Data Science

Leo

Mainly some useful facts without proof.

May 14, 2024

This note is made possible
by Hopcroft and Sven-Ake Wegner books.

Contents

1. High-Dimensional Space

λ^d denotes the d-dimensional Lebesgue measure on $\mathbb{R}^d$.

We say that X is uniformly distributed on $B \in \mathbb{R}^d$ if

$$\Pr(X \in A) = \frac{\lambda^d(A \cap B)}{\lambda^d(B)} \tag{1}$$

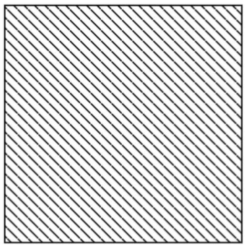
holds for every Borel set $A \in \mathbb{R}^d$. We write $X \sim U(B)$ in this case.

We say that X is (spherically) Gaussian distributed with mean zero and variance $\sigma > 0$, if

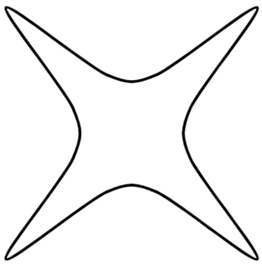
$$\Pr(X \in A) = \frac{1}{(2\pi\sigma^2)^{d/2}} \int_A \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) d\lambda^d(x) \tag{2}$$

holds for every Borel set $A \in \mathbb{R}^d$. We write $X \sim N(0, \sigma^2, \mathbb{R}^d)$ in this case. Note that $X = (X_1, ..., X_d) \sim N(0, \sigma^2, \mathbb{R}^d)$ holds iff $X_i \sim N(0, \sigma^2, \mathbb{R}^1)$ holds for all $i = 1, ..., d$.

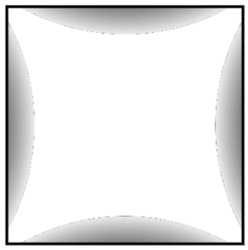
1.1. The curse of high dimensions



The hypercube is compact and convex.



Distances between corners grow with the dimension, while distances between faces remain constant.



The volume concentrates at the surface and in the middle of the faces.

1.2. Volume and Surface

Surface Concentration Theorem: The vast majority of the volume lies near the boundary of the hypersphere.

$$\frac{\text{volume}((1-\varepsilon)A)}{\text{volume}(A)} = (1-\varepsilon)^d \leqslant e^{-\varepsilon d}. \tag{3}$$

1.2.1. Volume and Surface of the Unit Ball

The surface area $A(d)$ and the volume $V(d)$ of a unit-radius ball in d-dimensions are given by

$$A(d) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \text{ and } V(d) = \frac{2\pi^{\frac{d}{2}}}{d\Gamma\left(\frac{d}{2}\right)}. \tag{4}$$

$$V_n(R) = \begin{cases} 1 & \text{if} n = 0, \\ 2R & \text{if} n = 1, \\ \frac{2\pi}{n} R^2 \times V_{n-2}(R) & \text{otherwise.} \end{cases} \tag{5}$$

Gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\ \mathrm{dt}$ have some properties:

- For positive x, $\Gamma(x+1) = x\Gamma(x)$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- For positive integer x, $\Gamma(x) = (x-1)!$
- For non-negative integer x, $\Gamma\left(\frac{1}{2} + x\right) = \frac{(2x-1)!!}{2^x}\sqrt{\pi}$ and $\Gamma\left(\frac{1}{2} - x\right) = \frac{(-2)^x}{(2x-1)!!}\sqrt{\pi}$

The volume of the unit ball reaches its maximum at $d = 5$.

