School of Electrical Engineering, KAIST

EE210 Probability and Introductory Random Processes

2023 Spring

Midterm

Tuesday, April 18, 2023

9:00–11:30 am

NAME: ______________________________

Student ID: ______________________________

- Don't forget to put your name and student ID.
- **Record all your solutions in this answer booklet. Only this answer booklet will be considered in the grading of your exam.**
- Be sure to show all relevant work and reasoning. A correct answer does not guarantee full credit, and a wrong answer does not guarantee loss of credit. You should clearly but concisely indicate your reasoning.

Problem	Your score	Max score
1		10
2		10
3		10
4		10
Total		40

<u>Problem 1</u> (10 Points)

In this problem, we consider an experiment of filling boxes with balls. Let's assume that there are k boxes. At each experiment, we randomly choose d boxes for some fixed integer $d \leq k$, and put a ball into each of the chosen d boxes. (So, we put total d balls into randomly selected d boxes (one ball at a box) at each experiment. The probability that any box is filled with a ball after the first experiment is equal to d/k.) We repeat this experiment n times. Each experiment is independent.

Let's define an indicator variable Z_i for $i = 1, \ldots, k$ such that $Z_i = 0$ if the i-th box is empty after n times of the experiments and $Z_i = 1$ otherwise, i.e., if it is filled with at least one ball. Let's define the number of boxes that are not empty as $Z = Z_1 + Z_2 + \cdots + Z_k$.

a) (5 points) Calculate $\mathbb{E}[Z]$.
Answer:

$$\mathbb{E}[Z] = k\left(1 - \left(1 - \frac{d}{k}\right)^n\right)$$

Reasoning for Problem 1(a):

Note that $\mathbf{P}(Z_1 = 1) = 1 - \mathbf{P}(Z_1 = 0)$ and $\mathbf{P}(Z_1 = 0) = \left(1 - \frac{d}{k}\right)^n$. Therefore, $\mathbf{P}(Z_1 = 1) = 1 - \left(1 - \frac{d}{k}\right)^n$. From the linearity of the expectation and symmetry between Z_i's, $\mathbb{E}[Z] = \sum_{i=1}^{k} Z_i = kZ_1$. Thus,

$$\mathbb{E}[Z] = k\left(1 - \left(1 - \frac{d}{k}\right)^n\right).$$

Partial point: $\mathbf{P}(Z_1 = 0) = \left(1 - \frac{d}{k}\right)^n$ (3 points)

b) (5 points) Calculate $\mathbb{E}[Z_1 Z_2]$.
(Hint: $\mathbb{E}[Z_1 Z_2] = \mathbf{P}(Z_1 = 1, Z_2 = 1) = \mathbf{P}(Z_1 = 1) - \mathbf{P}(Z_2 = 0) + \mathbf{P}(Z_1 = 0, Z_2 = 0)$.)

Answer: $\mathbb{E}[Z_1 Z_2] = 1 - 2\left(1 - \frac{d}{k}\right)^n + \left(\frac{(k-d)(k-d-1)}{k(k-1)}\right)^n$

Reasoning for Problem 1(b): Note that

$$\mathbb{P}(Z_1 = 0, Z_2 = 0) = \left(\frac{\binom{k-2}{d}}{\binom{k}{d}}\right)^n = \left(\frac{(k-d)(k-d-1)}{k(k-1)}\right)^n.$$

Thus, by using the hint and the calculation in a),

$$\mathbb{E}[Z_1 Z_2] = 1 - 2\left(1 - \frac{d}{k}\right)^n + \left(\frac{(k-d)(k-d-1)}{k(k-1)}\right)^n.$$

Partial point: $\mathbb{P}(Z_1 = 0, Z_2 = 0) = \left(\frac{\binom{k-2}{d}}{\binom{k}{d}}\right)^n$ (3 points)

Problem 2 (10 Points)

In this problem, we consider two sampling methods–importance sampling and rejection sampling–to approximate a distribution by samples from another distribution.

a) (5 points) Let $Y_1, \ldots, Y_n$ be independent random variables drawn from a common and known PDF f_Y. Let S be the set of all possible values of Y_i, $S = \{y | f_Y(y) > 0\}$. Let X be a random variable with known PDF f_X, such that $f_X(y) = 0$, for all $y \notin S$. Design a weight function $g(Y)$ in terms of $f_X(\cdot)$ and $f_Y(\cdot)$ such that when we define a new random Z by

$$Z = \frac{1}{n}\sum_{i=1}^{n} Y_i g(Y_i),$$

the random variable Z satisfies $\mathbb{E}[Z] = \mathbb{E}[X]$. This sampling method is called 'importance sampling.'

