Basics of Probability and Random Processes

Pattern Recognition and Machine Learning, Jul-Nov 2019

Indian Institute of Technology Madras

August 18, 2019

Definition of Probability

Probability of an event E

The probability of an event E is defined as the relative frequency of outcomes favourable to E; to the total number of outcomes in the sample space S of the experiment.

- Let the experiment be repeated N_S times.
- Let N_E be the outcomes favourable to event E.
- The probability of the event E is given by

$$P(\mathrm{E}) = \frac{N_E}{N_S} \tag{1}$$

Probability of sample space

- Let there be a total of N events in the sample space S, namely $A_1, A_2, \ldots, A_N$.
- $A_1, A_2, \ldots, A_N$ form a partition of the sample space as in Figure 1

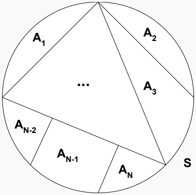

Figure 1: Sample space S partitioned by the events $A_1, A_2, \ldots, A_N$.

- The probability of the sample space S is given by

$$P(S) = \sum_{n=1}^{N} P(A_n) = 1 \tag{2}$$

Joint Probability

- The joint probability distribution of events A and B is given by

$$P(\mathrm{A},\mathrm{B}) = P(\mathrm{A} \cap \mathrm{B}) = \frac{N_{\mathrm{A}\cap\mathrm{B}}}{N_{\mathrm{S}}} \tag{3}$$

Conditional Probability

- The conditional probability is defined as the probability of occurrence of an event A given that an event B has occurred.

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{B})} = \frac{N_{\mathrm{A}}/N_{\mathrm{S}}}{N_{\mathrm{B}}/N_{\mathrm{S}}} \tag{4}$$

Bayes' Rule

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{B})} \tag{5}$$

$$P(\mathrm{B}|\mathrm{A}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{A})} \tag{6}$$

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}|\mathrm{B}) \times P(\mathrm{B}) = P(\mathrm{B}|\mathrm{A}) \times P(\mathrm{A}) \tag{7}$$

$$P(\mathrm{B}|\mathrm{A}) = \frac{P(\mathrm{A}|\mathrm{B}) \times P(\mathrm{B})}{P(\mathrm{A})} \tag{8}$$

Equation 8 is called as Bayes's rule.

Marginalization

- Let $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_N$ form a partition of the sample space S
- Along with the events $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_N$, let B be an event defined in sample space S as in Figure 3

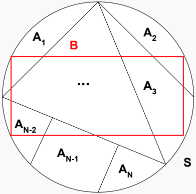

Figure 2: Sample space S with events $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_N$ and B.

- The marginalization of the event B over the events $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_N$ is given by

$$P(\mathrm{B}) = \sum_{n=1}^{N} P(\mathrm{A}_n, \mathrm{B}) = \sum_{n=1}^{N} P(\mathrm{B}|\mathrm{A}_n) P(\mathrm{A}_n) \tag{9}$$

Realization of Bayes' rule for classification problem

- Let A_i be one of the classes and $\bar{\mathrm{x}}$ be a feature vector. From Bayes' rule we have that

$$\overbrace{P(\mathrm{A}_i|\bar{\mathrm{x}})}^{\textit{Posterior}} = \frac{\overbrace{P(\bar{\mathrm{x}}|\mathrm{A}_i)}^{\textit{likelihood}} \times \overbrace{P(\mathrm{A}_i)}^{\textit{Prior}}}{\underbrace{P(\bar{\mathrm{x}})}_{\textit{Evidence}}} \tag{10}$$

Problem - 1

Consider an experiment of tossing a coin two times. The possible outcomes are $\{HH, HT, TH, TT\}$. Based on this, outcomes to the following events can be :

- Event A : First coin comes Head : $\{HT, HH\}$
- Event B : Atleast one head : $\{HT, TH, HH\}$

Also, notice that event A+B (First coin comes head OR at least one head) is same as event B : $\{HT, TH, HH\}$.

The event AB (First coin comes head AND at least one head) is same as event A : $\{HT, HH\}$

Problem - 2

An urn consists of a white balls and b black balls.

