UNIVERSITY OF ALBERTA
CMPUT 267 Fall 2024

Midterm Exam 1

Do Not Distribute

Duration: 75 minutes

Last Name: ______________________________

First Name: ______________________________

Carefully read all of the instructions and questions. Good luck!

---

1. **Do not turn this page** until instructed to begin.
2. You got exam version **1**. Please mark **1 in the special code section in coloum J** of your scantron.
3. Verify that your exam package includes 15 pages, along with a formula sheet and a blank page at the end.
4. **Only the scantron will be marked**. All of your answers must be clearly marked on the scantron.
5. Use **pencil only** to fill out the scantron (preferably an HB or #2 pencil).
6. **Erase mistakes completely** on the scantron to avoid misreading by the scanner.
7. **Mark answers firmly and darkly**, filling in the bubbles completely.
8. This exam consists of **25 questions**. Each question is worth **1 mark**. The exam is worth a total of **25 marks**.
9. Some questions may have **multiple correct answers**. To receive **full marks**, you must select **all correct answers**. If you select only **some** of the correct answers, you will receive **partial marks**. Selecting an incorrect option will cancel out a correct one. For example, if you select two answers—one correct and one incorrect—you will receive zero points for that question. If the number of incorrect answers exceeds the correct ones, your score for that question will be zero. **No negative marks** will be given.

Question 1. [1 MARK]

Given $\mathcal{D} = ((x_1, y_1), \ldots, (x_n, y_n))$, the loss function $\ell(\hat{y}, y) = (\hat{y} - y)^2$, and the predictor $f(x) = ax + b$, which of the following represents the total loss calculated by summing $\ell(f(x), y)$ over all data points $(x, y) \in \mathcal{D}$?

A. $\sum_{i=0}^{n}(ax_i + b - y_i)^2$

B. $\sum_{i=1}^{n}(ax_i + b - y_i)^2$

C. $\sum_{i=1}^{n+1}(ax_i + b - y_i)^2$

D. $\sum_{(x,y)\in\mathcal{D}}(ax + b - y)^2$

E. $\sum_{i=0}^{n+1}|ax_i + b - y_i|$

Question 2. [1 MARK]

Which of the following represents the set of all exponential functions of the form $f(x) = ae^{bx} + c$ where $a, b, c \in \mathbb{R}$ and $b \neq 0$?

A. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx} + c,\ a, b, c \in \mathbb{R},\ b \neq 0\}$

B. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx^2} + c,\ a, b, c \in \mathbb{R}\}$

C. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ax^b + c,\ a, b, c \in \mathbb{R}\}$

D. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^x,\ a \in \mathbb{R}\}$

E. $\{f \mid f : \mathbb{R} \to \mathbb{R},\ f(x) = ae^{bx} + cx,\ a, b, c \in \mathbb{R}\}$

Question 3. [1 MARK]

Which of the following is a valid definition of a function $\mathcal{A} : (\mathbb{R}^2 \times \mathbb{R})^2 \to \{f \mid f : \mathbb{R}^2 \to \mathbb{R}\}$?

A. $\mathcal{A}((a, b, c), (d, e, g)) = f$, where $f(x_1, x_2) = dx_1 + ex_2^2 + a$

B. $\mathcal{A}((a, b), (d)) = f$, where $f(x) = ax^2 + bx + d$

C. $\mathcal{A}((a, b), (d)) = f$, where $f(x_1, x_2) = ax_1^2 + bx_2 + dx_1$

D. $\mathcal{A}((a, b, c), (d, e, g)) = f$, where $f(x_1, x_2) = ax_1 + bx_2 + c$

E. $\mathcal{A}((a, b), (d)) = f$, where $f(x_1, x_2) = dx_1 + bx_2^2 + a$

Question 4. [1 MARK]

Let X be a discrete random variable uniformly distributed over the outcome space

$$\mathcal{X} = \{-1, 1, 2, 3, 5\}.$$

The probability mass function (pmf) of X is

$$p(x) = \frac{1}{5} \quad \text{for each } x \in \mathcal{X}.$$

Which of these statements are true?

