Permutations 排列

Involution 退化，内卷

- A permutation is an involution if all cycles have length one or two.
- A permutation is "fixed point free" if there are no cycles of length one.

Counting tricks

This is a useful trick: break counting problem into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.

Binominal coefficients

$\binom{n}{k} := \frac{n!}{k!(n-k)!}$ choose (unordered) k elements from a list of n without repeats

$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ Pascal's triangle

Partition problems $\binom{n}{n_1, n_2, \ldots, n_r} := \frac{n!}{n_1! n_2! \ldots n_r!}$

$$(x_1+x_2+\ldots+x_r)^n = \sum_{n_1,\ldots,n_r : n_1+\ldots+n_r=n} \binom{n}{n_1, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

Plus version

Integer partitions

How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?

Answer: $\binom{n+k-1}{n}$. Represent partition by $k-1$ bars and n stars, e.g., as $**|**||****|*$.

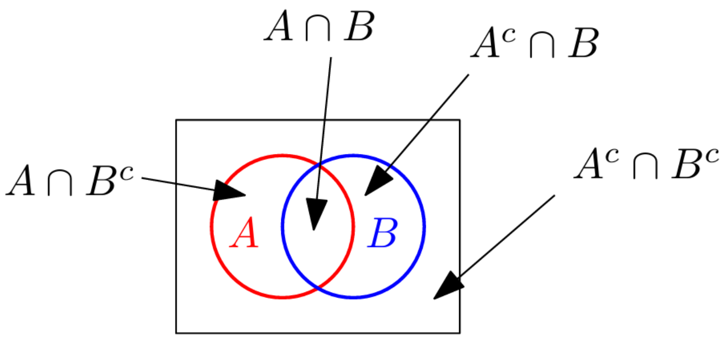
Interesting

Sample space

subset

Intersections, unions, complements

- $A \cup B$ means the union of A and B, the set of elements contained in at least one of A and B.
- $A \cap B$ means the intersection of A and B, the set of elements contained on both A and B.
- A^c means complement of A, set of points in whole sample space S but not in A.
- $A \setminus B$ means "A minus B" which means the set of points in A but not in B. In symbols, $A \setminus B = A \cap (B^c)$.
- $\cup$ is associative. So $(A \cup B) \cup C = A \cup (B \cup C)$ and can be written $A \cup B \cup C$.
- $\cap$ is also associative. So $(A \cap B) \cap C = A \cap (B \cap C)$ and can be written $A \cap B \cap C$.



DeMorgan's laws

$$(\cup_{i=1}^n E_i)^c = \cap_{i=1}^n (E_i)^c \quad (\cap_{i=1}^n E_i)^c = \cup_{i=1}^n (E_i)^c$$

Axioms of probability | Inclusion-exclusion identity

$$
\begin{aligned}
P(\cup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \dots \\
+ (-1)^{(r+1)} \sum_{i_1<i_2<\dots<i_r} P(E_{i_1}E_{i_2}\dots E_{i_r}) \\
+ \dots + (-1)^{n+1} P(E_1 E_2 \dots E_n).
\end{aligned}
$$

The notation $\sum_{i_1<i_2<\dots<i_r}$ means a sum over all of the $\binom{n}{r}$ subsets of size r of the set $\{1, 2, \dots, n\}$.

*Famous hat problem_L4

Conditional probability

$$P(E_1E_2E_3\dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\dots P(E_n|E_1\dots E_{n-1})$$

Baye's theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Independence

Cumulative distribution function

Random variable

Indicator

Say X is a **discrete** random variable if (with probability one) it takes one of a countable set of values.

For each a in this countable set, write $p(a) := P\{X = a\}$. Call p the **probability mass function**.

