Contents

Prob: Set theory

Basic

- **Set** is a collection of well-defined objects (called elements or members): x A.
- Set can be represented by a list, i.e. A = { -2, 2} , or by a set-builder, i.e. A = { x | $x^2 - 4 = 0$} .
- The **cardinality** of the set is the number of elements in that set # (A), a set can be finite set or infinite set, a set can be countable or uncountable.
- A **subset** of a set: A B if any element of A is also an element of B; apparently A A.
- If A B and A B we have A B.
- Absolute **complement** of A: A^c = { x | x A}
- Relative complement of A with respect to B: B – A = { x | x B and x A}
- **Union**: A B = { x | x A or x B} ; this relation can be extended to more than two sets.
- **Intersection**: A B = { x | x A and x B} ; again, this can be extended.
- If A B = , A and B are **disjoint** sets.

Intermediate

- Three basic laws of sets and De Morgan's laws
  - **Commutative** A B = B A; and A B = B A
  - **Associative** A (B C) = (A B) C; and A (B C) = (A B) C
  - **Distributive** A (B C) = (A B) (A C); and A (B C) = (A B) (A C)
  - **De Morgan** $(A\ B)^c = A^c\ B^c$; and $(A\ B)^c = A^c\ B^c$
    (De Morgan's law can be extended to more than two sets)
- **Inclusion-Exclusion principles**: suppose that A and B are finite sets, then:
  - If A B, # (A) ≤ # (B)
  - # (A B) = # (A) + # (B) - # (A B)
  - Extension: # (A B C) = # (A) + # (B) + # (C) - # (A B) - # (B C) - # (C A) + # (A B C)
- Cartesian product
  - A **Cartesian product** of two sets A, B is the set A × B = { (a,b) | a A and b B}
  - # (A × B) = # (A) # (B)

Prob: Permutation and Combination

- Fundamental **principle of counting**: if a choice consists of k steps, each step i can be made in n_i ways, the whole choice can be made in n_1 n_2 n_k ways.
- Permutation: a permutation is an **ordered arrangement** of objects. The number of permutation of n objects is n (n-1) (n-2) 1 = n! (0! = 1)
- Permutations of a set of distinct objects taken from a larger set: suppose we have n items, the number of ordered arrangement of k items can we from these n items is $_{\mathbf{n}}\mathbf{P}_{\mathbf{k}}$ = n! / (n-k)!
- Combination is a possible selection of a certain number of objects taken from a group **without regard to order**. The number of k-element subset of an n-element set is $_{\mathbf{n}}\mathbf{C}_{\mathbf{k}}$ = $_{\mathbf{n}}\mathbf{P}_{\mathbf{k}}$ / k! = n! / (k!∗ (n-k)!). Alternative not. (n k)
- Properties:
  - Symmetry $C_k^n = C_{n-k}^n$
  - Pascal $C_k^{n+1} = C_{k-1}^n + C_k^n$
  - Binomial theorem $(x+y)^n = \sum_{k=0}^n C_k^n x^{n-k} y^k$ (proof by induction)

  Application: number of subsets of a set with n elements $\sum_{k=0}^n C_k^n = (1+1)^n = 2^n$

- Combination with **repetition**: count the way to select k objects from n different categories with repetition allowed pictorially convert to k ***stars*** and n-1 ***bars*** problem of insert k stars into k+n-1 position: C_k^{k+n-1}
- Permutation with **indistinguishable** objects (**partition**): suppose we have n objects where n_1 of them are identical, n_2 of them are identical... and n_k of them are identical and n_1 + n_2 + ... + n_k = n. The number of permutation of n objects are then $\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$.

Prob: Probability

Basic

- A (random) **experiment** is a process whose **outcomes** cannot be predicted with certainty.
- The **sample space** S of an experiment is the set of all possible outcomes.
- An **event** E is a subset of the sample space.
- **Probability** is the measure of occurrence of an event.
  - Experimental probability uses the relative frequency & the law of large numbers (n is # of experiments)

$$P(E) = \lim_{n\to\infty} \frac{\#(E)}{n}$$

  - Theoretical or classical probability applies **only when all possible outcomes are equally likely**

$$P\left(E\right)=\frac{\#\left(E\right)}{\#\left(S\right)}$$

- Any function P satisfies the following axioms (**Kolmogorov** axioms) can be a **probability** measure.
  - For any event E, $0\leq P\left(E\right)\leq 1$. $P\left(S\right)=1$
  - For any sequence of mutually exclusive events { $E_{n\geq 1}$} and Ei Ej = for i j, we have

$$P\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}P\left(E_n\right)$$

- If P is a probability measure, then
  - If { E1, E2, . . . , En} is a finite set of mutually exclusive events then

$$P\left(\bigcup_{k=1}^{n}E_k\right)=\sum_{n=1}^{\infty}P\left(E_n\right)$$

  - P(E^c) = 1 – P(E)
  - If events A and B are mutually exclusive (sets A and B are disjoint): P(A B) = P() = 0
  - If A B then P(A) ≤ P(B)
  - With events A and B: P(A B) = P(A) + P(B) – P(A B)
    Extension: P(A B C) = P(A) + P(B) + P(C) – P(A B) – P(B C) – P(C A) + P(A B C)
    Extension: A = A (B B^c) = (A B) (A B^c) P(A) = P(A B) + P(A B^c)

Conditional probability and Bayes's formula

- Conditional probability:
  - Basic: if the occurrence of event A depends on the occurrence of B then the probability of A given B provided that P(B) 0, denoted as P(A| B), is:

$$P\left(A|B\right)=\frac{\#\left(A\cap B\right)}{\#\left(B\right)}=\frac{P\left(A\cap B\right)}{P\left(B\right)}$$

    Extension: P(A) = P(A| B)P(B) + P(A| B^c)P(B^c)
    Extension: P(B^c| A) = 1 – P(B| A)
  - Generalization of the basic formula

$$P\left(A_1\cap A_2\cap A_3\ldots\right)=P\left(A_1\right)+P\left(A_2|A_1\right)+P\left(A_3|A_1\cap A_2\right)+\ldots$$

- Bayes's formula
  - **Theorem**: since $P\left(A\cap B\right)=P\left(B\cap A\right)$ hence $P\left(A|B\right)\cdot P\left(B\right)=P\left(B|A\right)\cdot P\left(A\right)$, we have:

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

  - Extension: $P\left(B|A\right)=\frac{P(A\cap B)}{P(A\cap B)+P(A\cap B^c)}=\frac{P(A|B)\cdot P(B)}{P(A|B)\cdot P(B)+P(A|B^c)\cdot P(B^c)}$

- **Law of Total probability**: suppose that the sample space S is the union of mutually exclusive events $H_1, H_2, \ldots, H_n$ with $P(H_i) > 0$. Then for any event A we have:

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + \ldots = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + \ldots$$

- **Generalization of Bayes's rule**

$$P(H_i|A) = \frac{P(A|H_i) \cdot P(H_i)}{P(A)}$$

in which $P(A) = P(H_1)P(A|\ H_1) + P(H_2)P(A|\ H_2) + \ldots + P(H_n)P(A|\ H_n)$

- **Prior probability**: the **unconditional probability** P(A) is the probability of the event A prior to introducing new events that may affect A.
- **Posterior probability**: when the occurrence of an event B will affect event A, the **conditional probability** P(A| B) is known as the posterior probability of A.

• Independent events

- In term of conditional probability, two events A and B are said to be **independent** if and only if:

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

- General, A and B are independent events if and only if:

$$P(A \cap B) = P(A) \cdot P(B)$$

- If A and B are independent, then so A and B^c; A^c and B^c.
- Events $A_1, A_2, \ldots, A_n$ are said to be independent if for any $1 \le i_1 < i_2 < \ldots < i_k \le n$ we have:

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots \cdot P(A_{ik})$$

- Conditionally independent: two events A and B are said to be conditionally independent, given another event C with $P(C) > 0$, if:

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

If $P(B\ C) > 0$, conditional independence is equivalent to:

$$P(A|B \cap C) = P(A|C)$$

• Odds and conditional probability

- If E is the event of all **favorable outcomes,** then E^c is the event of **unfavorable outcomes**. Hence:

$$\text{odds in favor} = \frac{\#(E)}{\#(E^c)}; \quad \text{odds against} = \frac{\#(E^c)}{\#(E)}$$

Consequently:

$$P(E) = \frac{\#(E)}{\#(E) + \#(E^c)}; \quad \mathrm{P}(E^c) = \frac{\#(E^c)}{\#(E) + \#(E^c)}$$

$$\text{odds in favor} = \frac{P(E)}{1 - P(E)}; \quad \text{odds against} = \frac{1 - P(E)}{P(E)}$$

Prob: Random variables

Basis

- **Random variable**: is a function with domain the sample space and range a subset of real numbers, notation X(s) = x means x is the value associated with the outcome s by the random variable X.
- Notation:
  - X, Y, Z are random variables; x, y, z are possible values that the X, Y, Z random variables can take;
  - The statement "X = x" defines an event consisting of all outcomes with X-measurement equal to x;
  - P(X = x) is then the probability of the event X = x.
- Type of random variables:
  - **Discrete**: a random variable whose range is either finite or countably infinite;
  - **Continuous**: a random variable whose range is an interval in ;
  - **Mixed**: partially discrete and partially continuous.

