ST111

Probability A
Revision Guide

Written by Itamar E. Aharoni

Contents

Introduction

This revision guide for ST111 Probability A has been designed as an aid to revision, not a substitute for it. This guide is useful for revising through key definitions, theorems and some shorter proofs found in the course.

Disclaimer: Use at your own risk. No guarantee is made that this revision guide is accurate or complete, or that it will improve your exam performance.

Authors

Written by Itamar E. Aharoni.
Any corrections or improvements should be reported by email to comms@warwickmaths.org.

1 Definitions

1.1 Sample Space

A set whose elements, denoted *Sample Points*, correspond to possible outcomes of the experiment considered. Denoted with Ω.

1.2 Sample Points

Elements of a *Sample Space*.

1.3 Event

A property that the outcome of an experiment may or may not have. It is identified with the subspace of Ω of all *Sample Points* satisfying this property.

1.4 Probability Measure

Defined on a *Sample Space* Ω, it is a function $\mathbf{P}(A)$ defined over a collection of subsets of Ω[1] such that:

- For all A where $\mathbf{P}(A)$ is defined, $0 \leq \mathbf{P}(A) \leq 1$.
- $\mathbf{P}(\phi) = 0$.
- $\mathbf{P}(\Omega) = 1$.
- **Countable Additivity:** For any sequence of events $A_1, A_2, \ldots$ for which $\mathbf{P}(A_1), \mathbf{P}(A_2), \ldots$ are defined, if they are all mutually disjoint[2] then:

$$\mathbf{P}\left(\bigcup_{i=1}^{n \text{ or } \infty} A_i\right) = \sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(A_i) \tag{1}$$

1.5 Uniform Probability

If Ω is a finite set and all outcomes are equally likely, we define the probability measure by:

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \tag{2}$$

1.6 The Fundamental Multiplication Rule

Suppose Ω is a set of sequences of the form $(\omega_1, \omega_2, \ldots, \omega_k)$ with n_1 possible values for ω_1, n_2 possible values for ω_2 (which are irrelevant of ω_1), etc. Then:

$$|\Omega| = \prod_{i=1}^{k} n_i \tag{3}$$

1.7 Types of Sampling

- With repetition.
- Without repetition.
- Ordered, where we note the order of each individual we choose.
- Unordered, where we either *forget* the order of the individuals we chose or we choose all of the *simultaneously*.

Note: Unless explicitly stated, in a "random sample" all samples are equally likely and are ordered.

[1] If Ω is countable then $\mathbf{P}(A)$ can be defined for all $A \subseteq \Omega$, but not necessary if Ω is uncountably infinite.

[2] Meaning for all $i \neq j$ holds $A_i \cap A_j = \emptyset$.

1.8 Multinomial Coefficient

Suppose in a list of length k from a set of length n we have k_1 instances of the first element, k_2 instances of the second element, etc. The multinomial coefficient deordering this list is given as:

$$\frac{k!}{k_1!k_2!\dots k_n!} \tag{4}$$

Note: If you choose an ordered sample uniformly at random and deorder it, you don't get an unordered sample chosen uniformly at random as you have e ectively chosen the number of instances of each letter.

1.9 Hypergeometric Probabilities

Probabilities following the law:

$$\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} \tag{5}$$

Note: If $k << n_1, n_2$ then hypergeometric probabilities are well-approximated by binomial probabilities, and so we can just use those if the sample space is large enough

1.10 Binomial Probabilities

Probabilities following the law:

$$\binom{k}{k_1}\left[\frac{n_1}{n}\right]^{k_1}\left[\frac{n_2}{n}\right]^{k_2} \tag{6}$$

1.11 Conditional Probability

Let A, B be two events events from the same sample space with $\mathbf{P}(A) > 0$. Then the *conditional probability of B given A* is defined as:

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)} \tag{7}$$

1.12 Independent Events

Two events, A and B, are said to be independent if:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B) \tag{8}$$

Alternatively, if $\mathbf{P}(A), \mathbf{P}(B) > 0$, this is equivalent to:

$$\mathbf{P}(A|B) = \mathbf{P}(A) \tag{9}$$
$$\mathbf{P}(B|A) = \mathbf{P}(B) \tag{10}$$

The idea behind it is that A happening doesn't affect B and vice versa.