Gambler's ruin

Expectation of a discrete random variable

$$E[X] = \sum_{x:p(x)>0} xp(x).$$

SUM OVER STATE SPACE $E[X] = \sum_{s \in S} P\{s\}X(s).$

SUM OVER POSSIBLE X VALUES $E[X] = \sum_{x:p(x)>0} xp(x).$ $E[g(X)] = \sum_{x:p(x)>0} g(x)p(x).$

linearity of expectation

Variance $$\mathrm{Var}[X] = \sum_{x:p(x)>0} (x-\mu)^2 p(x).$$

Let X be a random variable with mean μ

$$E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2.$$

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

Y = X + b, Var[Y] = Var[X]

E[Ax] = aE[X], Var[aX] = a^2Var[X]

Bernoulli random variables $1 = 1^n = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$

Number of heads is binomial random variable with parameters (n, p).

$i\binom{n}{i} = n\binom{n-1}{i-1}.$

$$E[X] = \sum_{i=0}^{n} i \binom{n}{i} p^i q^{n-i} = \sum_{i=1}^{n} n \binom{n-1}{i-1} p^i q^{n-i}.$$

$$E[X] = np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{(i-1)} q^{(n-1)-(i-1)}.$$

$$E[X] = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{(n-1)-j} = np(p+q)^{n-1} = np.$$

$E[X^k] = npE[(Y+1)^{k-1}]$

$e = \lim_{n\to\infty}(1 + 1/n)^n$

A **Poisson random variable** X with parameter λ satisfies $P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda}$ for integer $k \geq 0$.

$$E[X] = \sum_{k=0}^{\infty} P\{X = k\}k = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}e^{-\lambda}.$$

Var[X] = λ

N(t): random non-decreasing integer-valued function

1. $N(0) = 0$.
2. **Independence**: Number of events (jumps of N) in disjoint time intervals are independent.
3. **Homogeneity**: Prob. distribution of # events in interval depends only on length. (Deduce: $E[N(h)] = \lambda h$ for some λ.)
4. **Non-concurrence**: $P\{N(h) \geq 2\} << P\{N(h) = 1\}$ when h is small. Precisely:
   - $P\{N(h) = 1\} = \lambda h + o(h)$. (Here $f(h) = o(h)$ means $\lim_{h \to 0} f(h)/h = 0$.)
   - $P\{N(h) \geq 2\} = o(h)$.

- Axioms should imply that $P\{N(t) = k\} = e^{-\lambda t}(\lambda t)^k / k!$.
- One way to prove this: divide time into n intervals of length t/n. In each, probability to see an event is $p = \lambda t/n + o(1/n)$.
- Use binomial theorem to describe probability to see event in exactly k intervals.
- Binomial formula: $\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p^k (1-p)^{n-k}$.
- This is approximately $\frac{(\lambda t)^k}{k!}(1-p)^{n-k} \approx \frac{(\lambda t)^k}{k!} e^{-\lambda t}$.

time till first event

- Can we work out the probability of no events before time t?
- We assumed $P\{N(h) = 1\} = \lambda h + o(h)$ and $P\{N(h) \geq 2\} = o(h)$. Taken together, these imply that $P\{N(h) = 0\} = 1 - \lambda h + o(h)$.
- Fix λ and t. Probability of no events in interval of length t/n is $(1 - \lambda t/n) + o(1/n)$.
- Probability of no events in first n such intervals is about $\left(1 - \lambda t/n + o(1/n)\right)^n \approx e^{-\lambda t}$.
- Taking limit as $n \to \infty$, can show that probability of no event in interval of length t is $e^{-\lambda t}$.
- $P\{N(t) = 0\} = e^{-\lambda t}$.

Geometric random variables

Let X be such that the first heads is on the Xth toss.

For example, if the coin sequence is $T, T, H, T, H, T, \ldots$ then $X = 3$.

Then X is a random variable. What is $P\{X = k\}$?

Answer: $P\{X = k\} = (1-p)^{k-1}p = q^{k-1}p$, where $q = 1 - p$ is tails probability.

Can you prove directly that these probabilities sum to one?

Say X is a **geometric random variable with parameter** p.

Geometric random variable expectation

Note $E[X-1] = \sum_{k=1}^{\infty} q^{k-1}p(k-1)$. Setting $j = k-1$, we have $E[X-1] = q\sum_{j=0}^{\infty} q^{j-1}pj = qE[X]$.

