Foundations of Machine Learning
AI2000 and AI5000

FoML-02
Probability - Bayes Theorem and Independence

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

So far in FoML

- What is ML?

So far in FoML

- What is ML?
- Learning Paradigms

Probability Theory

Probability Theory

- Provides a consistent framework for the quantification and manipulation of "Uncertainty"

Probability Theory

- Provides a consistent framework for the quantification and manipulation of “Uncertainty”
- Where does this ‘Uncertainty’ come from?

Uncertainty in ML

- Measurement Noise

Uncertainty in ML

- Measurement Noise
- Finite size of the datasets

Probability Theory

- Frequentist Interpretation

Probability Theory

- Frequentist Interpretation
  - Fraction of times the event occurs

Probability Theory

- Bayesian Approach

Probability Theory

- Bayesian Approach
  - Quantification of plausibility or strength of the belief of an event

Probability Theory

- Bayesian Approach
  - Quantification of plausibility or strength of the belief of an event
  - Modeling based approach

Probability Theory

- Bayesian Approach
  - Quantification of plausibility or strength of the belief of an event
  - Modeling based approach
  - Plays a central role in this course

Random Variable

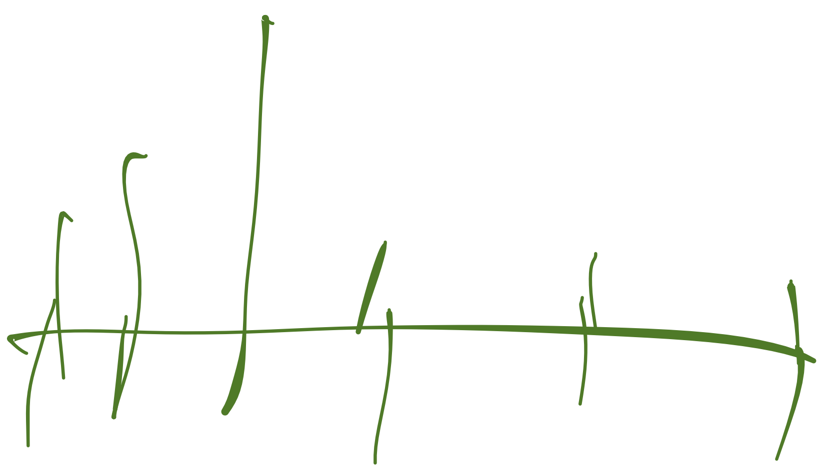
- Stochastic variable sampled from a set of possible outcomes

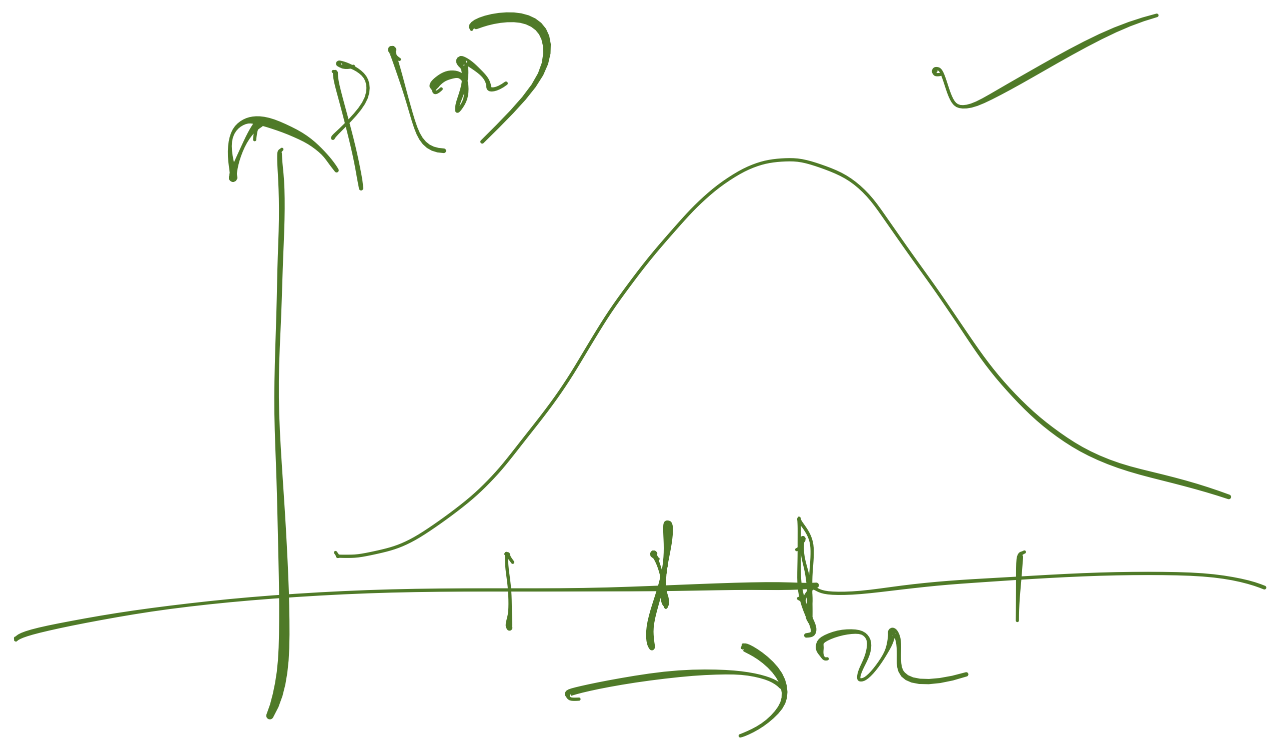
Random Variable

- Stochastic variable sampled from a set of possible outcomes
- Discrete or Continuous

Random Variable

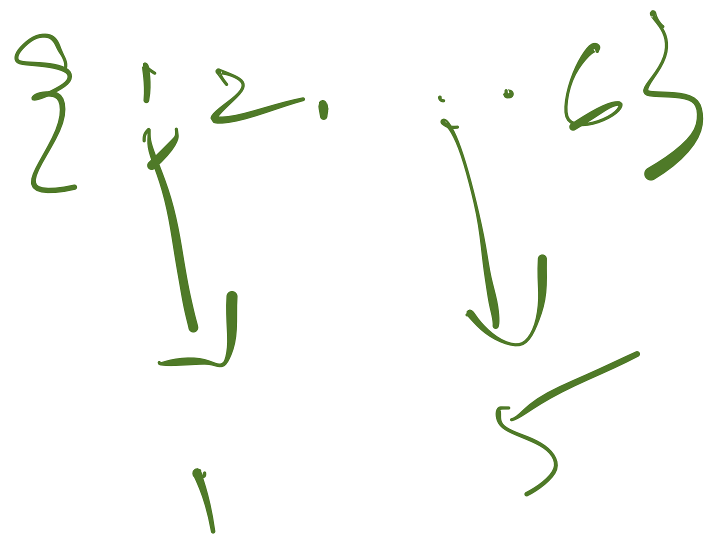
- Stochastic variable sampled from a set of possible outcomes
- Discrete or Continuous
- Probability distribution ρ(X)