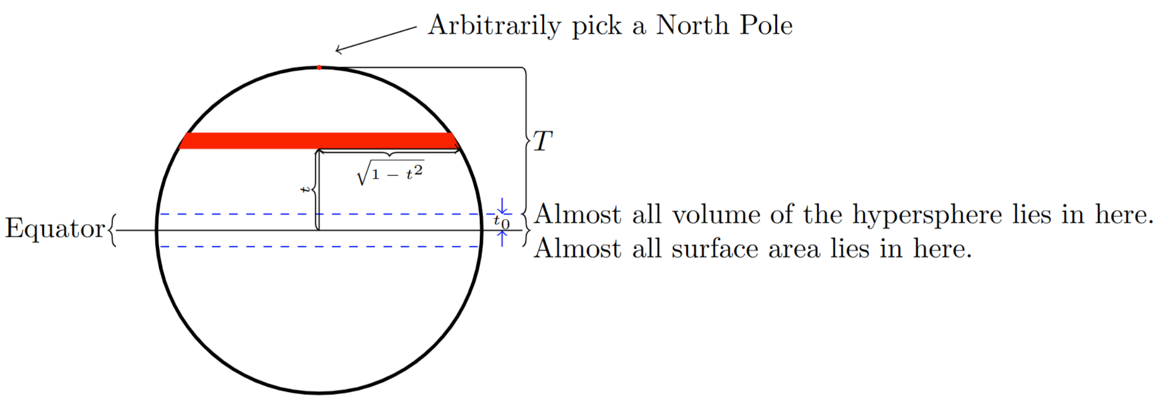
If $d = 2k$, then $V(d) = \frac{2\pi^{\frac{d}{2}}}{d(k-1)!} = \frac{2\pi^{\frac{d}{2}}}{d\left(\frac{d}{2}-1\right)!} = \frac{\pi^{\frac{d}{2}}}{\left(\frac{d}{2}\right)!}$. If $d = 2k+1$, then $V(d) = \frac{2\pi^{\frac{d}{2}}}{d\left(\Gamma\left(\frac{2k+1}{2}\right)\right)} = \frac{\pi^{\frac{d}{2}}}{\frac{2k+1}{2}\left(\frac{2k-1}{2}!\right)} = \frac{\pi^{\frac{d}{2}}}{\left(\frac{d}{2}\right)!}$.

$\lim_{d\to\infty} V(d) = 0$.

1.2.2. Volume near the equator

Defination (Arbitrary “North Pole” and “equator”) Pick a vector on the hypersphere arbitrarily, call it the North Pole. Then the intersection of the hyperplane that is perpendicular to it and the hypersphere is called the equator.

As d grows to infinity, almost all volume are near the equator.



Waist Concentration Theorem: Let $d \geqslant 3$ and $\varepsilon > 0$. Then we have

$$\frac{\lambda^d(W_\varepsilon)}{\lambda^d\left(\overline{\mathrm{B}}_1(0)\right)} \geqslant 1 - \frac{2}{\varepsilon\sqrt{d-1}} \exp\left(-\frac{\varepsilon^2(d-1)}{2}\right) \tag{6}$$

$$\frac{\lambda^d\left(\{(x_1, ..., x_d) \in \overline{\mathrm{B}}_1(0); |x_1| > \varepsilon\}\right)}{\lambda^d\left(\overline{\mathrm{B}}_1(0)\right)} \leqslant \frac{2}{\varepsilon\sqrt{d-1}} \exp\left(-\frac{\varepsilon^2(d-1)}{2}\right) \tag{7}$$

where $W_\varepsilon = \left\{(x_1, ..., x_d) \in \overline{\mathrm{B}}_1(0); |x_1| \leqslant \varepsilon\right\}$.

Proof: Bound $\int_{\frac{c}{\sqrt{d-1}}}^1 \left(1 - x_1^2\right)^{\frac{d-1}{2}} V(d-1)\,\mathrm{d}x$ and $\int_0^1 \left(1 - x_1^2\right)^{\frac{d-1}{2}} V(d-1)\,\mathrm{d}x$.

