

# IEOR 173 Lecture Notes

Ryan Lin

# Contents

## Chapter 1

1.1	Lecture 03	2
	Joint Distributions — 2 • Cov + Correlations — 2 • Hat Example — 3 • Start condition (Ch 3) — 3	

## Chapter 2

2.1	Lecture 04	5
	Properties of condition expectation — 5	

## Chapter 3

3.1	Lecture 05	7
	Packet Transmission (cont) — 7 • Computing Variance: — 7	

# Chapter 1

## 1.1 Lecture 03

### 1.1.1 Joint Distributions

$$P(x, y) = P(X \leq x, Y \leq y)$$

This is a cumulative distribution for joint distributions.

If  $X \perp Y$  (independent), (perpendicular), then the joint distribution is the product of the marginals.  
 $P(X, Y) = P(X)P(Y) \forall X, Y$

### 1.1.2 Cov + Correlations

Covariance is a measure of linear dependence.

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

Interpretation: If they vary together, Cov is larger.

Properties

1.  $\text{Cov}(X, X) = \text{Var}(X)$
2.  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3.  $\text{Cov}(cX, Y) = c\text{Cov}(X, Y)$
4.  $\text{Cov}(X, Y+Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

Correlation:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{sd(X) \cdot sd(Y)}$$

If

- $> 0$ , pos correlation
- $< 0$ , negative correlation
- $= 0$  No correlation

**Note:-**

Independent does imply uncorrelated, however uncorrelated does not imply independent.

Ex: If  $P(-1, 1) = P(0, 0) = P(1, 1) = \frac{1}{3}$ . then  $E[XY] = \frac{0}{3} = 0$ . However,  $E[X] = 0, EY = ?$ , therefore  $E[XY] = 0$  and  $\text{Cov}(X, Y) = 0$ . There is no linear relationship, however there is a dependency for  $y = x^2$ .

Therefore, we can generalize our variance formula

$$\text{Var}(A + \sum b_i X_i) = \sum b_i^2 \text{Var}(X_i) + \sum_{\forall i, j, i \neq j} b_i b_j \text{Cov}(X_i, X_j)$$

### 1.1.3 Hat Example

#### Example 1.1.1 (Lists of number)

Imagine go to party, men take off hats, women keep on hats. Each man randomly takes a hat and leaves. Let  $X = \#$  of matches in  $N$  pairs (man and his hat)

What is  $EX$  and  $Var(X)$ ?

If two men, the only possibility of matches are 0 and 2. The mean is 1. If three men, a similar pattern occurs.

$$EX = 1 \forall N$$

As  $N$  grows larger, the probability decreases, but there is more men to balance it out.

$$Var(X) = 1 \forall N \geq 2$$

We can prove this by using Indicators.

Let  $X = \sum^N I_i$ , where  $I_i = 1$  if it is a match  $\sim Bern(\frac{1}{N})$

$$EX = E[\sum I] = \sum E[I] = N \frac{1}{N} = 1$$

$$Var(I_i) = E[I^2] - [EI]^2 = p - p^2 = p(1 - p) = \frac{1}{N} - \frac{1}{N^2} = \frac{N-1}{N^2}$$

$$Cov(I_i, I_j) = E[I_i I_j] - EI_i EI_j$$

$$E[I_i I_j] = P(\text{both } i \text{ and } j \text{ match}) = P(I_i = 1, I_j = 1)$$

$$P(I_i = 1)P(I_j = 1 | I_i = 1) = \frac{1}{N} \cdot \frac{1}{N-1}$$

Conceptually, we can think of event  $j$  following event  $i$

$$Cov(I_i, I_j) = \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{N^2(N-1)}$$

$$Var(X) = Var(\sum I_i) = \sum Var(I_i) + \sum Cov(I_i, I_j) = N \cdot \frac{N-1}{N^2} + N(N-1) \frac{1}{N^2(N-1)} = 1$$

The calculation will be left as an exercise to the future reader.

As  $N \rightarrow \infty$ ,

$$Cov(I_i, I_j) \rightarrow 0$$

$$Var(I_i) \rightarrow \frac{1}{N}$$

$$X \rightarrow Bin(N, \frac{1}{N}) \sim Poisson(1)$$

Intuitively, if there are thousands of men there, whether or not 1 person gets his hat is not really going to affect the chances another gets his.

### 1.1.4 Start condition (Ch 3)

What if we have dependent random variables and want to compute conditional probabilities?

Conditions:

1. discrete:  $P(X, Y) = P(X = x, Y = y), P(x, y) = P(X \leq x, Y \leq y) = \sum_{\leq x} \sum_{\leq y} P(x, y)$
2. continuous:  $f(x, y) = F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$

3. marginal:  $F_X(x) = P(X \leq x) = F(X, \infty)$

4. conditional:  $P_{x|y}(X | y) = P(X = x | Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P_{x,y}(X,y)}{P_y(y)} = \frac{\text{joint}}{\text{marginal}}$

# Chapter 2

## 2.1 Lecture 04

### 2.1.1 Properties of condition expectation

$$E[X|X = x] = x \forall x$$

$$E[X|Y = y]$$

, is a number bu  $E[X|Y]$  is a random variable.

$$X \perp W, X \sim \exp(\lambda), W \sim \exp(\lambda)$$

$$f_x(x) = \lambda e^{-x} \text{ for } x \geq 0$$

$Y = X + W$ , condition on  $Y = y$ ,  $E[X|Y = y]$ . We need  $f_{x|y}(x|y)$

$$X|y + W = [y - X|y]$$

This should have the same distributions.

