Notes on Probability Theory

Ayush Bhandari
(Work in Progress)

Contents

I Axioms of Probability
I-A Set Theory
I-A1 Set Operations
I-A1a Subsets
I-A1b Mutually Exclusive Sets or Disjoint Sets
I-A2 Partitions
I-A3 Complements
I-B Probability Space
I-B1 The Axioms of Probability
I-B2 Properties
I-B3 Von Mises and the Frequency Interpretation
I-B4 Equality of Events
I-B5 The class \( \mathcal{F} \) of events
I-B5a Fields
I-B6 Axiomatic Definition of an Experiment
I-B6a Countable Spaces: Case when \( \mathcal{F} \) consists of \( N \) countable outcomes.
I-B6b Non-countable Spaces: The real line

II Repeated Trials
II-A Combined Experiments
II-A1 Cartesian Products or CP
II-A2 Cartesian Products of Experiments
II-A2a Independent Experiments
II-B Bernoulli Trials
II-B1 Success and Failure of an event \( \mathcal{A} \) in \( n \) independent trials
II-B1a Most Likely Number of Successes
II-B1b Success of an event in an interval
II-C Asymptotic Theorems
II-C1 Normal Distribution (for large values of \( n \) in \( p_n(k) \))
II-C1a DeMoivre–Laplace Theorem
II-C1b Law of Large Numbers
II-C1c Generalization of Bernoulli Trials
II-C1d Generalized DeMoivre–Laplace Theorem
II-C2 Poisson Distribution for Rare Events or when \( p \ll 1 \)
II-C3 Random Poisson Points
II-C4 Asymptotic Independence of Random Poisson Points

Ayush Bhandari is with the Biomedical Imaging Laboratory at the Ecole Polytechnique Fedéralé de Lausanne (EPFL), Station 17, CH-1015 Lausanne VD, Switzerland (e-mail: Ayush.Bhandari@{epfl.ch, googlemail.com}).
III Concept of Random Variable
I. Axioms of Probability

A. Set Theory

A set is a collection of objects called elements, example: \( A = \{ \zeta_1, \ldots, \zeta_n \} \) or \( A = \{ \text{all positive, even numbers} \} \). Similarly \( \zeta_k \in A \) and \( \zeta_k \notin A \) means that \( \zeta_k \) is and is not an element of \( A \), respectively. An empty or null set denoted by \( \{ \} \) or \( \emptyset \) is the one which contains no element. A set containing \( k \) elements will have \( 2^k \) subsets whose collection is a Power Set. Example Let \( f_k \) denote the faces of a die, then, 

\[
\emptyset, \ldots, \{ f_1 \}, \ldots, \{ f_1, f_2 \}, \ldots, \{ f_1, f_2, f_3 \}, \ldots, I
\]

are the \( 2^6 \) subsets.

1) Set Operations:

a) Subsets: \( B \subset A \) or \( A \supset B \) implies that \( B \) is a subset of \( A \) or every element of \( B \) is an element of \( A \).

\[
\{ \emptyset \} \subset A \subset A \subset I.
\]

The following are the consequences,

- Transitivity — If \( C \subset B \) and \( B \subset A \) then \( C \subset A \).
- Equality — \( A = B \leftrightarrow B \subset A \) and \( A \supset B \).
- Sums or Unions — The commutative and associate operation: \( A \cup B \) or \( A + B \). If \( B \subset A \) then \( A + B = A \) and therefore,

\[
A + A = A, \quad A + \emptyset = A, \quad A + I = I.
\]

- Products or Intersections: The commutative, associate and distributive operation: \( A \cap B \) or \( A \cap B \). If \( A \subset B \) then,

\[
A \cap A = A, \quad A \cap \emptyset = \emptyset, \quad A \cap I = A.
\]

b) Mutually Exclusive Sets or Disjoint Sets: Iff \( A \cap B = \emptyset \) and for several sets \( \{ A_k \}_{k \in K}, A_kA_j = \emptyset, \forall i \neq j \).

