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

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

Data-driven Intelligence & Learning Lab

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

Data-driven Intelligence
& Learning Lab

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

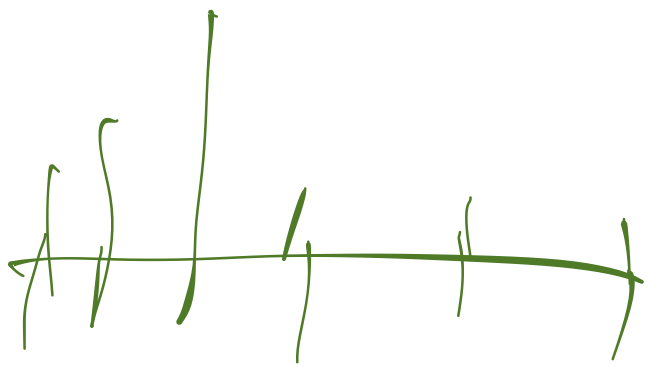
- Stochastic variable sampled from a set of possible outcomes

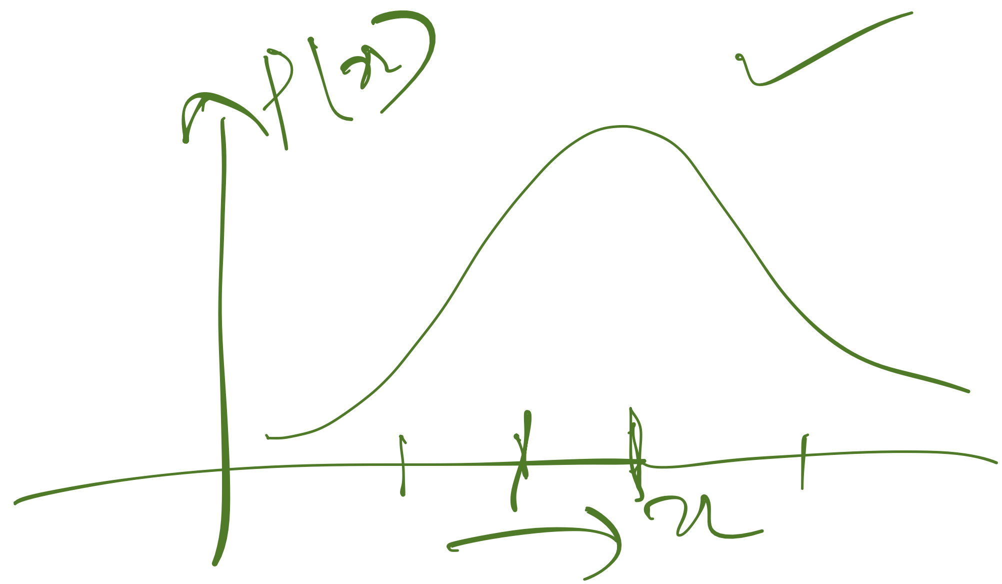
Random Variable

- 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

$\{1, 2, \ldots, 6\}$

$1 \downarrow \quad \ldots \downarrow$

$1 \quad 5$

Random Variable - Example (discrete)

