Final Part I

Thursday, June 18, 2020
2:30–4:10 pm

- Be sure to **show all relevant work and reasoning** in your answer sheet. 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.
- Please be clear in writing—we can't grade what we can't decipher!
- Don't forget to upload your answer sheet by 4:10 pm through KLMS. The system will be automatically closed at 4:10 pm. If the system does not work, you should email it to `ee210b_20spring@kaist.ac.kr` by 4:10 pm. Late submissions will not be accepted/graded.

Problem 1 **(15 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 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 a 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 $Z = Z_1 + Z_2 + \cdots + Z_k$. We aim to calculate the expectation of Z and variance of Z by the following steps.

a) (5 points) Calculate $\mathbb{E}[Z]$.
(Hint: You may use the fact that $\mathbb{E}[Z_1] = \mathbf{P}(Z_1 = 1)$.)

Solution: 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: $\mathbf{P}(Z_1 = 0) = \left(1 - \frac{d}{k}\right)^n$ (4 points)

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

Solution: 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: $\mathbb{P}(Z_1 = 0, Z_2 = 0) = \left(\frac{\binom{k-2}{d}}{\binom{k}{d}}\right)^n$ (4 points)

c) (5 points) Calculate $\mathrm{var}(Z)$.
(Hint: Note that $\mathbb{E}[Z^2] = \sum_{i=1}^{k} \mathbb{E}[Z_i^2] + \sum_{\{(i,j):i \neq j\}} \mathbb{E}[Z_i Z_j]$.)

Solution: By using $\mathrm{var}(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \sum_{i=1}^{k} \mathbb{E}[Z_i^2] + \sum_{\{(i,j):i \neq j\}} \mathbb{E}[Z_i Z_j] -$

$(\mathbb{E}[Z])^2$, and the result from a) and b),

$$
\begin{aligned}
&\operatorname{var}(Z)\\
&= k\left(1-\left(1-\frac{d}{k}\right)^n\right) + k(k-1)\left(1-2\left(1-\frac{d}{k}\right)^n + \left(\frac{(k-d)(k-d-1)}{k(k-1)}\right)^n\right) - k^2\left(1-\left(1-\frac{d}{k}\right)^n\right)\\
&= k\left(1-\frac{d}{k}\right)^n\left(1-k\left(1-\frac{d}{k}\right)^n\right) + k(k-1)\left(\frac{(k-d)(k-d-1)}{k(k-1)}\right)^n.
\end{aligned}
$$

Partial point: Using $\operatorname{var}(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$ (2 point)

Problem 2 (15 Points)

Assume that your course grade is determined by your midterm score X_1 and your final score X_2. Your scores X_1, X_2 are independent normal random variables with the same mean $\mu < 90$ and the same variance σ^2.

a) (5 points) Assume that the grade is determined by the average score $Z = \frac{X_1}{2} + \frac{X_2}{2}$. You earn grade 'A' if $Z > 90$. What is the probability $\mathbf{P}(A) = \mathbf{P}(Z > 90)$? Write down this probability in terms of the CDF for the standard normal. (Remind that the CDF for the standard normal is $\Phi(y) = \mathbf{P}(Y \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2} dt$.)

Solution: Note that Z is also a normal random variable with mean μ and variance $\sigma^2/2$. By standardizing Z,

$$\mathbb{P}(Z > 90) = 1 - \mathbb{P}(Z \leq 90) = 1 - \mathbb{P}\left(\frac{Z-\mu}{\sigma/\sqrt{2}} \leq \frac{90-\mu}{\sigma/\sqrt{2}}\right) = 1 - \Phi\left(\frac{90-\mu}{\sigma/\sqrt{2}}\right)$$

b) (5 points) A different student proposes that the better exam is the one that should count and the grades should be based on $M = \max(X_1, X_2)$. What is $\mathbf{P}(A) = \mathbf{P}(M > 90)$ for this case?

