Introduction to Probability

January 11, 2021 8:49 AM

Permutations and combinations

- The number of ways to choose k objects from n is $n(n-1)\ldots\frac{n-k+1}{k!}=\frac{n!}{k!(n-k)!}=\binom{n}{k}$.
  This is called binomial coefficient.
- Multinomial coefficient: the number of ways to place n objects in m buckets with n_i objects in bucket i is $\binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-n_2-\cdots-n_{m-1}}{n_m}=\frac{n!}{n_1!n_2!\ldots n_m!}$.
- $\binom{n}{k}=\binom{n-k}{i}\binom{k}{i}=\binom{n-k}{i}\binom{k}{k-i}$
- $\sum_k\binom{n}{k}^2=\binom{2n}{n}$.

Probability

- Sample space S: set of all possible outcomes of an experiment
  - Could be finite/infinite, discrete/continuous
- Event E: a subset of the sample space ($E\subset S$)
- A probability is a function that assigns to each $E\subset S$ a number $P(E)$ such that
  - $0\leq P(E)\leq 1$
  - $P(S)=1$
  - $P(E_1\cup E_2\cup\cdots)=P(E_1)+P(E_2)+\cdots$, if $E_i\cap E_j=\emptyset$ for all i,j (finite or infinite union or sum)
- Probability space (S,E,P) where S is the sample space, E is the set of possible events and P is a probability function
  - Often (not always) S is finite and all outcomes are equally likely, then $P(E)=\frac{\#outcomes\ in\ E}{\#outcomes\ in\ S}$
- Properties:
  - $P(E)+P(E^C)=P(S)=1$,
    - $P(E^C)=1-P(E)$
  - $P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)$
  - $P(E_1\cup E_2\cup\cdots\cup E_n)=\sum_{i=1}^{n}P(E_i)-\sum_{i<j}P(E_i\cap E_j)+\sum_{i<j<k}P(E_i\cap E_j\cap E_k)-\cdots+(-1)^{n-1}P(E_1\cap E_2\cap\cdots\cap E_n)$ (Generalization to n events)

Conditional probability

- Suppose $P(F)>0$, define $P(E|F)=\frac{P(E\cap F)}{P(F)}$ (conditional probability of E given that F occurs)
- Frequency interpretation: perform experiment repeatedly. Ignore all cases where F does not occur. Report fraction where E does occur
- $P(.|F)$ is a probability function where . is any event
- Note: by definition $P(E\cap F)=P(E|F)P(F)$

Independent events

- Definition: E and F are independent events if $P(E\cap F)=P(E)P(F)\Leftrightarrow P(E|F)=P(E)$
- More generally, $E_1,E_2,\ldots,E_n$ are independent if $P(E_{i_1},E_{i_2},\ldots,E_{i_r})=P(E_{i_1})P(E_{i_2})\ldots P(E_{i_r})$ for any subset $\{i_1,i_2,\ldots,i_r\}$
- Note: independence ($P(E\cap F)=P(E)P(F)$) is different from disjointedness ($P(E\cap F)=0$)

Theorem: Let $F_1,\ldots,F_n$ be a partition of S, i.e. and $F_i\cap F_j=\emptyset$ for all $i,j\in\{1,\ldots,n\}$. Let E be any event. Then:

- $P(E)=\sum_{i=1}^{n}P(E\cap F_i)=\sum_{i=1}^{n}P(E|F_i)P(F_i)$ (law of total probability)
- $P(F_j|E)=\frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n}P(E|F_i)P(F_i)}$ (Bayes theorem)

Monty hall problem:
door 1, 2, 3, one contains a car, other two contain goats.
If we pick door #1, the probability we picked a car is $\frac{1}{3}$
Monty reveals door 2 or door 3, showing a goat
Assume: Monty always reveals a goat and if you pick the car at first, he reveals a goat at random
Analysis 1:

Case	D1	D2	D3	P	Monty	Result if switch
a	g	g	c	1/3	Open 2	win
b	g	c	g	1/3	Open 3	win
c	c	g	g	1/3	Open 2 or 3	lose

$P(win\ by\ switching) = \frac{2}{3}$
Analysis 2:
We pick 1 and Monty opens 3
$$P(win\ by\ switching) = P(b|3) = \frac{P(3|b)P(b)}{P(3)} = \frac{P(3|b)P(b)}{P(3|a)P(a) + P(3|b)P(b) + P(3|c)P(c)}$$
$$= \frac{1 \cdot \frac{1}{3}}{0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}$$

If 100 doors and 99 goats, $P(win\ by\ switching) = \frac{99}{100}$ (except you choose first one correctly)

Discrete random variables

January 18, 2021 3:17 PM

Definition: a random variable ($r.v.$) is a function $X: S \to \mathbb{R}$

Notations

- A random variable will be capital letters $X, Y, Z, \ldots$
- Real numbers will be x, y, z
- $\{X = x\}$ would be an example of an event

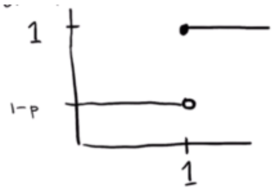
A random variable is discrete if it only takes values in a countable set $\{x_1, x_2, x_3, \ldots\} \subset \mathbb{R}$

- A discrete random variable is defined in terms of a probability mass function $(p.m.f.)$ p
  - $p(a) = P(X = a)$
  - $\Sigma_i p(x_i) = 1$
- Examples
  - Bernoulli r.v. ($X \sim Ber(p)$): fix $p \in [0,1]$,
    then $p(1) = P(X = 1) = p, p(0) = P(X = 0) = 1 - p$
    - Common usage: given an event E, let $I_E = \begin{cases} 1, if\ E\ occurs \\ 0, if\ E\ does\ not\ occur \end{cases}$

      Then I_E is a Bernoulli r.v. with $p = P(E)$

Definition: Cumulative distribution function ($c.d.f.$) of a random variable X is $F_X(a) = P(X \le a)$

- For Bernoulli random variable



Geometric random variable

- Definition: perform a sequence of trails, each successful with probability p (Bernoulli trials).
  Think of 1 as success, 0 as fail.
  Let $X = trial\ number\ of\ the\ first\ success$
  We say $X \sim Geom(p)$ (X is distributed as a geometric random variable) with
- $p(i) = P(X = i) = P(i - 1\ fails, then\ success) = (1 - p)^{i-1} p$
- Properties: $\sum_{i=1}^{\infty} p(i) = 1$
- No memory property: $P(X > m + n | X > m) = P(X > n)$

Binomial random variable

- Definition: perform n independent Bernoulli trials. Success with probability p and fail with $1 - p$
  Let $X = \#successes = \sum_{i=1}^{n} I_{si}$, we say $X \sim Bin(n, p)$ with
- $p(i) = P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$, n is number of sequences with i successes and $n - i$ fails
- $I_{si} = 1$ if trial i is a success,I_{si} means indicator of success at i

Poisson random variables with parameter $\lambda > 0$

- Arises as an approximation to binomial random variable. Suppose $X \sim Bin(n, p)$ with n large, p small but $\lambda = np$ is fixed, $X \sim Poisson(\lambda)$
- $p(i) = P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$, for $i = 0,1,2, \ldots$ s
- Comparing with binomial ($P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} = \frac{\lambda^i}{i!} \cdot \frac{n(n-1)\ldots(n-i+1)}{n^i} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i}$)