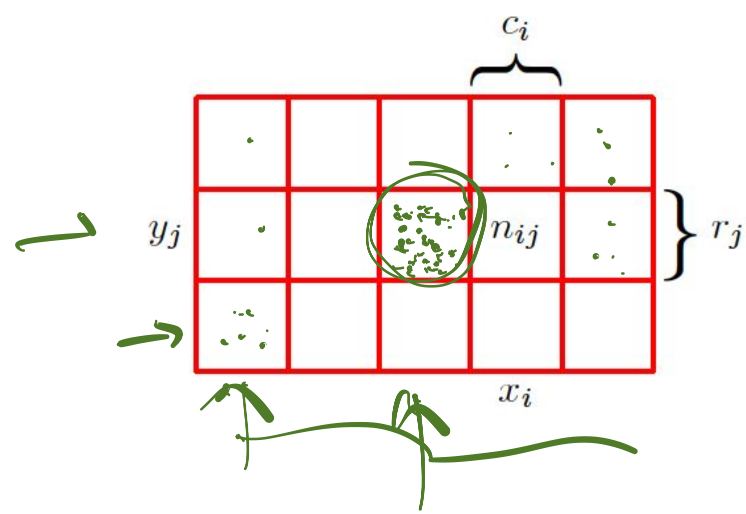
- Flipping a coin $\{H, T\}$

Two Discrete Random Variables

- X
- Y

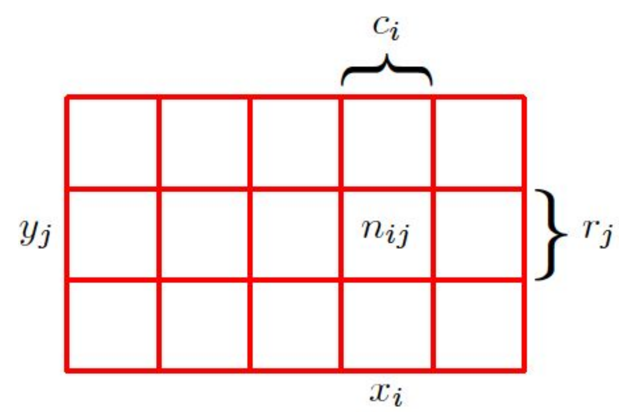
Two Discrete Random Variables

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



Two Discrete Random Variables

- Joint probability

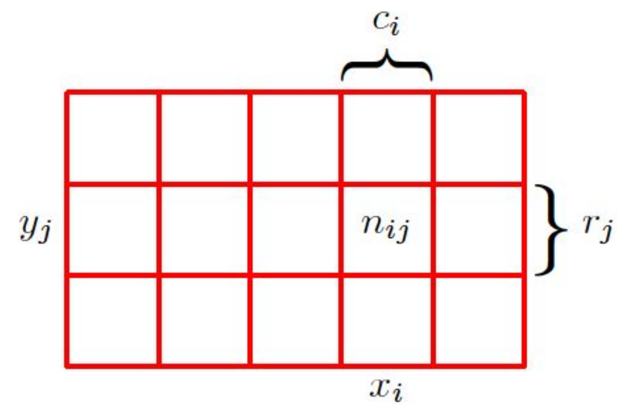


Two Discrete Random Variables

- Joint probability

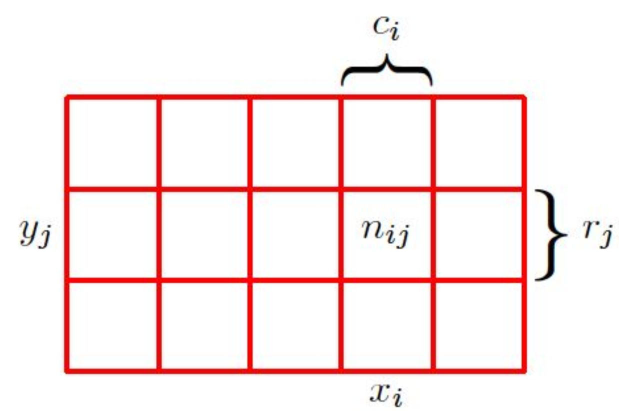
and

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



