Probability Review

APL 405

Why do we care about probability?

Uncertainty arises through:

- ❑Noisy measurements
- ❑Variability between samples
- ❑Finite size of data sets

Probability provides a consistent framework for the quantification and manipulation of uncertainty.

Sample Space

Sample space Ω is the set of all possible outcomes of an experiment.

Observations $\omega \in \Omega$ are points in the space also called sample outcomes, realizations, or elements.

Events $E \subset \Omega$ are subsets of the sample space.

Sample Space

Example: Flip a coin twice:

Sample space includes all possible outcomes

$\Omega = \{HH, HT, T\,H, T\,T\}$

Observation is any single element of the sample space

$\omega = HT \in \Omega$

Event is a subset of the sample space (eg. the event where both flips have the same outcome)

$E = \{HH, T\,T\} \subset \Omega$

Probability

The probability of an event E, P(E), satisfies three axioms:

1. P(E) ≥ 0 for every E
2. P(Ω) = 1
3. If E_1, E_2, . . . are disjoint then

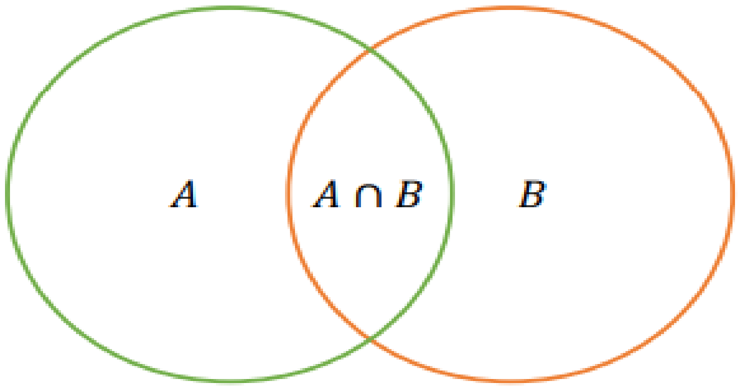
$$P\left(\bigcup_{i=1}^{\infty}(E_i)\right) = \sum_{i=1}^{\infty} P(E_i)$$

Joint and Conditional Probabilities

Joint Probability of A and B is denoted P(A, B).

Conditional Probability of A given B is denoted P(A|B).

Joint: $p(A, B) = p(A \cap B)$



Conditional: $p(A|B) = \frac{p(A \cap B)}{p(B)}$

$$p(A, B) = p(A|B)p(B) = p(B|A)p(A)$$

Conditional Example

Probability of passing the midterm is 60% and probability of passing both the final and the midterm is 45%.

What is the probability of passing the final given the student passed the midterm?

$$P(F|M) = P(M,F)/P(M)$$

$$= \frac{0.45}{0.60} = 0.75$$

Independence

Events A and B are independent if $P(A, B) = P(A)P(B)$.

Independence

Suppose you have 2 coins. Coin 1 always comes up Heads and Coin 2 always comes up Tails. You close your eyes, pick a coin and toss it. Then you replace it, pick again and toss again.

- Independent: Before seeing the result of any toss, you wonder about 2 events; A: first toss is Head, B: second toss is Head.

$$P(A, B) = 0.5 \times 0.5 = P(A)P(B)$$

- Not Independent: Now you wonder about the same events A and B but you toss the same coin twice.

$$P(A, B) = 0.5 \neq P(A)P(B)$$

Conditional Independence

Events A and B are conditionally independent given C if

$$P(A, B|C) = P(B|C)P(A|C)$$

Consider two coins: A regular coin and a coin which always outputs heads.

A = The first toss is heads,
B = The second toss is heads,
C = The regular coin is used,
D = The biased coin is used.

Then A and B are conditionally independent given C and given D

Conditional Dependence

Events A and B are conditionally independent given C if

$$P(A, B|C) \;=\; P(A|C)P(B|C)$$

Consider a coin which outputs heads if the first toss was heads, and tails otherwise.

A = The first toss is heads;

B = The second toss is heads;

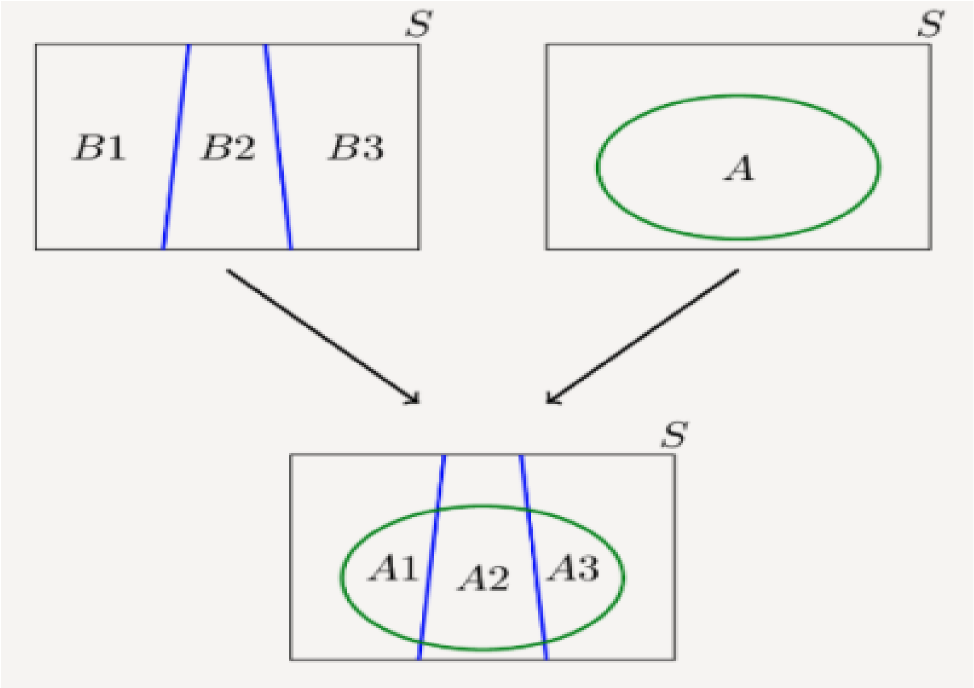
E = The eventually biased coin is used

Then A and B are conditionally dependent given E.

