- Office hours: { Chris: Monday 3-4, PAB 214
  Quinn: Monday 5:30-6:30, PAB 214 }

- Email: { coverstr@stanford.edu
  qmac@stanford.edu }

- (Stats 116): Intro to Probability Theory.

- Probability sample:
  1. Sample space Ω, a set representing all possible outcomes.

     E.g. paramagnet with $N=2$: $\Omega = \{\uparrow\uparrow, \uparrow\downarrow, \downarrow\uparrow, \downarrow\downarrow\}$

     Def: A subset $A \subset \Omega$ is an <u>event</u>.

  2. Probability function $f : \Omega \to \mathbb{R}$ s.t.

     i. $f(x) \in [0,1]$ for all $x \in \Omega$.

     ii. $\sum_{x \in \Omega} f(x) = 1$.

- Definition of probability:

  For all $A \subset \Omega$, $P(A) \equiv \sum_{x \in A} f(x)$.

  - $A \cup B$: elements of A or B
  - $A \cap B$: elements of A and B

  $\Rightarrow P(\bigcup_n A_n) = \sum_n P(A_n)$ if $A_1, A_2, \dots A_n$ are disjoint.

- Can now prove some of the rules of probability that you may already know - Problem 1.

- Conditional probability $P(A|B)$, the probability of A given B.

  : $P(A|B) \equiv \dfrac{P(A \cap B)}{P(B)}$.

  - Intuition: rearrange to get $P(A \cap B) = P(A|B)\,P(B)$.

- Law of total probability: suppose $A_1, \dots A_n$ partition Ω.

  Then $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots$

  $= \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

  - Usually we know the rhs and want to calculate the left.

- Now $P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A)$.
- Bayes' Rule: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$
  - $P(A)$: "prior probability"
  - $P(A|B)$: "posterior probability"
  - Tells us how to update our beliefs when we receive new information.
  - Denominator: use law of total probability.
- Problem 2.

- Independence: Events A and B are independent if

  $P(A \cap B) = P(A) \cdot P(B)$, i.e. $P(A|B) = P(A)$.

  - knowing that B occurred provides no information about whether A will occur.
- Conditional independence: A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C) \cdot P(B|C)$.
- C.I. does not imply I. Intuition: A and B may be dependent because B contains information about A, but C may contain the same information, so A and B conditionally independent given C.
- Does I. imply C.I.? Problem 3.

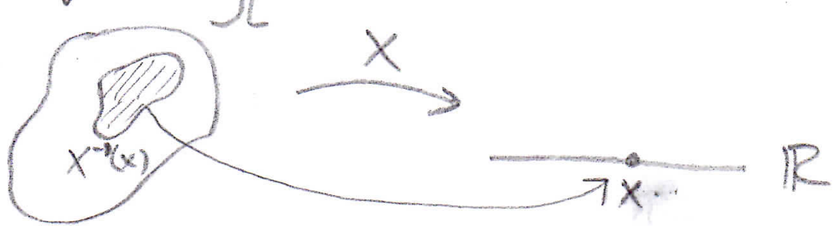
- Random Variables
  - Def: a random variable X is a function from Ω to $\mathbb{R}$.

    $X: \Omega \to \mathbb{R}$.
  - What's random about it? The argument - we don't know where it will be evaluated.
  - X is a way of summarizing the sample space.
  - Ex: paramagnet with $N=2$; X is the spin excess.

    $X(\uparrow\uparrow) = 2$; $X(\uparrow\downarrow) = 0$; $X(\downarrow\uparrow) = 0$;

    $X(\downarrow\downarrow) = -2$.

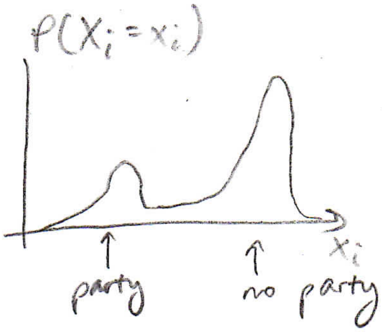
- Probability distribution of a random variable:



- $X^{-1}(x)$ is an event in Ω.

  $P(X=x) \equiv P(X^{-1}(x))$. Also $f_X(x)$.

- All the rules of probability also apply to probability distributions of random variables.
- In statistical mechanics: $\begin{cases} \text{elements of } \Omega \Leftrightarrow \text{microstates} \\ \text{values of } X \Leftrightarrow \text{macrostates.} \end{cases}$
- Abuse of notation: $P(X) = Ae^{-X^2}$ means $P(X=x) = Ae^{-x^2}$.