1.13 Mutual Independence

Events $A_1, A_2, \dots, A_n$ are said to be *mutually independent* if for all choices $A_{i_1}, A_{i_2}, \dots A_{i_k}$, we have:

$$\mathbf{P}\left(\bigcap_{l=1}^{k} A_{i_l}\right) = \prod_{l=1}^{k} \mathbf{P}(A_{i_l}) \tag{11}$$

1.14 Random Variable Notation

X is known as a random variable, a quantity whose value depends on what happens in the experiment. Formally, this can be defined as a function $X : \{\text{sets in } \Omega \to \mathsf{R}\}$.

Both the formula and graph are descriptions of the distribution of X.

1.15 Binomial Distribution

If X is a random variable such that:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{12}$$

We say that it has a ***binomial distribution*** where n, p are said to be the parameters of the experiment.

1.16 Poisson Distribution

If X is a random variable such that:

$$P(X = k) = e^{-\lambda} \left(\frac{\lambda^k}{k!}\right) \tag{13}$$

We say that it has a ***Poisson distribution*** where λ is said to be the parameter of the experiment.

1.17 Normal Density Function

The ***normal density function*** or the ***standard Gaussian density function*** is defined as follows:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{14}$$

Note: $\,^{\infty}_{-\infty} \phi(x)\, dx = 1$

1.18 Stirling's Formula

Stirling's formula is defined as follows:

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\;\; n^n \;\; e^{-n}} = 1 \tag{15}$$

2 Propositions

2.1 Short Event "Dictionary"

Let $A \subseteq \Omega$ be two events of an experiment with sample space Ω. The following events and conditions are equivalent:

$A \cap B$	**Both** A and B happen.
$A \cup B$	**At least one of** A or B happens.
A^C	A **doesn't** happen.
$A \subseteq B$	**If** B happens, **then** A happens as well.

2.2 Using Singletons to Calculate Probability

Let A be an event on Ω, where Ω is countable. Then for $\omega_i \in A, i \in \{1, 2, \ldots, |A|\}$, in which ω_i are all the elements of A, we have:

$$\mathbf{P}(A) = \sum_{i=1}^{|A|} \mathbf{P}(\omega_i) \tag{16}$$

2.3 Basic Properties of Probability Measures

Let A and B be events on sample space Ω, then the following always hold:

1. $\mathbf{P}\left(A^C\right) = 1 - \mathbf{P}(A)$
2. If $A \subseteq B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$
3. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
4. let $A_1, A_2, \ldots$ be a sequence of events, not necessarily disjoint. Then:

$$\mathbf{P}\left(\bigcup_{i=1}^{n \text{ or } \infty} A_i\right) \leq \sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(A_i) \tag{17}$$

Proof.

1. By definition, the sets A, A^C are disjoint and $A \cup A^C = \Omega$. Hence, by the countable additivity property:

$$\mathbf{P}(A) + \mathbf{P}\left(A^C\right) = \mathbf{P}\left(A \cup A^C\right) = \mathbf{P}(\Omega) \tag{18}$$

By property 1 of probability measures, $\mathbf{P}(\Omega) = 1$ and so:

$$\mathbf{P}(A) + \mathbf{P}\left(A^C\right) = 1 \implies \mathbf{P}\left(A^C\right) = 1 - \mathbf{P}(A) \tag{19}$$