- Probability of picking a white ball : $\left(\frac{a}{a+b}\right)$
- Probability of picking two white balls : $\left(\frac{a}{a+b}\right)\left(\frac{a-1}{a+b-1}\right)$
- Probability of picking two white balls in sequence and with replacement : $\left(\frac{a}{a+b}\right)^2$

Problem - 3

9 instruments are on a shelf. All are brand new. In an experiment, we pick 3 instruments, use them and replace them. What is the probability of repeating the experiment 3 times and no new instrument being left on the shelf?

- Let event A_1 : Probability that in the first experiment, all new instruments are picked : 1 (because all are new initially)
- Let event A_2 : Probability that in the second experiment, all new instruments are picked : $\left(\frac{6}{9}\right)\left(\frac{5}{8}\right)\left(\frac{4}{7}\right)$ (because only 6 new instruments are left after 1^{st} experiment)
- Let event A_3 : Probability that in the third experiment, all new instruments are picked : $\left(\frac{3}{9}\right)\left(\frac{2}{8}\right)\left(\frac{1}{7}\right)$ (because only 3 new instruments are left after 2 experiments)

Answer $= \mathrm{P}(A_1).\mathrm{P}(A_2).\mathrm{P}(A_3)$

Problem - 4

3 identical urns are there. Urn U_1 has a white balls and b black balls. Urn U_2 has c white balls and d black balls. Urn U_3 has d white balls and 0 black balls.

- What is the probability of choosing a white ball (event A)?
  - Probability of choosing a white ball from urn 1 : $\mathrm{P}(U_1 A) = \left(\frac{1}{3}\right)\left(\frac{a}{a+b}\right)$ (because first we will choose a urn out of the 3 urns and then choose a white ball from the chosen urn)
  - Similarly, $\mathrm{P}(U_2 A) = \left(\frac{1}{3}\right)\left(\frac{c}{c+d}\right)$ and $\mathrm{P}(U_3 A) = \left(\frac{1}{3}\right)\left(\frac{d}{0+d}\right)$
  - $\mathrm{P(A)} = \sum_{i=1}^{3} P(U_i A)$ (marginalizing over the urns, because we don't care which urn we choose the white ball from)

- What is the probability that a ball is chosen from U_1 given that it is white?

  $\mathrm{P}(U_1/\mathrm{A}) = \frac{P(A/U_1)P(U_1)}{P(A)} = \frac{\left(\frac{a}{a+b}\right)\cdot\left(\frac{1}{3}\right)}{P(A)}$

Problem - 5

Market share of 3 car manufacturer (M_1, M_2 and M_3) is $20\%, 30\%$ and 40% respectively. The probability that their car requires major repair in 1^{st} year is $5\%, 10\%$ and 15% respectively.

- What is the probability that a car requires major repair in 1^{st} year (event A)?
  - Since the car can be of any manufacturer, we will add the probability of car being faulty of each manufacturer.
  - $\mathrm{P(A)} = \mathrm{P}(\mathrm{A}/M_1)\mathrm{P}(M_1) + \mathrm{P}(\mathrm{A}/M_2)\mathrm{P}(M_2) + \mathrm{P}(\mathrm{A}/M_3)\mathrm{P}(M_3)$ $= 0.05 \times 0.2 + 0.1 \times 0.3 + 0.15 \times 0.5$

- What is the probability that the faulty car is of M_1 manufacturer?
  - Same as last question.
  - $\mathrm{P}(M_1/\mathrm{A}) = \frac{P(A/M_1)P(M_1)}{P(A)} = \frac{0.05 \times 0.2}{P(A)}$

Problem - 6

Assume that a test to detect a disease whose prevalence is $\frac{1}{100}$ has a False Positive rate of 8% and a True Positive Rate of 100%. What is the probability that a person who is found to test positive actually has the disease?

Let us assume the following events :
D : Person has the disease. $\overline{D}$: Person doesn't have the disease.
T : Person is tested positive. $\overline{T}$: Person is tested negative.

