INEQUALITIES, LLSE 11

COMPUTER SCIENCE MENTORS 70

November 13-17, 2017

1 Inequalities

1.1 Introduction

Markov's Inequality
For a non-negative random variable X with expectation $\mathrm{E}(X) = \mu$, and any $\alpha > 0$:

$$\mathrm{P}[X \geq \alpha] \leq \frac{\mathrm{E}(X)}{\alpha}$$

Solution: Proof of Markov's Inequality

$$\begin{aligned}
E(X) &= \sum_a a * Pr[X = a] \\
&\geq \sum_{a \geq \alpha} a * Pr[X = a] \\
&\geq \alpha \sum_{a \geq \alpha} Pr[X = a] \\
&= \alpha Pr[X \geq a]
\end{aligned}$$

Chebyshev's Inequality
For a random variable X with expectation $E(X) = \mu$, and any $\alpha > 0$:

$$P[|X - \mu| \geq \alpha] \leq \frac{\text{Var}(X)}{\alpha^2}$$

1.2 Questions

1. Use Markov's to prove Chebyshev's Inequality:

Solution: Define the random variable $Y = (X - \mu)^2$. Note that $E(Y) = E((X - \mu)^2) = \text{Var}(X)$ Also, notice that the event that we are interested in, $|X\mu|\alpha$ is exactly the same as the event $Y = (X\mu)^2\alpha^2$. Therefore, $P[|X\mu| \geq \alpha] = P[Y \geq \alpha^2]$. Moreover, Y is non-negative, so we can apply Markov's inequality to it to get:

$$P[Y \geq \alpha^2] \leq \frac{E(Y)}{\alpha^2} = \frac{\text{Var}(X)}{\alpha^2}$$

Substituting in $Y = (X - \mu)^2$ above and taking square roots yields the form we are used to.

2. Let X be the sum of 20 i.i.d. Poisson random variables $X_1, \ldots, X_{20}$ with $E(X_i) = 1$. Find an upper bound of $P[X \geq 26]$ using,

(a) Markov's inequality:

Solution:

$$\begin{aligned} P[X \geq a] &\leq \frac{E(X)}{a} \text{ for all } a > 0 \\ P[X \geq 26] &\leq \frac{20}{26} \\ &\approx 0.769 \end{aligned}$$

(b) Chebyshev's inequality:

Solution:

$$\begin{aligned} \mathrm{P}[|X - \mathrm{E}(X) \geq c] &\leq \frac{\sigma_X^2}{c^2} \\ \mathrm{P}[|X - 20| \geq 6] &\leq \frac{20}{36} \\ &\approx 0.5556 \end{aligned}$$

3. **Bound It**

A random variable X is always strictly larger than -100. You know that $\mathrm{E}(X) = -60$. Give the best upper bound you can on $\mathrm{P}[X \geq -20]$.

Solution: Notice that we do not have the variance of X, so Chebyshev's bound is not applicable here. Since X is also not a sum of other random variables, other bounds or approximations(chernoff, Hoeffding's inequality. Don't worry about them if they are not covered.) are not available. This leaves us with just Markov's Inequality. But Markov Bound only applies on a non-negative random variable, whereas X can take on negative values.

This suggests that we want to shift X somehow, so that we can apply Markovs Inequality on it. Define a random variable $Y = X + 100$, which means Y is strictly larger than 0, since X is always strictly larger than -100. Then, $\mathrm{E}(Y) = \mathrm{E}(X + 100) = \mathrm{E}(X) + 100 = -60 + 100 = 40$. Finally, the upper bound on X that we want can be calculated via Y, and we can now apply Markov's Inequality on Y since Y is strictly positive.

$\mathrm{P}[X \geq -20] = \mathrm{P}[Y \geq 80] \leq \frac{\mathrm{E}(Y)}{80} = \frac{40}{80} = \frac{1}{2}$

Hence, the best upper bound on $\mathrm{P}[X \geq -20]$ is $\frac{1}{2}$.

4. The citizens of the country USD (the United States of Drumpf) vote in the following manner for their presidential election: if the country is liberal, then each citizen votes for a liberal candidate with probability p and a conservative candidate with probability $1p$, while if the country is conservative, then each citizen votes for a conservative candidate with probability p and a liberal candidate with probability $1p$. After the election, the country is declared to be of the party with the majority of the votes.