2) Partitions: Partition \( A \) of \( I \) is a union of mutually exclusive subsets of \( I \):

\[
A_1 + \cdots + A_n = I, \quad A_iA_j = \emptyset, i \neq j
\]

and the partition is denoted by: \( A = [A_1, \ldots, A_n] \).

3) Complements: \( \overline{A} \) or \( A^c \) is a set of all elements of \( I \) which are not in \( A \).

\[
\overline{A} + A = I, \quad \overline{A}A = \emptyset, \quad A = \overline{A}, \quad I = \emptyset, \quad \emptyset = I.
\]

- Property of subsets: If \( B \subset A \) then \( \overline{A} \subset \overline{B} \).
- De Morgan's Law: \( \overline{A + B} = \overline{A} \cdot \overline{B} \) and \( \overline{A \cdot B} = \overline{A} + \overline{B} \)

- Duality Principle: Removing all the overbars preserves the identity.

B. Probability Space

The space \( I \) is called certain space whose elements are called experimental outcomes and whose subsets are events. The empty set \( \emptyset \) is an impossible event and any event \( \{ \zeta_k \} \) with single element \( \zeta_k \) is an elementary event.
**Trial** is the single performance of an experiment. At each trial, we observe a single outcome \( \zeta_k \). An event \( \mathcal{A} \) occurs during the trial if \( \zeta_k \in \mathcal{A} \). A certain event \( \zeta_m \in \mathcal{I} \) occurs on every trial and the impossible event \( \emptyset \) never occurs. At each trial one of \( \mathcal{A} \) or \( \mathcal{A}^c \) occurs. Other events that might occur are \( \mathcal{A} + \mathcal{B} \) or \( \mathcal{A} \mathcal{B} \).

1) **The Axioms of Probability:** The probability of an event \( \mathcal{A} \) is a number \( P(\mathcal{A}) \) which satisfies the following three conditions:

   I  \[ P(\mathcal{A}) \geq 0 \]

   II \[ P(\mathcal{I}) = 1 \]

   III \[ \mathcal{A} \mathcal{B} = \emptyset \Rightarrow P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}). \]

   The last axiom has an easy generalization \( \bigcap_{j,k \in K} A_j A_k = \emptyset, P\left( \bigcup_{k \in K} A_k \right) = \sum_{k \in K} P(\mathcal{A}_k). \) In the development of the theory of probability, all conclusions are directly or indirectly based on the above mentioned axioms.

2) **Properties:**

   1) The probability of an impossible event is 0 or \( P(\emptyset) = 0 \). \[ \text{The property of intersections asserts} \]

   \[ A \cap \emptyset = \emptyset \Rightarrow A \] and \( \emptyset \) are mutually exclusive. Also, since \( A + \emptyset = A \), we have:

   \[ P(A) = P(A + \emptyset) = P(A) + P(\emptyset) \Rightarrow P(\emptyset) = 0. \]

   2) The probability of an event is a positive number less than 1 or \( \forall \mathcal{A}, P(\mathcal{A}) = 1 - P(\overline{\mathcal{A}}) \leq 1 \). \[ \text{The property of complements states} \]

   \[ A + \overline{A} = \mathcal{I} \] and \( A \overline{A} = \emptyset \). Therefore, \( P(\mathcal{I}) = P(A + \overline{A}) = P(A) + P(\overline{A}) \) but \( P(\mathcal{I}) = 1 \) and the result follows.

   3) The subset property: If \( \mathcal{B} \subset \mathcal{A} \) then \( P(\mathcal{A}) \geq P(\mathcal{B}) \). \[ \text{The property of subsets states} \]

   \[ \mathcal{A} \subset \mathcal{B} \Rightarrow P(\mathcal{A}) \leq P(\mathcal{B}) = 0. \]

   4) Probability of events: \( \forall \mathcal{A}, \mathcal{B} \) we have \( P(\mathcal{A} + \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \mathcal{B}) \leq P(\mathcal{A}) + P(\mathcal{B}) \). \[ \text{The property of unions states} \]

   \[ \mathcal{A} \cap \mathcal{B} = \emptyset \Rightarrow P(\mathcal{A} + \mathcal{B}) = P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) \] but \( P(\mathcal{I}) = 1 \) and the result follows.

