MXB101

Lecture Notes

MAXWELL HANKS

PROBABILITY AND STOCHASTIC MODELLING

Events and Probability

WHAT IS PROBABILITY?

Classical Definition

- N – number of possible outcomes
- N_A – number of outcomes; event A
  - $Pr(A) = \frac{N_A}{N}$ } only if equally likely

Frequency Definition

- Say you perform n of N_A satisfy A: by observation:
  $Pr(A) \approx \frac{n_A}{n}$, $\therefore Pr(A) = \lim_{n \to \infty} \frac{n_A}{n}$
  - inductive, basis of statistical theory

DEFINITIONS

- A set A is an unordered collection of elements.
- We denote
  $A = \{a_1, a_2, \dots, a_n\}$, for $\underbrace{a \in A}_{\text{element of } A}$ and $\underbrace{A \notin A}_{\text{not element of } A}$
- Special sets; $\emptyset = \{\}$, no elements
- Universal set: $\mathcal{U} = \{\text{all elements in universe}\}$ (or Ω)

E.g., $A = \{\text{all positive even numbers}\}$
$A = \{2, 4, 6, \dots\}$
$A = \{a \in \mathbb{Z} : a > 0, \exists k \in \mathbb{Z}, a = 2k\}$

SET OPERATORS

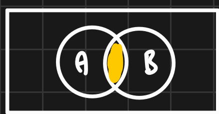
1. $\bar{A}$ or A^c complement of A; not A (depends on universal set)
   $\Rightarrow \bar{A} = \{x : x \notin A\}$
   $\Rightarrow \bar{\Omega} = \emptyset$
2. $A \cup B$ or $A + B$; union, A or B (or both)
   $\Rightarrow A \cup B = \{x : x \in A, \text{or } x \in B\}$
3. $A \cap B$ or AB; intersection; (both)
   $\Rightarrow A \cap B = \{x ; x \in A \text{ and } x \in B\}$

1.

$\bar{A}$

2.

3.

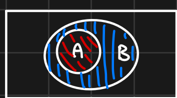


PROPERTIES

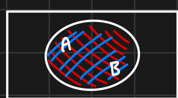
1. $A \cup B = B \cup A$ } commutativity
2. $A \cap B = B \cap A$
3. $(A \cup B) \cup C = A \cup (B \cup C)$ } associativity
4. $(A \cup B) \cap C = A \cup (B \cap C)$
5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ } distributive laws
6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
7. $\overline{A \cup B} = \bar{A} \cap \bar{B}$ } De Morgan's laws
8. $\overline{A \cap B} = \bar{A} \cup \bar{B}$

SET RELATIONS

1. $A = B$, sets are equal; $\begin{cases} x \in A, x \in B, \text{ and} \\ x \in B, x \in A \end{cases}$
2. $A \subset B$, proper subset; $\exists x \in B \rightarrow x \notin A$
3. $A \subseteq B$, subset, if $x \in A$; $x \in B$ (also equal)



$A \subset B$



$A = B$

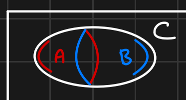
both $A \subseteq B$

PROPERTIES

1. $A \subseteq A \cup B$
2. $A \cap B \subseteq A$
3. $A = B$ iff $(A \subseteq B) \cap (B \subseteq A)$
4. $\emptyset \subseteq A \subseteq \mathcal{U} / \Omega$
5. if $A \subseteq B$, then $A \cap B = A$; $A \cup B = B$ **

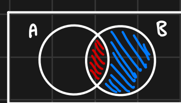
DISJOINT

1. if $A \cap B = \emptyset$, disjoint
2. if $A \cup B = C$, $A \cap B$ exhaust C



1. $A \cap \bar{A} = \emptyset$
2. $A \cup \bar{A} = \mathcal{U} = \Omega$
3. If $A \cap B \neq \emptyset$; $\boxed{A \cap B}$ and $\boxed{\bar{A} \cap B}$ exhaust B !!!

Partition



Events and Probability

EVENTS

- Let Ω (sample space) denote the set of all possible outcomes
- An event A is any subset Ω, (not continuous spaces)
  - if $A \subseteq \Omega$, $\bar{A} \subseteq \Omega$
  - if $A, B \subseteq \Omega$, then $A \cap B \subseteq \Omega$, and $A \cup B \subseteq \Omega$

AXIOMS OF PROBABILITY

- Three fundamental axioms;
  - A1. For $A \subseteq \Omega$, $Pr(A) \geq 0$
  - A2. $Pr(\Omega) = 1$
  - A3. If $AB = \phi$, $\underbrace{Pr(A \cup B) = Pr(A) + Pr(B)}_{\text{addition rule (disjoint)}}$

IMPLICATIONS

1. $Pr(\bar{A}) = 1 - Pr(A)$] Complement
2. $Pr(\phi) = 0$
3. $Pr(B) = Pr(A \cap B) + Pr(\bar{A} \cap B)$] Total probability
4. $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$] Addition
5. If $A \subseteq B$, $Pr(A) \leq Pr(B)$
6. 〃 $Pr(A \cap B) = Pr(A)$
7. 〃 $Pr(A \cup B) = Pr(B)$
   ] Subset
8. $Pr(A) \leq 1$
9. $Pr(\overline{A \cup B}) = Pr(\bar{A} \cap \bar{B})$
10. $Pr(\overline{A \cap B}) = Pr(\bar{A} \cup \bar{B})$
   ] De morgan

PROOFS

1. $Pr(\bar{A}) = 1 - Pr(A)$ RTP

   Since $\Omega = \bar{A} \cup A$, then

   $Pr(\bar{A} \cup A) = 1$ {A2}

   Since $A \cap \bar{A} = \phi$, then

   $Pr(\bar{A}) + Pr(A) = 1$ {A3}

   $\therefore Pr(\bar{A}) = 1 - Pr(A)$ QED

2. $Pr(\phi) = 0$

   Since $\phi = \bar{\Omega}$, $Pr(\phi) = 1 - Pr(\Omega)$ {complement}

   $= 0$, QED

3. $Pr(B) = Pr(A \cap B) + Pr(\bar{A} \cap B)$ RTP {Total prob}

   $A \cap B$ and $\bar{A} \cap B = \phi$ and exhaust B

   So $Pr(B) = Pr((A \cap B) \cup (\bar{A} \cap B))$ {exhaust}

   $\therefore = Pr(A \cap B) + Pr(\bar{A} \cap B)$

PROOFS

4. $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

   Note $A \cup B = (A \cap B) \cup (A \cup \bar{B}) \cup (\bar{A} \cap B)$ {disjoint}

   So:

   $Pr(A \cup B) = Pr(A \cap B) + Pr(A \cap \bar{B}) + Pr(\bar{A} \cap B)$

   $= Pr(A \cap B) + Pr(A) - Pr(A \cap B) + Pr(B) - Pr(A \cap B)$

   $= Pr(A \cap B) + Pr(A) + Pr(B) - 2Pr(A \cap B)$

   $= Pr(A) + Pr(B) - Pr(A \cap B)$ QED

5. $A \subseteq B$, then $Pr(A) \leq Pr(B)$ RTP

   Proof; $B = (\bar{A} \cap B) \cup A$ {$A, \bar{A} \cap B \in \phi$, exhaust B}

   So; $Pr(B) = Pr(\bar{A} \cap B) + Pr(A)$ {disjoint!}

   $\therefore Pr(A) \leq Pr(B)$ {A1; $Pr(\bar{A} \cap B) \geq 0$}

6. If $A \subseteq B$; $Pr(A \cap B) = Pr(A)$; $A \subseteq B$, $A \cap B = A$
7. If $A \subseteq B$; $Pr(A \cup B) = Pr(B)$; $A \subseteq B$, $A \cup B = B$

   QED

8. $Pr(A) \leq 1$

   $\rightarrow Pr(\Omega) = Pr(A) + Pr(\bar{A})$ {disjoint}

   Since $Pr(\Omega) = 1$, $Pr(A) + Pr(\bar{A}) = 1$

   $\therefore Pr(A) \leq 1$ {$Pr(\bar{A}) = 1 - Pr(A)$}

9. $Pr(\overline{A \cup B}) = Pr(\bar{A} \cap \bar{B})$

   $Pr(\overline{A \cup B}) = 1 - Pr(A \cup B)$

   $= 1 - Pr((A \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap B))$

   $Pr(\bar{A} \cap \bar{B}) = Pr(\bar{A}) + Pr(\bar{B}) - Pr(\bar{A} \cup \bar{B})$

   Since $\overline{A \cup B} = \bar{A} \cap \bar{B}$, QED

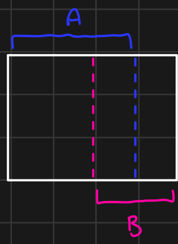
INDEPENDENCE

- A and B are independent if A does not affect probability of B
- If A, B are independent;
  - $Pr(A \cap B) = Pr(A)Pr(B)$
  - Independence is <u>NOT</u> same as disjoint!

Independence and Conditional Probability

CONDITIONAL PROBABILITY

- Conditional probability; probability of A given B; – $\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}}$



$Pr(A) = 3/4$
$Pr(B) = 1/2$
$Pr(A|B) = 1/2$

CHANGING THE SAMPLE SPACE

- All probability rules hold when conditioning on the same event (C)
  - $Pr(\bar{A}|C) = 1 - Pr(A|C)$ {Complement}
  - $Pr(A \cup B|C) = Pr(A|C) + Pr(B|C) - Pr(A \cap B|C)$ {Addition}
  - $Pr(A \cap B|C) = Pr(A|B \cap C)Pr(B|C)$ {Multiplication}

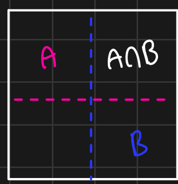
MULTIPLICATION RULE

- If A and B are not independent, then
  - $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$

    $\Longrightarrow \boxed{Pr(A \cap B) = Pr(A|B)Pr(B)}$
- This is general multiplication rule for dependent.

INDEPENDENCE

- One event does not affect the Pr of another;
  - $Pr(A|B) = Pr(A)$
  - $Pr(B|A) = Pr(B)$
  - $Pr(A|B) = Pr(A) = \frac{Pr(A \cap B)}{Pr(B)}$
  - $Pr(A)Pr(B) = Pr(A \cap B)$



CONDITIONAL INDEPENDENCE

- Sometimes A and B are not independent, i.e.,
  - $Pr(A|B) \neq Pr(A)$
- But they become independent when we condition on another event C;
  - $Pr(A|B \cap C) = Pr(A|C)$
- A and B are conditionally independent given C.
  - $Pr(A \cap B|C) = Pr(A|C)Pr(B|C)$
- Conversely, A and B may be conditionally dependent but unconditionally independent

DISJOINT EVENTS

- Never occur simultaneously; $Pr(A \cap B) = 0$; $A \cap B = \emptyset$ (mutually exclusive)
  - Therefore; $Pr(A|B) = 0$; $Pr(A \cap B) = 0$

SUBSET EVENTS

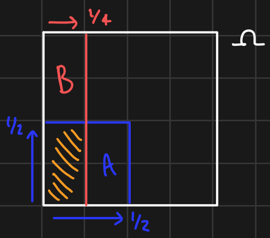
- If $A \subseteq B$; $Pr(A) \leq Pr(B) \Rightarrow A \cup B = B$, $A \cap B = A$

  $\therefore Pr(B|A) = 1$, $\therefore Pr(A|B) = \frac{Pr(A)}{Pr(B)}$

Independence and Conditional Probability

CONDITIONAL PROBABILITY

- Definition; $\boxed{Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)}, \ \forall \, Pr(B) \neq 0}$



$Pr(\Omega) = 1$
$Pr(A) = 1/4$
$Pr(B) = 1/4$
$Pr(A\cap B) = 1/8$
$Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)}$
$= \frac{1}{2}$

SUBSET PROPERTIES

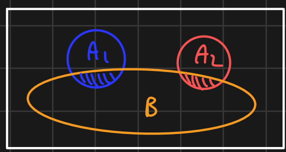
- If $B \subseteq A$, $Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)} = \frac{Pr(B)}{Pr(B)} = 1$ {Subset rule; $B \subseteq A$; $AB = B$}

also

- If $A \subseteq B$ then $Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)} = \frac{Pr(A)}{Pr(B)} \geqslant Pr(A)$ {Subset rule; $Pr(B) \leq 1$}

IMPORTANT

1. $Pr(A|B) \geqslant 0$; since if $Pr(B) \geqslant 0$, $Pr(A\cap B) \geqslant 0$
2. $Pr(B|B) = 1$; since $B \subseteq \Omega$; $Pr(\Omega|B) = 1$
3. For $A_1 \cap A_2 = \emptyset$, $Pr(A_1 \cup A_2 | B) = Pr(A_1|B) + Pr(A_2|B)$



↳ Consider $Pr(A_1 \cup A_2 | B) = \frac{Pr((A_1\cup A_2)\cap B)}{Pr(B)}$

$= \frac{Pr((A_1\cap B)\cup(A_2\cap B))}{Pr(B)}$ Since $A_1 \cap A_2 = \emptyset \Rightarrow \frac{Pr(A_1\cap B) + Pr(A_2\cap B)}{Pr(B)}$

$= Pr(A_1|B) + Pr(A_2|B)$

INDEPENDENCE

- Consider $Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)}$

  rearrange; $Pr(A\cap B) = Pr(A|B)Pr(B)$] – {multiplication rule}

- For independence; $Pr(A\cap B) = Pr(A)Pr(B)$ ← recovered from general case $Pr(A|B) = Pr(A)$

  ↳ A and B are independent if;
  - $Pr(A|B) = A$
  - $Pr(B|A) = B$
  - $Pr(A\cap B) = Pr(A)Pr(B)$

  All hold true!