2. The sets A and $B \cap A^C$ are disjoint. By the countable additivity property of probability measures:

$$\mathbf{P}(A) + \mathbf{P}\left(B \cap A^C\right) = \mathbf{P}\left(A \cup \left(B \cap A^C\right)\right) = \mathbf{P}(B) \tag{20}$$

$$\mathbf{P}(A) + \mathbf{P}\left(B \cap A^C\right) = \mathbf{P}(B) \tag{21}$$

Thus by property 1 of probability measures, $\mathbf{P}\left(B \cap A^C\right) \geq 0$ and so:

$$\mathbf{P}(A) \leq \mathbf{P}(B) \tag{22}$$

3. We define the following:

$$A' = A \cap B^C \qquad B' = B \cap A^C \qquad C = A \cap B \tag{23}$$

By our choice of sets, A', B', C are all disjoint and we have:

$$A' \cup B' \cup C = A \cup B \tag{24}$$

By countable additivity:

$$\mathbf{P}(A') + \mathbf{P}(B') + \mathbf{P}(C) = \mathbf{P}(A \cup B) \tag{25}$$

Next, we note that:

$$\mathbf{P}(A) = \mathbf{P}\left(\left(A \cap B^C\right) \cup (A \cap B)\right) = \mathbf{P}(A' \cup C) \tag{26}$$

Which, by countable additivity, yields:

$$\mathbf{P}(A) = \mathbf{P}(A') + \mathbf{P}(C) \longrightarrow \mathbf{P}(A') = \mathbf{P}(A) - \mathbf{P}(C) \tag{27}$$

Similarly:

$$\mathbf{P}(B') = \mathbf{P}(B) - \mathbf{P}(C) \tag{28}$$

Lastly, we can return to (25) to receive:

$$\mathbf{P}(A \cup B) = [\mathbf{P}(A) - \mathbf{P}(C)] + [\mathbf{P}(B) - \mathbf{P}(C)] + \mathbf{P}(C) \tag{29}$$

$$= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(C) \tag{30}$$

$$= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \tag{31}$$

4. We define the following:

$$B_1 = A_1 \tag{32}$$
$$B_2 = A_2 \cap A_1^C \tag{33}$$
$$B_3 = A_3 \cap (A_1 \cup A_2)^C \tag{34}$$
$$\vdots \tag{35}$$
$$B_k = A_k \cap (A_1 \cup A_2 \cup \ldots \cup A_{k-1})^C \tag{36}$$

By our construction of the sets, we have:

- $\bigcup_{i=1}^{n \text{ or } \infty} B_i = \bigcup_{i=1}^{n \text{ or } \infty} A_i$
- $B_1, B_2, \ldots$ are all mutually disjoint.
- $B_k \subseteq A_k \qquad \forall\, k \in \mathsf{N}$

Hence, again by the countable additivity property, we have:

$$\mathbf{P}\left(\bigcup_{i=1}^{n \text{ or } \infty} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^{n \text{ or } \infty} B_i\right) = \sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(B_i) \tag{37}$$

By part 2 of this proposition:

$$\sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(B_i) \leq \sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(A_i) \tag{38}$$

In conclusion:

$$\mathbf{P}\left(\bigcup_{i=1}^{n \text{ or } \infty} A_i\right) \leq \sum_{i=1}^{n \text{ or } \infty} \mathbf{P}(A_i) \tag{39}$$

□

2.4 The Inclusion-Exclusion Principle

Let $A_1, A_2, \ldots, A_n$ be events defined over sample space Ω, not necessarily disjoint. Define P_i as follows:

$$P_1 = \sum_{i=1}^{n} \mathbf{P}(A_i) \tag{40}$$
$$P_2 = \sum_{1 \leq i_1 < i_2 \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2}) \tag{41}$$
$$P_3 = \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \tag{42}$$
$$\vdots \tag{43}$$
$$P_k = \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) \tag{44}$$
$$\vdots \tag{45}$$
$$P_n = \mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_n) \tag{46}$$

Then we have:

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i+1} P_i \tag{47}$$

Proof. The proof will follow by induction on n.

