EK381 Exam 1 Formula Sheet

1. Foundations of Probability

Set Theory

- A set is a collection of elements.
- We usually use capital letters (such as A) to refer to sets and lowercase letters (such as x) to refer to elements.
- $x \in A$ means "x is an element of the set A."
- $x \notin A$ means "x is not an element of the set A."
- The empty set or null set is the set with no elements. Notation: ϕ or $\{\ \}$.
- The universal set Ω is the set of all elements (for the specific context).
- A subset A of a set B is a set consisting of some (or none or all) of the elements of B. Notation: $A \subset B$.
- Two sets A and B are equal if and only if $A \subset B$ and $B \subset A$.

Set Operations

- Complement: $A^c = \{x : x \notin A\}$.
- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$.
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
- Set Difference: $A - B = \{x : x \in A \text{ and } x \notin B\}$.

Other Set Concepts

- A collection of sets $A_1, \dots, A_n$ is mutually exclusive if $A_i \cap A_j = \phi$ for $i \neq j$.
- A collection of sets $A_1, \dots, A_n$ is collectively exhaustive if $A_1 \cup \cdots \cup A_n = \Omega$.
- A collection of sets $A_1, \dots, A_n$ is a partition if it is both mutually exclusive and collectively exhaustive.

De Morgan's Laws

$$(A \cup B)^c = A^c \cap B^c \qquad \left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c$$

$$(A \cap B)^c = A^c \cup B^c \qquad \left(\bigcap_{i=1}^n A_i\right)^c = \bigcup_{i=1}^n A_i^c$$

Probability Axioms

- An experiment is a procedure that generates observable outcomes.
- An outcome is a possible observation of an experiment.
- The sample space Ω is the set of all possible outcomes.
- An event is a subset of Ω: it is a set of possible outcomes.
- A probability measure $\mathbb{P}[\cdot]$ is a function that maps events to real numbers. It must satisfy the following axioms:
  1. Non-negativity: For any event A, $\mathbb{P}[A] \geq 0$.
  2. Normalization: $\mathbb{P}[\Omega] = 1$.
  3. Additivity: For any countable collective $A_1, A_2, \dots$ of mutually exclusive events,
  $$\mathbb{P}[A_1 \cup A_2 \cup \cdots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \cdots \ .$$
- These axioms imply the following properties:
  - Complement: $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.
  - Inclusion-Exclusion: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

Conditional Probability

- The conditional probability of event A given that B occurs is
$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \ .$$
- For $\mathbb{P}[B] = 0$, $\mathbb{P}[A|B]$ is undefined.
- Conditional probability satisfies the probability axioms:
  - Non-negativity: For any event A, $\mathbb{P}[A|B] \geq 0$.
  - Normalization: $\mathbb{P}[\Omega|B] = 1$.
  - Additivity: For any countable collective $A_1, A_2, \dots$ of mutually exclusive events,
  $$\mathbb{P}[A_1 \cup A_2 \cup \cdots |B] = \mathbb{P}[A_1|B] + \mathbb{P}[A_2|B] + \cdots \ .$$
- Multiplication Rule: For two events A and B, $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B|A] = \mathbb{P}[B]\,\mathbb{P}[A|B]$.
For n events $A_1, A_2, \dots, A_n$,
$$\mathbb{P}\left[\bigcap_{i=1}^n A_i\right] = \mathbb{P}[A_1]\,\mathbb{P}[A_2|A_1] \cdots \mathbb{P}[A_n|A_1 \cap \cdots \cap A_{n-1}] \ .$$
- Law of Total Probability: For a partition $B_1, \dots, B_n$ satisfying $\mathbb{P}[B_i] > 0$ for all i,
$$\mathbb{P}[A] = \sum_{i=1}^n \mathbb{P}[A|B_i]\mathbb{P}[B_i] \ .$$
- Bayes' Rule: This is a method to "flip" conditioning:
$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A]} \ .$$
Sometimes, it is useful to solve for the denominator using the total probability theorem. For a partition $B_1, \dots, B_n$ satisfying $\mathbb{P}[B_i] > 0$ for all i,
$$\mathbb{P}[B_j|A] = \frac{\mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}{\sum_{i=1}^n \mathbb{P}[A|B_i]\,\mathbb{P}[B_i]} \ .$$

Independence

- Two events A and B are independent if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$.
- Independence of A and B means that knowing if A occurs cannot help predict whether B also occurs (and vice versa).
- Events $A_1, \dots, A_n$ are independent if
  - All collections of $n-1$ events chosen from $A_1, \dots, A_n$ are independent.
  - $\mathbb{P}[A_1 \cap \cdots \cap A_n] = \mathbb{P}[A_1] \cdots \mathbb{P}[A_n]$
- Independence means that no subset of the events can be used to help predict the occurrence of any other subset of events.
- If $A_1, \dots, A_n$ only satisfy $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j]$ for all $i \neq j$, then we say they are pairwise independent (but not independent).
- If A and B are independent, then so are A and B^c, A^c and B, and A^c and B^c.

