ST2131
AY21/22 SEM 2
github/jovyntls

01. COMBINATORIAL ANALYSIS

tricky - E18, E20-22, E23, E26

The Basic Principle of Counting

- **combinatorial analysis** $\rightarrow$ the mathematical theory of counting
- **basic principle of counting** $\rightarrow$ Suppose that two experiments are performed. If exp1 can result in any one of m possible outcomes and if, for each outcome of exp1, there are n possible outcomes of exp2, then together there are mn possible outcomes of the two experiments.
- **generalized basic principle of counting** $\rightarrow$ If r experiments are performed such that the first one may result in any of n_1 possible outcomes and if for each of these n_1 possible outcomes, and if ..., then there is a total of $n_1 \cdot n_2 \cdot \cdots \cdot n_r$ possible outcomes of r experiments.

Permutations

factorials - $1! = 0! = 1$

N1 - if we know how to count the number of different ways that an event can occur, we will know the probability of the event.

N2 - there are $n!$ different arrangements for n objects.

N3 - there are $\frac{n!}{n_1!\, n_2! \ldots n_r!}$ different arrangements of n objects, of which n_1 are alike, n_2 are alike, ..., n_r are alike.

Combinations

N4 - $\binom{n}{r} = \frac{n!}{(n-r)!\, r!}$ represents the number of different groups of size r that could be selected from a set of n objects when the order of selection is not considered relevant.

N4b - $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}, \quad 1 \leq r \leq n$

Proof. If object 1 is chosen $\Rightarrow \binom{n-1}{r-1}$ ways of choosing the remaining objects.
If object 1 is not chosen $\Rightarrow \binom{n-1}{r}$ ways of choosing the remaining objects.

N5 - **The Binomial Theorem** - $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

Proof. by mathematical induction: $n = 1$ is true; expand; sub dummy variable; combine using N4b; combine back to final term

Multinomial Coefficients

N6 - $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \ldots n_r!}$ represents the number of possible divisions of n districnt objects into r distinct groups of respective sizes $n_1, n_2, \ldots, n_3$, where $n_1 + n_2 + \cdots + n_r = n$

Proof. using basic counting principle,

$$= \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-n_2-n_{r-1}}{n_r}$$

$$= \frac{n!}{(n-n_1)!\, n_1!} \times \frac{(n-n_1)!}{(n-n_1-n_2)!\, n_2!} \times \cdots \times \frac{(n-n_1-n_2-\cdots-n_{r-1})}{0!\, n_r!}$$

$$= \frac{n!}{n_1!\, n_2! \ldots n_r!}$$

N7 - **The Multinomial Theorem**: $(x_1 + x_2 + \cdots + x_r)^n$

$$= \sum_{(n_1, \ldots, n_r): n_1 + n_2 + \cdots + n_r = n} \frac{n!}{n_1!\, n_2! \ldots n_r!} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

Number of Integer Solutions of Equations

N8 - there are $\binom{n-1}{r-1}$ distinct *positive* integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying $x_1 + x_2 + \cdots + x_r = n, \quad x_i > 0, \quad i = 1, 2, \ldots, r$

! cannot be directly applied to *N8* as 0 value is not included

N9 - there are $\binom{n+r-1}{r-1}$ distinct *non-negative* integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying $x_1 + x_2 + \cdots + x_r = n$

Proof. let $y_k = x_k + 1 \Rightarrow y_1 + y_2 + \cdots + y_r = n + r$

02. AXIOMS OF PROBABILITY

Sample Space and Events

- **sample space** $\rightarrow$ The *set* of all outcomes of an experiment (where outcomes are not predictable with certainty)
- **event** $\rightarrow$ Any *subset* of the sample space
- **union** of events E and $F \rightarrow E \cup F$ is the event that contains all outcomes that are either in E or F (or both).
- **intersection** of events E and $F \rightarrow E \cap F$ or EF is the event that contains all outcomes that are both in E and in F.
- **complement** of $E \rightarrow E^c$ is the event that contains all outcomes that are *not* in E.
- **subset** $\rightarrow E \subset F$ is all of the outcomes in E that are also in F.
  - $E \subset F \wedge F \subset E \Rightarrow E = F$

DeMorgan's Laws

$$(\bigcup_{i=1}^{n} \mathbf{E_i})^{\mathbf{c}} = \bigcap_{i=1}^{n} \mathbf{E_i^c}$$