Random Variable - Example (discrete)

- Throwing a dice



Random Variable - Example (discrete)

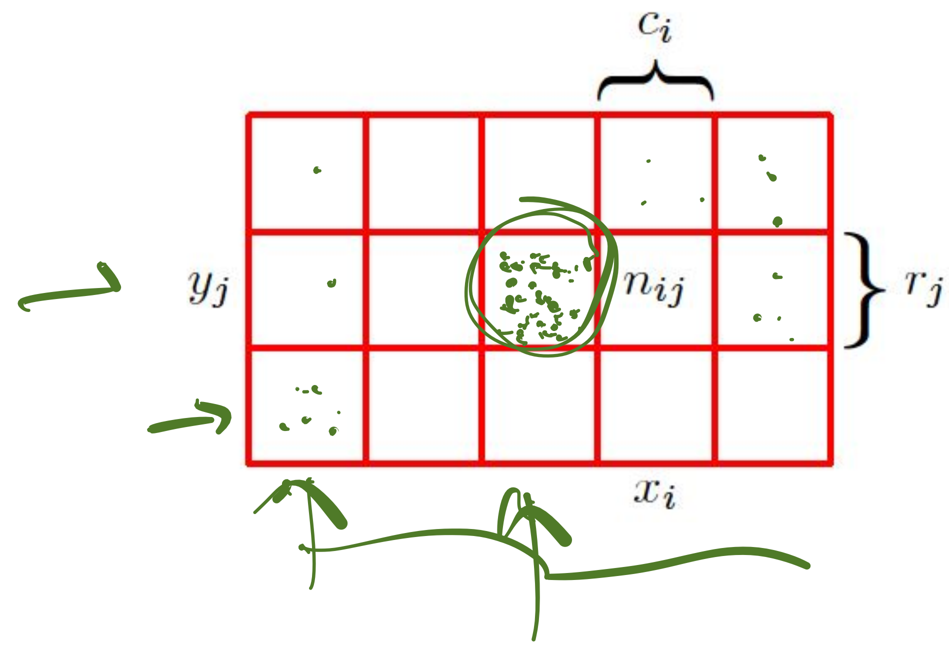
- Flipping a coin $\{H, T\}$

Two Discrete Random Variables

- X
- Y

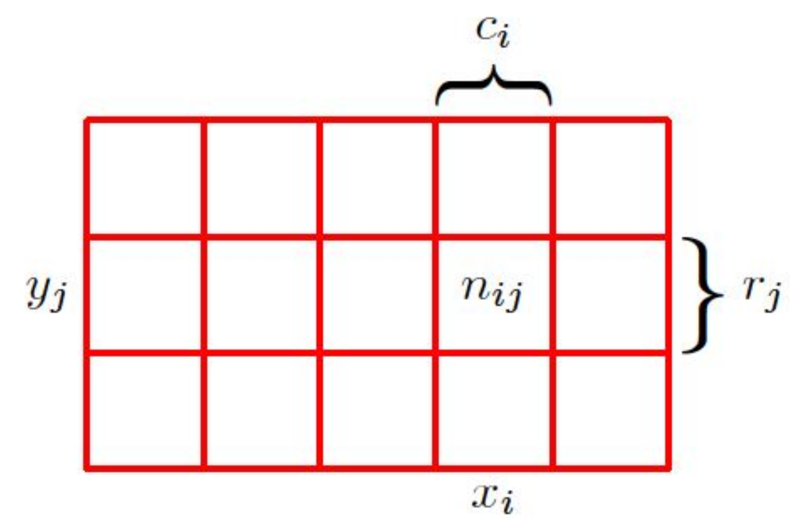
Two Discrete Random Variables

- X = $\{x_1 \ldots x_6\}$
- Y = $\{y_1 \ldots y_3\}$
- N trails: sample both



Two Discrete Random Variables

- Joint probability

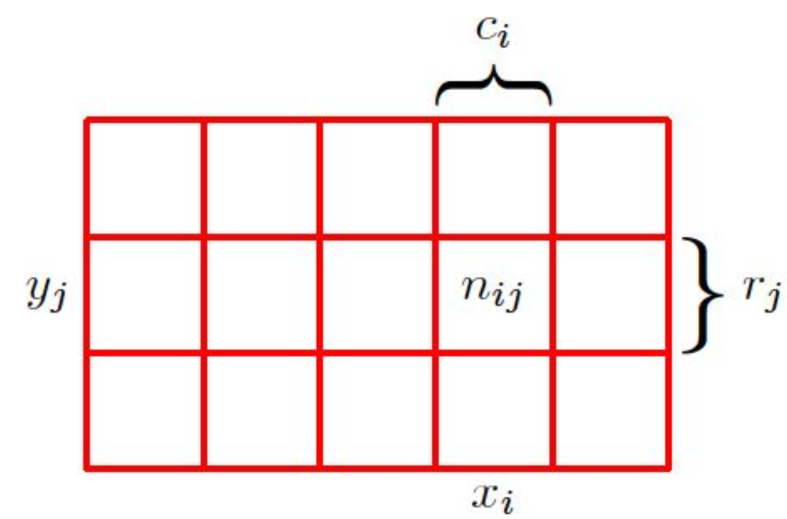


Two Discrete Random Variables

- Joint probability

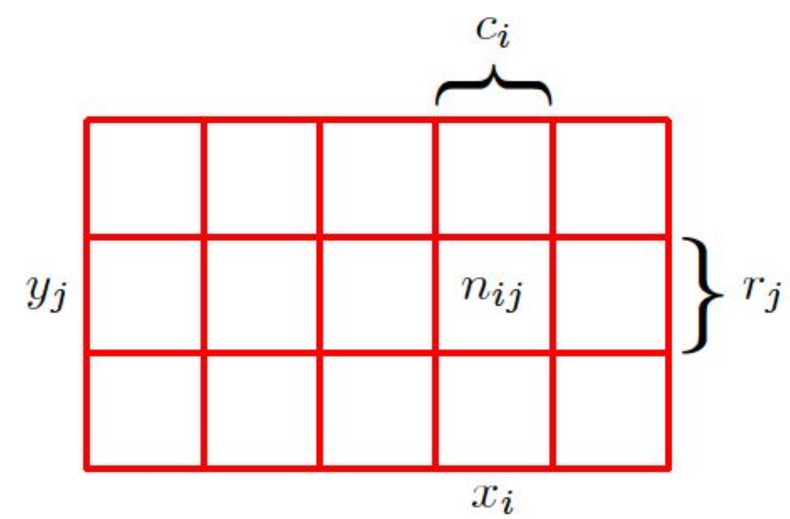
