Prob Stochastics

Saad Ansari

January 2022

1 Introduction to Probability

1.1 Basic Definitions

- Random Experiment (denoted by E): An experiment whose outcome cannot be predicted in advance.
- Sample Space (denoted by Ω): The collection of all possible outcomes of the random experiment E.
- Event (denoted by any capital letter): A subset of the sample space, which might be favourable to us.
- Power set: Set of all subsets of a *finite* sample space.
- Sigma field (denoted by f, generally): Any collection of *subsets* of Ω which satisfies:
  - $\Omega \in f$
  - If $A \in f$, then $\bar{A} \in f$.
  - If $A_1, A_2 \in f$, then $A_1 \cup A_2 \in f$.
- Probability (denoted by P(A)): A real number relating to an event A, which satisfies:
  - $P(\Omega) = 1$, because $\Omega \in f$
  - $P(A) \geq 0$, if $A \in f$
  - If $A_1, A_2, \ldots$ are mutually exclusive, then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

  Note: Probabilities can also be defined as functions that map all entries in a sigma field to a number between 0 and 1. Usually, the sigma field is taken to be the power set of the sample space.

$$P : f \rightarrow [0, 1]$$

1.2 Basic Properties of Probability

- Calculation of probability:
$$P(A) = \frac{n(A)}{n(\Omega)}$$

- $P(\phi) = 0$

- Union: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Generalised Union (JEE Inclusion-Exclusion Principle): If $A_i \in f$, for $i = 1, 2, 3...n$, then
$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \; - \sum_{i,j=1,i<j}^{n} P(A_i \cap A_j) \; + \; \cdots \; + \; (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

- Monotonicity: A sequence of events $\{A_n\}_{n=1}^{\infty}, A_n \in f$ are monotonically increasing if, for all n, $A_n \subseteq A_{n+1}$. For a monotonically increasing sequence,
$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$$
and
$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

- Similarly, a monotonically decreasing sequence is one in which $A_n \supseteq A_{n+1}$ holds true, along with
$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$
and
$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

1.3 Conditional Probability

- Conditional Probability: For two events A and B, if P(B) ¿ 0,
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Multiplication Rule: The probability that events $A_1, A_2, A_3, \ldots A_n$ occur together is
$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$
provided that $P(A_1) > 0$, $P(A_1 \cap A_2) > 0$, and so on.

- Independence: Events $A_1, A_2, \ldots A_n$ are mutually independent if and only if the probability of intersection of any $2, 3, \ldots, n$ of these sets is the product of their respective probabilities. For $r = 2, 3, \ldots, n$,
$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_r})$$

- Total Probability Theorem: If events $E_1, E_2, \dots E_n$ are mutually exclusive and exhaustive, then for an event A,

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$$

where $P(E_i) > 0$, for all $i = 1, 2, \dots, n$.

- Bayes' Theorem: If events $E_1, E_2, \dots E_n$ are mutually exclusive and exhaustive, then for an event A,

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{j=1}^{n} P(A|E_j)P(E_j)}$$

where $P(A) > 0$, and $P(E_i) > 0$, for all $i = 1, 2, \dots, n$.

Note: The Total Probability and Bayes' theorems are especially useful when event A is *dependent* on at least one of $E_1, E_2, \dots E_n$. The results are trivially true if A is independent.

1.4 Random variables and Distribution functions

Let (Ω, f, P) be a probability model. A real valued function X defined on sample space Ω is a random variable if:

$$\forall x \in \mathbb{R},\ \{\omega : X(\omega) \leq x\} \in f$$

or

$$X^{-1}((-\infty, x]) \in f$$

or that $X^{-1}((-\infty, x])$ is an event.

Example: For two coin tosses, $\Omega = \{HH, HT, TH, TT\}$. Taking f to be the power set of the sample space (Ω) and X as the *number of heads*, we have

$$X^{-1}((-\infty, x]) = \begin{cases} \phi, & x < 0 \\ \{TT\}, & 0 \leq x < 1 \\ \{TT, HT, TH\}, & 1 \leq x < 2 \\ \Omega, & x \geq 2 \end{cases}$$

Thus, X can be called a random variable as it has assigned a value for

Random variables can either be discrete or continuous.