   3) **Von Mises and the Frequency Interpretation:** The classical theory depends on the account of empirical experiments and the associate probability is defined as

   \[ P(\mathcal{A}) = \frac{n_{\mathcal{A}}}{N} = \text{outcomes favorable to event } \mathcal{A} \]

   \[ \text{Total number of trials} \]

   The Axioms of Probability are compatible with this definition.

   I \[ P(\mathcal{A}) \geq 0 \text{ since } n_{\mathcal{A}} \geq 0 \text{ and } n > 0. \]

   II \[ P(\mathcal{I}) = 1 \text{ because } \mathcal{I} \text{ occurs at every trial.} \]

   III \[ \text{If } \mathcal{A} \mathcal{B} = \emptyset \text{ then } n_{\mathcal{A} \cup \mathcal{B}} = n_{\mathcal{A}} + n_{\mathcal{B}}. \text{ Consequently,} \]

   \[ P(\mathcal{A} \cup \mathcal{B}) = \frac{n_{\mathcal{A} + \mathcal{B}}}{n} = \frac{n_{\mathcal{A}} + n_{\mathcal{B}}}{n} = P(\mathcal{A}) + P(\mathcal{B}). \]

   For the general case: \( n_{\mathcal{A} \cup \mathcal{B}} = n_{\mathcal{A}} + n_{\mathcal{B}} - n_{\mathcal{A} \cap \mathcal{B}}. \)

   4) **Equality of Events:** Two events are equal if they have the same elements.

   5) **The class \( \mathcal{F} \) of events:** Events are subsets of \( \mathcal{I} \) to which a probability has been assigned. In practice, it is meaningful to associate a probability with only a specific subclass \( \mathcal{F} \) of subsets of \( \mathcal{I} \). **Example** if in the die experiment, we are interested in the outcomes ‘odd’ and ‘even’, then it suffices to mark the subclass.
\( \mathcal{F} = \{ \emptyset, \text{odd, even, } \mathcal{I} \} \). In the subsequent discussion, \( \mathcal{F} \) will be chosen as the unions and intersections of the events.

\textit{a) Fields:} A field is a non-empty class of sets so that:

\[
\begin{align*}
\text{If } \mathcal{A} \in \mathcal{F} & \quad \text{then } \mathcal{A} \in \mathcal{F} \\
\text{If } \mathcal{A} \in \mathcal{F} \text{ and } \mathcal{B} \in \mathcal{F} & \quad \text{then } \mathcal{A} \cup \mathcal{B} \in \mathcal{F}
\end{align*}
\]

\( \mathcal{F} \) contains the impossible event and the certain event. This is because by definition, a field is non-empty. Assume that it contains at least one element \( \mathcal{A} \). Since \( \mathcal{A} \cup \mathcal{I} = \mathcal{I} \in \mathcal{F} \) we also have \( \mathcal{I} = \emptyset \in \mathcal{F} \).

All sets that can be written as unions and intersections of finitely many sets in \( \mathcal{F} \) are also in \( \mathcal{F} \). This, however, might not be true for infinitely many sets. 

\textit{Borel Fields:} Let \( \{ \mathcal{A}_k \}_{k=0}^{\infty} \in \mathcal{F} \) — an infinite sequence of sets in \( \mathcal{F} \). If the unions and intersections of \( \mathcal{A}_k \) belong to \( \mathcal{F} \) then \( \mathcal{F} \) is called a Borel Field.

\textit{Events:} Certain subsets of \( \mathcal{I} \) that form a Borel Field. The key idea is to assign probabilities to the finite unions and intersections of events together with their limits. Repeated application of the last AOP results in conclusion that \( P(\bigcup_{k \in K} \mathcal{A}_k) = \sum_{k \in K} P(\mathcal{A}_k) \). However, this extension does not result in:

\[
P(\bigcup_{k \in \mathbb{Z}} \mathcal{A}_k) = \sum_{k} P(\mathcal{A}_k).
\]

This is an additional condition which will be referred to as \textit{axiom of infinite additivity}.

