MA224: PROBABILITY TEST 1 OUTLINE

Chapter 1: Probability
Chapter 2: Discrete Distributions

1.1. Basic Concepts

- Random experiments: an experiment (action with multiple outcomes) for which the outcome cannot be predicted with certainty
- Sample space, universal set, or outcome space S: collection of all possible outcomes
  - Can be continuous or discrete (countable), infinite or finite
- Random variables: measurements on outcomes associated with random experiments
  - Usually denoted with capital letter
- Distribution of a random variable, or population: a description of the frequencies of different outcomes
  - Usually estimated through samples, collection of the observations that are obtained from repeated trials of the random experiment
  - Statistical inference: the process of making a conjecture about the distribution of a random variable based on a sample
- Probability of an outcome A, with frequency $f = \mathcal{N}(A)$ in n trials, is $P(A) = \dfrac{\mathcal{N}(A)}{n}$. This ratio is also called the relative frequency.
  - Frequency table, relative frequency table, histogram, or density histogram can be used to graphically/visually show frequencies of occurrences out of total trials
  - Relative frequencies can be unstable for small n, but tends to stabilize for a large n towards $P(A)$
- Probability mass function (p.m.f.) is a function that serves as a model for the probabilities of the outcomes of a random experiments
  - i.e., if random experiment repeated many times, it is expected that the relative frequency $\dfrac{\mathcal{N}(x_0)}{n} \to P(x = x_0) = f(x_0)$
  - Can construct a probability histogram, which should be close to the frequency histogram if n is large and the model is good
- Simpson's paradox: Relative frequencies are estimates towards probabilities, but you can't easily compare multiple groups of unlike relative frequencies to estimate a composite probability. (i.e., It's possible that to have two groups and two random experiments, in which one group has lower probabilities in both random experiments but a higher total probability than the other, based on the conditional probabilities of the two random experiments.)

1.2. Properties of Probability

- Event: Given an outcome space S, let $A \subset S$. Thus A is an event, a subset of S. When the outcome of the experiment is in A, then A has occurred.
- S and any event are sets, and follow set theory, and can be illustrated with Venn diagrams\
  - Null set: $\emptyset$
  - A subset of B: $A \subset B$
  - A union B: $A \cup B$
  - A intersection B: $A \cap B$
  - Complement of A: A'
  - $A_1, A_2, \ldots, A_k$ mutually exclusive events if $A_i \cap A_j = \emptyset, i \neq j$ (i.e., all events are disjoint sets)
  - $A_1, A_2, \ldots, A_k$ exhaustive events if $A_1 \cup A_2 \cup \cdots \cup A_k = S$

- Possible to have set of events that are both mutually exclusive and exhaustive
- Probability: a real-valued set function ($P : \{\mathbb{R}\} \to \mathbb{R}$) that assigns, to each event A in the sample space S the probability (real number) $P(A)$, which follows the following properties:
    - $P(A) \geq 0$
    - P(S)=1
    - If set of mutually exclusive events $A_1, A_2, \ldots, A_k$, then $P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$ (for any countable number of events)
- Properties of probability (know how to prove each one, proofs are easy):
    - $P(A) = 1 - P(A')$
    - $P(\emptyset) = 0$
    - If $A \subseteq B$, then $P(A) \leq P(B)$
    - $P(A) \leq 1$
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (for any event, not only mutually exclusive)
        - Can be extended to more elements: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cup B \cup C)$
- If m equally likely, mutually-exclusive and exhaustive outcomes, then the probability of any of those outcomes is $\frac{1}{m}$
    - If h mutually-exclusive, equally likely outcomes in event A, then $P(A) = \frac{h}{m}$

1.3. Methods of Enumeration

- Multiplication principle: If event E_1 has n_1 possible outcomes, and for each of these possible outcomes, event E_2 has n_2 possible outcomes, then the composite experiment E_1E_2 has n_1n_2 possible outcomes.
- Permutation: each of the $n!$ possible arrangements of n different objects.
- Each of the ${}_nP_r$ arrangements is called a permutation of n objects taken r at a time (ordered sample of size r)
- Different sampling methods (n is number of distinct objects, r is sample size):
    - Without replacement, ordered: ${}_nP_r = \frac{n!}{r!}$
    - Without replacement, unordered: ${}_nC_r = \frac{n!}{r!(n-r)!} = \binom{n}{r}$
    - With replacement, ordered: n^r
    - With replacement, unordered: ${}_{n-1+r}C_r = \frac{(n-1+r)!}{r!(n-1)!}$
- Number of distinguishable permutations:
    - Same as choosing without replacement, unordered (${}_nC_r$)
    - For more than two distinguishable types, multinomial coefficient number of possibilities (i.e., for k distinguishable types such that $n_1 + n_2 + \cdots + n_k = n$, $\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!n_2!\ldots n_k!}$ possible outcomes)
- Note: $\binom{n}{r} = \binom{n}{n-r}$

1.4. Conditional Probability

- Conditional probability of event A given that event B has occurred, is: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, given that $P(B) > 0$
  - Thus, $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$ (think multiplication rule)
  - Note that conditional probability follows the axioms for a probability function

1.5. Independent Events

- A pair of events is independent (statistically independent, stochastically independent, independent in a probabilistic sense) if the occurrence of one does not change the probability of the occurrence of the other
  - i.e., $P(A) = P(A|B)$
  - Thus, because $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, $P(A)P(B) = P(A \cap B)$ (special case of the multiplication rule)
- Theorem: If A and B are independent events, then so are (prove these as well, also easy proofs):
  - A and B'
  - A' and B
  - A' and B'
- Events A, B, and C are mutually independent IFF both conditions hold (can be extended to larger sets of events, where each pair, triplet, quartet, etc. satisfy the special multiplication rule):
  - A, B, and C are pairwise independent
  - $P(A \cap B \cap C) = P(A)P(B)P(C)$

1.6. Bayes's Theorem

- Consider a space that is partitioned into k mutually exclusive, exhaustive events $B_1, B_2, \ldots, B_k$ with known probabilities (prior probabilities), and a space A, such that $P(A|B_i), 1 \leq i \leq k$ is known
  - Once event A has occurred, the probability that the outcome was in event B_j is the posterior probability $P(B_j|A) = \dfrac{P(B_j \cap A)}{P(A)} = \dfrac{P(B_j)P(A|B_j)}{\sum\limits_{i=1}^{k} P(B_k)P(A|B_k)}, j = 1, 2, \ldots k$ (Bayes's Theorem)

2.1. Random Variables of the Discrete Type

- Given a random experiment with an outcome space S, a function X that maps every element of S to a real number is called a random variable ($X : S \to \mathbb{R}$)
  - The space of X is the set of real numbers $\{x : X(s) = x, s \in S\}$
  - In other words, X can be thought of as a numeric measurement (or designation) taken from a random experiment, a way of "mathematicalizing" an arbitrary outcome space by mapping it to the real number line
- Discrete (countable) types were mentioned in 1.1, this is the type dealt with in this chapter
- The p.m.f. of a discrete random variable X is a function $f(x) = P(X = x)$ with the following properties:
  - $f(x) > 0, x \in S$
  - $\sum\limits_{x \in S} f(x) = 1$

- $P(X \in A) = \sum_{x \in A} f(x)$, where $A \subseteq S$

- Hypergeometric distribution: when choosing n items from a collection of $N = N_1 + N_2$ objects, where N_1 and N_2 are the counts of the two distinguishable classes of objects, and the random variable X is the number of objects selected of type N_1, then the p.m.f. is:
$$f(x) = P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, 0 \le x \le N_1, x \le n \le N$$

2.2. Mathematical Expectation

- Mathematical expectation, or expected value of function $u(X)$ random variable X of the discrete type with space S is $E(u(X)) = \sum_{x \in S} u(x) f(x)$
  - Since $u(X)$ is also a function mapping X to another value, it can be thought of as another random variable (will produce same result, just a different way to think about it)
- When it exists, the mathematical expectation E satisfies the properties (be able to prove these):
  - If c is a constant, then $E(c) = c$
  - If c is a constant and u is a function, then $E[cu(X)] = cE[u(X)]$
  - If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$ (i.e., E is a linear operator)
- For a hypergeometric distribution, $E(X) = n\frac{N_1}{N} = np$ (makes sense)

2.3. The Mean, Variance, and Standard Deviation

- The mean of a random variable X is $\mu = E(X) = \sum_{x \in S} x f(x)$
  - This is the first moment about the origin
  - The first moment about the mean is $E(X - \mu) = 0$
- The variance of a random variable X is the second moment about the mean: $\sigma^2 = E((X-\mu)^2) = \sum_{x \in S} (x-\mu)^2 f(x)$
  - The standard deviation is $\sigma = (+)\sqrt{\sigma^2}$
  - Formula 2: $\sigma^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2$, because $\mu = E(X)$
  - Formula 3: $\sigma^2 = E(X(X-1)) + \mu - \mu^2$ (see derivation below under factorial moment)
- If random variable $Y = aX + b$ (linear mapping of random variable X) (be able to derive these and think about them intuitively):
  - $\mu_Y = a\mu_X + b$
  - $\sigma_Y^2 = a^2 \sigma_X^2$ (or $\sigma_Y = a\sigma_X$)
- r^{th} moment of the distribution about b is $E((X-b)^r)$
- r^{th} factorial moment is $E(X(X-1)(X-2)\ldots(X-r+1))$
  - Using the second factorial moment, $E(X(X-1)) = E(X^2) - E(X)$. Thus, $\sigma^2 = E((X-\mu)^2) = E(X^2) - \mu^2 = E(X(X-1)) + \mu - \mu^2$. This is sometimes easier to calculate than using other formulas for variance
- Variance of a hypergeometric distribution is $\sigma^2 = npq\left(\frac{N-n}{N-1}\right)$

- If a random experiment is actually performed n times, the collection of outcomes is called a sample
  - The empirical distribution has a p.m.f. $f(x) = \frac{1}{n}$, sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (sample mean and sample variance are used to estimate the population mean and standard deviation)

2.4. Bernoulli Trials and the Binomial Distribution

- Bernoulli experiment: random experiment with only two mutually-exclusive and exhaustive outcomes (usually success and failure)
  - Bernoulli trials: a sequence of Bernoulli experiments performed several independent times
  - Let $p = P(\text{success})$, $q = 1 - p$
  - A Bernoulli experiment (a single Bernoulli trial) has a Bernoulli distribution, with random variable X corresponding to success (1) or failure (0), with p.m.f. $f(x) = p^x q^{1-x}, x = 0, 1;$ $\mu = p$; and $\sigma^2 = pq$
- When a sequence of n Bernoulli trials carried out, it has a binomial distribution, with random variable X indicating number of successes
  - p.m.f.: $f(x) = \binom{n}{x} p^x q^{n-q}, x = 0, 1, 2, \ldots n$; $\mu = np$; $\sigma^2 = npq$
  - Bernoulli distribution is special case of binomial distribution with $n = 1$
  - Binomial distribution be represented shorthand as $b(n, p)$
  - This can be used to approximate the hypergeometric distribution (similar, but without replacement) when n is large, since events are essentially independent of one another
- Cumulative distribution function (c.d.f.), or distribution function, of a random variable X, is a function such that $F(x) = P(X \leq x), -\infty < x < \infty$
  - $f(x) = F(x) - F(x-1)$
  - $P(X > x) = 1 - F(x)$

2.5. The Moment-Generating Function

- Let X be a random variable of the discrete type with p.m.f. $f(x)$ and space S. If there is a positive number h such that $E(e^{tX}) = \sum_{x \in S} e^{tx} f(x)$ exists and is finite for $-h < t < h$, then the function of t defined by $M(t) = E(e^{tX})$ is called the moment-generating function of X (m.g.f.)
  - $M(0) = \sum_{x \in S} f(x) = 1$
  - If the space of S is the set of mutually-exclusive, exhaustive events $\{b_1, b_2, b_3, \ldots\}$, then $M(t) = f(b_1)e^{b_1 t} + f(b_2)e^{b_2 t} + f(b_3)e^{b_3 t} + \cdots$; use this pattern to identify probabilities of the outcomes given the m.g.f., or v.v.
  - Note that the m.g.f. is unique to a distribution; if a m.g.f. exists, there exists one and only one distribution of probability associated with it
  - If moments are given (i.e., formula for $E(X^r), r = 1, 2, \ldots$), then usually can find closed-form m.g.f. using Maclauren series
- Differentiating the m.g.f. r times gives the r^{th} moment of the distribution (around 0) at $t = 0$
  - $M'(t) = \sum_{x \in S} x e^{xt} f(x) \Rightarrow M'(0) = E(X) = \mu$
  - $M''(t) = \sum_{x \in S} x^2 e^{xt} f(x) \Rightarrow M''(0) = E(X^2) \Rightarrow M''(0) - M'(0)^2 = \sigma^2$

- (and so on)
- Using these formulas can be used to easily find μ, σ^2 for the binomial distribution (with m.g.f. $M(t) = \left[(1-p)+pe^t\right]^n$

- <u>Negative binomial distribution</u> is a distribution where X is number of trials until r successes occur
  - p.m.f.: $f(x) = \binom{x-1}{r-1} p^r q^{x-r}, x = r, r+1, \ldots$
  - m.g.f: $M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, (1-p)e^t < 1$ (don't have to memorize this, but this is used to find the μ, σ^2)
    - $\mu = \frac{r}{p}$, $\sigma = \frac{rq}{p^2}$
  - If $r = 1$, then called <u>geometric distribution</u>, p.m.f. $f(x) = pq^{x-1}, x = 1, 2, \ldots$
    - The sum of the p.m.f. can easily be verified to equal 1 using the infinite sum of a geometric series formula
    - $P(X > k) = q^k$
    - c.d.f: $F(x) = P(X \le x) = 1 - q^x$
    - $\mu = p^{-1}$, $\sigma^2 = \frac{q}{p^2}$

Summary of Counting Methods (from 1.3)

	With replacement	**Without replacement**
Ordered	n^r	${}_nP_r$
Unordered	${}_{n-1+r}C_r$	${}_nC_r$

Summary of Distributions

Name	**p.m.f.**	**m.g.f**	μ	σ^2	**Use**
Hypergeometric	$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$, $x = 0, 1, \ldots N_1, x \le n \le N$	doesn't look fun on Wikipedia	np	$npq\left(\frac{N-n}{N-1}\right)$	How many of first type ("successes") when drawing w/o replacement
Binomial $b(n, p)$	$f(x) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, \ldots n$	$M(t) = \left[(1-p)+pe^t\right]^n$	np	npq	How many successes in n Bernoulli trials
Bernoulli (binomial, $n = 1$)	$f(x) = p^x q^{1-x}$, $x = 0, 1$	$M(t) = (1-p)+pe^t$	p	pq	Chance of success in 1 Bernoulli trial
Negative binomial	$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$, $x = r, r+1, \ldots, r \ge 1$	$M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}$, $(1-p)e^t < 1$	$\frac{r}{p}$	$\frac{rq}{p^2}$	How many Bernoulli trials until r successes
Geometric (negative binomial,	$f(x) = pq^{x-1}$, $x = 1, 2, \ldots$	$M(t) = \frac{pe^t}{1-(1-p)e^t}$, $(1-p)e^t < 1$	$\frac{1}{p}$	$\frac{q}{p^2}$	How many Bernoulli trials until first success

$r = 1)$					