Properties

- All random variables (discrete, continuous, mixed) have a distribution function or a cumulative distribution function (cdf). If X is a random variable then the cumulative distribution function, is the function:

  $$F\left(x\right) = P\left(X \leq x\right)$$

  and, hence:

  $$P\left(X > a\right) = 1 - F\left(a\right)$$

  Any non-negative function satisfying the three following propositions can be a cdf:

  - If a ≤ b then F(a) ≤ F(b)
  - $\lim_{a \to +b} F\left(a\right) = F\left(b\right)$
  - $\lim_{a \to -\infty} F\left(a\right) = 0;\ \ \lim_{a \to \infty} F\left(a\right) = 1$

    Further properties: if a < b then

  - P(a < X ≤ b) = F(b) - F(a)
  - P(a ≤ X < b) = F(b⁻) - F(a⁻)
  - P(a ≤ X ≤ b) = F(b) - F(a⁻) P(X = a) = F(a) - F(a⁻)
  - P(a < X < b) = F(b⁻) - F(a)

    Type of graphs:

  - For a continuous distribution, the graph of cdf is continuous non-decreasing curve.

- For a discrete distribution, the graph of its cdf consists of a series of horizontal lines with jumps between them.
- A cdf of a mixed distribution contains both jumps and pieces of continuous increasing curves.

- Survival distribution function (sdf), a.k.a. reliability function, gives the probability that an object of interest will survive beyond any given specified time.

$$S(x) = P(X > x) = 1 - P(X \leq x) = 1 - F(x)$$

An example for the application of the survival distribution function: if a < b then

$$P(a < X \leq b) = F(b) - F(a) = S(a) - (b)$$

Noted: for a continuous random variable, P(X = a) = F(a) - F(a$^-$) = 0 P(X $\leq$ a) = P(X < a) + P(X = a) = P(X < a); finally we have:

$$P(X < a) = P(X \leq a) = F(a) = 1 - S(a)$$

Prob: (single) Discrete random variables

Basis

- **Discrete random variable** is a random variable whose range is either finite or countably infinite.
- **Probability mass function** (**pmf**) or **probability distribution** of a discrete random variable X:
  - p(x) = P(X = x): pmf gives the probability that a discrete random variable is **exactly equal** to some value.
  - pmf can be an equation, a table or a graph.
- **Cumulative distribution function** (**cdf**) is a function giving the probability that the random variable X is **less than or equal to** x, for every value of x. For a discrete random variable:
  - Definition: cdf is found by **summing up** the probabilities:

$$F(a) = P(X \leq a) = \sum_{x \leq a} p(x)$$

  - Extension: if the range of a discrete random variable X consists of the value $x_1 < x_2 < \ldots < x_n$ then: (1) $p(x_1) = F(x_1)$, and (2) $p(x_i) = F(x_i) - F(x_{i-1})$ for i = 2, 3, . . . , n.
- **Expected value** (**mean value**) of a discrete random variable: let the range of a discrete random variable X be a sequence of number $x_1, x_2, \ldots, x_k$ and let p(x) be the corresponding probability mass function, then the expected value of X is:

$$E(X) = \sum_{i=1}^{k} x_i p(x_i)$$

The expected value (or mean) is related to the physical property of **center of mass**.

- **Expected value** of a function of a discrete random variable
  - Definition: if X is a discrete random variable with range D and pmf p(x), then the expected value of the random variable g(X) is:

$$E\left(g\left(X\right)\right)=\sum_{x\in D}g\left(x\right)\cdot p\left(x\right)$$

where g: is a real function.

  - Extension: if X is a discrete random variable, then for any constants a and b, we have:

$$E\left(aX+b\right)=a\cdot E\left(X\right)+b$$

$$E\left(aX^{2}+bX+c\right)=a\cdot E\left(X^{2}\right)+b\cdot E\left(X\right)+c$$

  - Extension: if $g\left(x\right)=x^{n}$ then we call $E\left(x^{n}\right)=\sum_{x}x^{n}\cdot p\left(x\right)$ the $\mathrm{n^{th}}$-moment about the origin of X, or the $\mathrm{n^{th}}$-raw moment. Apparently, $E\left(X\right)$ is the first moment of X.

- **Variance** and **standard deviation** of a discrete random variable
  - Meaning: standard deviation measures how **spread out** the **pmf** is about its **expected value**
  - **Variance**: the expected square distance between the random variable and its mean
  - **Standard deviation**: the positive square root of the variance

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

  Extension:

$$\mathrm{Var}\left(X\right)=E\left(X\right)+E\left(X\cdot\left(X-1\right)\right)-\left(E\left(X\right)\right)^{2}$$

$$\mathrm{Var}\left(aX+b\right)=a^{2}\cdot\mathrm{Var}\left(X\right)$$

  - **Coefficient of variation** of a random variable X is defined as:

$$\mathrm{CV}\left(X\right)=\frac{\sigma_{X}}{E\left(X\right)}$$

Uniform discrete random variable

- **[Commonly used]** Uniform discrete random variable
  - Definition: let X be a discrete random variable defined on a sample space S with n elements $\{\mathrm{x_1, x_2, \ldots, x_n}\}$, and the pmf of X is

$$p\left(x_{i}\right)=\frac{1}{n}\quad\text{for}\quad i=1,2,\ldots,n$$

  - Properties: the expected value of X is

$$E\left(X\right)=\sum_{i=1}^{k}x_{i}p\left(x_{i}\right)=\frac{x_{1}+x_{2}+\ldots+x_{n}}{n}$$

The second moment of X is

$$E\left(X^2\right)=\sum_{i=1}^{k}x_i^2p\left(x_i\right)=\frac{x_1^2+x_2^2+\ldots+x_n^2}{n}$$

The variance of X is

$$\operatorname{Var}\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=\frac{x_1^2+x_2^2+\ldots+x_n^2}{n}-\frac{\left(x_1+x_2+\ldots+x_n\right)^2}{n}$$

- In case if x = { a, a + h, a + 2∗ h, ..., a + (n-1)∗ h} then we have:

$$E\left(X\right)=a+\frac{h\cdot\left(n-1\right)}{2}$$

$$E\left(X^2\right)=a^2+ah\cdot\left(n-1\right)+h^2\frac{\left(n-1\right)\left(2n-1\right)}{6}$$

$$\operatorname{Var}\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=\frac{h^2\cdot\left(n^2-1\right)}{12}$$

- In case if x = { a, a + 1, a + 2, ..., b} then we have:

$$p\left(x_i\right)=\frac{1}{b-a+1}$$

$$E\left(X\right)=\frac{a+b}{2}$$

$$E\left(X^2\right)=ab+\frac{\left(b-a\right)\left(2b-2a+1\right)}{6}$$

$$\operatorname{Var}\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=\frac{\left(b-a+1\right)^2-1}{12}$$

$$M_X\left(t\right)=E\left(e^{tX}\right)=\sum_{x=0}^{b-a}e^{t(x+a)}\cdot p\left(x+a\right)=\frac{e^{ta}}{b-a+1}\left(\frac{1-e^{t(b-a+1)}}{1-e^t}\right)$$

Binomial random variable

- **[Commonly used]** Binomial random variable
  - Background: a **Bernoulli trial** is an experiment with **exactly two outcomes: success and failure** with the corresponding **probability of p and q**. Moreover: 0 < p, q < 1 and **p + q = 1** (e.g. flips of a coin). A **Bernoulli experiment** is a sequence of independent Bernoulli trials.

    Let X represent the **number of success** that occur in **n independent Bernoulli trials**, then X is said to be a **binomial random variable** with parameters **(n,p)**. If n = 1 then X is a **Bernoulli random variable**. The main problem of a binomial experiment is to find the **probability of r success out of n trials**.

- The probability of having r successes out of n trials **in any order** is given by the **binomial mass function**:

$$p(r) = P(X = r) = \binom{n}{r} p^r q^{n-r}$$

Note that as q = 1 − p:

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

$$p(r) = \frac{p}{1-p} \times \frac{n-r+1}{r} \times p(r-1)$$

From the above formula, it is proven that $p(r)$ has a unique global maximum at $r = \lfloor (n+1) \cdot p \rfloor$ (and $(n+1) \cdot p - 1$ if $(n+1) \cdot p$ is an integer).

- The cumulative distribution function is given by:

$$F(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & 0 \le x \le n \\ 1 & x > n \end{cases}$$

where $\lfloor x \rfloor$ is the floor function.