$Pr(A) = 1/3$
$Pr(B) = 1/4$
$Pr(A\cap B) = 1/12$

$Pr(A|B) = 1/3 = Pr(A)$
$Pr(B|A) = 1/4 = Pr(B)$

Common errors

- Subsets are <u>not</u> independent
  $A \subseteq B$, $Pr(A|B) = \frac{Pr(A)}{Pr(B)} \geqslant Pr(A)$
- Disjoint events <u>not</u> independent
  $Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)} = \frac{0}{Pr(B)} = 0$

Related Ideas

- Mutual indendence; $Pr(A\cap B\cap C) = Pr(A)Pr(B)Pr(C)$
- Pairwise independence; $Pr(A\cap B) = Pr(A)Pr(B)$, $Pr(B\cap C) = \ldots$ etc.
- Conditional independence; $Pr(A\cap B|C) = Pr(A|C)Pr(B|C)$
  ↳ $Pr(A\cap B|C) = \ldots$ does <u>NOT</u> imply $Pr(A|B) = Pr(A)Pr(B)$

Total Probability

LAW OF TOTAL PROBABILITY

- We can write A as $AB \cup A\bar{B}$
- Noting $B \cap \bar{B} = \emptyset$, $B \cup \bar{B} = \Omega$; using addition rule:

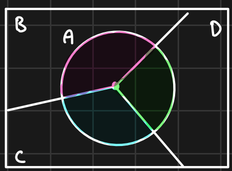
$$\boxed{Pr(A) = Pr(AB) + Pr(A\bar{B})}$$

- Applying multiplication rule:

$$\boxed{Pr(A) = Pr(A|B)Pr(B) + Pr(A|\bar{B})Pr(\bar{B})}$$

- Law of TP can be derived by partitioning Ω into a collection of disjoint events, $B_1, \dots B_n \Rightarrow \bigcup_1^n B_n = \Omega$

$$\boxed{Pr(A) = \sum_{i=1}^{n} Pr(A|B_i)Pr(B_i)}$$



$B \cap C \cap D = \emptyset$

$B \cup C \cup D = A$

BAYES' RULE

- Want to calculate $Pr(A|E)$, but only know $Pr(E|A)$
- Start with conditional probability rule:

$$Pr(A|E) = \frac{Pr(A \cap E)}{Pr(E)}$$

- Using general multiplication rule (numerator)

$$\boxed{Pr(A|E) = \frac{Pr(E|A)Pr(A)}{Pr(E)}}$$

- Using law of TP (denominator):

$$\boxed{Pr(A|E) = \frac{Pr(E|A)Pr(A)}{\sum_{i=1}^{n} Pr(E|B_i)Pr(B_i)}}$$

Combinatorics

FUNDAMENTALS

- With replacement?
  - YES → Order important?
    - YES → String $= C^n$
    - NO → Unordered Sample w/ repl. $= \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$
  - NO → Order important?
    - YES → Permutation ${}^nP_k = \frac{n!}{(n-k)!}$
    - NO → Combination ${}^nC_k = \frac{n!}{k!(n-k)!} = \binom{n}{k}$

$$\downarrow$$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{n-k}$$

Random Variables

DISCRETE DISTRIBUTIONS

- Let N be a discrete random variable:

$$\boxed{Pr(N=n) = p_n \forall n}$$

- For this to be a valid pmf:

$$\boxed{Pr(N=n) \geq 0, \forall n, \sum_n Pr(N=n) = 1}$$

CONTINUOUS DISTRIBUTIONS

- A continuous probability is defined using a probability density function (pdf), and also range of values taken.
- Let X be a CRV, the pdf of X is denoted $f(x)$.
- For $f(x)$ to be a valid pdf:

$$\boxed{f(x) \geq 0 \ \forall x, \int_x f(x)dx = 1}$$

- If X is continuous; $\underline{f(x) \neq Pr(X=x)}$
- To compute $Pr(a<X<b)$, integrate over range:

$$\boxed{Pr(a<X<b) = \int_a^b f(x)\,dx}$$

CUMULATIVE DISTRIBUTION FUNCTION

- The cumulative distribution function computes $Pr(X=x)$:

$$\boxed{F(x) = Pr(X \leq x) = \begin{cases} \int_{-\infty}^{x} f(y)dy = Pr(X<x), & X \text{ is cont.} \\ \sum_{y=-\infty}^{y=x} Pr(X=y), & X \text{ is discrete} \end{cases}}$$

- For a continuous random variable X we have that $Pr(X>x) = 1 - Pr(X<x) = 1 - F(x)$. Survival function/comp. cdf
- For a CRV X, the cdf holds:
  - $\lim_{x\to-\infty} F(x) = 0$
  - $\lim_{x\to\infty} F(x) = 1$
  - $F(x)$ is monotonically increasing and continuous.

CDF to PDF

- If we only have access to the cdf $F(x)$, $f(x)$ recovered:

$$\boxed{F(x) = \int_{-\infty}^{x} f(y)dy \Rightarrow f(x) = \frac{dF(x)}{dx} = \frac{d}{dx}\left[\int_{-\infty}^{x} f(y)dy\right]}$$

- Derived from FToC.

MEDIAN

- The median, m, for a CRV x, is a value such that:

$$\boxed{\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = 0.5}$$

QUANTILES

- Cdf gives $F(x) = p, p \in [0,1]$. F is monotone increasing.
- Of interest, x that produces p; $\boxed{x = F^{-1}(p) = Q(p)}$
  - Inverse cdf/ quantile function.

EXPECTED VALUE

- Central tendency; expected value for $g(x)$:

$$\boxed{\begin{array}{ll} \mathbb{E}[X] = \sum_x x\, Pr(X=x) & \text{Disrete} \\ \mathbb{E}[X] = \int_x x\, f(x)dx & \text{Continuous} \end{array}}$$

- If X is continuous, another identity arises:

$$\boxed{\mathbb{E}[X] = \int_{x<0} F(x)dx + \int_{x>0} (1-F(x))dx}$$

  - Result obtained by integration by parts to $\int_x x f(x) dx$.