- Interpretations of λ:
  If $X \sim Bin(n,p)$, then np represents the average number of successes in n trials

Expectation of a discrete random variable

- Def: for a discrete random variable X taking values $\{x_1, x_2, x_3, \ldots\}$, $E(X) = \Sigma_i x_i p(x_i) = \Sigma_i x_i P(X = x_i)$
- Examples
  - $X \sim Ber(p), E(X) = p$
  - $X \sim Bin(n,p), E(X) = np$
  - $X \sim Geom(p), E(X) = 1/p$
  - $X \sim Poisson(\lambda), E(X) = \lambda, E(X^2) = \lambda + \lambda^2$
- Suppose X is a discrete random variable with values $\{0,1,2,3,\ldots\}$, then $E(X) = \Sigma_0^\infty P(X > n)$
- $E(g(x)) = \Sigma_i g(x_i) P_X(x_i)$ where P_X is probability mass function of X

Joint distribution: X, Y have joint probability mass function $p(x,y) = P(\{X = x\} \cap \{Y = y\})$

- Marginal probability mass function of X is $P_X(x) = P(X = x) = \Sigma_y p(x,y)$
- For Y is $P_Y(y) = \Sigma_x p(x,y)$
- $\Sigma_{x,y} p(x,y) = \Sigma_x P_X(x) = \Sigma_y P_Y(y) = 1$

Sum of independent random variables

- If X, Y are independent Poisson random variables with parameters λ_1 and λ_2, $X \sim Poisson(\lambda_1), Y \sim Poisson(\lambda_2)$, then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$
- If $X \sim Bin(n,p)$ and $Y \sim Bin(m,p)$ are independent, then $X + Y \sim Bin(m + n, p)$

Conditional expectation

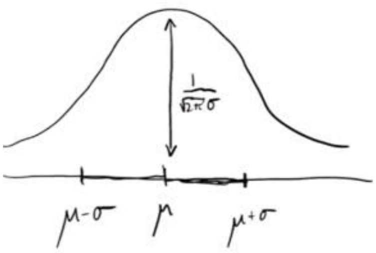
Let X, Y be two discrete random variables

- The conditional probability mass function of X given $Y = y$ is $P_{X|Y} = P(X = x|Y = y) = \frac{P(x,y)}{P_Y(y)}$
- The conditional expectation of X given $Y = y$ is $E[X|Y = y] = \Sigma_x x P_{X|Y}(x|y)$
  - $E[X|Y = y]$ depends on Y (is a function of y)
  - It is the average value of X in the sample space $\{Y = y\}$
  - Theorem: $E(X) = \Sigma_y P_Y(y) E[X|Y = y] = E(E(X|Y))$
  - Memoryless property gives that $E[X|X > x] = x + E[X]$

Continuous random variables

January 25, 2021 11:21 AM

Def: X is a continuous random variable if there exists a function $f(x)$, $x \in \mathbb{R}$ with $f(x) \geq 0\ \forall x$ and $P(X \in B) = \int_B f(x)dx, \forall B \subset \mathbb{R}$

- Interpretation of f:
    - For $B = \left[a - \frac{\epsilon}{2}, a + \frac{\epsilon}{2}\right)$ with ϵ small, $P(X \in B) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f(x)dx \approx \epsilon f(a)$
    - $f(a)$ indicates how likely it is for X to be near a, but $f(a)$ is not the probability of any event
    - It is possible $f(a) > 1$
    - f is called the probability density function of X
      Note: for all probability density function f, $\int_{-\infty}^{\infty} f(x)dx = 1$
- Examples
    - Uniform random variable on $[c, d]$ $X \sim Unif(c, d)$
      $f(x) = \frac{1}{d-c}$ for $x \in [c, d]$, 0 other wise
    - Exponential random variable with $\lambda > 0$ $X \sim Exp(\lambda)$
      $f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, x \leq 0 \end{cases}$
        - Half life of exponential random variable
          $X \sim Exp(\lambda)$ with probability density function $f(x) = \lambda e^{-\lambda x}$, τ is the time such that $P(X > \tau) = \frac{1}{2}$, i.e. $\tau = \frac{\log 2}{\lambda}$
            - No memory property gives: $P(X > 2\tau | X > \tau) = P(X > \tau) = \frac{1}{2}$
            - $P(X > s + t) = P(X > s)P(X > t)$
    - Normal (Gaussian) random variable $X \sim N(\mu, \sigma^2)$
        - μ is the mean value, σ^2 is the variance
        - $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}}$
            - $\int_{-\infty}^{\infty} f(x)dx = 1$
        - 

        - Standard normal ($X \sim N(0,1)$) has $\mu = 0$ and $\sigma = 1$
        - Scaling property: if $X \sim N(\mu, \sigma^2)$ and $Y = \frac{X-\mu}{\sigma}$, then $Y \sim N(0,1)$
        - $E(X) = \mu, E(X^2) = \mu^2 + \sigma^2$
        - If $X \sim N(\mu, \sigma^2)$, and $Y = aX + b$, then $Y \sim (a\mu_X + b, a^2\sigma^2)$

Cumulative distribution function: $F(a) = P(X \leq a) = P(X \in (-\infty, a]) = \int_{-\infty}^{a} f(x)dx$

- $F'(a) = f(a)$
- Example
    - Exponential random variable, for $a \geq 0, P(X \geq a) = e^{-\lambda a}$
      $F(a) = P(X \leq a) = 1 - e^{-\lambda a}$
        - It has the memoryless property ($P(X > s + t | x > s) = P(X > t)$)
    - Gaussian random variable $\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$
- Given $f_X(x)$, known $Y = X^2$, we can get the CDF of Y by $P(Y \leq y) = P(X^2 \leq y) = P(|X| \leq \sqrt{y})$

Cauchy distribution ($X \sim Cauchy$):

- Density of $X = \tan\theta$ where $\theta \sim Unif\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$
- Probability density function is $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$

Expectation

- Def: expectation for a continuous random variable X with probability density function f is $E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- It may not be the median which halves the mass
- Examples
  - $X \sim Unif(a,b)$, $E(X) = \frac{a+b}{2}$
  - $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$
    - On average, event occurs at time $\frac{1}{\lambda}$, so rate of occurrence is λ per unit time
    - $E(X^2|X > 1) = E((X+1)^2)$.
  - $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$
  - $X \sim Cauchy$, $E(X)$ is undefined, it has a median but not a mean
- Suppose X is a continuous random variable with probability density function f ($f(x) = 0 \; \forall \, x \le 0$). Then $E(X) = \int_0^{\infty} P(X > x)dx$
- Law of the unconscious statistician: for a continuous random variable X and function $g: \mathbb{R} \to \mathbb{R}$, then $E(g(x)) = \int_{-\infty}^{\infty} f(x)g(x)dx$ is the probability density function of X
- Linearity: $E(aX + b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx = aE(X) + b$

