DCP 1206: Probability
Lecture 16 — Covariance and Conditional Distributions

Ping-Chun Hsieh

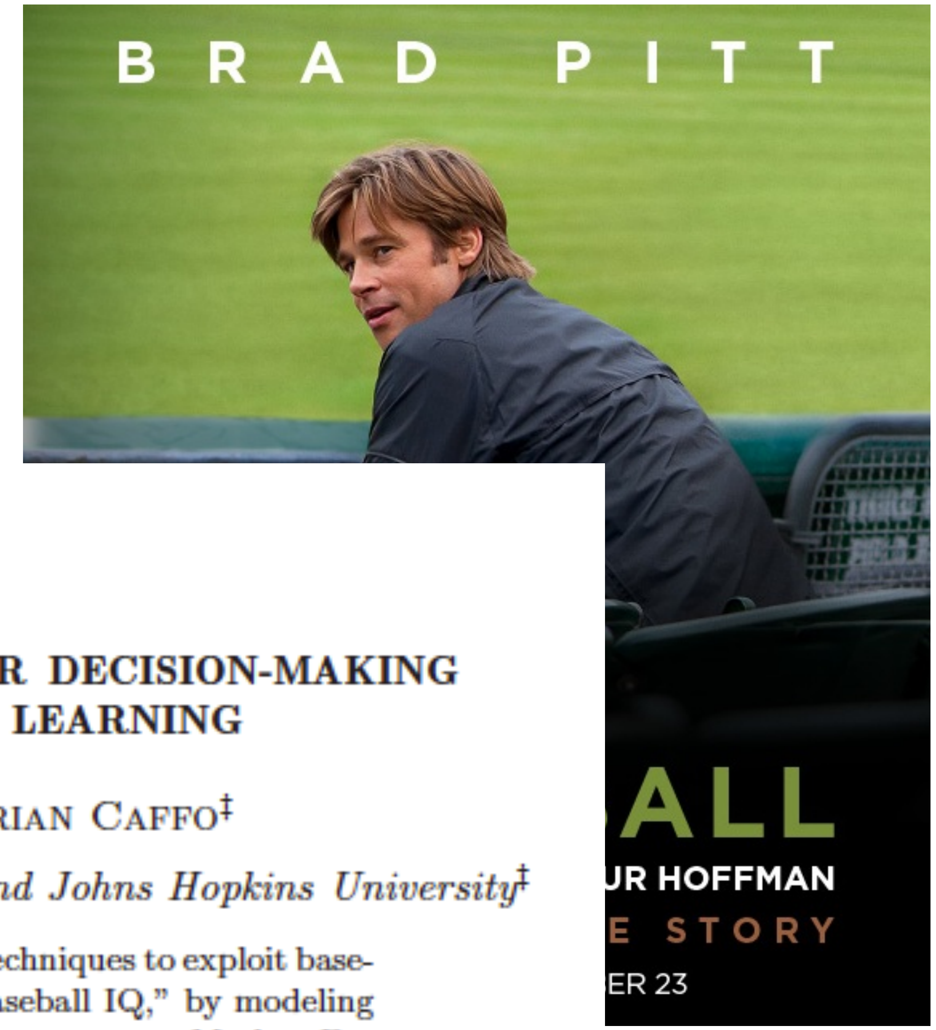
November 13, 2019

Announcements

- Midterm course evaluation (期中意見調查)
  - Google form: shorturl.at/ixAFP (posted on E3)
  - 11/8 (Friday) ~ 11/13 (Wednesday)
- HW4 is posted on E3
  - Part 1: due on 11/22 (Friday)
  - Part 2: due on 11/27 (Wednesday)
- Midterm exam booklet will be returned on Friday after class
  - 4:30pm - 5:30pm @ EC122

Baseball, Sabermetrics, and Precision Sports

12強》張奕好投與陳俊秀三分砲 中華隊7比0擊潰韓國



The Annals of Applied Statistics
2014, Vol. 8, No. 2, 926–955
DOI: 10.1214/13-AOAS712

MONEYBaRL: EXPLOITING PITCHER DECISION-MAKING USING REINFORCEMENT LEARNING

BY GAGAN SIDHU[*,†] AND BRIAN CAFFO[‡]

General Analytics[*1], *University of Alberta*[†] *and Johns Hopkins University*[‡]

This manuscript uses machine learning techniques to exploit baseball pitchers' decision making, so-called "Baseball IQ," by modeling the at-bat information, pitch selection and counts, as a Markov Decision Process (MDP). Each state of the MDP models the pitcher's current pitch selection in a Markovian fashion, conditional on the information immediately prior to making the current pitch. This includes the count prior to the previous pitch, his ensuing pitch selection, the batter's ensuing action and the result of the pitch.