- For $n = 2$, the proof lies in part 3 of proposition 2.3.
- Suppose the proposition holds for all $n = k$, we shall prove it holds for $n = k + 1$.
  Look at the union $\bigcup_{i=1}^{n+1} A_i$ as the union of A_{n+1} and $\bigcup_{i=1}^{n} A_i$. By the induction assumption:

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i+1} P_i \tag{48}$$

  By part 3 of proposition 2.3:

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbf{P}(A_{n+1}) - \mathbf{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \tag{49}$$

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{i=1}^{n} (-1)^{i+1} P_i + \mathbf{P}(A_{n+1}) - \mathbf{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \tag{50}$$

  By distributivity properties of intersections and unions:

$$\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1} = \bigcup_{i=1}^{n} (A_i \cap A_{n+1}) \tag{51}$$

  And so:

$$\mathbf{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) = \mathbf{P}\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) \tag{52}$$

  We can now reuse the inclusion-exclusion formula again to receive:

$$\mathbf{P}\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) = \sum_{i=1}^{n} (-1)^{i+1} Q_i \tag{53}$$

  Where $Q_k = \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})$. Note that:

  - $P_1 + \mathbf{P}(A_{n+1}) = \sum_{i=1}^{n} \mathbf{P}(A_i) = P'_1$
  - $P_k + Q_{k-1} = P'_k$ for $k \in \{2, 3, \ldots, n\}$
  - $Q_n = P'_{n+1}$

  And so, we can finally return to 50 to receive:

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{i=1}^{n+1} (-1)^{i+1} P'_i \tag{54}$$

  This means that the formula holds for $n = k + 1$, assuming it does for $n = k$.

Hence, by the induction principle, the formula holds for all n. □

2.5 Counting Distinct Subsets

Let S be a set with n elements and $0 \le k \le n$. Then the number of subsets of S with k distinct elements is given by:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{55}$$

Proof. By the ***Fundamental Theorem of Multiplication***, the number of distinct sequences of length k with terms from S is:[3]

$$\frac{n!}{(n-k)!} \tag{56}$$

Similarly, for each subset T of S with k elements, we have $k!$ different sequences associated with it. Hence we can use it to count the number of distinct sequences to receive that there are:

$$|\text{Number of subsests of size } k| \cdot k! \tag{57}$$

many such sequences.
As both equations 56 and 57 are measuring the same thing, we get that

$$\frac{n!}{(n-k)!} = |\text{Number of subsests of size } k| \cdot k! \tag{58}$$

$$|\text{Number of subsests of size } k| = \frac{n!}{k!\,(n-k)!} \tag{59}$$

□

2.6 Counting Ways of Sampling

Suppose we have a population of n people and we wish to choose k people. How many ways are there to do so?

	Ordered	Unordered
With replacement	n^k	$\binom{n+k-1}{k}$
Without replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Proof. We shall split the proof for each of the four methods used for counting:

- **Ordered, with replacement:**
  There are n ways of choosing each of the k people. Treating this as a sequence problem, we can use the ***fundamental multiplication rule*** to get that there are:

  $$\underbrace{n * n * n * \ldots * n}_{k \text{ times}} = n^k \tag{60}$$

  many ways.

- **Ordered, without replacement:**
  Treating this as a problem of choosing a sequence of distinct elements of length k from a set of size n, we can use the ***fundamental multiplication rule*** to get that there are:

  $$n * (n-1) * (n-2) - \ldots * (n-k+1) = \frac{n!}{(n-k)!} \tag{61}$$

  many ways.

- **Unordered, without replacement:**
  Treating this a problem of receiving a subset of size k from a set of size n, by proposition 2.5, we get an answer of:

  $$\binom{n}{k} \tag{62}$$

[3] We have n options for the first value, $(n-1)$ for the second, ... , and $(n-k+1)$ options for the last term, resulting in $n\,(n-1)\ldots(n-k+1)$ options.

