Probability

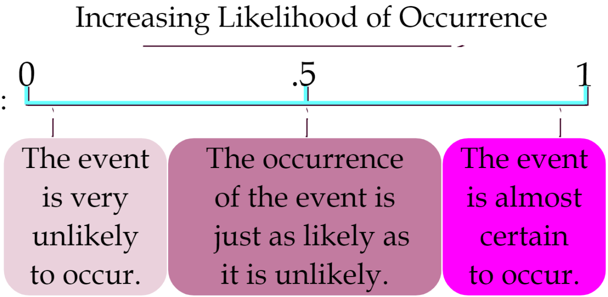
Tuan Nguyen

Ngày 27 tháng 1 năm 2021

Overview

Probability

- Probability:



Sample space and Events

- Sample Space (Ω), all results of an experiment.

  If you toss a coin twice $\Omega = \{HH, HT, TH, TT\}$

- Event: a subset of Ω

  First toss is head $= \{HH, HT\}$

Probability Measure

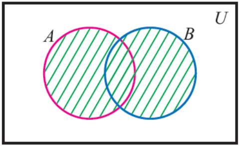
Defined over (Ω, S):

- P(α) >= 0 for all α ∈S. P(Ω) = 1
- If α, β are mutually exclusive then

$$P(\alpha \cup \beta) = P(\alpha) + P(\beta)$$

- In general

$$P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$$



Hình 1: $A \cup B$

Variables

- Deterministic variable

Code

```
int x = 3
float y = 3.14
```

- Stochastic (Random) variable

$$x \sim `\mathcal{N}(0,1)$$

$$x \sim$$

Notation

- The probability distribution over random variable X

$$P(X), P(X = x_i)$$

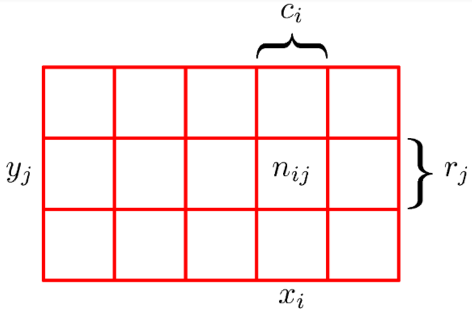
For example, when we roll a dice, X is the number of outcome.

$P(X=1) = P(X=2) = ... = P(X=6) = \frac{1}{6}$

- The probability distribution over random variable X, Y

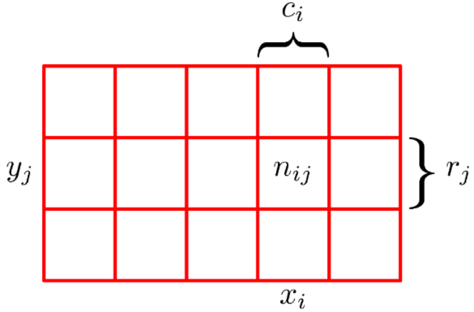
$$P(X, Y), P(X = x_i, Y = y_i)$$

Join distribution



Joint Probability

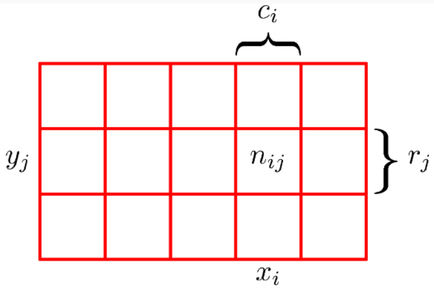
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{\sum_{kl} n_{kl}} = \frac{n_{ij}}{N}$$



Marginal Probability

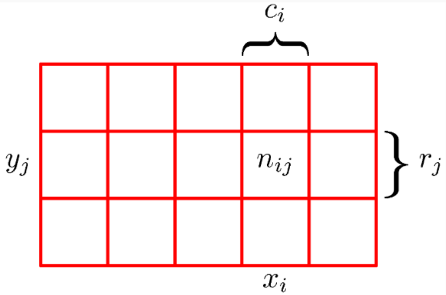
$$p(X = x_i) = \frac{\sum_j n_{ij}}{N} = \frac{c_i}{N}$$