Moments

- nth moments of X is $E(X^n) = \begin{cases} \int_{-\infty}^{\infty} x^n f(x)dx, if\ continuous \\ \Sigma_i x_i^n p(x_i), if\ discrete \end{cases}$
- Often write mean $\mu = E(x)$
- Variance $\sigma^2 = Var(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2$
  - $X \sim Bin(n,p)$, $Var(X) = np(1-p)$
  - $X \sim Poisson(\lambda)$, $Var(X) = \lambda$
  - $X \sim Exp(\lambda)$, $Var(X) = \frac{1}{\lambda^2}$
  - $X \sim N(\mu, \sigma^2)$, $Var(X) = \sigma^2$
  - $X \sim Unif(a,b)$, $Var(X) = \frac{(b-a)^2}{12}$
  - $Var(cX) = c^2 Var(X)$, $Var(c + X) = Var(X)$
  - If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$
    - Generally, $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- Standard deviation $\sigma = \sqrt{Var(X)}$
  - Measures the width of the distribution

If X, Y are jointly continuous with probability density function $f(x,y)$

- $P((X,Y) \in C) = \int\int_C f(x,y)dxdy$
- Normalization: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$
- Often $C = A \times B$ is regular, then $P(X \in A, Y \in B) = \int_B \int_A f(x,y)dxdy$
- Marginal probability density function of X is $\mathrm{P(X \in A) = P(X \in A, Y \in \mathbb{R})} = \int_{-\infty}^{\infty}\int_A f(x,y)dxdy$, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$
- Marginal probability density function of Y is $\mathrm{P(Y \in B) = P(X \in \mathbb{R}, Y \in B)} = \int_B \int_{-\infty}^{\infty} f(x,y)dxdy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$

2D law of unconscious statistician $E\left(g(x,y)\right) = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)g(x,y)dxdy, if\ continuous \\ \Sigma_{x,y} g(x,y)p(x,y), if\ discrete \end{cases}$

- $E(X+Y) = E(X) + E(Y)$

Independent random variables

- Def: X and Y are independent if $P(\{X \le a\} \cap \{Y \le b\}) = P(\{X \le a\})P(\{Y \le b\})$ for $a, b \in \mathbb{R}$
  - i.e. $\{X \le a\}$ and $\{Y \le b\}$ are independent
  - Cumulative distribution function: $F_{XY}(a,b) = F_X(a)F_Y(b) \forall\, a, b$
  - Probability mass function $p(x,y) = p_X(x)p_Y(y)$ for discrete, $f(x,y) = f_X(x)f_Y(y)$ for continuous
- If X, Y are independent random variables, then $E(XY) = E(X)E(Y)$
- If X, Y are independent, $Z = \max(X,Y)$, then $F_z(a) = P(\max(X,Y) \le a) = F_x(a)F_y(a)$
- Known $f_X(x)$ and $f_{Y|X}(y|x)$, then $f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$

Problem 5

Suppose that the number of customers visiting a fast food restaurant in a given day is $N \sim Poisson(\lambda)$. Assume that each customer purchases a drink with probability p, independently from other customers, and independently from the value of N. Let X be the number of customers who purchase drinks. Let Y be the number of customers that do not purchase drinks; so $X + Y = N$.

a. Find the marginal PMFs of X and Y.
b. Find the joint PMF of X and Y.
c. Are X and Y independent?
d. Find $E[X^2Y^2]$.

Solution

a. First note that $R_X = R_Y = \{0, 1, 2, \dots\}$. Also, given $N = n$, X is a sum of n independent $Bernoulli(p)$ random variables. Thus, given $N = n$, X has a binomial distribution with parameters n and p, so

$$\begin{aligned} X|N = n &\sim Binomial(n, p), \\ Y|N = n &\sim Binomial(n, q = 1 - p). \end{aligned}$$

We have

$$\begin{aligned} P_X(k) &= \sum_{n=0}^{\infty} P(X = k|N = n)P_N(n) & \text{(law of total probability)} \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} exp(-\lambda) \frac{\lambda^n}{n!} \\ &= \sum_{n=k}^{\infty} \frac{p^k q^{n-k} exp(-\lambda) \lambda^n}{k!(n-k)!} \\ &= \frac{exp(-\lambda)(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{(\lambda q)^{n-k}}{(n-k)!} \\ &= \frac{exp(-\lambda)(\lambda p)^k}{k!} exp(\lambda q) & \text{(Taylor series for } e^x) \end{aligned}$$

Covariance

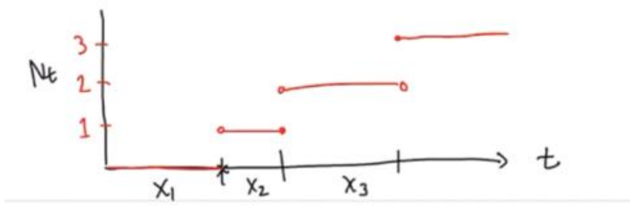
- Def: the covariance of X, Y is $Cov(X,Y) = E\big((X - E(X))(Y - E(Y))\big) = \sum P(x,y)(x - E(X))\big(y - E(y)\big)$
  - Note: $Cov(X,X) = Var(X)$
  - Formula: $Cov(X,Y) = E(XY) - E(Y)E(X) = \sum xyP(x,y) - E(X)E(Y)$
    - And $Cov(aX, bY) = abCov(X,Y)$
- If X and Y are independent, then $Cov(X,Y) = 0$. The opposite is not true
- Interpretation:
  - If $Cov(X,Y) > 0$, X, Y tend to be large together or small together
  - If $Cov(X,Y) < 0$, X tends to be large when Y is small

- Correlation coefficient: $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$
  - Cauchy Schwartz inequality: $|E(XY)|^2 \le E(X^2)E(Y^2)$
  - The Cauchy Schwartz inequality gives that $|\rho(X,Y)| \le 1$

Sum of independent variables

- If X, Y are continuous and independent, then $F_{X+Y}(a) = P(X+Y \le a) = \iint_{x+y\le a} f_X(x)f_Y(y)dxdy$
  Then, $F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy$
  Differentiating both sides with respect to a gives: $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$
- Density of the sum is the convolution of the densities
- If $X_i \sim Exp(\lambda)$, then $f_{X_1+X_2}(x) = \lambda^2 x e^{-\lambda x}$
  - More generally, $f_{X_1+\cdots+X_n}(x) = \begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, x \ge 0 \\ 0, x < 0 \end{cases}$
  - This is called the $Gamma(n, \lambda)$ random variable, with $E(X) = \frac{n}{\lambda}, Var(X) = \frac{n}{\lambda^2}$

Continuous time stochastic process

- Poisson process
  - For $t \ge 0$, let N_t be the number of jobs completed by time t, N_t is called the Poisson process
  - 

  - $P(N_t \ge n) = P(X_1 + .. + X_n \le t) = -\frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} + P(N_t \ge n-1)$, $E(N_t) = \lambda t$
  - So $P(N_t = m) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$, $N_t \sim Poisson(\lambda t)$, $f_{S_n}(s) = \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!}$
    - $E(S_n) = \frac{n}{\lambda}$ is the expected time of n-th event $S_n \sim Gamma\ (n, \lambda)$, $Var(S_n) = \frac{n}{\lambda^2}$
    - $E(N_t) = Var(N_t) = \lambda t$ is the number of events completed by time t
    - $S_n > t$ is equivalent to $N_t < n$
  - Given two Poisson process with parameter λ_1, λ_2
    - The probability of observing event 1 first is $\frac{\lambda_1}{\lambda_1+\lambda_2}$
  - No arrival in t means $P(S_1 > t) = e^{-\lambda t}$, $S_1 \sim Exp(\lambda)$.