A. The probability that X is less than 2 is 0.4.

B. The expected value is $\mathbb{E}[X] = 2$.

C. The variance is $\mathrm{Var}(X) = 4$.

D. The probability that X is greater or equal to 3 is 0.6.

E. The probability that X is positive is 0.8.

Question 5. [1 MARK]

Let $h(x, y) = x^2 + y^{-1}$ where $x, y \in \mathbb{R}$. What is

$$\sum_{y \in \mathcal{Y}} \int_{\mathcal{X}} h(x, y)\, dx$$

where $\mathcal{Y} = \{1, 3\}$ and $\mathcal{X} = [0, 3]$?

A. 15

B. 22

C. 22.5

D. 30

E. 48

Question 6. [1 MARK]

Let $h(x,y) = e^x + y$ where $x, y \in \mathbb{R}$. What is

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} h(x,y)\, dy\, dx$$

where $\mathcal{Y} = [0,1]$ and $\mathcal{X} = [0,1]$?

A. $-e + \frac{1}{2}$

B. $-e - \frac{1}{2}$

C. $e + \frac{1}{2}$

D. $e - \frac{1}{2}$

E. $\frac{1}{2}e - \frac{3}{2}$

Question 7. [1 MARK]

Suppose you have a random variable Z representing the time (in seconds) it takes for a computer to boot up. You know Z is distributed according to the continuous uniform distribution over the interval

$$\mathcal{Z} = [5, 15].$$

Let p be the probability density function (pdf) of Z. Which of the following statements are true?

A. $p(10)$ is the probability that $Z = 10$.

B. $p(z)$ represents the probability mass at the point z.

C. The probability that Z is between 8 and 12 is 0.4.

D. The probability that $Z = 5$ is 0.

E. The expected value of Z is 10.

Question 8. [1 MARK]

Suppose you toss three fair coins. Let the first coin be represented by the random variable $C_1 \in \{H,T\}$, the second coin by $C_2 \in \{H,T\}$, and the third coin by $C_3 \in \{H,T\}$. Which of the following sets represents the outcome space of the random variable $C = (C_1, C_2, C_3)$?

A. $\{(c_1, c_2, c_3) \mid c_1, c_2, c_3 \in \{H,T\}\}$

B. $\{H,T\}^3$

C. $\{(c_1, c_2) \mid c_1, c_2 \in \{H,T\}\}$

D. $\{(c_1, c_2, c_3) \mid c_2, c_3 \in \{H,T\},\ c_1 = T\}$

E. $\{(c_1, c_2, c_3, c_4) \mid c_1, c_2, c_3, c_4 \in \{H,T\}\}$

Question 9. [1 MARK]

Suppose your favorite German football team is playing two consecutive games. In each game, the outcomes are defined as follows:

$$W = \text{win}, \quad D = \text{draw}, \quad L = \text{loss}.$$

Thus, the outcome space for the two games is

$$\mathcal{X} = \{WW, WD, WL, DW, DD, DL, LW, LD, LL\}.$$

Which of the following are events?

A. $\{WD, DW\}$

B. $\{WW, WD, WL\}$

C. WL

D. $\{W, L\}$

E. $\varnothing$(empty set)

Question 10. [1 MARK]

On a winter day in Edmonton, the probability that a day is freezing is 0.2 and 0.8 that it is adequate. Assuming that the temperature on consecutive days is independent and identically distributed, what is the probability that exactly one out of three consecutive days is freezing?

A. 0.096

B. 0.128

C. 0.256

D. 0.384

E. 0.512

Question 11. [1 MARK]

Suppose that $X \in \mathbb{R}$ is distributed according to Laplace(0,1). Which of the following statements are true?