VARIANCE

- Spread of the function for X is variance:

$$\boxed{\begin{aligned} Var[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}}$$

  - $Var[X] \geq 0$, $Var[a] = 0$, $a \in \mathbb{R}$
- For continuous X, we have

$$\boxed{\begin{array}{c} \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_x (x-\mu)^2 f(x), \mu = \mathbb{E}[X] \\ \mathbb{E}[X^2] = \int_x x^2 f(x)dx \end{array}}$$

STANDARD DEVIATION

- Common measure of spread: $\boxed{sd(X) = \sqrt{Var(X)}}$

TRANSFORMATIONS

$$\boxed{\begin{array}{l} \mathbb{E}[aX+b] = a\mathbb{E}[X] + b \\ Var[aX+b] = a^2 Var[X] \end{array}}$$

Special Discrete Distributions

DISCRETE UNIFORM DISTRIBUTION

- Single equally likely trials
- Let $X \sim \text{Uniform}(a,b)$; then
  - $\Pr(X=x) = \frac{1}{b-a+1} \quad \forall x \in \{a, a+1, \ldots, b-1, b\}$
  - $\mathbb{E}[X] = \frac{a+b}{2}$
  - $\text{Var}[X] = \frac{(b-a+1)^2-1}{12}$

BERNOULLI DISTRIBUTION

- Two outcomes, one trial
- Let $X \sim \text{Bernoulli}(p)$;
  - $\Pr(X=x) = p^x(1-p)^{1-x}, \quad x = 0, 1$
  - $\mathbb{E}[X] = p$
  - $\text{Var}[X] = p(1-p) = pq \quad (q = 1-p)$

BINOMIAL DISTRIBUTION

- Repeated Bernoulli distribution.
- Let $X \sim Bi(n,p)$; n-trials, p-prob of success.
  - $\Pr(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, n\}$
- For $X \sim Bi(n,p)$, $X = \sum_{i=1}^{n} Y_i \sim \text{Bern}(p)$
  - $\mathbb{E}[X] = np$
  - $\text{Var}[X] = np(1-p) = npq$

GEOMETRIC DISTRIBUTION

- Distribution of number of trials, N, up to and including first success (Bernoulli)
- Let $X \sim \text{Geom}(p)$
  - $\Pr(N=n) = p(1-p)^{n-1}, \quad n \in \mathbb{Z}^+$
  - $\mathbb{E}[N] = \frac{1}{p}$
  - $\text{Var}[N] = \frac{1-p}{p^2}$
- Alternative: N trials to success $\Rightarrow$ Y fails to success
  - $\Pr(Y=y) = p(1-p)^y, \quad y \in \mathbb{Z}^+$
  - $\mathbb{E}[Y] = \mathbb{E}[N-1] = \frac{1-p}{p}$
  - $\text{Var}[Y] = \frac{1-p}{p^2}$

Proof: $\mathbb{E}[N] = \frac{1}{p} \Rightarrow \Pr(N=n) = p(1-p)^{n-1}$

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p\sum_{n=1}^{\infty} n(1-p)^{n-1}$$
$$= -p\sum_{n=1}^{\infty} \frac{d}{dp}(1-p)^n = -p\frac{d}{dp}\sum_{n=1}^{\infty}(1-p)^n$$

Recall: Sum of GP; $1 + r + r^2 + \cdots = \frac{1}{1-r}, \quad |r| > 1$

$$\mathbb{E}[N] = -p\frac{d}{dp}\left[\frac{1}{p} - 1\right] = \frac{1}{p}.$$ QED

Proof: $\text{Var}[N] = \frac{1-p}{p^2} \Rightarrow \mathbb{E}[N(N-1)] = \sum_{n=1}^{\infty} n(n-1)(1-p)^{n-1}p$

$$= p(1-p)\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} = p(1-p)\sum_{n=1}^{\infty}\frac{d^2}{dp^2}(1-p)^n$$
$$= p(1-p)\sum_{n=1}^{\infty}\left[\frac{1}{p} - 1\right] = -p(1-p)\frac{d}{dp}\left[\frac{1}{p^2}\right] = \frac{1-p}{p^2}$$

Alternative: $\text{Var}[N] = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \mathbb{E}[N^2] - \mathbb{E}[N] + \mathbb{E}[N] - \mathbb{E}[N]^2$

$$= \frac{1-p}{p^2}$$ QED

NEGATIVE BINOMIAL DISTRIBUTION

- Number of trials N, until k successes, $k \geq 1$, success p.
- Let $N \sim NB(k,p)$:
  - $\Pr(N=n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = \{k, k+1, \ldots\}$
  - $\mathbb{E}[N] = \frac{k}{p}$
  - $\text{Var}[N] = \frac{k(1-p)}{p^2}$
- For $N \sim NB(k,p)$ $N = \sum_{k=1} Y_k \sim \text{Geom}(p) \Rightarrow Y = N - k$
  - $\Pr(Y=y) = \binom{y+k-1}{k-1} p^k (1-p)^y, \quad y \in \mathbb{Z}^+ \cup 0$
  - $\mathbb{E}[Y] = \mathbb{E}[N-k] = \mathbb{E}[N] - k = \frac{k(1-p)}{p}$
  - $\text{Var}[Y] = \frac{k(1-p)}{p^2}$

POISSON DISTRIBUTION

- Counts number of events N over time λ.
- Let $N \sim \text{Pois}(\lambda)$
  - $\Pr(N=n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad n \in \mathbb{Z}^+ \cup 0$
  - $\mathbb{E}[N] = \text{Var}[N] = \lambda$

Proof: $\mathbb{E}[N] = \lambda \Rightarrow \Pr(N=n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad n \in \mathbb{Z}^+ \cup 0$

$$\mathbb{E}[N] = \sum_{n=0}^{\infty} n\frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1} n\frac{\lambda^n e^{-\lambda}}{n!}$$
$$= e^{-\lambda}\sum_{n=1}\frac{\lambda^n}{(n-1)!}$$
$$\therefore \mathbb{E}[N] = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!}, \quad \text{Let } y = n-1$$
$$\therefore \mathbb{E}[N] = \lambda e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!}$$

Recall $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$

$$\therefore \mathbb{E}[N] = \lambda e^{-\lambda} e^{\lambda}$$
$$\therefore \mathbb{E}[N] = \lambda$$ QED

Proof: $\text{Var}[N] = \lambda \Rightarrow \mathbb{E}[N(N-1)] = \sum_{n=0}^{\infty} n(n-1)\frac{\lambda^n e^{-\lambda}}{n!}$

$$= \sum_{n=2}^{\infty} n(n-1)\frac{\lambda^n e^{-\lambda}}{n!}$$
$$\therefore \mathbb{E}[N(N-1)] = \lambda^2 e^{-\lambda}\sum_{n=2}^{\infty}\frac{\lambda^{n-2}}{(n-2)!}$$

Let $y = n-2$, $\mathbb{E}[N(N-1)] = \lambda^2 e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} = \lambda^2$

$$\text{Var}[N] = \mathbb{E}[N^2] - \mathbb{E}[N]^2$$
$$= \mathbb{E}[N^2] - \mathbb{E}[N] + \mathbb{E}[N] - \mathbb{E}[N]^2$$
$$= \lambda^2 + \lambda - \lambda^2$$
$$\therefore \text{Var}[N] = \lambda$$ QED

MODELLING DATA

- Poisson ($\mathbb{E} = \text{Var}$)
- Binomial (underdispersed, $\mathbb{E} > \text{Var}$)
- Geom/NB (overdispersed, $\mathbb{E} < \text{Var}$)

Special Continuous Distributions

UNIFORM DISTRIBUTION

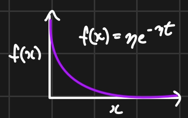
- Simplest continuous distribution
- Values between a and b.
- The probability of getting an outcome within some interval is the same as all other intervals of the same length.
- The probability density function is proportional to k.
- X is uniform over (a,b), $X \sim U(a,b)$.