- The Central Limit Theorem
  - Estimating your grade:
    - Random variables X_i : scores on individual assignments.
    - 

    - Assume the X_i are i.i.d.
    - Grade = $X = \sum_{i=1}^{N} X_i$.
  - We might expect X to have a complicated distribution, but according to the Central Limit Theorem,

    $p(X=x) \rightarrow$ Gaussian distribution as $N \rightarrow \infty$.

    - Mean $\langle X \rangle = N \langle X_i \rangle$
    - Variance $\Delta X^2 = N \Delta X_i^2$; std. dev. $\Delta X = \sqrt{N} \Delta X_i$.
  - Actually, the X_i don't need to be exactly identically distributed, and they don't have to be perfectly independent either.

Names:

1. (*) Consider a sample space Ω with events $A \subset \Omega$ and $B \subset \Omega$ along with probability function $f : \Omega \to \mathbb{R}$. Using the definition of probability, prove the following statements and then **restate each statement in plain English**:

   a. $P(\Omega) = 1$.

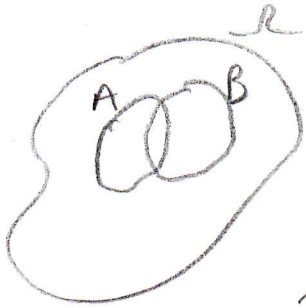
   $$P(\Omega) = \sum_{x \in \Omega} f(x) = 1.$$ "Something will happen."

   b. $P(A) = 1 - P(A^c)$, where $A^c = \{x \in \Omega \mid x \notin A\}$ is the complement of A.

   $$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

   "Either A happens or not."

   c. $P(A \cup B) = P(A) + P(B) - P(B \cap A)$. Hint: draw a picture. Additional hint: write $A \cup B$ as a union of disjoint sets.



   $$P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c)$$

   and $B = (B \cap A) \cup (B \cap A^c)$, so

   $$P(B) = P(B \cap A) + P(B \cap A^c) \checkmark$$

   The probability of A or B is the probability of A plus the probability of B, minus the probability of both.

2. (*) After reading about a rare disease online, you go to the doctor to take a blood test. The disease afflicts fraction d of the general population. Let D be the event that you have the disease, and let T be the event that you test positive for the disease. The blood test has accuracy a, meaning that $P(T|D) = a$ and also that $P(T^c|D^c) = a$.

   a. What is the probability that you have the disease, given that you test positive for the disease? Write the result in terms of d and a. Hint: use the law of total probability to expand the denominator of Bayes' rule.

$$P(D|T) = \frac{P(T|D)\,P(D)}{P(T)} = \frac{P(T|D)\,P(D)}{P(T|D)\,P(D) + P(T|D^c)\,P(D^c)}.$$

Now $P(T|D^c) = 1 - P(T^c|D^c) = 1 - a$, so

$$P(D|T) = \frac{a \cdot d}{ad + (1-a)(1-d)}.$$

b. Suppose $d = 1 - a$. What is the probability that you have the disease, given that you test positive for the disease? What is the intuitive explanation of this result?

$$P(D|T) = \frac{1}{2}.$$

Intuition: either we have the disease and the test succeeded, or we don't and the test failed.
If $d = 1-a$, these are equally likely.

3. (*) Consider a paramagnet of two spins with $p = 1/2$. Let A be the event that the first spin points up. Let B be the event that the second spin points up. Assume that these events are independent, so in particular, $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

a. Let C be the event that exactly one spin points up. What is $P(A|C)$? What is $P(B|C)$? What is $P(A \cap B \mid C)$?

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{1/4}{1/2} = 1/2$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/4}{1/2} = 1/2$$

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} = 0.$$

b. Are the spins still independent if we know the total energy of the system? Or in other words, does independence imply conditional independence? In plain English, explain why or why not.

No, since $P(A \cap B|C) \neq P(A|C)P(B|C)$.

Constraints introduce dependence between the spins, since one spin can be written as a function of the other spin and the constraint.

4. Consider a paramagnet of N spins.

   a. Write down one of the elements of the sample space, schematically. How many distinct elements are in the sample space?

