Customized Tufte Handout Template

Matthew He

September 18, 2024

This is a demo for a customized Tufte Handout stytle latex template, inspired by Anna Brandenberger, a former McGill undergraduate student.
Just like the original Tufte Handout template, it integrates the core feature of marginnote. Just like a comment section, marginnote is extremely handy when you want to present the flow of a math proof or some connection between concepts, etc.
It also preserve the ability for citation management, with the bibtex package.

Based on the original version, this template adds more customizable aesthetic components in, including section/subsection/paragraph title, as well as Definition/Theorem/Lemma blocks. These aesthetic components are heavily inspired by Anna Brandenberger.

1 Time-Homogeneous Markov Chains

1.1 Regular Chains

Definition 1. *Regularity A transition matrix P is regular if and only if there exists $n \in \mathbb{N}$ such that every entry of P^n is positive.*

Theorem 1. *A stochastic matrix P has an eigenvalue $\lambda^* = 1$. All other eigenvalues λ of P satisfy $|\lambda| \leq 1$, with strict inequality if P is regular.*

Proof. Let P be a $k \times k$ stochastic matrix. The rows of P sum to 1 by definition, so $P \cdot \mathbf{1} = \mathbf{1}$ and $\lambda^* = 1$ is a right eigenvalue of P. Suppose that $\bar{z}$ is the eigenvector corresponding to any other eigenvalue λ of P. Let $|z_m| = \max_{1 \leq i \leq k} |z_i|$ be the component of $\bar{z}$ of maximum absolute value. Then,

$$|\lambda| \cdot |z_m| = |\lambda z_m| = |(P \cdot \bar{z})_m| = \left| \sum_{i=1}^{k} P_{mi} z_i \right| \leq |z_m| \sum_{i=1}^{k} P_{mi} = |z_m|$$

and consequently $|\lambda| \leq 1$.

Assume that P is regular. Then $P^n > 0$ for some $n > 0$. P is a stochastic matrix, and it was shown above that P has an eigenvalue $\lambda^* = 1$. Moreover, all other eigenvalues λ of P satisfy $|\lambda| \leq 1$. We want to show that the inequality is strict. If λ is an eigenvalue of P, then λ^n is an eigenvalue of P^n. Let $\bar{x}$ be its corresponding eigenvalue, with $|x_m| = \max_{1 \leq i \leq k} |x_i|$. Then,

$$|\lambda|^n \cdot |x_m| = |(P^n \cdot \bar{x})_m| = \left| \sum_{i=1}^{k} P^n_{mi} x_i \right| \leq |x_m| \sum_{i=1}^{k} P^n_{mi} = |x_m|$$

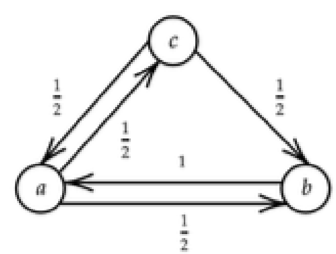
Since the entries of P^n are positive, the last inequality only holds if $|x_1| = \cdots = |x_k|$. Similarly, the first inequality only holds if

Example of Regularity:
The following matrix is regular,

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

because P^4 is positive,

$$P^4 = \begin{pmatrix} 9/16 & 5/16 & 1/8 \\ 1/4 & 3/8 & 3/8 \\ 1/2 & 5/16 & 3/16 \end{pmatrix}$$



$x_1 = \cdots = x_m$. But the constant vector whose components are the same is an eigenvector associated with the eigenvalue 1. Hence, if $\lambda \neq 1$, one of the inequalities must be strict. Thus, $|\lambda| < 1$.