- **Unordered, With replacement:**
  We can treat this question as a problem of finding the number of multisets of size k we can create from a given set of size n. Each multiset can be represented with a list tallying how many times we chose each element. We write $n-1$ horizontal lines to create n different cells, and in them place k crosses to tally the k individuals chosen.
  For example:

  $$\times\times \mid \times \mid \quad \mid \times \times \times$$

  Figure 1: A tally for a multiset of length 6 from a sample of 4

  In the tally above, we see a multiset of length 6, from a sample size of 4, in which we chose the first element twice, the second once, the third zero times and the fourth three times.

  As each multiset corresponds to one and only one such list, and each list corresponds to one and only one such multiset - we can just count all the lists. For each tally mark, we have $n+k-1$ places at which we can place it, and we have k such markings. Hence the number of such lists is just the number of ways to place k items in $n+k-1$ places, i.e.:

  $$\binom{n+k-1}{k} \tag{63}$$

  □

2.7 Sampling From a Two-Type Population

Suppose a population consists of n individuals each of whom is either type 1 or type 2. Let the number of people of type i be n_i, then $n_1 + n_2 = n$.
Suppose we choose a random sample of size k from this population, with k_i individuals of type i, such that $k_1 + k_2 = k$.
Then the probability of this event is:

$$\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} \qquad \text{If we sample without replacement.} \tag{64}$$

$$\binom{k}{k_1}\left(\frac{n_1}{n}\right)^{k_1}\left(\frac{n_2}{n}\right)^{k_2} \qquad \text{If we sample with replacement.} \tag{65}$$

Proof. We shall divide this proof into two parts, for each kind of sampling.

- Each unordered sample is equally likely and so by definition 1.7:

  $$\mathbf{P}\left(\text{We choose a sample of } k_1 \text{ and } k_2 \text{ people}\right) = \frac{\text{Number of subsets with } k_1 \text{ and } k_2 \text{ people}}{\text{All subsets of the population}} \tag{66}$$

  By 2.5, we know that the denominator is equal to $\binom{n}{k}$, and so all that is left is to calculate the numerator[4].

  We first choose k_1 individuals from the total possible n_1, and then k_2 individuals from the total possible n_2. Hence, by the *fundamental multiplication rule*, the number of such sequences is:

  $$\binom{n_1}{k_1}\binom{n_2}{k_2} \tag{67}$$

  Meaning:

  $$\mathbf{P}\left(\text{We choose a sample of } k_1 \text{ and } k_2 \text{ people}\right) = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} \tag{68}$$

[4] Yay

- Each ordered sample is equally likely and so by definition 1.7:

$$\mathbf{P}\left(\text{We choose a sample of } k_1 \text{ and } k_2 \text{ people}\right) = \frac{\text{Number of possible ordered samples with } k_1, k_2 \text{ people}}{\text{All possible ordered samples}} \tag{69}$$

By 2.6, we know that the denominator is equal to n^k. In order to choose an ordered subset of individuals with the correct amount of people from each type, we first pick a subset $S \subseteq \{1, 2, \ldots, k\}$ of size k_1. This will correspond to the elements of type 1 in our sample.

Now, for each of these subsets, we have many different possible ordered sequences. How do we construct the ordered sequences from this subset? Let us look of the i^{th} term of the sequence. If it is in S, then it is type 1, and if it isn't in S, then it is type 2. Hence, for each S, the number of such sequences is:

$$\prod_{i \in S} n_1 \prod_{i \notin S} n_2 \tag{70}$$

Meaning all such sequences are counted by:

$$\binom{k}{k_1} \prod_{i \in S} n_1 \prod_{i \notin S} n_2 \tag{71}$$

Resulting in:

$$\mathbf{P}\left(\text{We choose a sample of } k_1 \text{ and } k_2 \text{ people}\right) = \frac{\binom{k}{k_1} \prod_{i \in S} n_1 \prod_{i \notin S} n_2}{n^k} \tag{72}$$