To show this, we need to get the distribution function.

$$f_{x|y}(x|y) \frac{f_{x,y}(x, y)}{f_y(y)} = \frac{f_{x,w}(x, y - x)}{f_y(y)} = \frac{f_x(x)f_w(y - x)}{f_y(y)} = \lambda e^{-\lambda x} \lambda e^{-\lambda(y-x)} = \lambda^2 e^{-\lambda y}$$

For  $0 \leq x \leq y$

We also need to get the marginal distributions.

$$f_y(y) = \lambda^2 \int_{-\infty}^{\infty} f_{x,y}(x, y) dx = \lambda^2 \int_0^y e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y}$$

$$f_{x|y}(x|y) = \frac{f_{x,y}(x, y)}{f_y(y)} \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \frac{1}{y}$$

This is a uniform distribution from  $0 \leq x \leq y$ . Therefore  $E[X|Y = y] = \int_0^y (x \frac{1}{y}) dx = \frac{y}{2}$

Instead, if we take  $G(y) = E[X|Y]$ , which is a random variable, we can get  $E[E[X|Y]]$ .

We can do this by  $E[g(Y)] = \int_{-\infty}^{\infty} g(x) D(x) dx$ , which is the probability  $g(X)$  multiplied by the density  $d(x)$ .

$$E[E[X|Y]] = \int E[X|Y = y]f_y(y)dy$$

It turns out that this becomes  $E[X]$ .

### Example 2.1.1

$Y = \sum^N X_i, EX_i = \mu, N \perp X_i$ 's, or  $N$  is independent of the  $X_i$ s.

$$\begin{aligned} E[Y] &= E\left[\sum^N y_n\right] = E\left[E\left[\sum^N X_i|N\right]\right] = \sum_{\text{all } n} E\left[\sum^N X_i|N = n\right]p_n(n) = \sum_{\text{all } n} E\left[\sum^N m^n X_1\right]p_n(n) = \sum_{\text{all } n} n\mu p_n(n) \\ &= \mu \sum n p_n(n) = \mu EN \end{aligned}$$

Where  $N$  is discrete, integer valued, nonnegative.  $p_n(n)$  is the pmf for  $n$

### Example 2.1.2 (Mean of a geometric)

$$X \sim \text{geom}(p)$$

Find  $EX$ .

Because geometric dists are memoryless, we can condition on the first step.

Let  $I = I\{\text{1st trial is a success}\} \sim \text{Bern}(p)$

Compute  $EX$

$$\begin{aligned} EX &= E[E[X|I]] = \sum_{\text{all } i} E[X|I = i]p(I = i) = E[X|I = 1]p(I = 1) + E[X|I = 0]p(I = 0) \\ &= 1p + (1 + EX)(1 - p) \implies EX = \frac{1}{p} \end{aligned}$$

Basically saying, we either get the heads on this trial, or then we have to restart.

### Example 2.1.3 (Packet Transmission)

Serve 1 packet per time slot. Each slot we have  $A_i = \#$  arrivals in slot  $i$ . We assume i.i.d. and has  $p_0, p_1, p_2$ .

We assume that the buffer is size 2. (We can only hold 2 packets in the buffer)

Basically, for every time slot, we can only store up to 2 packets. Each timeslot, one packet leaves the buffer, and if there is too many packets, the incoming packet has to bounce.

$N$  = length of busy period

A busy is period is where there are arrivals to an empty system and ends when the system is emptied again.

Need  $E[N]$

First thing that we need to think about is conditioning. It would be helpful to know how many packets start the busy period, therefore we condition on  $X = \#$ packets that start the busy period (b.p), where  $X \in \{1, 2\}$ . Thus we also need to know  $P(X = 1) = P(A = 1|A = 1 \text{ or } = 2) = \frac{p_1}{p_1 + p_2}$  and  $P(X = 2) = P(A = 2|\dots) = \frac{p_2}{p_1 + p_2}$ .

$$E[N] = E[E[N|X]] = E[N|X = 1] \frac{p_1}{p_1 + p_2} + E[N|X = 2] \frac{p_2}{p_1 + p_2}$$

# Chapter 3

## 3.1 Lecture 05

### 3.1.1 Packet Transmission (cont)

#### Example 3.1.1 (Packet Transmission (cont))

Packet transmission appends  $A_i$  i.i.d.  $p_0, p_1, p_2$ .  $N$  = length of busy period. We still want  $E[N]$

$$X = \begin{cases} 1 \frac{p_1}{p_1 + p_2} \\ 2 \frac{p_2}{p_1 + p_2} \end{cases}$$

$$E[N] = E[E[N|X]] = E[N|X = 1] \frac{p_1}{p_1 + p_2} + E[N|X = 2] \frac{p_2}{p_1 + p_2}$$

We now need to compute  $E[N|X = n]$ .

