

# 01. COMBINATORIAL ANALYSIS

tricky - E18, E20-22, E23, E26

## The Basic Principle of Counting

- combinatorial analysis** → the mathematical theory of counting
- basic principle of counting** → 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** → 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  $n_1 \cdot n_2 \cdot \dots \cdot n_r$  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  $\frac{n!}{n_1! n_2! \dots n_r!}$  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** -  $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$ ,  $1 \leq r \leq n$

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

*Proof.* using basic counting principle,

$$\begin{aligned} &= \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-n_2-\dots-n_{r-1}}{n_r} \\ &= \frac{n!}{(n-n_1)! n_1!} \times \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \times \dots \times \frac{(n-n_1-n_2-\dots-n_{r-1})!}{0! n_r!} \\ &= \frac{n!}{n_1! n_2! \dots n_r!} \end{aligned}$$

**N7 - The Multinomial Theorem:**  $(x_1 + x_2 + \dots + x_r)^n = \sum_{(n_1, \dots, n_r): n_1 + n_2 + \dots + n_r = n} \frac{n!}{n_1! n_2! \dots n_r!} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}$

## Number of Integer Solutions of Equations

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

! 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

$(x_1, x_2, \dots, x_r)$  satisfying  $x_1 + x_2 + \dots + x_r = n$

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

# 02. AXIOMS OF PROBABILITY

## Sample Space and Events

- sample space** → The *set* of all outcomes of an experiment (where outcomes are not predictable with certainty)
- event** → 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** →  $E \subset F$  if all of the outcomes in  $E$  that are also in  $F$ .
  - $E \subset F \wedge F \subset E \Rightarrow E = F$

## DeMorgan's Laws

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

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

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

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

## Axioms of Probability

### definition 1: relative frequency

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

problems with this definition:

- $\frac{n(E)}{n}$  may not converge when  $n \rightarrow \infty$
- $\frac{n(E)}{n}$  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 defined and satisfies the following 3 axioms:

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

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

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

## Simple Propositions

**N1** -  $P(\emptyset) = 0$

**N2** -  $P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$  (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** -

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

*Proof.* Suppose an outcome with probability  $\omega$  is in exactly  $m$  of the events  $E_i$ , where  $m > 0$ . Then

**LHS:** the outcome is in  $E_1 \cup E_2 \cup \dots \cup E_n$  and  $\omega$  will be counted once in  $P(E_1 \cup E_2 \cup \dots \cup E_n)$

**RHS:**

- the outcome is in exactly  $m$  of the events  $E_i$  and  $\omega$  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  $\omega$  will be counted  $\binom{m}{2}$  times in  $\sum_{i_1 < i_2} P(E_{i_1} E_{i_2})$

- ... and so on

hence  $RHS = \binom{m}{1} \omega - \binom{m}{2} \omega + \binom{m}{3} \omega - \dots \pm \binom{m}{m} \omega$

$$\begin{aligned} &= \omega \sum_{i=0}^m \binom{m}{i} (-1)^i = \text{binomial theorem where } x = -1, y = 1 \\ &= 0 = LHS \end{aligned}$$

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

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

**N12** -

- $P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i)$
- $P\left(\bigcup_{i=1}^n E_i\right) \geq \sum_{i=1}^n P(E_i) - \sum_{j < i} P(E_i E_j)$
- $P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i) - \sum_{j < i} P(E_i E_j) + \sum_{k < j < i} P(E_i E_j E_k)$
- and so on.

*Proof.*  $\bigcup_{i=1}^n E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \dots \cup E_1^c E_2^c \dots E_{n-1}^c E_n$

$$P\left(\bigcup_{i=1}^n E_i\right) = P(E_1) + P(E_1^c E_2) + P(E_1^c E_2^c E_3) + \dots + P(E_1^c E_2^c \dots E_{n-1}^c E_n)$$

## Sample Space having Equally Likely Outcomes

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

Consider an experiment with sample space  $S = \{e_1, e_2, \dots, e_n\}$ . Then  $P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_n\}) = \frac{1}{n}$  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  $\{E_n, n \geq 1\} \rightarrow$

$E_1 \subset E_2 \subset \dots \subset E_n \subset E_{n+1} \subset \dots$

lim\_{n -> infinity} E\_n = union\_{i=1}^infinity E\_i

decreasing sequence of events {E\_n, n ≥ 1} -> E\_1 supset E\_2 supset ... supset E\_n supset E\_{n+1} supset ...

lim\_{n -> infinity} E\_n = intersection\_{i=1}^infinity E\_i

03. CONDITIONAL PROBABILITY AND INDEPENDENCE

tricky - E6, urns (p.37)

Conditional Probability

- N1 - if P(F) > 0. then P(E|F) = (P(E intersection F) / P(F))
- N2 - multiplication rule - P(E\_1 E\_2 ... E\_n) = P(E\_1) P(E\_2 | E\_1) P(E\_3 | E\_1 E\_2) ... P(E\_n | E\_1 E\_2 ... E\_{n-1})
- N3 - axioms of probability apply to conditional probability

- 1. 0 ≤ P(E|F) ≤ 1
- 2. P(S|F) = 1 where S is the sample space
- 3. If E\_i (i ∈ Z\_{≥1}) are mutually exclusive events, then

P(intersection\_{i=1}^infinity E\_i | F) = intersection\_{i=1}^infinity P(E\_i | F)

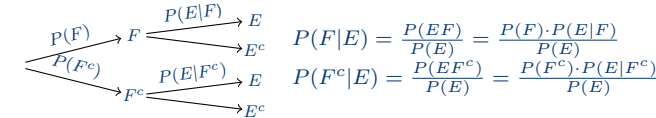
N4 - If we define Q(E) = P(E|F), then Q(E) can be regarded as a probability function on the events of S, hence all results previously proved for probabilities apply.