Conditional expectation

- If X, Y are jointly continuous random variables, then the conditional probability density function of X given $Y = y$ is $f_{X|Y} = \frac{f(x,y)}{f_Y(y)}$
- The conditional expectation of X given $Y = y$ is $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$
- Properties:
  - Linearity: $E[aX|Y=y] = aE[X|Y=y]$, $E[X_1 + X_2|Y=y] = E[X_1|Y=y] + E[X_2|Y=y]$
  - Monotonicity: if $X_1 \le X_2$, then $E[X_1|Y=y] \le E[X_2|Y=y]$
- $P(X|X>1) = \frac{f(x)}{P(x>1)}$. Memoryless property gives that $E[X|X>x] = x + E[X]$
- If X, Y independent, $f_{X|Y} = f_X$

If $Y = g(X)$, then, $F_Y(y) = F_X\left(g^{-1}(y)\right)$, $f_Y(y) = \frac{f_X(x)}{g'(x)}$.

Example 5.25

Let X and Y be two independent $Uniform(0,1)$ random variables. Find $P(X^3+Y>1)$.

Solution

Using the law of total probability (Equation 5.16), we can write

$$\begin{aligned} P(X^3+Y>1) &= \int_{-\infty}^{\infty} P(X^3+Y>1 | X=x) f_X(x) \ dx \\ &= \int_0^1 P(x^3+Y>1 | X=x) \ dx \\ &= \int_0^1 P(Y>1-x^3) \ dx && \text{(since } X \text{ and } Y \text{ are independent)} \\ &= \int_0^1 x^3 \ dx && \text{(since } Y \sim Uniform(0,1)\text{)} \\ &= \frac{1}{4}. \end{aligned}$$

Characteristic functions

February 5, 2021 1:45 PM

Moment generating functions

- Def: the moment generating function of a random variable X is $M(t) = E(e^{tx}) = \begin{cases} \Sigma e^{tx} p(x), X\ discrete \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx, X\ continuous \end{cases}$
  - Note that $E(e^{aX}) = \int e^{ax}\lambda e^{-\lambda x}dx$ if $X \sim Exp(\lambda)$
- Special cases
  - X discrete with values in $(0,1,2 \dots)$, then $M(t) = \sum_0^\infty (e^t)^n p(n)$ (let $z = e^t$, we have z transform)
  - X continuous with $f(x) = 0$ for $x < 0$, then $M(t) = \int_0^\infty e^{tx} f(x)dx$ (let $t = -s$, we have Laplace transform)
- Note: $\left.\frac{d^n}{dt^n}\right|_{t=0} M(t) = E(X^n)$ is the nth moment of X
  - Can also Taylor expand e^t, and find the coefficient of $\frac{t^k}{k!}$
- If X, Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$
  - The Laplace transform of convolution=product of Laplace transform ($\int_0^\infty e^{-sx} f_{X+Y}(x)dx = \int_0^\infty e^{-sx} f_X(x)dx \int_0^\infty e^{-sy} f_Y(y)dy$)
- $M(t)$ may not always exist
  - $X \sim Exp(\lambda)$ has $M(t) = \int_0^\infty e^{tx}e^{-\lambda x}dx = \frac{\lambda}{\lambda - s}$, is infinite for $t > \lambda$
  - $X \sim N(\mu, \sigma)$, $M_X(s) = e^{s\mu + \frac{\sigma^2 s^2}{2}}$
  - $X \sim Poisson(\lambda)$, $M_X(s) = e^{\lambda(e^s - 1)}$

Characteristic functions

- Def: $\phi(t) = M(it) = E(e^{itx}) = \begin{cases} \Sigma e^{itx} p(x), X\ discrete \\ \int_{-\infty}^{\infty} e^{itx} f(x)dx, X\ continuous \end{cases}$ is the characteristic function
  - If vector values, we have tx to be $t \cdot x$
- Properties
  - $\phi(t)$ always exists, $|\phi(t)| \le 1$
  - Always $\phi(0) = 1$
  - If X, Y independent, $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
    - Fourier transform of convolution=product of Fourier transform
  - If $Y = aX + b$, then $\phi_Y(t) = \phi_{aX+b}(t) = e^{itb}\phi_X(at)$
- Example
  - If $X \sim Exp(\lambda)$, $\phi_X(t) = \int_0^\infty e^{itx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - it}$
    - If $X_i \sim Exp(\lambda)$, $S_n = \Sigma X_i$, then $\phi_{S_n}(t) = \left(\frac{\lambda}{\lambda - it}\right)^n$, $\phi_{\frac{S_n}{n}}(t) = \phi_{S_n}\left(\frac{t}{n}\right) = \left(\frac{\lambda}{\lambda - \frac{it}{n}}\right)^n \to e^{\frac{it}{\lambda}}$
  - $X \sim N(0,1)$, $\phi_X(t) = e^{-\frac{t^2}{2}}$
  - $Y \sim N(\mu, \sigma^2)$, $\phi_Y(t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}$
  - Constant random variable $X = c \in \mathbb{R}$, $\phi_X(t) = e^{itc}$
- Note: $\phi(t)$ contains all info about distribution of X, $\left.\frac{d^n}{dt^n}\right|_{t=0} \phi(t) = i^n E(X^n)$.
  - So $E(X^n) = \frac{1}{i^n}\phi^{(n)}(0)$
- Inversion theorem: If X is a continuous random variable with probability density function f, then $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt$ at every x for which f' exists
  - For $X \sim Exp(\lambda)$, f' is discontinuous at 0, so inverse FT at 0 is $\frac{f(0^+) + f(0^-)}{2}$

Convergence of random variables

- Convergence in distribution: let Y_1, Y_2 be random variables with CDFs $F_{Y_1}, F_{Y_2}, \dots$ We say $Y_n \to Y$ for some random variable Y with CDF F_Y if $\lim_{n \to \infty} F_{Y_n}(x) = F_Y(x)$ for each x where $F_Y(x)$ is continuous
- Continuous theorem: let $X_1, X_2, \dots$ be random variables with CDFs $F_1, F_2, \dots$ and characteristic functions $\phi_1, \phi_2, \dots$

- If $F_n \to F$, then $\phi_n(t) \to \phi(t)$
- If $\phi_n(t) \to \phi(t)$ exists $\forall t \in \mathbb{R}$ with ϕ continuous at 0, then ϕ is the characteristic function of some random variable X and $F_n \to F$, i.e. $X_n \to X$
- Uniform random variable does not converge ($\phi(t)$ is discontinuous at 0)
- Exponential random variable converges to $Y = \frac{1}{\lambda}$,
  and $F_{Y_n}(b) - F_{Y_n}(a) = P(a < Y_n \le b) \to F_Y(b) - F_Y(a) = P(a < Y \le b)$