\textit{6) Axiomatic Definition of an Experiment:} An experiment is specified in terms of:

- Set \( \mathcal{I} \) of all experimental outcomes.
- The Borel Field of all event of \( \mathcal{I} \).
- The probabilities of these events.

What follows next is the determination of probabilities in experiments with finitely and infinitely many elements.

\textit{a) Countable Spaces: Case when \( \mathcal{I} \) consists of \( N \) countable outcomes:} Since \( N \) is finite, the probabilities of all the events can be expressed as the probabilities of their elementary events:

\[
P(\zeta_k) = p_k
\]

and from the axioms if:

\[
p_k \geq 0, \quad \sum_{k} p_k = 1.
\]

Analogy with the classical definition: If \( \mathcal{I} \) consists of \( N \) outcomes and the probabilities \( p_k \) of elementary events are all equal then \( p_k = 1/N \).

\textit{b) Non-countable Spaces: The real line:} If \( \mathcal{I} \) consists of infinitely many elementary events, then its probability can not be determined in terms of the probability of the elementary events. For example, if \( \mathcal{I} = \mathbb{R} \) then, it can shown that it is impossible to assign probability to all subsets of \( \mathcal{I} \) so to satisfy the axioms. However, one might consider events as bounded intervals and their countable unions and intersections, provided that the events form a field \( \mathcal{F} \). This field (usually the Borel Field) contains all open and closed intervals and all points.
Suppose \( \alpha(x) \) is a positive function such that:

\[
\alpha \geq 0, \quad \int_{\mathbb{R}} \alpha(x) \, dx = 1
\]

then, we define the probability of an event \( \{ x \leq x_k \} \) by:

\[
\mathbb{P}(x \leq x_k) = \int_{-\infty}^{x_k} \alpha(x) \, dx.
\]

Also, we have, \( \mathbb{P}(x \in (x_k, x_{k+1}]) = \int_{x_k}^{x_{k+1}} \alpha(x) \, dx \) and for the case when the intervals are mutually exclusive:

\[
\mathbb{P}(x \leq x_1) + \mathbb{P}(x_1 < x \leq x_2) = \mathbb{P}(x \leq x_2).
\]

Another interesting property is that when the function \( \alpha \) is bounded, we have,

\[
\lim_{x_k \to x_{k+1}} \mathbb{P}(x \in (x_k, x_{k+1}]) = \int_{x_k}^{x_{k+1}} \alpha(x) \, dx = 0.
\]

One can deduce that the probability of point mass, \( x_{k+1} \) is 0 and in this case, probability of all elementary events of \( \mathcal{I} \) is 0 although the probability of their unions is 1.

**C. Conditional Probability**

The Conditional Probability of an event \( \mathcal{A} \) assuming \( \mathcal{B} \) or \( \mathcal{A} \) given \( \mathcal{B} \) is denoted by

\[
\mathbb{P}_{\mathcal{B}}(\mathcal{A}) \overset{\text{def}}{=} \mathbb{P}(\mathcal{A} | \mathcal{B}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{B})}, \quad \mathbb{P}(\mathcal{B}) \neq 0.
\]

§ If \( \mathcal{B} \subset \mathcal{A} \) then \( \mathbb{P}(\mathcal{A} | \mathcal{B}) = 1 \).

§ If \( \mathcal{A} \subset \mathcal{B} \) then \( \mathbb{P}(\mathcal{A} | \mathcal{B}) = \frac{\mathbb{P}(\mathcal{A})}{\mathbb{P}(\mathcal{B})} \).

---The conditional probabilities for a fixed \( \mathcal{B} \) are indeed probabilities for the satisfy the AOP:

A1) Since \( \mathbb{P}(\mathcal{A}) \geq 0, \mathcal{A} = \mathcal{A} \cap \mathcal{B} \) and \( \mathbb{P}(\mathcal{B}) > 0 \), we have \( \mathbb{P}(\mathcal{A} | \mathcal{B}) \geq 0 \).

