1 Probability revision

1.1 Common random variables

Bernoulli distribution *Ber(p)*:
$P(X=x)=p^x(1-p)^{1-x}, x\in\{0,1\}$, mean $=p$, variance $=p(1-p)$
Binomial distribution *Bin(n,p)*:
$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, x\in\mathbb{Z}\cap[0,n]$, mean $=np$, variance $=np(1-p)$
Poisson distribution *Pois(λ)*:
$P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}, x\in\mathbb{Z}^{\geq 0}$, mean $=\lambda$, variance $=\lambda$
Geometric distribution *Geo(p)*:
$P(X=x)=p(1-p)^{x-1}, x\in\mathbb{Z}^{>0}$, mean $=\frac{1}{p}$, variance $=\frac{1-p}{p^2}$

1.2 Property of expectation

If $X\geq 0$, then $E[X]\geq 0$
If $X\geq 0$ and $E[X]=0$, then $P(X=0)=1$
If a and b are constants, then $E[a+bX]=a+bE[X]$
If X and Y are random variables, then $E[X+Y]=E[X]+E[Y]$
$E[X]$ is the constant that minimizes $E[(X-c)^2]$
If X is a non-negative random variable that takes integral value, then we have $E[X]=\sum_{i=1}^{\infty}P(X\geq i)$

1.3 Property of variance

$Var(X)\geq 0$
If $Var(X)=0$, then X is a constant.
If a and b are constants, then $Var(a+bX)=b^2Var(X)$
If X and Y are random variables, then $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$, for independent X and Y, their covariance is 0.
$Var(X)=E[X^2]-(E[X])^2$

1.4 Moment generating function

It is the unique identifier of a random variable.
$M_X(t)=E(e^{tX})$
Calculate k-th moment of X:
$E[X^k]=\frac{d^k}{dt^k}M_X(t)|_{t=0}$
Linear transformation:
$M_{aX+b}(t)=e^{bt}M_X(at)$
MGF of binomial distribution: $(pe^t+1-p)^n$
MGF of Poisson distribution: $exp[\lambda(e^t-1)]$
MGF of geometric distribution: $\dfrac{pe^t}{1-(1-p)e^t}$

1.5 Joint distribution

Definition: $p_{X,Y}(x,y)=P(X=x,Y=y)=P(\{\omega: X(\omega)=x, Y(\omega)=y\})$
Marginal distribution: $p_X(x)=P(X=x)=\Sigma_y P(X=x,Y=y)=P(\{\omega: X(\omega)=x\})$

1.6 Independence

Discrete random variables $X_1,\ldots,X_n$ are independent if for any $x_1,\ldots,x_n$, $P(X_1=x_1,\ldots,X_n=x_n)=\Pi_{i=1}^n P(X_i=x_i)$
If $f_1,\ldots,f_n$ are functions from $\mathbf{R}$ to $\mathbf{R}$, then $Y_i=f_i(X_i)$ are also random variables, and they preserve independence.
If X_i's are independent, then $E[\Pi_{i=1}^n X_i]=\Pi_{i=1}^n E[X_i]$
If $Z=\Sigma_{i=1}^n X_i$, then $Var(Z)=\Sigma_{i=1}^n Var(X_i)$, and $M_Z(t)=\Pi_{i=1}^n M_{X_i}(t)$

1.7 Conditional probability

Discrete: $p_{X|Y}(x|y)=\dfrac{P(X=x,Y=y)}{P(Y=y)}$
Continuous: $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$
Multiplication law: $p_{X,Y}(x,y)=p_{X|Y}(x|y)p_Y(y)=p_{Y|X}(y|x)p_X(x)$
Law of total probability: $p_X(x)=\Sigma_y p_{X|Y}(x|y)$
Bayes' formula: $p_{Y|X}(y|x)=\dfrac{p_{X|Y}(x|y)p_Y(y)}{\Sigma_y p_{X|Y}(x|y)p_Y(y)}$
Conditional independence: $P(X=x,Y=y|Z=z)=P(X=x|Z=z)P(Y=y|Z=z)$
Law of total expectation: $E[X]=E[E[X|Y]]$

2 Markov Chain preliminaries

2.1 Definition

A stochastic process with discrete state space S and discrete time set T satisfying the Markovian property.