The necessary Markovian probabilities can be estimated by the relevant observed conditional proportions in MLB pitch-by-pitch game data. These probabilities could be pitcher-specific, using only the data from one pitcher, or general, using the data from a collection of pitchers.

Quick Review

- Given 2 random variables X, Y: what have we learned so far?
  1. Joint CDF
  2. Marginal CDF
  3. Joint PMF / PDF
  4. Marginal PMF / PDF
  5. Independence
  6. Expected value involving both X, Y
  7. Covariance and correlation
  8. Conditional distribution
  9. Distribution of $X + Y$

This Lecture

1. Expected Value Regarding 2 Random Variables

2. Covariance and Correlation Coefficient

3. Conditional Distributions

- Reading material: Chapter 8.2-8.3 and 10.2-10.3

1. Expected Value Regarding Two Random Variables

LOTUS for 2 Discrete Random Variables

Expected Value of a Function of 2 Discrete RVs:

1. Let X, Y be 2 discrete random variables with sets of possible values S_X, S_Y and joint PMF $p(x, y)$

2. Let $g(\cdot\,,\cdot)$ be a function from $\mathbb{R}^2 \to \mathbb{R}$

The expected value of $g(X, Y)$ is

$$E[g(X, Y)] = \sum_{\substack{x \in S_X \\ y \in S_Y}} \big(g(x,y) \big) \big(\overset{\text{PMF}}{p(x,y)} \big)$$

$$= \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) \cdot p(x,y)$$

LOTUS for 2 Continuous Random Variables

Expected Value of a Function of 2 Continuous RVs:

1. Let X, Y be 2 continuous random variables with joint PDF $f_{XY}(x, y)$

2. Let $g(\,\cdot\,,\cdot\,)$ be a function from $\mathbb{R}^2 \rightarrow \mathbb{R}$

The expected value of $g(X, Y)$ is

$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \underbrace{(g(x,y))}_{\text{function value}} \underbrace{(f_{XY}(x,y))}_{\text{PDF}} \, dx\, dy$$

Useful Property (I)

- **Linearity Property**:

$$E[\alpha \cdot g_1(X, Y) + \beta \cdot g_2(X, Y)] = \alpha E[g_1(X, Y)] + \beta E[g_2(X, Y)]$$

- Remark: X, Y are NOT required to be independent
- Remark: This results holds for both discrete and continuous cases
- Proof: $E[\alpha \cdot g_1 + \beta \cdot g_2] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (\alpha g_1 + \beta g_2)(f(x,y))\,dx\,dy$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \alpha g_1 \cdot f(x,y)\,dx\,dy = \alpha \cdot E[g_1(X,Y)]$$

$$+ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \beta g_2 \cdot f(x,y)\,dx\,dy = \beta \cdot E[g_2(X,Y)]$$

Useful Property (II)

- **Property under independence**: Suppose X, Y are independent random variables. Then, we have

$$\underbrace{E[g(X) \cdot h(Y)]}_{\ell(X,Y)} = E[g(X)] \cdot E[h(Y)]$$

- Remark: This result holds for both discrete and continuous cases
- Proof:

$$E[g(X) \cdot h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (g(x) \cdot h(y)) (f_{XY}(x,y)) \, dx\,dy \qquad f_{XY}(x,y) \text{ " } f_X(x) \cdot f_Y(y)$$

$$= \int\int g(x) h(y) f_X(x) f_Y(y) \, dx\,dy$$

$$= \int h(y) f_Y(y) \left(\int g(x) f_X(x) \, dx \right) dy$$

$$= \int h(y) f(y) E[g(X)] \, dy$$

$$= E[g(X)] \cdot E[h(Y)]$$

$E[XY] = E[X]E[Y]$ If X, Y Are Independent

$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$

Corollary: Let X, Y be 2 independent random variables. Then,

$$E[XY] = E[X]E[Y]$$

$g(X) = X$

$h(Y) = Y$

- Question: How about the reverse argument?

