MAT377 Lecture Notes

Arky!! :3c

'24 Fall Semester

Contents

§1 Day 1: Introduction to Probability (Sep. 4, 2024)

Link to textbook.

We start with a sampler problem that on the surface, seems unrelated to probability. Let $v_1, \dots, v_n \in \mathbb{R}^n$ be unit vectors on the unit sphere, i.e.t $\|v_i\| = 1$. If we are to pick $\varepsilon_i = \{-1, 1\}$ at random, what is our expectation on how large will

$$\sum_{i=1}^{n} \varepsilon_i v_i$$

be? We could brute force and average out over all probabilities as follows,

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left(\sum_{i=1}^{n} \varepsilon_i v_i \right) = \sum_{i=1}^{n} \left(\underbrace{\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \varepsilon_i}_{=0} \right) v_i = 0.$$

Now, consider that

$$\begin{aligned} \frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left| \sum_{i=1}^{n} \varepsilon_i v_i \right|^2 &= \frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \sum_{i_1, i_2 = 1}^{n} \varepsilon_{i_1} \varepsilon_{i_2} \langle v_{i_1}, v_{i_2} \rangle \\ &= \sum_{i_1, i_2 = 1}^{n} \left(\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \varepsilon_{i_1} \varepsilon_{i_2} \right) \langle v_{i_1}, v_{i_2} \rangle \end{aligned}$$

To simplify the bracketed summation, we could consider the following two cases:

- If $i_1 \neq i_2$, we would have that

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \varepsilon_{i_1} \varepsilon_{i_2} = \frac{2^{n-2}}{2^n} \sum_{\substack{\varepsilon_{i_1} \in \{-1,1\} \\ \varepsilon_{i_2} \in \{-1,1\}}} \varepsilon_{i_1} \varepsilon_{i_2} = 0.$$

- If $i_1 = i_2$, we would have

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \varepsilon_{i_1} \varepsilon_{i_2} = \frac{2^{n-1}}{2^n} \sum_{\varepsilon_i \in \{-1,1\}} \varepsilon_{i_1} \varepsilon_{i_2} = 1.$$

By linearity of expectation, we obtain

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left| \sum_{i=1}^{n} \varepsilon_i v_i \right|^2 = n,$$

and

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left| \sum_{i=1}^{n} \varepsilon_i v_i \right| = c\sqrt{n},$$

where c is a positive real constant.

We now abstract a few probability terms;

- Ω is a sample space, i.e. the set of possible outcomes.
- Let P denote probability, i.e. a mapping of subsets of Ω to $[0,1]$ (read: probability of getting these subsets of Ω); the probability of an event ε out of S occuring is given by $P(\varepsilon \in S, S \subset \{-1,1\}^n) = \frac{1}{|S|}$, assuming that each event in S is equally likely. With this, we have three important properties of P to define:
  1. $P(\Omega) = 1$; the chance of an event in the probability space happening is 1.
  2. Let $\mathcal{F}$ be a collection of subsets $A_1, \dots, A_n$. Then
  $$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i),$$
  given that $A_i \cap A_j = \emptyset$ for all $1 \leq i, j \leq n$. This is linearity of expectation.
  3. $P(A^C) = 1 - P(A)$, which is a property of set complement.
- When our collection $\mathcal{F}$ of subsets of Ω satisfy the following properties, we call it a σ-algebra:
  1. $\emptyset \in \mathcal{F}$,
  2. Closed under countable union: $A_n \in \mathcal{F} \implies \bigcup_{i=1}^{n} A_i \in \mathcal{F}$,
  3. Closed under complement: $A_i \in \mathcal{F} \implies A_i^C \in \mathcal{F}$.

  In a finite sample space, the power set $\mathcal{F} = \mathcal{P}(\Omega)$ is one such example of a σ-algebra. As an example, let $\Omega = [0,1)$; then $P([a,b]) = b - a$ (wlog, let $a < b$). Now, let $\mathcal{F} = \mathcal{P}([0,1))$. Define the equivalence $x \sim y$ if $x - y \in \mathbb{Q}$.
- We now introduce the axiom of choice; Let A be a set containing one element of each equivalence class from the above defined equivalence. Consider $\tau_q A := \{A\} + q$; let us claim that
  $$\bigcup_{q \in \mathbb{Q}} \tau_q A = [0,1),$$
  which is a countable union of $[0,1)$, since $\mathbb{Q}$ is countable. We have that $P(A) = P(\tau_q A)$ because intervals don't change size under shifting by q. However, observe that
  $$P([0,1)) = P\left(\bigcup_{q \in \mathbb{Q}} \tau_q A\right) = \sum_{q \in \mathbb{Q}} P(\tau_q A).$$
  Then either
  $$\begin{aligned} P(A) = 0 &\implies P(\tau_q A) = 0 \implies P([0,1)) = 0, \text{or} \\ P(A) \neq 0 &\implies P(\tau_q A) \to \infty \implies P([0,1)) \to \infty, \end{aligned}$$
  which doesn't make sense (for now). This shows that we need to pick our σ-algebra properly; observing that the intersection of two σ-algebras is also a σ-algebra, it is appropriate to let $\mathcal{F}$ be the smallest σ-algebra containing $[a,b)$. This is called a *Borel Set.*[1]