A2) Since \( \mathcal{B} \subset \mathcal{I} \), we have \( \mathbb{P}(\mathcal{I} | \mathcal{B}) = 1 \).

A3) If \( \mathcal{A} \cap \mathcal{B} = \emptyset \) then \( \mathbb{P}(\mathcal{A} \cup \mathcal{B} | \mathcal{B}) = \mathbb{P}(\mathcal{A} | \mathcal{B}) + \mathbb{P}(\mathcal{B} | \mathcal{B}) \).

1) **Total Probability and Bayes' Theorem**

**Total Probability Theorem:** Total Probability Theorem:

If \( \mathcal{A} = [\mathcal{A}_1, \ldots, \mathcal{A}_n] \) is a partition of \( \mathcal{I} \) and \( \mathcal{B} \) is an arbitrary event then,

\[
\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{B} | \mathcal{A}_1) \mathbb{P}(\mathcal{A}_1) + \cdots + \mathbb{P}(\mathcal{B} | \mathcal{A}_n) \mathbb{P}(\mathcal{A}_n).
\]

Note that \( \bigcap_k \mathcal{A}_k = \emptyset \) and therefore, \( \mathbb{P}(\mathcal{B}) = \sum_k \mathbb{P}(\mathcal{B} \cap \mathcal{A}_k) \). And since, \( \mathbb{P}(\mathcal{B} \cap \mathcal{A}_k) = \mathbb{P}(\mathcal{B} | \mathcal{A}_k) \mathbb{P}(\mathcal{A}_k) \), the proof follows. Also, because,

\[
\mathbb{P}(\mathcal{B} \cap \mathcal{A}_k) = \mathbb{P}(\mathcal{B} | \mathcal{A}_k) \mathbb{P}(\mathcal{A}_k) = \mathbb{P}(\mathcal{A}_k | \mathcal{B}) \mathbb{P}(\mathcal{B}),
\]

\( \text{a priori} \) \( \text{a posteriori} \)
we have the so called Bayes’ Theorem:

\[
P(A_k | B) = \frac{P(B | A_k) P(A_k)}{P(A_1) P(A_1) + \ldots + P(B | A_n) P(A_n)},
\]

in terms of a priori

2) Independence: Two events are independent if \(P(A \cap B) = P(A) P(B)\). If so is the case then the events (a) \(A, B\) and (b) \(A, B\) are also independent.

II. Repeated Trials

The key idea of this section is to study the probability of an event that occurs \(k\)–times in an experiment that is repeated \(n\)–times. Quantitatively, if the probability of certain event \(A\) of an experiment \(I\) is \(p\). If this experiment is repeated \(n\)–times, then the probability that it occurs \(k\)–times, in any order, is:

\[
p_n(k) = \binom{n}{k} p^k (1-p)^{n-k}.
\]

A. Combined Experiments

1) Cartesian Products or CP: Given two sets, say \(I_1\) and \(I_2\) with elements: \(\{a_k\}_{k \in K}\) and \(\{b_m\}_{m \in M}\), respectively, their Cartesian Product is defined as:

\[
I = I_1 \times I_2, \quad \{a_k b_m\} \in I, \quad a_k \in I_1 \text{ and } b_m \in I_2.
\]

Now if \(A \subset I_1\) and \(B \subset I_2\) then \(C = A \times B\) consists of all pairs \(\{a_k b_m\}, a_m \in A, b_m \in B\). Alternatively,

\[
A \times B = (A \times I_2) \cap (I_1 \times B).
\]

2) Cartesian Products of Experiments: CP product of two experiments results in a new experiment of form:

\[
\underbrace{I_1 \times I_2} = I \quad \text{(New Experiment)}
\]

whose events are all CP of form:

\[
\begin{align*}
A \times B & = \begin{cases} 
A \text{ is an event of } I_1 \\
B \text{ is an event of } I_2
\end{cases}, \\
\text{Their unions and intersections}
\end{align*}
\]