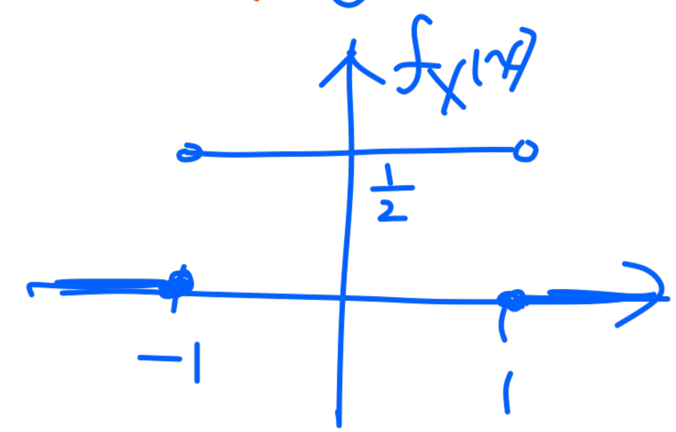
independence $\Rightarrow E[XY] = E[X]E[Y]$

$E[XY] = E[X]E[Y] \nRightarrow X, Y$ Independent

- Example: Let X be a continuous uniform r.v. on $[-1,1]$.
  - Define $Y = X^2$
  - $E[X] =$? $E[Y] =$?
  - $E[XY] =$?
  - Are X, Y independent?

Unif(a,b) $\frac{1}{b-a}$

$$f_X(x) = \begin{cases} \frac{1}{2} & , -1<x<1 \\ 0 & , \text{else} \end{cases}$$



$E[X] = 0$

$$E[Y] = E[X^2] = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\, dx = \frac{1}{6}x^3 \Big|_{-1}^{1} = \frac{1}{3}$$

$$E[XY] = E[X^3] = \int_{-1}^{1} x^3 \cdot \frac{1}{2}\, dx = \frac{1}{8}x^4 \Big|_{-1}^{1} = 0$$

$E[XY] = 0$, $E[X] \cdot E[Y] = 0$

Example: X, Y Not Independent and $E[XY] \neq E[X]E[Y]$

- Example: Let X be a continuous uniform r.v. on $[-1,1]$.
  - Define $Y = X$
  - Are X, Y independent?
  - $E[X] = ?\ E[Y] = ?$
  - $E[XY] = ?$

$E[XY] = E[X^2] = \frac{1}{3}$

$E[X] = 0$

Remark:

Please see last page!

$E[Y] = 0$

$E[XY] \neq E[X] \cdot E[Y]$.

When Do We See $E[XY]$?

- Example: Var$[X + Y]$

$$= \left(E[(X+Y)^2] \right) - \left(\left(E[X+Y] \right)^2 \right)$$

$$= E[X^2 + 2XY + Y^2] - \left(\left(E[X+Y] \right)^2 \right)$$

$$= E[X^2] + 2 \cdot E[XY] + E[Y^2] - \left(\left(E[X+Y] \right)^2 \right)$$

More on $E[XY]$: Cauchy-Schwarz Inequality

- Recall: Cauchy Inequality in high school

$a_1, a_2, \cdots, a_n \in \mathbb{R}$

$b_1, b_2, \cdots, b_n \in \mathbb{R}$

比例時等號成立

$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \cdots = \frac{a_n}{b_n}$

Cauchy inequality:

平方和乘積

$$\left((a_1)^2 + (a_2)^2 + \cdots + (a_n)^2\right)\left((b_1)^2 + (b_2)^2 + \cdots + (b_n)^2\right)$$

$$\geq \left((a_1)(b_1) + (a_2)(b_2) + \cdots + (a_n)(b_n)\right)^2$$

乘積和平方

Cauchy-Schwarz Inequality

- **Cauchy-Schwarz Inequality**: Let X, Y be two random variables. Then, we have

$$E[X^2] \cdot E[Y^2] \geq (E[XY])^2$$

平方和 平方和 乘積和

$Y = \alpha X$ with probability 1

- **Question:** Under what condition do we have "="?