[1] this is confusing. ill check later

§2 Day 2: Expectations and Distributions (Sep. 9, 2024)

Course administrative details first; starting next week, office hours will be held on Monday from 11:15am to 12:15pm. Recap of last lecture:

- A probability space Ω is the set of all possible outcomes of an "experiment," i.e. a countable set of individual events $\{\omega_1, \dots, \omega_n\}$ (we will cover continuous probability later on).
- $\mathcal{F} = \mathcal{P}(\Omega)$ is the set of all subsets of Ω.
- $P(A) = \sum_{\omega \in \Omega} P(\omega)$ is the probability of an outcome in $A \in \mathcal{F}$ occurring.

A random variable X is a function $\Omega \to \mathbb{R}$, aka the measurement of the event, and the expectation of the random variable, EX, is given by $\sum_{\omega \in \Omega} X(\omega)P(\omega)$. Is expectation well behaved? No. For example, consider the St. Petersburg Paradox; suppose you are playing a game in the casino; every time you flip a coin, your prize money doubles if it lands on heads (read: double or nothing lfg!!!). Then we may consider the set of outcomes to be the number of consecutive heads, i.e.

$$\begin{aligned} \Omega &= \{1, 2, 3, \dots\}, \\ P(n) &= \frac{1}{2^n}, \\ X(n) &= 2^n. \end{aligned}$$

Clearly, the chance of getting n heads in a row is 2^{-n}, and assuming your prize money started at 1 dollar, you would win 2^n dollars for said n heads. Taking the expectation of this game, we find

$$EX = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum 1 = \infty.$$

It doesn't make sense to expect to win infinite amounts of money from this game[2] unless you had unlimited wealth to start with. With this in mind, we insist on

$$\sum_{\omega \in \Omega} |X(\omega)| \, P(\omega) < \infty$$

within the context of this class.

Theorem 2.1 (Linearity of Expectation)**.** X is linear; i.e., $E[ax + by] = aEx + bEy$.

We start with a lemma:

Lemma 2.2. Let us have a bijective map $\pi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Then

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} C_{\pi(n,m)}$$

if all $c_n \geq 0$ or if either side is absolutely convergent.

[2] martingale strat lfg.,..

We proceed to prove this with casework.

- Suppose $c_n \geq 0$; then using the bijective nature of π, we may choose large enough N, M such that
$$\sum_{n=1}^{K} c_n \leq \sum_{n=1}^{N} \sum_{m=1}^{M} c_{\pi(n,m)}$$
for any choice of K. Conversely, we may pick
$$\sum_{n=1}^{N} \sum_{m=1}^{M} c_{\pi(n,m)} \leq \sum_{n=1}^{K} c_n$$
for any N, M by picking $k \geq \max_{\substack{1\leq n\leq N \\ 1\leq m\leq M}} \{\pi(n,m)\}$. Now, let $M \to \infty$; we have
$$\sum_{n=1}^{N} \sum_{m=1}^{\infty} c_{\pi(n,m)} \leq \sum_{n=1}^{\infty} c_n,$$
then let $N \to \infty$ to get
$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_{\pi(n,m)} \leq \sum_{n=1}^{\infty} c_n.$$
As per earlier, we also see that LHS is greater or equal to RHS, which implies equality. $\square$

- Now, suppose $\sum_{n=1}^{\infty} |c_n| < \infty$. Let $c_n = a_n - b_n$, where $a_n = c_n 1(c_n \geq 0)$ and $b_n = c_n 1(c_n < 0)$. Then we obtain
$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_\pi(n,m), \qquad \sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} b_\pi(n,m)$$
as per our proof above. Summing both, we conclude that equality holds for absolute convergence as well. $\square$

For now, let X take values $\{a_1, a_2, \dots\}$ (countably many). Consider
$$P'(a_n) = P(X = a_n) = P(\underbrace{\{\omega \mid X(\omega) = a_n\}}_{X^{-1}(a_n)})$$
as the probably of a pre-image (or, $P' = P \circ X$). We see that P' is a probability on $\mathbb{R}$ (concentrated on $\{a_1, a_2, \dots\}$), and $0 \leq P'(a_n) \leq 1$ for any n; from now, we will call P' a *distribution* of X. Here are some examples of distributions:

- The Bernoulli distribution: let $0 \leq p \leq 1$. Then consider a coin with p chance to land on heads, and $1 - p$ on tails; then Ber_p is given by $\Omega = \{H, T\}$,
$$X(H) = 1, P(X = 1) = p,$$
$$X(T) = 0, P(X = 0) = 1 - p.$$

- Flip N coins, with $X = \{0, 1, \dots, N\}$ being the number of heads we obtain. Then
$$P(X = \ell) = \binom{N}{\ell} p^\ell (1-p)^{N-\ell},$$

and the expected value is given by

$$EX = \sum_{\ell=0}^{N} \ell \binom{N}{\ell} p^{\ell}(1-p)^{N-\ell}.$$

Using linearity of expectation, we see $EX = EX_1 + \cdots + EX_n = Np$ by separating each coinflip.

Expectation enjoys the change of variables property;[3]

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{n=1}^{\infty} a_n P'(a_n).$$

To see this, consider partioning the probability space Ω into $X^{-1} = \{\omega_{nm} \mid 1 \leq m \leq M_n\}$ in terms of their measurement from X (where $X(\omega_{ni}) = X(\omega_{nj}) = a_n$ for any $1 \leq i, j \leq M_n$)[4], and write

$$\sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{n=1}^{\infty} \sum_{m=1}^{M_n} X(\omega_{nm})P(\omega_{nm}), \qquad \text{(by Lemma)}$$

where we may note that mapping each individual $\omega \in \Omega$ to some index nm is bijective since it is a partition. We continue by writing

$$\begin{aligned} &= \sum_{n=1}^{\infty} \sum_{m=1}^{M_n} a_n P(\omega_{nm}) = \sum_{n=1}^{\infty} a_n \left(\sum_{m=1}^{M_n} P(\omega_{nm}) \right) \\ &= \sum_{n=1}^{\infty} a_n P(X = a_n), \end{aligned}$$

where we may note $P(X = a_n) = P'(a_n)$. □

The probability distribution of any given random variable X also approaches 0 at its tail. Specifically, we have that $\lim_{t \to \infty} P(x \geq t) = 0$. To prove this, we start by observing that $P(x \geq t)$ is monotone decreasing; consider

$$P(X \geq n) = \sum_{m=n}^{\infty} P(m \leq X < m+1).$$

Clearly, the sum is convergent, as the sum of probabilities is equal to 1. Using the fact that the tail of a convergent series approaches 0, we conclude that $P(m \leq X < m+1) \to 0$ as $m \to \infty$, and so $P(X \geq n) \to 0$ as $n \to \infty$.

[3] read: sum of value of outcome multiplied by the chance it occurs over all ω is the same as going over each value individually and multiplying the chance you roll into it

[4] read M_n as a counter of how many outcomes in Ω have the same measurement of a_n

Lemma 2.3 (Expectation of Random Variable in terms of Integral)**.** The expectation of a random variable X may be expressed as $EX = \int_0^\infty P(X \geq t)\, dt$ for $X \geq 0$.[5]

Let us start by considering the case where X takes integer values only;

$$\begin{aligned} EX = \sum_{n=1}^{\infty} nP(X = n) &= \sum_{n=1}^{\infty}\sum_{m=1}^{n} P(X = n) \\ &= \sum_{m=1}^{\infty}\sum_{n=m}^{\infty} P(X = n) \\ &= \sum_{m=1}^{\infty} P(X \geq m). \end{aligned}$$

For the general case, let us start by writing $a_n = \int_0^\infty 1(t \leq a_n)\, dt$ by the layer cake decomposition. Then

$$\begin{aligned} EX = \sum_{n=1}^{\infty} a_n P(X = a_n) &= \sum_{n=1}^{\infty}\left(\int_0^\infty 1(t \leq a_n)\, dt\right) P(X = a_n) \\ &\overset{(*)}{=} \int_0^\infty \left(\sum_{n=1}^{\infty} 1(t \leq a_n) P(X = a_n)\right) dt && \text{(Fubini)} \\ &= \int_0^\infty \sum_{a_n \geq t} P(X = a_n)\, dt \\ &= \int_0^\infty P(X \geq t)\, dt \end{aligned}$$