- Discrete Random Variable: X is a discrete random variable if we can associate a number $p_X(x) = P(X = x)$ with each outcome x in the range space R_X (sigma field f, which is the input space or domain space in case of probability), such that $0 \leq P(X = x) \leq 1$ and $\sum_{x \in R_x} P(X = x) = 1$

  $(x, p_X(x)), x \in R_X$ is called a probability distribution and $p_X(x)$ is called the probability mass function (PMF).

- Continuous Random Variable: X is a continuous random variable if a probability density function (PDF) $f(x)$ can be associated with it such that:

$$f(x) \geq 0,\ \forall x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$P(a \leq x \leq b) = \int_{a}^{b} f(x)dx,\ -\infty < a < b < \infty$$

- Cumulative Distribution Functions (CDF): The CDFR is defined as $F_X(x) = P(X \leq x)$
  For discrete variables, $F_X(x) = \sum_{x_i \leq x} p_X(x_i)$, where $p(x_i)$ is the PMF.
  For Continuous variables, $F_X(x) = \int_{-\infty}^{x} f(u)du$, where $f(x)$ is the PDF.
  For any CDF,
  - $\forall x, 0 \leq F_X(x) \leq 1$
  - $F_X(x)$ is non-decreasing.
  - $F_X(x)$ is right-continuous.
  - $\lim_{x\to\infty} F_X(x) = 1$ and $lim_{x\to-\infty} F_X(x) = 0$
  -

1.5 Expectations and (Bruh) Moments

If X is a random variable with probability distribution $(x, p(x))$ and

$$\sum_{i=1}^{\infty} |x_i| p(x_i) = \int_{-\infty}^{\infty} |x| f(x)dx < \infty$$

Then, the expected value of X exists, and is denoted by μ

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i p(x_i) = \int_{-\infty}^{\infty} x f(x)dx$$

The variance of X, denoted by σ, also exists

$$\sigma^2 = E(X-\mu)^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 p(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = E(X^2) - E(X)^2$$

Moments about the origin are defined as:

$$\mu_r^{'} = E(X^r) = \sum_{i=1}^{\infty} (x_i)^r p(x)dx = \int_{-\infty}^{\infty} x^r p(x)$$

$$\mu = \mu_1^{'}$$

Moments about the mean (μ) are defined as:

$$\mu_r = E(X-\mu)^r = \sum_{i=1}^{\infty} (x_i - \mu)^r p(x)dx = \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx$$

$$\sigma^2 = \mu_2$$

Moment Generating Function (MGF): A function that gives $\mu_r^{'}$ when differentiated r times.

$$\begin{aligned} M_X(t) &= E(e^{tX}) = E(1 + tX + \frac{t^2X^2}{2!} + \cdots) \\ &= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots \end{aligned} \qquad \therefore \frac{d_r}{dt^r}M_X(t)|_{t=0} = E(X^r) = \mu_r^{'}$$

Also, Expectations are linear, and Moments are linear too, but in a weird way:

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2Var(X)$$

Thus, for a random variable $Z = \frac{X-\mu}{\sigma}$,

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = 0 \qquad Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X) = 1 \tag{1}$$

This is called a standardised random variable, and this process is called standardisation.

Chebyshev's inequality: For any random variable X with mean μ and variance σ^2, for any $k > 0$,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

2 Probability Distributions

2.1 Discrete Distributions

Consider a trail in which there can be either success (with probability p), and failure (with probability q) such that $p + q = 1$. Let X be a random variable that counts the number of successes. For a single instance, X can be either 0 or 1. This is called a Bernoulli trail, denoted by $X \quad Ber(p)$. Some properties of a Bernoulli trail:

$$PMF \longrightarrow P(X = 1) = p, P(X = 0) = q$$

$$MGF \longrightarrow M(t) = q + pe^t$$

$$E(X) = p$$

$$Var(X) = E(X^2) - E(X)^2 = p - p^2 = pq$$

The Bernoulli trail can also be understood as a *single instance* of any event, which has two outcomes, a success and a failure. Now let's look at some distributions based on the Bernoulli trail.