Solution: Note that

$$\begin{aligned}
\mathbf{P}(M > 90) &= 1 - \mathbf{P}(M \leq 90) = 1 - \mathbf{P}(X_1 \leq 90, X_2 \leq 90) \\
&= 1 - \mathbf{P}(X_1 \leq 90)\mathbf{P}(X_2 \leq 90) = 1 - \left(\mathbf{P}\left(\frac{X_1-\mu}{\sigma} \leq \frac{90-\mu}{\sigma}\right)\right)^2 \\
&= 1 - \left(\Phi\left(\frac{90-\mu}{\sigma}\right)\right)^2
\end{aligned}$$

c) (5 points) Assume that the mean and the standard deviation are $\mu = 74$ and $\sigma = 16$, respectively. Use the table below for the CDF values of the standard normal to calculate the expected increase in the number of A's awarded by using $M = \max(X_1, X_2)$ instead of $Z = \frac{X_1}{2} + \frac{X_2}{2}$ in a class of 100 students. (You may use $\sqrt{2} \approx 1.41$.)

Solution: For $\mu = 74$ and $\sigma = 16$, $\mathbb{P}(Z > 90) = 1 - \Phi(\sqrt{2}) \approx 1 - 0.9207 = 0.0793$, and $\mathbb{P}(M > 90) = 1 - (\Phi(1))^2 \approx 1 - 0.8413^2 = 0.2922$. Therefore, the expected number of A's for a class of size 100 increases by 29-8=21, approximately.

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \leq y)$, where Y is a standard normal random variable, for y between 0 and 1.99. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

Problem 3 (10 Points)

Let $X_1, X_2, \ldots$ be independent normal random variables with mean 1 and variance 8. Let N be a positive (finite) integer random variable with $\mathbb{E}[N] = 2$ and $\mathbb{E}[N^2] = 6$. We assume that the random variables $N, X_1, X_2, \ldots$ are independent. Let $S = \sum_{i=1}^{N} X_i$.

a) (5 points) Find the variance of S.

Solution: Using the law of total variance,

$$\begin{aligned}\mathrm{var}(S) &= \mathbb{E}[\mathrm{var}(S|N)] + \mathrm{var}(\mathbb{E}[S|N]) \\ &= \mathbb{E}[8N] + \mathrm{var}(1 \cdot N) \\ &= 8\mathbb{E}[N] + (6-4) = 18.\end{aligned}$$

Partial point: Use of law of total variance (1 point), correct calculation for $\mathbb{E}[\mathrm{var}(S|N)]$ (2 points), correct calculation for $\mathrm{var}(\mathbb{E}[S|N])$ (2 points)

b) (5 points) Calculate the covariance $\mathrm{cov}(N, S)$. Are N and S are uncorrelated?

Solution: The covariance of N and S is

$$\begin{aligned}\mathrm{cov}(N, S) &= \mathbb{E}[NS] - \mathbb{E}[N]\mathbb{E}[S] \\ &= \mathbb{E}[\mathbb{E}[NS|N]] - \mathbb{E}[N]\mathbb{E}[\mathbb{E}[S|N]] \\ &= \mathbb{E}[\mathbb{E}[\sum_{i=1}^{N} X_i N|N]] - \mathbb{E}[N]\mathbb{E}[\mathbb{E}[\sum_{i=1}^{N} X_i|N]] \\ &= \mathbb{E}[N^2 X_1] - \mathbb{E}[N]\mathbb{E}[N X_1] \\ &= \mathbb{E}[N^2]\mathbb{E}[X_1] - (\mathbb{E}[N])^2\mathbb{E}[X_1] = \mathrm{var}(N)\mathbb{E}[X_1] = (6-4) \cdot 1 = 2.\end{aligned}$$

Since $\mathrm{cov}(N, S) > 0$, N and S are correlated.

Partial point: Covariance formula (1 point), correct calculation for $\mathbb{E}[NS]$ (2 points), correct calculation for $\mathbb{E}[N]\mathbb{E}[S]$ (2 points)