Sum



Sum rule

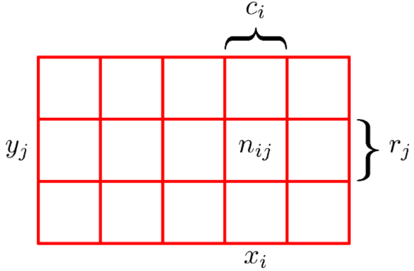
$$p(X = x_i) = \frac{\sum_j n_{ij}}{N} = \sum_j \frac{n_{ij}}{N} = \sum_j p(X = x_i, Y = y_j)$$



Conditional

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

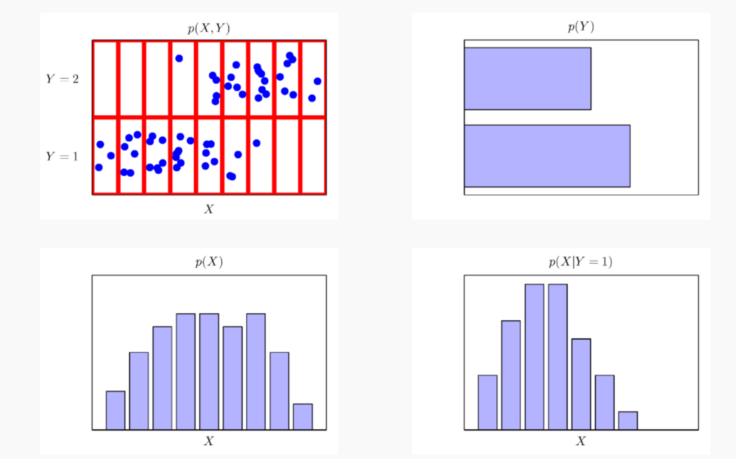
Product rule



Product rule

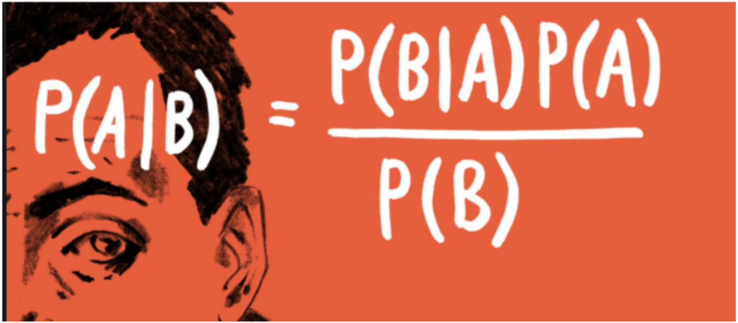
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} = p(Y = y_j | X = x_i) p(X = x_i)$$

Visualization



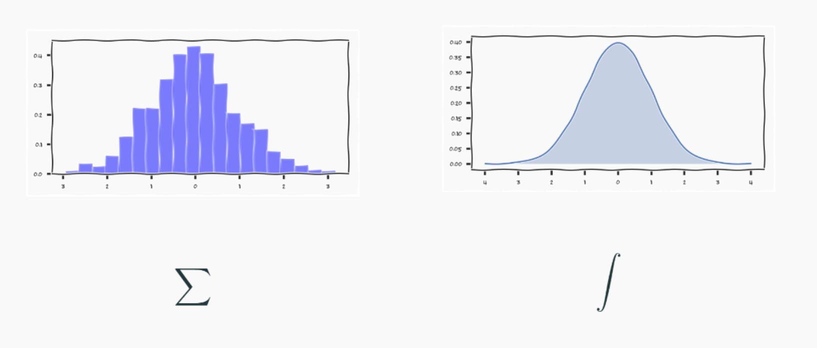
Bayes rule

$$
\begin{aligned}
p(X, Y) &= p(Y|X)p(X) \\
p(X, Y) &= p(X|Y)p(Y) \\
p(X|Y)p(Y) &= p(Y|X)p(X) \\
p(X|Y) &= \frac{p(Y|X)p(X)}{p(Y)} \\
&= \frac{p(Y|X)p(X)}{\sum_X p(Y|X)p(X)}
\end{aligned}
$$



Ex: Có một quick test covid, nếu một người bị covid thì 95% sẽ test ra dương tính, nếu một người không bị covid thì 90% sẽ ra âm tính. Tỉ lệ bị covid ở Việt Nam là 0.01%. Nếu bạn xét nghiệm dương tính, xác suất bạn bị covid là bao nhiêu?