- **Binomial Distribution** ($X \quad Bin(n,p)$): Let n independent Bernoulli trails be conducted. If X counts the number of successes in n trails, then X can take the values $0, 1, 2, \ldots n$.

$$P(X = x) = p^x q^{n-x}, \quad \text{if } x = 0, 1, \ldots n, \text{ and } 0, \text{ otherwise}$$

$$MGF \longrightarrow M(t) = (q + pe^t)^n$$

$$E(X) = np, E(X^2) = n(n-1)p^2 + np, Var(X) = npq$$

- **Geometric Distribution** ($X \quad Geo(p)$): Let independent Bernoulli trails be conducted until there is a success. If X counts the number of trials to get a success, then X can take the values $1, 2, \ldots \infty$.

$$P(X = x) = \binom{n}{x} q^{x-1} p, \quad \text{if } x = 1, 2, 3, \ldots \infty, \text{ and } 0, \text{ otherwise}$$

$$MGF \longrightarrow M(t) = \frac{pe^t}{1 - qe^t}$$

$$E(X) = 1/p, Var(X) = \frac{q}{p^2}$$

- **Negative Binomial Distribution** ($X \quad NB(r,p)$): Let independent Bernoulli trails be conducted until there are r successes. If X counts the number of trials to get r successes, then X can take the values $r, r+1, \ldots \infty$.

$$P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad \text{if } x = r, r+1, \ldots \infty, \text{ and } 0, \text{ otherwise}$$

$$MGF \longrightarrow M(t) = (\frac{pe^t}{1 - qe^t})^r$$

$$E(X) = r/p, Var(X) = \frac{rq}{p^2}$$

- **Hypergerometric Distribution**: Let a box contain N balls, M of which are marked. If n balls are drawn at random from the box, and the marked balls are counted, then

$$P(X = x) = \frac{\binom{N-M}{n-x}\binom{M}{x}}{\binom{N}{n}}, \quad \text{if } max(0, M + n - N) \leq x \leq min(M, n)$$

$$E(X) = \frac{n}{N} M$$

$$Var(X) = \frac{nM}{N^2(N-1)}(N - M)(N - n)$$

Note: If a function f exists, such that $\lim_{h \to 0} \frac{f(h)}{h} = 0$, then $f = o(h)$ and any linear combination of any number of $o(h)$ functions is $o(h)$.

2.2 Poisson Process

Let infinitely many Binomial trials be conducted per unit time, $n \to \infty$, but let the probability p of each trial succeeding be infinitesimal such that $\lambda = np$ is finite. So, $p = \frac{\lambda}{n}$, and probability distribution for t units of time is:

$$P(X = k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \cdot (\frac{\lambda}{n})^k \cdot (1 - \frac{\lambda}{n})^{n-k} \tag{2}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{n!}{n^k(n-k)!} \cdot (1 - \frac{\lambda}{n})^n \cdot (1 - \frac{\lambda}{n})^{-k} \tag{3}$$

$$= \frac{\lambda^k}{k!} \cdot 1 \cdot (1 + \frac{1}{-\frac{n}{\lambda}})^{(-\frac{n}{\lambda})(-\lambda)} \cdot 1^{-k}. \tag{4}$$

$$= \frac{\lambda^k}{k!} e^{-\lambda} \tag{5}$$

Over a period of time t, the total number of trials will be nt, and $\lambda' = \lambda t$, so effectively,

$$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$E(X) = \lambda t, Var(X) = \lambda t, MGF \to e^{\lambda t(e^t - 1)}$$

A function, $N(t)$ which counts the number of events occurring in $(0, t]$ follows Poisson Process $(PP(\lambda))$ if it has:

- Independent increment: Events occurring in disjoint time intervals are independent
- Stationary increment: The distribution only depends on the length of interval, not position.