Proof. to show LHS $\subset$ RHS: let $x \in (\bigcup_{i=1}^n E_i)^c$
$\Rightarrow x \notin \bigcup_{i=1}^n E_i \Rightarrow x \notin E_1$ and $x \notin E_2 \ldots$ and $x \notin E_n$
$\Rightarrow x \in E_1^c$ and $x \in E_2^c \ldots$ and $x \in E_n^c$
$\Rightarrow x \in \bigcap_{i=1}^n E_i^c$
to show RHS $\subset$ LHS: let $x \in \bigcap_{i=1}^n E_i^c$

$$(\bigcap_{i=1}^{n} \mathbf{E_i})^{\mathbf{c}} = \bigcup_{i=1}^{n} \mathbf{E_i^c}$$

Proof. using the first law of DeMorgan, negate LHS to get RHS

Axioms of Probability

definition 1: relative frequency

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

problems with this definition:

1. $\frac{n(E)}{n}$ may not converge when $n \to \infty$
2. $\frac{n(E)}{n}$ may not converge to the same value if the experiment is repeated

definition 2: Axioms

Consider an experiment with sample space S. For each event E of the sample space S, we assume that a number $P(E)$ is definned and satisfies the following 3 axioms:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (i.e., events for which $E_i E_j = \emptyset$ when $i \neq j$),

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

$P(E)$ is the probability of event E.

Simple Propositions

N1 - $P(\emptyset) = 0$

N2 - $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$ (aka axiom 3 for a finite n)

N3 - **strong law of large numbers** - if an experiment is repeated over and over again, then with probability 1, the proportion of time during which any specific event E occurs will be equal to $P(E)$.

N6 - the definitions of probability are mathematical definitions. They tell us which set functions can be called **probability functions**. They do not tell us what value a probability function $P(\cdot)$ assigns to a given event E.

probability function $\iff$ it satisfies the 3 axioms.

N7 - $P(E_c) = 1 - P(E)$

N8 - if $E \subset F$, then $P(E) \leq P(F)$

N9 - $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

N10 - Inclusion-Exclusion identity where $n = 3$

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\ &\quad - P(EF) - P(EG) - P(FG) \\ &\quad + P(EFG) \end{aligned}$$

N11 - **Inclusion-Exclusion identity** -

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots \\ &\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r}) + \ldots \\ &\quad + (-1)^{n+1} P(E_1 E_2 \ldots E_n) \end{aligned}$$

Proof. Suppose an outcome with probability ω is in exactly m of the events E_i, where $m > 0$. Then
LHS: the outcome is in $E_1 \cup E_2 \cup \cdots \cup E_n$ and ω will be counted once in $P(E_1 \cup E_2 \cup \cdots \cup E_n)$
RHS:

- the outcome is in exactly m of the events E_i and ω will be counted exactly $\binom{m}{1}$ times in $\sum_{i=1}^{n} P(E_i)$
- the outcome is contained in $\binom{m}{2}$ subsets of the type $E_{i_1} E_{i_2}$ and ω will be counted $\binom{m}{2}$ times in $\sum_{i_1 < i_2} P(E_{i_1} E_{i_2})$
- $\ldots$ and so on

hence RHS = $\binom{m}{1}\omega - \binom{m}{2}\omega + \binom{m}{3}\omega - \cdots \pm \binom{m}{m}\omega$
$= \omega \sum_{i=0}^{m} \binom{m}{i}(-1)^i =$ binomial theorem where $x = -1, y = 1$
$= 0 =$ LHS

e.g. For an outcome with probability ω and $n = 3$

- *Case 1.* $w = P(E_1 E_2)$
  LHS = ω
  RHS = (ω + ω + 0) - (ω + 0 + 0) + 0 = ω
- *Case 2.* $\omega = P(E_1 \cap E_2 \cap E_3)$
  LHS = ω
  RHS = (ω + ω + ω) - (ω + ω + ω) + ω = ω

N12 -

(i) $P(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^{n} P(E_i)$

(ii) $P(\bigcup_{i=1}^n E_i) \geq \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i E_j)$

(iii) $P(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i E_j) + \sum_{k<j<i} P(E_i E_j E_k)$

(iv) and so on.

Proof. $\bigcup_{i=1}^{n} E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \cdots \cup E_1^c E_2^c \ldots E_{n-1}^c E_n$
$P(\bigcup_{i=1}^{n} E_i) = P(E_1) + P(E_1^c E_2) + P(E_1^c E_2^c E_3) + \cdots + P(E_1^c E_2^c \ldots E_{n-1}^c E_n)$

Sample Space having Equally Likely Outcomes

tricky - 14, 15, 16, 18, 19, 20

Consider an experiment with sample space $S = \{e_1, e_2, \ldots, e_n\}$. Then $P(\{e_1\}) = P(\{e_2\}) = \cdots = P(\{e_n\}) = \frac{1}{n}$ or $P(\{e_i\}) = \frac{1}{n}$.