Let $\varepsilon = \frac{c}{\sqrt{d-1}}$ and we can know that: For $c > 0$ and $d \geqslant 3$, at least a $1 - \frac{2}{c}e^{-\frac{c^2}{2}}$ fraction of the volume of the d-dimensional unit ball has $|x_1| \leqslant \frac{c}{\sqrt{d-1}}$.

Let $d \geqslant 3$ and assume that we draw $n \geqslant 2$ points $x^{(1)}, ..., x^{(n)}$ at random from the d-dimensional unit ball. Then

1. $\Pr\left(\left\|x^{(i)}\right\| \geqslant 1 - \frac{2\ln n}{d} \text{ for all } i\right) \geqslant 1 - \frac{1}{n}$, and
2. $\Pr\left(\left|\left\langle x^{(i)}, x^{(j)}\right\rangle\right| \leqslant \sqrt{\frac{6\ln n}{d-1}} \text{ for all } i \neq j\right) \geqslant 1 - \frac{1}{n}$.

1.2.3. The High-Dimensional Unit Cube

A point picked at random in a unit cube will be within distance t of the equator defined by $H = \left\{x \mid \sum_{i=1}^{d} x_i = \frac{d}{2}\right\}$ with probability at least $1 - \frac{1}{4t^2}$.

1.3. Gaussians in High Dimension

If $X, Y \sim N\left(0, 1, \mathbb{R}^d\right)$ and $\xi \in \mathbb{R}^d$ be a constant, then $\forall d \geqslant 1$

1. $E\left(\|X\|^2\right) = d$
2. $\mathrm{Var}\left(\|X\|^2\right) = 2d$
3. $\left|E\left(\|X\| - \sqrt{d}\right)\right| \leqslant \frac{1}{\sqrt{d}}$
4. $\mathrm{Var}(\|X\|) \leqslant 2$
5. $E\left(\|X - Y\|^2\right) = 2d$
6. $\left|E\left(\|X - Y\| - \sqrt{2d}\right)\right| \leqslant \frac{1}{\sqrt{2d}}$
7. $\mathrm{Var}(\|X - Y\|) \leqslant 3$
8. $E(\langle X, Y\rangle) = 0$ and $\mathrm{Var}(\langle X, Y\rangle) = d$
9. $E(\langle X, \xi\rangle) = 0$ and $\mathrm{Var}(\langle X, \xi\rangle) = \|\xi\|^2$
10. $E(|\langle X, \xi\rangle|) = \sqrt{\frac{2}{\pi}}$

Gaussian Annulus Theorem: Let $x \in \mathbb{R}^d$ be drawn at random with respect to the spherical Gaussian distribution with zero mean and unit variance. Then

$$\Pr\left(\left||x| - \sqrt{d}\right| \geqslant \varepsilon\right) \leqslant 2e^{-c\varepsilon^2}. \tag{8}$$

holds for every $0 \leqslant \varepsilon \leqslant \sqrt{d}$ where $c = \frac{1}{16}$.

Proof: Let a random vector $X \sim \mathcal{N}(0, 1)$ be given and let $X_1, ..., X_d$ denote its coordinate functions. We have $X_i \sim \mathcal{N}(0, 1)$ for every $i = 1, ..., d$. We define $Y_i := \frac{X_i^2 - 1}{2}$ and see that

$$\begin{aligned}
\Pr\left[\|X\| - \sqrt{d} \geqslant \varepsilon\right] &\leqslant \Pr\left[\left|\|X\| - \sqrt{d}\right| \cdot \left(\|X\| + \sqrt{d}\right) > \varepsilon \cdot \sqrt{d}\right] \\
&= \Pr\left[\left|X_1^2 + \cdots + X_d^2 - d\right| > \varepsilon\sqrt{d}\right] \\
&= \Pr\left[\left|\left(X_1^2 - 1\right) + \cdots + \left(X_d^2 - 1\right)\right| > \varepsilon\sqrt{d}\right] \\
&= \Pr\left[\left|\frac{X_1^2 - 1}{2} + \cdots + \frac{X_d^2 - 1}{2}\right| > \frac{\varepsilon\sqrt{d}}{2}\right] \\
&= \Pr\left[|Y_1 + \cdots + Y_d| > \frac{\varepsilon\sqrt{d}}{2}\right]
\end{aligned} \tag{9}$$