In the combined experiment \(I\), we can assign probabilities to events above so that

\[
\begin{align*}
P(A \times I_2) &= P_1(A) \text{ and } P(I_2 \times B) = P_2(B), \\
P(A \times B) &= P(A \times I_2) P(I_2 \times B) = P_1(A) P_2(B).
\end{align*}
\]

a) Independent Experiments:

\[
P(A \times B) = P(A \times I_2) P(I_2 \times B) = P_1(A) P_2(B).
\]

And since the elementary event \(\{a_k, b_m\}\) can be written as the CP: \(\{a_k\} \times \{b_m\}\), we have:

\[
P(\{a_k, b_m\}) = P(\{a_k\}) P(\{b_m\}).
B. Bernoulli Trials

Let \(|\mathcal{F}|\) denote the cardinality of some set \(\mathcal{F}\). If \(|\mathcal{A}| = n\) then the total number of subsets \(\mathcal{A}_k \subset \mathcal{A}\) such that \(|\mathcal{A}_k| = k\) is given by,

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]

1) Success and Failure of an event \(\mathcal{A}\) in \(n\) independent trials: Let \(\mathcal{A}\) be an event of experiment \(\mathcal{I}\). It may be the only event of \(\mathcal{I}\) or one of the many events of \(\mathcal{I}\). Now let:

\[
P(\mathcal{A}) = p, \quad P(\mathcal{\overline{A}}) = q = 1 - p.
\]

Repeating the experiment \(n\)-times, we have,

\[
\mathcal{I}^n = \mathcal{I}_1 \times \cdots \times \mathcal{I}_n.
\]

The probability \(P_k(\mathcal{A}) = p_n(k)\) such that the event \(\mathcal{A}\) occurs exactly \(k\)-times in any order is given by:

\[
P_k(\mathcal{A}) = p_n(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{Fundamental Theorem.}
\]

a) Most Likely Number of Successes: The function:

\[
p_n(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}
\]

is symmetric about \(k = n/2\). Consequently, there is a point which maximizes the value of \(p_n(k)\) for fixed \(n \in \mathbb{Z}\). Note that:

\[
\frac{p_n(k-1)}{p_n(k)} = \frac{kq}{(n-k+1)p} = 1
\]

\[
\Rightarrow k_{\text{max}} = (n+1)p
\]

Since \(p \in \mathbb{R}\), and \(k \in \mathbb{Z}\), the integer value of \(k\) that maximizes \(p_n(k)\) is \(k_{\text{max}} = \lfloor (n+1)p \rfloor\). In case, \(k^* = k_{\text{max}} \in \mathbb{Z}\), it turns out that \(k = k^*\) and \(k - 1 = k^*\) and therefore, two values of \(k\) maximize \(p_n(k)\).

b) Success of an event in an interval:

\[
P(\{k_1 \leq k \leq k_2\}) = \sum_{k=k_1}^{k=k_2} p_n(k)
\]

Probability that \(\mathcal{A}\) occurs in order

C. Asymptotic Theorems

1) Normal Distribution (for large values of \(n\) in \(p_n(k)\)): We will use the following definition of the Normal Distribution \(g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\) and an auxiliary function,

\[
G(x) = \int_{-\infty}^{x} g(t) \, dt
\]

and \(G(-\infty) = 0\). Also, \(G(-x) = 1 - G(x)\). This function is linked with the error function as \(\text{erf}(x) = G(x) - \frac{1}{2}\). Also, for large \(x\), we can use the approximation: \(G(x) \simeq 1 - \frac{\text{erf}(x)}{x}\).
a) DeMoivre–Laplace Theorem:

**Theorem 1** (DeMoivre–Laplace Theorem). If $npq \gg 1$ then for $k$ in the $\sqrt{npq}$ neighborhood of $np$, $p_n(k)$ can be approximated as:

\[
\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/(2npq)}, \quad p + q = 1, \quad p > 0, \quad q > 0
\]

\[
= \frac{1}{\sigma \sqrt{2\pi}} e^{-(k-\eta)^2/2\sigma^2)}
\]

\[
= \frac{1}{\sigma} g \left( \frac{k-\eta}{\sigma} \right)
\]