For example, if $N(t)$ denotes the number of customers in a shop, then $N(t)$ is a Poisson process if the PDF for customers entering the shop is the same for the two time intervals 11:00 am-11:30 am and 11:30 pm-12:00 am, and any other half-hour interval too. Also, distributions for any two intervals should be independent of each other.

2.3 Compound Poisson Process

If $N(t)$ is a Poisson Process and Y_i are i.i.d. (independently identically distributed) then,

$$X(t) = \sum_{i=1} N(t) Y_i$$

is a compound Poisson Process. Note that if $Y_i = 1 \forall i$, then $X(t) = N(t)$.

$$E(X(t) = E(\sum_{i=1}^{n} Y_i | N(t)) = n \tag{6}$$

$$= E(N(t))E(Y_i) = \lambda t E(Y_i) \tag{7}$$

$$V(X(t)) = \lambda t E(Y_i^2)$$

2.4 Continuous Probability Distributions

- **Exponential Distribution:**$(Exp(\lambda))$

$$\text{PDF} \rightarrow f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0, \lambda > 0 \end{cases}$$

$$\text{CDF} \rightarrow F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$$\text{MGF} \rightarrow M(t) = (1 - \frac{t}{\lambda})^{-1}, t < \lambda$$

$$E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$$

- **Gamma Distribution:**$(Gamma(\lambda, r)$

$$\text{PDF} \rightarrow f(x) = \begin{cases} 0 & x < 0 \\ \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1} & x \geq 0, \lambda > 0, r > 0 \end{cases}$$

Where $\Gamma(r) = (r-1)!$

$$M(t) = (1 - \frac{t}{\lambda})^{-r}, E(X) = \frac{r}{\lambda}, Var(X) = \frac{r}{\lambda^2}$$

Note: $Gamma(\lambda, 1) \equiv Exp(\lambda)$

- **Normal Distribution** $(N(\mu, \sigma^2)$**:**

$$\text{PDF} \rightarrow f(x) = \frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi}\sigma}$$

$$M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, E(X) = \mu, Var(X) = \sigma^2$$

$$f(\mu - x) = f(\mu + x)$$

When $\mu = 0$ and $\sigma = 1$, Normal distribution becomes a *Standard* Normal Distribution $(N(0, 1))$:

$$\text{PDF} \rightarrow \phi(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$$

$$\text{CDF} \rightarrow \Phi(x) = \int_{-\infty}^{x} \phi(u)du$$

$$M(t) = e^{\frac{1}{2}t^2}, E(X) = 0, Var(X) = 1$$

Also, for a Standard Normal Distribution, $\Phi(x) + \Phi(-x) = 1$

2.5 Joint Probability Distributions

A function is a Joint PMF/PDF for variables X,Y if:

$$p_{XY}(x, y) = f_{XY}(x, y) \geq 0, \forall x, y$$

$$\sum_X \sum_Y p(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

$$\text{Marginal density of } X \rightarrow f_X(x) = \sum_Y f_{XY}(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \text{ and vice versa}$$

$$\text{Conditional density of } X \text{ given } Y = y \rightarrow f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} \text{ and vice versa}$$

For two random variables X, Y:

$$E(E(X|Y)) = E(X)$$
$$E(E(g(X)|Y)) = E(g(X))$$
$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$

Independence:

$$f_{XY}(x, y) = f_X(x) f_Y(y), \forall x \forall y$$

Covariance:

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X \mu_y$$

Correlation:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$-1 \leq \rho_{XY} \leq 1$, and $|\rho_{XY}| = 1$ if and only if $Y = \alpha + \beta X$ for some real numbers α and $\beta \neq 0$. If $\rho_{XY} = 0$, then X and Y are uncorrelated.

If X, Y are independent, then $Cov(X, Y) = \rho_{XY} = 0$, but the converse is not true.