- The expected value of the binomial distribution is given by:

$$E(X) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np$$

- The variance of the binomial distribution is calculated from:

$$E(X \cdot (X-1)) = \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = n(n-1)p^2$$

$$\mathrm{Var}(X) = E(X) + E(X \cdot (X-1)) - (E(X))^2 = np(1-p)$$

- Moment generating functions:

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot p(x) = \left(q + pe^t\right)^n$$

$$M_X'(t) = npe^t \cdot \left(q + pe^t\right)^{n-1}$$

$$M_X''(t) = npe^t \cdot \left(q + pe^t\right)^{n-2} \cdot \left(q + npe^t\right)$$

Geometric random variable

- **[Commonly used]** Geometric random variable
  - Background: a geometric random variable models the number of successive independent Bernoulli trials that must be performed to obtain the "first" success.
  - Definition: let X be the **number of trials needed to achieve the first success**, then the pmf of X is:

$$p\left(n\right)=p\left(1-p\right)^{n-1}$$

    and X is called a **geometric random variable with parameter p**.
  - The expected value of the binomial distribution is given by:

$$E\left(X\right)=\sum_{n=1}^{\infty}np\left(1-p\right)^{n-1}=p\sum_{n=1}^{\infty}n\left(1-p\right)^{n-1}=\frac{1}{p}$$

  - Moment generating functions:

$$M_X\left(t\right)=E\left(e^{tX}\right)=\sum_{x=0}^{n}e^{tx}\cdot p\left(x\right)=\frac{pe^t}{1-e^t\left(1-p\right)}$$

  - The variance of the binomial distribution is calculated from:

$$E\left(X\cdot\left(X-1\right)\right)=\sum_{n=1}^{\infty}n\left(n-1\right)p\left(1-p\right)^{n-1}=2\frac{1-p}{p^2}$$

$$\mathrm{Var}\left(X\right)=E\left(X\right)+E\left(X\cdot\left(X-1\right)\right)-\left(E\left(X\right)\right)^2=\frac{1-p}{p^2}$$

  - Extension: observe that for k = 1, 2, ... we have

$$P\left(X\geq k\right)=\sum_{n=k}^{\infty}p\left(1-p\right)^{n-1}=\left(1-p\right)^{k-1}$$

$$P\left(X<k\right)=1-P\left(X\geq k+1\right)=1-\left(1-p\right)^{k}$$

    Then the cdf of X is given by:

$$F\left(x\right)=P\left(X\leq x\right)=\begin{cases}0 & x<1\\ 1-\left(1-p\right)^{\lfloor x\rfloor} & x\geq 1\end{cases}$$

Negative binomial random variable

- **[Commonly used]** Negative binomial random variable
  - Background: a negative binomial random variable models the number of successive independent Bernoulli trials that must be performed to obtain the r-th success.

- Definition: consider a Bernoulli experiment where a success occurs with probability p and a failure occurs with probability q. Let X be the **number of trials needed to achieve the r-th success**, X is called a **negative binomial distribution with parameters r and p**.
  By using courting law, the pmf of X can be computed as:

$$p(n) = P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

- Extension: the negative binomial distribution is also defined in terms of the random variable Y: **number of failures before the r-th success**. This formulation is statistically equivalent to the one given in term of X since Y = X − r:

$$P(Y = y) = \binom{y+r-1}{r-1} p^r (1-p)^y$$

  In this form, the number of success is fixed and we are interested in the number of failure before reaching the fixed number of success.
- The exacted value:

$$E(Y) = \sum_{y=0}^{\infty} y \binom{y+r-1}{r-1} p^r (1-p)^y = \frac{r(1-p)}{p}$$

$$E(X) = E(Y + r) = \frac{r}{p}$$

- The variance:

$$\text{Var}(X) = \text{Var}(Y) = \frac{r(1-p)}{p^2}$$

- Moment generating functions:

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot p(x) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$$

Hyper-geometric random variable

- **[Commonly used]** Hyper-geometric random variable
  - Background: consider a population of **N** objects where objects can be divided exactly into two types A and B. Suppose that the number of object A is **n** and the number of object B is N-n. A random sample of size **r** is equally likely selected without replacement. The **hyper-geometric random variable** X counts the total number **k** of objects of type A in the sample with k = 0, 1, . . . , r.
    By using courting law, the pmf of X can be computed as:

$$p(k) = P(X = k) = \frac{\binom{n}{k} \cdot \binom{r-k}{N-n}}{\binom{r}{N}}$$

- The exacted value:

$$E(X) = \sum_{k=0}^{r} kP(X = k) = \frac{nr}{N}$$

$$E\left(X^2\right) = \sum_{k=0}^{r} k^2 P(X = k) = \frac{nr}{N}\left[\frac{(n-1)(r-1)}{N-1} + 1 - \frac{nr}{N}\right]$$

- The variance:

$$\operatorname{Var}(X) = \frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$$

Poisson random variable

- **[Commonly used]** Poisson random variable
  - Background: the Poisson random variable is used to model the **number of occurrences of some phenomenon in a fixed interval of space or time**.
  - A random variable X is said to be a Poisson random variable with parameter > 0 if its pmf has the form:

$$p(k) = P(X(\omega) = k) = e^{-\lambda\omega}\frac{(\lambda\omega)^k}{k!}$$

with $\omega > 0$, k = 0, 1, 2, ... and λ indicates the average number of successes per unit time or space. Elaboration:

$$P(k \text{ events in time period}) = e^{-\frac{\text{events}}{\text{time}} \cdot \text{time period}} \cdot \frac{\left(\frac{\text{events}}{\text{time}} \cdot \text{time period}\right)^k}{k!}$$

Normally, the pmd is reduced to $p(k) = P(X = k) = e^{-\lambda}\frac{(\lambda)^k}{k!}$ with λ is simply the rate parameter.
Check the condition of pmf:

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda\omega}\sum_{k=0}^{\infty}\frac{(\lambda\omega)^k}{k!} = e^{-\lambda\omega}e^{\lambda\omega} = 1$$

  - The expected value of the Poisson distribution is given by:

$$E(X) = \sum_{k=1}^{\infty} ke^{-\lambda\omega}\frac{(\lambda\omega)^k}{k!} = \lambda\omega$$

  - The variance of the Poisson distribution is calculated from:

$$E(X \cdot (X-1)) = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda\omega}\frac{(\lambda\omega)^k}{k!} = (\lambda\omega)^2$$

$$\operatorname{Var}(X) = E(X) + E(X \cdot (X-1)) - (E(X))^2 = \lambda\omega$$

  - Moment generating functions:

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot p(x) = e^{\lambda\omega\left(e^t - 1\right)}$$

- In the simple form $p(k) = P(X = k) = e^{-\lambda}\frac{(\lambda)^k}{k!}$ to predict the probability of k events occur in a time interval, it is apparent that $E(X) = \mathrm{Var}(X) = \lambda$.
- Poisson approximation to the binomial: if the binomial probability of a binomial random variable X has n very large and p very small then for any fixed nonnegative integer k, the binomial probability:

$$p_X(k) = \binom{n}{k} p^k q^{n-k} \quad \text{for} \quad k = 0, 1, 2, \ldots, n$$

converges to the Poisson probability of a Poisson random variable Z with parameter $\lambda = np$

$$p_Z(k) = e^{-\lambda}\frac{(\lambda)^k}{k!} \quad \text{for} \quad k = 0, 1, 2, \ldots$$

when we take the limit as $n \to \infty$ and $p = \lambda/n$ while λ is kept constant.

Prob: (single) Continuous random variables

Basis

- **Continuous random variable** is a function with an **uncountable infinite range** such as an interval.
- A random variable X is continuous if there exists a non-negative function f (not necessarily continuous) defined for all real numbers and having the property that for any set B of real numbers we have:

$$P(X \in B) = \int_B f(x) \cdot dx$$

- The function f is called the **probability density function (pdf)** of the random variable X.
  - $P[X \in (-\infty, \infty)] = \int_{-\infty}^{\infty} f(x) \cdot dx$
  - $P(a \leq X \leq b) = \int_a^b f(x) \cdot dx$
  - $P(X = a) = \int_a^a f(x) \cdot dx = 0$
  - $P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b)$
  - $P(X \leq a) = P(X < a) = \int_{-\infty}^{a} f(x) \cdot dx = F(a)$
  - $P(X \geq a) = P(X > a) = \int_a^{\infty} f(x) \cdot dx = 1 - F(a)$
  - The pdf does not represent a probability but how likely X will be near a. Let $\epsilon > 0$, then the probability that X will be contained in an interval of length ϵ around the point a is approximately:

$$P(a \leq X \leq a + \epsilon) = F(a + \epsilon) - F(a) = \int_a^{a+\epsilon} f(t) \cdot dt \approx \epsilon \cdot f(a)$$

$$P\left(a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}\right) = P\left(a - \frac{\epsilon}{2} \leq X \leq a\right) + P\left(a \leq X \leq a + \frac{\epsilon}{2}\right) \approx 2 \cdot \frac{\epsilon}{2} \cdot f(a)$$

- The **cumulative distribution function** or simply the **distribution function** (cdf) F(t) of the random variable X is defined as:

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x) \cdot dx$$

  - Geometrically, F(t) is the area under the graph of f to the left of t.
  - $0 \leq F(t) \leq 1$; $F(t) \to 0$ when $t \to -\infty$; $F(t) \to 1$ when $t \to \infty$.
  - if $a \leq b$ then $F(a) \leq F(b)$; $P(a < X \leq b) = F(b) - F(a)$.
  - $F'(t) = f(t)$ whenver the derivative exists.