The equality for DeMoivre–Laplace Theorem holds in the limit case of $n \to \infty$. This result can further be extended to the case of $n$–trials where the occurrences of the event $\mathcal{A}$ is in the interval $k_1 \leq k \leq k_2$. In which case, one has,

\[
\sigma^2 \gg 1, \quad \sum_{k=k_1}^{k=k_2} p_n(k) = \sum_{k=k_1}^{k=k_2} \binom{n}{k} p^k (1-p)^{n-k} \approx G \left( \frac{k_2-\eta}{\sigma} \right) - G \left( \frac{k_1-\eta}{\sigma} \right)
\]

**Error Correction:** The summation above contains $k_2 - k_1 + 1$ terms while the integral contains the interval $k_2 - k_1$. There is no problem for the case when $k_2 - k_1 \gg 1$, however, when this is not true, a correction in the integral limits:

\[
G \left( \frac{k_2 + \frac{1}{2} - \eta}{\sigma} \right) - G \left( \frac{k_1 - \frac{1}{2} - \eta}{\sigma} \right)
\]

accounts for the symmetric distribution of the extra 1 in the summation.

b) **Law of Large Numbers:** According to the Von Mises model, if there are $k$ favorable outcomes in $n$–trials for an event $\mathcal{A}$, then $\mathbb{P}(\mathcal{A}) = k/n$. This has another interpretation. If an event $\mathcal{A}$ with probability $\mathbb{P}(\mathcal{A}) = p$ occurs appears $k$–times in $n$–trials then $k \simeq np$. A version of this empirical observation for the Bernoulli Trails takes form of the following theorem—theoretical justification of the empirical observation:

**Theorem 2** (Law of Large Numbers). For any $\epsilon > 0$,

\[
\lim_{n \to \infty} \mathbb{P} \left( \left| \frac{k}{n} - p \right| \leq \epsilon \right) = 1.
\]

The key idea behind its proof is that

\[
\left| \frac{k}{n} - p \right| \leq \epsilon = n (p - \epsilon) \leq k \leq n (p + \epsilon)
\]

and since $p_n(k_1 \leq k \leq k_2) \approx G \left( \frac{k_2-\eta}{\sigma} \right) - G \left( \frac{k_1-\eta}{\sigma} \right)$, we have,

\[
p_n (n (p - \epsilon) \leq k \leq n (p + \epsilon)) \approx G \left( \frac{\epsilon n}{\sigma} \right) - G \left( \frac{-\epsilon n}{\sigma} \right).
\]

Now, as $n \to \infty, \epsilon \to \infty$ and we have $G(\infty) = 0$. Also, $G(x) + G(-x) = 1$. Consequently, we have

\[
\lim_{n \to \infty} \mathbb{P} \left( \left| \frac{k}{n} - p \right| \leq \epsilon \right) = 2G(\infty) - 1 = 1.
\]
c) Generalization of Bernoulli Trials: Suppose that the experiment \( \mathcal{I} \) is partitioned into \( r \) events \( \mathcal{A}_m \) so that \( \mathcal{A} = [\mathcal{A}_1, \ldots, \mathcal{A}_r] \) with probability,

\[
P(\mathcal{A}_m) = p_m, \quad p_1 + p_2 + \cdots + p_r = 1.
\]

Repeating this experiment \( n \) times we compute the probability, \( p_n(k_1, \ldots, k_r) \) that the event \( A_m \) occurs \( k_m \) times, with \( k_1 + k_2 + \cdots + k_r = n \), we have,

\[
p_n(k_1, \ldots, k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r} = n! \prod_{m=1}^{r} \frac{p_m^{k_m}}{k_m!}.
\]

The interpretation of this result is as follows: given \( n \) items, of which \( k_1 \) are alike, \( k_2 \) are alike and so on. Since event \( \mathcal{A}_1 \) occurs \( k_1 \) times, \( \mathcal{A}_2 \) occurs \( k_2 \) times and so on, the total number of arrangements are:

\[
\binom{n}{k_1} \times \binom{n-k_1}{k_2} \times \cdots = \binom{n}{k_1, \ldots, k_r} = \frac{n!}{k_1! \cdots k_r!} \text{ Multinomial Coeff.}
\]