N1 - for any event E, $P(E) = \frac{\text{\# of outcomes in } E}{\text{\# of outcomes in } S} = \frac{\text{\# of outcomes in } E}{n}$

increasing sequence of events $\{E_n, n \geq 1\} \rightarrow$ $E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \ldots$

$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$

decreasing sequence of events $\{E_n, n \ge 1\} \to$
$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \ldots$

$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$

03. CONDITIONAL PROBABILITY AND INDEPENDENCE

tricky - E6, urns (p.37)

Conditional Probability

N1 - if $P(F) > 0$. then $P(E|F) = \frac{P(E\cap F)}{P(F)}$

N2 - **multiplication rule** - $P(E_1E_2\ldots E_n) =$
$P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\ldots P(E_n|E_1E_2\ldots E_{n-1})$

N3 - **axioms of probability** apply to conditional probability

1. $0 \le P(E|F) \le 1$
2. $P(S|F) = 1$ where S is the sample space
3. If E_i ($i \in \mathbb{Z}_{\ge 1}$) are mutually exclusive events, then
   $P(\bigcup_1^\infty E_i|F) = \sum_1^\infty P(E_i|F)$

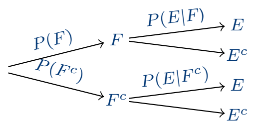
N4 - If we define $Q(E) = P(E|F)$, then $Q(E)$ can be regarded as a probability function on the events of S, hence all results previously proved for probabilities apply.

- $Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1E_2)$
- $P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1E_2|F)$
- **theorem of total probability**:
  - $Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c)$
  - $P(H|F_n) = \sum_{i=0}^{k} P(H|F_nc_i)P(c_i|F_n)$

Total Probability & Bayes' Theorem

conditioning formula - $P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$

tree diagram -



$P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(F)\cdot P(E|F)}{P(E)}$

$P(F^c|E) = \frac{P(EF^c)}{P(E)} = \frac{P(F^c)\cdot P(E|F^c)}{P(E)}$

Total Probabilility

theorem of total probability - Suppose $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$, then $P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(F_i)P(E|F_i)$

Bayes Theorem

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(F_j)P(E|F_j)}{\sum_{i=1}^{n} P(F_i)P(E|F_i)}$$

application of bayes' theorem

$$P(B_1 \mid A) = \frac{P(A|B_1)\cdot P(B_1)}{P(A|B_1)\cdot P(B_1)+P(A|B_2)\cdot P(B_2)}$$

Let A be the event that the person test positive for a disease.
B_1: the person has the disease. B_2: the person does not have the disease.

true positives: $P(B_1 \mid A)$	false negatives: $P(\bar{A} \mid B_1)$
false positives: $P(A \mid B_2)$	true negatives: $P(\bar{A} \mid B_2)$

Independent Events

N1 - E and F are independent $\iff$ $P(EF) = P(E)\cdot P(F)$

N2 - E and F are independent $\iff$ $P(E|F) = P(E)$

N3 - if E and F are independent, then E and F^c are independent.

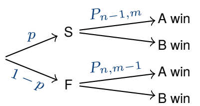
N4 - if E, F, G are independent, then E will be independent of any event formed from F and G. (e.g. $F \cup G$)

N5 - if E, F, G are independent, then $P(EFG) = P(E)P(F)P(G)$

N6 - if E and F are independent and E and G are independent,
$\not\Rightarrow E$ and FG are independent

N7 - For independent trials with probability p of success, probability of m successes before n failures, for $m, n \ge 1$,

method 1



method 2

$P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k(1-p)^{m+n-1-k}$
$= P(\text{exactly } k \text{ successes in } m+n-1 \text{ trials})$

recursive approach to solving probabilities: see page 85 alternative approach

04. RANDOM VARIABLES

- **random variable** $\to$ a real-valued function defined on the sample space

Types of Random Variables

- X is a **Bernoulli r.v.** with parameter p if $\to$
  $p(x) = \begin{cases} p, & x = 1, \text{ ('success')} \\ 1-p, & x = 0 \quad \text{('failure')} \end{cases}$
- Y is a **Binomial r.v.** with parameters n and $p \to \quad Y = X_1 + X_2 + \cdots + X_n$
  where $X_1, X_2, \ldots, X_n$ are independent Bernoulli r.v.'s with parameter p.
  - $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$
  - $P(k$ successes from n independent trials each with probability p of success$)$
  - e.g. number of red balls out of n balls drawn with replacement
- **Negative Binomial** $\to X =$ number of trials until k successes are obtained
  - e.g. number of balls drawn (with replacement) until k red balls are obtained
- **Geometric** $\to X =$ number of trials until a success is obtained
  - $P(X = k) = (1-p)^{k-1}\cdot p$ where k is the number of trials needed
  - e.g. number of balls drawn (with replacement) until 1 red ball is obtained
- **Hypergeometric** $\to X =$ number of trials until success, *without replacement*
  - e.g. number of red balls out of n balls drawn without replacement

Summary

binomial	$X =$ number of successes in n trials with replacement
negative binomial	$X =$ number of trials until k successes
geometric	$X =$ number of trials until a success
hypergeometric	$X =$ number of successes in n trials without replacement

Coupon Collector Problem

Q. Suppose there are N distinct types of coupons. If T denotes the number of coupons needed to be collected for a complete set, what is $P(T = n)$?