- The **expected value**:

  - Background: as with discrete random variables, the **expected value** of a continuous random variable is a **measure of location** which defines the **balancing point of the distribution**.
  - Definition: suppose that a continuous random variable X has a density function f(x) defined in [a,b] then we have:

$$E(X) = \int_{a}^{b} x \cdot f(x) \cdot dx;\ E\left(X^2\right) = \int_{a}^{b} x^2 \cdot f(x) \cdot dx$$

    If the domain of f consists of all real numbers, we define the expected value of X by the improper integral provided that that integral converges:

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx;\ ;\ E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 \cdot f(x) \cdot dx$$

    Sometimes for technical purposes, the expectation is expressed in terms of an integral of probabilities, especially when random variables X have **only positive values** (then the second term vanishes):

$$E(x) = \int_{0}^{\infty} P(X > y) \cdot dy - \int_{-\infty}^{0} P(X < y) \cdot dy$$

  - If X is a continuous random variable and g is a function defined for the values of X, then Y = g(X) is also a random variable. The expected value of g(X) is defined based on the pdf f(x), if existed, as follows:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) \cdot dx$$

    For any constant a and b: $E(aX + b) = a \cdot E(X) + b$

- The variance, a measure of the spread of the random variable about its expected value, is determined by:

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

$$\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$$

- The standard deviation X to be the square root of the variance.

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

- Coefficient of variation of a random variable X is defined as:

$$\mathrm{CV}(X) = \frac{\sigma_X}{E(X)}$$

pdf and cdf of a function

- pdf and cdf of a function of a continuous random variable
  - Let X be a continuous random variable with pdf f(x)
  - Let g(x) be a monotone and differentiable function of x; suppose that $g^{-1}(Y) = X$, this definition leads to:

$$P(Y \leq y) = P(g(X) \leq y) = P\left(X \leq g^{-1}(y)\right)$$

  - Then the random variable $Y = g(X)$ has a cdf and pdf given by:

$$F_Y(y) = F_X\left(g^{-1}(y)\right)$$

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \times \left|\frac{d}{dy} g^{-1}(y)\right|$$

Continuous uniform distribution function

- **[Commonly used]** Continuous uniform distribution function
  - Definition: a continuous random variable X is said to be uniformly distributed over the interval a ≤ x ≤ b if its pdf is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

  - The cdf is given by:

$$F(x) = \int_{-\infty}^{x} f(t) \cdot dt = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

  - The expected value of X and the second moment is:

$$E(X) = \int_a^b x \cdot f(x) \cdot dx = \frac{a+b}{2}$$

$$E\left(X^2\right) = \int_a^b x^2 \cdot f(x) \cdot dx = \frac{a^2+b^2+ab}{3}$$

  - The variance of X is:

$$\mathrm{Var}(X) = E\left(X^2\right) - [E(X)]^2 = \frac{(b-a)^2}{12}$$

Normal random variable

- **[Commonly used]** Normal random variable
  - Definition: a normal random variable with parameters μ and σ^2 has a pdf as follows

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

    The distribution is symmetric about the point μ and is a foundation for the ***Central Limit Theorem***.
    If X is a normal distribution with (μ, σ^2) then Y = aX+b is a normal distribution with $(a\mu + b, a^2\sigma^2)$
    If X is a normal distribution then $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.
  - **Standardization**: Let $Z = \frac{X-\mu}{\sigma}$ be the standard normal distribution then $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.
    The cdf of a standard normal distribution is denoted by:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \cdot dt$$

    It can be proven that $\Phi(x) = 1 - \Phi(-x)$ for $-\infty < x < \infty$ which implies P(Z ≤ -x) = P(Z > x).
    It can be seen that $\Phi(x)$ is the area under the standard curve to the left of x. Normally, the values of $\Phi(x)$ for x ≥ 0 are pre-computed. For x < 0, the above formulation can be used.
    The probabilities involving normal random variables can be reduced to the ones involving standard normal variable:

$$P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

    Extension:

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

  - **Normal approximation to the binomial distribution**: when the number of trials in a binomial distribution is very large, evaluation of $p(r) = \binom{n}{r} p^r q^{n-r}$ becomes tedious. Then, the normal distribution is employed to approximate the binomial distribution.
    Let X denote the number of successes that occur with n independent Bernoulli trials, each with probability p of success, hence X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. It is proven that:

$$\lim_{n\to\infty} \left(\frac{X - np}{\sqrt{np(1-p)}}\right) = \frac{N - np}{\sqrt{np(1-p)}}$$

    where N is the normal random variable with $\mu = np$ and $\sigma^2 = np(1-p)$.
    A rule-of-thumb for the normal distribution to be a good approximation to the binomial distribution is to have np > 5 and nq > 5.
    Continuity correction is used when approximating a discrete random variable with a continuous random variable. For example:

$$P\left(a<X<b\right)=P\left(a+1\leq X\leq b-1\right)\approx P\left(a+0.5\leq N\leq b-0.5\right)$$

and then:

$$P\left(a+0.5\leq N\leq b-0.5\right)=P\left(\frac{a+0.5-\mu}{\sigma}\leq\frac{N-\mu}{\sigma}\leq\frac{b-0.5-\mu}{\sigma}\right)=P\left(a^{*}\leq Z\leq b^{*}\right)$$

and finally:

$$P\left(a^{*}\leq Z\leq b^{*}\right)=\Phi\left(b^{*}\right)-\Phi\left(a^{*}\right)$$

Exponential random variable

- **[Commonly used]** Exponential random variable
  - Definition: an exponential random variable with parameters $\lambda>0$ (the rate parameter) has a pdf as follows:

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

    The exponential random variables are often used to model arrival times, waiting times or equipment failure times. The exponential distribution deals with the time between occurrences of successive events in a continuous time flow.
  - The expected value and the second moment of X can be found as:

$$E\left(X\right)=\int_{0}^{\infty}x\lambda e^{-\lambda x}\cdot dx=\frac{1}{\lambda};\quad \mathrm{E}\left(X^{2}\right)=\int_{0}^{\infty}x^{2}\lambda e^{-\lambda x}\cdot dx=\frac{2}{\lambda^{2}}$$

  - Thus, the variance is calculated as:

$$\mathrm{Var}\left(X\right)=E\left(X^{2}\right)-\left[E\left(X\right)\right]^{2}=\frac{1}{\lambda^{2}}$$

  - The cdf of an exponential random variable X is given by:

$$F\left(x\right)=P\left(X\leq x\right)=\int_{0}^{\infty}\lambda e^{-\lambda u}\cdot du=1-e^{-\lambda x}$$

    for $x\geq 0$ and 0 otherwise.
  - The most important property of the exponential distribution is known as the memoryless property:

$$P\left(X>s+t|X>s\right)=P\left(X>t\right)$$

    for $s,t\geq 0$. So, the exponential distribution forgets that it is larger than s.
  - Relationship between the Poisson random variable and the exponential random variable through example: Poisson-discrete to predict number of phone calls in an hour; exponential-continuous to predict the time before a new phone call comes.

$$N\sim\text{Poisson}\left(\lambda\right)\leftrightarrow X\sim\text{Exponential}\left(\beta=\frac{1}{\lambda}\right)$$

Gamma distribution

- **[Commonly used]** Gamma distribution
  - The Gamma function is defined as:

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} \cdot dy \quad \text{with } \alpha > 0$$

    For $\alpha > 1$, we can prove $\Gamma(\alpha) = \alpha \cdot \Gamma(\alpha - 1)$. Hence, if n is a positive integer $\Gamma(n) = (n-1)!$
  - A Gamma random variable with parameters $(\alpha > 0, \lambda > 0)$ has a pdf defined as follows:

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

    α is called the shape parameter because changing α changes the sahpe of the density function. λ is called the scale parameter because it rescales the density function without changing its shape. If X is a Gamma distribution with parameters (α, λ) then cX is also a Gamma distribution with parameters $(\alpha, \lambda/c)$ where c $>$ 0 is a constant. The Gamma distribution is skewed right.