Random variable



Hình 2: Discrete variable vs Continuous variable

Probability Mass Function (PMF)

The probability mass function for X, the number of heads that appear in two tosses of a fair coin

x	$p(x)$
0	0.25
1	0.5
2	0.25

Table 1: Frequency function of X

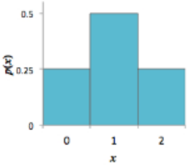
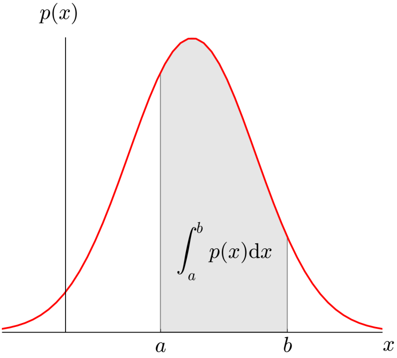


Figure 1: Histogram of X

Hình 3: Probability mass function

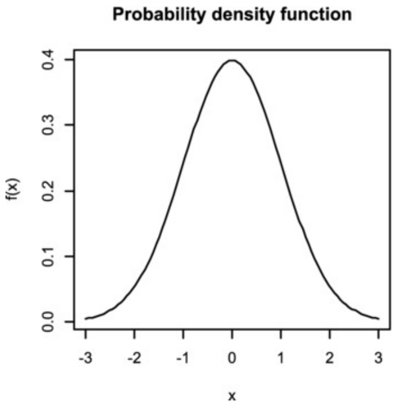
Probability density function (PDF)

$$P(a \leq X \leq b) = \int_a^b f(x)dx \geq 0$$



Hình 4: Probability density function

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



Hình 5: Cumulative probability function

Expectation

The expectation

$$\mu = \mathbb{E}[x]$$

- Discrete random variable

$$\mathbb{E}[x] = \sum_{x} xP(X = x)$$

- Continuous random variable

$$\mathbb{E}[x] = \int_{-\infty}^{+\infty} xf(x)dx$$

Expectation (cont.)

Recall our example: Repair costs for a particular machine are represented by the following probability distribution:

x	\$50	200	350
$P(X = x)$	0.3	0.2	0.5

Variance

The variance

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

- Discrete random variable

$$Var(X) = \sum_x (x - \mu)^2 P(X = x)$$

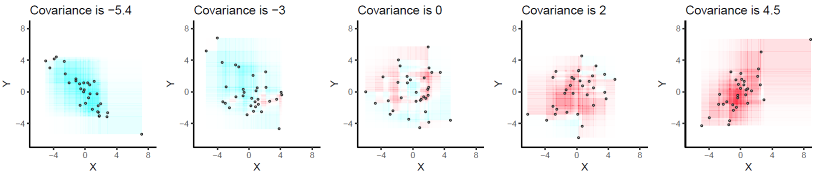
- Continuous random variable

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

Covariance

For two random variables x and y, the covariance is defined by

$$cov[x,y] = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] = \mathbb{E}[xy] - \mathbb{E}[x]\,\mathbb{E}[y]$$



Hình 6: Covariance of x and y

The correlation between two random variables X; Y is given by

$$corr[x,y] = \frac{cov[x,y]}{\sqrt{Var(X)Var(Y)}} \leq 1$$

Sums and Transformations of Random Variables

Consider two random variables X,Y with states x,y$\in \mathcal{R}^D$

- $\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y]$
- $\mathbb{E}[x - y] = \mathbb{E}[x] - \mathbb{E}[y]$
- $V[x + y] = V[x] + V[y] + cov[x, y] + cov[y, x]$
- $V[x + y] = V[x] + V[y] - cov[x, y] - cov[y, x]$