Stat 4202: Mathematical Statistics II

Lecturer: **Professor Andrew Kerr**

Notes by: Farhan Sadeek

Spring 2025

1 January 6, 2025

STAT 4202 will rely a lot on STAT 4201. So we need to have a pretty good understanding of those concepts.

1.1 Review of Probability Theory

Definition 1

The **Sample Space**, denoted by $\mathcal{S}$, is the set of all outcomes from an experiment.

Definition 2

An **Event**, usually denoted by a capital letter such as A or B, is a subset of the Sample Space.

The probability function

- $P(A) \geq 0$
- $P(\mathcal{S}) = 1$
- For disjoint sets $A_1, A_2, \cdots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

If an event A is a subset of another event B, then the probability of A is less than or equal to the probability of event B. That is to say, if $A \subseteq B$, then $P(A) \leq P(B)$

The complement of an event A, denoted by A^c, has a probability equal to one minus the probability of the event A. That is,

$$P(A^c) = 1 - P(A)$$

A partition of a sample space $\mathcal{S}$ is an exhaustive, non-overlapping collection of events $A_1, A_2, \cdots, A_n$ that is exhaustive and mutually exclusive:

$$\bigcup_{i=1}^{n} A_i = \mathcal{S}$$

and

$$A_i \cap A_j = \emptyset \quad \forall i \neq j$$

For any partition, we have

$$\sum_{i=1}^{n} P(A_i) = 1$$

Two events A and B are **independent** if the outcome of one doesn't affect the likelihood of the occurrence of the other. For two independent events, we have

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

The **conditional probability** of A given B is given by

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

Lemma 3

Note that if A and B are independent, then

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(B)}{P(B)} \\ &= P(A) \end{aligned}$$

Corollary 4

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$

1.2 Random Variables

Definition 5

A random variable is a function that takes outcomes from the sample space $\mathcal{S}$ to the real numbers $\mathbb{R}$. That is, a random variable is a function $X : \mathcal{S} \to \mathbb{R}$.

We then use a probability mass function (pmf) in the discrete case or a probability density function (pdf) in the continuous case:

pmf:	$f_X(x) = P(X = x)$	when X is discrete
pdf:	$\int_a^b f_X(x)\mathrm{d}x = P(a \leq X \leq b)$	when X is continuous

The cumulative distribution function (cdf) gives the probability of observing a value less than or equal to a given value x:

$$F_X(x) = P(X \leq x)$$

When X is a continuous random variable, the pdf is the derivative of the cdf:

$$f_X(x) = F'_X(x)$$

1.3 Expected Value and Variance

For random variable X, the **expected value** is denoted by $E(X)$ and is given by:

$$E(X) = \begin{cases} \sum_x x f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

The **variance** of a random variable X is denoted by $\mathrm{Var}(X)$ and is given by:

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right]$$

1.4 Covariance

The **covariance** of two random variables X and Y is denoted by:

$$\mathrm{Cov}(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

If two random variables X and Y are independent, then

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

So, we will be using these formulas to estimate the mean and the variance throughout the semester.

2 January 8, 2025

2.1 Statistical Models

In statistics, we often model data $X_1, X_2, \cdots, X_n$ as a random sample from a population. We assume that the data are independent and identically distributed (iid) random variables. The goal is to estimate the parameters of the population distribution.

Definition 6

A **parameter** of a distribution are values that describe a certain characteristic of the given distribution.

Some examples of parameters include:

- The mean height of **all** OSU incoming freshmen.
- The proportion of registered voters that voted for a particular candidate.
- The standard deviation of waiting times for **all** customers shopping at a store during a week.

Fact 7

If $X_1, X_2, \dots, X_n \overset{iid}{\approx} f_X(x)$ then $\mu = E(X_i)$ is a parameter, which is the mean of the distribution. The variance is also a parameter: $\sigma^2 = E[(X - \mu)^2]$

Example 8

Suppose we are examining the efficacy difference between a newly developed drug and an existing drug. We look at the differences, Δi, from a series of n comparative samples. Note that these will all come from some distribution:

$$\Delta_1, \Delta_2, \dots, \Delta_n, \overset{iid}{f}(x)$$

Fact 9

Here the independed is a really important to look for we will look thorogh that through the semester.

For a parametric model

$$\{f_g(x)_{\theta \in \mathbb{R}}\}$$

Which is indexed by a vector θ of parameters.

Example 10

Suppose we wanted to estimate the height and weight of all incoming students at Ohio State. We could take a random sample of n of the incoming students and observe the height (H) and weight (W) of each student, giving the following sample data:

$$(H_1, W_1), (H_2, W_2), \dots, (H_n, W_n)$$

We can then consider the following model:

$$N(\mu, \Sigma)$$

3 January 8, 2025

We went over the **Recitation Logistics** and **Quiz 1**.

4 January 10, 2025 (In-Person)

We wanted to check how to get estimators. We will do the backwards this week for.

4.1 Unbiased Estimator

Definition 11

An **estimator** $\hat{(}\theta)$