Two Discrete Random Variables

- If I am interested only on X

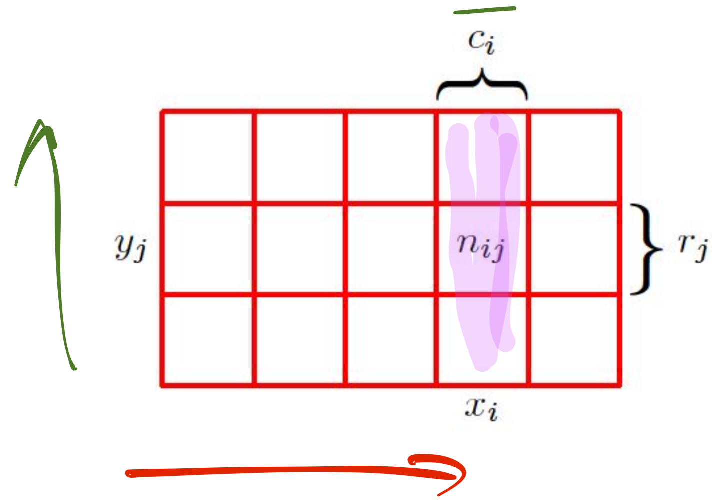


Two Discrete Random Variables

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

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

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

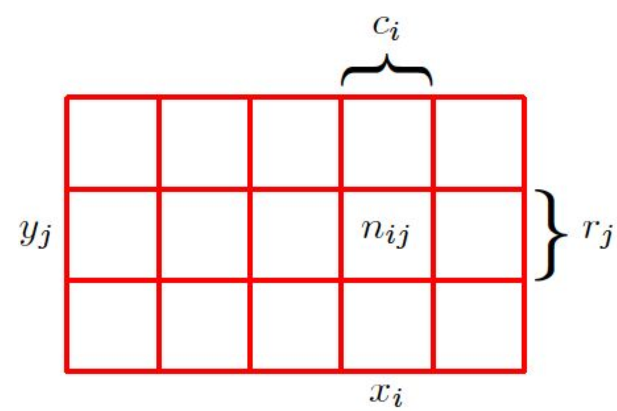


భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Sum rule of Probability

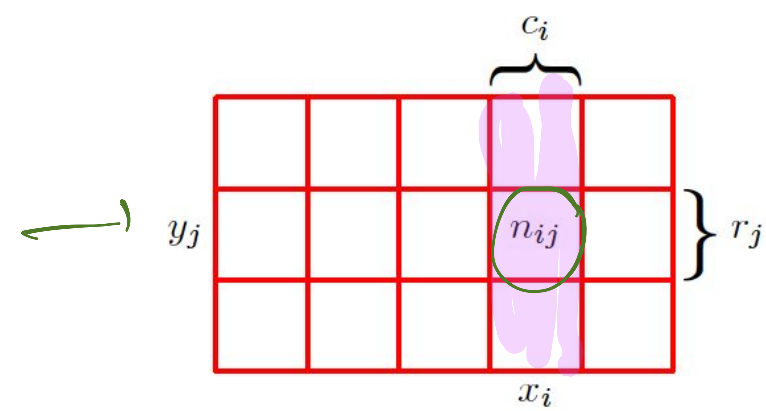
$$p(X = x_i) = \sum_{j=1}^{3} p(X = x_i, Y = y_j)$$