where we used $\|X\| + \sqrt{d} \geqslant \sqrt{d}$ for the inequality. Since we have $\mathbb{E}(X_i) = 0$ and $\operatorname{Var}(X_i) = 1$ we get

$$\mathbb{E}(Y_i) = \mathbb{E}\left(\frac{X_i^2 - 1}{2}\right) = \frac{1}{2}(\mathbb{E}(X_i^2) - 1) = 0 \tag{10}$$

for every $i = 1, ..., d$. We now estimate for $k \geqslant 2$ the k-th moment of Y_i. For this we firstly note that

$$\left|X_i^2(\omega) - 1\right|^k \leqslant X_i^{2k}(\omega) + 1 \tag{11}$$

holds for every $\omega \in \Omega$. Indeed, if $|X_i(\omega)| \leqslant 1$, then $0 \leqslant X_i^2(\omega) \leqslant 1$ and thus $\left|X_i^2(\omega) - 1\right|^k = \left(1 - X_i^2(\omega)\right)^k \leqslant 1$. If otherwise $|X_i(\omega)| > 1$, then $X_i^2(\omega) - 1 > 0$ and therefore $\left|X_i^2(\omega) - 1\right|^k = \left(X_i^2(\omega) - 1\right)^k \leqslant X_i^{2k}(\omega)$. We employ the above and estimate

$$\begin{aligned}
\left|\mathbb{E}\left(Y_i^k\right)\right| &= \left|\mathbb{E}\left(\left(\frac{X_i^2 - 1}{2}\right)^k\right)\right| = \frac{1}{2^k}\left|\mathbb{E}\left(\left(X_i^2 - 1\right)^k\right)\right| \\
&\leqslant \frac{1}{2^k}\int_\Omega \left|X_i^2 - 1\right|^k \mathrm{d}P \leqslant \frac{1}{2^k}\int_\Omega \left(X_i^{2k} + 1\right)\mathrm{d}P = \frac{1}{2^k}\left(\mathbb{E}\left(X_i^{2k}\right) + 1\right) \\
&= \frac{1}{2^k}\left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} t^{2k}\exp(-t^2/2)\,dt + 1\right) = \frac{1}{2^k}\left(\frac{(2k)!}{2^k k!} + 1\right) \\
&= \frac{(2k-1)(2k-2)\cdots 5\cdot 3\cdot 1}{(2k)(2k-2)\cdots 6\cdot 4\cdot 2}\cdot k! + \frac{1}{2^k} \\
&\leqslant 1\cdot 1\cdots 1\cdot \frac{3}{4}\cdot\frac{1}{2}\cdot k! + \frac{k!}{4\cdot 2} \\
&= \frac{k!}{2}.
\end{aligned} \tag{12}$$

We make use of Bernstein's inequality: $\mathrm{P}[|Y_1 + \cdots + Y_d| \geqslant a] \leqslant 2\exp\left(-C\min\left(\frac{a^2}{d}, a\right)\right)$. with $a := \frac{\varepsilon\sqrt{d}}{2} > 0$ to continue our estimate. Since $\sqrt{d} \geqslant \varepsilon$ holds, we get

$$\begin{aligned}
\mathrm{P}\left[\left|\|x\| - \sqrt{d}\right| \geqslant \varepsilon\right] &\leqslant \mathrm{P}\left[|Y_1 + \cdots + Y_d| \geqslant \frac{\varepsilon\sqrt{d}}{2}\right] \\
&\leqslant 2\exp\left(-C\min\left(\frac{\left(\varepsilon\sqrt{d}/2\right)^2}{d}, \frac{\varepsilon\sqrt{d}}{2}\right)\right) \\
&\leqslant 2\exp\left(-C\min\left(\frac{\varepsilon^2}{4}, \frac{\varepsilon^2}{2}\right)\right) \\
&= 2\exp\left(-C\frac{\varepsilon^2}{4}\right).
\end{aligned} \tag{13}$$