    The Gamma distribution can be used to model a number of physical quantities such as service time, lifetimes of objects or repair times.
  - The cdf of the Gamma distribution is:

$$F(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^x e^{-\lambda y} y^{\alpha-1} \cdot dy$$

  - The expected value and the second moment of X can be found as:

$$E(X) = \frac{1}{\Gamma(\alpha)} \int_0^\infty x \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} \cdot dx = \frac{\alpha}{\lambda};$$

$$E\left(X^2\right) = \frac{1}{\Gamma(\alpha)} \int_0^\infty x^2 \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} \cdot dx = \frac{\alpha(\alpha+1)}{\lambda^2}$$

  - Thus, the variance is calculated as:

$$\mathrm{Var}(X) = E\left(X^2\right) - [E(X)]^2 = \frac{\alpha}{\lambda^2}$$

  - It can be seen that when $(\alpha, \lambda) = (1, \lambda)$ the Gamma distribution becomes the exponential distribution since $\Gamma(1) = 1$.

Prob: Relationship between random variables

Basis

- Jointly distributed random variables:
  - Suppose that X and Y are two random variables defined on the sample space S.
  - The **joint cumulative distribution function** of X and Y is defined as:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = P(\{e \in S : X(e) \leq x \text{ and } Y(e) \leq y)$$

- The **individual cdf** (the **marginal distribution**) are obtained from the joint cdf as follows:

$$F_X(x) = P(X \le x) = P(X \le x, Y < \infty) = F_{XY}(x, \infty)$$

  Similarly, and extension:

$$F_Y(y) = F_{XY}(\infty, y)$$

$$F_{XY}(\infty, \infty) = 1; \quad F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = 0$$

$$P(X > x, Y > y) = 1 - P(\{X \le x \cup \{Y \le y) = 1 - F_X(x) - F_Y(y) + F_{XY}(x, y)$$

  Also, if $a_1 < a_2$ and $b_1 < b_2$ then (similar to the concept of area):

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{XY}(a_2, b_2) - F_{XY}(a_1, b_2) - F_{XY}(a_2, b_1) + F_{XY}(a_1, b_1)$$

- Jointly pmf of discrete random variables:
  - If X and Y are both discrete random variables, the join probability mass function of X and Y is:

$$p_{XY}(x, y) = P(X = x, Y = y)$$

    Normally, the join pmf is given under a tabular format.
  - Then the marginal pmf of X and Y can be obtained as follows:

$$p_X(x) = \sum_{y: p_{XY}(x,y) > 0} p_{XY}(x, y)$$

$$p_Y(y) = \sum_{x: p_{XY}(x,y) > 0} p_{XY}(x, y)$$

    * Jointly pmf of continuous random variables:
      † Two random variables X and Y are jointly continuous if there exists a function $f_{XY}(x, y) \ge 0$ (the joint probability density function of X and Y) that for very subset C of R^2 we have:

$$P((X, Y) \in C) = \iint_{(X,Y) \in C} f_{XY}(x, y) \cdot dx \cdot dy$$

      † If $C = \{(x, y) : x \in A, y \in B$ then:

$$P(x \in A, y \in B) = \int_B \int_A f_{XY}(x, y) \cdot dx \cdot dy$$

      As a result of this equation, we can write:

$$F_{XY}(x, y) = P\left(X \in \left((-\infty, x], Y \in ((-\infty, y]\right) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u, v) \cdot du \cdot dv\right.$$

      And by using differentiation (whenever the partial derivatives exist):

$$f_{XY}(x, y) = \frac{\partial^2}{\partial y \partial x} F_{XY}(x, y)$$

† The marginal pmf of X and Y can be obtained as follows:

$$P(x \in A) = \int_A \int_{-\infty}^{\infty} f_{XY}(x, y) \cdot dy \cdot dx \quad \rightarrow f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) \cdot dy$$

$$P(y \in B) = \int_B \int_{-\infty}^{\infty} f_{XY}(x, y) \cdot dx \cdot dy \quad \rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) \cdot dx$$

* Joint pdfs and joint cdfs for three or more random variables are obtained as generalizations of the above definitions and conditions.

Independent random variables

- Independent random variables
  * Two random variables X, Y defined on the same sample space S are independent if and only if for any two sets of real numbers A and B we have:

$$P(x \in A, y \in B) = P(x \in A) \cdot P(y \in B)$$

  * In term of the pmf (for discrete) and pdfs (for continuous), the above theorem is stated as:

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y) \quad \text{and} \quad f_{XY}(x, y) = f_X(x) \cdot f_X(x)$$

  * In the more general words: X and Y are independent if and only if their joint pdf can be factored into two independent part:

$$p_{XY}(x, y) = g(x) \cdot h(y) \quad \text{or} \quad f_{XY}(x, y) = g(x) \cdot h(y)$$

  * Sum of two discrete independent random variables X, Y with pmf $p_X(x)$ and $p_Y(y)$, respectively: the pmf of the random variable X + Y (the convolution of p_X and p_Y) is given by:

$$p_{X+Y}(n) = P(X + Y = n) = \sum_{k=0}^{n} P(X = k) \cdot P(Y = n - k) = p_X(n) \cdot p_Y(n)$$

  Moreover, the mean and variance of the random variable X + Y are:

$$E(X + Y) = E(X) + E(Y)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

$$E(XY) = E(X) \cdot E(Y)$$

  * Sum of two continuous independent random variables X, Y with pdf $f_X(x)$ and $f_Y(y)$ defined for all x and y, respectively: the pdf of the random variable X + Y (the convolution of f_X and f_Y) is given by:

$$f_{X+Y}(n) = P(X + Y \leq n) = \int_{-\infty}^{\infty} f_X(n - y) \cdot f_Y(y) \cdot dy = (f_X \cdot f_Y)(n)$$

  Similarly, the mean and variance of the random variable X + Y are:

$$E(X + Y) = E(X) + E(Y)$$

$$\mathrm{Var}\left(X+Y\right)=\mathrm{Var}\left(X\right)+\mathrm{Var}\left(Y\right)$$

$$E\left(XY\right)=E\left(X\right)\cdot E\left(Y\right)$$

Conditional distribution

- Conditional distribution: recall the conditional probability of event E provided that the probability of event F is non zero:

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

  In a similar way, the conditional probability of two random variables can be derived.

- Discrete case:
  - Definition: the conditional probability mass function of X given that Y $=$ y and $p_Y\left(y\right)>0$ is:

    $$p_{X|Y}\left(x|y\right)=P\left(X=x|Y=y\right)=\frac{P\left(X=x,Y=y\right)}{P\left(Y=y\right)}=\frac{p_{XY}\left(x,y\right)}{p_Y\left(y\right)}$$

  - The conditional cumulative distribution of X given that Y $=$ y is defined by:

    $$F_{X|Y}\left(x|y\right)=P\left(X\leq x|Y=y\right)=\sum_{a\leq x}p_{X|Y}\left(a|y\right)$$

  - Extension: another practical use of the above formula is the following formula to calculate the marginal distribution of a variable X provided that the conditional distribution of X given Y $=$ y and the marginal distribution of Y are known:

    $$p_X\left(x\right)=\sum_y p_{XY}\left(x,y\right)=\sum_y p_{X|Y}\left(x|y\right)\cdot p_Y\left(y\right)$$

    And, of course when X and Y are independent, the conditional pmf and cdf are the same as the unconditional ones:

    $$p_{X|Y}\left(x|y\right)=p_X\left(x\right)\quad\text{and}\quad F_{X|Y}\left(x|y\right)=\sum_{a\leq x}p_X\left(a\right)$$

  - The conditional expectation of X given that Y $=$ y is given by:

    $$E\left(X|Y=y\right)=\sum_x x\cdot P\left(X=x|Y=y\right)=\sum_x x\cdot p_{X|Y}\left(x|y\right)$$

    In case when X and Y are independent, $p_{X|Y}\left(x|y\right)=p_X\left(x\right)$, hence $E\left(X|Y=y\right)=E\left(X\right)$.

- Continuous case:

- Definition: suppose that X and Y are two continuous random variables with joint density $f_{XY}(x, y)$, the conditional probability density function and conditional cumulative distribution function of X given that Y = y and $f_Y(y) > 0$ are:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{\partial}{\partial x} F_{X|Y}(x|y)$$

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{\infty} f_{X|Y}(t|y) \cdot dt$$

Again, X and Y are independent with $f_Y(y) > 0$ if and only if:

$$f_{X|Y}(x|y) = f_X(x)$$

- The conditional expectation of X given that Y = y is given by:

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) \cdot dx$$

Again, if X and Y are independent $f_{X|Y}(x|y) = f_X(x)$, hence $E(X|Y = y) = E(X)$.