2.2 Markovian property

Given X_n, what happened afterwards $(t>n)$ is independent with what happend before $(t<n)$. Mathematically, for any set of state $i_0,\ldots,i_{n-1},i,j\in S$ and $n\geq 0$, $P(X_{n+1}=j|X_n=i,X_{n-1}=i_{n-1},\ldots,X_0=i_0)=P(X_{n+1}=j|X_n=i)$

2.3 One-step transition probability

Let $p_{i,j}^{n,n+1}=P(X_{n+1}=j|X_n=i)$ denote the one-step transition probability.

2.4 m-step transition probability

Given all one-step probabilities, we can calculate m-step probabilities.

$$\begin{aligned}
p_{i,j}^{n,n+m} &= p(X_{n+m}=j|X_n=i) && (1)\\
&= \sum_{k\in S} p(X_{n+m}=j, X_{n+1}=k|X_n=i) && (2)\\
&= \sum_{k\in S} p(X_{n+m}=j|X_{n+1}=k, X_n=i)P(X_{n+1}=k|X_n=i) && (3)\\
&= \sum_{k\in S} p(X_{n+m}=j|X_{n+1}=k)p(X_{n+1}=k|X_n=i) && (4)\\
&= \sum_{k\in S} p_{k,j}^{n+1,n+m}p_{i,k}^{n,n+1} && (5)\\
&= \sum_{k_1,k_2,\ldots,k_{m-1}\in S}[p_{i,k_1}^{n,n+1}p_{k_1,k_2}^{n+1,n+2}\ldots p_{k_{m-1},j}^{n+m-1,n+m}] && (6)
\end{aligned}$$

Intuitively, the m-step probability is to consider the sum of probabilities over all m-length paths from state i to state j from time n to time $n+m$ From (3) to (4) we exploit Markovian property, from (5) to (6) is to expand iteratively according to previous steps.

2.5 Specification of Markov Chain

We define a Markov Chain by the index set T, the state space S, and all one-step transition probabilities $p_{i,j}^{n,n+1}$. The one-step transition probabilities could be denoted by a one-step transition probability matrix $\mathbf{P}^{n,n+1}=(p_{i,j}^{n,n+1})$

2.6 Properties of transition probability matrix

All entries are non-negative.
Row sum is always 1.
Chapman-Kolmogorov equations:
$\mathbf{P^{n,n+m+1}}=\mathbf{P}^{n,n+1}\mathbf{P}^{n+1,n+m+1}=\mathbf{P}^{n,n+m}\mathbf{P}^{n+m,n+m+1}$

2.7 Stationary Markov Chain

A Markov Chain is stationary if the one-step transition probability matrix does not change when n changes. Mathematically, we have $p_{i,j}^{n,n+1}=p_{i,j}\ \forall n\in T$
The Chapman-Kolmogorov equation reduces to $\mathbf{P}^{n,n+m}=\mathbf{P}^{(m)}=\mathbf{P}^m$

3 First step analysis

3.1 Initial distribution

Suppose the head of the Markov Chain X_0 follows a certain distribution π_0, then $X_n|X_0\sim\pi_0$ will follow a distribution of $\pi_0\mathbf{P}^n$.

3.2 Terminologies

Absorbing state: A state i which satisfies that $p_{i,j}=0\ \forall j\neq i$.
Stopping time: $T=\min\{n\geq 0: X_n=i\}$

