Machine Learning Foundations

Chapter 6: Probability

Harish Guruprasad Ramaswamy
IIT Madras

Outline for Chapter 6 : Probability

6.1 : Discrete Random Variables

6.2 : Continuous Random Variables

6.3 : Maximum Likelihood and other advanced topics

Outline for Chapter 6 : Probability

6.1 : Discrete Random Variables

1. Probability space
2. Conditioning
3. Random variables
4. **Expectation and Variance**
5. Multiple Random Variables
6. Bernoulli, Binomial, Poisson and Geometric RVs

6.2 : Continuous Random Variables

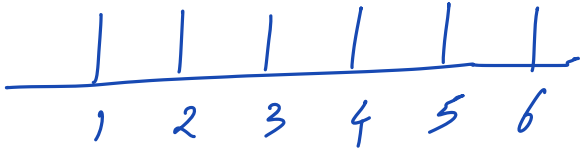
6.3 : Maximum Likelihood and other advanced topics

Expectation: Definition and Properties

$$X : \Omega \to \mathbb{R}$$

$$E[X] = \sum_{x \in \text{Range}(X)} x \cdot f_X(x)$$

e.g. : SDTE



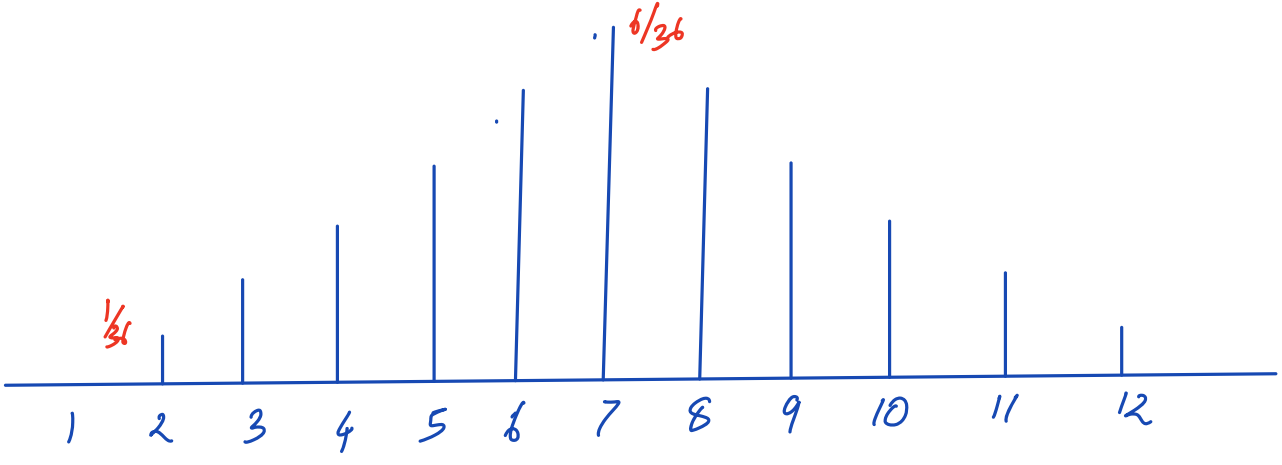
$$EX = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6$$

$$= 3.5$$

Examples

example 2: DDTE.

X = Sum of the faces



$$EX: \frac{1}{36}\cdot 2 + \frac{2}{36}\cdot 3 + \frac{3}{36}\cdot 4 \quad \cdots \quad + \frac{1}{36}\cdot 12$$

$$= 7$$

Examples

Ex: 3 DDTE Y: Abs of diff of dice throw

$f_Y(0) = 6/36$

$f_Y(1) = 10/36$

$f_Y(2) = 8/36$

$f_Y(3) = 6/36$

$f_Y(4) = 4/36$

$f_Y(5) = 2/36$

$$EY = \frac{10 + 16 + 18 + 16 + 10}{36}$$

$$= \frac{70}{36}$$

Conditional Expectation

$(\Omega, \mathcal{F}, P)$

$A \subseteq \Omega$

$X : \Omega \to \mathbb{R}$

$$f_{X|A}(x) = P(X = x \mid A)$$

$$= P(\{\omega : X(\omega) = x\} \mid A)$$

$$E[X|A] = \sum_{x \in \text{Range}(X)} x \, f_{X|A}(x)$$

Examples

Example: DDTE

X = Sum of faces

A : First face = 2

$E[X|A] = ?$

$f_{X|A}(1) = 0$

$f_{X|A}(2) = 0$

$f_{X|A}(3) = 1/6$

$f_{X|A}(4) = 1/6$

$f_{X|A}(5) = 1/6$

$f_{X|A}(6) = 1/6$

$f_{X|A}(7) = 1/6$

$f_{X|A}(8) = 1/6$

$E[X|A] = 5.5$

Examples

example 2:

DDTE

X = Sum of faces

A : Diffesence = 0

$f_{X|A}(1) = 0$

$f_{X|A}(2) = 1/6$

$f_{X|A}(3) = 0$

$f_{X|A}(4) = 1/6$

$f_{X|A}(6) = 1/6$

$E[X|A] = \frac{2}{6} + \frac{4}{6} + \frac{6}{6} + \frac{8}{6} + \frac{10}{6} + \frac{12}{6}$

$= 7$

Linearity of Expectation

X, Y

(i) $E[X+Y] = EX + EY$

(ii) $E[aX] = a\,EX$

Example: X = # of heads in 10 coin toss exp.

$X_i = 1$ (Toss i is heads)

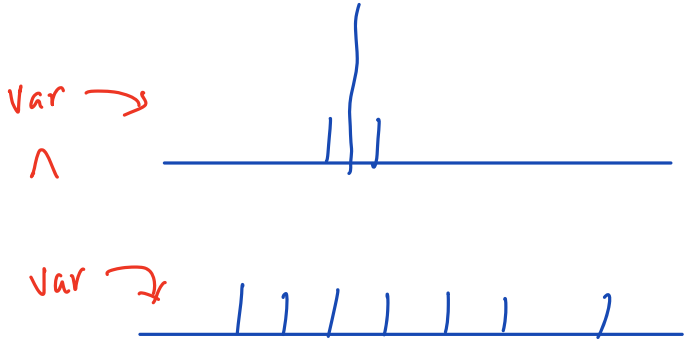
$X = \sum_{i=1}^{10} X_i$ | $EX = \sum_{i=1}^{10} EX_i = 10 \cdot \frac{1}{2} = 5$

Variance: Definition and Properties

$$X : \Omega \to \mathbb{R}$$

$$\begin{aligned}
\mathrm{Var}[X] &= E\left[(X - EX)^2\right] \\
&= E\left[X^2 + (EX)^2 - 2X \cdot EX\right] \\
&= E[X^2] + (EX)^2 - E[2X\, EX] \\
&= E[X^2] + (EX)^2 - 2 \cdot EX \cdot EX \\
&= E[X^2] - (EX)^2
\end{aligned}$$

Variance: Definition and Properties



Var [X] : How far is X from its mean on average

$\Downarrow$

Measure of deviation or spread.

Examples

$$\text{Var}[X] = E[(X - EX)^2]$$

Example: DDTE $\quad X$ = first dice face result.

$$EX = 3.5$$

$$Y = (X - EX)^2$$

X	Y
1	6.25
2	2.25
3	0.25
4	0.25
5	2.25
6	6.25

$$EY = \text{Var}[X] = \left(\frac{1}{6}(6.25) + \frac{1}{6}\cdot(2.25) + \frac{1}{6}(0.25)\right) \times 2$$

$$=$$