Finally, we recall that the constant $C = \frac{1}{4}$ and thus the exponent equals $-c\varepsilon^2$ with $c = \frac{1}{16}$.

Proposition: Let $x \in \mathbb{R}^d$ be drawn at random with respect to the spherical Gaussian distribution with zero mean and unit variance. Then

$$\Pr\left(\left|\|x\|^2 - d\right| \geqslant \varepsilon\right) \leqslant 2e^{-c\min\left(\frac{\varepsilon^2}{2d},\varepsilon\right)}. \tag{14}$$

holds for every $0 < \varepsilon$ where $c = \frac{1}{8}$.

Proof: We define the random variables Y_i as in the proof of Last Theorem and then proceed using the Bernstein inequality as in the last part of the previous proof. This way we obtain

$$\mathrm{P}\left[\|X\|^2 - d\right| \geqslant \varepsilon\right] = \mathrm{P}\left[|Y_1 + \cdots + Y_d| \geqslant \frac{\varepsilon}{2}\right] \leqslant 2\exp\left(-C\min\left(\frac{\varepsilon^2}{4d}, \frac{\varepsilon}{2}\right)\right) \tag{15}$$

which leads to the claimed inequality taking $C = \frac{1}{4}$ into account.

Let $x, y \in \mathbb{R}^d$ be drawn at random with respect to the spherical Gaussian distribution with zero mean and unit variance. Then for every $\varepsilon > 0$ and for all $d \geqslant 1$ the estimate

$$\Pr\left(\left|\left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|}\right\rangle\right| \geqslant \varepsilon\right) \leqslant \frac{\frac{2}{\varepsilon} + 7}{\sqrt{d}} \tag{16}$$

holds.

Let $X \sim N\left(0, 1, \mathbb{R}^d\right)$ and $0 \neq \xi \in \mathbb{R}^d$ be fixed. Then we have

$$\mathrm{P}(|\langle X, \xi\rangle| \geqslant \varepsilon) \leqslant \frac{4}{\sqrt{2\pi}} \frac{\|\xi\|}{\varepsilon\sqrt{d}} \tag{17}$$

for every $\varepsilon > 0$.

Let $x, y \in \mathbb{R}^d$ be drawn at random with respect to the spherical Gaussian distribution with zero mean and unit variance. Then for every $\varepsilon > 0$ and for all $d \geqslant 1$ we have

$$\Pr\left(\left|\left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|}\right\rangle\right| \leqslant \frac{\varepsilon}{2\sqrt{d}}\right) \leqslant \varepsilon + 2\exp(-cd) \tag{18}$$

where $c = \frac{1}{16}$.

Combining the last two results we can think of the scalar products for a fixed high dimension d, with high probability, to be small numbers which are however bounded away from zero.

Let $x, y \in \mathbb{R}^d$ be drawn at random with respect to the spherical Gaussian distribution with zero mean and unit variance. Then we have the following.

1. $\forall \varepsilon > 18, \exists d_0 \in \mathbb{N}, \forall d \geqslant d_0 : \Pr\left(\left|\|x - y\| - \sqrt{2d}\right| \geqslant \varepsilon\right) \leqslant \frac{18}{\varepsilon}$.
2. $\forall \varepsilon > 18, d \geqslant \varepsilon^2 : \Pr\left(\left|\|x - y\| - \sqrt{2d}\right| \geqslant \varepsilon\right) \leqslant \frac{18}{\varepsilon} + \frac{8}{\sqrt{d}}$.
3. $\forall \varepsilon > 18, d \geqslant \varepsilon^2 : \Pr\left(\left|\|x - y\|^2 - 2d\right| \geqslant \varepsilon\sqrt{2d}\right) \leqslant \frac{18}{\varepsilon} + \frac{8}{\sqrt{d}}$.

