Lecture 12: Midterm Review

- Brief review of probability
- Practice questions

Sample Space

- Ω: sample space = set of all possible outcomes
- $\omega \in \Omega$: a particular outcome

Let A and B be collections of outcomes or events

Example: Dice

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
$$A = \{1, 2, 3\}$$

Set Theory

- $A \cup B = \{\omega : \omega \in A \quad \text{or} \quad \omega \in B\}$ (union)
- $A \cap B = \{\omega : \omega \in A \quad \text{and} \quad \omega \in B\}$ (intersection)
- $A^c = \{\omega : \omega \notin A\}$ (complement)
- $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$ (De Morgan's laws)

Events are **mutually exclusive** if

$$A \cap B = \emptyset \quad \text{(empty set)}$$

Definition of Probability

Let $\mathcal{F}$ be a well-defined collection of subsets of Ω,

1. $P(A) \geq 0$, for every event $A \in \mathcal{F}$
2. $P(\Omega) = 1$
3. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

Example: Dice

$$\begin{aligned} \Omega &= \{1,2,3,4,5,6\} \\ P(\{\omega = i\}) &:= 1/6, \text{ for } \ i = 1,2,3,4,5,6 \\ P(\omega = \{1,2\}) &= P(\omega = \{1\}) + P(\omega = \{2\}) \\ &= 1/3 \end{aligned}$$

Conditional Probability

Given two events, A and B, what is the probability that A will occur given that B has occurred?

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

Independent Events

Two events are said to be **independent** if

$$P(A|B) = P(A)$$

Equivalently, A and B are independent if

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

Example: Dice

Suppose that A is an event related to the roll of a die and B is an event related to the current temperature.

Bayes' Rule

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

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

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

Random Variables

A **random variable** X is a mapping of Ω to the real numbers.

Example: Digital Thermometer

$$\Omega = \{\text{weather patterns}\}$$
$$X = \text{mercury level on a thermometer}$$

Continuous Distributions

When X takes values in real numbers, the probability distribution is completely described by the **cumulative distribution function**:

$$F(x) = P(X \leq x) \equiv P(\omega : X(\omega) \leq x)$$

If $F(x)$ is differentiable, then the **probability density function** is the function $f(x)$ satisfying

$$F(x) = \int_{-\infty}^{x} f(u)du$$

Example: Uniform on [0, 1]

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

$$f(x) = 1, \quad 0 \leq x \leq 1$$

Discrete Distributions

Discrete random variables: mapping Ω to an enumerable set of values.

The **probability mass function**

$$p(k) = P(X = k)$$

plays the role of a density, and the probability distribution function

$$P(x) = \sum_{k \leq x} p(k)$$

plays the role of a cumulative distribution function.

Example: Poisson Random Variable

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad \lambda > 0, \quad k = 0, 1, ...$$

$$P(n) = \sum_{k=0}^{n} p(k)$$

Expectation

The expected value of a random variable X is defined to be

$$E(X) = \int_{\mathcal{X}} x f(x) dx, \quad \text{(continuous case)}$$

$$E(X) = \sum_{k \in \mathcal{X}} k P(X = k), \quad \text{(discrete case)}$$

More generally, if g is an arbitrary function, then

$$E(g(X)) = \int_{\mathcal{X}} g(x) f(x) dx, \quad \text{(continuous case)}$$

$$E(g(X)) = \sum_{k \in \mathcal{X}} g(k) P(X = k), \quad \text{(discrete case)}$$

Example: Second Moment

$$g(x) = x^2 \quad \rightarrow \quad E[X^2]$$

Example: Indicator Function

$$g(x) = I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \quad \rightarrow \quad E[I_A(x)] = P(x \in A)$$

Bivariate Random Variables

Joint CDF and PDF

$$F(x,y) = P(\{X \leq x\} \cap \{Y \leq y\})$$

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(x,y) dx dy$$

Independent random variables

$$F(x,y) = F(x)F(y), \quad \forall x, y$$

or

$$p(x,y) = p(x)p(y), \quad \forall x, y$$

Uncorrelated random variables: $E(XY) = E(X)E(Y)$

Conditional Distribution

Conditional density:

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

If X and Y are independent, then $p(x|y) = p(x)$

Conditional expectation:

$$E[X|Y] = \int_{\mathcal{X}} xp(x|y)dx$$

Note: $E[X|Y]$ is a function of y, the value that the random variable Y takes.