To resolve $(*)$ without the use of Fubini's theorem, we may write

$$\begin{aligned} \sum_{n=1}^{\infty}\int_0^\infty 1(t \leq a_n) P(X < a_n)\, dt &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\int_{m-1}^{m} 1(t \leq a_n) P(X = a_n)\, dt \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\int_{m-1}^{m} 1(t \leq a_n) P(X \geq a_n)\, dt \\ &= \sum_{m=1}^{\infty}\lim_{N\to\infty}\sum_{n=1}^{N}\int_{m-1}^{m} 1(t \leq a_n) P(X \geq a_n)\, dt \\ &= \sum_{m=1}^{\infty}\lim_{N\to\infty}\int_{m-1}^{m}\left(\sum_{n=1}^{N} 1(t \leq a_n) P(X \geq a_n)\right) dt \\ &= \sum_{m=1}^{\infty}\int_{m-1}^{m}\left(\lim_{N\to\infty}\sum_{n=1}^{N} 1(t \leq a_n) P(X \geq a_n)\right) dt \\ &= \int_0^\infty \left(\sum_{n=1}^{\infty} 1(t \leq a_n) P(X = a_n)\right) dt, \end{aligned}$$

which we conclude by removing the auxiliary summations, since it is enough to know that $\sum_{n=1}^{N} 1(t \leq a_n) P(X = a_n) \to \sum_{n=1}^{\infty} 1(t \leq a_n) P(X = a_n)$ uniformly in $t \in [m-1, m]$; i.e.,

$$\left|\sum_{n=N+1}^{\infty} 1(t \leq a_n) P(X = a_n)\right| \leq \sum_{n=N+1}^{\infty} P(X = a_n) \to 0$$

as $N \to \infty$ as per earlier (since the tail goes to 0).

[5] intuition: layer cake formula, but compile them together in level sets.

We also briefly went over examples multinomial distributions at the end of class;

- Suppose $X_1, \ldots, X_n$ are independent, and let $P(X_i = j) = p_j$ for $j = 1, \ldots, k$. Let $\Omega = \{n_1, \ldots, n_k), n_j \geq 0, n_1 + \cdots + n_k = n\}$ (read: k-sided dice rolled n times, where n_j denotes the number of times j came up). Then

$$P((n_1, \ldots, n_k)) = \binom{n}{n_1, \ldots, n_k} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}.$$

- The geometric distribution; let $0 < p < 1$ denote the probability of getting a head, and let us toss a coin until we get a heads. Let the outcome of X denote the number of tosses it took. Then

$$P(X = n) = (1-p)^{n-1} p,$$

and we may check $\sum_{n=1}^{\infty} P(x = n) = 1$ by geometric series.

- The Poisson distribution; let $\lambda > 0$. Then

$$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

for $n = 0, 1, 2, \ldots$.

§3 Day 3: Distributions, Stability Property, Moments (Sep. 11, 2024)

Recall the Poisson distribution

$$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

where $n = 0, 1, \dots$ and $\lambda > 0$. We introduce the stability property: let us consider the independent random variables X_1, X_2 distributed as follows,

$$\begin{aligned} X_1 &\sim \text{Poiss}_{\lambda_1}, \\ X_2 &\sim \text{Poiss}_{\lambda_2}. \end{aligned}$$

Then we have that $X_1 + X_2 \sim \text{Poiss}_{\lambda_1+\lambda_2}$. To prove this, write

$$\begin{aligned} P(X_1 + X_2 = n) &= \sum_{m=0}^{n} P(X_1 = m, X_2 = n - m) \\ &= \sum_{m=0}^{n} P(X_1 = m) P(X_2 = n - m) \\ &= \sum_{m=0}^{n} \frac{\lambda_1^m}{m!} e^{-\lambda_1} \frac{\lambda_2^{m-2}}{(m-n)!} e^{-\lambda_2} \\ &= \frac{1}{n!} \underbrace{\sum_{m=0}^{n} \frac{n!}{m!(n-m)!} \lambda_1^m \lambda_2^{n-m}}_{(\lambda_1+\lambda_2)^n} e^{-(\lambda_1+\lambda_2)} \\ &= \frac{1}{n!} (\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)}. \end{aligned}$$

Binomials also have a related property; let

$$\begin{aligned} X_1 &\sim \text{Bin}(n_1, p), \\ X_2 &\sim \text{Bin}(n_2, p). \end{aligned}$$

Given that X_1, X_2 are independent, we know that $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$. To prove this, we may just write

$$\begin{aligned} X_1 &= y_1 + \dots + y_{n_1}, \\ X_2 &= y_{n_1+1} + \dots + y_{n_1+n_2}, \\ X_1 + X_2 &= y_1 + \dots + y_{n_1+n_2} \sim \text{Bin}(n_1 + n_2, p). \end{aligned}$$

Moreover, we also have $\text{Bin}(n, \frac{\lambda}{n}) \overset{n\to\infty}{\to} \text{Poiss}_\lambda$. This is called the *law of little numbers*. To prove this, we have

$$\begin{aligned} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \frac{\lambda^k}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n-k}}_{\to e^{-\lambda}} \\ &\overset{n\to\infty}{\to} \frac{\lambda^k}{k!} e^{-\lambda} = \text{Poiss}_\lambda. \end{aligned}$$

There are two examples of Poisson distributions that we will go over: shark attacks and radioactive decay. (but we didn't go over it ig?)