□

2.8 The Multiplication Rule for Conditional Probability

Suppose $A_1, A_2, \ldots, A_n$ is a finite sequence of events with:

$$\mathbf{P}\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) > 0 \tag{73}$$

$$\tag{74}$$

$$\text{Then: } \mathbf{P}\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) = \mathbf{P}\left(A_1\right) \mathbf{P}\left(A_2 | A_1\right) \mathbf{P}\left(A_3 | A_2 \cap A_1\right) \ldots \mathbf{P}\left(A_n | A_1 \cap \ldots \cap A_{n-1}\right) \tag{75}$$

Proof. By definition of conditional probability, the RHS is equal to:

$$\mathbf{P}\left(A_1\right) \frac{\mathbf{P}\left(A_1 \cap A_2\right)}{\mathbf{P}\left(A_1\right)} \frac{\mathbf{P}\left(A_1 \cap A_2 \cap A_3\right)}{\mathbf{P}\left(A_1 \cap A_2\right)} \cdots \frac{\mathbf{P}\left(A_1 \cap \ldots \cap A_n\right)}{\mathbf{P}\left(A_1 \cap \ldots \cap A_{n-1}\right)} \tag{76}$$

Which cancels to achieve the LHS. □

2.9 The Law of Total Probability

Let $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω such that $\mathbf{P}\left(A_i\right) > 0 \, \forall \, i$ and let $B \subseteq \Omega$ be another event. Then:

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}\left(A_i\right) \mathbf{P}\left(B | A_i\right) \tag{77}$$

Proof. We know that:

$$B = \bigcup_{i=1}^{n} \left(A_i \cap B\right) \tag{78}$$

where $(A_i \cap B)$ are all disjoint since A_i are all disjoint. Hence, by the additivity property of probability measures:

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(A_i \cap B) = \sum_{i=1}^{n} \mathbf{P}(A_i) \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(A_i)} \tag{79}$$

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(A_i) \mathbf{P}(B|A_i) \tag{80}$$

□

2.10 Bayes Formula

Let $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω with $\mathbf{P}(A_i) > 0 \, \forall \, i$ and let $b \subseteq \Omega$ be another event. Then for all j:

$$\mathbf{P}(A_j|B) = \frac{\mathbf{P}(B|A_j)\mathbf{P}(A_j)}{\mathbf{P}(B)} \tag{81}$$

$$= \frac{\mathbf{P}(B|A_j)\mathbf{P}(A_j)}{\sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i)} \tag{82}$$

Proof. The first equality follows from being a rearrangement of the definition for conditional probability. The second is obtained by inputting the law of total probability into the first equality. □

2.11 Independence of complements

Let $E_1, E_2, \ldots, E_n$ be independent events. Fix $S \subseteq \{1, 2, \ldots, n\}$ and define:

$$A_i = \begin{cases} E_i & \text{if } i \in S \\ E_i^C & \text{if } i \notin S \end{cases} \tag{83}$$

Then $A_1, A_2, \ldots, A_n$ are mutually independent.

Proof. Exercise. □

2.12 Binomial Probability

Fix $p \in (0, 1)$ and let $A_1, \ldots, A_n$ be independent events with $\mathbf{P}(A_i) = p$ for all i. Then the probability that exactly k of them occur out of n trials is:

$$P(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{84}$$

Proof. Let E_k be the event that exactly k of $A_1, \ldots, A_n$ happen.

$$E_k = \bigcup_{S \subseteq \{1, \ldots n\}} \left(\bigcup_{i \in S} A_i \bigcap_{i \notin S} A_i^C \right) \tag{85}$$

This is a disjoint union[5] and by the additivity property of probability measures:

$$\mathbf{P}(E_k) = \sum_{S \subseteq \{1, \ldots n\}} \mathbf{P} \left(\bigcup_{i \in S} A_i \bigcap_{i \notin S} A_i^C \right) \tag{86}$$

[5] As different Ss refer to different experimental results.