Conditional Independence

- The events A and B are conditionally independent given C if
$$\mathbb{P}[A \cap B|C] = \mathbb{P}[A|C]\,\mathbb{P}[B|C] \ .$$
- Conditional independence means that, given C occurs, knowing that A occurs cannot help predict whether B also occurs (and vice versa).
- Events $A_1, \dots, A_n$ are conditionally independent given B if
  - All collections of $n-1$ events chosen from $A_1, \dots, A_n$ are conditionally independent given B.
  - $\mathbb{P}[A_1 \cap \cdots \cap A_n|B] = \mathbb{P}[A_1|B] \cdots \mathbb{P}[A_n|B]$
- Independence does not imply conditional independence.
- Conditional independence does not imply independence.

Counting

- If an experiment is composed of m subexperiments and the i^{th} subexperiment consists of n_i outcomes (that can be freely chosen), then the total number of outcomes is $n_1\, n_2 \cdots n_m$.
- Counting techniques are especially useful in scenarios where all outcomes are equally likely, since the probability of an event can be expressed as
$$\mathbb{P}[A] = \frac{\#\text{ outcomes in } A}{\#\text{ outcomes in } \Omega}$$

Sampling

- A sampling problem consists of n distinguishable elements with k selections to be made.
  - Selections may be made either with or without replacement.
  - The final outcome is either order dependent or order independent.

	Order: Dependent	Order: Independent
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

2. Discrete Random Variables

- A random variable is a mapping that assigns real numbers to outcomes in the sample space.
- Random variables are denoted by capital letters (such as X) and their specific values are denoted by lowercase letters (such as x).
- The range of a random variable X is denoted by R_X.

Probability Mass Function

- The probability mass function (PMF) specifies the probability that a discrete random variable X takes the value x:
$$P_X(x) = \mathbb{P}[X = x].$$
- The PMF satisfies the following basic properties:
  1. Non-negativity: $P_X(x) \geq 0$ for all x.
  2. Normalization: $\sum_{x \in R_X} P_X(x) = 1$.
  3. Additivity: For any event $B \subset R_X$, the probability that X falls in B is
$$\mathbb{P}[\{X \in B\}] = \sum_{x \in B} P_X(x).$$

Cumulative Distribution Function

- The cumulative distribution function (CDF) returns the probability that a random variable X is less than or equal to a value x:
$$F_X(x) = \mathbb{P}[X \leq x].$$
- The CDF satisfies the following basic properties:
  - Normalization: $\lim_{x \to \infty} F_X(x) = 1$.
  - Non-negativity: $F_X(x)$ is a non-decreasing function of x.
  - For $b \geq a$, $F_X(b) - F_X(a) = \mathbb{P}[a < X \leq b]$.
- For discrete random variables, $F_X(x)$ is piecewise constant and jumps at points x where $P_X(x) > 0$ by height $P_X(x)$.

Expected Value

- The expected value of a discrete random variable X is
$$\mathbb{E}[X] = \sum_{x \in R_X} x \, P_X(x).$$
- This is also known as the mean or average.
- Sometimes denoted as $\mu_X = \mathbb{E}[X]$.

Variance

- The variance measures how spread out a random variable is around its mean,
$$\mathsf{Var}[X] = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \sum_{x \in R_X} (x - \mu_X)^2 P_X(x).$$
- Alternate formula: $\mathsf{Var}[X] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2$.
- Standard Deviation: $\sigma_X = \sqrt{\mathsf{Var}[X]}$.
- The variance is sometimes written as $\sigma_X^2 = \mathsf{Var}[X]$.