- Conditional variance:
  - Conditional variance of X given that Y = y is defined as follows:

$$\text{Var}(X|Y = y) = E\left[(X - E(X|Y))^2 |Y = y\right]$$

  - Properties:

$$\text{Var}(X|Y) = E\left(X^2|Y\right) - (E(X|Y))^2$$

$$\text{E}(\text{Var}(X|Y)) = E\left(X^2\right) - E\left((E(X|Y))^2\right)$$

$$\text{Var}(E(X|Y)) = E\left((E(X|Y))^2\right) - (E(X))^2$$

$$\text{Var}(X) = \text{E}(\text{Var}(X|Y)) + \text{Var}(E(X|Y))$$

Joint pdf of functions

- Joint pdf of functions of two random variables
  - Let X and Y be jointly continuous random variables with joint probability density function $f_{XY}(x, y)$ with $A = \{(x, y) : f_{XY}(x, y) > 0$.
  - Let U = g_1(X,Y) and V = g_2(X,Y) and suppose that there are 1-1 transformations: u = g_1(x,y) and v = g_2(x,y) and x = x(u,v) and y = y(u,v).
  - Suppose that g1 and g2 have continuous partial derivatives at all point (x,y) A and the following Jacobian determinant is non-zero for all (x,y) A:

$$J(x, y) = |\begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix}| = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \neq 0$$

  - Then the random variable U, V are continuous random variables with joint density function given by:

$$f_{UV}(u, v) = f_{XY}\left(g_1^{-1}(u, v), g_2^{-1}(u, v)\right) |J\left(g_1^{-1}(u, v), g_2^{-1}(u, v)\right)|$$

Expected value of a function

- (unconditional) Expected value of a function of two random variables
  - Suppose that X and Y are two random variables taking values in S_X and S_Y, respectively. For a function g : S_X . S_Y R, the expected value of g(X,Y) is:

$$E\left(g\left(X,Y\right)\right)=\sum_{x\in S_X}\sum_{y\in S_Y}g\left(x,y\right)\cdot p_{XY}\left(x,y\right)$$

    if X, Y are discrete with the joint pmf $p_{XY}\left(x,y\right)$; and:

$$E\left(g\left(X,Y\right)\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g\left(x,y\right)\cdot f_{XY}\left(x,y\right)\cdot dxdy$$

    if X, Y are continuous with the joint pdf $f_{XY}\left(x,y\right)$.
  - Properties:
    Summation

$$E\left(X_1\pm X_2\pm\ldots\pm X_n\right)=E\left(X_1\right)\pm E\left(X_2\right)\pm\ldots\pm E\left(X_n\right)$$

    Multiplication

$$E\left(g\left(X\right)\cdot h\left(Y\right)\right)=E\left(g\left(X\right)\right)\cdot E\left(h\left(Y\right)\right)$$

    Markov's inequality

$$\text{If } X\geq 0\text{and } c>0 \text{ then } P\left(X\geq c\right)\leq\frac{E\left(X\right)}{c}$$

- (conditional) Expected value of a function of two random variables
  - Suppose that X and Y are two random variables taking values in S_X and S_Y, respectively. Let g(x) be any function, the conditional expected value of g given Y = y, in the discrete case is:

$$E\left(g\left(X\right)|Y=y\right)=\sum_{x\in S_X}g\left(x\right)\cdot p_{X|Y}\left(x|y\right)$$

    For the continuous case, the corresponding formula is:

$$E\left(g\left(X\right)|Y=y\right)=\int_{-\infty}^{\infty}g\left(x\right)\cdot f_{X|Y}\left(x|y\right)\cdot dx$$

Covariance

Covariance is a measure quantifying the strength of a relationship between two random variables.

- Covariance between two random variable X and Y is defined by:

$$\text{Cov}\left(X,Y\right)=E\left[\left(X-E\left(X\right)\right)\cdot\left(Y-E\left(Y\right)\right)\right]=E\left(XY\right)-E\left(X\right)\cdot E\left(Y\right)$$

  Recall that if X and Y are independent then $E\left(XY\right)=E\left(X\right)\cdot E\left(Y\right)$, hence $\text{Cov}\left(X,Y\right)=0$.

- Properties of covariance:

$$\mathrm{Cov}\left(X, Y\right) = \mathrm{Cov}\left(Y, X\right)$$

$$\mathrm{Cov}\left(aX, Y\right) = a \cdot \mathrm{Cov}\left(X, Y\right)$$

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}\left(X_i, Y_j\right)$$

$$\mathrm{Cov}\left(X, X\right) = \mathrm{Var}\left(X\right)$$

$$\mathrm{Var}\left(X + Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right) + 2 \cdot \mathrm{Cov}\left(X, Y\right)$$

Coefficient of correlation

- Coefficient of correlation, a measure quantifying the linear dependence of two random variables X and Y with $\mathrm{Var}\left(X\right) > 0$ and $\mathrm{Var}\left(Y\right) > 0$ or the degree of linearity between X and Y, is defined as:

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

- Elaboration:
  - A value of $\rho\left(X, Y\right)$ near $+1$ or -1 indicates a high degree of linearity between X and Y; whereas a value near 0 implies a lack of such linearity.
  - Positive sign means X and Y are positively correlated; negative sign means X and Y are negatively correlated; 0 means uncorrelated.
  - Correlation is a scaled version of covariance.
- Properties of coefficient of correlation:

$$-1 \leq \rho\left(X, Y\right) \leq 1$$

$$\rho\left(X, Y\right) = \pm 1 \text{ if and only if } Y = aX + b$$

Prob: Miscellaneous

Central tendency

There are several measures of central tendency:

- **Mean**: a measure of the center of the data (a measure of central tendency)

- **Standard deviation**: a measure of how data is scattered around the mean (a measure of dispersion)

The **median** of a set of data is **the number where half of the data falls below it**.

- Consider a discrete random variable X: the median is the first number M such that F(M) = P(X ≤ M) > 0.5. If there is M satisfying F(M) = 0.5 then the median is the average of M and the next value of X.
- Consider a continuous random variable X: the median is the number M such that P(X ≤ M) = P(X ≥ M) = 0.5; that is F(M) = 0.5.

The **mode** of a random variable is defined as **the value** that **maximizes the probability mass function** p(x) (discrete case) or **the probability density function** f(x) (continuous case).

- In the discrete case: the mode is the value that is most likely to be sampled, i.e. p(x) = maximum
- In the continuous case: the mode is where f(x) is at its peak, i.e. f(x) = maximum

Percentiles and quartiles:

- A percentile is the value of a variable below which a certain percent of observations falls.
- For a random variable X and 0 < p < 1, the p^{th}-percentile (or the p^{th}-quantile) is the number x satisfying

$$P\left(X < x\right) \leq p \leq P\left(X \leq x\right)$$

- For a continuous random variable, this is the solution of F(x) = p.
- Note that the 50^{th}-percentile is the median.

Moment generating functions

- For a random variable X and a positive integer n, $E\left(X^n\right)$ is called moments.
- Sometimes, the mean $E\left(X\right)$ and the variance $\mathrm{Var}\left(X\right)$, which both are functions of moments, are difficult to find.
- Special functions, called moment-generating functions, can sometimes help:
  - finding moments and functions of moments;
  - identifying the distribution of a random variable.
- The **moment generating function** of a random variable X, denoted by $M_X\left(t\right)$, is defined by:

$$M_X\left(t\right) = E\left(e^{tX}\right)$$

  Provided that the expectation exists for t in some neighborhood of 0.

  For a discrete random variable with a pmf $p_X(x)$:

$$M_X(t) = E\left(e^{tX}\right) = \sum_x e^{tx} \cdot p_X(x)$$

For a continuous random variable with a pdf f(x):

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \cdot dx$$

- Properties:

$$M_{aX+b}(t) = e^{bt} \cdot M_X(at)$$

$$E(X^n) = M_X^n(0) = \left(\frac{d^n}{dt^n} M_X(t)\,|_{t=0}\right.$$

- Distribution similarity: if random variables X and Y have moment generating function $M_X(t) = M_Y(t)$ then X and Y have the same distributions.
- Distribution of sums of independent random variables
    - Let X_1, X_2, ... , X_n be independent random variables
    - The moment generating function of Y = X_1 + X_2 + ... + X_n is given by:

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{X_1 t} e^{X_2 t} \dots e^{X_n t}\right) = \prod_{k=1}^{n} E\left(e^{X_k t}\right) = \prod_{k=1}^{n} M_{X_k}(t)$$

- For any random variables X_1, X_2, ... , X_n, the joint moment generating function is defined by:

$$M(t_1, t_2, \dots, t_n) = E\left(e^{X_1 t_1 + X_2 t_2 + \dots + X_n t_n}\right)$$

Central limit theorem

- Theorem: the sum of a large number of independent identically distributed random variables is well-approximated by a normal random variable.
    - Let X_1, X_2, ... , X_n be a sequence of independent and identically distributed random variables, each with mean μ and variance σ^2 . Then:

    $$P\left(\frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right) \le a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} \cdot dx$$

    as $n \to \infty$.
    - Elaboration: regardless of the underlying distribution of the variables X_i, so long as they are independent, the distribution of $\frac{\sqrt{n}}{\sigma}\left(\frac{X_1+X_2+\dots+X_n}{n} - \mu\right)$ converges to the same, normal, distribution. Hence, the sample mean can be approximated by a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.
    - Similarly, a sum of n independent and identically distributed random variables with common mean μ and variance σ^2 can be approximated by a normal distribution with mean $n\mu$ and variance $n\sigma^2$.