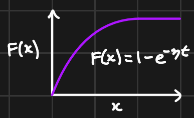
- pdf: $f(x) = \frac{1}{b-a},\ x \in (a,b)$
- cdf: $F(x) = \frac{x-a}{b-a},\ x \in (a,b)$
  - $P(X<x) = \int_a^x \frac{1}{b-a}\,dy = \frac{1}{b-a}[y]_a^x$

- Expected – $\mathbb{E}[X] = \frac{a+b}{2}$
- Variance – $Var[X] = \frac{(b-a)^2}{12}$
- Median – $m = \frac{a+b}{2}$

EXPONENTIAL DISTRIBUTION

- Used to model time between events
- Has non-negative values $(0 \to \infty)$
- Requires a rate parameter η, $\eta > 0$.
- $T \sim Exp(\eta)$.





- pdf: $f(t) = \eta e^{-\eta t},\ t > 0$
- cdf: $F(t) = 1 - e^{-\eta t},\ t > 0$
- $P(T<t) = \int_0^t \eta e^{-\eta t}\,dx = 1 - e^{-\eta t}$
- If $T \sim Exp(\eta)$, $N \sim Pois(\eta t)$;
  - $P(T>t) = P(N=0) = e^{-\eta t}$

- Expected: $\mathbb{E}[T] = \frac{1}{\eta}$
- Variance: $Var[T] = \frac{1}{\eta^2}$

- $$\begin{aligned}\mathbb{E}[T] &= \int_0^\infty t\eta e^{-\eta t}\,dt = \eta \int_0^\infty t e^{-\eta t}\,dt \\ &= \eta\left[-\frac{1}{\eta}\left[te^{-\eta t}\right]_0^\infty\right] + \frac{1}{\eta}\int_0^\infty e^{-\eta t}\,dt \\ &= \eta\left[-\frac{1}{\eta^2}\left[e^{-\eta t}\right]_0^\infty\right] = \boxed{\frac{1}{\eta}}\end{aligned}$$

- Memoryless property: let T be the time between events
  - $T \sim Exp(\eta)$

$$\boxed{Pr(T>s+t \mid T>t) = Pr(T>s)}$$

Proof: $Pr(T>s+t \mid T>t) = \frac{Pr(T>s+t \cap T>t)}{Pr(T>t)}$

$$= \frac{Pr(T>s+t)}{Pr(T>t)} = \frac{1-F(s+t)}{1-F(t)} = \frac{e^{-\eta(s+t)}}{e^{-\eta t}} = e^{-\eta s}$$

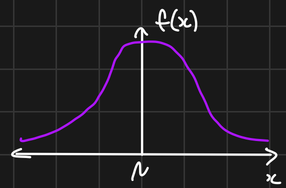
$$= 1 - F(s) = Pr(T>s)$$

NORMAL DISTRIBUTION

- Describes many random situations, measurement/error.
- Can be used to approximate other distributions
- Frequently appears in statistical problems.

- Is unbounded, requires μ and σ^2 (mean, variance)
- Normal distribution: $X \sim N(\mu, \sigma^2)$
- Expected: $\mathbb{E}[X] = \mu$
- Variance: $Var[X] = \sigma^2$

- pdf: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^2\right),\ x \in \mathbb{R}$
- cdf: $F(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right),\ erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$



$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^2\right)$

STANDARD NORMAL DISTRIBUTION

- If $X \sim N(\mu, \sigma^2)$, consider a transformation:

$$Z = \frac{X-\mu}{\sigma}$$

- Then we can show that $Z \sim N(0,1)$, called SND
- Allows to disregard μ and σ^2.

Proof: $\mathbb{E}[Z] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\mathbb{E}[X-\mu] = 0.$

$Var[Z] = Var\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}Var[X-\mu] = \frac{1}{\sigma^2}Var[X] = 1$

Central Limit Theorem

INTRODUCTION

- The CLT states that the sum of independent and identical (iid) distributed random variables, when properly standardised, can be approximated with a normal distribution, as the number of elements get large
- CLT can hold under weaker conditions than iid.

- Let $X_1, X_2, \ldots, X_n$ be iid random variables from the distribution defined by a random variable X.
- Let $\mu = \mathbb{E}[X]$, $\sigma^2 = Var[X]$, $\bar{X} = \frac{\sum_i^n X_i}{n}$, Consider:

$$\boxed{W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}$$

- CLT states as $n \to \infty$, $W \to N(0,1)$.

Proof: $\mathbb{E}[\bar{X}] = \mu$, $std[\bar{X}] = \frac{\sigma}{\sqrt{n}}$ RTP

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{\sum_i^n x_i}{n}\right] = \frac{1}{n}\mathbb{E}\left[\sum_i^n x_i\right]$$

$$= \frac{1}{n}\left[\mathbb{E}[x_1] + \mathbb{E}[x_2] + \cdots \mathbb{E}[x_n]\right] = \frac{n\mu}{n} = \mu \quad \text{QED}$$

$$Var[\bar{X}] = Var\left[\frac{\sum_i^n x_i}{n}\right] = \frac{1}{n^2} Var[X_1 + X_2 + \cdots X_n]$$

$$= \underbrace{\frac{1}{n^2}\left[Var[X_1] + Var[X_2] + \cdots + Var[X_n]\right]}_{\{\text{Independence}\}} = \frac{n\sigma^2}{n^2}$$

$$\therefore Var[\bar{X}] = \frac{\sigma^2}{n}$$

$$std[\bar{X}] = \sqrt{Var[\bar{X}]} = \frac{\sigma}{\sqrt{n}} \quad \text{QED}$$

APPLICATION

- Let $X_1, X_2, \ldots, X_n$ be iid random variables from the distribution defined by a random variable X.
- Let $\mu = \mathbb{E}[X]$, $\sigma^2 = Var[X]$
- Let $Y = X_1 + X_2 + \cdots X_n$
- When n is large, $Y \approx N(n\mu, n\sigma^2)$

- E.g., $X_1, X_2, \ldots X_n$ be iid random variables from $Y \sim Exp(\eta)$
- Let $Y = \sum_k^n X_k$, Y is a gamma distribution:

$$\boxed{f(y) = \frac{\eta^k}{(k-1)!} y^{k-1} e^{-\eta y}, \ y > 0}$$

- Normal approximation: $Y \approx N\left(\frac{k}{\eta}, \frac{k}{\eta^2}\right)$

BINOMIAL APPROXIMATION (NORMAL)

- For n is large, binomial is harder to calculate.
- Normal / Poisson distribution can be used instead.
  - E.g., $X \sim Bi(1000, 0.3)$

$$P(X \geq 700) = \sum_{x=700}^{1000} \binom{1000}{x} 0.3^x 0.7^{n-x}$$

NORMAL APPROXIMATION (CLT)

- $X \sim Bi(n,p)$
- X can be a sum of n iid Bernoulli RVs, $Y_1, Y_2, \ldots, Y_n$ ($Y_i \sim Ber(p)$) for $i \in [1,n]$. $\Longrightarrow X = \sum_i^n Y_i = Y_1 + Y_2 + \cdots + Y_n$.
- Thus, a normal approximation for X can be made for large n:

$$\boxed{X \approx W \sim N(np, np(1-p))}$$

- For $\mu = np$, $\sigma^2 = np(1-p)$.
- This approximation works for $np > 5$, $n(1-p) > 5$.

