INEQUALITIES, LLSE 11

COMPUTER SCIENCE MENTORS 70

April 16-18, 2018

1 Normal Distribution

Gaussian (Normal) Distribution: $N(\mu, \sigma^2)$

- Mean: μ
- Variance: σ^2
- $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
- If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_X, \sigma_X^2)$, then $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- The CLT (Central Limit Theorem) states that for a sequence of iid random variables: $X_1, X_2, ..., X_n$, each with mean μ and variance σ^2,

$$\frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

approaches the standard normal distribution $Z \sim N(0, 1)$

1.1 Questions

1. **Coin Flipping**
Suppose you have two coins, one that has heads on both sides and another that has tails on both sides. You pick one of the two coins uniformly at random and flip it. You repeat this process 400 times, each time picking one of the two coins uniformly at random and then flipping it, for a total of 400 flips.
Use the CLT to approximate the probability of getting more than 220 heads.

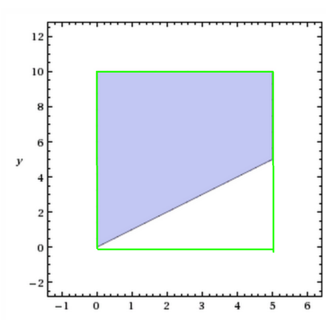
Solution: Let X be the number of heads we get. We have a $\frac{1}{2}$ probability of getting the coin with 2 heads and thus getting a head and we have a $\frac{1}{2}$ probability of getting the coin with 2 tails and not getting heads. Thus $X\sim\text{Bin}(400, \frac{1}{2})$.
Now, we know that $E[X] = 400 * \frac{1}{2} = 200$ and $\text{Var}(X) = 200 * \frac{1}{2} * (1 - \frac{1}{2}) = 100$, so standard deviation $\sigma = \sqrt{100} = 10$
Using CLT, we approximate $\frac{X-E[X]}{\sigma} = \frac{X-200}{10}$ as $Z\sim N(0,1)$
Therefore, $P(X > 200) = P(\frac{X-200}{10} > 2) \approx P(Z > 2) = 1 - P(Z \leq 2) \approx 1 - 0.9772$ $= 0.0228$.

2 More Continuous Practice

2.1 Questions

1. Alice and Bob are throwing baseballs and they want to see who can throw a baseball further. The distance Alice throws a baseball is modeled as a uniform distribution between 0 and 5 while the distance Bob throws a baseball is modeled as a uniform distribution between 0 and 10. Assume Alice and Bob throw independently, what is the probability that Bob's throw will be further than Alice's?

Solution: Let $A\sim U[0,5]$ be the distance of Alice's throw and let $B\sim U[0,10]$ be the distance of Bob's throw.
We know the density of A is $\frac{1}{5}$ and the density of B is $\frac{1}{10}$
We can draw the following graph where x is Alice's throw and y is Bob's throw.



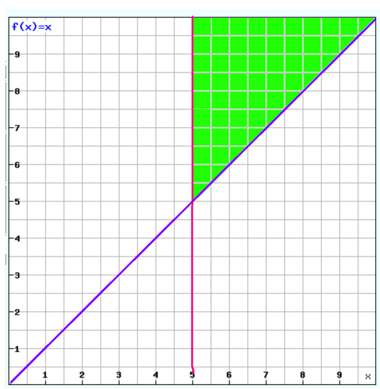
The green outline is the entire sample space while the shaded region is the region of interest. As we can see, the blue region takes up $\frac{3}{4}$ of the total sample space.

A more algebraic approach is to use double integrals.
We fix Alice's throw to be between 0 and 5 and we only consider Bob's throw if it

is greater than Alice's.
The joint pdf $f_{A,B} = f_A * f_B = \frac{1}{50}$ as A and B are independent.
Therefore,
$P(B > A)$
$= \int_0^5 \int_a^{10} f_{A,B} db da$
$= \int_0^5 \int_a^{10} \frac{1}{50} db da$
$= \int_0^5 (\frac{1}{50} b)\Big|_a^{10} da$
$= \int_0^5 \frac{1}{5} - \frac{a}{50} da$
$= (\frac{1}{5}a - \frac{a^2}{100})\Big|_0^5 = 1 - \frac{25}{100} = \frac{3}{4}$

2. Now Alice has improved her throwing abilities and her throwing distance is now also uniform on the interval 0 to 10, which is the same as Bob.
Given that Alice's throw was greater than 5, what is the probability that Bob throws further than her?

Solution: Let $A \sim U[0, 10]$ be the distance of Alice's throw and let $B \sim U[0, 10]$ be the distance of Bob's throw.
We know the density of A is $\frac{1}{10}$ and the density of B is $\frac{1}{10}$
We want $P(B > A | A > 5)$
We can draw the following graph where x is Alice's throw and y is Bob's throw.