For simplicity  $E[N_1] = E[N|X = 1]$ :

It would help to know how many packets we get at time 1

$$E[N_1] = E[E[N_1|A]]$$

where  $A$  is the number of arrivals in the slot of the busy period.

$$\begin{aligned} E[N] &= E[N_1|A = 0]p_0 + E[N_1|A = 1]p_1 + E[N_1|A = 2]p_2 \\ &= 1p_0 + (1 + E[N_1])p_1 + (1 + E[N_2])p_2 \end{aligned}$$

We can use the  $1 + E[N_1]$  because of the i.i.d. and the memory less state.

$$E[N_2] = E[N_2|A = 0]p_0 + E[N_2|A = 1]p_1 + E[N_2|A = 2]p_2$$

$$E[N_2] = (1 + E[N_1])p_0 + (1 + E[N_2])p_1 + (1 + E[N_2])p_2$$

The last term is because the buffer only has storage for 2 packets. If one goes out, then one gets replaced and the other gets sent away.

We can thus solve for  $E[N_1]$  and  $E[N_2]$ , which we can plug into the original equation.

This is a very simple Markov Chain.

### 3.1.2 Computing Variance:

$\text{Var}(x)$  in two methods



1. Use  $\text{Var}(x) = E[X^2] - E[X]^2$ , condition for  $E[X^2] + E[X]$
2. Condition Var formula

Example for method 1:

$$X \sim \text{geo}(p)$$

$$E[X^2] = E[E[X^2]|I] \text{ where } I = \text{indicator of success} \sim \text{Bern}(p)$$

$$\begin{aligned} E[X^2] &= E[X^2|I=1]p + E[X^2|I=0](1-p) \\ &= 1p + (E[(X+1)^2])(1-p) = p + E[X^2 + 2X + 1](1-p) \\ &= p + (E[X^2] + E[2X] + 1)(1-p) \end{aligned}$$

$$\text{Solve for } E[X^2] = \frac{n-p}{p^2}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

How do we get the method 2 formula?

*Proof.* Conditional Variance Formula

$$\begin{aligned} [E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])] &= E[E[X|Y]^2] - (E[X|Y])^2 + E[E[X|Y]^2] - E[E[E[X|Y]^2]]^2 \\ &= E[E[X^2|Y]] - E[E[X|Y]]^2 + E[X^2] - E[X]^2 \\ &= \text{Var}(X) \end{aligned}$$

**Theorem 3.1.1**  $X \perp Y \implies \text{Var}(XY) = \text{Var}(X)$

$$X \perp Y \implies \text{Var}(XY) = E[Y^2] \text{Var}(x) + E[X]^2 \text{Var}(Y) = E[X^2] \text{Var}(Y) + E[Y]^2 \text{Var}(x)$$

$$\text{Var}(XY) = E[\text{Var}(XY|y)] + \text{Var}(E[XY|y])$$

For the first term:

$$\text{Var}(XY|Y=y) = \text{Var}(X_y|Y=y) = y^2 \text{Var}(X) \implies \text{Var}(XY|y) = y^2 \text{Var}(X)$$

We can do this because  $X$  is independent of  $Y$ .

For the second term:

$$E[XY|Y=y] = yE[X] \implies E[XY|y] = YE[X]$$

Therefore,

$$\text{Var}(XY) = E[Y^2 \text{Var}(X)] + \text{Var}(YE[X]) = \text{Var}(X)E[Y^2] + E[X]^2 \text{Var}(Y)$$

**Theorem 3.1.2**

$X_i$  is i.i.d.  $N$  is an integer valued random variable independent of the  $X_i$  s. Find the  $\text{Var}(\sum^N X_i)$

$$\text{Var}(\sum^N X_i) = E[\text{Var}(\sum^N X_i|N)] + \text{Var}(E[\sum^N X_i|N])$$

For the first term:

$$\text{Var}(\sum^N X_i|N = n) = \text{Var}(\sum^n X_n|N = n) = \sum^n \text{Var}(X|N = n) = \sum^n \text{Var}(X) = n \text{Var}(X_i)$$

We proved the second term earlier.

$$E[\text{Var}(\sum^N X_i)] = E[N \text{Var}(X_i)] + \text{Var}(NE[X_i]) = E[N] \text{Var}(X_i) + \text{Var}(N)E[X_i]^2$$

If  $N \perp X$

$$\text{Var}(NX) = E[X^2] \text{Var}(N) + E[N^2] \text{Var}(X)$$

This one is like the catering example, where everyone orders the same thing. The other is if all the  $X_i$  are i.i.d., therefore ordering different things. Therefore the first is more variable than the second.