Theorem 3.1 (Doeblin)**.** Let X_i be independent random variables distributed by Ber_{p_i}, where $0 < p_i < 1$. Let us have $S_n = X_1 + \cdots + X_n$ with $\lambda = p_1 + \cdots + p_n$. Then

$$\left| P(S_n \in A) - \sum_{n \in A} \frac{\lambda^n}{n!} e^{-\lambda} \right| \leq \sum_{i=1}^{n} p_i^2,$$

where $A \subset \{0, 1, \dots\}$.

To prove this, let y be a random variable where $y \sim \text{Poiss}_p$, then $P(y = 0) = e^{-p} > 1 - p$. Define $\Omega^{\perp} = \{-1, 0, 1, 2, \dots\}$; then we have $P_p(-1) = 1 - p, P_p(0) = e^{-p} - 1 + p, P_p(k) = \frac{p^k}{k!} e^{-p}$ for $k = 1, 2, 3, \dots$. Moreover, define

$$X(\omega) = \begin{cases} 0 & \omega = -1 \\ 1 & \omega \geq 0 \end{cases}, \qquad y(\omega) = \begin{cases} 0 & \omega = 0, 1 \\ \omega & \omega \geq 1 \end{cases}.$$

Then $P(x = y) = 1 - p + pe^{-p} \geq 1 - p + p(1 - p) = 1 - p^2$, so $P(x = y) \leq p^2$. Take $\Omega = (\Omega_+)^n$ and $X_i(\omega) = X(\omega)$, and let us have

$$P(\omega) = \prod_{i=1}^{n} P_{p_i}(\omega_i)$$

where X_i are independently distributed by Ber_p, and y_i are independently distributed by Poiss_{p_i}. Finally, let us have

$$\begin{aligned} S_n &= X_1 + \cdots + X_n, \\ S'_n &= y_1 + \cdots + y_n, \end{aligned}$$

then $P(S_n \neq S'_n) \leq \sum_{i=1}^{n} P(X_i \neq y_i) \leq \sum_{i=1}^{n} p_i^2$, which means $S'_n \sim \text{Poiss}_\lambda$.[6] □

We now define *moments*. For a random variable X, $E(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega)$, as long as $E|X| < \infty$. Then moments are given by EX^n where $n = 1, 2, \dots$. (??) For example, let $X = \text{Poiss}_\lambda$. Then we have for $n = 1$,

$$\begin{aligned} EX &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n}{(n-1)!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} \frac{\lambda^{n+1}}{n!} e^{-\lambda} \\ &= \lambda. \end{aligned}$$

For $n = 2$, we have

$$\begin{aligned} EX^2 &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} \\ &= E(X(X-1)) + EX \\ &= \lambda^2 + \lambda. \end{aligned}$$

[6]reminder: review this proof, i'm stupid and don't really get it.

§4 Day 4: Independence and Dependence (Sep. 16, 2024)

Let (Ω, P) be our probability space. We define

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

to be the conditional probability, i.e. probability of A given B, as long as $P(B) > 0$ (this is called Bayes' Rule). If $P(A \mid B) = P(A)$, then A is said to be independent of B. In particular, if $A_1, \dots, A_n$ are independent, then

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

If the above is true for only pairs of events A_i, A_j, then we say that they are pairwise independent.

Let Ω_i, P_i be probability spaces, and consider $\Omega = \Omega_1 \times \cdots \times \Omega_n = \prod_{i=1}^{n} \Omega_i$, where we define a probability event in $\omega \in \Omega$ to be $(\omega_1, \dots, \omega_n) = \omega$ with $\omega_i \in \Omega_i$. Specifically, we have

$$P(\omega) := \prod_{i=1}^{n} P_i(\omega_i).$$

For example, let $A = A_1 \times \cdots \times A_n$, and $A_i \in \Omega_i$. then

$$P(A) := \sum_{\omega \in A} P(\omega) = \sum_{\substack{\omega_i \in A_i \\ i=1,\dots,n}} \prod_{i=1}^{n} P_i(\omega_i) = \prod_{i=1}^{n} \sum_{\omega_i \in A_i} P_i(\omega_i) = \prod_{i=1}^{n} P_i(A_i).$$

Let us have random variables $X_i : \Omega_i \to \mathbb{R}$ where $1 \le i \le n$. Then $X_i^{(\omega)} = f_i(\omega_i)$ are independent if $P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$. In other words, $X_i^{-1}(A_i)$ are independent.[7] We may continue simplifying the expression as follows,

$$\prod_{i=1}^{n} P(X_i \in A_i) = \prod_{i=1}^{n} P_i(f_i(\omega_i) = x_i) = \prod_{i=1}^{n} P(X_i = x_i).$$