A. The probability density function of X is $p(x) = \frac{1}{\sqrt{2\pi}} e^{-|x|}$ for $x \in \mathbb{R}$.

B. The probability that $X = 0$ is $\frac{1}{2}$.

C. X is a continuous random variable.

D. The variance of X is 1.

E. The mean of X is 0.

Question 12. [1 MARK]

Suppose you have two discrete random variables X and Y with outcome spaces

$$X \in \{\text{heads}, \text{tails}\} \quad \text{and} \quad Y \in \{\text{fair}, \text{biased}\}.$$

The joint probability mass function (pmf) of X and Y is given by

$$p(\text{heads}, \text{fair}) = \frac{1}{4}, \qquad p(\text{heads}, \text{biased}) = \frac{1}{8},$$
$$p(\text{tails}, \text{fair}) = \frac{1}{4}, \qquad p(\text{tails}, \text{biased}) = \frac{3}{8}.$$

Which of the following is the conditional pmf of X given Y = fair (hint: product rule can be helpful)?

A. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{2}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{1}{2}$

B. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{4}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{3}{4}$

C. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{8}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{3}{8}$

D. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{3}{4}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{1}{4}$

E. $p_{X|Y}(\text{heads}|\text{biased}) = \frac{1}{3}, \quad p_{X|Y}(\text{tails}|\text{biased}) = \frac{2}{3}$

Question 13. [1 MARK]

Suppose Y is a random variable with outcome space

$$\mathcal{Y} = \{1, 2, 3, 4\},$$

and probability mass function given by

$$p(1) = 0.1, \quad p(2) = 0.3, \quad p(3) = 0.4, \quad p(4) = 0.2.$$

Let the function f be defined as follows:

$$f(1) = 2, \quad f(2) = 5, \quad f(3) = 1, \quad f(4) = 4.$$

What is $\mathbb{E}[f(Y)]$?

A. 2.9

B. 3.1

C. 2.5

D. 3.3

E. 2.7

Question 14. [1 MARK]

Let X be a random variable taking values in $\mathbb{R}$, and define

$$Y = X^2.$$

Which of the following statements are true?

A. Y is a random variable.

B. The outcome space of Y is $[0, \infty)$.

C. The outcome space of Y is $\mathbb{R}$.

D. Y is not a random variable.

E. Y has the same probability distribution as X.

Question 15. [1 MARK]

Suppose X_1, X_2, X_3 are independent random variables, each with $X_i \sim \mathcal{N}(2, 5)$. Let $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$. Which of the following statements are true?

A. $\mathbb{E}[\bar{X}] = 2$.

B. $\mathrm{Var}(\bar{X}) = \frac{5}{3}$.

C. $\mathbb{E}[\bar{X}] = \frac{2}{3}$.

D. The variance of $\frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$ is smaller than the variance of $\bar{X}$.

E. The variance of $\frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$ is equal or higher than the variance of $\bar{X}$.

Question 16. [1 MARK]

Suppose you have a dataset where each instance represents a movie. The dataset includes the following variables: production budget (in millions of dollars), runtime (in minutes), the number of prizes the movie received, and the movie's generated revenue represented as a number with digits behind the decimal point. You wish to predict the movie's generated revenue. Which of the following statements are true?

A. The features can be represented as an element of $\mathbb{N}^3$.

B. The label set is finite.

C. This is a regression problem.

D. This is a classification problem.

E. The label set can be represented as a subset of $\mathbb{R}$.

Question 17. [1 MARK]

Let X represent whether a bus arrives on time, where $X = 1$ (on time) with probability 0.75 and $X = 0$ (late) with probability 0.25. Given that $X = 1$, the conditional distribution of the number of passengers N boarding the bus is given by

$$\mathbb{P}_{N|X}(N = 10 \mid X = 1) = 0.2, \quad \mathbb{P}_{N|X}(N = 20 \mid X = 1) = 0.5, \quad \mathbb{P}_{N|X}(N = 30 \mid X = 1) = 0.3.$$

What is the conditional expectation $\mathbb{E}[N \mid X = 1]$?