	$X=a_1$	$X=a_2$
$Y=b_1$	p_{11}	p_{21}
$Y=b_2$	p_{12}	p_{22}

$$E[X^2] = p_{11} \cdot a_1^2 + p_{21} a_2^2 + p_{12} a_1^2 + p_{22} a_2^2$$

$$E[Y^2] = p_{11} (b_1^2) + p_{21} (b_1^2) + p_{12} (b_2^2) + p_{22} (b_2^2)$$

$$E[XY] = p_{11} (a_1 b_1) + p_{21} (a_2 b_1) + p_{12} (a_1 b_2) + p_{22} (a_2 b_2)$$

Proof of Cauchy-Schwarz Inequality

$$E[X^2] \cdot E[Y^2] \ge (E[XY])^2$$

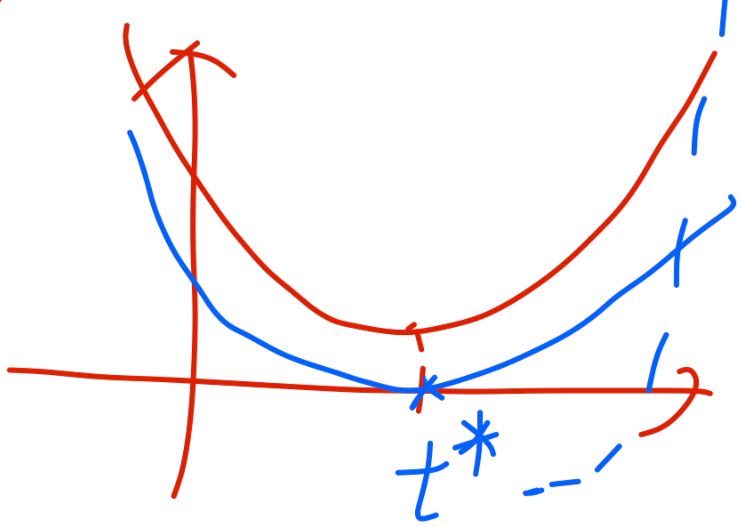
- **Hint:** Start from that $E[(tX + Y)^2] \ge 0$, for all $t \in \mathbb{R}$
- **Proof:**

$$0 \le E[(tX+Y)^2] = E[t^2X^2 + 2\cdot tXY + Y^2]$$

$$= (E[X^2])t^2 + (2\cdot E[XY])t + (E[Y^2])$$

Review:

$f(x) = ax^2 + bx + c \ge 0$, $a > 0$



$\Leftrightarrow b^2 - 4ac \le 0$

$(2\cdot E[XY])^2 - 4(E[X^2])(E[Y^2]) \le 0$

$\Rightarrow E[X^2]\cdot E[Y^2] \ge (E[XY])^2$

2. Covariance and Correlation Coefficient

Motivating Example for “Covariance”

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - $X =$ traveling time from NCTU to Mackay
  - $Y =$ traveling time from Mackay to Train Station
  - We want to know $\text{Var}[X+Y]$
  - Question: Given $\text{Var}[X]$ and $\text{Var}[Y]$, can we get $\text{Var}[X+Y]$?

Covariance and Where to Find Them

Variance $= E[(X - E[X])^2]$

- Property:

$a, b \in \mathbb{R}$

Covariance

$$\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\boxed{E[(X - E[X])(Y - E[Y])]}$$

$$\text{Var}[aX+bY] = E\left[\left((aX+bY) - E[aX+bY]\right)^2\right]$$

$$= E\left[\left((aX - a \cdot E[X]) + (bY - b \cdot E[Y])\right)^2\right]$$

$$= a^2 \cdot E\left[(X - E[X])^2\right] + b^2 \cdot E\left[(Y - E[Y])^2\right]$$

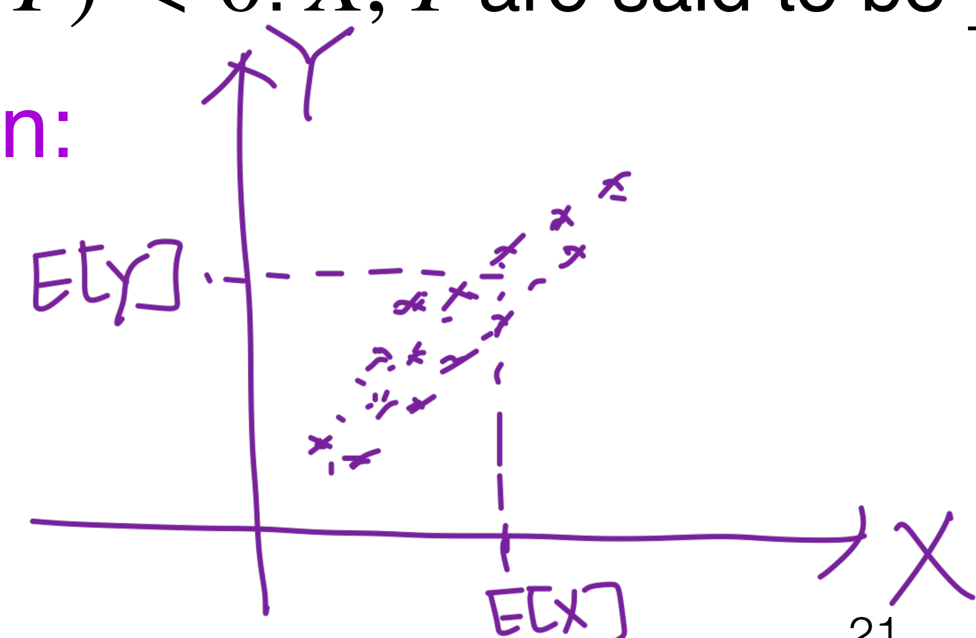
$$+ 2ab \cdot E\left[(X - E[X])(Y - E[Y])\right]$$