   $\underbrace{\uparrow\downarrow\downarrow\uparrow \cdots \downarrow}_{N}$; 2^N elements $\rightarrow$ "microstates"

   b. How many distinct *events* are in the sample space? Hint: way more than you thought.

   Each element can be included in the event or not, so there are 2^{2^N} events.

   c. Which of the following does the statement "At least two spins point up" describe?

      i. An element of the sample space.
      ii. An event.
      iii. A random variable.

   ii. Could be "$A \equiv \{x \in \Omega \mid N_\uparrow(x) \geq 2\}$ occurred."
   iii. Could be "$N_\uparrow \geq 2$."

   d. Write down some examples of random variables that we have used in this course to describe the paramagnet. How many distinct values are in the image of each random variable? What is the largest number of distinct values that could be in the image of a random variable associated with the paramagnet?

Random Variables	Size of image
$N_\uparrow$	$N+1$
X	$N+1$
$\vdots$	

There could be up to 2^N elements in the image, but such a random variable wouldn't be useful to us.

Values of random variables $\rightarrow$ "macrostates".

5. The **Central Limit Theorem** states that the probability distribution for the sum $X = \sum_i X_i$ of a large number of independent, identically distributed random variables X_i is a **Gaussian distribution** with mean $\langle X \rangle$ and variance $\Delta X^2 \equiv \langle X^2 \rangle - \langle X \rangle^2$ given by

$$\langle X \rangle = \sum_i \langle X_i \rangle, \tag{1}$$

$$\Delta X^2 = \sum_i \Delta X_i^2. \tag{2}$$

Remarkably, this result holds (with a few caveats) irrespective of the probability distribution from which the summands X_i are drawn.

a. We showed last time that

$$I \equiv \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \tag{3}$$

so that a properly normalized probability density function with mean $\mu = 0$ is given by

$$f_X(x) = \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha x^2}. \tag{4}$$

Express the standard deviation $\sigma \equiv \Delta X$ in terms of α and rewrite $f_X(x)$ in terms of σ.

$$\langle x^2 \rangle = \sqrt{\frac{\alpha}{\pi}} \int_{-\infty}^{\infty} x^2 e^{-\alpha x^2} dx$$

$$= -\sqrt{\frac{\alpha}{\pi}} \frac{\partial}{\partial \alpha} \underbrace{\int_{-\infty}^{\infty} e^{-\alpha x^2} dx}_{I}$$

$$= -\sqrt{\frac{\alpha}{\pi}} \frac{\partial}{\partial \alpha} \sqrt{\frac{\pi}{\alpha}} = \frac{1}{2\alpha}.$$

Since $\mu = 0$, $\sigma = \sqrt{\langle x^2 \rangle} \iff \alpha = \frac{1}{2\sigma^2}$.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$ or in general,

$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}}$$

b. We want to write a Gaussian approximation to the (binomial) probability distribution of the spin excess X. First, discuss: does this random variable satisfy the conditions of the Central Limit Theorem? In particular, in what way is X a "sum of a large number of independent, identically distributed random variables?"

• Random variables: $X_i \equiv \begin{cases} 1, & \text{spin } i \text{ is up} \\ -1, & \text{spin } i \text{ is down}. \end{cases}$

Then $X \equiv N_\uparrow - N_\downarrow = \sum_{i=1}^{N} X_i$.

c. Using Eqs. 1 and 2, evaluate the mean $\langle X \rangle$ and standard deviation ΔX of the spin excess X in terms of N and p. Are the results consistent with those that you calculated from the binomial distribution last class?

$\langle X \rangle = N \langle X_i \rangle = N \cdot \left[1 \cdot p + (-1)(1-p) \right] = N \cdot (2p-1)$. ✓

$\Delta X^2 = N \Delta X_i^2 = N \left[1 \cdot p + 1 \cdot (1-p) - (2p-1)^2 \right] = N \cdot 4p(1-p)$

so $\Delta X = 2\sqrt{Np(1-p)}$ ✓

d. Write down Gaussian approximations to...

i. the probability density function $P_N(X = x)$ for $p = 1/2$.

ii. the multiplicity $g_N(X = x)$ for $p = 1/2$.

i. $\langle X \rangle = 0$, $\Delta X = \sqrt{N} \Rightarrow P_N(X=x) = \frac{1}{\sqrt{2\pi N}} e^{-x^2/2N}$

ii. $g_N(X=x) = 2^N P_N(X=x) = \frac{2^N}{\sqrt{2\pi N}} e^{-x^2/2N}$