A. 20

B. 21

C. 22

D. 23

E. 24

Question 18. [1 MARK]

You are given data consisting of the delivery time (in hours) for a package, the distance to the destination (in kilometers), the package weight (in kilograms), and the prevailing weather condition (e.g., "clear", "rainy", "snowy"). Suppose you want to predict the delivery time of a package. Which of the following options correctly identifies the features and the label for this problem?

A. Features: delivery time, distance, package weight; Label: weather condition.

B. Features: package weight, weather condition; Label: delivery time.

C. Features: distance, package weight, weather condition; Label: delivery time.

D. Features: delivery time, weather condition; Label: package weight.

E. Features: distance, weather condition; Label: delivery time.

Question 19. [1 MARK]

Suppose you are given a predictor $f(x) = 4 + x$, which models the relationship between the number of years of experience x and the salary y (in tens of thousands of dollars) of an employee. You are provided with the following dataset of 4 (x, y) pairs: $\mathcal{D} = \{(1, 4.2), (2, 5.7), (3, 6.8), (4, 8.5)\}$. Let the loss function be the absolute loss $\ell(f(x), y) = \left|f(x) - y\right|$. Calculate $\hat{L}(f) = \frac{1}{4}\sum_{i=1}^{4} \ell(f(x_i), y_i)$. Which of the following statements are true?

A. $\hat{L}(f) = 0.425$

B. $\hat{L}(f) = 0.450$

C. $\hat{L}(f) = 0.400$

D. $\hat{L}(f) = 0.500$

E. $\hat{L}(f)$ is an estimate of the expected absolute loss $L(f)$.

Question 20. [1 MARK]

Let X be a random vector in $\mathbb{R}$ representing a feature extracted from an email, and let $Y \in \{\text{spam}, \text{not spam}\}$ be the corresponding label. Let $g(X)$ be a classifier that outputs a predicted label, and define the loss function as

$$l(g(X), Y) = \begin{cases} 1, & \text{if } g(X) \neq Y, \\ 0, & \text{if } g(X) = Y. \end{cases}$$

Which of the following statements about the expected loss are true?

A. The expected loss can be expressed as

$$\mathbb{E}[l(g(X), Y)] = \int_{\mathbb{R}} \sum_{y \in \{\text{spam,not spam}\}} l(g(x), y)\, p(x, y)\, dx.$$

B. The expected loss can be written as

$$\mathbb{E}[l(Y, g(X))] = \int_{\mathbb{R}} \sum_{y \in \{\text{spam,not spam}\}} l(y, g(x))\, p(x \mid y)\, dx.$$

C. The expected loss is given by

$$\mathbb{E}[l(g(X), Y)] = \int_{\mathbb{R}} l(g(x), y)\, p(x)\, dx.$$

D. The expected loss is computed as

$$\mathbb{E}[l(g(X), Y)] = \int_{\mathbb{R}} l(g(x), y)\, dx.$$

E. The expected loss is calculated with respect to the distribution $\mathbb{P}_{X,Y}$.

Question 21. [1 MARK]

Suppose you roll a six-sided die 4 times and obtain the following outcomes:

$$X_1 = 2, \quad X_2 = 5, \quad X_3 = 3, \quad X_4 = 6.$$

The sample mean is $\bar{X} = 4$. What is the sample variance of these 4 rolls?

A. $\frac{10}{3}$

B. 3.0

C. 4.0

D. $\frac{8}{3}$

E. 2.5

Question 22. [1 MARK]

Consider the convex function $h(z) = \sum_{i=1}^{3} \big(5z - a_i\big)^2$, where $a_1 = 2$, $a_2 = 4$, $a_3 = 6$, and $z \in \mathbb{R}$. Which value of z minimizes $h(z)$?

A. $z = 0.8$

B. $z = 1.0$

C. $z = 0.6$

D. $z = 1.2$

Question 23. [1 MARK]

Consider the function $f(x) = \ln(3x + 1) + 2x^2$, where $x \in \mathbb{R}$. What is the first derivative $f'(x)$?