POISSON APPROXIMATION

- $X \sim Bi(n,p)$
- Assume $np < 5$, a Poisson approximation can be made:

$$\boxed{X \approx Y \sim Pois(np)}$$

- If instead $n(1-p) < 5$, consider $W = n - X$ (number of fails):

$$\boxed{W \approx Y \sim Pois(n(1-p))}$$

APPROXIMATING POISSON PROBABILITIES

- Consider $X = \sum_i^n Y_i$, $Y_i \sim Pois(\xi)$ and are independent.
- Then, $X \sim Pois(\lambda = n\xi)$, thus by CLT:

$$\boxed{X \approx Y \sim N(\lambda, \lambda)}$$

- Approx works for $\lambda > 10$-20.

Bivariate Distributions

TWO RANDOM VARIABLES

- For the DRVs X and Y, the distribution is given by the joint pmf: $\Pr(X=x, Y=y)\ \forall x,y$.
- For two CRVs X and Y, distribution is given by joint pdf $f(x,y)$, defined over some range of X, Y
- May be represented by tabular form, conditions:
  - $\Pr(X=x, Y=y) \geq 0\ \forall x,y$, $f(x,y) \geq 0$
  - $\sum_x \sum_y \Pr(X=x, Y=y) = 1$, $\int_x \int_y f(x,y)\,dy\,dx = 1$

MARGINAL PMFs/PDFs

- From joint PMF, marginal pmf for X, Y

$$\boxed{\Pr(X=x) = \sum_y \Pr(X=x, Y=y),\ \Pr(Y=y) = \sum_x \Pr(X=x, Y=y)}$$

- Compute marginal pdf: $\boxed{f(x) = \int_y f(x,y)\,dy}$

CONDITIONAL PMF

- With joint pmfs, and marginal pmfs: conditional:

$$\boxed{\Pr(X=x \mid Y=y) = \frac{\Pr(X=x, Y=y)}{\Pr(Y=y)}}$$

- Computing CP $\forall X$ for some Y, gives cond. pmf of X given Y
- We must have $\sum_x \Pr(X=x \mid Y=y) = 1$
- Conditional pdf of X given $Y=y$: $f(x \mid y) = \frac{f(x,y)}{f(y)}$
- Hence $\int_x f(x \mid y) = 1 \Rightarrow X \mid Y=y$

INDEPENDENCE

- For DRVs X, Y to be independent, $\Pr(X=x \mid Y=y) = \Pr(X=x)\ \forall x,y$
- Above holds, $\Pr(X=x, Y=y) = \Pr(X=x)\Pr(Y=y)$
- Otherwise general multiplication rule:

$$\boxed{\Pr(X=x, Y=y) = \Pr(X=x \mid Y=y)\Pr(Y=y)}$$

- CRVs are independent iff:
  1. $f(x,y)$ decomposes to $g(x), h(y) \Rightarrow f(x,y) \propto g(x)h(y)$. Hence $f(x,y) = f(x)f(y)$ Independence: $f(x \mid y) = f(x)$
  2. The joint range of X, Y do not depend on eachother

CONDITIONAL EXPECTATION/VARIANCE

- Given a conditional distribution $X \mid Y=y$, compute $\mathbb{E}[X]/\mathrm{Var}[X]$:

$$\boxed{\mathbb{E}[X \mid Y=y] = \int_x x f(x \mid y)\,dx}$$

- For DRV, replace integration with summation.
- Conditional Variance: $\boxed{\mathrm{Var}[X \mid Y=y] = \mathbb{E}[X^2 \mid Y=y] - \mathbb{E}[X \mid Y=y]^2}$
- If X and Y are independent, $\mathbb{E}[X \mid Y=y] = \mathbb{E}[X]$

TOTAL EXPECTATION

- Obtain marginal expectations by law of total expectation.

$$\boxed{\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]}$$

RANDOM SUMS/PRODUCTS

- Given two random variables X, Y, holds $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- If X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

↳ $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

↳ $\mathrm{Var}[XY] = \mathrm{Var}[X]\mathrm{Var}[Y] + \mathbb{E}[X]^2\mathrm{Var}[Y] + \mathbb{E}[Y]^2\mathrm{Var}[X]$.

COVARIANCE

- Measure of dependence between two variables:

$$\boxed{\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}}$$

- We also have $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$
- Joint expectation of X, Y is:

$$\boxed{\mathbb{E}[XY] = \sum_x \sum_y xy \Pr(X=x, Y=y) \Big/ \mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dy\,dx}$$

- Positive: increase in one, increase in another
- Negative: increase in one, decrease in another
- Zero, X and Y are independent.

VARIANCE SUMS

- If X, Y are independent; $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
- If X, Y are dependent; $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X,Y)$

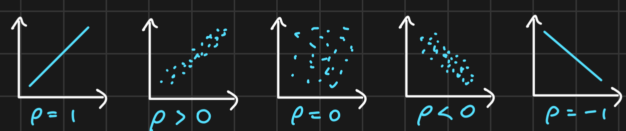
↳ $\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\mathrm{Cov}(X,Y) + 2\mathrm{Cov}(Y,Z) + 2\mathrm{Cov}(X,Z)$

CORRELATION

- Direction and strength of relationship of X, Y

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

- $\rho(X,Y) > 0$, X, Y positive linear
- $\rho(X,Y) < 0$, X, Y negative linear
- $\rho(X,Y) = 0$, independent
- $\rho(X,Y) = 1$, perfect positive linear
- $\rho(X,Y) = -1$, perfect negative linear.



Vectors and Matrices

VECTORS

- Scalar: single number
- Vector: row or column
  - ↳ E.g., $\underset{\sim}{x} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$, $\underset{\sim}{y} = \begin{pmatrix} 0 \\ 7 \end{pmatrix}$
  - $\therefore \underset{\sim}{r} = \underset{\sim}{x} + \underset{\sim}{y} = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$

VECTOR EQUALITY

- Two vectors are equal iff
  - They have the same dimensions
  - All corresponding elements are equal

↳ $\begin{pmatrix} 0.5 \\ 1 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$ Not same dimension

↳ $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 \end{pmatrix}$ Not same dimension.