Marginalization and Law of Total Probability

Law of Total Probability

$$P(A) = \sum_{B} P(A, B) = \sum_{B} P(A|B)P(B)$$



Bayes' Rule

Bayes' Rule:

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

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

$$Posterior \propto Likelihood * Prior$$

Bayes' Example

Suppose you have tested positive for a disease. What is the probability you actually have the disease?

This depends on the prior probability of the disease:

- $P(T = 1|D = 1) = 0.95$ (likelihood)
- $P(T = 1|D = 0) = 0.1$ (likelihood)
- $P(D = 1) = 0.1$ (prior)

So $P(D = 1|T = 1) = ??$

Bayes' Example

$P(D = 1|T = 1) =?$

Use Bayes' Rule:

$$P(T = 1) = P(T = 1|D = 1)P(D = 1) + P(T = 1|D = 0)P(D = 0)$$
$$= 0.95 * 0.1 + 0.1 * 0.90 = 0.185$$
$$P(D = 1|T = 1) = \frac{P(T = 1|D = 1)P(D = 1)}{P(T = 1)} = \frac{0.95 * 0.1}{P(T = 1)} = 0.51$$

Random Variable

How do we connect sample spaces and events to data?

A random variable is a mapping which assigns a real number $X(\omega)$ to each observed outcome $\omega \in \Omega$.

For example, let's flip a coin 10 times. $X(\omega)$ counts the number of Heads we observe in our sequence. If ω = HHT HT HHT HT then $X(\omega) = 6$. We often shorten this and refer to the random variable X.

Probability Distribution Statistics

Expectation: First Moment, μ

$$E_x[x] = \sum_{i=1}^{\infty} x_i p(xi)$$ (univariate discrete r.v.)

$$E_x[x] = \int_{-\infty}^{\infty} xp(x)dx$$ (univariate continuous r.v)

Variance: Second (central) Moment, σ^2

$$Var[x] = \int_{-\infty}^{\infty} (x - \mu)^2 \, p(x)dx$$

$$= E_x[(x - \mu)^2]$$

$$= E_x[x^2] - E_x[x]^2$$

Expectation as Monte Carlo average

We consider *n*-samples of a random variable *X's,* $(x_1,\ldots,x_n)$, and compute the mean of these over the number of samples, then would have the *Monte Carlo estimate* of $E[x]$ as

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Expectation Practice

- What is the expected value of a fair die?
- X = value of roll

$$E_X[X] = \sum_{a \in \{1,2,3,4,5,6\}} \frac{1}{6} a$$

$$= \frac{1}{6} \sum_{a=1} a$$

$$= 21/6 = 7/2$$

Linearity of Expectations

There are two powerful properties regarding expectations.

- ❑ $E[X+Y] = E[X] + E[Y]$. This holds even if the random variables are dependent.
- ❑ $E[cX] = cE[X]$, where c is a constant.

Note we cannot say anything in general about $E[XY]$.

Linearity of Expectation Practice

What is the expected value of the sum of two dice?

X_1 = value of roll 1

X_2 = value of roll 2

$$E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{7}{2} + \frac{7}{2} = 7$$

Else we compute like $2 * \frac{1}{36} + 3 * \frac{2}{36} + \ \ldots$

Linearity of Expectation Practice 2

Suppose there are n students in class, and they each complete an assignment. We hand back assignments randomly. What is the expected number of students that receive the correct assignment?

When n = 3? In general?

X = Number of students that get their assignment back

X_i = Student *i* gets their assignment back

$$E[X] = E[X_1 + X_2 + \ldots + Xn]$$
$$= E[X_1] + E[X_2] + \ldots + E[X_n]$$
$$= \frac{1}{n} * n = 1$$

Variances

Knowing the expectation can only tell us so much. We have another quantity used to describe how far off we are from the expected value.

It is defined as follows for a random variable X with expectation as μ,

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

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

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

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

Variance Properties

Constants get squared:

$$Var[cX] = c^2 Var[X]$$

For independent random variables X and Y , we have

$$E[XY] = E[X]E[Y]$$

and

$$Var[X + Y] = Var[X] + Var[Y]$$

Variance Practice

Consider a particle that starts at position 0. At each time step, the particle moves one step to the left or one step to the right with equal probability. What is the variance of the particle at time step n?

$X = X_1 + X_2 + ... + Xn$

Each X_i is 1 or -1 with equal probability

$$Var(X_i) = 1$$

$$Var(X) = \sum Var(X_i) = n$$

The expected squared distance from 0 is *n*.

Discrete and Continuous Random Variables

Discrete Random Variables

- Takes countably many values, e.g., number of heads
- Distribution defined by probability mass function (PMF).
- Marginalization: $p(x) = \sum_y p(x, y)$

Continuous Random Variables

- Takes uncountably many values, e.g., time to complete task
- Distribution defined by probability density function (PDF)
- Marginalization: $p(x) = \int_y p(x, y) dy$

I.I.D.

Random variables are said to be independent and identically distributed (i.i.d.) if they are sampled from the same probability distribution and are mutually independent.

This is a common assumption for observations. For example, coin flips are assumed to be i.i.d.