ then $\boldsymbol{\Omega} = \mathbf{R}^D$ and $\mathscr{F}$ contains subsets of $\mathbf{R}^D$
  - $P : \mathscr{F} \to [0,1]$ assigns probabilities to subsets of $\mathbf{R}^D$.

Expected Values

$$X \sim p(x)$$ ← X is a random variable with density p(x)

Statistics

$$\mathbb{E}[X] = \sum_x p(x)\, x$$

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

(distribution implied by X)

Machine Learning

$$\mathbb{E}_{p(x\,|\,y)}[f(x)] = \sum_x p(x\,|\,y)\, f(x)$$

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

(explicitly define distribution)

Mean, Variance, Covariance

Mean

$$\mu_X = \mathbb{E}[X]$$

Variance

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

Covariance

$$\Sigma_{X,Y} = \mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$