Conditional Probability

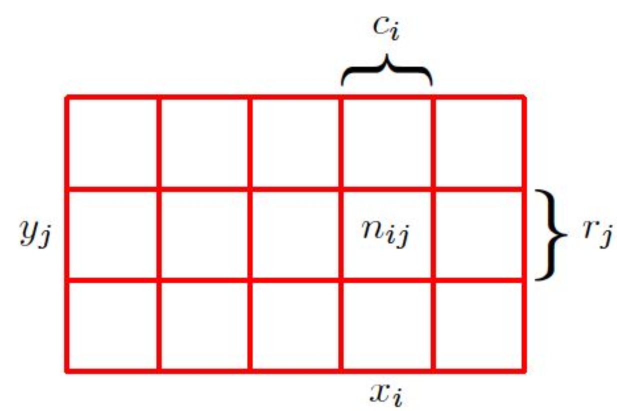
- Conditional probability of Y given X

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



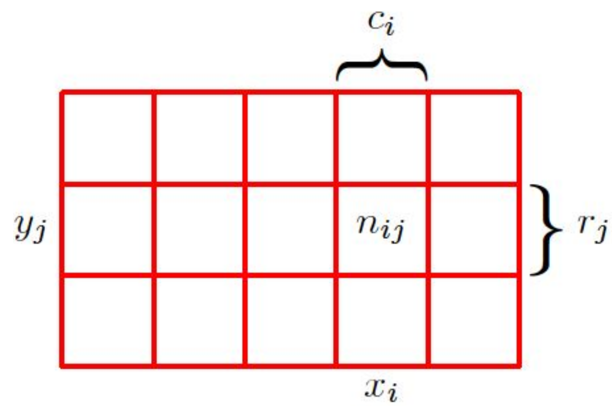
Product Rule of probability

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



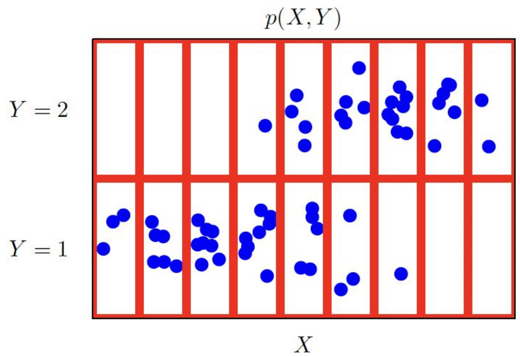
Product Rule of probability

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

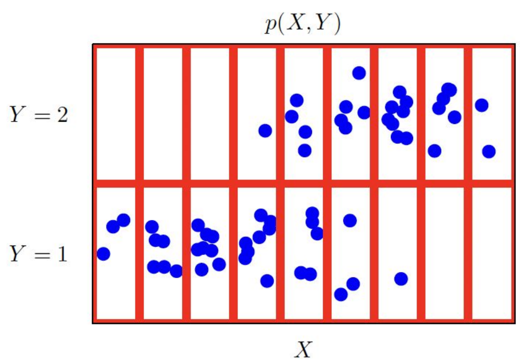


$$p(Y = y_j, X = x_i) = p(Y = y_j / X = x_i) \cdot p(X = x_i)$$

Example: Marginal & Conditional distributions

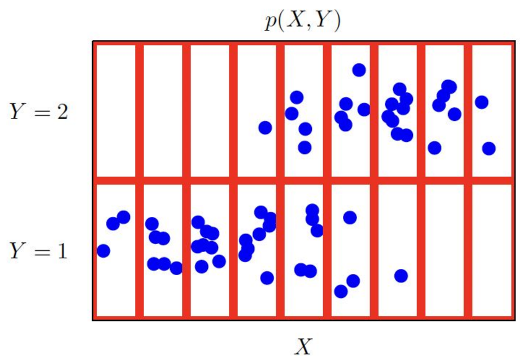


Example: Marginal & Conditional distributions



- X
- Y
- 60 trails - sample both

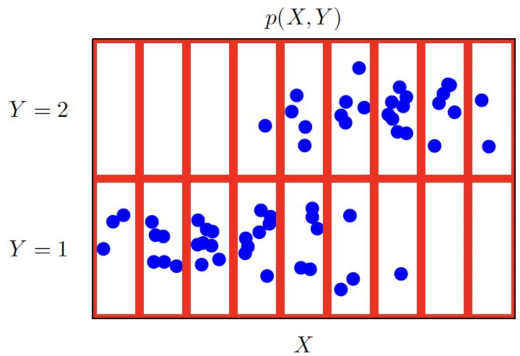
Example: Marginal & Conditional distributions

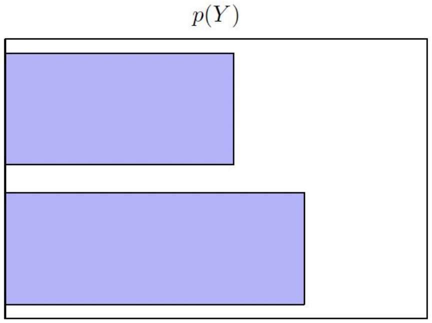


- Marginal distribution p(Y)

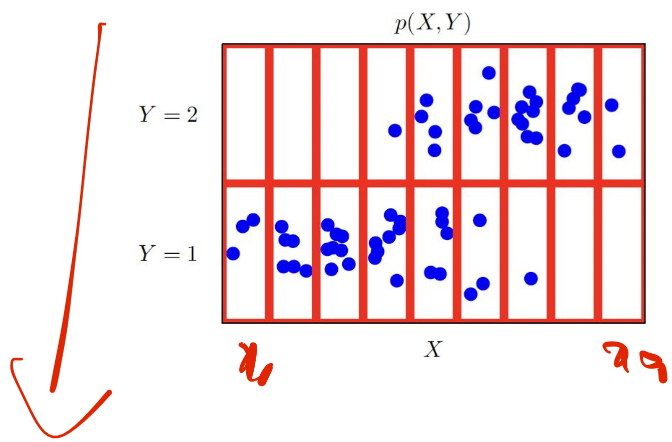
$$P(y) = \sum_{x \in X} p(x, y)$$

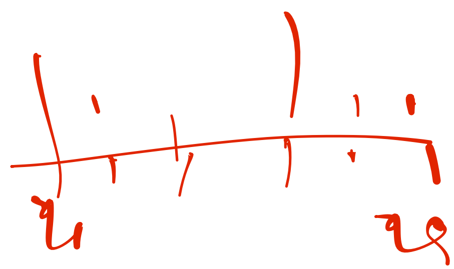
Example: Marginal & Conditional distributions