- Weak law of large numbers: let $X_1, X_2, \ldots$ be independent and identically distributed. Assume $\mu = E(X) < \infty$ (not Cauchy). Let $S_n = X_1 + \cdots + X_n$, then $\frac{S_n}{n} \to \mu$
- Strong law of large number: $P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1$
- Central limiting theorem (convergence to a random variable that is not constant)
  - Let X_i be independent and identically distributed with $E(X_i) < \infty$ and $Var(X_i) = \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. Then, $\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$
    - i.e. $\lim_{n\to\infty} P\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$
  - Note: distribution of X_i is arbitrary, as long as $\mu, \sigma < \infty$
  - This implies that $S_n \approx n\mu + \sigma\sqrt{n}Z$
    - i.e. $\frac{1}{n} S_n \approx \mu + \frac{\sigma}{\sqrt{n}} Z$
  - Interpretation: the typical fluctuation of $S_n - n\mu$ is roughly $\sigma\sqrt{n}$
  - It can be viewed as $\frac{X - n\mu}{\sqrt{n\,Var(X)}}$
    - For binomial distribution, $\frac{X - n\mu}{\sqrt{np(1-p)}} \to N(0,1)$
    - For discrete cases $P(X > n) = P(X \ge n + 0.5) = P\left(Z \ge \frac{n + 0.5 - n\mu}{\sqrt{nVar(X)}}\right)$
      - $P(a \le x \le b) = P(a - 0.5 \le x \le b + 0.5)$.

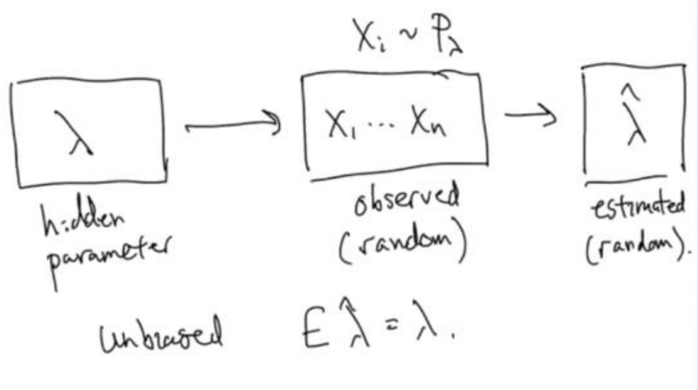
Markov's inequality: $P(X \ge a) \le \frac{E(X)}{a}$

Chebyshev's inequality: $P(|X - \mu| \ge k) \le \frac{\sigma^2}{k^2}$

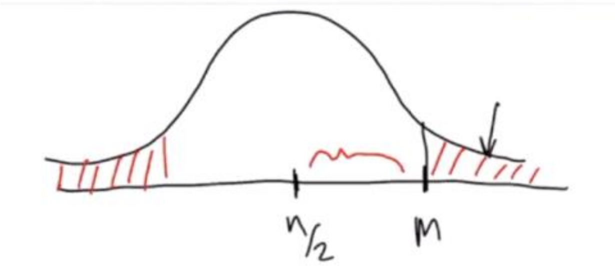
Statistical estimation, hypothesis testing

February 26, 2021 2:57 PM

Statistical estimation

- Given samples from some distribution P_λ depending on an unknown parameter λ, recover λ from samples $X_1, \dots X_n$
- Def: an estimator is a function of data
  - Sample mean: $\overline{X} = \frac{1}{n}\Sigma_{i=1}^{n} X_i$
  - Sample variance: $s^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}\left(X_i - \overline{X}\right)^2$
    - $n-1$ makes s^2 unbiased estimation for σ^2 $E(s^2) = \sigma^2$
  - $\overline{X}$ is an unbiased estimate of μ, $E(\overline{X}) = E(X_i)$
  - $\overline{X}$ has lower variance, $Var(X) = \frac{1}{n^2} Var\left(\Sigma_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$
  - Distribution of $\overline{X}$ is more narrowly centered around μ as n increases
    - Consistent with law of large numbers and central limiting theorem
- 

Hypothesis testing

- Consider a hypothesis H generating data, we want to know if the data is consistent with the hypothesis
- We check $P(observation\ or\ less \mid H)$ ($P(observation|H) = 0$ in most cases)
- 

- reject the hypothesis when it is outside the 95% CI
  - Note: the interval shrinks when $n \to \infty$

Confidence interval

- Assume $X_i \sim N(\mu, \sigma^2)$, independent and identically distributed, σ^2 known and μ not known
- Law of large number says
  - $\overline{X} \approx \mu,$
  - $\overline{X} - \mu = \frac{1}{n}\Sigma(X_i - \mu)$ has variance $\frac{\sigma^2}{n}$
  - $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to N(0,1)$
  - $P(|Z| < 1.96) \approx 0.95.$
- This means that $\overline{X} \in \left[\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}\right]$ with probability 95%

- i.e. $\mu \in \left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$ with probability 95%
- This is the 95% confidence interval for μ

- We usually reject if $P\left(\left|\overline{X} - \mu\right| > a\right) = 2P\left(\frac{\left|\overline{X}-\mu\right|}{\sigma} > \frac{a}{\sigma}\right) = 2P\left(Z > \frac{a}{\sigma}\right) = 0.05$
  - $\overline{X}$ is the sample mean, μ is the hypothesis mean, we want to find a first, by distribution of $\overline{X}$, reject the hypothesis when it is outside the 95% CI
    - Note: the interval shrinks when $n \to \infty$
  - Given a, we can reject if $\left|\overline{X} - \mu\right| > a$, and we would be 95% right
    - 95% sure that the hypothesis is wrong
  - 0.05 is the p value
  - If $\left|\overline{X} - \mu\right| \leq a$, we conclude nothing (this happens 95% of the time under the hypothesis)
  - Can also think about in an estimation perspective ($Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$)
    - $\left|\overline{X} - \mu\right| \leq \frac{1.96\sigma}{\sqrt{n}}$ holds with probability 95%

Def: a statistic is a number you compute to determine a hypothesis test

Now suppose μ, σ^2 both unknown, let $X_1, \ldots X_n \sim N(\mu, \sigma^2)$ with sample mean $\overline{X}$ and sample variance s^2. Then $T = \frac{\overline{X}-\mu}{s/\sqrt{n}}$ has a student-t distribution with $n-1$ degree of freedom

- This means that $T = \frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$, we want to find $a \in \mathbb{R}$ such that $P(|T| > a) = 0.05$, and reject if $|T| > a$
- To find the 95% CI, $0.95 = P\left(\frac{\left|\overline{X}-\mu\right|}{s/\sqrt{n}} \leq a\right)$, so the interval is $\mu \in \left[\overline{X} - \frac{as}{\sqrt{n}}, \overline{X} + \frac{as}{\sqrt{n}}\right]$
-

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
n-1 3	0.765	0.978	1.250	1.638	2.353	3.182 a	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140

