Discussion 6B

CS 70, Summer 2024

1 Inequality Practice

(a) X is a random variable such that $X \geq -5$ and $\mathrm{E}[X] = -3$. Find an upper bound for the probability of X being greater than or equal to -1.

(b) Y is a random variable such that $Y \leq 10$ and $\mathrm{E}[Y] = 1$. Find an upper bound for the probability of Y being less than or equal to -1.

(c) You roll a die 100 times. Let Z be the sum of the numbers that appear on the die throughout the 100 rolls. Compute $\mathrm{Var}(Z)$. Then use Chebyshev's inequality to bound the probability of the sum Z being greater than 400 or less than 300.

2 Crazy High moments

Continuing 1(c), let Z be the sum of numbers that appear on the die throughout the 100 rolls.

(a) For any n independent random variables $X_1, X_2, \dots, X_n$ with $\mathrm{E}[X_1] = \mathrm{E}[X_2] = \cdots = \mathrm{E}[X_n] = 0$. Prove that

$$\mathrm{E}[(X_1 + X_2 + \cdots + X_n)^4] = \sum_{i=1}^{n} \mathrm{E}[X_i^4] + \binom{4}{2} \cdot \sum_{i<j} \mathrm{E}[X_i^2] \cdot \mathrm{E}[X_j^2].$$

You may use

$$(X_1 + X_2 + \cdots + X_n)^4 = \sum_{i=1}^{n} X_i^4 + \binom{4}{2} \sum_{i<j} X_i^2 X_j^2 + \binom{4}{1} \sum_{i \neq j} X_i X_j^3 + 2 \cdot \binom{4}{2} \sum_{\substack{i,j,k \text{ distinct} \\ j<k}} X_i^2 X_j X_k + 4! \cdot \sum_{i<j<k<t} X_i X_j X_k X_t$$

without proof.

(Hint: use linearity of expectation and independence. For example, we know when $i \neq j$, $\mathrm{E}[X_i X_j^3] = \mathrm{E}[X_i] \cdot \mathrm{E}[X_j^3] = 0$ because X_i, X_j are independent $\mathrm{E}[X_i] = 0$.)

(b) Let Z_i be the number on the i-th roll, and let $X_i = Z_i - \mathrm{E}[Z_i]$. What is the value of the fourth moment $\mathrm{E}[(Z - \mathrm{E}[Z])^4]$? (To avoid the headache with numerical calculations, you may use $(1-3.5)^4 + (2-3.5)^4 + \cdots + (6-3.5)^4 = \frac{707}{8}$, $35^2 = 1225$, $12^2 = 144$, $100 \cdot \frac{707}{48} + \binom{4}{2} \cdot \binom{100}{2} \cdot \frac{1225}{144} = \frac{1524775}{6} \approx 254129$, and $254129/6250000 \approx 0.04066..$)

(c) Use Markov's inequality to upper bound $\Pr[Z > 400]$ with the calculation in (b). (Hint: Recall how we use Markov's inequality to prove Chebyshev's inequality.) Compare it with your result in 1(c). What do you see? Could you guess what would happen if we use even higher moments?

(You may use $50^4 = 6250000$ and $290573/6250000 \approx 0.0464917$.)

Remark. If you are interested, google Chernoff bound. It uses what is called the moment generating function, which is essentially equivalent to using all the higher moments at the same time.

3 Estimating μ and σ^2

Suppose you have n independent and identically distributed samples $X_1, X_2, \ldots, X_n$ drawn from some unknown distribution with expectation μ and variance σ^2. You would like to estimate the mean and variance of this unknown distribution.

(a) We saw in lecture a good estimation of the mean is $\hat{\mu} = \frac{X_1 + \cdots + X_n}{n}$. Show that $E[\hat{\mu}] = \mu$.

(b) What is $\mathrm{E}[X_i^2]$ and $\mathrm{E}[X_i X_j]$ $(i \neq j)$? Express them using μ and σ^2.

(c) One reasonable estimation of the variance is $\hat{\sigma}^2 = \frac{(X_1-\hat{\mu})^2 + (X_2-\hat{\mu})^2 + \cdots + (X_n-\hat{\mu})^2}{n}$. Calculate $E[\hat{\sigma}^2]$, and simplify to show that this estimation does not satisfy the property $E[\hat{\sigma}^2] = \sigma^2$. (Hint: Write $\hat{\sigma}^2$ in terms of only $X_1, X_2, \ldots, X_n$. Then use the results from (b).)

(d) Propose a modified estimation $\hat{\sigma}^2_{\text{unbias}}$ which does satisfy the property.