Kind of makes sense. $X-1$ is "number of extra tosses after first." Given first coin heads (probability p), $X-1$ is 0. Given first coin tails (probability q), conditional law of $X-1$ is geometric with parameter p. In latter case, conditional expectation of $X-1$ is same as a priori expectation of X.

Thus $E[X]-1 = E[X-1] = p\cdot 0 + qE[X] = qE[X]$ and solving for $E[X]$ gives $E[X] = 1/(1-q) = 1/p$.

$\mathrm{Var}[X] = q/p^2$

Negative binomial random variable parameter

- Consider an infinite sequence of independent tosses of a coin that comes up heads with probability p.
- Let X be such that the rth heads is on the Xth toss.
- For example, if $r = 3$ and the coin sequence is $T, T, H, H, T, T, H, T, T, \ldots$ then $X = 7$.
- Then X is a random variable. What is $P\{X = k\}$?
- Answer: need exactly $r - 1$ heads among first $k - 1$ tosses and a heads on the kth toss.
- So $P\{X = k\} = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} p$. Can you prove these sum to 1?

Expectation of binomial random variable

- Consider an infinite sequence of independent tosses of a coin that comes up heads with probability p.
- Let X be such that the rth heads is on the Xth toss.
- Then X is a **negative binomial random variable with parameters** (r, p).
- What is $E[X]$?
- Write $X = X_1 + X_2 + \ldots + X_r$ where X_k is number of tosses (following $(k-1)$th head) required to get kth head. Each X_k is geometric with parameter p.
- So $E[X] = E[X_1 + X_2 + \ldots + X_r] = E[X_1] + E[X_2] + \ldots + E[X_r] = r/p$.
- How about $\mathrm{Var}[X]$?
- Turns out that $\mathrm{Var}[X] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \ldots + \mathrm{Var}[X_r]$. So $\mathrm{Var}[X] = rq/p^2$.

continuous random variable and probability density function

Say X is a **continuous random variable** if there exists a **probability density function** $f = f_X$ on $\mathbb{R}$ such that $P\{X \in B\} = \int_B f(x)dx := \int 1_B(x)f(x)dx$.

A famous paradox

- Uniform probability measure on $[0,1)$ should satisfy **translation invariance:** If B and a horizontal translation of B are both subsets $[0,1)$, their probabilities should be equal.
- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- By translation invariance, $\tau_r(B)$ has same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0,1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0,1)$ contain **one** point from each class. For each $x \in [0,1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.
- Then each x in $[0,1)$ lies in $\tau_r(A)$ for **one** rational $r \in [0,1)$.
- Thus $[0,1) = \cup \tau_r(A)$ as r ranges over rationals in $[0,1)$.
- If $P(A) = 0$, then $P(S) = \sum_r P(\tau_r(A)) = 0$. If $P(A) > 0$ then $P(S) = \sum_r P(\tau_r(A)) = \infty$. Contradicts $P(S) = 1$ axiom.

uniform random variable

- Fix $\alpha < \beta$ and suppose X is a random variable with probability density function $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta]. \end{cases}$
- Then for any $\alpha \le a \le b \le \beta$ we have $P\{X \in [a, b]\} = \frac{b-a}{\beta-\alpha}$.
- Intuition: all locations along the interval $[\alpha, \beta]$ are equally likely.
- Say that X is a **uniform random variable on** $[\alpha, \beta]$ or that X is **sampled uniformly from** $[\alpha, \beta]$.

$E[X] = (\alpha+\beta)/2$
$Var(X) = (\beta-\alpha)^2/12$

(Standard) normal random variable

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

E[X] = 0
Var(X) = 1

$$Y = \sigma X + \mu ?$$

- Say Y is normal with parameters μ and σ^2 if $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$.
- What are the mean and variance of Y?
- $E[Y] = E[X] + \mu = \mu$ and $\mathrm{Var}[Y] = \sigma^2 \mathrm{Var}[X] = \sigma^2$.