Now, suppose X, Y are independent and f, g are functions. Then we claim that $f(X), g(Y)$ are independent. To check this, let us write

$$\begin{aligned} P(f(X) = a, g(Y) = b) &= P(X = f^{-1}(a), y = g^{-1}(b)) \\ &= P(X \in f^{-1}(a))P(Y \in g^{-1}(b)) \\ &= P(f(X) = a)P(g(Y) = b). \end{aligned}$$

We can also do this with grouping; let $\{1, \dots, n\} = U_{k=1}^{m} I_k$ with I_k disjoint; i.e., we're sorting $[n]$ into disjoint subsets I_k. Then let $y_k = f_k(\{x_i\}_{i \in I_k})$ for some function $f_k : \mathbb{R}^{|I_k| \to \mathbb{R}}$, and we have that y_k are independent. To prove this, observe that

$$\begin{aligned} P(y_1 \in A_1, \dots, y_m \in A_m) &= P(f_1 \in A_1, \dots, f_m \in A_m) \\ &= P(\{X_i\}_{i \in I_1} \in f_1^{-1}(A_1), \dots, \{X_i\}_{i \in I_m} \in f_m^{-1}(A_m)) \\ &= \prod_{j=1}^{m} P(\{X_i\}_{i \in I_j} \in f_j^{-1}(A_j)) \\ &= \prod_{j=1}^{m} P(y_j \in A_j). \end{aligned}$$

[7] note on board: for any $A_1, \dots, A_n$ borel sets, intervals are enough, like $(-\infty, x_i]$. confusion?

We need to show that $P(\{X_i\}_{i\in I_1} = b_1, \{X_i\}_{i\in I_2} = b_2) = P(\{X_i\}_{i\in I_1} = b_1)P(\{X_i\}_{i\in I_2} = b_2)$; but as per earlier, this is true.

Now, suppose our random variables X_is are independently binomial distributed. Then

$$X_1 + \cdots + X_{m_1} \sim \mathrm{Bin}(m_1, p) \sim \mathrm{Poiss}_{\lambda_1}, \qquad (\lambda_1 = pm_1)$$

$$X_{m_1+1} + \cdots + X_{m_1+m_2} \sim \mathrm{Bin}_{m_2,p} \sim \mathrm{Poiss}_{\lambda_2} \qquad (\lambda_2 = pm_2)$$

We may combine the groupings above to get $X_1 + \ldots X_{m_1+m_2} \sim \mathrm{Bin}(m_1 + m_2, p) \sim \mathrm{Poiss}_{\lambda_1+\lambda_2}$.

Lemma 4.1. If X and Y are indepnendent and $E[|X|] < \infty$, $E[|Y|] < \infty$, then $E[XY] = E[X]E[Y]$.[8]

First, assume $X, Y > 0$. Let us directly write

$$\begin{aligned} E[XY] &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)P(\omega) \\ &= \sum_{n,m} a_n b_m P(X = a_n, Y = b_m) \\ &= \sum_{n,m} a_n b_m P(X = a_n)P(Y = b_m) \\ &= \sum_{n} a_n P(X = a_n) \sum_{m} b_m P(Y = b_m) \\ &= E[X]E[Y]. \end{aligned}$$

In the case that the random variables are not necessarily non-negative, we may simple consider

$$\begin{aligned} X &= X1(X \geq 0) - |X|\,1(X < 0) = X_+ - X_-, \\ Y &= Y1(Y \geq 0) - |Y|\,1(Y < 0) = Y_+ - Y_-. \end{aligned}$$

However, do note that in the OPPOSITE direction that $E[XY] = E[X]E[Y]$ does NOT imply that X, Y are independent. It is true that $E[f(X)g(Y)] = E[f(X)]E[g(y)]$ for "lots of" f, g would imply that X, Y independent (if this is true for *all* f, g, then it is independent), but this is unreliable.

Using Fubini's theorem, we may consider X, Y on non-discrete probability spaces, and write

$$\begin{aligned} E[f(X,Y)] &= \sum_{n,m} f(a_n, b_m)P(X = a_n, Y = b_m) \\ &\overset{\text{if indep.}}{=} f(a_n, b_m)P(X = a_n)P(Y = b_m) \\ &\overset{\text{if ``nice''}}{=} \sum_n \left[\sum_m f(a_n, b_m)P(Y = b_m)\right] P(X = a_n). \end{aligned}$$