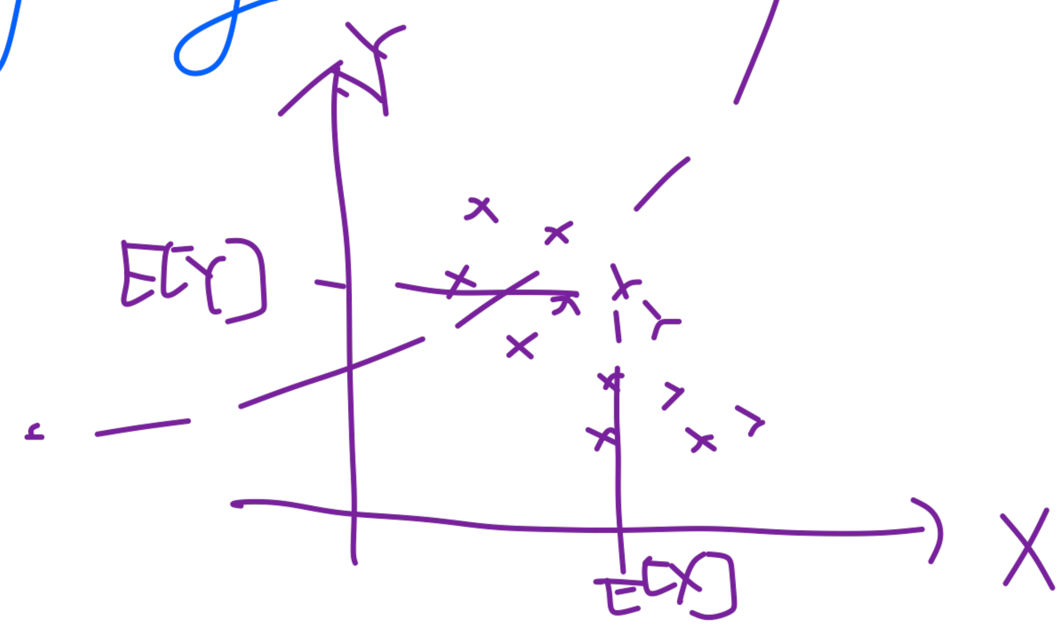
Covariance (Formally)

- **Covariance**: Let X, Y be two random variables. Then, the covariance of X and Y is defined as

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

- $\text{Cov}(X, X) =$ Var(X)
- $\text{Cov}(X, Y) = 0$: X, Y are said to be uncorrelated
- $\text{Cov}(X, Y) > 0$: X, Y are said to be positively correlated
- $\text{Cov}(X, Y) < 0$: X, Y are said to be negatively correlated
- Intuition:





Connecting Variance and Covariance

- Recall:

$$\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\underbrace{E[(X - E[X])(Y - E[Y])]}$$

- **Property**: Let X, Y be two random variables. Then,

$$\text{Var}[aX + bY] = \underline{a^2\text{Var}[X]} + \underline{b^2\text{Var}[Y]} + \underline{2ab \cdot \text{Cov}(X, Y)}$$

Another Expression of Covariance

- Let X, Y be two random variables. Then, the covariance of X and Y can also be written as

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

乘積的平均　　平均的乘積

- Question: How to show this?

$$\mathrm{Cov}(X,Y) = E[(X-E[X])\cdot(Y-E[Y])]$$

$$= E[(XY) + (-X\cdot E[Y]) + (-Y\cdot E[X]) + (E[X]\cdot E[Y])]$$

...