DeMoivre-Laplace Limit Theorem

--Lecture 19

?

- **DeMoivre-Laplace limit theorem (special case of central limit theorem):**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

Exponential random variables

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

$$P\{X < a\} = 1 - e^{-\lambda a} \text{ and } P\{X > a\} = e^{-\lambda a}.$$

$$E[X^0] = E[1] = 1, \quad E[X] = 1/\lambda, \quad E[X^2] = 2/\lambda^2 \qquad \mathrm{Var}[X] = E[X^2] - (E[X])^2 = 1/\lambda^2.$$

$$E[X^n] = n!/\lambda^n.$$

Minimum of independent exponentials is exponential

- **CLAIM:** If X_1 and X_2 are independent and exponential with parameters λ_1 and λ_2 then $X = \min\{X_1, X_2\}$ is exponential with parameter $\lambda = \lambda_1 + \lambda_2$.

Relationship to Poisson random variables

Memoryless property: If X represents the time until an event occurs, then *given* that we have seen no event up to time b, the conditional distribution of the remaining time till the event is the same as it originally was.

gamma function Γ

$\Gamma(\alpha) := E[X^{\alpha-1}] = \int_0^\infty x^{\alpha-1}e^{-x}dx = (\alpha-1)!.$

Γ distribution

Say that random variable X has gamma distribution with parameters (α, λ) if $f_X(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1}e^{-\lambda x}\lambda}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}.$

Cauchy distribution

- A standard **Cauchy random variable** is a random real number with probability density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
- There is a "spinning flashlight" interpretation. Put a flashlight at $(0,1)$, spin it to a uniformly random angle in $[-\pi/2, \pi/2]$, and consider point X where light beam hits the x-axis.
- $F_X(x) = P\{X \leq x\} = P\{\tan\theta \leq x\} = P\{\theta \leq tan^{-1}x\} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x$.
- Find $f_X(x) = \frac{d}{dx}F(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Beta distribution

- Suppose I have a coin with a heads probability p that I don't know much about.
- What do I mean by not knowing anything? Let's say that I think p is equally likely to be any of the numbers $\{0, .1, .2, .3, .4, \ldots, .9, 1\}$.
- Now imagine a multi-stage experiment where I first choose p and then I toss n coins.
- Given that number h of heads is $a-1$, and $b-1$ tails, what's *conditional* probability p was a certain value x?
- $P\Big(p = x | h = (a-1)\Big) = \frac{\frac{1}{11}\binom{n}{a-1}x^{a-1}(1-x)^{b-1}}{P\{h=(a-1)\}}$ which is $x^{a-1}(1-x)^{b-1}$ times a constant that doesn't depend on x.

Joint probability mass functions

marginal distributions.

- Given random variables X and Y, define $F(a, b) = P\{X \leq a, Y \leq b\}$.
- The region $\{(x, y) : x \leq a, y \leq b\}$ is the lower left "quadrant" centered at (a, b).
- Refer to $F_X(a) = P\{X \leq a\}$ and $F_Y(b) = P\{Y \leq b\}$ as **marginal** cumulative distribution functions.
- Question: if I tell you the two parameter function F, can you use it to determine the marginals F_X and F_Y?
- Answer: Yes. $F_X(a) = \lim_{b\to\infty} F(a, b)$ and $F_Y(b) = \lim_{a\to\infty} F(a, b)$.

Summing two random variables

- Say we have independent random variables X and Y and we know their density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \leq a\}$.
- This is the integral over $\{(x, y) : x + y \leq a\}$ of $f(x, y) = f_X(x)f_Y(y)$. Thus,
- $$P\{X + Y \leq a\} = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy$$
  $$= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy.$$
- Differentiating both sides gives $f_{X+Y}(a) = \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$.

covariance and correlation

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y]). \quad \mathrm{Var}(X) = \mathrm{Cov}(X, X)$$

- **General statement of bilinearity of covariance:**

$$\mathrm{Cov}(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j).$$

- Special case:

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{(i,j):i<j} \mathrm{Cov}(X_i, X_j).$$

Correlation of X and Y defined by

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Conditional expectation

$$E[E[X|Y]] = E[X].$$

Conditional variance

$$\mathrm{Var}(X) = \mathrm{Var}(E[X|Y]) + E[\mathrm{Var}(X|Y)].$$

- Sometimes think of the expectation $E[Y]$ as a "best guess" or "best predictor" of the value of Y.
- It is best in the sense that at among all constants m, the expectation $E[(Y-m)^2]$ is minimized when $m = E[Y]$.
- But what if we allow non-constant predictors? What if the predictor is allowed to depend on the value of a random variable X that we can observe directly?
- Let $g(x)$ be such a function. Then $E[(y - g(X))^2]$ is minimized when $g(X) = E[Y|X]$.

Moment generating functions

$$M(t) = M_X(t) := E[e^{tX}]$$

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.
- If $Z = X + b$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{tX+bt}] = e^{bt}M_X(t)$.
- Latter answer is the special case of $M_Z(t) = M_X(t)M_Y(t)$ where Y is the constant random variable b.

characteristic function

$$\phi(t) = \phi_X(t) := E[e^{itX}].$$

- Let X be a random variable and X_n a sequence of random variables.
- We say that X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- **Lévy's continuity theorem (see Wikipedia):** if $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all t, then X_n converge in law to X.
- **Moment generating analog:** if moment generating functions $M_{X_n}(t)$ are defined for all t and n and $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all t, then X_n converge in law to X.

Lecture27

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

Central limit theorem $\lim_{n\to\infty} P\{a \leq B_n \leq b\} \to \Phi(b) - \Phi(a).$ $B_n = \frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$

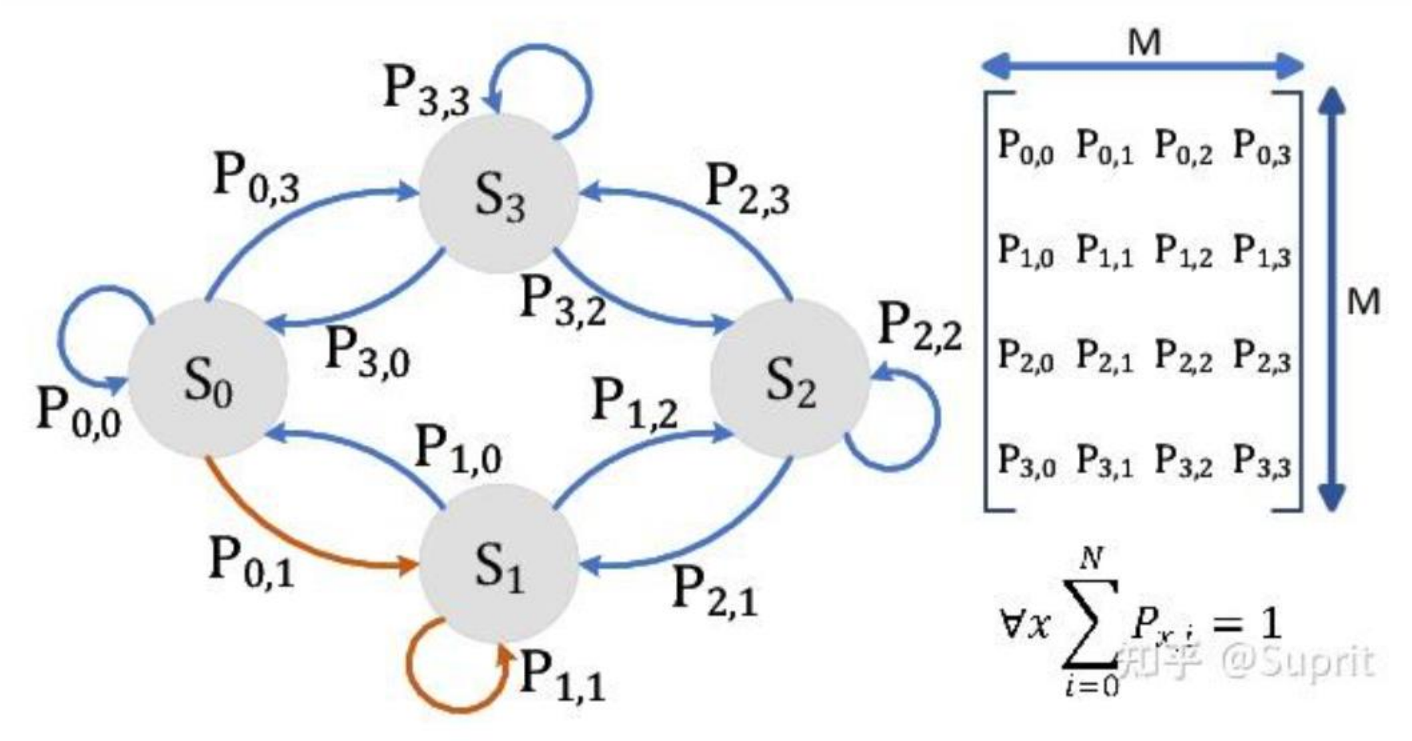
- Recall: **weak law of large numbers** states that for all $\epsilon > 0$ we have $\lim_{n\to\infty} P\{|A_n - \mu| > \epsilon\} = 0$.
- The **strong law of large numbers** states that with probability one $\lim_{n\to\infty} A_n = \mu$.