Bivariate Normal Distribution: (BVN($\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$)) If $X = N(\mu_1, \sigma_1^2)$ and $Y = N(\mu_2, \sigma_2^2)$

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1 - \rho^2)} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right)} \tag{8}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho)^2}\left(\left(\frac{y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right)^2} \tag{9}$$

$$E(X) = \mu_1, E(Y) = \mu_2, Var(X) = \sigma_1^2, Var(Y) = \sigma_2^2$$

For n variables $X_1, X_2, \ldots X_n$

$$E(a_0 + \sum_{i=1}^{n} a_i X_i) = a_0 + \sum_{i=1}^{n} a_i \mu_i$$

$$Var(a_0 + \sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

$$M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$\sum_{i=1}^{m} Bin(n_i, p) = Bin(\sum_{i=1}^{m} n_i, p)$$

$$\sum_{i=1}^{m} Poiss(\lambda_i) = Poiss(\sum_{i=1}^{m} \lambda_i)$$

$$\sum_{i=1}^{m} Geo(p) = NB(m, p)$$

$$\sum_{i=1}^{n} NB(n_i, p) = Bin(\sum_{i=1}^{m} n_i, p)$$

$$\sum_{i=1}^{n} Gamma(\alpha_i, \beta) = Gamma(\sum_{i=1}^{m} \alpha_i, \beta)$$

$$\sum_{i=1}^{n} \chi^2_{r_i} = \chi^2_{\sum_{i=1}^{m} r_i}$$

$$a_0 + \sum_{i=1}^{m} a_i N(\mu_i, \sigma_i^2) = N(a_0 + \sum_{i=1}^{m} a_i \mu_i, \sum_{i=1}^{m} a_i^2 \sigma_i^2)$$

Central Limit Theorem: For i.i.d. (independent and identically distributed) random variables ${}_{i=1}^{n} X_i$ with mean μ and variance σ^2 then

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

has approximately $N(0, 1)$ for $n \to \infty$.

Law of Large Numbers: For ${}_{i=1}^{n} X_i$ with common mean μ and common variance σ^2. Let $S_n = \sum_{i=1}^{n} X_i$. For any $\epsilon > 0$,

$$P(\left|\frac{S_n}{n} - \mu\right| < \epsilon) \to 1, \text{ as } n \to \infty$$

This was half a course.

3 Discrete-Time Stochastic Processes

3.1 Stochastic Processes and Definitions

A stochastic process is a family of random variables $\{X(t), t \in T\}$ defined on a given probability space, indexed by the parameter t, where $t \in T$.

Values of $X(t)$ are called states, and the set of all possible values form the state space S of the process.
Consider a shop where customers arrive at random points in time, queue up for service and leave after service is over. There can be four types of stochastic processes:

1. Discrete state, discrete parameter - Number of customers waiting in the shop when $k^t h$ customer leaves.
2. Discrete state, continuous parameter - Number of customers waiting in the shop at time t.
3. Continuous state, discrete parameter - Time that the kth customer has to wait for service completion.
4. Continuous state, continuous parameter - Time required to complete all queued jobs at time t.

The most basic form of a stochastic process is a Discrete-time Markov Chain.

3.2 DTMC - Discrete-Time Markov Chain

A discrete state, discrete parameter stochastic process which takes on a finite number of possible values, and the probability distribution for the next state is only dependent on the current state.

$$\begin{aligned} llP(X_{n+1} = j|X_0 = i_0, X_1 = i_1, \ldots X_n = i_n) &= P(X_{n+1} = j|X_n = i) && \text{(Last state only)} \\ &= p_{ij}^{(1)}(n) && \text{(Transition probability for 1 timestep)} \\ &= p_{ij}^{(1)} && \text{(Independent of n or time)} \end{aligned}$$

For a DTMC, there exists a value p_{ij} for every pair of i, j to indicate probability of next state being j if current state is i. Thus, we can create a matrix called a *Transition Probability Matrix (TPM)* which contains all the data.