3.3 Trick: move one step forward

Consider the case of gambler's ruin with winning probability p. Intuition suggests that $P(X_T=0|X_1=2)=P(X_T=0|X_0=2)$
Define the process $Y_n=X_{n+1}$, hence $P(X_T=0|X_1=2)=P(Y_{T-1}=0|Y_0=2)$, then we have $Y_{T-1}=X_T=0 \implies T_Y=T-1$. Since $\{Y_n\}$ and $\{X_n\}$ have the same probabilistic structure, we have $P(X_T=0|X_1=2)=P(Y_{T-1}=0|Y_0=2)=P(Y_{T_Y}=0|Y_0=2)=P(X_T=0|X_0=2)$, which completes our proof. By induction, the identity could be generalised from 1 step to any finite k steps. Similar process could be applied to derive the relation $E[T|X_0=i]+1=E[T|X_1=i]$
The implication of this technique is that now we can express our quantity of interest as a linear combination of other related quantities, differentiated by the initial condition. We then will obtain a system of linear equations by moving one step forward, and solve the quantity.

3.4 General analysis procedure

1.Define $|S|$ terms $a_i=E[\sum_{n=0}^T g(X_n)|X_0=i]$ for each $i\in S$
2.Apply the law of total expectation to a_i, conditional on X_1, we have $a_i=\sum_{k\in S}(g(i)+E[\sum_{n=1}^T g(X_n)|X_0=i,X_1=k])P(X_1=k|X_0=i)$
3.$P(X_1=k|X_0=i)=p_{i,k}$ in $\mathbf{P}$.
4.Consider the process $\{Y_n\}$. It is stochastically equivalent with $\{X_n\}$. Further by Markovian property, we have $E[\sum_{n=1}^T g(X_n)|X_0=i,X_1=k]=E[\sum_{n=1}^T g(X_n)|X_1=k]=E[\sum_{n=0}^{T_Y} g(Y_n)|Y_0=k]=E[\sum_{n=0}^T g(X_n)|X_0=k]=a_k$
5.Combine the two parts, we have $a_i=\sum_{k\in S}(g(i)+a_k)p_{i,k}=g(i)+\sum_{k\in S}a_k p_{i,k}$.
6.Solve the system and get the result.

3.5 Delve into the gambler's ruin

General case: The gambler will stop with N dollars in hand, or he is broke.
For each game, the winning probability is p. Denote $q=1-p$.
The gambler starts with $0<k<N$ in hand.
Chance of getting broke: $u_k=P(X_T=0|X_0=k)$
If game is fair, then $p/q=1$, solve and we obtain $u_k=1-\frac{k}{N}$, which agrees with our intuition that when k increases, the probability of getting broke is smaller. When N increases, the probability of getting broke converges to 1 even if the game is fair. It implies that suppose the gambler plays without stopping($N=+\infty$), then he will get broke.
If game is not fair, solve and we obtain $u_k=1-\frac{1-(q/p)^k}{1-(q/p)^N}$

If $q > p$, then it is approximately $1 - \frac{1}{(q/p)^{N-k}}$, again, probability converges to 1 if $N \to +\infty$.
If $q < p$, then it is approximately $(q/p)^k - (q/p)^N$, when $N \to +\infty$, probability converges to $(q/p)^k$, which means we have a chance of not getting broke.
In both cases, probability reduces with k increasing. It agrees with our intuition that with more one has at the beginning, the smaller probability that one will get broke.
In both cases, probability reduces with p increasing. It agrees with our intuition that with larger winning probability, the less likely one will get broke.
Quitting time: $v_k = E[T|X_0 = k]$
If game is fair, solve and we obtain $v_k = k(N-k)$, although we know that $P(stop) \geq P(broke) = 1$, the expectation of time to stop still goes to infinity.
If game is not fair, solve and we obtain $v_k = \frac{1}{p-q}[\frac{N(1-(q/p)^k)}{1-(q/p)^N} - k]$.
If $p > q$, $v_k \to \frac{N-k}{p-q} \to +\infty$ when $N \to +\infty$. It agrees with our intuition that with higher winning probability, it takes more time to stop the game. With k approaching N, v_k decreases. It agrees with our intuition that with higher k, we are more likely to reach N early and quit the game.
If $p < q$, $v_k \to \frac{k}{q-p}$ when $N \to +\infty$, which implies that when the winning probability is less than $\frac{1}{2}$, the game is expected to stop in finite time. When k increases, v_k increases. It agrees with our intuition that with higher k, one will sustain longer before getting broke.