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} \quad \forall\, x_i, y_j$$

and



Two Discrete Random Variables

- If I am interested only on X

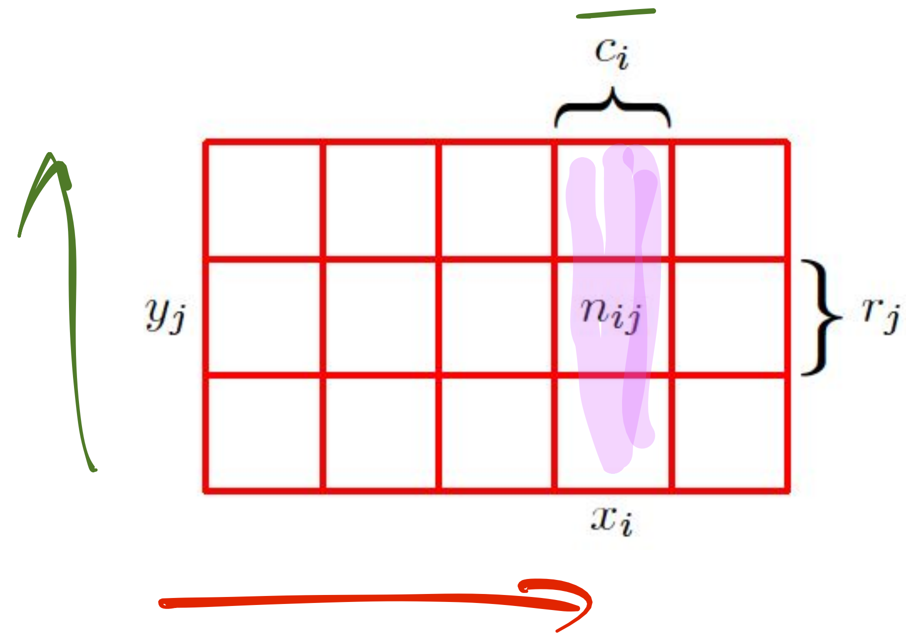


Two Discrete Random Variables

- If I am interested only on X
- Marginal probability of X

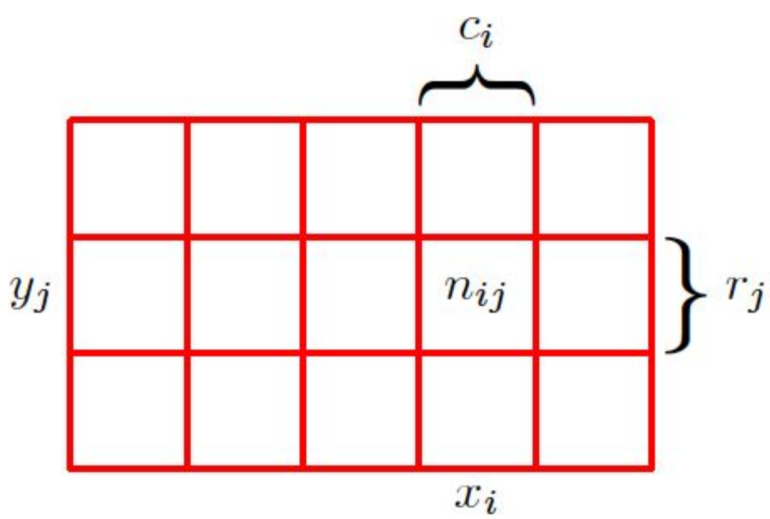
$$p(X = x_i) = \frac{c_i}{N}$$

$$c_i = \sum_{j=1}^{3} n_{ij}$$



Sum rule of Probability

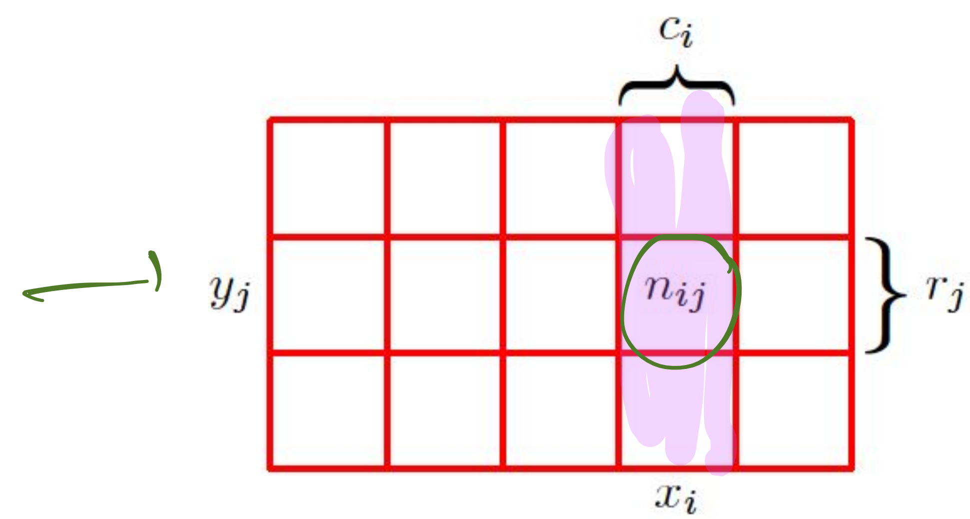
$$p(X = x_i) = \sum_{j=1}^{3} p(X = x_i, Y = y_j)$$