Random Walks & Markov chains

January 18, 2021 3:17 PM

Example: Gambler's Ruin

- Gambler has k dollars and bank has b dollars. Play fair game betting \$1, until one goes broke
- Let $q(k) = P(reach\ N\ before\ 0\ starting\ at\ k)$
  $= \frac{1}{2}P(win|win\ 1st\ game) + \frac{1}{2}P(win|lose\ 1st\ game)$
  $= \frac{1}{2}\big(q(k+1) + q(k-1)\big)$
- This gives $P(win) = \frac{k}{N}$ and $P(lose) = 1 - \frac{k}{N}$
- If unfair with probability p for win, we will have $q(k) = pq(k+1) + (1-p)q(k-1)$
  - This gives $q(k) = \frac{\alpha^k - 1}{\alpha^N - 1}$, where $\alpha = \frac{1-p}{p}$
  - It satisfies the $\frac{1}{2}$ probability case

Simple random walks on $\mathbb{Z}^d$ (points in d-dimensional space with integer components)

- Let e_j be unit vectors in $\mathbb{Z}^d$, walk take steps X_i with probability mass vector $P(X_i = e_j) = P(X_i = -e_j) = \frac{1}{2d}$
- Determine $u = P(walk\ will\ return\ to\ origin) = P(\exists n\ such\ that\ S_n = 0)$, let M be the number of visits to 0 (counting $S_0 = 0$)
  - $P(return\ twice|return\ once) = P(return\ once)$
  - $P(M = k) = u^{k-1}(1-u)$, $E(M) = \frac{1}{1-u}$
  - M is recurrent if $u = 1$, $E(M) = \infty$ (always come back), transient if $u < 1$, $E(M) < \infty$
  - To find u, we need to find $E(M)$, since $u = 1 - \frac{1}{E(M)}$
    - $E(M) = \Sigma \binom{2n}{n} p^n (1-p)^n$. It converges if $4p(1-p) < 1$, using Stirling formula, this gives $p \neq 1/2$

Characteristic functions for vector functions

- For $X \in \mathbb{R}^d$, $t \in \mathbb{R}^d$, $\phi(t) = E(e^{i<t,X>})$
- Character function of $S_n = \phi_n(k) = E\big(e^{i<k,S_n>}\big) = E\big(e^{i<k,X_1+\cdots+X_n>}\big)$
  $= \phi_1(k) \dots \phi_n(k) = \phi(k_1, k_2, \dots, k_n)$
- Given $P(X_i = e_j) = \frac{1}{2d}$, we have $\phi_1(k) = \frac{1}{d}\Sigma_{j=1}^d \cos k_j$, and $\phi_n(k) = \left(\frac{1}{d}\Sigma_{j=1}^d \cos k_j\right)^n$
  - Then $P(S_n = b) = \left(\frac{1}{2\pi}\right)^d \int \phi_n(t) e^{it\cdot b} dt_1 \dots dt_d$, $E(M) = \left(\frac{1}{2\pi}\right)^d \int \frac{dt_1 \dots dt_d}{1 - \phi_1(t)}$
  - If $d = 1$, $\phi(t) = \cos t$, $E(M) = \infty$, reccurent
  - In general, $\int \frac{dt_1 \dots dt_d}{1-\phi_1(t)} = \begin{cases} \infty, n \le 2 \\ < \infty, else \end{cases}$

Theorem: random walk in $\mathbb{Z}^d$ is recurrent for $d = 1,2$, transient for $d > 2$

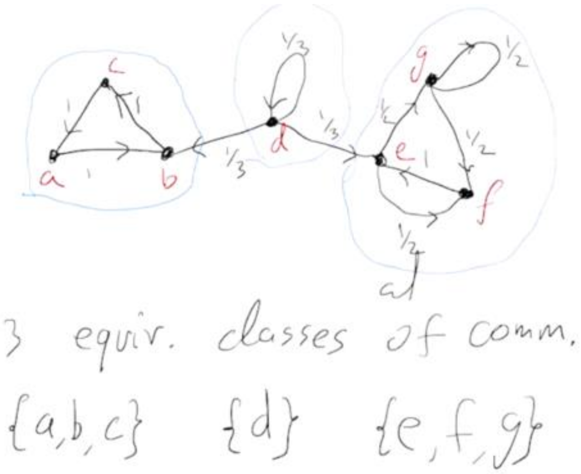
- A drunk person will eventually walk home
- A drunk bird will not. In $\mathbb{Z}^3$, $P(return\ to\ 0) = 1 - \frac{1}{EM} = 0.34$

Stochastic process:

- A stochastic process is a sequence of random variables $X_0, X_1, \dots, X_n$
- Transition probabilities (one step): $P_{ij} = P(X_{n+1} = j|X_n = i)$ (can depend on n)

Markov chains

- A Markov chain is a sequence of random variables $X_0, X_1, \dots$ such that
  - $P_{ij} = P(X_{n+1} = j|X_n = i) = P(X_{n+1} = j|X_n = i, X_{n-1} = i-1, \dots, X_0 = i_0)$
  - Markov property: condition on $X_n = i$ is the same as condition on $X_1, \dots, X_n$

    - Assumption: P_{ij} does not depend on n
- State space={possible values for X}
- The transition matrix of a Markov chain is $P = \left(P_{ij}\right)_{i,j} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$
    - The rows always sum to 1 (stochastic matrix)
- N-step transition probability
    - $P_{ij}^n = P(X_n = j|X_0 = i) = P\left(X_{t+n} = j\middle|X_t = i\right)$ for any t
        - $P = \left(P_{ij}\right)_{ij}$ and $P^n = \left(P_{ij}^n\right)_{ij}$ are both matrices
    - Chapman-Kolmogorov's theorem: P^n is the nth power of P
- Classification of states
    - A state i is called absorbing or a sink if $P_{ii} = 1$
        - 0 or N is absorbing in Gambler's ruin
    - j is accessible from i if $P_{ij}^n > 0$ for some n
    - State i, j are communicating if each is accessible from the other ($i \leftrightarrow j$)
        - Communication is an equivalent relation
            - $i \leftrightarrow j \leftrightarrow k$, then $i \leftrightarrow k$
            - $i \leftrightarrow i$ for all states
        - 

        - A Markov chain is irreducible if for all states i, j, $i \leftrightarrow j$
            - Equivalently, a Markov chain is irreducible if for all i, j, $\exists n$ such that $P_{ij}^n \neq 0$.
    - A state i is recurrent if condition on $X_0 = i$, the chain returns to i with probability 1. Otherwise, the state is transient.
        - i is recurrent if $f_i = 1$, $\Sigma_n P_{ii}^n = \infty$.
        - i is transient if $f_i < 1$, $\Sigma_n P_{ii}^n < \infty$. f_i is the probability of return
            - Note, if we let N_i be the total number of visits to state i, $N_i = \Sigma 1_{X_n = i}$, $M = E[N_i|X_0 = i] = \Sigma_n P_{ii}^n$
            - If $X_n = i$, by Markov property, $P(\exists n' > n\colon X_n = i|X_n = i) = f_i$
                - From i, we have probability of f_i to return, and $1 - f_i$ not return
            - $N \sim Geom(1 - f_i)$, $M = \frac{1}{1-f_i}$
        - Let $i \leftrightarrow j$, then i is recurrent if and only if j is recurrent (recurrent is a class property)
        - If a state in an irreducible Markov chain is recurrent, the Markov chain is recurrent.
    - Periodicity
        - A state i has period d if $d = GCD\{n\colon P_{ii}^n \neq 0\}$, i is aperiodic if $d = 1$
        - Period of a state is also a class property
- Behavior as $n \to \infty$
    - Let $V^{(n)}$ be the distribution for X_n
    - Then $V_j^{(n)} = P(X_n = j) = \Sigma P(X_n = j|X_0 = i)P(X_0 = i) = \Sigma V_i^{(0)} P_{ij}^n$
        - P^n is the nth matrix power