$$= E[XY] - E[X]\cdot E[Y]$$

$$E[XY] = E[X]\cdot E[Y]$$

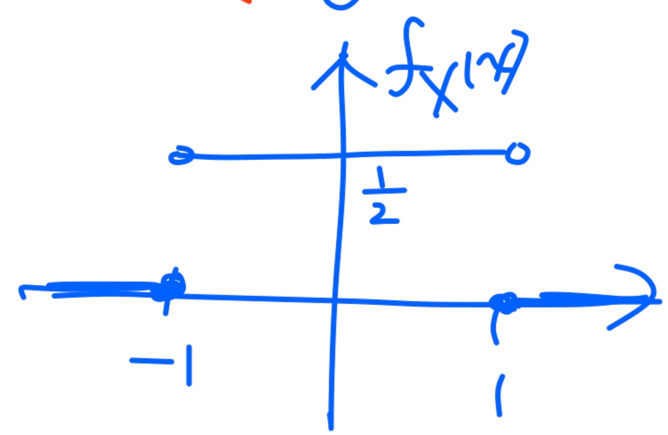
- Question: If X, Y are independent, then $\mathrm{Cov}(X, Y) =$ __0__
- Question: How about the reverse argument?

Recall: $E[XY] = E[X]E[Y] \nRightarrow X, Y$ Independent

- Example: Let X be a continuous uniform r.v. on $[-1,1]$.
  - Define $Y = X^2$
  - $E[X] = ?\ E[Y] = ?$
  - $E[XY] = ?$
  - Are X, Y independent?

$E[X$

$$f_X(x) = \begin{cases} \frac{1}{2} & , -1<x<1 \\ 0 & , \text{else} \end{cases}$$



$$E[X] = 0$$

$$E[Y] = E[X^2] = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\, dx = \frac{1}{6} x^3 \Big|_{-1}^{1} = \frac{1}{3}$$

$$E[XY] = E[X^3] = \int_{-1}^{1} x^3 \cdot \frac{1}{2}\, dx = \frac{1}{8} x^4 \Big|_{-1}^{1} = 0$$

$$E[XY] = 0 \quad , \quad E[X]E[Y] = 0 \Rightarrow E[XY] - E[X]E[Y] = 0$$

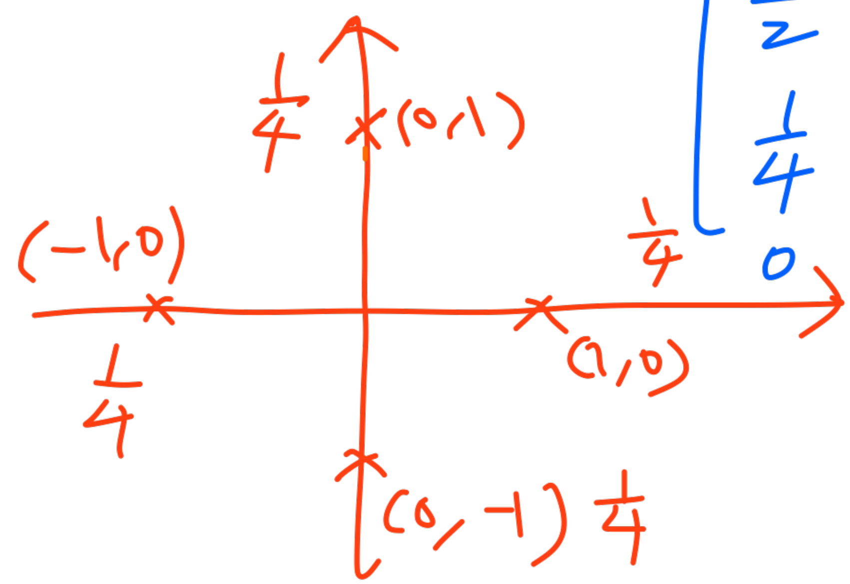
X, Y are uncorrelated

Example: Uncorrelated $\nRightarrow$ Independence

$E[XY] = E[X] \cdot E[Y]$

- Example: The pair of random variables (X, Y) takes the values $(1,0), (0,1), (-1,0), (0,-1)$, each with probability $\frac{1}{4}$
  - Cov(X, Y) = ?
  - Are X, Y independent?

$$P_Y(y) = \begin{cases} \frac{1}{4}, & 1 \\ \frac{1}{2}, & 0 \\ \frac{1}{4}, & -1 \\ 0, & \text{else} \end{cases}$$