↳ $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ Not equal (elements)

VECTOR ARITHMETIC

- Add / subtract vectors in same dimension.

↳ $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \\ a_3 - b_3 \end{pmatrix}$

- Multiply every element w/ scalar

↳ $\lambda \times \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \lambda a_3 \end{pmatrix}$

DOT PRODUCT

- Multiply elements and sum.

↳ $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3$

- If $\underset{\sim}{a} \cdot \underset{\sim}{b} = 0$, $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are orthogonal (perpendicular)

PROPERTIES

1. $\underset{\sim}{x} + \underset{\sim}{y} = \underset{\sim}{y} + \underset{\sim}{x}$
2. $(\underset{\sim}{x} + \underset{\sim}{y}) + \underset{\sim}{z} = \underset{\sim}{x} + (\underset{\sim}{y} + \underset{\sim}{z})$
3. $\underset{\sim}{x} + \underset{\sim}{0} = \underset{\sim}{0} + \underset{\sim}{x} = \underset{\sim}{x}$
4. $\underset{\sim}{x} + (-\underset{\sim}{x}) = (-\underset{\sim}{x}) + \underset{\sim}{x} = \underset{\sim}{0}$
5. $\lambda(\underset{\sim}{x} + \underset{\sim}{y}) = \lambda \underset{\sim}{x} + \lambda \underset{\sim}{y}$
6. $(\lambda + \theta)\underset{\sim}{x} = \lambda \underset{\sim}{x} + \theta \underset{\sim}{x}$
7. $(\lambda\theta)\underset{\sim}{x} = \lambda(\theta \underset{\sim}{x})$
8. $1\underset{\sim}{x} = \underset{\sim}{x}$

LINEAR COMBINATION / INDEPENDENCE

- $\underset{\sim}{x} = c_1 \underset{\sim}{v}_1 + c_2 \underset{\sim}{v}_2 + \dots + c_n \underset{\sim}{v}_n$
- Vectors $\underset{\sim}{v}_1, \underset{\sim}{v}_2, \dots, \underset{\sim}{v}_n$ are linearly independent if $\forall c_i = 0$ is solution.

MAGNITUDE

- For $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix}$, $|\underset{\sim}{x}| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$

MATRICES

- Matrix: table of numbers

↳ $\vec{M} = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \end{pmatrix}$ (2 rows × 3 columns)

- Equality: only equal when dimension and elements are equal.

TRANSPOSE MATRIX

- Transpose: rows and columns swapped

↳ $\vec{M} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \Rightarrow \vec{M}^T = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$

MATRIX MULTIPLICATION

- To multiply two matrices together, inner dimensions must match.
  - ↳ E.g., $(4 \times 2) \times (2 \times 3) = (4 \times 3)$
- $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$

IDENTITY MATRIX

- Identity matrix $\vec{I}$ has no effect when multiplying by another.

↳ $\vec{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\vec{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

MATRIX INVERSE

- Let $\vec{A}$ be a square matrix, $\vec{A}^{-1}$ is the inverse:

$$\boxed{\vec{A} \times \vec{A}^{-1} = \vec{I}}$$

MATRIX POWERS

- To take power of a matrix, multiply itself:

↳ $\vec{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\vec{A}^2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(b+d) & cb + d^2 \end{pmatrix}$

MATRIX PROPERTIES

1. $\vec{A} + \vec{B} = \vec{B} + \vec{A}$
2. $\vec{A} + (\vec{B} + \vec{C}) = (\vec{A} + \vec{B}) + \vec{C}$
3. $\vec{A} + \vec{0} = \vec{0} + \vec{A} = \vec{A}$
4. $\vec{A} + -\vec{A} = -\vec{A} + \vec{A} = \vec{0}$
5. $\lambda(\vec{A} + \vec{B}) = \lambda\vec{A} + \lambda\vec{B}$
6. $(\lambda + \theta)\vec{A} = \lambda\vec{A} + \theta\vec{A}$
7. $\lambda\theta\vec{A} = \lambda(\theta\vec{A})$
8. $1\vec{A} = \vec{A}$
9. $(\vec{A}^T)^T = \vec{A}$
10. $(\vec{A} + \vec{B})^T = \vec{A}^T + \vec{B}^T$
11. $\boxed{(\vec{A}\vec{B})^T = \vec{B}^T\vec{A}^T}$
12. $(\lambda\vec{A})^T = \lambda\vec{A}^T$
13. If $\vec{A}$ is symmetric, $\vec{A} = \vec{A}^T$
14. $\vec{A}\vec{B} \neq \vec{B}\vec{A}$
15. $(\vec{A}\vec{B})\vec{C} = \vec{A}(\vec{B}\vec{C})$
16. $\vec{A}(\vec{B} + \vec{C}) = \vec{A}\vec{B} + \vec{A}\vec{C}$
17. $(\vec{A} + \vec{B})\vec{C} = \vec{A}\vec{C} + \vec{B}\vec{C}$
18. $\lambda(\vec{A}\vec{B}) = (\lambda\vec{A})\vec{B} = \vec{A}(\lambda\vec{B})$

Markov Chains

CONDITIONAL PROBABILITY

- Addition rule: $Pr(A \cup B|C) = Pr(A|C) + Pr(B|C) - Pr(AB|C)$
- Multiplication rule: $Pr(AB|C) = Pr(A|BC)Pr(B|C)$
- Complement rule: $Pr(\bar{A}|C) = 1 - Pr(A|C)$

MARKOV CHAIN

- Discrete time and state stochastic process (disjoint/disc. state)
- Let X_t be a RV describing the state at step t.
- Each step changes probabilistically (may not change states)

MARKOV PROPERTY

- Markov property states that:

$$\boxed{Pr(X_t = x_t | X_{t-1} = x_{t-2}, \dots, X_0 = x_0) = Pr(X_t = x_t | X_{t-1} = x_{t-1})}$$

- X_t is conditionally independent of $X_{t-2}, X_{t-3}, \dots, X_0 | X_{t-1}$

HOMOGENEOUS MARKOV CHAINS

- A Markov chain is homogeneous if:

$$\boxed{Pr(X_{t+n} = j | X_t = i) = Pr(X_n = j | X_0 = i) = p_{ij}^{(n)}}$$

- n-step cond. prob. do not depend on step t.