2 Experiments with random outcomes

2.1 Ingredients of a Probability Model

Definition 2. *These are ingredients of a probability model.*

- *The **sample space** Ω is the set of all possible outcomes of the experiment. Elements of Ω are called **sample points** and typically denoted by ω.*
- *Subsets of Ω are called **events**. The collection of events in Ω is denoted by $\mathcal{F}$.*
- *The **probability measure** (also called **probability distribution** or simply **probability**) P is a function from $\mathcal{F}$ into the real numbers. Each event A has a probability of P(A), and P satisfies the axioms on the right.*

*The triple $(\Omega, \mathcal{F}, P)$ is called a **probablity space**. Every mathematically precise model of a random experiment or collection of experiments must be of this kind.*

Kolmogorov Axioms (early 1930s)

1. $0 \leq P(A) \leq 1$ for event A.
2. $P(\Omega) = 1$ and $P(\emptyset) = 0$.
3. If A_1, A_2...... is a sequence of pairwise disjoint events
   $(E_i \cap E_j = \emptyset)$ for $i \neq j$,
   then
   $P(E_1 \cup E_2 \cup E_3......) = \sum_{i=1}^{\infty} P(E_i)$
   or
   $$P(\bigcup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} P(A_i)$$

*Axiom 3 can also be stated in terms of finite union of events as
$P(E_1 \cup E_2 \cup E_3......) = \sum_{i=1}^{n} P(E_i)$
*Axiom 3 states that we can calculate probability of an event by summing up probabilities of its disjoint decomposed events.

2.2 Random Sampling

Theorem 2 (Random Sampling)**.** *Let S be a finite sample space with N equally likely events and let E be an event in S. Then*

$$P(E) = \frac{n}{N}$$

This important theorem can reduce the problem of finding probabilities to a counting problem.

Counting Rule 1: Multiplication Rule

Sampling with replacement, order matters. Consider k sets, Set 1 and Set 2 ... Set k. Set 1 has n_1, Set 2 has n_2 ... Set k has n_k distinct objects. Then the number of ways to form a set by choosing one object from each set is $n_1 n_2 ... n_k$.

Counting Rule 2: Factorial Rule

Sampling without replacement, order matters. The number of ways to arrange n distinct objects is $n!$.
$0! = 1$and $1! = 1$

Counting Rule 3: Permutation Rule

Sampling without replacement, order matters. The number of ways to arrange r chosen from n distinct object at a time without replacement, where the order matters, is known as **permutations** of n objects taken r at a time.

It is given by: ${}^nP_r = \dfrac{n!}{(n-r)!}$

Counting Rule 4: Combination Rule

Sampling without replacement, order irrelevant. The number of ways to select r object from n distinct total objects at a time without replacement, where order does not matter, is known as **combination** of n objects taken r at a time. It is given by: $\binom{n}{r} = \dfrac{n!}{(n-r)!r!}$

$\binom{n}{r}$ is also called binomial coefficient.

2.3 *Consequences of the rules of probability*

Decomposing an event

If $A_1, A_2, ...$ are pairwise disjoint events and A is their union, then $P(A) = P(A_1) + P(A_2) + ...$ Calculation of the probability of a complicated event A almost always involves decomposing A into smaller disjoint pieces whose probabilities are easier to find. Both finite and infinite decomposition is possible.

Theorem 3 (Events and complements)**.**
For any event A ,$P(A)^c = 1 - P(A)$

Theorem 4. $P(\emptyset) = 0$

Theorem 5. $P(A \cup B^C) = P(A) - P(A \cap B^c)$

Proof:
Express A as the union of disjoint events as $A = (A \cap B^C) \cup (A \cap B)$
$P(A) = P(A \cap B^C) + P(A \cap B)$
by Axiom 3,
$\Rightarrow P(A \cup B^C) = P(A) - P(A \cap B^c)$

Theorem 6 (Monotonicity of probability)**.** *If* $A \subset B$ *then* $P(A) \leq P(B)$

Proof:
$B = A \cup (A^C \cap B)$, $P(B) = P(A) + P(A^C \cap B)$ – Axiom 3
As $P(A^C \cap B) \geq 0$ – Axiom 1, $\Rightarrow P(B) \geq P(A) or P(A) \leq P(B)$

Theorem 7 (Inclusion-exclusion formulas)**.**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ - P(B \cap C) + P(A \cap B \cap C)$$

General Formula:

$$\begin{aligned} P(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) \\ + \sum_{1 \le i_1 < i_2 < i_3 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ - \sum_{1 \le i_1 < i_2 < i_3 \le i_4 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\ + ... + (-1)^{n+1} P(A_{i_1} \cap ... \cap A_{i_n}) \\ = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < ... < i_k \le n} P(A_{i_1} \cap ... \cap A_{i_k}) \end{aligned}$$