4 Classification of states

4.1 Accessibility

Definition: State j is accessible from state i if $\mathbf{P}^m_{i,j} > 0$ for some $m > 0$. Note that state i is accessible from itself since we define $\mathbf{P}^0 = \mathbf{I}$. If state i is also accessible from state j, then they communicate with each other, and we denote by $i \leftrightarrow j$. Communication is an equivalence relation, which means that it exhibits reflexivity, symmetry and transitivity.

4.2 Communication class

Each communication class is an equivalence class defined by the communication relation. It contains all states that are communicating with each other. Suppose we have two classes $\mathcal{C}_1, \mathcal{C}_2$, and $i \in \mathcal{C}_1 \to j \in \mathcal{C}_2$, then any state in $\mathcal{C}_2$ is accessible from any state in $\mathcal{C}_1$. Note that converse cannot hold in this case, otherwise they will become one communication class.

4.3 Reducible chain

Definition: An MC is irreducible if all states communicate with each other, otherwise reducible.

4.4 Return probability

Probability that starting from state i, and revisit state i at the n-th step, which is $P^n_{i,i}$. If state i is transient, then $P^n_{i,i} \to 0$ when $n \to +\infty$, which means that in long term we will never revisit state i.

4.5 First return probability

The probability that starting from state i, the first revisit to state i occurs at the n-th step. Mathematically, we denote by $f^n_{ii} = P(X_1 \neq i, X_2 \neq i, \ldots, X_{n-1} \neq i, X_n = i|X_0 = i)$. We define $f^0_{ii} = 0$. Since first revisit at different steps are disjoint events, and if a state is transient, we may not return to it in long run, we naturally have $P(\text{first revisit at step n} < +\infty) = \sum_{n=1}^{\infty} f^n_{ii} \leq 1$, and $f^n_{ii} \leq P^n_{ii}$ $\forall n \in \mathbf{Z}^{>0}$.
Condition on the first revisit, we have $P^n_{ii} = \sum_{k=0}^{n} f^k_{ii} P^{n-k}_{ii}$.

4.6 Recurrent and transient states

Consider $f_{ii} = \sum_{n=0}^{\infty} f^n_{ii} = \lim_{N\to\infty} \sum_{n=0}^{N} f^n_{ii}$. It denotes the probability of revisiting i in the future.
Definition: A state i is recurrent if $f_{ii} = 1$, and transient if $f_{ii} < 1$.
If the flow will revisit state i for sure, then state i is recurrent. Otherwise, the flow will finally leave state i and never come back.
Let N_i denote the number of times that the flow revisits state i. $N_i = \sum_{n=1}^{\infty} I(X_n = i)$, then recurrence means $E[N_i|X_0 = i] = \infty$, and transient means $E[N_i|X_0 = i] < \infty$.
Theorem: For a state i, consider the expected number of visits to i, $E[N_i|X_0 = i]$, then if $f_{ii} < 1$(i.e.i is transient), then $E[N_i|X_0 = i] = \frac{f_{ii}}{1-f_{ii}}$. If $f_{ii} = 1$(i.e.i is recurrent), then $E[N_i|X_0 = i] = \infty$.
Note that $E[N_i|X_0 = i] = E[\sum_{n=1}^{\infty} I(X_n = i|X_0 = i)] = \sum_{n=1}^{\infty} E[I(X_n = i|X_0 = i)] = \sum_{n=1}^{\infty} P(X_n = i|X_0 = i) = \sum_{n=1}^{\infty} P^n_{ii}$.
Theorem: State i is recurrent iff $\sum_{n=1}^{\infty} P^n_{ii} = \infty$ For recurrent states, it means a significant probability to return. For transient states, the series will converge to 0, which means that the probability of coming back vanishes in the long run.