Conditional Probability

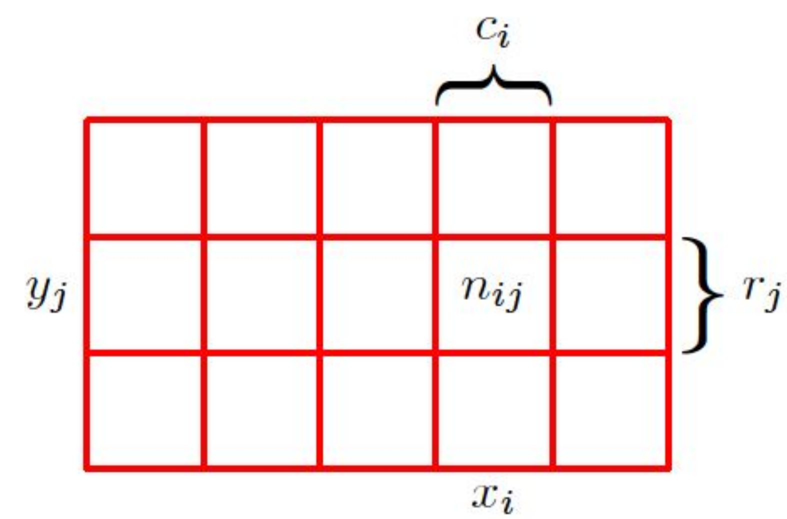
- Conditional probability of Y given X

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$



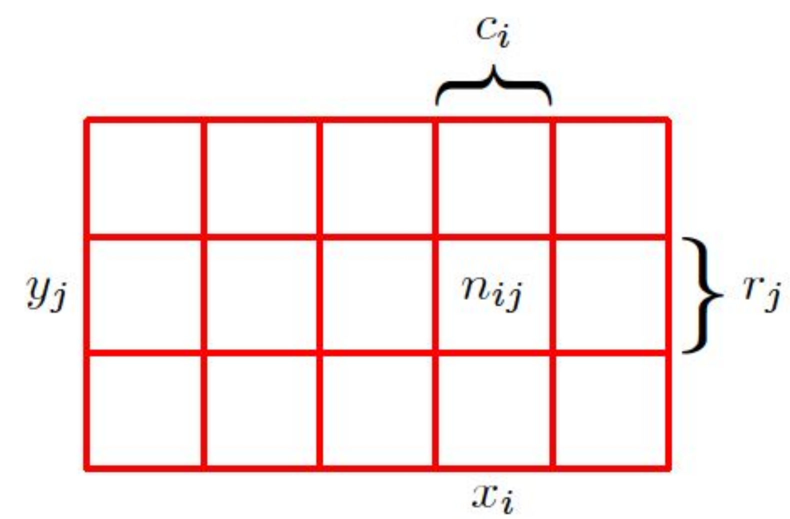
Product Rule of probability

$$p(Y = y_j / X = x_i) =$$



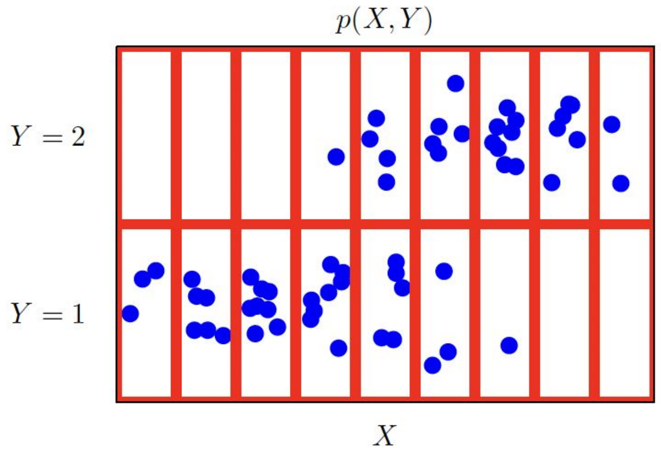
Product Rule of probability

$$p(Y = y_j / X = x_i) =$$

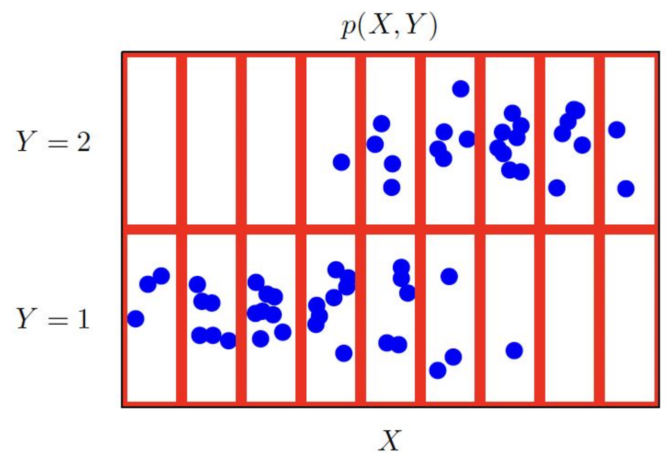


$$p(Y = y_j, X = x_i) = p(Y = y_j / X = x_i) \cdot p(X = x_i)$$

Example: Marginal & Conditional distributions

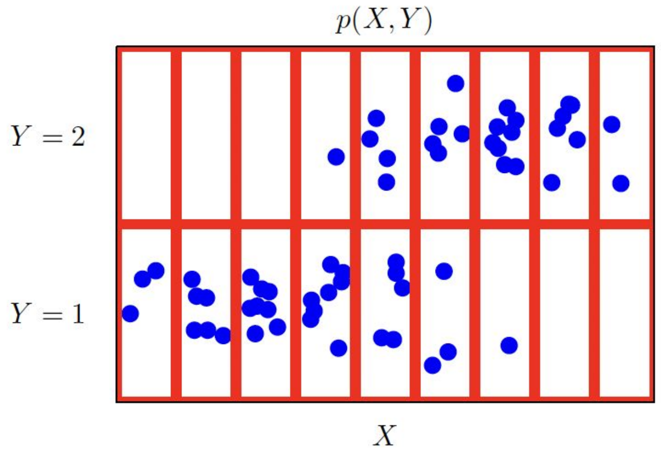


Example: Marginal & Conditional distributions



- X
- Y
- 60 trails - sample both

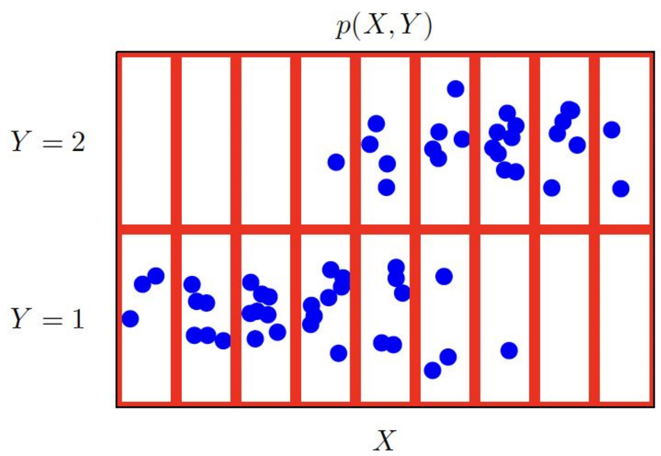
Example: Marginal & Conditional distributions