1.4. Random projection and Johnson-Lindenstrauss Lemma

Random Projection Theorem: Let $U \in \mathbb{R}^{k \times d}$ be a random matrix with $U \sim N(0,1)$. Then for every $x \in \mathbb{R}^d \setminus \{0\}$ and every $0 < \varepsilon < 1$ we have

$$\Pr\Big(\Big|\|Ux\| - \sqrt{k}\|x\|\Big| \geqslant \varepsilon\sqrt{k}\|x\|\Big) \leqslant 2e^{-ck\varepsilon^2} \tag{19}$$

where $c = \frac{1}{16}$.

Johnson-Lindenstrauss Lemma: Let $0 < \varepsilon < 1$ and $n \geqslant 1$ and let $k \geqslant \frac{48}{\varepsilon^2} \ln n$. Let $U \sim N\big(0, 1, \mathbb{R}^{k \times d}\big)$ be a random matrix. Then we have

$$\Pr\Big((1-\varepsilon)\sqrt{k}\|x_i - x_j\| \leqslant \|Ux_i - Ux_j\| \leqslant (1+\varepsilon)\sqrt{k}\|x_i - x_j\| \text{ for all } i, j\Big) \geqslant 1 - \frac{1}{n} \tag{20}$$

1.5. Separating Gaussian data

Let $\mu_1, \mu_2 \in \mathbb{R}^d$ and $\Delta := \|\mu_1 - \mu_2\|$. Let $X_1 \sim N\big(\mu_1, 1, \mathbb{R}^d\big)$ and $X_2 \sim N\big(\mu_2, 1, \mathbb{R}^d\big)$ be random vectors.

1. $\forall d \geqslant 1 : \Big|E\Big(\|X_1 - X_2\| - \sqrt{\Delta^2 + 2d}\Big)\Big| \leqslant \frac{3d + 8\Delta^2 + \sqrt{24}\Delta d^{1/2}}{2(\Delta^2 + 2d)^{3/2}}$.
2. $\forall d \geqslant 1 : \mathrm{Var}(\|X_1 - X_2\|) \leqslant \left(\frac{3d + 8\Delta^2 + \sqrt{24}\Delta d^{1/2}}{2(\Delta^2 + 2d)^{3/2}}\right)^2 + \frac{3d + 8\Delta^2 + \sqrt{24}\Delta d^{1/2}}{\Delta^2 + 2d}$.
3. $\Pr\Big(\Big|\|X_1 - X_2\| - \sqrt{\Delta^2 + 2d}\Big| \geqslant \varepsilon\Big) \leqslant \frac{54}{\varepsilon} + \frac{8}{\sqrt{d}} + \frac{20}{\varepsilon\sqrt{d}}$ holds for $\varepsilon > 54$ and $d \geqslant \frac{\varepsilon^2}{9}$.

Separation Theorem: Let $\mu_1, \mu_2 \in \mathbb{R}^d$ and put $\Delta := \|\mu_1 - \mu_2\|$. Let $\varepsilon_1 > 18$ and $\varepsilon_2 > 54$ be constant and assume that there is $d_0 \geqslant 1$ such that $\sqrt{\Delta^2 + 2d} - \sqrt{2d} > \varepsilon_1 + \varepsilon_2$ holds for $d \geqslant d_0$. Let $X_i \sim N\big(\mu_i 1, \mathbb{R}^d\big)$ and $S_i := \Big\{x_i^{(1)}, ..., x_i^{(n)}\Big\}$ be samples of X_i for $i = 1, 2$ such that $S_1 \cap S_2 = \emptyset$. Let $x, y \in S_1 \cup S_2$. Then

$$\liminf_{d \to \infty} \mathrm{P}\Big(x, y \text{ come from same Gaussian } \Big|\parallel x - y \parallel - \sqrt{2d}\Big| < \varepsilon_1\Big) \geqslant \frac{1 - 18/\varepsilon_1}{1 + 54/\varepsilon_2} \tag{21}$$

holds.