Here let the y-axis be Bob's throw and let the x-axis be Alice's throw.
Here, the purple line represents the region in which Bob throws further than Alice.
The red line down the middle represents the fact that Alice's throw was greater

than 5, so we focus our attention on everything that is right of the red line.
The green shaded region represents the probability that Bob's throw is greater than Alice's, conditioned on the fact that Alice's throw was greater than 5.
Therefore, we can see that the green shaded region is $\frac{1}{4}$ of the entire region to the right of the red line.

We can also take an algebraic approach and use double integrals.
Here we fix Alice's throw to be between 5 and 10 as we know she threw for more than 5.
As like the previous problem, we only consider when Bob throws a distance greater than Alice.
The joint pdf $f_{A,B} = f_A * f_B = \frac{1}{100}$ as A and B are independent.
Therefore,
$P(B > A|A > 5) = \frac{P(B>A,A>5)}{P(A>5)} = 2 * P(B > A, A > 5)$
$= 2 \int\limits_5^{10} \int\limits_a^{10} f_{A,B} dbda$
$= 2 \int\limits_5^{10} \int\limits_a^{10} \frac{1}{100} dbda$
$= 2 \int\limits_5^{10} (\frac{1}{100} b)\Big|_a^{10} da$
$= 2 \int\limits_5^{10} \frac{1}{10} - \frac{a}{100} da$
$= 2(\frac{1}{10}a - \frac{a^2}{200})\Big|_5^{10} = 2((1 - \frac{100}{200}) - (\frac{1}{2} - \frac{25}{200})) = 2(1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{8}) = 2 * \frac{1}{8} = \frac{1}{4}$

3 Inequalities

3.1 Introduction

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

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

Proof of Markov's Inequality

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

Alternate Proof of Markov's Inequality
Consider the indicator random variable Y which equals 1 if $X \geq a$ and 0 otherwise. Now consider the relationship between X and aY

- If $X < a$, then $Y = 0$, which means $aY = a * 0 = 0$
  Because X is a non-negative random variable, $X \geq 0$, so $aY \leq X$ in this case.
- If $X \geq a$, then $Y = 1$, which means $aY = a * 1 = a \leq X$

Thus, we have $aY \leq X$.
We take expectation on both sides to get:

$$E[aY] \leq X$$

$$aE[Y] \leq E[X]$$

$$E[Y] \leq \frac{E[X]}{a}$$

Now we note that the expectation of an indicator random variable is the probability that it is equal to 1 and we have the proof:

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

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

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

Chebyshev's Inequality can be used to estimate the mean of an unknown distribution. Often times we do not know the true mean, so we take many samples $X_1, X_2, ..., X_n$. Our sample mean is thus $S_n = \frac{X_1+X_2+...+X_n}{n}$
We can upper-bound the probability of that our sample mean deviates too much from our true mean as:

$$P(|\hat{\mu} - \mu| > \epsilon) \leq \delta$$

where ϵ is known as our error and δ is known as our confidence.

3.2 Questions

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

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

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

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 $\mathrm{E}(X_i) = 1$. Find an upper bound of $\mathrm{P}[X \geq 26]$ using,

   (a) Markov's inequality:

Solution:

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

(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. 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 $1-p$, while if the country is conservative, then each citizen votes for a conservative candidate with probability p and a liberal candidate with probability $1 - p$. 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:

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

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

Solution: Again let X_i be the indicator variable that person i votes as a liberal. Let $\hat{p}$ be our sample proportion.
We have:

$$\hat{p} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

We want our sample proportion to be within 0.01 of p with probability 0.95. Thus, we want to bound:

$$\mathrm{P}[|\hat{p} - p| \geq 0.01]$$

We know that $\mathrm{Var}(\hat{p}) = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathrm{Var}(X_i) = p(1-p) \leq \frac{1}{4}$.
Therefore, our variance $\mathrm{Var}(\hat{p}) \leq \frac{1}{4n}$

$$\mathrm{P}[|\hat{p} - p| \geq 0.01] \leq \frac{\mathrm{Var}(\hat{p})}{0.01^2} \leq \frac{1}{4 * 0.01^2 n} = 0.05$$

Solving for n, we get: $n \geq 50000$.

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:

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

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

7. Let $X_1, X_2, ..., X_n$ be n iid Geometric random variables with parameter p. Using Chebyshev's inequality, provide an upper-bound on: $P(|\frac{X_1+X_2+...+X_n}{n} - \frac{1}{p})| \geq a)$
Recall the variance for a Geometric Distribution with parameter p is $\frac{1-p}{p^2}$

Solution: Let $S_n = \frac{X_1+X_2+...+X_n}{n}$.
$E[S_n] = \frac{1}{n} * E[X_1 + X_2 + ... + X_n] = \frac{1}{n} * n * E[X_1] = \frac{1}{p}$
$var(S_n) = \frac{1}{n^2} var(X_1 + X_2 + ... + X_n) = \frac{1}{n^2} * n * var(X_1) = \frac{1-p}{np^2}$
Therefore, $P(|S_n - E[S_n]| \geq a) \leq \frac{var(S_n)}{a^2} = \frac{1-p}{p^2 n a^2}$

8. Suppose we have a sequence of iid random variables $X_1, X_2, ..., X_n$
Let $A_n = \frac{X_1+X_2+...+X_n}{n}$ be the sample mean.
Show that the true mean of $X_i = \mu$ is within the interval $[\mu - 4.5\frac{\sigma}{\sqrt{n}}, \mu + 4.5\frac{\sigma}{\sqrt{n}}]$ with 95% probability.