Practice 1

Remove all spades from a standard deck of 52 cards. Now shuffle and deal three cards (without replacement).

1. What is the probability that the first card is a King or a club? 0.3846
2. What is the probability that all three cards belong to Aces, Kings, Queens, or Jacks? 0.0241
3. What is the probability that the last card is black given that the first card was red? 0.3422

Practice 2

A club has 8 women and 6 men. In how many ways can a committee of four, consisting of two men and two women, be formed?

$$\binom{8}{2} \times \binom{6}{2} = 28 \times 15 = 420$$

Practice 3

In Canada, about 0.35% of women over 40 will develop breast cancer in any given year. A common screening test for cancer is the mammogram, but this test is not perfect. In about 11% of patients with breast cancer, the test gives a **false negative**: it indicates a woman does not have breast cancer when she does have breast cancer. Similarly, the test gives a **false positive** in 7% of patients who do not have breast cancer: it indicates these patients have breast cancer when they actually do not.[1]

If we tested a random woman over 40 for breast cancer using a mammogram and the test came back positive — that is, the test suggested the patient has cancer — what is the probability that the patient actually has breast cancer?

[1] The probabilities reported here were obtained using studies reported at `www.breastcancer.org` and `www.ncbi.nlm.nih.gov/pmc/articles/PMC1173421`.

Practice 4

Consider the cumulative distribution function $F(x) = x^3$ on the interval $[0, 1]$. For $x > 1$, $F(x) = 1$ and for $x < 0$, $F(x) = 0$

1. What is $E(X)$? The density is $f(x) = 3x^2$ on $[0, 1]$, so the expected value is

$$\int_0^1 x \times 3x^2 dx = 3/4$$

2. What is the standard deviation of x?

$$E(X^2) = \int_0^1 x^2 \times 3x^2 dx = 3/5,$$

and the variance is $V(X) = E(X^2) - [E(X)]^2$, so the standard deviation is

$$\sqrt{V(X)} = \sqrt{3/5 - (3/4)^2} = 0.1936$$

Practice 5

A professor says that about 30% of the class will get A or better for the final grades. In a class of 90, what is the probability that more than 25 people will get grades A or $A+$? 1-0.3632 = 0.6368

Practice 6

Suppose the price of gas per gallon anywhere in the United States has a uniform (continuous) distribution with support between 2 and 5 and zero anywhere else, so that the pdf is $f(x) = 1/3, 2 \leq x \leq 5$. Now let $X_1, X_2, \ldots, X_n$ be a random sample of price of gas per gallon around the country and let Y be the minimum price of gas per gallon.

1. What is the pdf of Y (You must add the support of Y to get full points) $f(y) = \frac{n(5-y)^{n-1}}{3^n}$, for $2 \leq y \leq 5$
2. Let $n = 3$, what is the expected minimum price of gas per gallon, that is $E(Y)$? 2.75
3. What is the standard deviation of the minimum price of gas per gallon? 0.5809
4. What is the value of $E(3Y^2 - 10)$
   $E(3Y^2 - 10) = 3E(Y^2) - 10 = 3(2.75^2 + 0.5809^2) - 10 = 13.6998$

Practice 7

Suppose that an automobile dealer pays an amount X (in thousands of dollars) for a used car and then sells it for an amount Y. Suppose that the random variables X and Y have the following joint p.d.f:

$$f(x,y) = \begin{cases} \frac{1}{36}x & \text{for } 0 < x < y < 6 \\ 0 & \text{otherwise} \end{cases}$$

Determine the dealer's expected gain from the sale. 1.5

Practice 8

Consider a distribution of a random variable Y with expectation $1/2$ and variance $1/12$. Now let $Y_1, Y_2, \ldots, Y_{50}$ be a random sample from that distribution,

1. What is the probability that the sample average is between 0.52 and 0.58? 0.2871
2. What is the probability that the sample average is at least 27.5? 0.1103
3. What is the probability that the sum of the sample is at most 23.4? 0.2166

Practice 9

Let $f(x;\theta) = \theta x^{\theta-1}$ for $0 \leq x \leq 1, \theta > 0$. You draw two random values: $1/4$ and $1/3$.

1. What is the maximum likelihood estimate of θ? 0.8049
2. What is the maximum likelihood estimate of $\sqrt{\theta}$? 0.8971

Good Luck!