   (a) Assume that $p = \frac{3}{4}$ and suppose that 100 citizens of USD vote in the election and that USD is known to be conservative. Provide a tight bound on the probability that it is declared to be a Liberal country.

Solution: Let X_i be the indicator that voter i votes as a Liberal. We are interested in boudning the quantity $\mathrm{P}[S_{100} \geq 51]$ where $S_{100} = X_1 + X_2 + \ldots + X_{100}$. We have:

$$\mathrm{P}[S_n \geq 51] = \mathrm{P}[X - 25 \geq 26] \leq \mathrm{P}[|X - 25| \geq 26] \leq \frac{\mathrm{Var}(X)}{26^2} = \frac{75}{4 \cdot 26^2} = 0.03$$

(b) Now let p be unknown; we wish to estimate it. Using the CLT, determine the number of voters necessary to determine p within an error of 0.01, with probability at least 0.95.

Solution: For now, we let consider general error and want the probability to be at least 1β . We are thus interested in:

$$\mathrm{P}[|\frac{S_n}{n}p| \geq \alpha]$$

Note that by the CLT,

$$\frac{S_n}{n}p \approx \sqrt{\frac{p \cdot (1p)}{n}}Z, \text{ where } Z \sim N(0, 1).$$

$$\mathrm{P}[|\frac{S_n}{n}p| \geq \alpha] \approx \mathrm{P}[|Z| \geq \sqrt{\frac{p \cdot (1p)}{n}}\alpha \leq \mathrm{P}[|Z| \geq 2\alpha]$$

Now, we have:

$$\mathrm{P}[|Z| \geq 2\alpha\sqrt{n}] = 2\mathrm{P}[Z \geq 2\alpha\sqrt{n}] = 2 \cdot \mathrm{P}[X \leq 2\alpha\sqrt{n}] = \beta$$

Now, we substitute in $\alpha = 0.01$, $\beta = 0.05$, and see that: $n = \frac{1.962}{(2 \cdot 0.01)^2} = 98^2 = 9604$.

5. **Squirrel Standard Deviation**
As we all know, Berkeley squirrels are extremely fat and cute. The average squirrel is 40% body fat. The standard deviation of body fat is 5%. Provide an upper bound on the probability that a randomly trapped squirrel is either too skinny or too fat? A skinny squirrel has less than 27.5% body fat, and a fat squirrel has more than 52.5% body fat?

Solution: We use Chebyshev's inequality. We are looking for the probability we fall within 2.5 standard deviations of the mean. By Chebyshev's inequality, the probability we are within this range is $\frac{1}{2.5^2}, or \frac{4}{25} = 0.16$ If we were to use Markov's inequality, we would have probabilities over 1, which yields a non-helpful value.

6. Consider a random variable Y with expectation μ whose maximum value is $\frac{3\mu}{2}$, prove that the probability that Y is 0 is at most $\frac{1}{3}$.

Solution:

$$\begin{aligned}
\mu &= \sum_a a * \text{P}[Y = a] \\
&= \sum_{a \neq 0} a * \text{P}[Y = a] \\
&\leq \sum_{a \neq 0} \frac{3\mu}{2} * \text{P}[Y = a] \\
&= \frac{3\mu}{2} * \sum_{a \neq 0} \text{P}[Y = a] \\
&= \frac{3\mu}{2} * (1 - \text{P}[Y = 0])
\end{aligned}$$