Solution: Using Chebyshev's Inequality, we have:
$P(|A_n - \mu| > 4.5\frac{\sigma}{\sqrt{n}}) \leq \frac{var(A_n)}{(4.5\frac{\sigma}{\sqrt{n}})^2}$
$var(A_n) = var(\frac{X_1+X_2+...X_n}{n}) = \frac{1}{n^2} var(X_1 + X_2 + ... + X_n) = \frac{1}{n^2} * n * var(X_1) = \frac{\sigma^2}{n}$
$P(|A_n - \mu| > 4.5\frac{\sigma}{\sqrt{n}}) \leq \frac{\frac{\sigma^2}{n}}{20\frac{\sigma^2}{n}} = \frac{1}{20}$.
The probability that you are outside of the confidence interval is 0.05, so therefore, the probability that the true mean lies within the interval is $1 - 0.05 = 0.95$.

3.3 Covariance

3.4 Introduction

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

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

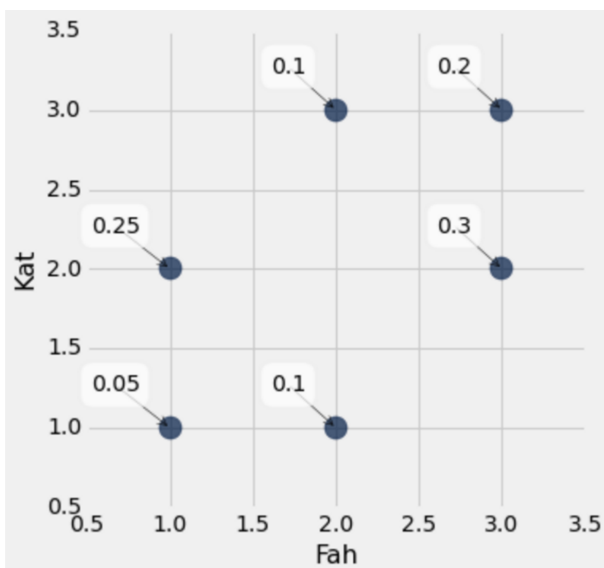
3.5 Questions

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) $= 1 \cdot .3 + 2 \cdot .2 + 3 \cdot .5 =$ **2.2**
E(Kat) = $1 \cdot .15 + 2 \cdot .55 + 3 \cdot .3 =$ **2.15**

E(KatFah) = $1 \cdot 1 \cdot .05 + 1 \cdot 2 \cdot .25 + 2 \cdot 1 \cdot .1 + 2 \cdot 3 \cdot .1 + 3 \cdot 2 \cdot .3 + 3 \cdot 3 \cdot .2 =$ **4.95**
cov(Kat, Fah) = $4.95 - 2.2 \cdot 2.15 =$ **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 E$(XY) =$ E$(X) \cdot$ E(Y), so we get 0 when we subtract

4. Prove that the converse is not necessarily true. In other words, give an example of 2 random variables whose covariance is 0 but are not independent.

Solution: Let $X \sim U(-1, 1)$ and let $Y = X^2$

We have $E[X] = \frac{-1+1}{2} = 0$
The pdf of X is $\frac{1}{2}$, so
$E[Y] = E[X^2] = \int_{-1}^{1} x^2 f_X(x)dx = \int_{-1}^{1} \frac{1}{2}x^2 = \frac{1}{3}$
We also have $E[XY] = E[X * X^2] = E[X^3] = \int_{-1}^{1} x^3 f_X(x)dx = \int_{-1}^{1} \frac{1}{2}x^3 dx = 0$
cov$(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 * \frac{1}{3} = 0$
However, Y is clearly dependent on X as $Y = X^2$, and thus we provide a valid counterexample.

5. 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: E$(A) = \frac{1}{6}$ for one die, by linearity of expectation, two dice make $\frac{1}{3}$, same for E(B) E$(A) = \frac{1}{3}$, 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).

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

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

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

Solution:

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

4 Linear Least Squares Estimator

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

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

4.1 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:

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

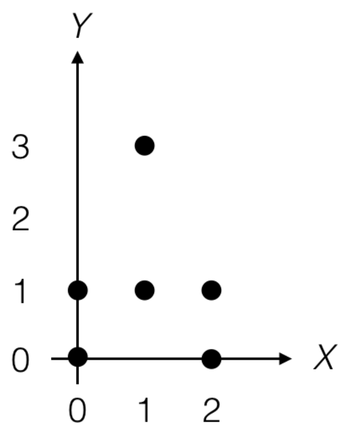
$$\text{Var}(Y) = \alpha^2\text{Var}(X) + \text{Var}(Z)$$

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

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 $\mathrm{L}[Y \mid X] \neq \mathrm{E}[Y \mid X]$.