Answer: $g(Y_i) = \frac{f_X(Y_i)}{f_Y(Y_i)}$

Reasoning for Problem 2(a):

When we set $g(Y_i) = \frac{f_X(Y_i)}{f_Y(Y_i)}$, we have

$$\mathbb{E}[Y_i g(Y_i)] = \int_S y \frac{f_X(y)}{f_Y(u)} f_Y(y) dy = \int_S y f_X(y) dy = \mathbb{E}[X].$$

Thus,

$$\mathbb{E}[Z] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i g(Y_i)] = \mathbb{E}[X].$$

b) (5 points) Let $Y_1, \ldots, Y_n$ be independent random variables drawn from a common and known PDF f_Y. Let S be the set of all possible values of Y_i, $S = \{y | f_Y(y) > 0\}$. Let X be a random variable with known PDF f_X, such that $f_X(y) = 0$, for all $y \notin S$. Furthermore, assume that there exists a constant $M > 0$ such that $f_X(x) \le M f_Y(x)$ for all values of x. For each Y_i, we sample another random variable U_i independently from uniform distribution over $[0, 1]$ and accept the sample Y_i only if $U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)}$. Otherwise, we reject the sample Y_i. This method is called rejection sampling and the remaining samples after the rejection approximates the samples from $f_X(x)$.

Calculate the acceptance probability $P(U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)})$ in terms of $M > 0$.

Answer:

$P(U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)}) = 1/M$ (3 points)
Expected number of trails = M (2 points)

Reasoning for Problem 2(b):

$$\begin{aligned}
P\left(U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)}\right) &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left(U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)}\right) \Big| Y\right]\right] \\
&= \mathbb{E}\left[P\left(\mathbb{1}\left(U_i < \frac{f_X(Y_i)}{M f_Y(Y_i)}\right) \Big| Y\right)\right] \\
&= \mathbb{E}\left[\frac{f_X(Y_i)}{M f_Y(Y_i)}\right] \\
&= \int_S \frac{f_X(y)}{M f_Y(y)} f_Y(y) dy \\
&= \frac{1}{M}
\end{aligned}$$

Problem 3 (10 Points)

An investment bank is managing \$1 billion (\$1000 million), which it invests in various financial instruments ("assets") related to the housing market. Because the bank is investing with borrowed money, its actual assets are only \$50 million (5%). Accordingly, if the bank loses more than 5%, it becomes insolvent (unable to pay debts owed).

a) (3 points) The bank considers investing in a single asset, whose gain (over a 1-year period, and measured in percentage points) is modeled as a normal random variable R, with mean 7 and standard deviation 10. (That is, the asset is expected to yield a 7% profit.) What is the probability that the bank will become insolvent, i.e., loses more than 5% of its assets? Use the table below to get the number until the 3rd decimal place by rounding it.

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Answer: $P(\text{the bank becomes insolvent}) = 0.115$

Reasoning for Problem 3(a):

Note that $\mathbb{E}[R] = 7$ and $\text{var}(R) = 10^2 = 100$. Thus, we have $P(R \leq -5) = P\left(\frac{R-7}{10} \leq \frac{-5-7}{10}\right) = \Phi(-1.2) \approx 1 - 0.88493 = 0.11507$.

b) (3 points) In order to safeguard its position, the bank decides to diversify its investments. It considers investing \$250 million in each of four different assets, with the i-th one having a gain R_i, which is again normal with mean 7 and standard deviation 10; the bank's gain will be $(R_1+\cdots+R_4)/4$. These four assets are chosen to reflect the housing sectors at different countries. Assume that R_i's are modeled to be independent. What is the probability that the bank becomes insolvent?

Answer: $P(\text{the bank becomes insolvent}) = 0.008$

Reasoning for Problem 3(b):

Note that $\mathbb{E}[R] = (\mathbb{E}[R_1] + \cdots + \mathbb{E}[R_4])/4 = 7$, and $\text{var}(R) = \frac{1}{16}(\text{var}(R_1) + \cdots \text{var}(R_4)) = \frac{4 \cdot 100}{16} = 25$. Thus, we have

$$P(R \leq -5) = P\left(\frac{R-7}{5} \leq \frac{-5-7}{5}\right) = \Phi(-2.4) = 1 - 0.99180 = 0.0082$$

c) (4 points) Based on the calculations in part (b), the bank goes ahead with the diversified investment strategy. It turns out that a global economic phenomenon can affect the housing sectors in different countries simultaneously, and therefore the gains R_i are in fact positively correlated. Suppose that for every i and j where $i \neq j$, the correlation coefficient $\rho(R_i, R_j)$ is equal to 0.07. What is the probability that the bank becomes insolvent? For simpleness in the calculation, just assume that $(R_1 + \cdots + R_4)/4$ is normal. Calculate the probability until the 3rd decimal place.
(Hint: $\text{cov}(R_i, R_j) = \rho(R_i, R_j)\sqrt{\text{var}(R_i)\,\text{var}(R_j)}$ and the fact that $\sqrt{484} = 22$.)