2. Random walks and Markov chains

2.1. Markov chains basics

A **Markov chain** is a process which moves among the elements of a set Ω (Ω is a set of possibly infinite states $x_1, x_2, ...$) in the following manner: when at $x_t \in \Omega$, the next position x_{t+1} is chosen according to a fixed probability distribution $P(x_t, \cdot)$ depending only on x_t. What we said in the last sentence is essentially the **Markovian Property**, which we can write down formally:

Markovian Property:

$$\Pr(x_t = y | x_1, x_2, ..., x_{t-1}) = \Pr(x_t = y | x_{t-1}). \tag{22}$$

We can use the following notation: $P(x, y) \coloneqq \Pr(x_{t+1} = y | x_t = x)$. The numbers $P(x, y)$ are called the **transition probabilities** of the chain. We denote μ_t as the probability distribution over Ω after t transitions and μ_t is a row vector. We have $\mu_{t+1} = \mu_t P$. Thus we also have $\mu_t = \mu_0 P^t$.

A **Random Walk** on a graph $G = (V, E)$ is a sequence $s_1, s_2, ..., s_t$ for which $s_{i+1} \in N(s_i)$ where $N(v)$ is the neighborhood of v. Note that the transition matrix of a graph is going to have

$$P(x, y) = \frac{1}{d_x} \tag{23}$$

where d_x is the degree of vertex x.

The transition matrix of any Markov Chain is **stochastic**, i.e. the entries in every row add up to 1. A transition matrix is **doubly stochastic** if the rows and the columns add to 1.

A **stationary distribution** π^* has the following property:

$$\forall x, \pi^*(x) = \sum_y \pi^*(y) P(y, x). \tag{24}$$

or even better:

$$\pi^* = \pi^* P \tag{25}$$

The **period** of a state i is the greatest common divisor of the set $\{n \in \mathbb{N} : p(i, i) > 0\}$. If every state has period 1 then the Markov chain (or its transition probability matrix) is called **aperiodic**.

A Markov Chain is **irreducible** if:

$$\forall x, y, \exists t = t(x, y) : P^{t'}(x, y) > 0, \forall t' > t \tag{26}$$

which means that all states communicate with each other.

A Markov Chain is **ergodic** if:

$$\exists t^* : \forall t > t^*, \Pr(x, y) > 0, \forall x, y \tag{27}$$

If a Markov Chain is ergodic then it is going to have a unique stationary distribution no matter what the initial distribution was.

If a finite Markov Chain is irreducible then it has a unique stationary distribution.

Suppose a probability distribution π on Ω satisfies

$$\pi(x)P(x,y) = \pi(y)P(y,x) \text{ for all } x,y \in \Omega. \tag{28}$$

This equation is called the **detailed balance equations**.

Let P be the transition matrix of a Markov chain with state space Ω. Any distribution π satisfying the detailed balance equations is stationary for P.

The **time reversal** of an irreducible Markov chain with transition matrix P and stationary distribution π is the chain with matrix

$$\hat{P}(x,y) := \frac{\pi(y)P(y,x)}{\pi(x)} \tag{29}$$

If a chain with transition matrix P is reversible, then $\hat{P} = P$.

We define **Hitting time** of x to y as:

$$h_{x,y} := E(\# \textit{ steps to go from } x \textit{ to } y) \tag{30}$$

For $x = y$, we call this **recurrence times** and $h_{x,x} = \frac{1}{\pi^*(x)}$.