By (2.11), events A_i, A_j^C are all mutually independent and so:

$$\mathbf{P}(E_k) = \sum_{S\subseteq\{1,\dots n\}} \prod_{i\in S} \mathbf{P}(A_i) \prod_{i\notin S} \mathbf{P}\left(A_i^C\right) = \sum_{S\subseteq\{1,\dots n\}} p^{|S|}(1-p)^{n-|S|} \tag{87}$$

$$\mathbf{P}(E_k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{88}$$

□

2.13 Binomial Maximum Value

Fix $n > 1, p \in (0,1)$, and let $M = \lfloor (n+1)p \rfloor$.
If $(n+1)p \in \mathbb{Z}$, then the function $P(X=k)$ reaches its maximum value at:

$$k = M, M-1 \tag{89}$$

Otherwise, the maximum value is at:

$$k = M \tag{90}$$

Proof. Define r_k to be the ratio between successive values of k

$$r_k = \frac{P(X=k+1)}{P(X=k)} = \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k}p^k(1-p)^{n-k}} \tag{91}$$

Which after cancellation of common terms, is equal to:

$$\left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) \tag{92}$$

and so, we consider 2 cases:

- $k < M$:
We get that $r_k > 1$ and so the following element will be bigger than this one and this k is not a maximum, unless $(n+1)p \in \mathbb{Z}$, in which case $r_{k-1} = 1$ and we reach a maximum value.
- $k \geq M$:
We get that $r_k < 1$ and so the previous element will be bigger than this one and this k is not a maximum.

□

2.14 Law of Large Numbers

Fix $p \in (0,1)$. For $n = 1, 2, \dots$, let X_n be the binomial distribution with parameters n, p. Then $\forall \varepsilon > 0$:

$$\lim_{n\to\infty} \mathbf{P}[n(p-\varepsilon) \leq X_n \leq n(p+\varepsilon)] = 1 \tag{93}$$

Which is equivalent to[6]

$$\lim_{n\to\infty} \mathbf{P}\left[(p-\varepsilon) \leq \frac{X_n}{n} \leq (p+\varepsilon)\right] = 1 \tag{94}$$

Proof. We shall prove the first version of the theorem, by showing that the remaining part of the function converges to zero. We prove this by proving the part to the right of $n(p+\varepsilon)$ converges to zero, as the proof for the left side is analogous.

[6] This representation also shows us the characteristic frequency according to the frequentist model.

Let k be the smallest integer such that $k \geq n\,(p+\varepsilon)$ Take some $j \in \mathsf{Z}$, $j > k$:

$$P(X_n = j) = P(X_n = k)\left(\frac{P(X_n = k+1)}{P(X_n = k)}\right)\left(\frac{P(X_n = k+2)}{P(X_n = k+1)}\right)\dots\left(\frac{P(X_n = j)}{P(X_n = j-1)}\right) \tag{95}$$

Recalling the definition of r_k from the proof of proposition (2.13):

$$P(X_n = j) = P(X_n = k) r_k r_{k+1} \dots r_j \tag{96}$$

As $k > M$, we recall that r_k is decreasing and so:

$$P(X_n = k) r_k r_{k+1} \dots r_j \leq P(X_n = k)\; r_k^{j-k} \tag{97}$$

And so

$$P(X_n > n\,(p+\varepsilon)) = \sum_{j=k}^{n} P(X_n = j) \leq \sum_{j=k}^{n} P(X_n = k)\; r_k^{j-k} \tag{98}$$

$$\leq P(X_n = k) \sum_{j=k}^{n}\; r_k^{j-k} \tag{99}$$

As the sum $\sum_{j=k}^{n}\; r_k^{j-k}$ is just a geometric series with positive ratio smaller than one, we can write:

$$P(X_n = k) \sum_{j=k}^{n}\; r_k^{j-k} = P(X_n = k)\frac{1}{1-r_k} \tag{100}$$

Finally, we now wish to examine the properties of each of these two factors as $n \to \infty$.