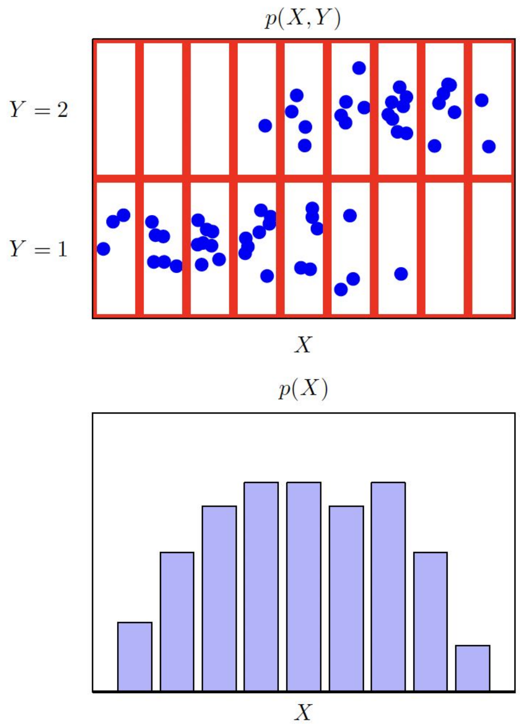
Example: Marginal & Conditional distributions



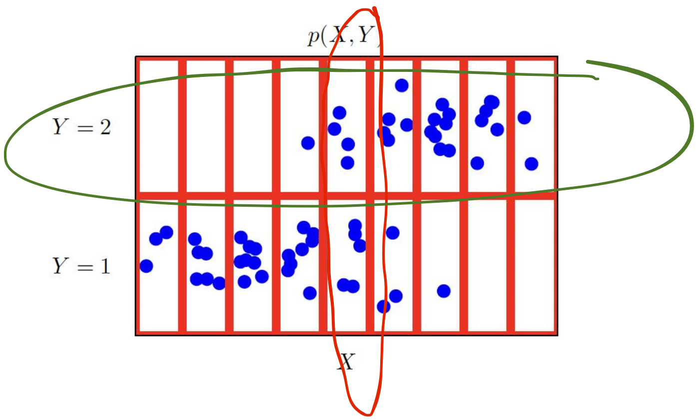


- Marginal distribution p(X)