Common series

$$\sum_{x=1}^{n} x = \frac{n \cdot (n+1)}{2}$$

$$\sum_{x=1}^{n} x^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{6}$$

$$\sum_{x=1}^{n} x^3 = \frac{n^2 \cdot (n+1)^2}{4}$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \quad \text{for } -1 < x < 1$$

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots \quad \text{for } -1 < x < 1$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \quad \text{for } -1 < x < 1$$

Common integrals

$$\int_0^\infty t^n \cdot e^{-t} dt = n!$$

Stoch: Stochastic processes

Basic

- A stochastic process is a mathematical model of a probabilistic experiment that evolves in time and generates a sequence of numerical values.
- Each numerical value in the sequence is modelled by a random variable, hence a stochastic process is simply a (finite or infinite) sequence of random variables.
- In the study of a stochastic process, we tend to focus on:
  - the dependency in the sequence of values generated by the process (e.g. how future values depend on past values);
  - the long-term average, involving the entire sequence of generated values;
  - the likelihood or frequency of certain boundary events.
- The following major categories of stochastic processes will be considered in this note:
  - Arrival-type processes: processes have the character of an arrival such as job completions, customer purchases... We focus more on models in which the times between successive arrivals (the interarrival times) are independent random variables; (1) Bernoulli process: arrivals occur in discrete time and the interarrival times are geometrically distributed, and (2) Poisson process: arrivals occur in continuous time and the interarrival times are exponentially distributed.
  - Markov processes: processes evolve in time and the future evolution exhibits a probabilistic dependence on the past. We focus more on a special type of dependence: the next value depends on past values only through the current value.

Bernoulli stochastic process

- Informal definition: the Bernoulli process consists of a sequence of Bernoulli trials, where each trial produces a 1 (a success) with probability p and a 0 (a failure) with probability 1-p, independently of what happens in other trials.
- Formal definition: the Bernoulli process is a sequence X_1, X_2,... of independent Bernoulli random variable X_i as:
  - $P\left(X_i = 1\right) = P\left(\text{success at the } i\text{−thtrial}\right) = p$
  - $P\left(X_i = 0\right) = P\left(\text{failure at the } i\text{−thtrial}\right) = 1 - p$ for each i.
- Basic properties:
  - The **number S of successes in n independent trials** is a **binomial distribution** with parameters p and n. Its pmf, mean and variance are:

$$p_S\left(k\right) = \binom{n}{k} p^k \left(1-p\right)^{n-k} \quad \text{for} \quad k = 0,\ 1,\ 2, \ldots, n; E\left(S\right) = np; \quad \mathrm{Var}\left(S\right) = np\left(1-p\right)$$

- The **number T of trials up to (and including) the first success** is a **geometric distribution** with parameter p. Its pmd, mean and variance are:

$$p_T(t) = p(1-p)^{t-1} \quad \text{for} \quad t = 1, 2, \ldots; \;\; E(T) = \frac{1}{p}; \quad \text{Var}(T) = \frac{1-p}{p^2}$$

  - Advanced properties:
    - † Fresh-start: for any given time n, the sequence of random variables X_{n+1},X_{n+2},... (the future of the process) is also a Bernoulli process, and is independent from X_1, X_2,... , X_n (the past of the process).
    - † Memoryless: let recall that the time T until the first success is a geometric variable, suppose that we have already have n time steps without any success, then the number of future trials until the first success is still descried by the same geometric pmf:

$$P(T-n=t|T>n) = p(1-p)^{t-1} = P(T=t) \quad \text{for } t = 1, 2, \ldots$$

  - Interarrival times:
    - † In the Bernoulli process, let Y_k denote the time of the k-th success and let T_k denote the k-th interarrival time (note that T_k are independent geometric random variables with common parameter p).

$$T_1 = Y_1; \quad T_k = Y_k - Y_{k-1} \quad \text{for} \quad k = 2, 3, \ldots; \; Y_k = \sum_{i=1}^{k} T_k$$

    - † Mean and variance of Y_k are given by:

$$E(Y_k) = \sum_{i=1}^{k} E(T_k) = \frac{k}{p}; \quad \text{Var}(Y_k) = \sum_{i=1}^{k} \text{Var}(T_k) = \frac{k(1-p)}{p^2}$$

    - † The pmf of Y_k known as the Pascal pmf of order k is given by:

$$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k} \quad \text{for } t = k, k+1, \ldots$$

  - Splitting and merging of Bernoulli processes [?]

Poisson stochastic process

- Informal definition: the Poisson process can be viewed as a continuous-time analog of the Bernoulli process and applies to situations where there is no natural way of dividing time into discrete periods.
- Formal definition: an arrival process is called a Poisson process with rate λ if it has the following properties:
  - Time homogeneity: the probability $P(k, \tau)$ of k arrivals during an interval of length τ is the same for all intervals of the same length τ (in other word, they obey the same probability law).
  - Independence: the number of arrivals during a particular interval is independent of the history of arrivals outside this interval.
  - Small interval probabilities: the probability $P(k, \tau)$ satisfy

$$P(0, \tau) = 1 - \lambda\tau + O_0(\tau)$$

$$P(1, \tau) = \lambda\tau + O_1(\tau)$$

in which $O_i(\tau)$ are functions of τ that satisfy

$$\lim_{\tau \to 0} \frac{O_i(\tau)}{\tau} = 0$$

The $O_i(\tau)$ terms are meant to be negligible in comparison to τ when the interval length τ is very small. They can be though as the $O(\tau^2)$ terms in the Taylor series expansion of $P(k, \tau)$.

- The above requirements lead to:
  - The number N_τ of arrival in a Poisson process with rate λ over an interval of length τ is a Poisson distribution with parameter $\lambda\tau$. Its pmf, mean and variance are:

$$p_{N_\tau}(k) = P(k, \tau) = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!} \quad \text{for} \quad k = 0,\ 1,\ 2, \ldots; E(N_\tau) = \lambda\tau; \quad \text{Var}(N_\tau) = \lambda\tau$$

  - The time T until the first arrival is an exponential distribution with parameter λ . Its pdf, mean and variance are:

$$f_T(t) = \lambda e^{-\lambda t} \text{ for } \ t \geq 0; \quad E(T) = \frac{1}{\lambda}; \quad \text{Var}(T) = \frac{1}{\lambda^2}$$

Advanced properties:

- Independence of non-overlapping sets of times
- Fresh-start property: the portion of the Poisson process that starts at any particular time t $>$ 0 is a probabilistic replica of the Poisson process starting at time 0, and is independent of the portion of the process prior to time t.
- Memoryless: remaining time until the next arrival has the same exponential distribution. Mathematically, the exponential cdf of T $>$ t:

$$P(T > t) = e^{-\lambda t}$$

  For all positive scalars s:

$$P(T > t + s | T > t) = e^{-\lambda s}$$

Interarrival times:

- Similar to the Bernoulli process, let $\mathrm{Y_k}$ denote the time of the k-th success and let $\mathrm{T_k}$ denote the k-th interarrival time (note that $\mathrm{T_k}$ are independent geometric random variables with common parameter p).

$$T_1 = Y_1; \quad T_k = Y_k - Y_{k-1} \quad \text{for} \quad k = 2, 3, \ldots; \ Y_k = \sum_{i=1}^{k} T_k$$

- Mean and variance of $\mathrm{Y_k}$ are given by:

$$E(Y_k) = \sum_{i=1}^{k} E(T_k) = \frac{k}{\lambda}; \quad \text{Var}(Y_k) = \sum_{i=1}^{k} \text{Var}(T_k) = \frac{k}{\lambda^2}$$

- The pdf of Y_k known as the Erlang pdf of order k is given by:

$$f_{Y_k}(y) = e^{-\lambda y} \frac{(\lambda y)^{k-1}}{(k-1)!} \quad \text{for } t = k, k+1, \ldots$$

Splitting and merging of Poisson processes [?]

Markov chains

- The Bernoulli and Poisson processes studied in the preceding chapter are memoryless, in the sense that the future does not depend on the past. We consider now processes where the future **depends on** and can **be predicted** to some extent by what happened in the past.
- We focus on models where the effect of the past on the future is summarized by a **state**, which changes over time according to given **probabilities**.
- One of those kinds of models is **Markov chain**, which will be discussed in more details later.

Stoch: Discrete-time Markov chains

Discrete-time Markov chains

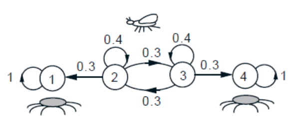
- At each time step n, the Markov chain has a **state**, denoted by Xn which belongs to a finite set S of possible states, called the **sate space**. Assumed that S = { 1, 2, . . . , m} .
- The Markov chain is described in terms of its **transition probabilities** p_{ij}: whenever the state happens to be i, there is probability p_{ij} that the next state is equal to j.

$$p_{ij} = P(X_{n+1} = j | X_n = i)\,; \quad i, j \in S$$

- Markov property: the key assumption of Markov processes is that the transition probabilities p_{ij} apply whenever state i is visited, no matter what happened in the past.

$$P(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i_n) = p_{ij}$$

  for all times n, all states $i, j \in S$, and all possible sequences $i_0, \ldots, i_{n-1}$ of earlier states.