- Then $\left(V_0^{(n)}, V_1^{(n)}\right) = \left(V_0^{(n)}, V_1^{(n)}\right) P^n$
- Note: for any Markov chain, $\lambda = 1$ is always an eigen value for P, since row of P add to 1
- For every Markov chains, all eigen values have $|\lambda| \le 1$

2-state Markov chain

- Suppose $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$
- Then $\lambda_1 = 1, \pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right), \lambda_2 = 1-p-q, f = (1,-1)$
- $V^{(0)} = \pi + bf, V^{(n)} = (\pi + bf)P^n = \pi + b\lambda_2^n f$
- If $|\lambda_2| < 1$, then $V^{(n)}$ converges to π
  - π is the limiting distribution of V^n
  - π_i is the asymptotic proportion of time in state i
- If $|\lambda_2| = 1$
  - $p = q = 0$, reducible
  - $p = q = 1$, periodic with period 2

Let T_i be the return time to state i, $T_i = \inf\{n \ge 1:\ X_n = i\}$

- A recurrent state i is
  - Positive recurrent if $E(T_i | X_0 = i) < \infty$
  - Null recurrent if $E(T_i | X_0 = i) = \infty$
- Random walk in $\mathbb{Z}, \mathbb{Z}^2$ are null recurrent
- For any finite space Markov Chain, any recurrent state is positive recurrent
- Given π_i the stationary distribution, the mean return time is $\frac{1}{\pi_i}$.

An aperiodic, positive recurrent state is called ergodic

- If every state is ergodic, then the Markov chain is ergodic
- In any irreducible ergodic Markov chain, we have $\pi_j = \lim_{n\to\infty} P_{ij}^n$ for any i
  - Moreover, π is the unique solution to $\begin{cases} \pi P = \pi \\ \Sigma\pi_j = 1, \pi_j = \Sigma_i \pi_i P_{ij} \end{cases}$
- Let $N_j(n)$ =#visist to j up to time n. If the Markov Chain is irreducible and ergodic, then $\frac{N_j(n)}{n} \to \pi_j$
- If a Markov Chain is irreducible and ergodic, then $\pi_j = \frac{1}{m_j}$, where $m_j = E(T_j | X_0 = j)$
  - Note: positive recurrent means $m_j < \infty$

π is called the stationary measure or stationary distribution for the Markov chain

- $V^{(n)} \to \pi$ exponentially fast

If $P(X_n = j) \to V_j$, then $P(X_{n+1} = j) = \Sigma P(X_{n+1} = j | X_n = i) P(X_n = i) = \Sigma P(X_n = i) P_{ij}$
Taking $n \to \infty$, $V_j = \Sigma V_i P_{ij}$, so $V = VP$

If $V^{(0)} = \pi$, i.e. at time 0, $P(X_0 = i) = \pi_i$, then at any n, $V^{(n)} = V^{(0)} P^n = \pi$
In this case, every X_n has the same distribution, π is also called the equilibrium distribution

On $\mathbb{Z}^d$, there is no limit, since random walk is null-recurrent $P(X_n = x) \to 0$

If the Markov Chain is reducible, then limit and stationary distribution depends on the communicating class

If the Markov chain is periodic, then $\pi = \pi P$ still has a unique solution, but P_{ij}^n does not converge

If P is doubly stochastic (rows and columns sum to 1), then $\pi = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$

Time reversal

- Given Markov chain $(X_0, \ldots, X_N)$, consider the backward chain $Y_0, \ldots, Y_N$, given by $Y_i = X_{N-i}$, Y_n is a Markov Chain
- Given X with stationary distribution and $P(X_0 = i) = \pi_i$
  - with transition probability $Q_{ij} = P_{ji} \times \frac{\pi_j}{\pi_i}$
  - Y is the reverse or dual Markov Chain of X
- A Markov Chain is reversible if $Q_{ij} = P_{ij}$ for all i, j, or equivalently, $\pi_i P_{ij} = \pi_j P_{ji}$
  - Note: stationary, then mass out = mass in at each vertex
  - Reversible, then i sends to j the same as j sends to i
- If X is an irreducible ergodic Markov Chain and for some vector μ has $\mu_i P_{ij} = \mu_j P_{ji}$ (detailed balance equation)for all i, j, and $\Sigma \mu_i = 1$, then $\mu = \pi$ and X is reversible
  - If a Markov Chain is reversible, we can find π using detailed balance
  - If solved, then we can deduce π and reversibility
  - If not solvable, then the Markov Chain is not reversible
- Doubly stochastic Markov Chain is reversible only if $p = \frac{1}{2}$

A graph is a pair (V,E) where V is the set of vertices/nodes, E is the set of edges (pair of vertices)

- Simple graph: graph with no loops or double edges

Random walk on a graph G:

- State space: V
- $P_{ij} = \begin{cases} \frac{1}{\deg(i)}, (i,j) \in E \\ 0, else \end{cases}$, where $deg(i)$ is the number of edges containing i.
- In any finite graph, the stationary measure π is $\pi_i = \frac{\deg(i)}{2|E|}$, moreover, this Markov chain is reversible
  - $\Sigma_i \deg(i) = 2|E|$, since every edge is counted twice

Birth and death chains

- Assume arrivals at rate λ, departure at rate 1
  - Times of arrivals are Poisson process with rate λ
  - $P_{n,n+1} = \frac{\lambda}{\lambda+1}, P_{n,n-1} = \frac{1}{\lambda+1}$
- If $\lambda < 1$, π is a geometric distribution, size of queue is $Geom(\lambda) - 1$
- If $\lambda \geq 1$, no stationary distribution
- Every birth and death chain is reversible, but not always have a stationary distribution
  - $\lambda < 1$, positive recurrent
  - $\lambda = 1$, null recurrent
  - $\lambda > 1$, transient