Transition Probability Matrix (TPM): A matrix that contains all the transition probabilities of a DTMC, where element p_{ij} (at row i, column j) in the matrix denotes the probability that the next state will be j, given that the current state is i.

$$P = P^{(1)} = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ \vdots & \ddots & \\ a_{n1} & & a_{nn} \end{bmatrix}$$

3.3 n-step Transition Probability

$$p_{ij}^{(n)} = P(X_{m+n} = j|X_m = i) = P(X_n = j|X_o = i)$$
$$P^{(n)} = \left[(p_{ij}^{(n)})\right]$$

Chapman-Kolmogorov equation:

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)} = \sum_k p_{ik}^{(n)} p_{kj}^{(m)}$$

PMF at step n:

$$\begin{bmatrix} p_1^{(n)} \cdots p_k^{(n)} \end{bmatrix} = \begin{bmatrix} p_1^{(n)} \cdots p_k^{(n)} \end{bmatrix} \begin{bmatrix} p_{11}^{(1)} & \cdots & p_{1k}^{(1)} \\ \vdots & \ddots & \\ p_{k1}^{(1)} & & \end{bmatrix}^n$$

$$\text{or } p^{(n)} = p^{(0)} P^n$$

3.4 Classification of states:

$i \to j$: j is accessible from i.
$i \leftrightarrow j$: i and j communicate with each other (this is true if and only if $i \to j$ and $j \to i$.
If $i \leftrightarrow j, j \leftrightarrow k$ then $i \leftrightarrow k$.

Irreducibility: A TPM P is irreducible if every state communicates with every other state, or that $i \leftrightarrow j \quad \forall i, \forall j \neq i$, and reducible otherwise.
Period of a state i ($d(i)$):

$$d(i) = gcd n \geq 1 | p_{ii}^{(n)} > 0, (0 if p_{ii} = 0 \forall n \geq 1)$$

If $d(i) = 1$, i is aperiodic.

First recurrence: $f_{ii}^{(n)}$ is defined as the probability that the first recurrence of state i is after n steps.

$$f_{ii}^{(n)} = P(X_n = i, X_{n-1} \neq i, \ldots | X_0 = i)$$

Summing $f_{ii}^{(n)}$ over all n, we get

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

$f_i i$ here is the probability that the state i will ever recur. If $f_{ii} = 1$, then the state i is recurrent. Else, it is transient.
Occurence: Let

$$\begin{aligned} I_n &= 1, & X_n &= i \\ &= 0, & X_n &\neq i \end{aligned}$$

$$\begin{aligned} E(\sum_{n=1}^{\infty} I_n | X_o = i) &= \sum_{n=1}^{\infty} E(I_n | X_0 = i) \\ &= \sum_{n=1}^{\infty} [1 \cdot P(X_n = i | X_0 = i) + 0] \\ &= \sum_{n=1}^{\infty} P(X_n = i | X_0 = i) \\ &= \sum_{n=1}^{\infty} p_{ii}^{(n)} \end{aligned}$$

State i is recurrent if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ (which is true if and only if $f_{ii} = 1$). Else, i is transient.
Max Recurrence Time:

$$m_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$$

If $m_{ii} = \infty$ then i is null recurrent, else it is non-null or positive recurrent.
Postulates of Recurrence:

1. $i \leftrightarrow j, i$ recurrent $\implies j$ recurrent.
2. $i \leftrightarrow j, i$ transient $\implies j$ transient.
3. In a *finite state* Markov Chain, all states cannot be transient.
4. In a finite state, *irreducible* Markov Chain, all states are recurrent.
5. In an irreducible Markov Chain, all states are recurrent or transient.