The transition probabilities p_{ij} must be of course non-negative and sum to one; all of the element of a Markov chain model can be encoded in a transition probability matrix or illustrated in a transition probability graph:

$$\sum_{j=1}^{m} p_{ij} = 1 \text{ for all } i; \quad P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & p_{ij} & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

- Given a Markov chain model, we can compute the probability of any particular sequence of future states:

$$P\left(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\right) = P\left(X_0 = i_0\right) \cdot p_{i_0 i_1} \cdot p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

$$P\left(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n | X_0 = i_0\right) = p_{i_0 i_1} \cdot p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

Graphically, a state sequence is similar to a path in the transition probability graph; hence, the probability of such a path (given the initial state) is given by the product of the probabilities associated with the connection traversed by the path.

- The probability law of the state at some future time, conditioned on the current state called the **n-step transition probabilities** is defined by:

$$r_{ij}\left(n\right) = P\left(X_n = j | X_0 = i\right)$$

The n-step transition probabilities can be calculated using the following recursion, **Chapman-Kolmogorov**, equation:

$$r_{ij}\left(n\right) = \sum_{k=1}^{m} r_{ik}\left(n-1\right) \cdot p_{kj}; \quad \text{for } n > 1; \quad \text{starting with } r_{ij}\left(1\right) = p_{ij}$$

- In addition to the transition matrix, a Markov chain is also characterized by its initial probability distribution which can be represented by a vector μ with entries:

$$\mu_i = P\left(X_0 = i\right)$$

Using the Chapman-Kolmogorov, it is clear that:

$$P\left(X_0 = i\right) = \mu_i$$

$$P\left(X_1 = j\right) = \sum_{i=1}^{m} \mu_i \cdot p_{ij} = \left(\mu^T P\right)_j$$

$$P\left(X_2 = j\right) = \sum_{i=1}^{m} P\left(X_2 = j | X_0 = i\right) = \sum_{i=1}^{m} \mu_i \cdot p_{ij}^2 = \left(\mu^T P^2\right)_j$$

$$P\left(X_n = j\right) = \left(\mu^T P^n\right)_j$$

Theorem: Let { X_0, X_1, X_2,...} be a Markov chain with m$*$ m transition matrix P . If the probability distribution of X_0 is given by the 1$*$ m row vector μ^T , then the probability distribution of X_n is given by the 1$*$ m row vector $\mu^T P^n$ **.**

$$X_o \sim \mu^T \rightarrow X_n \sim \mu^T P^n$$

Classification of states

- We wish to classify the states of a Markov chain with a focus on the long-term frequency with which they are visited:
  - A state j is **accessible** from a state i if for some n, the n-stp transition probability $r_{ij}(n)$ is positive.
  - A state i is **recurrent** if for every j that is accessible from i, i also accessible from j. Let A(i) be the set of states that are accessible from i, for all j that belong to A(i), we have that i belongs to A(j). A recurrent state is also called **absorbing** in the sense that it is infinitely repeated once reached.
  - A state is **transient** if it is not recurrent. There are states j A(i) such that i is not accessible from j.

  - If i is a recurrent state, the set of states A(i) form a recurrent class meaning that states in A(i) are all accessible from each other and no state outside A(i) is accessible from them.
  - A Markov chain can be decomposed into one or more recurrent classes plus possibly some transient states.
- Periodicity: a recurrent class is said to be **periodic** if its states can be grouped in d $>$ 1 disjoint subsets so that all transitions from one subset lead to the next subset. A recurrent class that is not periodic is said to be **aperiodic**.

Steady-state behavior

- In Markov chain models, a **long-term state occupancy** behavior is normally of interested, that is in the **n-step transition probabilities** $r_{ij}(n)$ when n is very large. Sometimes, $r_{ij}(n)$ can converge to **steady-state** values that are independent of the initial state.
- Consider a Markov chain with a single recurrent class which is aperiodic. Then, the states j are associated with steady-state probabilities π_j that have the following properties:
  - $\lim_{n\to\infty} r_{ij}(n) = \pi_j$ for all i, j
  - π_j are the unique solution of the system of equations

$$\pi_j = \sum_{k=1}^{m} \pi_k p_{kj} \quad j = 1,\ 2, \ldots, m; \sum_{k=1}^{m} \pi_k = 1$$

  - $\pi_j = 0$ for all transient states j; $\pi_j > 0$ for all recurrent states j

    π_j form a probability distribution on the state space, called the stationary distribution of the chain. That means if $P(X_0 = j) = \pi_j$ with $j = 1, 2, \ldots, m$ then $P(X_n = j) = \pi_j$ for all n and j . Examples on p. 416-423 in Ref. [1] and p. 237-241 in Ref. [2].
- Long-term frequency interpretation

Recall that probabilities are often interpreted as relative frequencies in an infinitely long string of independent trials. The steady-state probabilities of a Markov chain admit a similar interpretation as expected state frequencies.

For a Markov chain with a single class that is aperiodic, the steady-state probabilities π_j satisfy:

$$\pi_j = \lim_{n \to \infty} \frac{v_{ij}(n)}{n}$$

where $v_{ij}(n)$ is the expected value of the number of visits to state j within the first n transitions, starting from state i. π_j is the long-term expected fraction of time that the state is equal to j.

The expected number of transitions in n transitions that take the state from j to k $q_{jk}(n)$ can be calculated as:

$$\lim_{n \to \infty} \frac{q_{jk}(n)}{n} = \pi_j p_{jk}$$

$\pi_j p_{jk}$ can be viewed as the long-term expected fraction of transitions that move the state from j to k.

Absorption probabilities and expected time to absorption

- Aaa

Random walk models

- Aaa

Stoch: Continuous-time Markov chains

Continuous-time Markov chains

- Aaa

Brownian motion

- Aaa

Birth-Death process

Aaa

References:

1) ***Marcel B. Finan*** Lecture Notes in Actuarial Mathematics – A Probability Course for the Actuaries

2) ***Dimitri P. Bertsekas*** & ***John N. Tsitsiklis*** Introduction to Probability – Lecture notes MIT 6.041-6.431

3) ***Anders Tolver*** An introduction to Markov chains

4) ***Rachel Fewster*** Stochastic processes – Stats 325

5) ***Andasari*** Stochastic modelling using Python

My experiences

Differential equations

- Boundary value problem (BVP): a differential equation together with a set of additional constraints, called the boundary conditions.
    - Dirichlet BC: a BC specifies the value of the function itself.
    - Neumann BC: a BC specifies the value of the normal derivative of the function.
    - Cauchy C: a BC has a form of a curve or surface that gives a value to the normal derivative and the function itself.
- Initial value problem (IVP): an ordinary differential equation together with a specified value, called the initial conditions at a given point in the domain of the solution.

  Partial differential equations (PDE)
    - Hyperbolic, e.g. wave propagation equation
    - Parabolic, e.g. diffusion equation
    - Elliptic, e.g. elasticity and Laplace's equations
    - Laplace's equation: a second-order PDE, eg $\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$

  Finite element method (IBVP)
    - provide the local approximation of the original complex PDE;
    - math term: construct an integral of the inner product of the residual and the weight functions, then set the integral to be zero;
    - simple term: fitting trial functions into the PDE, the residual is the error caused by the trial functions, the weight functions are polynomial approximation functions that project the residual;
    - steady state problems -> a set of algebraic equations -> solved by numerical linear algebra methods;
        * linear systems of equations: Gaussian elimination, iterative method
        * nonlinear systems of equations: Newton-Raphson method
    - transient problems -> a set of ordinary differential equations -> solved by numerical integration using standard techniques
        * Newmark
        * Backward/Forward Euler
        * Crank-Nicolson
        * Runge-Kutta.
- Step to solve:
    - strong form (PDE) -> (integration by part) -> weak form
    - weak form -> discretization in a finite-dimensional space into piecewise polynomial functions
    - numerical approximation (integration) – Gauss quadrature: implemented on a computer
- Generalized FEM uses local spaces consisting of functions, not necessarily polynomial. Partition of Unity method is used to bond these local spaces together to form the approximating subspace.

- XFEM: GFEM + PUM + enriching the solution space with discontinuous functions

Constitutive relations define the dependence of the stress tensor in a body on kinematic variables such as strain tensor or the rate-of-deformation tensor

- the stress tensor is a mapping of a force vector to a direction vector or a normal vector of a surface that the force is acting on;
- similar to the strain tensor.

Smoothed particle hydrodynamics (SPH)

- an introduction of a kernel function to represent an arbitrary field as a convolution regarding to that kernel function -> any physical quantity of any particle can be obtained by taking the volume integral of the kernel function multiplied by the local value of that quantities of other particles that lie within the range of the kernel (the characteristic radius)
- the convolution integral is approximated/discretized using a Riemann summation over the particles
- the kernel functions can be Gaussian function, quintic spline...