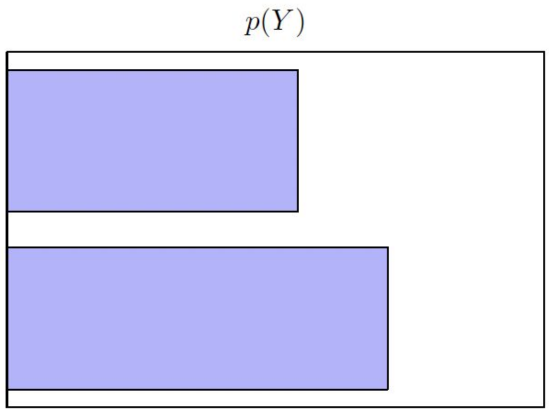


- Marginal distribution p(Y)

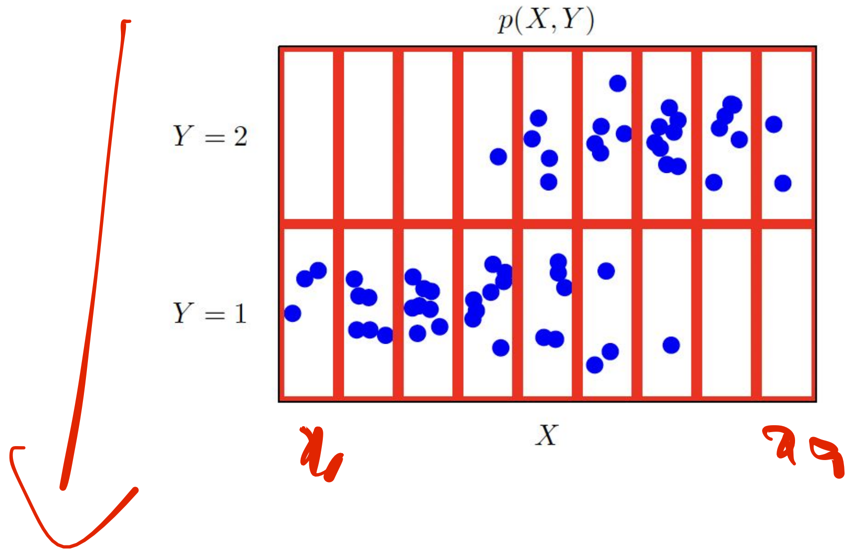
$$p(y) = \sum_{x \in X} p(x, y)$$

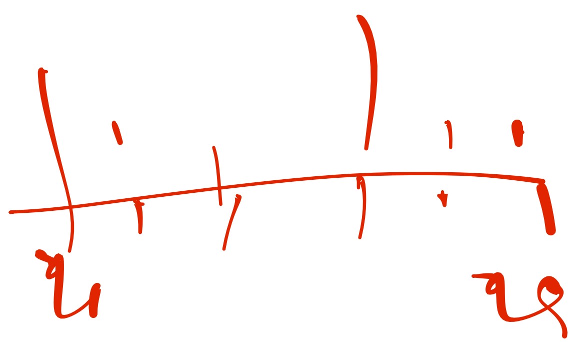
Example: Marginal & Conditional distributions





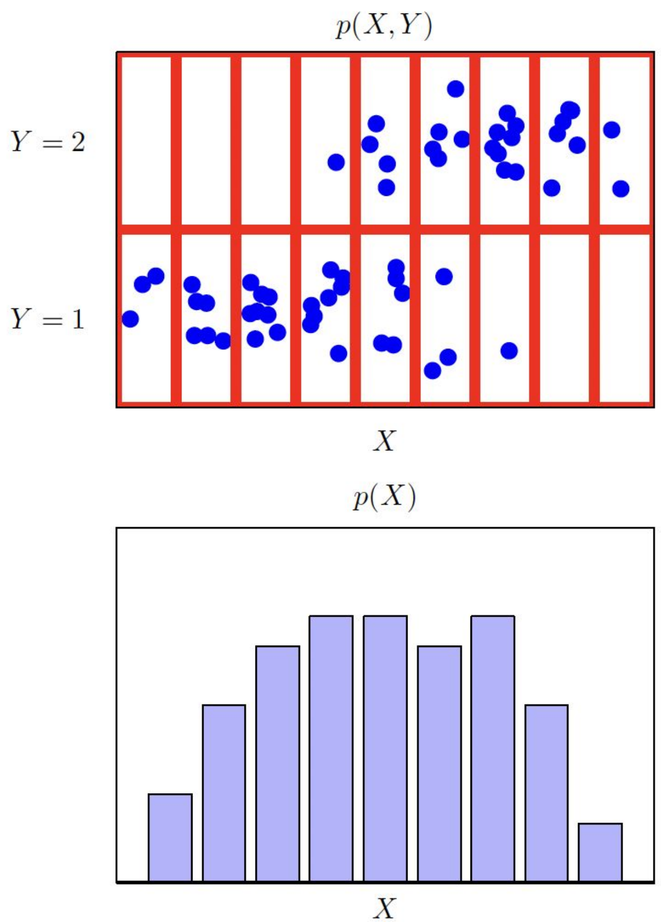
Example: Marginal & Conditional distributions