A. $f'(x) = \frac{1}{3x+1} + 4x$

B. $f'(x) = \frac{3}{3x+1} + 2x$

C. $f'(x) = \frac{3}{(3x+1)^2} + 4x$

D. $f'(x) = \frac{3}{3x+1} + 4x$

E. $f'(x) = \frac{3}{3x+1} + 4$

Question 24. [1 MARK]

Consider the function $f(x) = \frac{1}{12}x^4 - 4x^2 + 2$, where its second derivative is: $f''(x) = x^2 - 8$. Which of the following statements are true?

A. f is convex on $\mathbb{R}$.

B. $f''(x) = x^2 - 8$ is negative for some $x \in \mathbb{R}$.

C. f is not convex on $\mathbb{R}$.

D. f is convex on a subset of $\mathbb{R}$.

Question 25. [1 MARK]

Suppose you are working within the empirical risk minimization (ERM) framework for supervised learning. Let f be a predictor with parameters θ and let the loss function be $\ell(f(x), y)$. The empirical risk is defined as

$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), y_i)$$

over a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$. Which of the following statements correctly describe the estimation and optimization steps in ERM?

A. The estimation step involves computing the empirical risk, which serves as an approximation of the true risk

$$L(f) = \mathbb{E}_{(x,y)\sim\mathbb{P}_{X,Y}}[\ell(f(x), y)]$$

because the true data distribution is unknown.

B. The optimization step involves finding the model parameters θ that minimize the empirical risk $\hat{L}(f)$.

C. The optimization step always guarantees that the global minimum of the empirical risk is found.

D. The estimation step is necessary because we use the training data to approximate the expected loss over the true distribution.

E. The expected risk $L(f)$ is computed with respect to the true distribution $\mathbb{P}_{X,Y}$, which is generally not directly available.

For your notes (1/4)

For your notes (2/4)

For your notes (3/4)

For your notes (4/4)

Formula Sheet

Integration

$$\int_a^b x^d dx = \frac{x^{d+1}}{d+1}\Big|_a^b = \frac{b^{d+1} - a^{d+1}}{d+1} \quad \text{for } d \neq -1$$

Derivatives

$$\begin{aligned}
&f(x) = x^a, && f'(x) = \frac{df}{dx}(x) = ax^{a-1}\\
&f(x) = \exp(x), && f'(x) = \frac{df}{dx}(x) = \exp(x)\\
&f(x) = \ln(x), && f'(x) = \frac{df}{dx}(x) = \frac{1}{x}\\
&f(x) = g(h(x)), \quad u = h(x) && f'(x) = \frac{df}{dx}(x) = \frac{dg}{du}\frac{du}{dx}(x) = g'(u)h'(x) \quad \triangleright \text{Chain rule}\\
&f(x) = g(x)h(x), && f'(x) = \frac{df}{dx}(x) = g'(x)h(x) + g(x)h'(x) \quad \triangleright \text{Product rule}
\end{aligned}$$

Probability

Univariate: $\mathbb{P}(X \in \mathcal{E}) \stackrel{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}} p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{E}} p(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Multivariate: $\mathbb{P}(X \in \mathcal{E}_X, Y \in \mathcal{E}_Y) \stackrel{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}_x} \sum_{y \in \mathcal{E}_y} p(x,y) & \text{if } X, Y \text{ are discrete} \\ \int_{\mathcal{E}_x} \int_{\mathcal{E}_y} p(x,y)dydx & \text{if } X, Y \text{ are continuous} \\ \int_{\mathcal{E}_x} \sum_{y \in \mathcal{E}_y} p_{Y|X}(y|x)p_X(x)dx & \text{if } X \text{ is continuous, } Y \text{ is discrete} \\ \sum_{x \in \mathcal{E}_x} \int_{\mathcal{E}_y} p_{Y|X}(y|x)p_X(x)dy & \text{if } X \text{ is discrete, } Y \text{ is continuous} \end{cases}$