TRANSITION PROBABILITY MATRIX

- A HMC is characterised by a TPM $\underline{\underline{P}}$.
  - Square matrix
  - (i,j)th element of $\vec{P}$ is: $p_{i,j} = Pr(X_t = j | X_{t-1} = i)$
  - $p_{ij} \geq 0 \ \forall i,j \in \mathbb{R}^+$
  - $\sum_j p_{i,j} = 1 \ \forall i \in \mathbb{R}^+$

$$\underline{\underline{P}} = \begin{array}{cc} & X_{t+1} \\ X_t & \begin{bmatrix} & \\ & \end{bmatrix} \end{array}$$

STEPS TO CREATE P

- Identify possible states of MC, write row/column
- Work through matrix row by row
- For each row, identify which states are not possible at the next step and enter these as 0.
- Then, for the possible states at the next stage, identify what must happen to that state. Calculate Pr of that happening.
- Use the fact that $\Sigma_j p_{i,j} = 1$

N-STEP TRANSITION PROBABILITIES

- A TPM $\vec{P}$, holds conditional probabilities of moving from state $i \to j$ in one step $\forall i,j$.
- $\vec{P}^2$ holds conditional probabilities of moving $i \to j$ in 2 steps
- $\vec{P}^n$ provides n-step transition probabilities.

UNCONDITIONAL STATE PROBABILITIES

- Consider the (unconditional) probability of state j at time n:

$$\boxed{Pr(X_n = j) = p_j^{(n)}}$$

- Denote the vector of state probabilities at time n as $\underset{\sim}{s}^{(n)}$ with jth element given by $p_j^{(n)}$:

$$\boxed{\underset{\sim}{s}^{(n)T} = \underset{\sim}{s}^{(n-1)T}\underline{\underline{P}}, \quad \underset{\sim}{s}^{(n)T} = s^{(0)T}\underline{\underline{P}}^n}$$

STEADY STATE STATIONARY DISTRIBUTION

- At steady state, the probability of being in a particular state does not change from one step to the next.
- $\underset{\sim}{s}^{(n+1)} = \underset{\sim}{s}^{(n)} \Rightarrow \underset{\sim}{s}^{(n+1)T} = \underset{\sim}{s}^{(n)T}\underline{\underline{P}} \Rightarrow \boxed{s^{(n)T} = s^{(n)T}\underline{\underline{P}}}$
- The probability of being in any states in steady state is $\underset{\sim}{\pi}$
  - $\underset{\sim}{\pi}^T = \underset{\sim}{\pi}^T \underline{\underline{P}} \quad (\underset{\sim}{\pi}^T = [\pi_0 \ \pi_1 \dots \pi_i])$
  - $\sum_0 \pi_i = 1$
  - $n \to \infty, \ \underline{\underline{P}}^n = \underset{\sim}{\pi}^T$

Poisson Processes

- Stochastic process: events occurring "randomly" in time
- Continuous time, discrete state

SETUP

- Let $X(t)$ be the number of events occurring "randomly" in $(0,t)$. Without loss of generality, $\Pr(X(0)=0)=1$. Let h be a very small time interval such that at most 1 event can happen during that time. Then:
  - $\Pr(X(t+h)=n+1 \mid X(t)=n) \approx \eta h$
  - $\Pr(X(t+h)=n \mid X(t)=n) \approx 1-\eta h$
  - $\Pr(X(t+h)>n+1 \mid X(t)=n) \approx 0$

POISSON PROCESS

$$\boxed{X(t) \sim \text{Poisson}(\eta t)}$$

- $\Pr(X(t)=n) = p_n(t) = \dfrac{e^{-\eta t}(\eta t)^n}{n!}, \quad n \in \mathbb{Z}^+$

EXPONENTIAL DISTRIBUTION

- Let T be the time between events of a Poisson process

$$\boxed{T \sim \text{Exp}(\eta)}$$

- Thus the pdf of T is $f(t) = \eta e^{-\eta t}, \; t>0$

PROPERTIES

- Memoryless property: $\Pr(T>x+y \mid T>x) = \Pr(T>y)$
- Non-overlapping time intervals of a Poisson are indep.
- Exactly 1 event $E \in (0,a)$, $E \sim U(0,a)$

PROOF

Let X be time occurred in $(0,a)$

Let $N(t_1,t_2)$ be # of events in Poise (t_1,t_2)

Consider CDF of X: $\quad (N(t_1,t_2) \sim \text{Pois}(t_1,t_2))$

$$F(x) = \Pr(X<x \mid N(0,a)=1) = \Pr(N(0,x)=1 \mid N(0,a)=1)$$

$$F(x) = \frac{\Pr(N(0,x)=1 \cap N(0,a)=1)}{\Pr(N(0,a))=1} \quad \{\text{cond. rule}\}$$

$$= \frac{\Pr(N(0,x)=1 \cap N(x,a)=0)}{\Pr(N(0,a)=1)}$$

$$= \frac{\Pr(N(0,x)=1)\Pr(N(x,a)=0)}{\Pr(N(0,a)=1)} \quad \{\text{non overlapping intervals}\}$$

$$= \frac{(\eta x)e^{-\eta x}e^{-\eta(a-x)}}{(\eta a)e^{-\eta a}} \Rightarrow = \text{CDF}(x) = \frac{x}{a}$$

$$f(x) = \frac{dF}{dx} = \boxed{\frac{1}{a}, \quad \{0<x<a\}}$$

PROPERTIES CONT

- Exactly n events that occur in $(0,t)$, then number of events that occur in $(0,s<t)$ has $X \sim \text{Bin}(n, \frac{s}{t})$

PROOF

Let X be # of events in $(0,s)$ when n events occur in $(0,t)$, $s<t$

$$\Pr(X=x \mid N(0,t)=n) = \Pr(N(0,s)=x \mid N(0,t)=n)$$

$$= \frac{\Pr(N(0,s)=x \cap N(0,t)=n)}{\Pr(N(0,t)=n)} \quad \{\text{cond rule}\}$$

$$= \frac{\Pr(N(0,s)=x \cap N(s,t)=n-x)}{\Pr(N(0,t)=n)}$$

$$= \frac{\Pr(N(0,s)=x)\Pr(N(s,t)=n-x)}{\Pr(N(0,t)=n)}$$

$$= \frac{(\eta s)^x e^{-\eta s}}{x!}\left(\frac{[\eta(t-s)]^{n-x} e^{-\eta(t-s)}}{[n-x]!}\right) \Big/ \frac{(\eta t)^n e^{-\eta t}}{n!}$$

$$= \frac{n!}{x!(n-x)!} \cdot \frac{\cancel{\eta^x}\cancel{\eta^n}\cancel{\eta^{-x}}}{\cancel{\eta^n}} \cdot \frac{\cancel{e^{-\eta s}}\cancel{e^{-\eta t}}\cancel{e^{\eta s}}}{\cancel{e^{-\eta t}}} \cdot \frac{s^x(t-s)^{n-x}}{t^x t^{n-x}}$$

$$= \frac{n!}{x!(n-x)!}\left(\frac{s}{t}\right)^n\left(1-\frac{s}{t}\right)^{n-x} \Rightarrow \boxed{\text{Hence } X \sim \text{Bin}(n, \tfrac{s}{t})}$$