We say that the above is "nice" if $f \geq 0$, or $E[f(x,y)] < \infty$, or

$$\sum_n \left[\sum_m |f(a_n, b_m)|\, P(Y = b_m)\right] P(X = a_n) < \infty.$$

[8] i'm sick of the no bracket nonsense

Alternatively, if we dont have our "nice" cases, we have

$$\sum_n \left[\sum_m f(a_n, b_m) P(Y = b_m \mid X = a_n)\right] P(X = a_n).$$

Now, we inntroduce the conditional distribution $P(y = b_m \mid X = a_n)$, where the distribution is Y given $X = a_n$. We can write the expectation

$$E[g(Y) \mid X = a_n] = \sum_m g(b_m) P(y = b_m \mid x = a_n),$$

i.e. the conditional expectation of $g(Y)$ given $X = a_n$. For example, let $X_1, X_2, \ldots$ be i.i.d. Ber_p and N indep. Poiss_λ. Then $Y = X_1 + \cdots + X_N$ has

$$P(Y = k) = \sum_{n=0}^{\infty} P(Y = k, N = n) = \sum_{n=0}^{\infty} P(y = k \mid N = k) P(N = n).$$

If $N = n$, then $Y = X_1 + \cdots + X_n$, and we have

$$P(Y = k \mid N = n) = P(X_1 + \cdots + X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Specifically,

$$\begin{aligned} P(y) &= \sum_{n=0}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \frac{(\lambda k)^k}{k!} \left(\sum_{n=0}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} \lambda^{n-k}\right) e^{-\lambda} \\ &= \frac{(\lambda p)^k}{k!} e^{-p\lambda} \sim \mathrm{Poiss}_{p\lambda}. \end{aligned}$$

Now for another example; let $X_1, X_2, \ldots$ be i.i.d. Ber_{y_2}; i.e. let $x_i \in \{0,1\}^{\mathbb{N}}$; let $x \in [0,1) = \Omega$, $X = 0, X_1, X_2, \ldots$. Let P be on $[0,1)$. Then $P([a,b)) = b - a$ where $b > a$; we claim that they are i.i.d. $\mathrm{Ber}_{\mathrm{y}_2}$, which is proven by subdividing the intervals (whatever this means).

Let $P(X_1 = x_1, \ldots, X_n = x_n)$. Then this is equal to

$$\begin{aligned} &= P(X_n = x_n \mid x_1 = x_1, \ldots, X_{n-1} = x_{n-1}) P(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) \\ &= \prod_{k=0}^{n-1} P(X_{k+1} = x_{k+1} \mid X_1 = x_1, \ldots, X_k = x_k). \end{aligned}$$

In this specific kind of system where the probability of X_{k+1} only depends on the ones the step right before, we call it a *Markov Chain*, i.e. a probabilisitic version of dynamical systems.

§5 Day 6: (Sep. 23, 2024)

Let us throw n balls into n boxes. Then consider N to be the number of empty boxes, and we have

$$P(N = k) = \frac{1}{n^n}\binom{n}{k}\sum_{\ell=0}^{k}(-1)^\ell\binom{k}{\ell}(k-\ell)^n,$$

with accompanying expectation

$$E[N] = \sum_{k=0}^{n} P(N = k).$$

However, we may simply the expression as follows; let N instead be written as a sum of indicators, i.e.

$$N = \sum_{i=1}^{N} 1(\text{the } i\text{th box is empty}),$$

yielding

$$E[N] = \sum_{i=1}^{N} P(\text{the } i\text{th box is empty}) = nP(\text{the } i\text{th box is empty}) = n\frac{(n-1)^n}{n^n}.$$

Next example; the longest increasing subsequence of a random permutation. Let S_n be the set of bijections $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. Then $|S_n| = n!$ where each σ has probability $\frac{1}{n!}$, and we define an increasing subsequence to be given by $\sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_k)$ for $i_1 < i_2 < \cdots < i_k$. Let $L_n(k)$ be the longest increasing subsequence of σ. Then L_n is a random variable; it remains to ask how big L_n is (Ulam's Problem). It is proven (though not in this class) that

$$\frac{L_n}{\sqrt{n}} \to 2$$

as $n \to \infty$, i.e. the expectation $E[L_n] \sim 2\sqrt{n}$ [9].

We prove a looser bound for now; let N_k be the number of increasing subsequences of length k. Then

$$N_k = \sum_{i_1<\cdots<i_k} 1(\sigma(i_1) < \cdots < \sigma(i_k)),$$

and we may write

$$\begin{aligned} E[N_k] &= \sum_{i_1<\cdots<i_k} P(\sigma(i_1) < \cdots < \sigma(i_k)) \\ &= \binom{n}{k}\frac{1}{k!} = \frac{n!}{(n-k)!(k)!^2}. \end{aligned}$$

We now present Chebyshev's inequality,

$$P(X \geq x) \leq \frac{E[X1(X \geq x)]}{x},$$

where $x > 0$. Using $X1(X \geq x) \geq x1(X \geq x)$, we have $E[X1(X \geq x)] \geq xP(X \geq x)$. Returning to earlier, we obtain $P(N_k > 0) \leq E[N_k] = \frac{n!}{(n-k)!(k!)^2}$.