Marginal pmf: $p_X(x) \stackrel{\text{def}}{=} \begin{cases} \sum_{y \in \mathcal{Y}} p(x,y) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{Y}} p(x,y)dy & \text{if } Y \text{ is continuous} \end{cases}$

Marginal: $\mathbb{P}_X(X \in \mathcal{E}_X) \stackrel{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}_X} p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{E}_X} p_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Conditional pmf: $p_{Y|X}(y|x) \stackrel{\text{def}}{=} \frac{p(x,y)}{p_X(x)}$ such that $p_X(x) > 0$

Conditional: $\mathbb{P}_{Y|X}(Y \in \mathcal{E}_Y | X = x) \stackrel{\text{def}}{=} \begin{cases} \sum_{y \in \mathcal{E}_Y} p_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{E}_Y} p_{Y|X}(y|x)dy & \text{if } Y \text{ is continuous} \end{cases}$

Product Rule: $p(x,y) = p_{Y|X}(y|x)p_X(x)$

Bayes' Rule: $p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$

Independence: $p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$

Distribution	Parameters	pmf or pdf	Expectation and Variance
Bernoulli	$\alpha \in [0,1]$	$p(x) = \alpha^x(1-\alpha)^{1-x}$, $x \in \{0,1\}$	$\mathbb{E}[X] = \alpha$, $\mathrm{Var}[X] = \alpha(1-\alpha)$
Discrete Uniform	$n \in \mathbb{N}$	$p(x) = \frac{1}{n}$, $x \in \{1,\dots,n\}$	$\mathbb{E}[X] = \frac{n+1}{2}$, $\mathrm{Var}[X] = \frac{n^2-1}{12}$
Continuous Uniform	$a,b \in \mathbb{R}, a<b$	$p(x) = \frac{1}{b-a}$, $x \in [a,b]$	$\mathbb{E}[X] = \frac{a+b}{2}$, $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$
Normal	$\mu \in \mathbb{R}, \sigma^2 > 0$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x \in \mathbb{R}$	$\mathbb{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$
Laplace	$\mu \in \mathbb{R}, b > 0$	$p(x) = \frac{1}{2b} \exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$, $x \in \mathbb{R}$	$\mathbb{E}[X] = \mu$, $\mathrm{Var}[X] = 2b^2$

Expectation and Variance

Univariate: $\mathbb{E}[X] \stackrel{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}} xp(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} xp(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Function: $\mathbb{E}[f(X)] \stackrel{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}} f(x)p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} f(x)p(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Variance: $\mathrm{Var}[X] \stackrel{\text{def}}{=} \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$

Multivariate: $\mathbb{E}[f(X,Y)] \stackrel{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} f(x,y)p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x,y)p(x,y)\,dy\,dx & \text{if } X \text{ and } Y \text{ are continuous} \\ \int_{\mathcal{X}} \sum_{y\in\mathcal{Y}} f(x,y)p_{Y|X}(y|x)p_X(x)\,dx & \text{if } X \text{ is continuous, } Y \text{ is discrete} \\ \sum_{x\in\mathcal{X}} \int_{\mathcal{Y}} f(x,y)p_{Y|X}(y|x)p_X(x)\,dy & \text{if } X \text{ is discrete, } Y \text{ is continuous} \end{cases}$

Conditional: $\mathbb{E}[f(Y)|X=x] \stackrel{\text{def}}{=} \begin{cases} \sum_{y\in\mathcal{Y}} f(y)p_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{Y}} f(y)p_{Y|X}(y|x)\,dy & \text{if } Y \text{ is continuous} \end{cases}$

Expectation and Variance Properties

1. $\mathbb{E}[cX] = c\mathbb{E}[X]$
2. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
3. $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$
4. $\mathrm{Var}[c] = 0$
5. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$.

If X and Y are independent:

6. $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$
7. $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Estimation

Sample Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

If X_i are i.i.d.: $\mathbb{E}[\bar{X}] = \mathbb{E}[X_1]$, $\quad \mathrm{Var}\left[\bar{X}\right] = \frac{\mathrm{Var}[X_1]}{n}$