Proof:
$A \cup B = (A \cap B^C) \cup (A \cap B) \cup (A^C \cap B)$
$P(A \cup B) = P(A \cap B^C) + P(A \cap B) + P(A^C \cap B)$
$P(A \cup B) = (P(A) - P(A \cap B)) + P(A \cap B) + (P(B) - P(A \cap B))$
By Theorem 3, therefore $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Proof :
Write E as the union of its simple events (elementary outcomes).
$E = \cup_{i=1}^{n} E_i$
As the simple events are disjoint,
$P(E) = \sum_{n}^{i=1} P(E_i)$ by Axiom 3.
Similarly, $S = \cup_{i=1}^{N} E_i$ and $P(S) = \sum_{N}^{i=1} P(E_i)$ by Axiom 3.
Since all event E_i are equally likely (have the same probability of occurrence)
$\sum_{i=1}^{N} P(E_i) = NP(E_i)$ also $P(S) = 1$ by Axiom 2
Hence, $NP(E_i) = 1$ and $P(E_i) = \frac{1}{N}$
Therefore, $P(E) = \sum_{i=1}^{n} P(E_i) = \sum_{i=1}^{n} \frac{1}{N} = \frac{n}{N}$

Side notes ssss

3 Conditional probability and independence

3.1 Condition probability

Definition 3 (Conditional probability)**.**
Let B be an event in the sample space Ω such that $P(B) > 0$. Then for all events A the ***conditional probability*** *of A given B is defined as*

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

Fact:
Let B be an event in the sample space Ω such that $P(B) > 0$. Then, as a function of the event A, the conditional probability $P(A|B)$ satisfies the Kolmogorov Axioms. Especially, we have
$P(\cup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$
where, $B_i \cap B_j = \emptyset$ for $i \neq j$

Theorem 8. *Suppose that we have an experiment with **finitely many equally likely outcomes** and B is not the empty set. Then, for any event A*

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

Or simply, we have

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

Theorem 9 (Multiplication rule for n events)**.** *If $A_1, ..., A_n$ are events and all the conditional probabilities below make sense then we have*

$$P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1) \cdots P(A_n|A_1 \cdots A_{n-1})$$

Hints:

1. If required to find $P(A \cap B)$, look for either $P(A)$ or $P(B)$ and one of the conditional probabilities.
2. In word problems "of those that" implies a conditional probability.
3. Do not confuse "and" with "given that"

Nots: This implies that problems involving the intersection of several events can be simplified to a great extent by conditioning backwards.

Three special cases of connditional probability

1. Let A and B be two disjoint events, then, $A \cap B = \emptyset$ and $P(B|A) = 0$, since $P(A \cap B) = 0$
2. Let A and B be two events, such that $B \subset A$. Then, $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)}$
3. Let A and B be two events, such that $A \subset B$. Then, $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$

Calculating probability by decomposition

For example, a general version of the reasoning can be:

$$P(A) = P(AB) + P(AB^c) = P(A|B)P(B) + P(A|B^c)P(B^c). \quad (1)$$

The idea is the decomposition of a complicated event A into disjoint pieces that are easier to deal with. Above we used the pair $\{B, B^c\}$ to split A into two pieces. $\{B, B^c\}$ is an example of a *partition*.

Definition 4 (Partition)**.** *A finite collection of event $\{B_1, \ldots, B_n\}$ is a **partition** of Ω if the sets B_i are pairwise disjoint and together they make up Ω. That is , $B_iB_j = \emptyset$ whenever $i \neq j$ and $\cup_{i=1}^{n} B_i = \Omega$*

This equation is true for the same reason as the eq. (1).
Namely, set algebra gives

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} AB_i$$

Theorem 10 (The Law of Total Probability)**.** *Suppose that $B_1, \ldots, B_n$ is a partition of Ω with $P(B_i) > 0$ for $i = 1, \ldots, n$. Then for any event A we have*

$$P(A) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$