[9] for more, see here :3

Theorem 5.1 (Stirling's Formula)**.** $n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$.

To prove this, start by considering the Gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^t \, dt$, which gives $\Gamma(n) = (n-1)!$. Let us consider the following;

$$\begin{aligned}
\Gamma(n+1) &= \int_0^\infty t^n e^{-t} \, dt \\
&= \int_0^\infty e^{n \log t - t} \, dt && \text{(Substitute } t = nx\text{)} \\
&= n e^{n \log n} \underbrace{\int_0^\infty e^{n(\log x - x)} \, dx}_{\text{of the form } \int e^{nf(x)} \, dx} \\
&\approx \int e^{nf(x^*) + \frac{f''(x^*)}{2}(x-x^*)^2} \, dx \\
&= e^{-n} \int e^{-\frac{n}{2}(x-1)^2} \, dx && \text{(Substitute } x = 1 + \tfrac{y}{\sqrt{n}}\text{)} \\
&= \frac{e^{-n}}{\sqrt{n}} \int_{-\infty}^\infty e^{-\frac{y^2}{2}} \, dy \\
&= \frac{e^{-n}}{\sqrt{n}} \sqrt{\int_{-\infty}^\infty \int_{-\infty}^\infty e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy} \\
&= \frac{e^{-n}}{\sqrt{n}} \sqrt{\int_0^\infty \int_0^{2\pi} e^{-\frac{r^2}{2}} r \, d\theta \, dr} \\
&= \frac{e^{-n}}{\sqrt{n}} \sqrt{2\pi \int_0^\infty r e^{-\frac{r^2}{2}} \, dr} \\
&= \frac{e^{-n}}{\sqrt{n}} \sqrt{2\pi}.
\end{aligned}$$

(I don't know where the hell this went. Oh well.)

We may now insert Stirling's formula into Ulam's problem to obtain

$$P_n(N_k > 0) \leq \frac{n!}{(n-k)!(k!)^2} \approx c \frac{n^{n+\frac{1}{2}} e^{-n}}{k^{2k+1} e^{-2k} (n-k)^{n-k+\frac{1}{2}} e^{-}(n-\frac{1}{2})},$$

which cancels nicely. Using $k! \geq k^k e^{-k}$, we get $P(N_k > 0) \leq (\frac{e\sqrt{n}}{k})^{2k}$. We may make the bound nicer by writing

$$\left(\frac{e\sqrt{n}}{k}\right)^{2k} \leq \left(\frac{e}{3}\right)^{6\sqrt{n}}, \qquad (k = 3\sqrt{n})$$

and using $(\frac{e}{3})^6 \leq e^{-\frac{1}{2}}$ yields that it is less than $e^{-\frac{n}{2}}$. Thus, we have that $P(L > \sqrt[3]{n}) \leq \sum_{m=3\sqrt{n}}^\infty e^{-\frac{\sqrt{m}}{2}}$.

We now cover Erdös-Renyi random graphs. Let us have a graph on vertices $V = \{v_1, \ldots, v_n\}$, and edges $E = \{e_{ij}\} \subset V \times V$. We have that e_{ij} is in E with probability p, and not there with probability $1-p$, considered independently over all undirected pairs (i, j).

Define a clique to be a complete subgraph of any graph $G(n,p)$, and let us have $\omega(G)$ to be the clique number, i.e. the size of the largest clique of G. This is approximately $C_p \log n$, where C_p is some constant. Let us have N_k as the number of cliques of size k. To calculate the expectation, let us have

$$N_k = \sum_{V' \subset V |V'|=k} 1_{\text{all } e_{ij} for all i,j \in V' \text{ exists}}.$$

Then we have

$$E[N_k] = \binom{n}{k} p^{\binom{k}{2}} =: f(k).$$

Observe that we have

$$\frac{f(k+1)}{f(k)} = \frac{\binom{n}{k+1} p^{\binom{k+1}{k}}}{\binom{n}{k} p^{\binom{k}{2}}} = \frac{n-k}{k+1} p^k.$$

Observing that $f(1) = n$ and $f(n) = p^{\frac{n(n-1)}{2}} << 1$, we see that f is unimodal. In particular, there is a unique point k_0 such that $f(k_0) \geq 1 > f(k_0+1)$. Thus,

$$\left(\frac{n}{k} - 1\right)^k p^{\frac{k(k-1)}{2}} \leq f(k) \leq n^k p^{\frac{k(k-1)}{2}}.$$

In particular, the right hand side is less than 1 if $np^{\frac{k-1}{2}} < 1$, and this evaluates out to $k > C_p \log n$. The left hand side is greater than 1 when $k \leq \frac{\log(\frac{n}{k}-1)}{|\log p|} + 1$.