$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

$$E[XY] = (1\cdot 0)(\tfrac{1}{4}) + (0\cdot 1)(\tfrac{1}{4}) + (-1\cdot 0)(\tfrac{1}{4}) + (0\cdot -1)(\tfrac{1}{4}) = 0$$

$$P_X(x) = \begin{cases} \frac{1}{4}, & 1 \\ \frac{1}{2}, & 0 \\ \frac{1}{4}, & -1 \\ 0, & \text{else} \end{cases}$$

$\underline{E[X] = 0}$

$\underline{E[Y] = 0}$

<u>Review</u>: $\forall A, B$

$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$

e.g. $P(X=0, Y=0) = 0$, $P(X=0) P(Y=0) = \frac{1}{4}$

Example: Uncorrelated $\not\Rightarrow$ Independence

- Example: Let θ be a continuous uniform random variable on $[0,2\pi]$. Define $X = \cos\theta$ and $Y = \sin\theta$
  - Cov(X, Y) = ?
  - Are X, Y independent?

A Property of Covariance

- **Property**:

$$\left(\mathrm{Cov}(X, Y)\right)^2 \le \mathrm{Var}[X] \cdot \mathrm{Var}[Y]$$

- **Question:** How to show this?

$$\mathrm{Cov}(X,Y) = E[\underbrace{(X - E[X])}_{\tilde{X}}\underbrace{(Y - E[Y])}_{\tilde{Y}}]$$

$$\mathrm{Var}[X] = E[(\underbrace{X - E[X]}_{\tilde{X}})^2]$$

$$\mathrm{Var}[Y] = E[(\underbrace{Y - E[Y]}_{\tilde{Y}})^2]$$

By Cauchy-Schwarz,

$$E[\tilde{X}\tilde{Y}] \le E[\tilde{X}^2] \cdot E[\tilde{Y}^2]$$

Any Issue With Covariance?

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - From NCTU to Mackay: X minutes
  - From Mackay to Train Station: Y minutes
  - Question: $\text{Cov}(X, Y)$ = ?
  - Question: What if time is measured in "seconds"? Any change in the covariance?

$$\text{Cov}(X, Y)$$

$$\text{Cov}(60X, 60Y) = E[60X \cdot 60Y] - E[60X]E[60Y]$$

Covariance is Sensitive to the Units

- Property: $\underline{\text{Cov}(aX, aY) = a^2 \cdot \text{Cov}(X, Y)}$
  - a: scaling factor due to change of unit

- Question: Any suggested solution?

Correlation Coefficient

- **Correlation Coefficient**: Let X, Y be two random variables with finite variance $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$. Then, the correlation coefficient of X and Y is defined as

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Question: Do we have $\rho(X, Y) = \rho(aX, aY)$, for any $a \neq 0$?

A Property of Correlation Coefficient

- **Property**:

$$-1 \leq \rho(X, Y) \leq 1$$

- Question: How to prove this?

Example: Correlation Coefficient

- Example: Let X be a continuous uniform r.v. on $[0,1]$.
  - Define $Y = X^2$
  - $\rho(X, Y) = $?

A Brief Summary of Covariance

$\text{Cov}(X, Y)$

- definition
- another expression
- uncorrelated?
- correlation coefficient?
- independence?
- Connection with $\text{Var}(X + Y)$, $\text{Var}(X)$, $\text{Var}(Y)$?

3. Conditional Distributions

Example: Using Joint PMF to Find Conditional PMF

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - $X =$ traveling time from NCTU to Mackay
  - $Y =$ traveling time from Mackay to Train Station
  - $P(X = 10 \mid Y = 15)$ = ?

Joint PMF	X=10	X=15	X=20
Y=10	0.1	0.1	0.05
Y=15	0.1	0.3	0.1
Y=20	0.05	0.1	0.1

Conditional PMF (Formally)