A. $P(T > n-1) = P(T \ge n) = P(T = n) + P(T > n)$
$\Rightarrow P(T = n) = P(T > n-1) - P(T > n)$ Let
$A_j = \{\text{no type } j \text{ coupon is contained among the first } n\}$
$P(T > n) = P(\bigcup_{j=1}^{N} A_j)$

Using the inclusion-exclusion identity,

$P(T > n) = \sum_j P(A_j)$ - coupon j is not among the first n collected
$- \sum_{j_1}\sum_{j_2} P(A_{j_1}A_{j_2})$ - coupon j_1 and j_2 are not the first n
$+ \cdots + (-1)^{k+1}\sum_{j_1}\sum_{j_2}\cdots\sum_{j_k} P(A_{j_1}A_{j_2}\cdots A_{j_n}) + \ldots$
$+(-1)^{N+1}P(A_1A_2\cdots A_N)$

$P(A_{j_1}A_{j_2}\cdots A_{j_n}) = (\frac{N-k}{N})^n$

Hence $P(T > n) = \sum_{i=1}^{N-1} \binom{N}{i}\left(\frac{N-i}{N}\right)^n (-1)^{i+1}$

Probability Mass Function

- for a *discrete* r.v., we define the **probability mass function** (pmf) of X by
  $p(a) = P(X = a)$
  - cdf, $F(a) = \sum p(x)$ for all $x \le a$
  - if X assumes one of the values $x_1, x_2, \ldots$, then $\sum_{i=1}^{\infty} p(x_i) = 1$
  - the pmf $p(a)$ is positive for at most a countable number of values of a
- e.g.

a	1	2	4
$p(a)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

- **discrete** variable $\to$ a random variable that can take on at most a countable number of possible values

Cumulative Distribution Function

- for a r.v. X, the function F defined by $F(x) = P(X \le x), \quad -\infty < x < \infty$, is called the **cumulative distribution function (cdf)** of X.
  - aka *distribution function*
  - $F(x)$ is defined on the entire real line
- e.g. $F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \le a < 2 \\ \frac{3}{4}, & 2 \le a < 4 \\ 1, & a \le 4 \end{cases}$

Expected Value

- aka population mean/sample mean, μ
- if X is a discrete random variable having pmf $p(x)$, the **expectation** or the **expected value** of X is defined as $E(X) = \sum_x x\cdot p(x)$

N1 - if a and b are constants, then $E(aX + b) = aE(X) + b$

N2 - the n^{th} moment of of X is given as $E(X^n) = \sum_x x^n \cdot p(x)$

- I is an indicator variable for event A if $I = \begin{cases} 1, \text{ if } A \text{ occurs} \\ 0, \text{ if } A^c \text{ occurs} \end{cases}$. then $E(I) = P(A)$.

Proof of N1. $E(aX + b) = \sum_x (aX + b)p(x)$
$= a\cdot\sum_x xp(x) + b\cdot\sum_x p(x) = a\cdot E(X) + b$

finding expectation of f(x)

- method 1, using pmf of Y: let $Y = f(X)$. Find corresponding X for each Y.
- method 2, using pmf of X: $E[g(x)] = \sum_i g(x_i)p(x_i)$
  - where X is a discrete r.v. that takes on one of the values of x_i with the respective probabilities of $p(x_i)$, and g is any real-valued function g

Variance

If X is a r.v. with mean $\mu = E[X]$, then the variance of X is defined by

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] \\ &= \sum x_i(x_i-\mu)^2\cdot p(x_i) \qquad \text{(deviation} \cdot \text{weight)} \\ &= E(x^2) - [E(x)]^2 \end{aligned}$$

- $Var(aX + b) = a^2Var(x)$

commutative	$E \cup F = F \cup E$	$E \cap F = F \cap E$
associative	$(E \cup F) \cup G = E \cup (F \cup G)$	$(E \cap F) \cap G = E \cap (F \cap G)$
distributive	$(E \cup F) \cap G = (E \cap F) \cup (F \cap G)$	$(E \cap F) \cup G = (E \cup F) \cap (F \cup G)$
DeMorgan's	$(\bigcup_{i=1}^{n} E_i)^c = \bigcap_{i=1}^{n} E_i^c$	$(\bigcap_{i=1}^{n} E_i)^c = \bigcup_{i=1}^{n} E_i^c$