Gambler's ruin with m transient states, K absorbing states

- $P = \begin{pmatrix} A & B \\ 0 & I_k \end{pmatrix}$, then A is $m \times m$, B is $m \times k$, $I_k = k \times k$ identity matrix
- Let $P_i(A) = P(A|X_0 = i)$, $q_i = P_i$(end at absorbing state a)
  - $q_a = 1, q_b = 0$ for $b \neq a$ absorbing.
  - $q_i = \Sigma_j P_{ij} q_j = P_{ia} + (Aq)_i$ ($q_j = P(end\ at\ a|X_0 = j)$)by Markov property after 1 step
  - This gives $q = (P_{ia})_i + Aq = (col\ a\ of\ B) + Aq = (I_m - A)^{-1}(col\ a\ of\ B)$.
- Let $N_j = \#visits\ to\ j$, then $S_{ij} = E_i N_j = E(N_j | X_0 = i)$
  - $S = (I - A)^{-1}$, since $N_j = \Sigma_k\ jumps(k \to j) + 1_{X_0 = j}$, $S_{ij} = \Sigma_k S_{ik} P_{kj} + \delta_{ij}$
- Let $f_{ij} = P_i(hit\ j\ at\ least\ once) = P(N_j \neq 0 | X_0 = i)$
  - Then $S_{ij} = E_i N_j = E_i(N_j | N_j = 0) P_i(N_j = 0) + E_i(N_j | N_j > 0) P_i(N_j > 0) = S_{jj} f_{ij}$
  - So $f_{ij} = \frac{S_{ij}}{S_{jj}}$, but $f_{ii} = \frac{S_{ii} - 1}{S_{ii}}$

Branching process

- Family tree
  - Let Z_n =size of generation n. Assume individual has a random number of children independent of all others, $P(k\ children) = p(k)$ given.
  - Two options
    - $Z_n > 0$ for all n.
    - $Z_n = 0$ for some n_0, then $Z_n = 0$ for all $n \geq n_0$, 0 is an absorbing state
- Nuclear explosion
  - Each generation of neutrons has a random size
  - Each neutron has 0 or 3 children with probability $p(0), p(3)$
  - If Z_n grows very quickly, we have explosion
    - This is possible if $p(3) > \frac{1}{3}$, critical mass is the size needed such that $p(3) > \frac{1}{3}$
  - If Z_n stays non-zero but small, we have reaction
- Let $\mu = E(Y)$, where Y =number of children of an individual, assume $p(1) \neq 1$, then $P(survival) > 0 \Leftrightarrow \mu > 1, P(survival) = 0 \Leftrightarrow \mu \leq 1$, where survival means $Z_n > 0$ for all n, extinction means $Z_n = 0$ for $n \geq n_0$.
  - If $Z_n = k$, then $Z_n = \Sigma_i Y_i$, $E(Z_{n+1}|Z_n = k) = \Sigma E(Y_i) = \mu k$
  - If $Z_0 = 1$, then $E(Z_1) = \mu, E(Z_n) = \mu^n$, so $\mu > 1$ $E\big(Z_n\big) \to \infty$
- Let $f(t)$ be the probability generating function for Y, $f(t) = E(t^Y) = \sum_{n=0}^{\infty} p(n)t^n$.
  - $f(1) = 1, f(0) = p(0)$.
  - $f' \geq 0$ (increasing), $f'(t) = \sum_{n=0}^{\infty} nt^{n-1}p(n)$, $f'(1) = \mu$.
  - $f'' \geq 0$ (convex)
  - If $\alpha = P(extinction)$, then α is the smallest solution of $\alpha = f(\alpha)$ in $[0,1]$
    - If $\mu \leq 1, \alpha = 1$.
    - If $\mu > 1, \alpha < 1$.
- Below each individual, we see a copy of the whole branching process

Metropolis Markov chain:

- Given some state space S and target distribution π, construct a connected graph on S
- Steps of the Markov Chain
  - Assume $X_n = x$, pick an edge e uniformly in the graph
  - If e far from x, do nothing, $X_n = x$.
  - If $e = (x, y)$, then jump to y with probability $\mathrm{P} = \min\left(\frac{\pi_y}{\pi_x}, 1\right)$, stay at x with probability $1 - P$.
- Reversible with respect to π.
- In hard square model $S = \{0,1\}^V$, V is the number of vertices, 0 is free, 1 is occupied
  - If $\sigma \in S$ has $\sigma_u = \sigma_v = 1$ for neighboring u, v, then $\pi_\sigma = 0$
  - If no adjacent ones, $\pi_\sigma = Z^{-1}\lambda^{N(\sigma)}$
    - $N(\sigma) = \sum_u \sigma_u$.
    - $Z = \sum_\sigma \lambda^{N(\sigma)}$ is the normalizing factor
  - Regardless of Z, we always have $\frac{\pi_\sigma}{\pi'_\sigma} = \lambda^{N(\sigma)-N(\sigma')}$
    - Graphically, σ connected to σ' if they differ at a single vertex u
    - To pick the edge, pick uniformly a vertex u, $\sigma' = \sigma$ with u flipped
    - If σ' has 1 less particle, $\frac{\pi_{\sigma'}}{\pi_\sigma} = \frac{1}{\lambda}$
    - If σ has 1 more particle, $\frac{\pi_{\sigma'}}{\pi_\sigma} = \lambda$.
    - If $\lambda < 1$:
      - If u full, remove particle
      - If u empty, add particle with probability λ
    - If $\lambda \geq 1$:
      - If u full, remove with probability $\frac{1}{\lambda}$
      - If u empty, add with probability 1
  - Can get from σ to the empty config and from there to any state

- There is some λ_c such that if $\lambda < \lambda_c$, a large box is unordered, $Cov(\sigma_u, \sigma_v) \sim 0$ for u, v far. If $\lambda > \lambda_c$, then get order $|Cov(\sigma_u, \sigma_v)| \geq C$, for some constant.

Ising model

- Each atom has a magnetic field. If most atoms are aligned, get a magnet
- Simply to 2 directions $\{1, -1\}$
- If all independent N atoms, get total magnetism=0
- let σ_x =spin of atom x, $M = \Sigma_x \sigma_x \approx N(0, N)$, $|M| \approx \sqrt{N}$
- If a state $\sigma = (\sigma_x)$ has energy $H(\sigma)$ (Hamiltonian), then Boltzmann distribution is $P_\beta = \frac{e^{-\beta H}}{Z_\beta}$
  - $\beta = \frac{1}{T}$ is the inverse temperature, Z_β is the normalizing (partition) function
  - If $\beta < 1$, high temperature, all σ equally likely
  - If $\beta > 1$, low temperature, low energy states more likely
  - Hamiltonian: $H(\sigma) = -\sum_{x \sim y} \sigma_x \sigma_y$.
- A ferromagnet can stay magnetic up to some temperature T_C. Above it, no longer a magnetic
- On d-dimensional grid ($d > 1$), there is a critical β_C such that
  - if $\beta > \beta_C$, then $M = \sum \sigma_x$ has $|M| = cN$
    - c is a function of β, N is the total size
  - If $\beta < \beta_C$, $|M| = \sqrt{N}$
  - In 2D, $\beta_C = \frac{\log(1+\sqrt{2})}{2}$.
- Dynamics (Glauber)
  - Pick uniformly an x, pick new value for σ_x. Let σ^+, σ^- be σ_x changed to 1 or -1, make $\sigma_x = 1$ with $P = \frac{e^{-\beta H(\sigma^+)}}{e^{-\beta H(\sigma^+)} + e^{-\beta H(\sigma^-)}}$. (i.e. pick σ_x by its distribution conditioned on all other spins). Otherwise, keep $\sigma_x = -1$.
  - If $\beta > \beta_C$, then mixed after $O(N \log N)$ steps
  - If $\beta < \beta_C$, then mixed after $O(e^{CN})$ steps
  - If $\beta < \beta_C$ with boundary all 1, then mixed after $O(N^C)$ steps