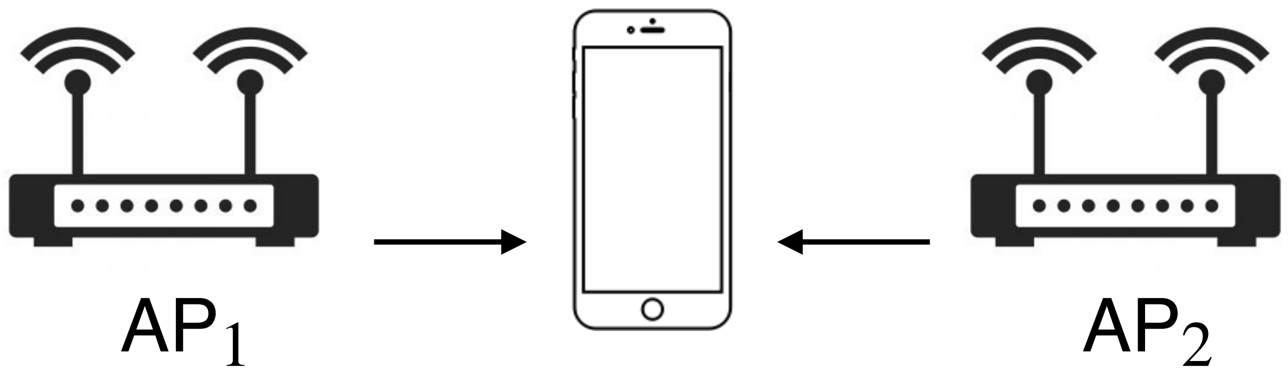
- **Conditional PMF**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$. When $P(Y = y) > 0$, the conditional PMF of X given $Y = y$ is

$$p_{X|Y}(x \,|\, y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

- Question: Conditional PMF of Y given $X = x$?

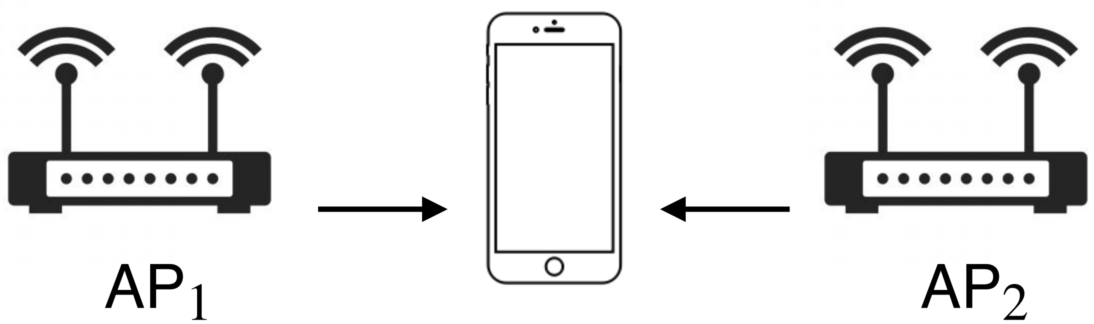
- Question: $\sum_x p_{X|Y}(x \,|\, y) =$

Example: Conditioning and Sum of Poisson



- Let N_1 and N_2 be the # of bits transmitted by AP$_1$ and AP$_2$ in a time interval T, respectively
  - N_1 and N_2 are Poisson with rates λ_1 and λ_2, respectively.
  - Moreover, N_1 and N_2 are independent
  - Define $M = N_1 + N_2$
  - Question: Conditional PMF $p_{N_1|M}(n \mid m)$ = ?

Example: Conditioning and Sum of Poisson



- Conditional PMF $p_{N_1|M}(n \mid m)$

Conditional CDF: Discrete Case (Formally)

- Question: Given $p_{X|Y}(x\,|\,y)$, how to find $P(X \le t\,|\,Y = y)$?

- **Conditional CDF**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$. When $P(Y = y) > 0$, the conditional CDF of X given $Y = y$ is defined as

$$F_{X|Y}(x\,|\,y) = P(X \le x\,|\,Y = y) =$$

Conditional PDF (Formally)

- **Conditional PDF**: Let X, Y be two continuous random variables with joint PDF $f_{XY}(x, y)$. When $f_Y(y) > 0$, the conditional PDF of X given $Y = y$ is

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

- Question: Conditional PDF of Y given $X = x$?

Example: Find Conditional PDF From Joint PDF

- Example:

$$f(x,y) = \begin{cases} 2 & , \text{ if } 0 < y < x < 1 \\ 0 & , \text{ otherwise} \end{cases}$$

- $f_{X|Y}(x \mid y) = ?$

Conditional CDF: Continuous Case (Formally)

- **Conditional CDF**: Let X, Y be two continuous random variables and $f_{X|Y}(x \mid y)$ be the conditional PDF of X given $Y = y$. The conditional CDF of X given $Y = y$ is

  $F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) =$

1-Minute Summary

1. Expected Value Regarding 2 Random Variables

- Independence: $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- Cauchy-Schwarz Inequality

2. Covariance and Correlation Coefficient

- $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
- Uncorrelated vs independence
- $\rho(X, Y) = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$, and $-1 \leq \rho(X, Y) \leq 1$

3. Conditional Distributions

- Conditional PMF / CDF / PDF