Functions of a Random Variable

- A function $Y = g(X)$ of a discrete random variable X is itself a discrete random variable.
- Range: $R_Y = \{g(x) : x \in R_X\}$.
- PMF: $P_Y(y) = \sum_{x : g(x) = y} P_X(x)$.
- Expected Value of a Function $Y = g(X)$:
$$\mathbb{E}[g(X)] = \mathbb{E}[Y] = \sum_{y \in R_Y} y \, P_Y(y) = \sum_{x \in R_X} g(x) \, P_X(x)$$
- **Linearity of Expectation:**
$$\mathbb{E}[aX + b] = a \, \mathbb{E}[X] + b$$
$$\mathbb{E}[g(X) + h(Y)] = \mathbb{E}[g(X)] + \mathbb{E}[h(Y)]$$
- **Variance of a Linear Function:**
$$\mathsf{Var}[aX + b] = a^2 \, \mathsf{Var}[X]$$

Important Families of Discrete RVs

Bernoulli Random Variables

- X is a Bernoulli(p) random variable if it has PMF
$$P_X(x) = \begin{cases} 1 - p & x = 0, \\ p & x = 1 \, . \end{cases}$$
- Range: $R_X = \{0, 1\}$.
- Expected Value: $\mathbb{E}[X] = p$.
- Variance: $\mathsf{Var}[X] = p(1 - p)$.
- Interpretation: single trial with success probability p.

Geometric Random Variables

- X is a Geometric(p) random variable if it has PMF
$$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$
- Range: $R_X = \{1, 2, \ldots\}$.
- Expected Value: $\mathbb{E}[X] = \frac{1}{p}$.
- Variance: $\mathsf{Var}[X] = \frac{1-p}{p^2}$.
- Interpretation: # of independent Bernoulli(p) trials until first success.

Binomial Random Variables

- X is a Binomial(n, p) random variable if it has PMF
$$P_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$
- Range: $R_X = \{0, 1, \ldots, n\}$.
- Expected Value: $\mathbb{E}[X] = np$.
- Variance: $\mathsf{Var}[X] = np(1 - p)$.
- Interpretation: # of successes in n independent Bernoulli(p) trials.

Discrete Uniform Random Variables

- X is a Discrete Uniform(a, b) random variable if it has PMF
$$P_X(x) = \begin{cases} \dfrac{1}{b - a + 1} & x = a, a+1, \ldots, b, \\ 0 & \text{otherwise.} \end{cases}$$
- Range: $R_X = \{a, a+1, \ldots, b\}$.
- Expected Value: $\mathbb{E}[X] = \frac{a+b}{2}$.
- Variance: $\mathsf{Var}[X] = \frac{(b-a)(b-a+2)}{12} = \frac{(b-a+1)^2 - 1}{12}$.
- Interpretation: equally likely to take any (integer) value between a and b.

Poisson Random Variables

- X is a Poisson(λ) random variable if it has PMF
$$P_X(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, \ldots \\ 0 & \text{otherwise.} \end{cases}$$
- Range: $R_X = \{0, 1, \ldots\}$.
- Expected Value: $\mathbb{E}[X] = \lambda$.
- Variance: $\mathsf{Var}[X] = \lambda$.
- Interpretation: # of arrivals in a fixed time window.

Conditioning for Discrete Random Variables

- The conditional PMF of X given an event B is
$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{\mathbb{P}[\{X \in B\}]} & x \in B \\ 0 & x \notin B \end{cases}$$
where $\mathbb{P}[\{X \in B\}] = \sum_{x \in B} P_X(x)$.
- The conditional PMF satisfies the basic PMF properties
  1. Non-negativity: $P_{X|B}(x) \geq 0$ for all x.
  2. Normalization: $\sum_{x \in B} P_{X|B}(x) = 1$.
  3. Additivity: For any event $C \subset R_X$, the conditional probability that X falls in C given that X falls in B is
$$\mathbb{P}\big[\{X \in C\} \big| \{X \in B\}\big] = \sum_{x \in C} P_{X|B}(x).$$
- The conditional expected value of X given an event B is
$$\mathbb{E}[X|B] = \sum_{x \in B} x \, P_{X|B}(x) \, .$$
- The conditional expected value of a function $g(X)$ given an event B is
$$\mathbb{E}\big[g(X)|B\big] = \sum_{x \in B} g(x) \, P_{X|B}(x) \, .$$
- The conditional variance of X given an event B is
$$\mathsf{Var}\big[X|B\big] = \mathbb{E}\Big[\big(X - \mathbb{E}[X|B]\big)^2 \Big| B\Big] = \sum_{x \in B} \big(x - \mathbb{E}[X|B]\big)^2 P_{X|B}(x)$$
$$= \mathbb{E}\big[X^2|B\big] - \big(\mathbb{E}[X|B]\big)^2$$