Jensens inequality statement

- Let X be random variable with finite mean $E[X] = \mu$.
- Let g be a **convex** function. This means that if you draw a straight line connecting two points on the graph of g, then the graph of g lies below that line. If g is twice differentiable, then convexity is equivalent to the statement that $g''(x) \geq 0$ for all x. For a concrete example, take $g(x) = x^2$.
- **Jensen's inequality:** $E[g(X)] \geq g(E[X])$.
- Similarly, if g is **concave** (which means $-g$ is convex), then $E[g(X)] \leq g(E[X])$.

Markov chain

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij}.$$



Entropy

- If a random variable X takes values $x_1, x_2, \ldots, x_n$ with positive probabilities $p_1, p_2, \ldots, p_n$ then we define the **entropy** of X by

$$H(X) = \sum_{i=1}^{n} p_i(-\log p_i) = -\sum_{i=1}^{n} p_i \log p_i.$$

Noiseless coding theorem

$$\sum_{i=1}^{N} n_i p(x_i) \geq H(X) = -\sum_{i=1}^{N} p(x_i) \log p(x_i).$$

conditional entropy

$$H(X, Y) = H(Y) + H_Y(X). \qquad H_Y(X) \leq H(X)$$

Martingale definition

Say X_n sequence is a **martingale** if $E[|X_n|] < \infty$ for all n and $E[X_{n+1}|X_0, X_1, X_2, \ldots, X_n] = X_n$ for all n.

Stopping time definition

- Let T be a non-negative integer valued random variable.
- Think of T as giving the time the asset will be sold if the price sequence is $X_0, X_1, X_2, \ldots$.
- Say that T is a **stopping time** if the event that $T = n$ depends only on the values X_i for $i \leq n$. In other words, the decision to sell at time n depends only on prices up to time n, not on (as yet unknown) future prices.

Doobs Optional Stopping Theorem

- **Doob's Optional Stopping Theorem:** If the sequence $X_0, X_1, X_2, \ldots$ is a **bounded** martingale, and T is a stopping time, then the expected value of X_T is X_0.

Risk neutral probability of outcomes known at fixed time T

$$\frac{\text{Price}\{\text{Contract paying 1 dollar at time } T \text{ if } A \text{ occurs }\}}{\text{Price}\{\text{Contract paying 1 dollar at time } T \text{ no matter what }\}}.$$

布朗运动、伊藤引理、BS 公式（前篇） - 知乎
期权定价的魔法：Black-Scholes公式（一）_black scholes推导-CSDN博客
什么是Black-Scholes期权定价模型？ - 知乎
Etc…