3.5 Gambler's Ruin Problem

Let a gambler have Rs. i in his pocket in the beginning. At every step, he places a bet of Rs. 1 and then gains $Z_i = +1$ rupees (with probability p) or -1 rupees (with probability q, such that $p + q = 1$). He quits the game, when he reaches either N rupees or 0 rupees. Let X_n denote the money the gambler has after n turns, Then X_n forms a DTMC. And for a given n,

$$X_n = i + Z_1 + Z_2 + \ldots + Z_n$$
$$p_{i,i+1} = p, p_{i,i-1} = q \quad \forall 0 < i < N$$
$$p_{0,0} = 1, p_{N,N} = 1$$
$$TPM = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ q & 0 & p & \cdots & 0 \\ 0 & q & 0 & p & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & 1 \end{bmatrix}$$

Here,

$$0 \to \text{recurrent state}$$
$$1, 2, \ldots, N-1 \to \text{transient states}$$
$$N \to \text{recurrent state}$$

For any state i, probability of winning (reaching state N) is:

$$\begin{aligned} P_i &= pP_{i+1} + qP_{i-1} \\ (p+q)P_i &= pP_{i+1} + qP_{i-1} \\ p(P_{i+1} - P_i) &= q(P_i - P_{i-1}) \end{aligned}$$

$$\begin{aligned} (P_{i+1} - P_i) &= \frac{q}{p}(P_i - P_{i-1}) \\ (P_2 - P_1) &= \frac{q}{p}(P_1 - P_0) && = \frac{q}{p}P_1 \\ (P_3 - P_2) &= \frac{q}{p}(P_2 - P_0) && = \left(\frac{q}{p}\right)^2 P_1 \end{aligned}$$

$$\begin{aligned} \text{So, } (P_i - P_{i-1}) &= \left(\frac{q}{p}\right)^{i-1} P_1 \\ \text{similarly, } (P_{i-1} - P_{i-2}) &= \left(\frac{q}{p}\right)^{i-2} P_1 \\ &\vdots \\ (P_2 - P_1) &= \left(\frac{q}{p}\right) P_1 \end{aligned}$$

Adding the equations,

$$\begin{aligned} P_i &= P_1 \left(1 + \left(\frac{q}{p}\right)^2 + \left(\frac{q}{p}\right)^3 + \dots \left(\frac{q}{p}\right)^{i-1}\right) \\ P_i &= P_1 \left(\frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)}\right) \\ P_N = 1 &= P_1 \left(\frac{1 - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)}\right) \\ P_1 &= \left(\frac{1 - \left(\frac{q}{p}\right)}{1 - \left(\frac{q}{p}\right)^N}\right) \end{aligned}$$

Plugging this value for P_i,

$$P_i = \left(\frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}\right)$$

Thus for starting state i, the probability of winning is,

$$P_i = \left(\frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} \right)$$

3.6 Limiting Probabilities

- For a *regular* Markov Chain, a limiting probability distribution exists, independent of the initial probability distribution, such that $\pi = [\pi_1 \pi_2 \ldots \pi_N] = [\pi_1 \pi_2 \ldots \pi_N]P$ where $\pi_j > 0 \forall 0 \leq j \leq N$ and $\sum_j \pi_j = 1$.
- For a non-regular Markov Chain, the solution to $\pi = \pi P$, (if it exists), describes a stationary probability distribution which does not change *if it is the initial distribution.* A limiting probability distribution is always a stationary distribution.
- As $n \to \infty$, $p^{(i)} \to \pi$.
- **A TPM P is regular if for some k, P^k has all elements greater than 0.**
- A Markov Chain is called *Ergodic* if it is both irreducible and aperiodic. A *finite state* DTMC is ergodic if and only if it is regular.
- For a *Doubly Stochastic Matrix* (NxN), (where both rows and columns sum to 1, $\pi = \left[\frac{1}{N} \frac{1}{N} \ldots \frac{1}{N}\right]$, or $pi_i = \frac{1}{N} \forall i$.