We can see that we need to find $P(D/T)$.
P(D/T) = $\frac{P(T,D)}{P(T)} = \frac{P(T/D)P(D)}{P(T)}$

To find $P(T)$, we use

$$\begin{aligned} P(T) &= P(T, D) + P(T, \overline{D}) \quad \text{(marginalizing over joint)} \\ &= P(T/D)P(D) + P(T/\overline{D})P(\overline{D}) \\ &= 1 \times 0.01 + 0.08 \times 0.99 \end{aligned}$$

Cumulative distribution

Open and closed intervals

$$\text{Closed interval} : [a, b] = \{x : a \leq x \leq b\}$$
$$\text{Half open interval} : (a, b] = \{x : a < x \leq b\}$$
$$\text{Open interval} : (a, b) = \{x : a < x < b\}$$

Cumulative distribution function (CDF)

Cumulative distribution function of a random variable (RV) X is

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

where, $P(X \leq x)$ is the probability that the RV X takes a value less than or equal to x

$$F_X(-\infty) = 0 \implies \text{The event is impossible to occur}$$
$$F_X(\infty) = 1 \implies \text{The event is certain to occur}$$

Cumulative distribution

Probability from CDF

Probability of a random variable X in the given interval can be obtained from CDF

$$P(X \in (-\infty, b)) = P(X \in (-\infty, a]) + P(X \in [a, b])$$
$$P(X \in [a, b)) = P(X = a) + F_X(b) - F_X(a)$$

CDF of a continuous RV

$$F_X(x) = \int_{-\infty}^{x} f_X(x).dx$$

where, $f_X(x)$ is the density function of the RV X

Example 1

A particle is moving along a straight line. The instantaneous velocity is given by

$$V = \frac{dx}{dt}$$

The particle is described by its position and velocity (x_0, v_0)
Kinetic energy of the particle is given by

$$K = \frac{1}{2} m V^2$$

X and V are the random variables correspond to position and velocity respectively. Given the density function $f_V(v)$, what is $f_K(k)$?

Example 1

The density function of the RV V is given by
$F_K(b)$ be the probability that the RV K has takes a value $\leq$ b

$$
\begin{aligned}
B &= \{K : -\infty < k \leq b\} \\
\phi = g^{-1}(b) &= \left\{ -\sqrt{\frac{2b}{m}} \leq v \leq \sqrt{\frac{2b}{m}} \right\} \\
P(v \in g^{-1}(b)) &= F_K(b) \\
&= P\left[v = \sqrt{\frac{2b}{m}}\right] + F_v\left(\sqrt{\frac{2b}{m}}\right) - F_v\left(\sqrt{\frac{2b}{m}}\right) \\
f_K(k) &= f_v\left(\sqrt{\frac{2k}{m}}\right)\frac{\partial}{\partial k}\left(\sqrt{\frac{2k}{m}}\right) + f_v\left(\sqrt{\frac{-2k}{m}}\right)\frac{\partial}{\partial k}\left(\sqrt{\frac{2k}{m}}\right) \\
&= \frac{f_v\left(\sqrt{\frac{2k}{m}}\right) + f_v\left(\sqrt{\frac{-2k}{m}}\right)}{\sqrt{2mk}}
\end{aligned}
$$

General case

$$Y = \phi(X)$$

$$\chi(Y) = \phi^{-1}(Y)$$

$$f_Y(y) = \sum_{i=1}^{K} f_X(\chi_i(y)) \cdot |\chi_i{}'(y)|$$

According to the above equation, in Example 1,

$$K = \frac{1}{2}mV^2$$

$$\chi_1(k) = \sqrt{\frac{2k}{m}}$$

$$\chi_2(k) = \sqrt{\frac{-2k}{m}}$$

$$|\chi_1^{'}| = |\chi_2^{'}| = \frac{1}{\sqrt{2mk}}$$

Statistical Averages

Ensemble average

It is the average of the outcomes of a stochastic process at a given instance of time.

- The ensemble average is often called as mean or expectation. It is given as:

$$\mu_x(t) = \mathrm{E}[\mathrm{X}(\mathrm{t})]$$

$$= \int_{-\infty}^{\infty} x(t) f_x(x(t)) dx(t)$$

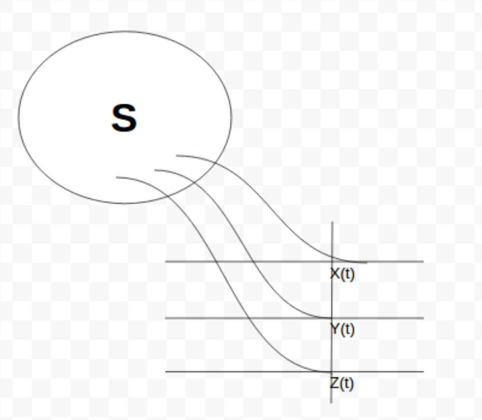

Figure 3: Sample space S with mapping X,Y and Z

Here, ensemble average at time instance 't' is the average of X(t),Y(t) and Z(t).