1. Let us look at $P(X_n = k)$. After M, we know that $r_k < 1$ we get that $P(X_n = i)$ decreases. Hence:

$$1 = \sum_{i=0}^{n} P(X_n = i) \geq \sum_{i=M}^{k} P(X_n = i) \tag{101}$$

   Therefore:

$$\sum_{i=M}^{k} P(X_n = i) \geq \sum_{i=M}^{k} P(X_n = k) = (k - M + 1)\, P(X_n = k) \geq (n\varepsilon - 1)\, P(X_n = k) \tag{102}$$

   We can now connect both ends of the inequality

$$1 \geq (n\varepsilon - 1)\, P(X_n = k) \tag{103}$$

$$P(X_n = k) \leq \frac{1}{(n\varepsilon - 1)} \tag{104}$$

   As $n \to \infty$ the RHS goes to 0, and as $P(X_n = k) \geq 0$, we get that $P(X_n = k) \to 0$.

2. Let us look at $\frac{1}{1-r_k}$:
   Recall from the proof of (2.13) that:

$$r_k = \left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) \leq \left(\frac{n - n\,(p+\varepsilon)}{np}\right)\left(\frac{p}{1-p}\right) \tag{105}$$

   After cancelling common terms and rearranging, we get that:

$$\left(\frac{n - n\,(p+\varepsilon)}{np}\right)\left(\frac{p}{1-p}\right) \leq 1 - \frac{\varepsilon}{1-p} \tag{106}$$

   And so:

$$\frac{1}{1-r_k} \leq \frac{1-p}{\varepsilon} \tag{107}$$

Ergo, as the product of a bounded and a null sequence, we get that (100) is null as well. □

2.15 Binomial to Poisson

Fix $\lambda > 0$ and let X_n be a binomial distribution with parameters $n = p = \frac{\lambda}{n}$. Then:

$$\lim_{n \to \infty} P(X_n = k) = e^{-\lambda} \left(\frac{\lambda^k}{k!} \right) \tag{108}$$

Proof. First, for $k = 0$:

$$P(X_n = 0) = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda} \tag{109}$$

Generally, we will now use r_k again:

$$r_k = \left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) = \left(\frac{n-k}{k+1}\right)\left(\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}}\right) \to \frac{\lambda}{k+1} \tag{110}$$

We can now use the same trick as in the proof of proposition (2.14):

$$P(X_n = k) = P(X_n = 0)\left(\frac{P(X_n = 1)}{P(X_n = 0)}\right)\left(\frac{P(X_n = 2)}{P(X_n = 1)}\right)\dots\left(\frac{P(X_n = k)}{P(X_n = k-1)}\right) \tag{111}$$

$$P(X_n = k) = P(X_n = 0) r_1 r_2 \dots r_{k-1} \tag{112}$$

Lastly, we can input (109) and (110) into (112) to get that:

$$P(X_n = k) \to e^{-\lambda} \frac{\lambda}{1} \cdot \frac{\lambda}{2} \cdot \ldots \cdot \frac{\lambda}{k} = e^{-\lambda} \left(\frac{\lambda^k}{k!} \right) \tag{113}$$

□

2.16 The De-Moivre & Laplace Theorem

Fix $p \in (0,1)$ and let X_n be have binomial distribution with parameters n, p. Fix $z_1, z_2 \in \mathsf{R}$ such that $z_1 > z_2$. As $n \to \infty$:

$$P(np + z_1\sqrt{np\,(1-p)} \leq X_n \leq np + z_2\sqrt{np\,(1-p)}) \to \int_{z_1}^{z_2} \phi\,(x)\,dx \tag{114}$$

where $\phi\,(x)$ is the normal density function.

Note: This theorem was given without proof.

This guide would not be possible without our wonderful sponsors: