STAT 230 Midterm Review

Zhilin (Catherine) Zhou

February 28, 2020

Contents

1 Intro **2**

2 Discrete random variables and Continuous random variables **2**

3 Coungting Techniques **2**
3.1 Permutation 2
3.2 Combination 3

4 Probability Rules **3**
4.1 Independent and Dependent Events 4
4.2 Conditional Probability 4
4.3 Series and Sums 5

5 Discrete Random Variables **6**
5.1 Discrete Random Variables 6
5.1.1 Probability function 6
5.1.2 Cumulative Density Function 7
5.2 Continuous Random Variables 7
5.2.1 Probability Density Function 7
5.2.2 Cumulative Density Function 7
5.3 Summary of Probability Functions for Discrete Random Variables 8
5.3.1 Physical Setup 8
5.4 Combining Other Models with the Poisson Process 9

6 Expected Value and Variance **11**

7 Continuous Variable **13**
7.0.1 Cumulative Distribution Function 13
7.0.2 Probability Density Function 13
7.0.3 Defined Variables or Change of Variable 14
7.0.4 Expectation, Mean, and Variance 14

1 Intro

Classical Definition:

$$\frac{\text{number of ways the } \textbf{event} \text{ can occur}}{\text{number of outcomes in } \textbf{S}}$$

All points in the sample space S are equally likely.

Relative Frequency:
The (limiting) **proportion** (or fraction) **of times** the event occurs in a very long series of repetitions of an experiment or process.

Subjective Probability:
A measure of **how sure** the person making the statment is that the event will happen.

Note: All these definitions have serious **limitations**:

- Classical Definition: hard to define **"equally likely"**
- Relative Frequency: we can never repeat an experiment **indefinitely**, or obtain a long series of **repetitions** due to time, cost, or other limitations
- Subjective Probability: it gives no rational basis for people to agree on a right answer

2 Discrete random variables and Continuous random variables

Definition 2.0.1. ***Discrete***
S is discrete if it consists of a finite or countably infinite set of simple events. A countably infinite sequence is one that can be put into a one-to-one correspondence with the positive integers.

Definition 2.0.2. ***Simple event***
An event in discrete sample space is a subset $A \subset S$, if the event is indivisible so it contains only one point, we call it simple event.

Definition 2.0.3. ***Compound event***
An event A made up of two or more simple events such as is called a compound event.

3 Coungting Techniques

Theorem 3.1. ***Addition Rule***
Suppose we can do job 1 in P ways and job 2 in q ways. Then we can do either job 1 ***OR*** *job 2 (but not both), in $p + q$ ways.*

Theorem 3.2. ***Multiplication Rule***
Suppose we can do job 1 in p ways, ***for each of these ways****, we can do job 2 in q ways. Then we can do both job 1* ***AND*** *job 2 in $p \times q$ ways.*

3.1 Permutation

Generalization:

- $n \times (n-1) \times \cdots \times 1$ arrangements of length n using each symbol once and only once. This product is denoted by $n!$

- $n \times (n-1) \times \cdots \times (n-k+1)$ arragements of length k using each symbol at most once. This product is denoted by $n^{(k)}$
- $n \times n \times \cdots \times n = n^k$ arrangements of length k using each symbol as often as we wish

3.2 Combination

Properties: should be able to prove

1. $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
2. $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!}$
3. $\binom{n}{k} = \binom{n}{n-k}$ for all $k = 0, 1, \ldots, n$
4. If we define $0! = 1$, then the formulas hold with $\binom{n}{0} = \binom{n}{n} = 1$
5. $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
6. **Binomial Theorem**: $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$

4 Probability Rules

1. $P(S) = 1$
2. For any event A, $0 \leq P(A) \leq 1$
3. If A and B are two events with $A \subseteq B$, then $P(A) \leq P(B)$
4. (a) **Addition Law of Probability or the Sum Rule**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

   (b) **Probability of the Union of Three Events**

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

   (c) **Probability of the Union of n Events**

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = &\sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) \\ &- \sum_{i<j<k<l} P(A_iA_jA_kA_l) + \cdots \end{aligned}$$

   where the subscripts are all distince, for example $i < j < k < l$

5. (a) **Probability of the Union of Two Mutually Exclusive Events**
   Let A and B be mutually exclusive events. Then

$$P(A \cup B) = P(A) + P(B)$$

   (b) **Probability of the Union of n Mutually Exclusive Events**
   In general, let $A_1, A_2, \ldots, A_n$ be mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

6. **Probability of the Complement of an Event**

$$P(A) = 1 - P(\bar{A})$$

7. **Product Rules**
Let A, B, C, D... be arbitrary events in a sample space. Assume that $P(A) > 0$, $P(A \cap B) > 0$, and $P(A \cap B \cap C) > 0$, then

$$\begin{aligned} P(A|B) &= P(A)P(B|A) \\ P(ABC) &= P(A)P(B|A)P(C|AB) \\ P(ABCD) &= P(A)P(B|A)P(C|AB)P(D|ABC) \\ &\cdots \end{aligned}$$

8. Let $A_1,\ A_2, \cdots A_k$ be a partition of the sample space S into disjoint events, that is

$$A_1 \cup A_2 \cup \cdots \cup A_k = S \text{ and } A_i \cap A_j = \varnothing \text{ if } i \neq j$$

Let B be an arbitrary event in S. Then

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \cdots + P(BA_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned}$$

Theorem 4.1. ***De Morgan's Laws***

(a) $A \bar{\cup} B = \bar{A} \cap \bar{B}$

(b) $A \bar{\cap} B = \bar{A} \cup \bar{B}$

4.1 Independent and Dependent Events

Definition 4.1.1. ***Independent and Dependent Events***
Events A and B are independent events iff

$$P(A \cap B) = P(A)P(B)$$

If the events are not independent, they are dependent

4.2 Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ provided } P(B) > 0$$

Theorem 4.1. *A and B are two events defined on a sample space S s.t. $P(A) > 0,\ P(B) > 0$. A and B are **independent** events iff*

$$P(A|B) = P(A) \text{ or } P(B|A) = P(B)$$

Example 4.2.1. ***Testing for HIV***
Tests used to diagnose medical conditions are often imperfect, and give false positive or false negative results, as described in Problem 2.6 of Chapter 2. A fairly cheap blood test for the Human Immunodeficiency Virus (HIV) that causes AIDS (Acquired Immune Deficiency Syndrome) has the following characteristics: the false negative rate is 2% and the false positive rate is 0.5%. It is assumed that around 0.04% of Canadian males are infected with HIV. Find the probability that if a male tests positive for HIV, he actually has HIV.

Solution. We can define

$$A = \text{male has HIV}$$
$$B = \text{blood test is positive}$$

We need to find the value of $P(A|B)$ (when we know that the blood test is positive, it is a male). We know that

$$P(A) = 0.0004 \qquad P(\overline{A}) = 0.9996$$
$$P(B|A) = 0.98 \qquad P(B|\overline{A}) = 0.05$$

Since by Theorem 4.2, $P(AB) = P(A)P(B|A)$, we can conclude that

$$P(AB) = 0.0004 \times 0.98 \approx 0.000392$$
$$P(\overline{A}B) = P(\overline{A})P(B|\overline{A}) = 0.9996 \times 0.05 \approx 0.004998$$
$$P(B) = P(AB) + P(\overline{A}B) \approx 0.00539$$
$$P(A|B) = \frac{P(AB)}{P(B)} \approx 0.0727$$

■

Theorem 4.2. ***Bayes' Theorem***
Suppose A and B are events defined on a sample space S with $P(B) > 0$, we have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(AB) + P(\overline{A}B)}$$

4.3 Series and Sums

1. **Geometric Series**

$$\sum_{i=0}^{n-1} t^i = 1 + t + t^2 + \cdots + t^{n-1} = \frac{1-t^n}{1-t} \text{ for } t \neq 1$$

If $|t| < 1$, then

$$\sum_{x=0}^{\infty} t^x = 1 + t + t^2 + \cdots = \frac{1}{1-t}$$

Note: other identities can be obtained from this one by **differentiation**.

2. **Binomial Theorem**

$$(1+t)^n = 1 + \binom{n}{1}t^1 + \binom{n}{2}t^2 + \cdots + \binom{n}{n}t^n = \sum_{x=0}^{n}\binom{n}{x}t^x$$

n is a positive integer and t is any real number.
If $|t| < 1$, then

$$(1+t)^n = \sum_{x=0}^{n}\binom{n}{x}t^x$$

3. **Multinomial Theorem**
   A generalization of the Binomial Theorem:

$$(t_1 + t_2 + \cdots + t_k)^n = \sum \frac{n!}{x_1!x_2!\dots x_k!} t_1^{x_1} t_2^{x_2} \dots t_k^{x_k}$$

   where x_i are all non-negative integers s.t. $\sum_{i=1}^{k} x_1 = n$ where n is a positive integer.

4. **Hypergeometric Identity**

$$\sum_{x=0}^{\infty} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

   Note: There will not be an infinite number of terms if a and b are positive integers since the terms become 0 eventually.

$$\binom{4}{5} = \frac{4^{(5)}}{5!} = \frac{4 \times 3 \times 2 \times 1 \times 0}{5!} = 0$$

5. **Exponential Series**
   Let $f(x) = e^x$, then $f^{(k)}(0) = 1$ for $k = 1,\ 2, \dots$, therefore,

$$e^t = \frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!} \,\forall\, t \in \mathbb{R}$$

   We can use the limit definition of the exponential function:

$$e^t = \lim_{n \to \infty} (1 + \frac{t}{n})^n$$

6. **Special Series**

$$\begin{aligned} 1 + 2 + \cdots + n &= \frac{n(n+1)}{2} \\ 1^2 + 2^2 + \cdots + n^2 &= \frac{n(n+1)(2n+1)}{6} \\ 1^3 + 2^3 + \cdots + n^3 &= [\frac{n(n+1)}{2}]^2 \end{aligned}$$

5 Discrete Random Variables

5.1 Discrete Random Variables

5.1.1 Probability function

1. The value is always non-negative.
2. Every value of the function lies in the interval $[0, 1]$
3. The domain of the function is countable
4. The sum of hte function over all values of x equals 1.

5.1.2 Cumulative Density Function

1. The value is always non-negative.
2. Every value of the function lies in the interval $[0, 1]$
3. The limit of the function as $x \to \infty$ equals 1.
4. The limit of hte function as $x \to -\infty$ equals 0.
5. The domain of the function is $\mathbb{R}$
6. The fucntion is non-decreasing for all $x \in \mathbb{R}$
7. The function is right-continuous for all $x \in \mathbb{R}$

5.2 Continuous Random Variables

5.2.1 Probability Density Function

1. The value is always non-negative.
2. The domain of the function is $\mathbb{R}$.
3. The area bounded by the graph of the curve of the function and the x-axis equals 1.

5.2.2 Cumulative Density Function

1. The value is always non-negative.
2. Every value of the function lies in the interval $[0, 1]$
3. The limit of the function as $x \to \infty$ equals 1.
4. The limit of the function as $x \to -\infty$ equals 0.
5. The domain of the function is $\mathbb{R}$.
6. The function is non-decreasing for all $x \in \mathbb{R}$.
7. The function is right-continuous for all $x \in \mathbb{R}$.
8. The function is continuous for all $x \in \mathbb{R}$.

Definition 5.2.1. ***Probability distribution***
Let $S = \{a_1, a_2, \ldots\}$ be a discrete sample space, assign probabilities $P(a_i), i = 1, 2, 3, \ldots$ to the a_i's such that the following two conditions hold:

1. $0 \leq P(a_i) \leq 1$

2. $\sum_{\textit{all } a_i} P(a_i) = 1$

The set of probabilities $\{P(a_i), i = 1, 2, \ldots\}$ is called a probability distribution on S.

Definition 5.2.2. ***Comulative Distribution Function***
c.d.f. of X is usually denoted by F(X)

$$F(X) = P(X \leq x) \textit{ defined for all } x \in \mathbb{R}$$

In general, F(x) can be obtained from f(x) using

$$F(X) = P(X \leq x) = \sum_{u \leq x} f(u)$$

1. *F(x) is a non-decreasing function of x for all* $x \in \mathbb{R}$
2. $0 \leq F(X) \leq 1$ *for all* $x \in \mathbb{R}$
3. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$
4. *If* $x \in A$ *and* $x - 1 \in A$, $f(x) = F(x) - F(x-1) = P(X = x)$

5.3 Summary of Probability Functions for Discrete Random Variables

Notation	Probability Function	Expected Value	Variance
$Hypergeometric(N, r, n)$ $r < N, n < N$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-(N-r)), \ldots, \min(n, r)$	$\frac{nr}{N}$	$\frac{nr}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$
$Binomial(n, p)$ $0 < p < 1, q = 1 - p$	$f(x) = \binom{n}{x}p^x q^{n-x}$ $x = 0, 1, 2, \ldots, n$	np	npq
$NP(k, p)$ $0 < p < 1, q = 1 - p$	$f(x) = \binom{x+k-1}{x}p^k q^x = \binom{-k}{x}p^k(-q)^x$ $x = 0, 1, 2, \ldots$	$\frac{kq}{p}$	$\frac{kq}{p^2}$
$Geometric(p)$ $0 < p < q, q = 1 - p$	$f(x) = pq^x$ $x = 0, 1, 2, \ldots$	$\frac{q}{p}$	$\frac{q}{p^2}$
$Poisson(\mu)$ $\mu > 0$	$f(x) = \frac{\mu^x e^{-\mu}}{x!}$ $x = 0, 1, 2, \ldots$	μ	μ

5.3.1 Physical Setup

1. **Hypergeometric**
   We have a collection of N objects which can be classified into two distinct types. Call one type "success" (S) and the other type "failure" (F). There are r sucesses and $N - r$ failures. Pick n objects at random **without replacement**. Let X be the number of successes obtained. Then X has a Hypergeometric distribution.

2. **Binomial**
   Suppose an "experiment" has two types of distinct outcomes – "success" (S) and "failure" (F), and let their probabilities be p (for S) and $1 - p$ (for F). Repeat the experiment n **independent** times. Let X be the number of successes obtained. Then X has a **Binomial distribution**. We write $X \sim Binomial(n, p)$ as a shorthand for "X is distributed according to a Binomial Distribution with n repetitions and probability p of success".

   The n individual experiments are often called "trials" or "Bernoulli trials" and the process is called a "Bernoulli process" or a "Bionomial Process".

3. **Negative Binomial**
   It is almost the same as Bionomial: an experiment (trial) has two distinct types of outcome (S and F) and is **repeated independently** with the same probability p, of success at each time. Continue doing the experiment until a specified number, k, of success have been obtained. Let X be the number of failures obtained before the k'th success. Then X has a negative Binomial Distribution. We can denote it as $X \sim NegativeBionmial(k, p)$.

   Definition 5.3.1. ***Bernoulli Trials***

   the Binomial, Negative Binomial and Geometric models all involve trials (experiments) which:

(a) are independent

(b) have 2 distinct types of outcome (S and F)

(c) have the same probability p of "success" (S) each time.

4. **Gemoetric**
Consider the Negative Binomial Distribution with $k = 1$. In this case we repeat independent Bernoulli trials with two types of outcomes (S and F) each time, and the same probability, p, of success each time until we obtain the first success. Let X be the number of failures obtained before the first sucess. We write $X \sim Geometric(p)$

5. **Poisson**
It is as a limiting case of the Binomial distribution as $n \to \infty$ and $p \to 0$. In particular, we keep the product np fixed at some constant value, μ, while letting $n \to \infty$. This automacially makes $p \to 0$. Let us see what the limit of the Binomial probability function $f(x)$ in this case.

   Consider a situation in which a certain type of event occurs at random points in time (or space) according to the following conditions:

   (a) **Independence**: the number of occurences in non-overlapping intervals are independent.

   (b) **Individuality**: for sufficiently short time periods of length Δt, the probability of 2 or more events occuring in the interval is close to zero, i.e. events occur singly not in clusters. More precisely, as $\Delta t \to 0$, the probability of two or more events in the interval of length Δt must go to zero faster than $\Delta t \to 0$ or
   $$P(\text{2 or more events in } (t, t + \Delta t)) = o(\Delta t) \text{ as } \Delta t \to 0$$

   (c) **Homogeneity or Uniformity**: events occur at a uniform or homogeneous rate λ over time so that the probability of one occurrence in an interval $(t, t + \Delta t)$ is approximately $\lambda\Delta t$ for small Δt for any value t. More precisely,
   $$P(\text{one event in } (t, t + \Delta t)) = \lambda\Delta t + o(\Delta t)$$

   Suppose a process satisfies the three conditions above, then assume events occur t the average rate of λ per unit time. Let X be the number of times an event occur in a time period of t units, then $X \sim Poisson(\mu = \lambda t)$

Interpretation of μ and λ

1. μ refers to the **intensity** or **rate of occurrence** parameter for the events.
2. $\lambda t = \mu$ represents the **average number of occurrences** in t units of time.
3. IMPORTANT: the value of λ **depends on** the units used to measure time.

5.4 Combining Other Models with the Poisson Process

Example:

A very large (essentially infinite) number of ladybugs is released in a large orchard. They scatter randomly so that on average a tree has 6 ladybugs on it.Trees are all the same size:

1. Find the probability a tree has > 3 ladybugs on it.

Since the ladybugs are **randomly scattered**, we can use the Poisson distirbution to solve this. In this case, $\lambda = 6$ and $v = 1$ (i.e. any tree has a "volume" of one unit), so $\mu = 6$ and

$$\begin{aligned} P(X > 3) = 1 - P(X \leq 3) &= 1 - [f(0) + f(1) + f(2) + f(3)] \\ &= [\frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!} + \frac{6^3 e^{-6}}{3!}] \\ &= 0.8488 \end{aligned}$$

2. When 10 trees are picked at random, what is the probability 8 of these trees have > 3 ladybugs on them?

   We can use the Binomial Distribution where "success" means to have > 3 ladybugs on a tree. We have $n = 10$, and

$$f(8) = \binom{10}{8}(0.8488)^8(1 - 0.8488)^{10-8} = 0.2772$$

3. Trees are checked until 5 wtih > 3 ladybugs are found. Let X be the total number of trees checked. Find the probability function, $f(x)$

   We can use the Negative Binomial Distribution. We need the number of success, k, to be 5, and the number of failures to be $(x - 5)$. Then,

$$f(x) = \binom{x-1}{4}(0.8488)^5(1 - 0.8488)^{x-5} \text{ for } x = 5, 6, 7, \ldots$$

4. Find the probability a tree with > 3 ladybugs on it has exactly 6.

   This is a conditional probability. Let A = {6 ladybugs} and B = {a tree with > 3 ladybugs}. Then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{6^6 e^{-6}}{6!}}{0.8488} = 0.1892$$

5. On 2 trees there are a total of t ladybugs. Find the probability that x of these are on the first of these 2 trees.

   This is also a conditional probability. Let A ={x on the first tree}, B ={$t - x$ on second tree}, and C ={2 trees has a total of t ladybugs}. Then we have

$$\begin{aligned} P(X = x) = P(A|C) &= \frac{P(AC)}{P(C)} = \frac{P(AB)}{P(C)} \\ &= \frac{P(A)P(B)}{P(C)} \end{aligned}$$

   We can use Poisson Distribution at this point to calculate each, with $\mu = 6 \times 2 = 12$in the denominator since there are 2 trees.

$$\begin{aligned} P(A|C) &= \frac{(\frac{6^x e^{-6}}{x!})(\frac{6^{t-x} e^{-6}}{(t-x)!})}{\frac{12^t e^{-12}}{t!}} \\ &= \frac{t!}{x!(t-x)!}\left(\frac{6}{12}\right)^x\left(\frac{6}{12}\right)^{t-x} \\ &= \binom{t}{x}\left(\frac{1}{2}\right)^x\left(1 - \frac{1}{2}\right)^{t-x} \quad \text{for } x = 0, 1, \ldots, t \end{aligned}$$

6 Expected Value and Variance

Definition 6.0.1. ***Sample***
A set of observed outcomes $x_1, \ldots, x_n$ for a random variable X

Definition 6.0.2. ***Sample Mean***
We refer to the fact that this is the average for a particular sample.

Unless somebody deliberately "cooked" the study, we would not expect to get precisely the sample mean if we repeated it another time.

Definition 6.0.3. ***Median***
A value such that half the results are below it and half above it, when the results are arranged in numerical order.

Definition 6.0.4. ***Mode***
The value which occurs most often. There is no guarentee there will only be a single mode.

Definition 6.0.5. ***Expected Value***
*Let X be a discrete random variable with $\mathrm{range}(X)$ = A and probability function $f(x)$. The **expected value** (also called the mean or the expectation) of X is given by*

$$E(X) = \sum_{x \in A} x f(x)$$

Theorem 6.1. *Let X be a discrete random variable with $\mathrm{range}(X)$ = A and probability function $f(x)$. The **expected value** of some function $g(X)$ of X is given by*

$$E[g(X)] = \sum_{x \in A} g(x) f(x)$$

Notes:

1. You can interpret $E[g(X)]$ as the average value of $g(X)$ in an infinite series of repetitions of the process where X is defined.
2. $E[g(x)]$ is also known as the "expected value" of $g(X)$. However, this value can be a value $g(X)$ never takes.
3. When calculating expectations, look at your answer to be sure it makes sence. Suppose for example that X takes values from 1 to 10. Then since

$$1 = \sum_{x=1}^{10} (1) P(X = x) \leq \sum_{x=1}^{10} x P(X = x) = E(X) \leq \sum_{x=1}^{10} (10) P(X = x) = 10(1) = 10$$

   you should know you've made an arror if you get $E(X) > 10$ or $E(X) < 1$. In physical terms, $E(X)$ is the balance point for the probability histogram of $f(x)$

Example 6.0.1. ***Diagnostic Medical Tests***
Often there are cheaper, less accurate tests for diagnosing the presence of some conditions in a person, along with more expensive, accurate tests. Suppose we have two cheap tests and one expensive test, with the following characteristics. All three tests are positive if a person has the condition (there are no "false negatives"), but the cheap tests give "false positives". Let a person be chosen at random, and let D = {person has the condition}. For the three tests the probability of a false positive and cost are:

Test	*P(positive test$\|\bar{D}$)*	*Cost (in dollars)*
1	*0.05*	*5*
2	*0.03*	*8*
3	*0*	*40*

We want to check a large number of people for the condition, and have to choose among three testing strategies:

1. *Use Test 1, followed by Test 3 if Test 1 is positive*
2. *Use Test 2, followed by Test 3 if Test 2 is positive*
3. *Use Test 3*

Determine the expeceted cost per person under each of strategies 1, 2, and 3. We will then choose the strategy with the lowest expecetd cost. It is known that about 0.001 of the population have the condition

$$P(D) = 0.001, \qquad P(\bar{D}) = 0.999$$

Assume that given D or $\bar{D}$, tests are independent of one another.

Solution. For a person tested chosen at random and tested, define the random variable X as follows:

$$\begin{aligned} X &= 1 \qquad \text{if the initial test is negative} \\ X &= 2 \qquad \text{if the initial test is positive} \end{aligned}$$

Let $g(x)$ be the total cost of testing the person, the expected cost per person is then

$$E[g(X)] = \sum_{x=1}^{2} g(x) f(x)$$

The probability function $f(x)$ for X and the function $g(x)$ differ for strategies 1, 2, 3. Consider for example strategy 1. Then

$$\begin{aligned} P(X = 2) &= P(\text{initial test positive}) \\ &= P(D) + P(\text{positive}|\bar{D})P(\bar{D}) \\ &= 0.001 + (0.005)(0.999) \\ &= 0.0510 \end{aligned}$$

The rest of the probabilities, associated with the values of $g(X)$ and $E[g(X)]$ are obtained below.

1. Strategy 1

$$\begin{aligned} f(2) &= 0.0510 \text{ obtained above} \\ f(1) &= P(X = 1) = 1 - f(2) = 1 - 0.0510 = 0.949 \\ g(1) &= 5 \qquad g(2) = 45 \\ E[g(X)] &= 5(0.949) + 45(0.0510) = \$7.04 \end{aligned}$$

2. Strategy 2

$$\begin{aligned} f(2) &= 0.001 + (0.03)(0.999) = 0.03097 \\ f(1) &= 1 - f(2) = 0.96903 \\ g(1) &= 8 \qquad g(2) = 48 \\ E[g(X)] &= 8(0.96903) + 48(0.03097) = \$9.2388 \end{aligned}$$

3. Strategy 3

$$f(2) = 0.001, \qquad f(1) = 0.999$$
$$g(2) = g(1) = 40$$
$$E[g(X)] = \$40.00$$

Therefore, the cheapest strategy is strategy 1. ■

7 Continuous Variable

7.1 General Notation

7.1.1 Cumulative Distribution Function

Properties of a cumulative distribution function are the same for continuous variables as for discrete variables.

1. $F(x)$ is defined for all real x.
2. $F(x)$ is a non-decreasing function of x for all real x.
3. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
4. $P(a < X \leq b) = F(b) - F(a)$

7.1.2 Probability Density Function

Definition 7.1.1. *The* ***probability density function*** *(p.d.f) $f(x)$ for a continuous random variable X is the derivative*

$$f(x) = \frac{dF(x)}{dx}$$

where $F(x)$ is the cumulative distribution function for X

If the derivative of F does not exist at $x = a$ we usually define $f(a) = 0$ for convenience. Assume $f(x)$ is a continuous function of x at all points for which $0 < F(x) < 1$

Properties of a probability density function

1. $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$ (This follows from the definition of $f(x)$)
2. $f(x) \geq 0$ (since $F(x)$ is non-decreasing, its derivative is non-negative)
3. $\int_{-\infty}^{\infty} f(x)dx = \int_{\text{all } x} f(x)dx = 1$ (This is because $P(-\infty \leq X \leq \infty) = 1$)
4. $F(x) = \int_{-\infty}^{x} f(u)du$ (This is just property 1 with $a = -\infty$)

Definition 7.1.2. ***Quantiles and Percentiles***
Suppose X is a continuous random variable with cumulative distribution fucntion $F(x)$. The pth quantile of X (or the pth quantile of the distribution) is the value $q(p)$, such that

$$P[X \leq q(p)] = p$$

or

$$F(q(p)) = p$$

The value $q(p)$ is also called the 100th percentile of the distribution. If $p = 0.5$ then $m = q(0.5)$ is called the median of X or the median of the distribution.

7.1.3 Defined Variables or Change of Variable

Sometimes we want to find the probability density function or comulative distribution function for some other random variable Y which is a function of X. It is based on the fact that the cumulative distribution function $F_Y(y)$ for Y equals $P(Y \leq y)$, and this can be rewritten in terms of X since Y is a function of X. Thus:

1. Write the comulative distribution function of Y as a function of X
2. Use $F_X(x)$ to find $F_Y(y)$. Then if you want the probability density function $f_Y(y)$, you can differentiate the expression for $F_Y(y)$.
3. Find the range of values of y.

7.1.4 Expectation, Mean, and Variance

Definition 7.1.3. *When X is a continuous random variable we define*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$