This implies that $\text{P}[Y = 0] \leq \frac{1}{3}$

1.3 Covariance

1.4 Introduction

The **covariance** of two random variables X and Y is defined as:

$$\text{Cov}(X, Y) := \text{E}((X - \text{E}(X)) \cdot (Y - \text{E}(Y)))$$

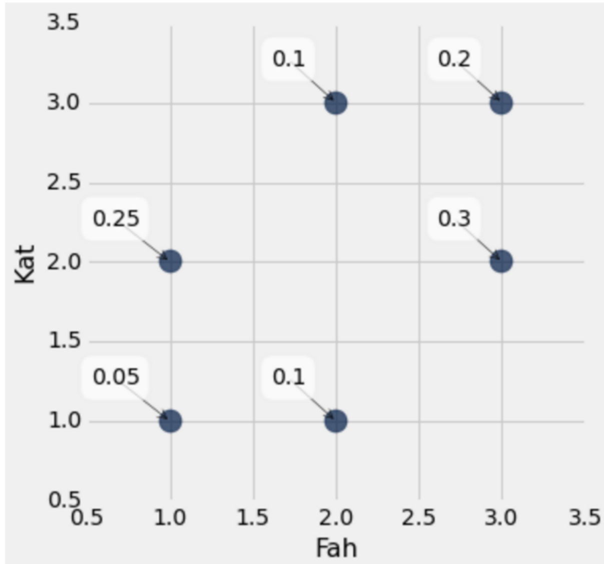
1.5 Questions

1. Prove that $\text{Cov}(X, X) = \text{Var}(X)$:

Solution:

$$\text{Cov}(X, X) = \text{E}(X \cdot X) - \text{E}(X) \cdot \text{E}(X) = \text{E}(X^2) - \text{E}(X)^2$$

2. Consider the following distribution with random variables Fah and Kat:



Find the covariance of Fah and Kat.

Solution: E(Fah) $= 1 \cdot .3 + 2 \cdot .2 + 3 \cdot .5 = \mathbf{2.2}$
E(Kat) = $1 \cdot .15 + 2 \cdot .55 + 3 \cdot .3 = \mathbf{2.15}$
E(KatFah) = $1 \cdot 1 \cdot .05 + 1 \cdot 2 \cdot .25 + 2 \cdot 1 \cdot .1 + 2 \cdot 3 \cdot .1 + 3 \cdot 2 \cdot .3 + 3 \cdot 3 \cdot .2 = \mathbf{4.95}$
cov(Kat, Fah) = $4.95 - 2.2 \cdot 2.15 = \mathbf{0.22}$

3. Prove that if X and Y are independent, then Cov$(X, Y) = 0$:

Solution:

$$\text{Cov}(X, Y) = \text{E}(X \cdot Y) - \text{E}(X) \cdot \text{E}(Y)$$

Remember that a property of expectation is that if X and Y are independent, then $\text{E}(XY) = \text{E}(X) \cdot \text{E}(Y)$, so we get 0 when we subtract

4. Roll 2 dice. Let A be the number of 6's you get, and B be the number of 5's, find Cov(A, B)

Solution: $\text{E}(A) = \frac{1}{6}$ for one die, by linearity of expectation, two dice make $\frac{1}{3}$, same for $\text{E}(B)$ $\text{E}(A) = \frac{1}{3}, \text{E}(B) = \frac{1}{3}$
AB can be either 0 (if no 5's or 6's show up) or 1 (get a 5 and a 6).

$$\begin{aligned}\text{E}(AB) &= 1 \cdot \text{P[get a 5 and a 6]} \\ &= \text{P[first die = 5 and second die = 6]} + \text{P[first die = 6 and second die = 5]} \\ &= \frac{1}{36} + \frac{1}{36}\end{aligned}$$

$$\begin{aligned}\text{Cov}(AB) &= \text{E}(AB) - \text{E}(A)\cdot\text{E}(B)\\ &= \frac{1}{18} - \frac{1}{9}\\ &= -\frac{1}{18}\end{aligned}$$

5. Prove that $\text{Cov}(X+Y,Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$:

Solution:

$$\begin{aligned}\text{Cov}(X+Y,Z) &= \text{E}((X+Y)\cdot Z) - \text{E}(X+Y)\cdot\text{E}(Z)\\ &= \text{E}(X\cdot Z) + \text{E}(Y\cdot Z) - (\text{E}(X)\cdot\text{E}(Z) + \text{E}(Y)\cdot\text{E}(Z))\\ &= \text{E}(X\cdot Z) - \text{E}(X)\cdot\text{E}(Z) + \text{E}(Y\cdot Z) - \text{E}(Y)\cdot\text{E}(Z)\\ &= \text{Cov}(X,Z) + \text{Cov}(Y,Z)\end{aligned}$$

2 Linear Least Squares Estimator

Theorem: Consider two random variables, X, Y with a given distribution $\text{P}[X=x, Y=y]$. Then

$$\text{L}[Y|X] = \text{E}(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \text{E}(X))$$

2.1 Questions

1. Assume that

$$Y = \alpha X + Z$$

where X and Z are independent and $\text{E}(X) = \text{E}(Z) = 0$. Find $\text{L}[X|Y]$.