- Q(E\_1 union E\_2) = Q(E\_1) + Q(E\_2) - Q(E\_1 E\_2)
- P(E\_1 union E\_2 | F) = P(E\_1 | F) + P(E\_2 | F) - P(E\_1 E\_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\_n c\_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 -



Total Probability

theorem of total probability - Suppose F\_1, F\_2, ..., F\_n are mutually exclusive events such that union\_{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) = (P(EF\_j) / P(E)) = (P(F\_j) P(E|F\_j) / sum\_{i=1}^n P(F\_i) P(E|F\_i))

application of bayes' theorem

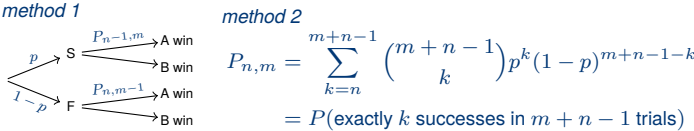
P(B\_1 | A) = (P(A|B\_1) \* P(B\_1) / (P(A|B\_1) \* P(B\_1) + P(A|B\_2) \* P(B\_2)))

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   A)	false negatives: P(A-bar   B_1)
false positives: P(A   B_2)	true negatives: P(A-bar   B_2)

Independent Events

- N1 - E and F are independent <=> P(EF) = P(E) \* P(F)
- N2 - E and F are independent <=> 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 union 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, <math>\nRightarrow</math> E and FG are independent
- N7 - For independent trials with probability p of success, probability of m successes before n failures, for m, n ≥ 1,



recursive approach to solving probabilities: see page 85 alternative approach

04. RANDOM VARIABLES

- random variable -> a real-valued function defined on the sample space

Types of Random Variables

- X is a Bernoulli r.v. with parameter p if -> p(x) = { p, x=1, ('success') / 1-p, x=0 ('failure') }
- Y is a Binomial r.v. with parameters n and p -> Y = X\_1 + X\_2 + ... + X\_n where X\_1, X\_2, ..., X\_n are independent Bernoulli r.v.'s with parameter p.
  - P(X = k) = (n choose 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 -> X = number of trials until k successes are obtained
  - e.g. number of balls drawn (with replacement) until k red balls are obtained
- Geometric -> X = number of trials until a success is obtained
  - P(X = k) = (1-p)^{k-1} \* p where k is the number of trials needed
  - e.g. number of balls drawn (with replacement) until 1 red ball is obtained
- Hypergeometric -> 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

Properties

- N1 - if X ~ Binomial(n, p), and Y ~ Binomial(n - 1, p), then E(X^k) = np \* E[(Y + 1)^{k-1}]
- N2 - if X ~ Binomial(n, p), then for k ∈ Z^+, P(X = k) = ((n-k+1)p / (k(1-p))) \* P(X = k - 1)

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 ≥ n) = P(T = n) + P(T > n) -> P(T = n) = P(T > n - 1) - P(T > n) Let A\_j = {no type j coupon is contained among the first n} P(T > n) = P(union\_{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 j\_2} P(A\_{j\_1} A\_{j\_2}) - coupon j\_1 and j\_2 are not the first n  
+ ... + (-1)^{k+1} sum\_{j\_1 j\_2} ... sum\_{j\_k} P(A\_{j\_1} A\_{j\_2} ... A\_{j\_n}) + ...  
+ (-1)^{N+1} P(A\_1 A\_2 ... A\_N)  
P(A\_{j\_1} A\_{j\_2} ... A\_{j\_n}) = ((N-k) / N)^n

Hence P(T > n) = sum\_{i=1}^{N-1} (N choose i) ((N-1) / N)^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 ≤ a
- if X assumes one of the values x\_1, x\_2, ..., then sum\_{i=1}^infinity p(x\_i) = 1
- the pmf p(a) is positive for at most a countable number of values of a

e.g. a/p(a) | 1/2, 2/4, 4/4

- discrete variable -> 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 ≤ x), -infinity < x < infinity, 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) = { 0, a < 1 / 1/2, 1 ≤ a < 2 / 3/4, 2 ≤ a < 4 / 1, a ≤ 4 }

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 \* 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 \* p(x)

- I is an indicator variable for event A if I = { 1, if A occurs / 0, if A^c occurs }. then E(I) = P(A).

Proof of N1. E(aX + b) = sum\_x (aX + b)p(x) = a \* sum\_x xp(x) + b \* sum\_x p(x) = a \* 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 μ = E[X], then the variance of X is defined by Var(X) = E[(X - μ)^2]

= sum x\_i (x\_i - μ)^2 \* p(x\_i) (deviation \* weight)  
= E(x^2) - [E(x)]^2

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

Poisson Random Variable

a r.v.  $X$  is said to be a **Poisson r.v.** with parameter  $\lambda$  if for some  $\lambda > 0$ ,

$$P(X = i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

- notation:  $X \sim \text{Poisson}(\lambda)$
- $\sum_{i=0}^\infty P(X = i) = 1$
- **Poisson Approximation of Binomial** - if  $X \sim \text{Binomial}(n, p)$ ,  $n$  is large and  $p$  is small, then  $X \sim \text{Poisson}(\lambda)$  where  $\lambda = np$ .
  - For  $n$  independent trials with probability  $p$  of success, the number of successes is approximately a *Poisson r.v.* with parameter  $\lambda = np$  if  $n$  is large &  $p$  is small.

Mean and Variance

if  $X \sim \text{Poisson}(\lambda)$ , then  $E(X) = \lambda, \quad Var(X) = \lambda$

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