Example: Marginal & Conditional distributions



Example: Marginal & Conditional distributions

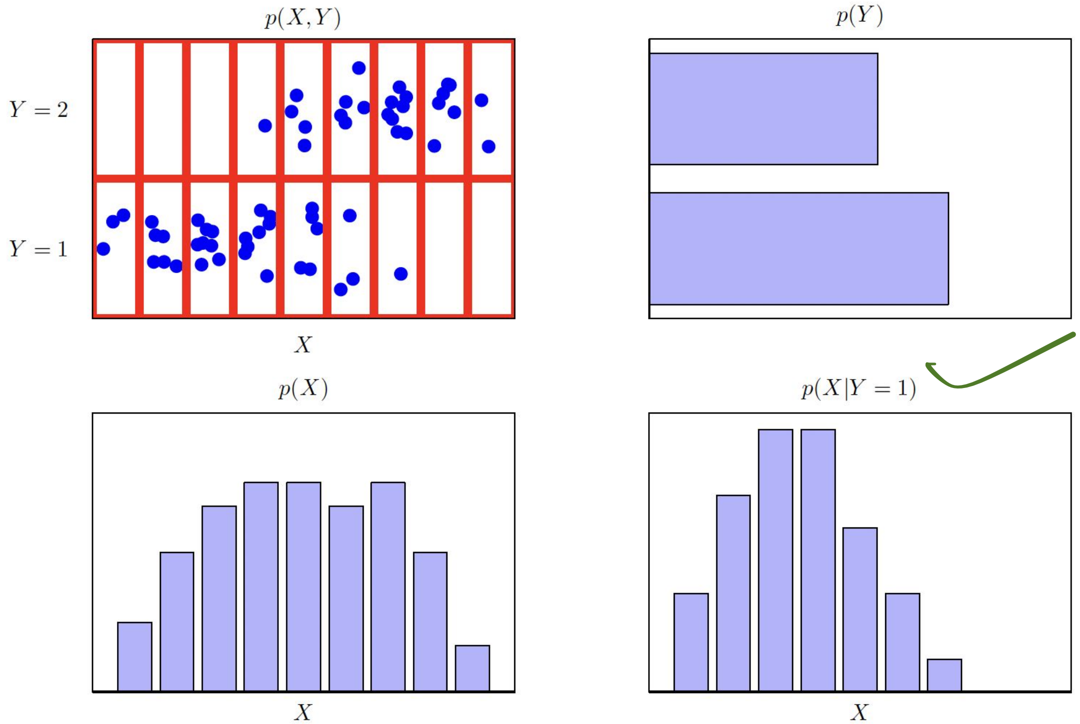


- Conditional distribution of X

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

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

Example: Marginal & Conditional distributions



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

Data-driven Intelligence & Learning Lab

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

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

Data-driven Intelligence & Learning Lab

Continuous Random Variable

- $p(X)$: Probability density over X

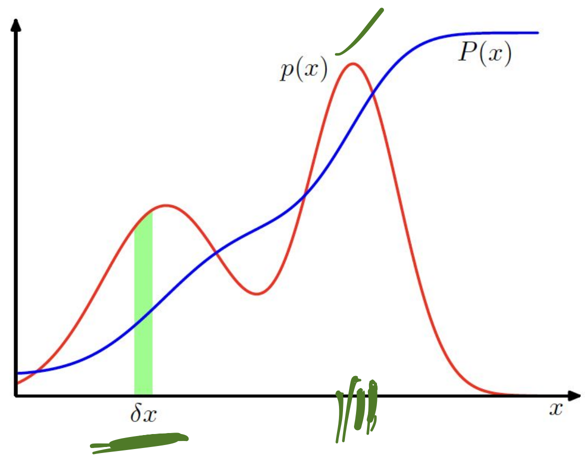
Continuous Random Variable

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

$p(x) \geq 0$

$\int_{x}^{x+dx} p(x)\,dx$

$\int_{a}^{b} p(x)\,dx =$



Continuous Random Variable

- Non-negativity
- Normalization

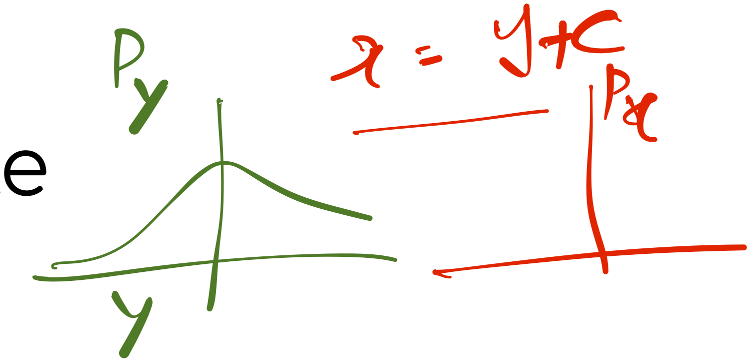
Continuous Random Variable

- Change of variables

Continuous Random Variable

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

Continuous Random Variable



$x = y + c$

$y = 0$

$c = 2$

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

$$P_y(y)\,dy = P_x(x)\,dx$$

$$P_x(x) = P_y(y)\left|\frac{dy}{dx}\right|$$

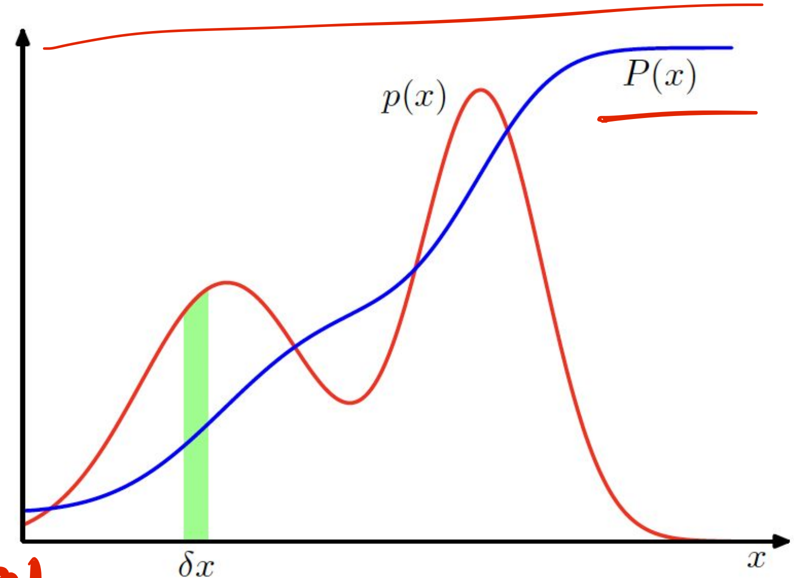
Continuous Random Variable

- Cumulative distribution function

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

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

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



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