Solution:

$$\begin{aligned}\text{Cov}(X,Y) &= \text{E}(XY) - \text{E}(X)\text{E}(Y)\\ &= \text{E}(X\cdot(\alpha X + Z)) = \alpha\text{E}(X^2)\end{aligned}$$

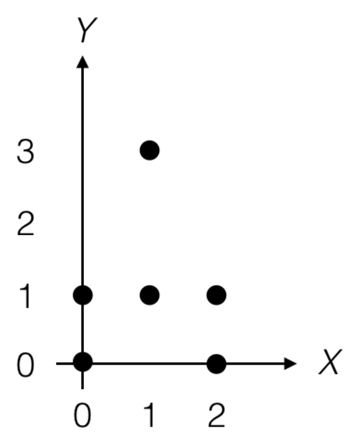
$$\text{Var}(Y) = \alpha^2\text{Var}(X) + \text{Var}(Z)$$

$$= \alpha^2 \mathrm{E}(X^2) + \mathrm{E}(Z^2)$$

Therefore,

$$\mathrm{L}[X|Y] = \frac{\alpha \mathrm{E}(X^2)}{\alpha^2 \mathrm{E}(X^2) + \mathrm{E}(Z^2)} \cdot Y$$

2. The figure below shows the six equally likely values of the random pair (X, Y). Specify the functions of:
   - $L[Y \mid X]$
   - $E(X \mid Y)$
   - $L[X \mid Y]$
   - $E(Y \mid X)$



Solution: Let's calculate some useful properties of the distribution first and then see how we can use them to calculate the estimates.

$$|\Omega| = 6 \implies P[\text{one point}] = \frac{1}{6}$$

$$\mathrm{E}(X) = 0\left(\frac{2}{6}\right) + 1\left(\frac{2}{6}\right) + 2\left(\frac{2}{6}\right)$$
$$= 1$$
$$\mathrm{E}(Y) = 0\left(\frac{2}{6}\right) + 1\left(\frac{3}{6}\right) + 30\left(\frac{1}{6}\right)$$
$$= 1$$

$$\begin{aligned}\mathrm{E}(XY) &= 0\left(\frac{3}{6}\right)+1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+3\left(\frac{1}{6}\right)\\ &= 1\\ \mathrm{Cov}(X,Y) &= 0 \implies \mathrm{L}[Y|X]=E(Y)\end{aligned}$$

- $L[Y \mid X]$: Using the LLSE formula: $\mathrm{L}[Y \mid X] = \mathrm{E}[Y] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathrm{E}[Y]) = \mathrm{E}[Y]$. Therefore $\boxed{L[Y \mid X] = 1}$
- $E[X \mid Y]$: Notice the symmetry across $X = 1$. For all values of y, $E[X|Y = y]$ is the same; therefore $\boxed{E[X|Y] = E[X] = 1}$.
- $L[X \mid Y]$: The MMSE estimator for X given Y is a linear function, therefore $\boxed{\mathrm{L}[X \mid Y] = \mathrm{E}[X \mid Y] = 1}$
- $E[Y \mid X]$ For this one we can't make use of symmetry or directly apply what we calculated above. We must go back to the definition of conditional expectation. We can calculate $\mathrm{E}[Y \mid X = x]$ for every point x, and that entirely defines the expression:

$$\mathrm{E}(Y \mid X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ 2 & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 2 \end{cases}$$

The above equation is sufficient, but we can go further by realizing that these points are part of a flipped absolute value function centered around $x = 1$: $\boxed{\mathrm{E}[Y \mid X] = \frac{-3}{2}|X - 1| + 2}$. Indeed, this is not linear, which is why $\mathrm{L}[Y \mid X] \neq \mathrm{E}[Y \mid X]$.