Time average

It is the average value of a single outcome of a stochastic process across time.

- Time average is given as:

$$\mu_x = \mathrm{E[X]}$$

$$= \int x f_x(x) dt$$

Ergotic process

Definition of ergodic process

A process is ergodic if its ensemble average is equal to its time average.

That is, if $\mu_x = \mu_t$, the process is ergodic.

Moments about the origin

- The n^{th} moment about the origin is given as:

$$E[X^n] = \int x^n f_x(x)dx$$

Expectation is the first order moment.

- The n^{th} moment about the mean is given as:

$$E[(X-\mu_x)^n] = \int (x-\mu_x)^n f_x(x)dx$$

- Auto-covariance:

$$E[(X-\mu_x)^2] = \int (x-\mu_x)^2 f_x(x)dx$$

It is the second order moment

Autocorrelation

Definition

Autocorrelation is the correlation of the same process with itself at different instances of time.

- It gives the similarity of a random process at different time instances t_1 and t_2.
- It is given as: $R_x(|t_2 - t_1|) = \mathrm{E}(\mathrm{X}(t_1),\mathrm{X}(t_2)]$

Stationary process

Wide sense stationarity

A random process X(t) is said to be wide sense stationary if the following conditions are satisfied:

- Expectation, that is $\mu_x = E(X(t)]$ is independent of time.
- Autocorrelation is only a function of time lag t_2-t_1.
  - $R_X(|t_1 - t_2|) = R_X(|t_2 - t_1|)$
  - $R_X(|t_2 - t_1|) = R_X(|t_4 - t_3|)$, if $t_2 - t_1 = t_4 - t_3$.

Strict sense stationarity

A random process X(t) is said to be strict sense stationary if the above conditions are satisfied for all higher order moments.

Problem

A station attempts to transmit a packet. Let p be the probability that it will not collide. Find the expected number of collisions for successful packet transmission.

Ans:

X	0	1	2	3	...	n
p(X)	p	pq	q^2p	q^3p	...	q^np

This is an ordered series.

$$E[X] = \sum x p_x(x)$$

$$= \sum_{i=0}^{\infty} ipq^i$$

$$= 0 + pq + 2pq^2 + 3pq^3 + ...$$

$$= p[q + 2q^2 + 3q^3 + ...]$$

Let $S = q + 2q^2 + 3q^3 + ...$

$$Sq = q^2 + 2q^3 + 3q^4 + ...$$

$$S(1-q) = q + q^2 + q^3 + ...$$

$$S = \frac{q}{(1-q)^2}$$

Hence, $E[X] = \frac{q}{p}$

$$Var(X) = E[(X - \mu_x)^2]$$

$$= E[X^2] - (E[X])^2$$

$$= E[X^2] - (\tfrac{q}{p})^2$$

- Find $E[X^2]$.

  Ans: $E[X^2] = pq \sum_{k\geq 1}^{\infty} k^2 q^{k-1}$

  Hint: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$
  Take derivative on both sides.

Cross co-variance

Cross co-variance

Cross-covariance of two random variables X and Y is given as:
$Cov(X, Y) = E[(X(t) - \mu_x)(Y(t) - \mu_y)]$

Correlation coefficient

Correlation coefficient for two random variables X and Y is given as:
$S_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}; \quad S_{XY} \leq 1$

Some properties

- $Cov(X, Y) = Cov(Y, X)$
- $X = Y \implies S_{XY} = 1$
- X and Y are independent $\implies E[X] = E[Y]$
- X and Y are independent $\implies$ $Cov(X, Y) = E[XY] - E[X]E[Y] = 0$

Some properties

- $E[aX] = aE[X]$
- $E[a_1X_1 + a_2X_2 + ...] + b = a_1E[X_1] + a_2E[X_2] + ..$
- $Var(aX) = a^2Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- X and Y are independent $\implies$
  $Var(X + Y) = Var(X) + Var(Y)$