We define the **cover time** of a vertex v of a graph G as follows:

$$C_{v(G)} = E(\# \textit{ steps to visit all vertices when we start at } v) \tag{31}$$

We define the cover time of a graph G as:

$$C(G) = \max_{v \in G} C_v(G) \tag{32}$$

We define **commute time** from x to y the expected number of steps it takes to go from x to y and back to x.

$$C_{x,y} = E(\# \textit{ of steps to go from } x \textit{ to } y \textit{ and back to } x) = h_{x,y} + h_{y,x} \tag{33}$$

2.2. Markov Chain Monte Carlo

Use Markov chain to sample from a given probability distribution is called **Markov chain Monte Carlo**.

2.2.1. Metropolis-Hasting algorithm

Symmetric base chain. Suppose that Ψ is a symmetric transition matrix. In this case, Ψ is reversible with respect to the uniform distribution on Ω.

The **Metropolis chain** for a probability π and a symmetric transition matrix Ψ is defined as

$$P(x,y) = \begin{cases} \Psi(x,y)\left[\frac{\pi(y)}{\pi(x)} \wedge 1\right] & \text{if } y \neq x, \\ 1 - \sum_{z:z\neq x} \Psi(x,z)\left[\frac{\pi(z)}{\pi(x)} \wedge 1\right] & \text{if } y = x. \end{cases} \quad (34)$$

General base chain. The Metropolis chain can also be defined when the initial transition matrix is not symmetric.

$$P(x,y) = \begin{cases} \Psi(x,y)\left[\frac{\pi(y)\Psi(y,x)}{\pi(x)\Psi(x,y)} \wedge 1\right] & \text{if } y \neq x, \\ 1 - \sum_{z:z\neq x} \Psi(x,z)\left[\frac{\pi(z)\Psi(z,x)}{\pi(x)\Psi(x,z)} \wedge 1\right] & \text{if } y = x. \end{cases} \quad (35)$$

2.2.2. Gibbs Sampling

State spaces are contained in a set of the form S^V, where V is the vertex set of a graph and S is a finite set. The elements of S^V, called **configurations**, are the functions from V to S.

The **Gibbs Sampling** is a Markov chain with probability distribution

$$P(x,y) = \begin{cases} \frac{1}{d}\frac{\pi(y)}{\sum_{z:z\sim_j x} \pi(z)} & \text{if } y \sim_j x \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

where $x \sim_j y$ if $x_i = y_i$ for all $i \neq j$.

3. Some useful inequality

3.1. Markov's inequality

For $a > 0$, we have $\Pr(x \geqslant a) \leqslant \frac{E[x]}{a}$.

3.2. Chebyshev's inequality

For $a > 0$, we have $\Pr(|x - E(x)| \geqslant c) \leqslant \frac{\mathrm{Var}(x)}{c^2}$.

3.3. Law of Large Numbers

Let $x_1, x_2, ..., x_n$ be n independent samples of a random variable x. Then

$$\Pr\left(\left|\frac{x_1 + x_2 + ... + x_n}{n} - E(x)\right| \geqslant a\right) \leqslant \frac{\mathrm{Var}(x)}{na^2}. \tag{37}$$

3.4. Master Tail Bounds Theorem

Let $x = x_1 + x_2 + ... + x_n$ be n mutually independent random variables with zero mean and variance at most σ^2. Let $0 \leqslant a \leqslant \sqrt{2}n\sigma^2$ and $\frac{a^2}{4n\sigma^2} \leqslant s \leqslant \frac{n\sigma^2}{2}$ is a positive even integer and $|E(x_i^r)| \leqslant \sigma^2 r!$ for $r = 3, 4, ..., s$. Then

$$\Pr(|x| \geqslant a) \leqslant 3e^{-\frac{a^2}{12n\sigma^2}}. \tag{38}$$

Bibliography