3.7 Mean Time Spent in Transient States:

- Let s_{ij} be the expected number of time periods that a Markov Chain spends in state j, given that it starts in state i.
- Let δ_{ij} such that,

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$$

- Let $I_{n,j}$ such that,

$$I_{n,j} = \begin{cases} 1, & X_n = j \\ 0, & \text{otherwise} \end{cases}$$

- So,

$$
\begin{aligned}
s_{ij} &= \delta_{ij} + E\left(\sum_{n=1}^{\infty} I_{n,j} | X_0 = i\right) \\
&= \delta_{ij} + \sum_{n=1}^{\infty} E\left(I_{n,j} | X_0 = i\right) \\
&= \delta_{ij} + \sum_{n=1}^{\infty} P\left(X_n = j | X_0 = i\right) \\
&= \delta_{ij} + \sum_{n=1}^{\infty} p_{ij}^{(n)} \\
&= \delta_{ij} + \sum_{n=1}^{\infty} \sum_{k} p_{ik} p_{kj}^{(n-1)} \\
&= \delta_{ij} + \sum_{k} p_{ik} \sum_{n=1}^{\infty} p_{kj}^{(n-1)} \\
&= \delta_{ij} + \sum_{k} p_{ik} \left[\delta kj + \sum_{n=2}^{\infty} p_{kj}^{(n-1)}\right] \\
&= \delta_{ij} + \sum_{k} p_{ik} \left[\delta kj + \sum_{n=1}^{\infty} p_{kj}^{(n)}\right] \\
s_{ij} &= \delta_{ij} + \sum_{k} p_{ik} s_{kj}
\end{aligned}
$$

- So, in matrix form,

$$
\begin{aligned}
S &= I + P^T S \\
(I - P^T) S &= I \\
S &= (I - P^T)^{-1}
\end{aligned}
$$

3.8 Branching Process:

Let a population exist, each member of which produces j offspring in its lifetime, independently of other processes, with probability $p_j < 1 \forall j \geq 0$. The population of the species in any given generation is given by a distribution called a Branching Process, where

- X_0 is the size of the zeroth generation (usually 1),
- X_1 denotes all the offspring of the first generation,
- X_i denotes the size of the i^{th} generation.
- The population will either die out ($X_n = 0$), or converge to infinity ($X_n = \infty$).

- For a generation n, where Z_i represents the number of offspring of individual i in generation $n-1$,

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i, \text{ and } E(Z_i) = \mu, Var(Z_i) = \sigma^2$$

- Mean and Variance of the number of offsprings of a single individual,

$$\mu = \sum_{j=0}^{\infty} jp_j = E(\sum_{i=1}^{X_{n-1}} Z_i | X_{n-1}) = E(\mu X_{n-1}) = \mu E(X_{n-1}) = \mu^n E(X_0) = \mu^n$$

$$\sigma^2 = \sum_{j=0}^{\infty} (j-\mu)^2 p_j = \begin{cases} \sigma^2 \mu^{n-1} \left(\dfrac{1-\mu^n}{1-\mu} \right), & \mu \neq 1 \\ n\sigma^2, & \mu = 1 \end{cases}$$

- Probability of going extinct after the n^{th} generation,

$$\begin{aligned} u_{n+1} &= P(X_{n+1}) = 0) \\ &= \sum_j P(X_{n+1} = 0 | X_1 = j) p_j \\ &= \sum_j P(X_n = 0)^j p_j \\ &= \sum_j u_n^j p_j \\ u_{n+1} &= \sum_j u_n^j p_j \end{aligned}$$

- Probability of ultimate extinction, π_0, probability that the population will eventually die out,

$$\lim_{n \to \infty} P(X_n = 0 | X_0 = 1)$$

  For $\mu < 1$,

$$\begin{aligned} \mu^n = E(X_n) &= \sum_{j=1}^{\infty} jP(X_n = j) \\ &\geq \sum_{j=1}^{\infty} 1 \cdot P(X_n = j) \\ &= P(X_n \geq 1) \end{aligned}$$

So, $\pi_0 \to 1$ for $n \to \infty$ if $\mu < 1$. Also,

$$\begin{aligned}
\pi_0 &= P(X_\infty = 0) \\
&= \sum_{j=0}^{\infty} P(X_\infty = 0 | X_1 = j) p_j \\
&= \sum_{j=0}^{\infty} \pi_0^j p_j \\
\pi_0 &= \sum_{j=0}^{\infty} \pi_0^j p_j
\end{aligned}$$

The smallest integer that satisfies the above equation is the true value of π_0.