Answer: $P(\text{the bank becomes insolvent}) = 0.015$

Reasoning for Problem 3(c): Let us first calculate the covariance between R_i and R_j.

$$\text{cov}(R_i, R_j) = \rho(R_i, R_j)\sqrt{\text{var}(R_i)\,\text{var}(R_j)} = 0.07\sqrt{10^2 \cdot 10^2} = 7$$

Note that we have

$$\begin{aligned}\text{var}(R) = \text{var}\left(\frac{1}{4}\sum_{i=1}^{4} R_i\right) &= \frac{1}{16}\left(\sum_{i=1}^{4}\text{var}(R_i) + \sum_{i\neq j}\text{cov}(R_i, R_j)\right)\\ &= \frac{1}{16}(4 \cdot 100 + 12 \cdot 7) = \frac{484}{16}.\end{aligned}$$

Note that $\sqrt{\frac{484}{16}} = \frac{22}{4} = 5.5$. Thus, we have

$$P(R \leq -5) = P\left(\frac{R-7}{5.5} \leq \frac{-5-7}{5.5}\right) \approx \Phi(-2.18) = 1 - 0.98537 = 0.01463$$

<u>Problem 4</u> (10 Points)

Consider the communication of binary-valued messages using a noisy channel from Alice to Bob. Specifically, Alice wants to send one of two symbols, either 0 or 1, to Bob, with equal probabilities. Alice and Bob are connected by a noisy channel, and Alice can send some numerical value X through the channel. Due to the noise in the channel, Bob receives some random output $Y = X + N$ where the random variable N represents the noise that is independent of X and distributed by a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 9$.

a) (5 points) Suppose that Alice encodes the symbol 0 with the value $X = -3$ and the symbol 1 with the value $X = 3$. Bob then decodes the message according to the following rule:

   - If $Y \geq 0$, then conclude that the symbol 1 was sent.
   - If $Y < 0$, then conclude that the symbol 0 was sent.

   Calculate the probability that Bob gets the wrong guess of the message from this scheme. Remind that the probability that Alice sends symbol 0 (or 1) is 1/2. Reduce your calculations to a single numerical value (until the 3rd decimal place) by using the standard normal table in page 6.

   Answer: $P(\text{Bob gets the wrong guess of the message}) = 0.159$

Reasoning for Problem 4(a):

$$\begin{aligned}
\Pr(\text{error}) &= \Pr(\text{error}|X=-3)\Pr(X=-3) + \Pr(\text{error}|X=3)\Pr(X=3) \\
&= \frac{1}{2}\Pr(Y \geq 0|X=-3) + \frac{1}{2}\Pr(Y < 0|X=3) \\
&= \frac{1}{2}\Pr(N \geq 3) + \frac{1}{2}\Pr(N < -3) \\
&= \Pr(N \geq 3) = \Pr\left(\frac{N-0}{3} \geq \frac{3-0}{3}\right) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587.
\end{aligned}$$

b) (5 points) Next, consider the following modification of the previous communication scheme. Now, Alice and Bob use the communication channel three times, and Alice sends $(X_1, X_2, X_3) = (-3, -3, -3)$ for symbol 0 and $(X_1, X_2, X_3) = (3, 3, 3)$ for symbol 1. Bob receives $Y_i = X_i + N_i$ for $i = 1, 2, 3$ where each N_i is a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 9$. Assume that each N_i is independent of each other and independent of X_i's. Each Y_i is decoded according to the rule in a), and Bob uses a majority voting rule to determine which symbol was sent:

- If two or more components of (Y_1, Y_2, Y_3) are greater than or equal to 0, then conclude that symbol 1 was sent
- If two or more components of (Y_1, Y_2, Y_3) are less than 0, then conclude that symbol 0 was sent

What is the probability that Bob gets the wrong guess of the message from this scheme? Just write down the formula (you don't need to calculate the value of the probability).

Answer:

$P(\text{Bob gets the wrong guess of the message}) = \binom{3}{2}(0.1587)^2(1-0.1587)+\binom{3}{3}(0.1587)^3$

Reasoning for Problem 4(b): By symmetry, Pr(error|sent 0) = Pr(error|sent 1).

$$\begin{aligned}\Pr(\text{error}|\text{sent } 0) &= \binom{3}{2}(0.1587)^2(1-0.1587)+\binom{3}{3}(0.1587)^3 \\ &= 1 - 3(0.1587)^1(1-0.1587)^2 - (1-0.1587)^3.\end{aligned}$$