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

Data-driven Intelligence & Learning Lab

Bayes Theorem

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

Bayes Theorem

- Product rule
- Symmetry property
- Bayes rule
- Denominator

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

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

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

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

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

$p(x/y)$

Bayes Theorem

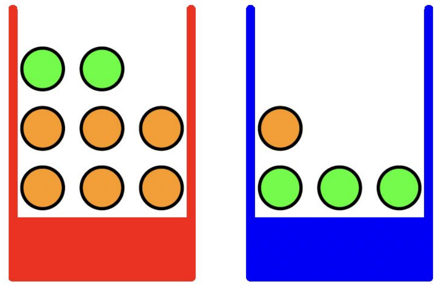
$$p(y/x) = \frac{p(x/y) \cdot p(y)}{p(x)}$$

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

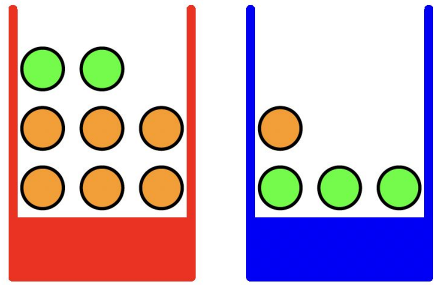
Example

Boxes and Fruits

- Random variables



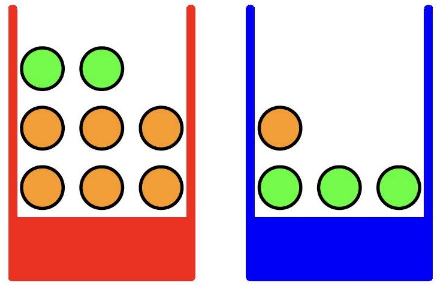
Boxes and Fruits



- Random variables
  - Box - B
  - Fruit - F

Boxes and Fruits

- Prior Box distribution

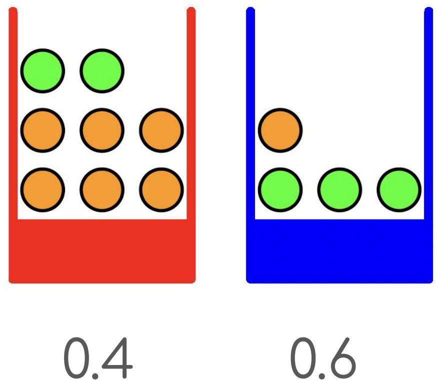


0.4 0.6

Boxes and Fruits

- Prior Box distribution
    -
- Conditional of F given B

- Marginal of F



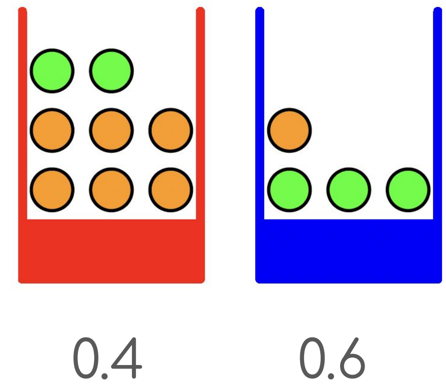
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Boxes and Fruits

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

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



Independence

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

Data-driven Intelligence
& Learning Lab

Independent Random variable

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

Next
Expectation, Variance, and Gaussian Distribution

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

Data-driven Intelligence
& Learning Lab