4.7 Summary of recurrence

Recurrent $\iff f_{ii} = 1 \iff \sum_{n=1}^{\infty} P^n_{ii} = \infty \iff E[N_i|X_0 = i] = \infty$
Transient $\iff f_{ii} < 1 \iff \sum_{n=1}^{\infty} P^n_{ii} < \infty \iff E[N_i|X_0 = i] < \infty$

4.8 States of the same class

Theorem: States in the same communication class are either all transient or all recurrent. We could interpret transient/recurrent status as a status for the whole class. An MC with finite states must have at least one recurrent class.

5 Long run performance

5.1 Periodicity

Definition: For a state i, consider $\{n \geq 1 : P^n_{ii} > 0\}$, we define the period of state i, $d(i)$, as the greatest common divisor of the set. If the set is empty, we define $d(i) = 0$. If $d(i) = 1$, we say that state i is aperiodic.
Theorem: If $i \leftrightarrow j$, then $d(i) = d(j)$.
Theorem: $\exists N$ such that $P^{Nd(i)}_{ii} > 0$, and $\forall n \geq N$, $P^{nd(i)}_{ii} > 0$.
Theorem: If $\exists m > 0$ such that $P^m_{ji} > 0$, then for sufficiently large n, we have $P^{m+nd(i)}_{ji} > 0$.

5.2 Regular Markov Chain

Definition: An MC is regular if $\exists k > 0$, such that all elements of $\mathbf{P}^k$ are strictly positive.
It means that the flow can achieve any state from any state at step k.
Theorem: If a Markov Chain is irreducible, aperiodic and with finite states, then it is regular.

5.3 Main theorem

Theorem: Suppose $\mathbf{P}$ is a regular transition probability matrix with states $S = \{1, 2, \ldots, N\}$, then
1. $\lim_{n\to\infty} p^n_{ij}$ exists.
2. The limit does not depend on i. Hence, we can denote it by $\pi_j = \lim_{n\to\infty} p^n_{ij}$.
3. $\sum_{k=1}^{N} \pi_k = 1$. We call it as the limiting distribution.
4. The limits $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$ are the solutions of the system $\pi_j = \sum_{k=1}^{N} \pi_k P_{kj}, j = 1, 2, 3, \ldots, N, \sum_{k=1}^{N} \pi_k = 1$. In matrix form, it is to solve $\pi P = \pi, \sum_{k=1}^{N} \pi_k = 1$
5. The limiting distribution π is unique.
We should interpret π_j as the marginal probability that the flow is in state j for a long run. In other words, suppose we observe the flow at some time, then the probability that it stays in state j is π_j. That probability is not related to the original state and the time point. It gives the limit of $\mathbf{P}^n$, in long run, each row of $\mathbf{P}^n$ is the same, which is π. π can be seen as the long run proportion of time in every state.

5.4 Stationary distribution

When the MC is not regular, it is possible that $|S| = \infty$ and $\pi = 0$, but it is also possible that we can still find a non-trivial π. It means, if the initial states have a distribution π, then after any steps, the chain also has a distribution π on the states, which is called the stationary distribution.
Definition: Consider a Markov Chain with state space $S = \{1, 2, \ldots\}$ and the transition probability matrix P. A distribution $(p_1, p_2, \ldots)$ on S is called a stationary distribution, if it satisfies that if $P(X_n = i) = p_i, i = 1, 2, \ldots$, then $P(X_{n+1} = i) = p_i, i = 1, 2, \ldots$
It means that if the initial states have a distribution π, then after any steps, the chain still has a distribution π on the states. Note that a irregular MC may have more than one stationary distributions. To solve for possible π, note that here we also have to solve the system $\pi P = \pi$

6 Helpers

Binomial expansion: $(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$
Maclaurin expansion: $f(x) = \sum_{i=0}^{\infty} \dfrac{x^i}{i!} f^{(i)}(0)$
Arithmetic series: $\sum_{i=1}^{n} a_i = \dfrac{n(a_1 + a_n)}{2}$
Geometric series: $\sum_{i=1}^{n} a_i = \dfrac{a(1 - r^n)}{1 - r} (r \neq 1)$ $\sum_{i=1}^{\infty} a_i = \dfrac{a}{a - r} (|r| < 1)$