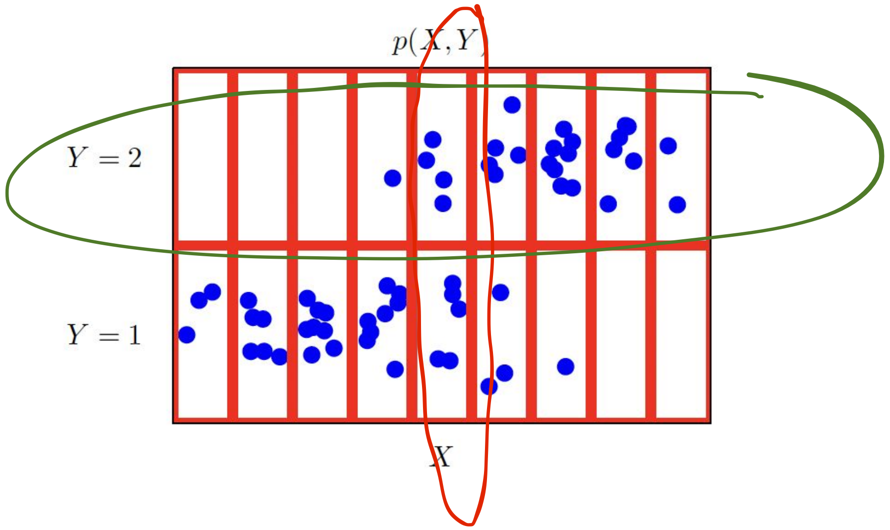


- Marginal distribution p(X)

Example: Marginal & Conditional distributions



Example: Marginal & Conditional distributions

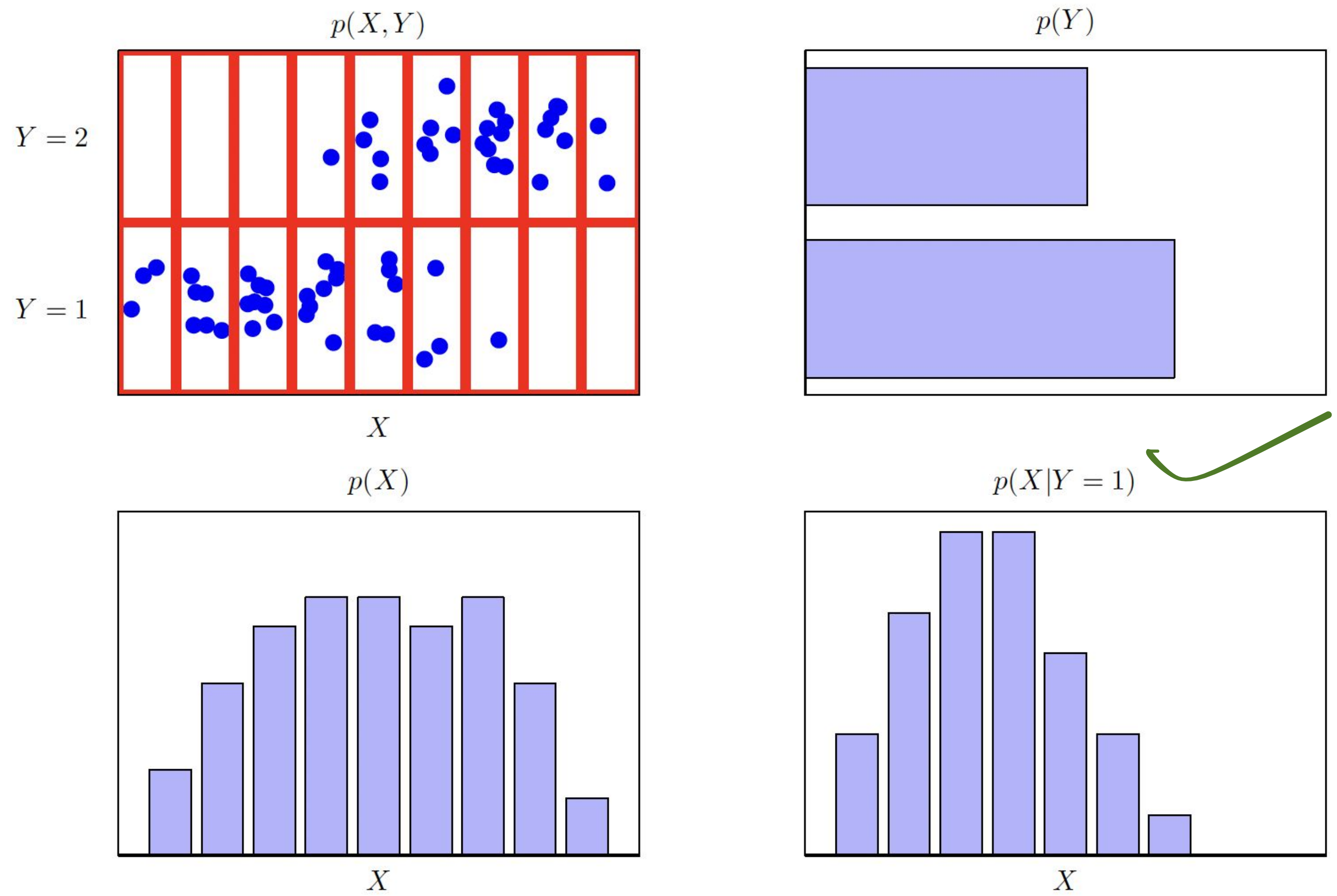


- Conditional distribution of X

$p(Y / X = x_i)$

$p(X / Y = y_j)$

Example: Marginal & Conditional distributions



Example: Marginal & Conditional distributions

$$\sum_{y \in Y} p(Y = y_i / X = x_i) = ?$$

$$\sum_{y \in X} p(Y = y_i / X = x_j) =$$

'x'

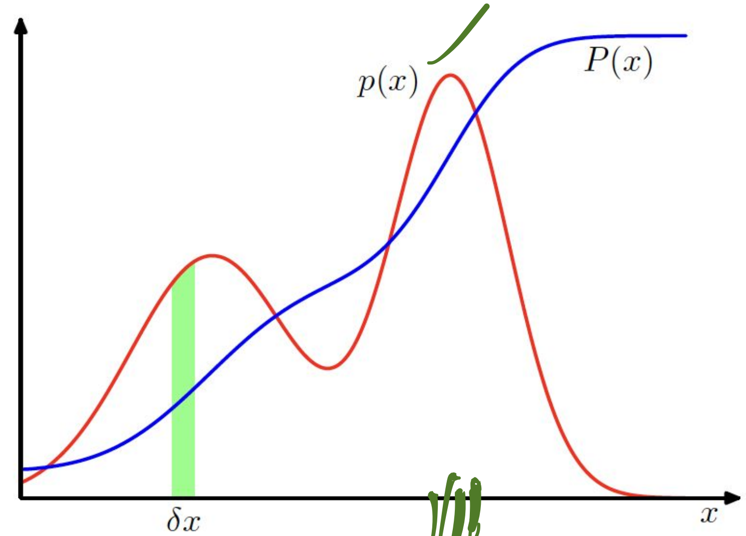
Continuous Random Variable

Continuous Random Variable

- p(X): Probability density over X

Continuous Random Variable

- p(X): Probability density over X
- Probability of x falling in (x, x+dx)
- Probability over a finite interval (a, b)



Continuous Random Variable

- Non-negativity
- Normalization

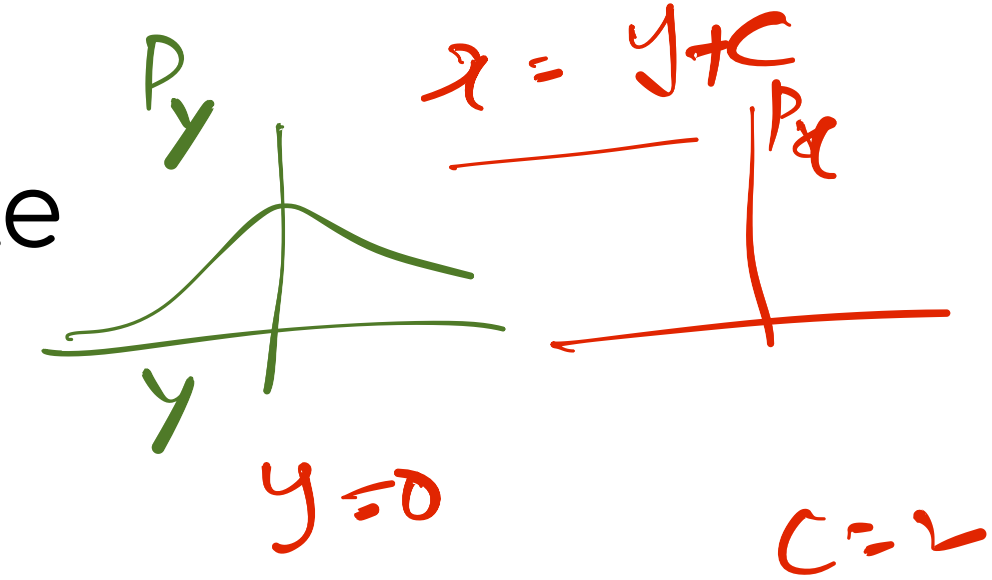
Continuous Random Variable

- Change of variables

Continuous Random Variable

- Change of variables
- $x = g(y)$

Continuous Random Variable



- Change of variables
- x = g(y)
- Probabilities in (x, x+dx) must be transformed to (y, y+dy)

$$p_y(y)\,dy = p_x(x)\,dx$$

$$p_x(x) = p_y(y)\left|\frac{dy}{dx}\right|$$

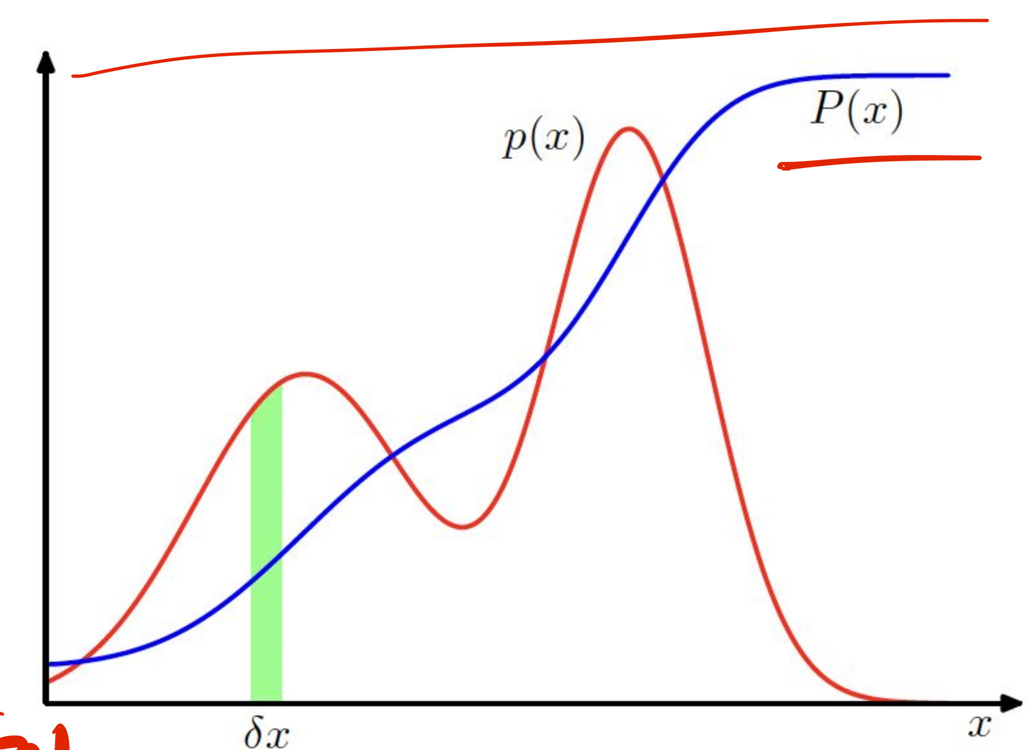
Continuous Random Variable

- Cumulative distribution function

$$P(x) = p(X \leq x)$$

$$P(x) = \int_{-\infty}^{x} p(x)\, dx$$

$$p(x) = \frac{d}{dx} P(x)$$



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Rules of Probability Theory

	Discrete	Continuous
Additivity	$p(X \in A) = \sum_{x \in A} p(x)$	$\int_{x \in X} p(x)\, dx$
Positivity	$p(x) \geq 0$	$p(x) \geq 0$
Normalization	$\sum_{x \in X} p(x) = 1$	$\int_{\mathcal{X}} p(x)dx = 1$
Sum Rule	$p(x) = \sum_{y \in Y} p(x, y)$	$p(x) = \int_{y} p(x, y)\, dy$
Product Rule	$p(x, y) = p(x/y) \cdot p(y)$	$p(x, y) = p(x/y) \cdot p(y)$

Bayes theorem

Bayes Theorem

- Product rule $$p(x, y) = p(x/y) \cdot p(y)$$

Bayes Theorem

- Product rule
- Symmetry property
- Bayes rule
- Denominator

$$p(x, y) = p(x/y) \cdot p(y)$$

$$p(y,x) = p(y/x) \cdot p(x)$$

$$p(y/x) = \frac{p(x/y)\, p(y)}{p(x)}$$

$$\sum_y p(y/x) = 1 = \frac{1}{p(x)} \sum_y p(x/y)\, p(y)$$

$$\sum_y p(x,y) \leftarrow p(x) = \sum_y p(x/y)\, p(y)$$

Bayes Theorem

$p(x/y)$

$$p(y/x) = \frac{p(x/y) \cdot p(y)}{p(x)}$$

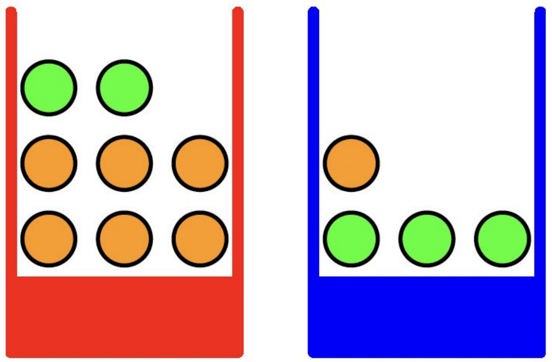
- Prior probability
- Posterior probability of Y
- Likelihood of X = x given Y = y
- Evidence for X = x

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

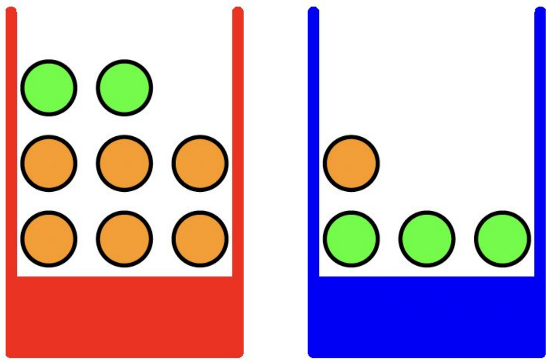
Example

Boxes and Fruits



- Random variables

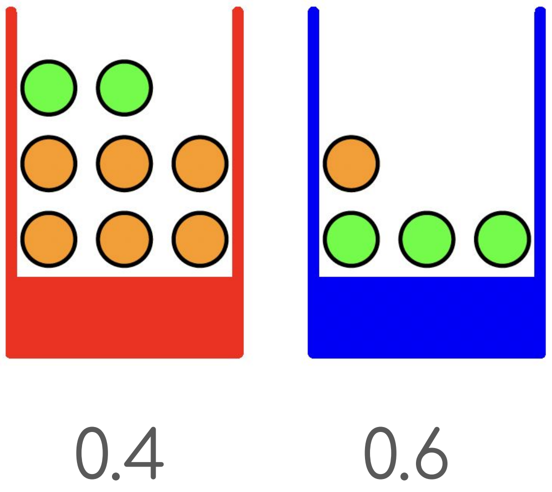
Boxes and Fruits



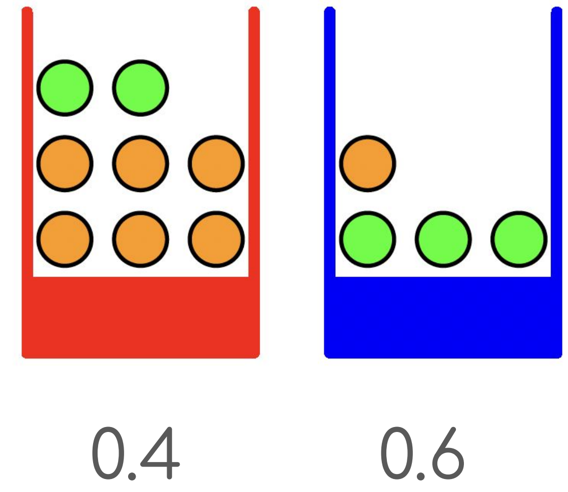
- Random variables
  - Box - B
  - Fruit - F

Boxes and Fruits

- Prior Box distribution



Boxes and Fruits



- Prior Box distribution
  - $P(B=r) = 4/10 \quad P(B=b) = 6/10$
- Conditional of F given B

$P(F=a/B=r) = 2/8 = 1/4 \qquad P(F=a/B=b) = 3/4$

$P(F=o/B=r) = 3/4 \qquad P(F=o/B=b) = 1/4$

- Marginal of F

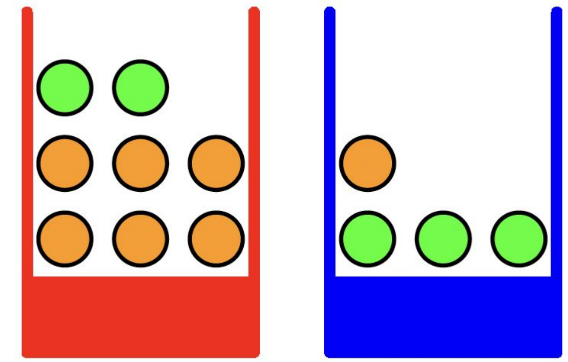
$P(F=a) = \sum_B P(B,F) = \sum_B P(F/B)\,P(B) = \frac{4}{10}\cdot\frac{1}{4} + \frac{6}{10}\cdot\frac{3}{4} = 11/20$

$= 9/20$

$P(F=o) = \sum P(B,F)$

Boxes and Fruits

- Marginals ρ(F=a) = 11/20 & ρ(F=0) = 9/20
- Posterior probability of Box given observed fruit



0.4 0.6

$$p(B = r / F = o) =$$

$$\frac{p(F=o \mid B=r) \cdot p(B=r)}{p(F=o)} = \frac{3/4 \cdot 4/10}{9/20} = \frac{3}{10} \cdot \frac{20}{9} = 2/3 \simeq 67\%$$

$$p(B=r) = 40\%$$

Independence

Independent Random variable

- Two random variables X and Y are independent iff measuring X gives no information about Y (and vice versa)

$$P(Y/X) = P(Y) \checkmark \qquad P(X/Y) = P(X)$$

$$\frac{P(X,Y)}{P(X)} = P(Y)$$

$$P(X,Y) = P(X)P(Y)$$

Next

Expectation, Variance, and Gaussian Distribution