Having associated their respective probabilities, we have

\[
p_n(k_1, \ldots, k_r) = n! \frac{p_1^{k_1}}{k_1!} \cdots \frac{p_r^{k_r}}{k_r!}.
\]

d) Generalized DeMoivre–Laplace Theorem:

\[
p_n(k_1, \ldots, k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r} \approx \frac{\exp \left( \sum_{m=1}^{r} \frac{(k_m - \eta_m)^2}{2 \eta_m} \right)}{\sqrt{(2\pi)^r n^{-1}}} = \frac{\exp \left( \frac{1}{2} \sum_{m=1}^{r} \frac{(k_m - \eta_m)^2}{\eta_m} \right)}{\sqrt{(2\pi)^r n^{-1}}} \prod_{m=1}^{r} \frac{p_m^{k_m}}{k_m!}.
\]

2) Poisson Distribution for Rare Events or when \( p \ll 1 \): The Poisson Distribution is used to approximate the case when the event \( \mathcal{A} \) is rare, meaning that \( P(\mathcal{A}) \ll 1 \).

**Theorem 3** (Poisson Theorem). If \( p \ll 1, n \gg 1 \) and \( k \sim np \), then,

\[
\binom{n}{k} p^k q^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.
\]

This theorem can be generalized to \( r \) partitions as:

\[
p_n(k_1, \ldots, k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r} \approx \prod_{m=1}^{r} \frac{(np_m)^{k_m} e^{-np_m}}{k_m!}.
\]

3) Random Poisson Points: Let \( \Delta \) be an arbitrary subinterval of \( \mathcal{I} = (\frac{T}{2}, \frac{T}{2}) \). Now place \( n \in \mathbb{Z} \) at random in the interval \( \mathcal{I} \). We are interested in the following probability:

\[
P(\mathcal{A}) = P(\{k \text{ points in any order } \in \Delta\})
\]
which is given by
\[ P(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad p = \frac{k}{\Delta}. \]

Assuming that \( n \gg 1 \) and \( \Delta \ll T \) we can use the Poisson Theorem and
\[ P(k) \simeq e^{-\eta} \frac{\eta^k}{k!}, \quad \eta = np = n\frac{\Delta}{T}, \quad k \sim \eta. \]

4) Asymptotic Independence of Random Poisson Points: Consider two non-overlapping intervals \( \Delta_1 \) and \( \Delta_2 \). The probability that out of \( n \) points, \( k_1 \) points are in \( \Delta_1 \) and \( k_2 \) points are in \( \Delta_2 \) is:
\[ P(k_1 \in \Delta_1, k_2 \in \Delta_2) = \frac{n!}{k_1!k_2!(n-(k_1+k_2))!} p_1^{k_1} p_2^{k_2} (1-(p_1+p_2))^{n-(k_1+k_2)} \]
\[ P(k_1 \in \Delta_1, k_2 \in \Delta_2) \neq P(k_1 \in \Delta_1) \cdot P(k_2 \in \Delta_2) \quad \text{(not independent)} \]
where \( p_1 = \Delta_1/T \) and \( p_2 = \Delta_2/T \).

However, if \( n \) and \( T \) are increased indefinitely while keeping \( n/T \sim \lambda \) constant, we have,
\[ P(k_1 \in \Delta_1, k_2 \in \Delta_2) = e^{-\eta_1} \frac{\eta_1^{k_1}}{k_1!} e^{-\eta_2} \frac{\eta_2^{k_2}}{k_2!}, \quad \eta = np_m \]
\[ P(k_1 \in \Delta_1, k_2 \in \Delta_2) = P(k_1 \in \Delta_1) \cdot P(k_2 \in \Delta_2) \quad \text{(independent).} \]

In conclusion, Poisson Random Points are independent in the asymptotic limit.

III. Concept of Random Variable