國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 18: further topics on random variable

李祈均

Final Projects

- Team-based project
  - 360 degree evaluation (final presentation)
- Find your topic

- Presentation
  - How do you collect data (show us you have actually collected the data)
  - How ‘good’ is your inference
  - Probabilistic reasoning processs
  - Q&A from every body

- Idea: probabilistic reasoning -> inference -> real data collection

Final Projects : Exemplary Topics

- 測量每一次等電梯要等幾秒才搭得到 (detla 貨梯 or 客梯)
  - 測量：每次從按電梯到真的搭到電梯的時間
  - infer: 哪一台電梯
  - Decide which distribution
    - Estimate: parameter
    - Make probabilistic reasoning (inference)

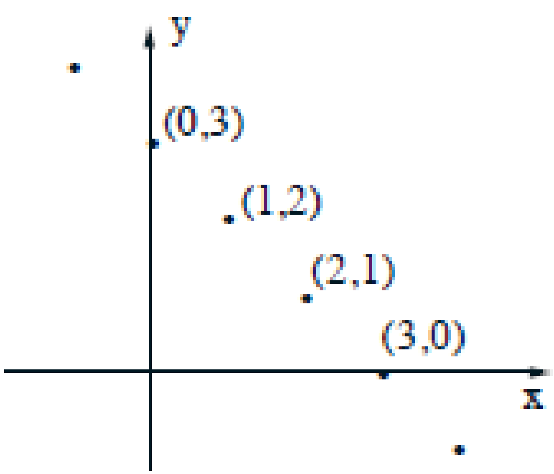
- 測量小吃部小七有幾秒收銀員結帳一個人
  - 測量： 時間
  - infer : 哪間小七

- 二十分鐘內進小七的人數
  - 測量：人的數量
  - infer: 哪一家小七

- 珍珠奶茶裏面的珍珠量
  - 測量：珍珠的數量
  - infer: 哪一家飲料店

The distribution of X+Y

- Before jumping into continuous case, let's take a look at discrete case:
  - $W = X + Y$; X, Y, are independent



Say we want to compute $W = 3$

We look at each possible values of (X, Y) that would make the sum into 3

And add up the probabilities for each of these associations

$$p_W(w) = P(X + Y = w)$$

$$= \sum_x P(X = x)P(Y = w - x)$$

$$= \sum_x p_X(x)p_Y(w - x)$$

What is the mechanics of these?

- Put the two pmf's on top of each other
- Flip the pmf of Y
- Shift the flipped pmf by w (to right if $w > 0$)
- Cross-multiply and add

Continuous case?

- Extremely similar to discrete case, as always..

Say $\mathrm{Z} = X + Y$

First note the following:

$$P(Z \leq z|X = x) = P(X + Y \leq z|X = x)$$

$$= P(x + Y \leq z|X = x) = P(x + Y \leq z)$$

$$= P(Y \leq z - x)$$

If we take differentiation on both sides:

$$f_{Z|X}(z|x) = f_Y(z - x)$$

- Knowing that, then invoke multiplication rule for joint PDF:

$$f_{X,Z}(x,z) = f_X(x) f_{Z|X}(z|x) = f_X(x) f_Y(z-x)$$

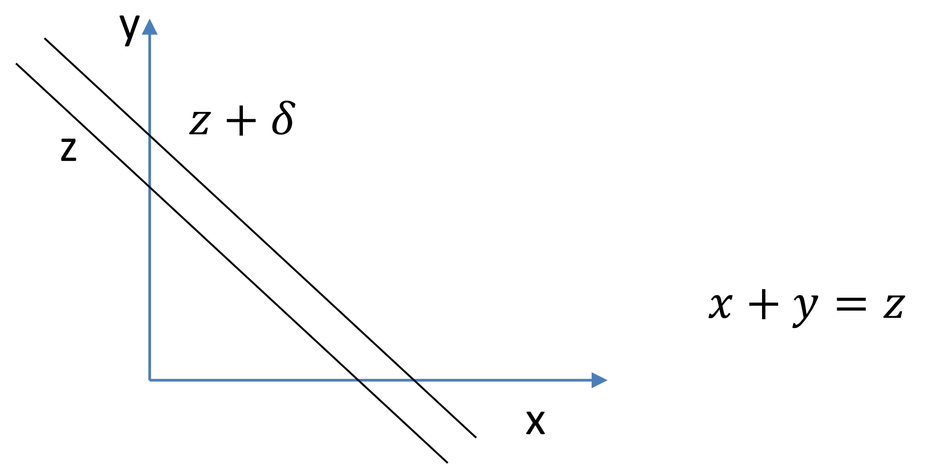
Now to find the $f_Z(z)$ -> marginal PDF, integrate out x

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

$$p_W(w) = \sum_x p_X(x) p_Y(w-x)$$

Both cases are similar, and intuitions are exactly the same:

This operation is called **convolution**



- Essentially, we are trying to compute the probability using the following:

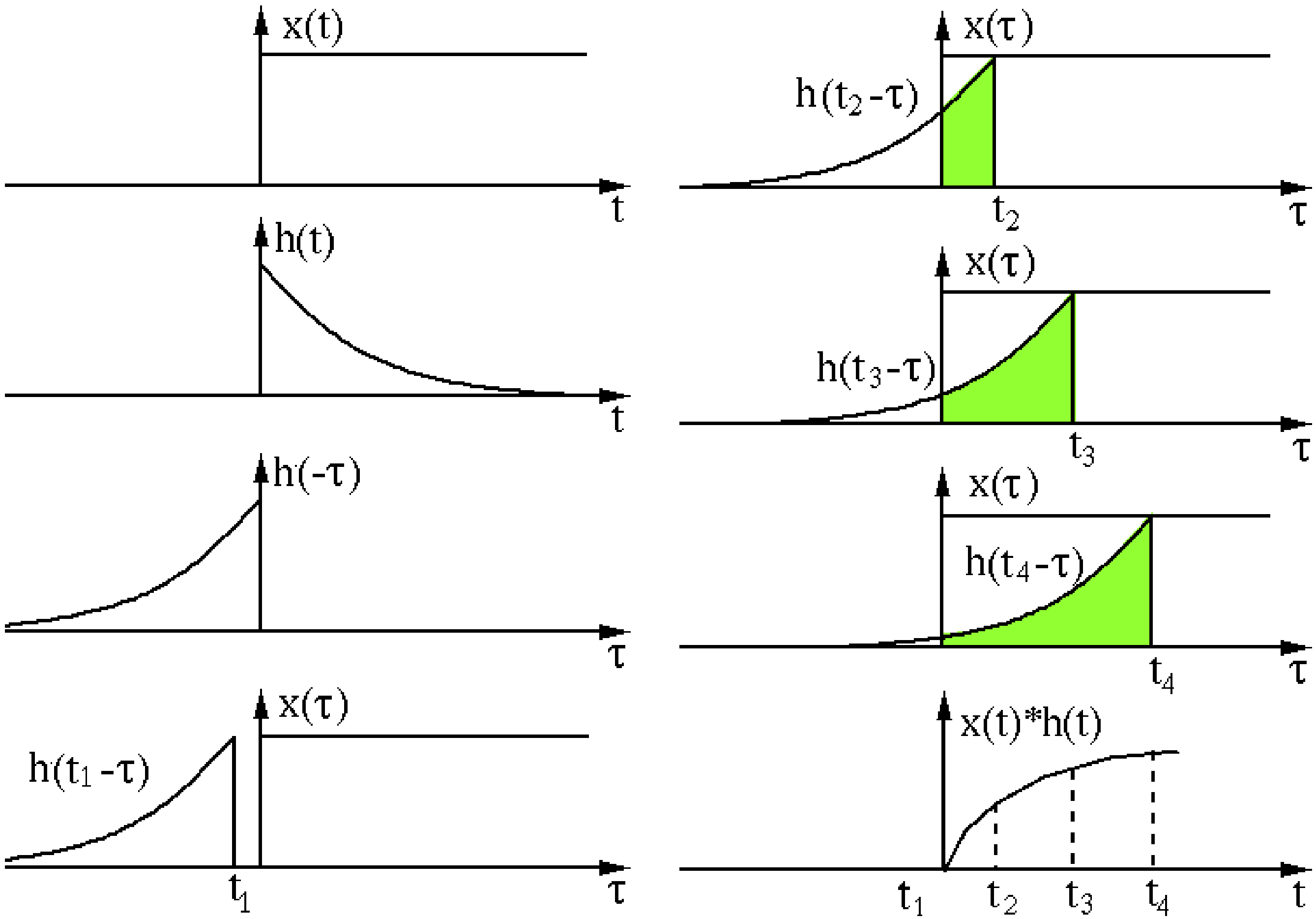
$$P(z \le X + Y \le z+\delta) \approx f_Z(z)\delta$$

So,

$$f_Z(z)\delta = P(z \le X + Y \le z+\delta)$$

$$= \int_{-\infty}^{\infty}\int_{z-x}^{z-x+\delta} f_X(x) f_Y(y)\,dy\,dx \approx \delta \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

Graphical representation of convolution



New topic: quantitative measure of the strength and direction of the relationship between two r.v.s

- Definition of **covariance (expected value operator)**

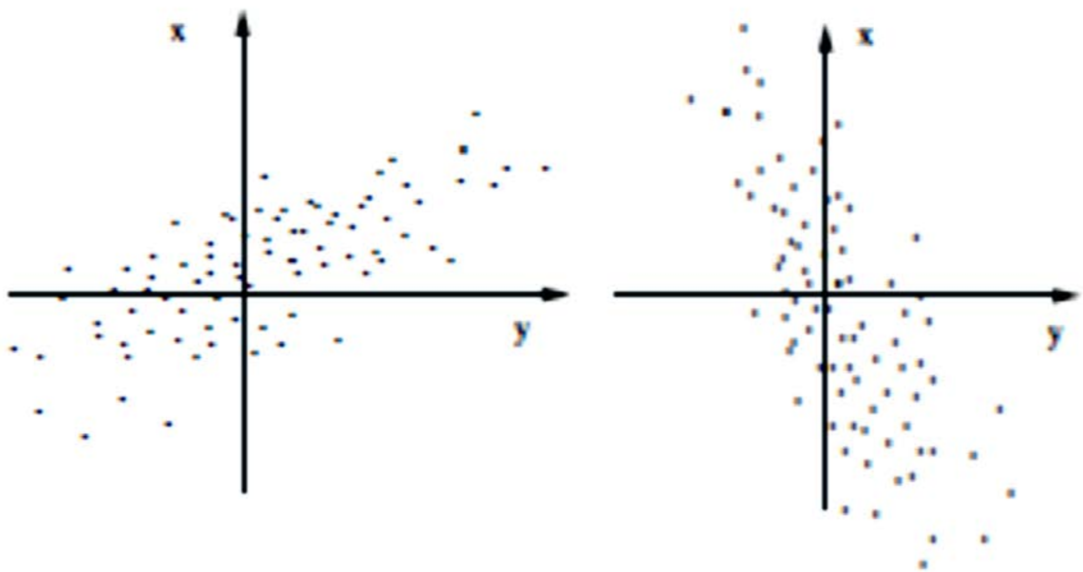
$$cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Or alternatively,

$$\begin{aligned} cov(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E\big[XY - XE[Y] - YE[X] + E[X]E[Y]\big] \\ &= \boldsymbol{E[XY] - E[X]E[Y]} \end{aligned}$$

- Uncorrelated random variables
  - Covariance is zero

- Roughly speaking: positive or negative covariance indicates that the values of $X - E[X]$ and $Y - E[Y]$ in an experiment tend to have the same or opposite sign



Properties of covariance

These can be derived easily from the definition

- $cov(X, X) = var(X)$
- $cov(X, aY + b) = a\, cov(X, Y)$
- $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$

Note, if X,Y are independent:

$E[XY] = E[X]E[Y]$, then $cov(X, Y) = E[XY] - E[X]E[Y] \; = \; 0$

Is converse true? (think about it for a moment)

Textbook example

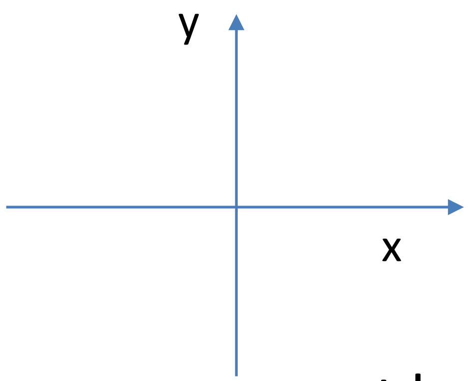
The pair of random variables X,Y takes the value (1,0), (0,1), (-1,0), (0,-1), each with probability ¼

What is $E[XY]$?

- Since at least one of X or Y take on the value of 0
- $E[XY] = 0$

What is $E[X], E[Y]$?

- Since X, Y are symmetric around 0
- $E[X], E[Y] = 0$



So are they uncorrelated? Yes

Are the independent ? No (knowing X is non-zero.., y can not be)

- Let's generalize this concept:

As long as you have the following condition, they are uncorrelated:

$$E[X|Y = y] = E[X], \text{for all y}$$

Using total expectation theorem:

$$E[XY] = \sum_y y p_Y(y) E[X|Y = y] = E[X] \sum_y y p_Y(y)$$
$$= E[X]E[Y]$$

Variance of the sum of random variables

$$var(X_1 + X_2) = var(X_1) + var(X_2) + 2cov(X_1, X_2)$$

Generally,

$$var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i) + \sum_{\{(i,j)|i \neq j\}} cov(X_i, X_j)$$

- Let's derive it
  - First set $\tilde{X}_i = X_i - E[X_i]$

$$var\left(\sum\nolimits_{i=1}^{n} X_i\right) = E\left[\left(\sum_{i=1}^{n} \tilde{X}_i\right)^2\right]$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{X}_i\tilde{X}_j\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\tilde{X}_i\tilde{X}_j\right]$$

$$= \sum_{i=1}^{n}E\left[\tilde{X}_i^{\ 2}\right] + \sum_{\{(i,j)|i\neq j\}}E\left[\tilde{X}_i\tilde{X}_j\right]$$

$$= \sum_{i=1}^{n}var(X_i) + \sum_{\{(i,j)|i\neq j\}}cov(X_i, X_j)$$

Recall that we have essentially used this formula to derive the variance of the hat problem back in chapter 2!

Correlation coefficients

- A dimensionless version of covariance
  - Can also be imagined as normalized version of covariance
  - General notion is the same as covariance
  - Definition as follows:

$$\rho = E\left[\frac{(X - E[X])}{\sigma_X} * \frac{(Y - E[Y])}{\sigma_Y}\right] = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

  - $-1 \leq \rho \leq 1$
  - $|\rho|$ provides a normalized measure of the association
  - Independence means that $\rho = 0$, but reverse is not true

- Property of correlation coefficients

$$|\rho| = 1 \leftrightarrow (X - E[X]) = c(Y - E[Y])$$

  - Linearly related makes coefficient to be 1 or -1!

- Example

Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the numbers of heads and tails, respectively, and let us look at the correlation coefficients of X and Y.

We have

$$X + Y = n$$

$$(E[X] + E[Y] = n)$$

intuitively, the correlation coefficient is -1

- From the above two equation (first – second)

$$X - E[X] = -(Y - E[Y])$$

First calculate the covariance:

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

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

Now compute correlation coefficient

$$\rho = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{-var(X)}{\sqrt{var(X)var(X)}} = -1$$

Can we prove that $-1 \leq \rho \leq 1$?

- First we have to know a very important inequality
  - Schwarz inequality, for any two random variables X,Y

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

Now:

$$0 \leq E\left[\left(X - \frac{E[XY]}{E[Y^2]}Y\right)^2\right]$$

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

Hence, proved.

- Now, if we let:

$$\tilde{X} = X - E[X]$$

$$\tilde{Y} = Y - E[Y]$$

Using the Schwarz inequality:

$$(\rho(X,Y))^2 = \frac{\left(E\left[\tilde{X}\tilde{Y}\right]\right)^2}{E\left[\tilde{X}^2\right]E\left[\tilde{Y}^2\right]} \leq 1$$

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

$$\frac{(E[XY])^2}{E[X^2]E[Y^2]} \leq 1$$

Another problem on correlation coefficients

Show that $\rho(aX + b, Y) = \rho(X, Y)$

$$\rho(aX + b, Y) = \frac{cov(aX + b, Y)}{\sqrt{var(aX + b)var(Y)}}$$

$$= \frac{E[(aX + b - E[aX + b])(Y - E[Y])]}{\sqrt{a^2 var(X)var(Y)}}$$

$$= \frac{E[(aX + b - aE[X] - b)(Y - E[Y])]}{a\sqrt{var(X)var(Y)}}$$

$$= \frac{aE[(X - E[X])(Y - E[Y])]}{a\sqrt{var(X)var(Y)}}$$

$$= \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \rho(X,Y)$$

As an example where this is relevant, consider the homework and exam scores. We expect the homework and exam scores to be positively correlated. In this example, the above property will mean that the correlation coefficient will not change whether the exam is out of 105 points, 10 points, or any other number of points.

Example

Let X be a discrete random variable with PMF, p_X and let Y be a continuous random variable, independent from X, with PDF, f_Y. Derive a formula for the PDF of the random variable $X + Y$

- Let $Z = X + Y$
  - Now use the 2-step CDF approach

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z)$$

Let's now use total probability theorem, we have

$$F_Z(z) = \sum_x p_X(x) p(x + Y \leq z)$$

$$= \sum_x p_X(x) p(Y \leq z - x)$$

$$= \sum_x p_X(x) F_Y(z - x)$$

Now it's time to differentiate both sides with respect to z to obtain the PDF of Z

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \sum_x p_X(x) f_Y(z - x)$$

Another example problem

An ambulance travels back and forth, at a constant specific speed v, along a road of length l.

We model the location of the ambulance at any moment in time to be uniformly distributed over the interval $(0, l)$

Also at any moment in time, an accident (not involving the ambulance itself) occurs at a point uniformly distributed on the road; that is the accident's distance from one of the fixed ends of the road is also uniformly distributed over $(0, l)$

Assume the location of the accident and the location of the ambulance are independent

- Further assume that the ambulance is capable of immediate *U-TURN,* compute the CDF and PDF of the ambulance’s travel time T to the location of the accident

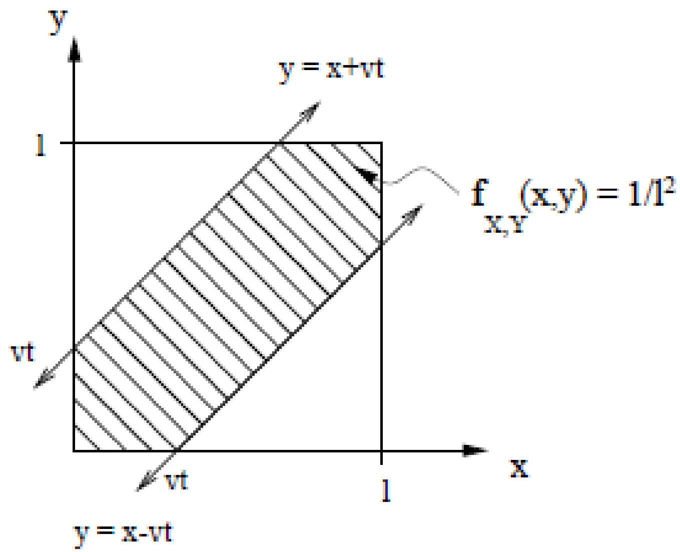
Two step approach

- $P(T \leq t) = P(|X - Y| \leq vt)$
  - X and Y are the locations of the ambulance and accident (both uniform over $[0, l]$
  - Since X,Y are independent

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{l^2} & , \text{if } 0 \leq x, y \leq l \\ 0, \text{otherwise} & \end{cases}$$

$$P(T \leq t) = P(|X - Y| \leq vt) = P(-vt \leq Y - X \leq vt)$$
$$= P(X - vt \leq Y \leq X + vt)$$

- That corresponds to the shaded area in the following figure:



- Since the joint density if uniform over that entire region: we have:

$$F_T(t) = \left(\frac{1}{l^2}\right) * (\text{shaded area}) = \begin{cases} 0, t < 0 \\ \frac{2vt}{l} - \frac{(vt)^2}{l^2}, 0 \le t \le \frac{l}{v} \\ 1, t \ge \frac{l}{v} \end{cases}$$

- After this, simply differentiate with respect to t

$$f_T(t) = \begin{cases} \dfrac{2v}{l} - \dfrac{2v^2 t}{t^2}, 0 \leq t \leq \dfrac{l}{v} \\ 0, \text{otherwise} \end{cases}$$

Lecture outline

- Conditional expectation
  - Law of iterated expectation
  - Law of total variance
- Transforms: moment generating function

A little background review

Conditional expectations

- Given the value of y of a r.v. Y

$$E[X|Y = y] = \sum_{x} x p_{X|Y}(x|y)$$

  – Continuous?

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)$$

- Let’s try it with an example

Stick example

- Stick of length l
  - Break uniformly chosen at point Y
  - Break again uniformly chosen a point X
- First, what is ?

$$E[X|Y = y] = \frac{y}{2}$$

  - That is a ***number***
  - But before you do any experiment, that y value is unknown, Y is random itself
  - So conditioned expectation can be seem as a ***random variable***!

$$E[X|Y] = \frac{Y}{2}$$

- Conditional expectation can be seen as a random variable instead of a number!
  - Once you do the experiment, and get a number for y, that conditional expectation becomes a number!
  - It's a subtle concept, and abstraction (useful abstraction)
  - It is a random variable, then we can take expectation of this random variable, let try it:

$$E\big[E[X|Y]\big] = \sum_y E[X|Y = y]p_Y(y)$$
$$= E[X]$$

What are these:
essentially, the same thing noted in chapt2,3
Total expectation theorem

- That is called law of iterated expectation

$$E\big[E[X|Y]\big] = E[X]$$

Now back to stick problem, if we want to compute E[X], how?

$$E[X] = E\big[E[X|Y]\big]$$

$$= E[Y/2] = \frac{1}{2}E[Y] = \frac{l}{4}$$

Conditional variance

- Now we know conditional expectation, now move onto conditional variance

Standard conditional variance that we know:

$$var(X|Y = y) = E[(X - E[X|Y = y])^2|Y = y]$$

- The same concept that we can imagine cases when we actually don't know the value Y takes, and make conditional variance a r.v.

$$var(X|Y)$$

- That is a random variable
  - Once you know $Y = y$, that conditional variance is specified

- Law of iterated expectation
  - Expected value of a conditional expectation is the unconditional expectation

- However, law of total variance is a little different:
- Let's list it first:

- Law of total variance

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

Let's prove it:

Recall:

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

Now, we can write:

$$var(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

Now, we can take expected value on both sides

$$E[var(X|Y)] = E[X^2] - E[(E[X|Y])^2]$$

Then, let's take variance of r.v. $E[X|Y]$

$$var(E[X|Y]) = E[(E[X|Y])^2] - (E[X])^2$$

Okay, now we take sum or 3 and 4

$$E[var(X|Y)] + var(E[X|Y]) = E[X^2] - (E[X])^2 = var(X)$$

- Law of total variance

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

Intuition around the proofs,

not quite, but useful to go through the drill

How to intuitively think about conditional variance

- Some prelude, but save for later
- You can imagine about an inference problem:
  - You want to know X, but you only measure Y,
  - So you estimate X based on Y
  - Conditional variance can be thought as estimation error…

Back to a toy problem

- Say we give a quiz to a class consisting of many sections, so the experiment goes as follows,

I pick a student at random, and I look at two variables, one is the quiz score (X) of the randomly selected student, and another one is the section number ($Y = y$, where $Y = \{1,2\}$) of the student that I have selected, we know that:

Section1: 10 students

Section2: 20 students

Quiz average in section 1: 90

Quiz average in section 2: 60

- What is expected value of X ($E[X]$)
  - Pretty straightforward, randomly selected, equally-like to pick any student

$$E[X] = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{90 * 10 + 60 * 20}{30} = 70$$

- Now, let’s look at conditional expectation
  - Simple case first:

$$E[X|Y = 1] = 90$$
$$E[X|Y = 2] = 60$$

  - Abstract case?

$$E[X|Y] = \begin{cases} 60, \text{with } p = 1/3 \\ 90, \text{with } p = 2/3 \end{cases}$$

  - Now this is a random variable, with a distribution, let’s compute the expected value $E\big[E[X|Y]\big]$

$$E\big[E[X|Y]\big] = \frac{1}{3} * 60 + \frac{2}{3} * 90 = 70 = E[X]$$

- That simply means, to get overall average, you take averages of each section, and weighted this average by the associated probability of picking a student from that class

Now let's try to find $var(X)$ from conditional variance?

- Someone goes and calculate the variance of quiz scores inside each of the section, gives the following

$$\frac{1}{10}\sum_{i=1}^{10}(x_i - 90)^2 = 10, \frac{1}{20}\sum_{i=11}^{30}(x_i - 60)^2 = 20$$

- Now we know,

$$var(X|Y = 1) = 10, var(X|Y = 2) = 20$$

- So what is $var(X|Y)$

$$var(X|Y) = \begin{cases} 10, \text{with } p = 1/3 \\ 20, \text{with } p = 2/3 \end{cases}$$

  – This is a random variable we can take the expectation:

$$E[var(X|Y)] = \frac{1}{3} * 10 + \frac{2}{3} * 20 = \frac{50}{3}$$

Also, we know:

$$E[X|Y] = \begin{cases} 60, \text{with } p = 1/3 \\ 90, \text{with } p = 2/3 \end{cases}$$

– $var(E[X|Y]) = \frac{1}{3}(90 - 70)^2 + \frac{2}{3}(60 - 70)^2 = 200$

Now, we have all the information available to use law of total variance

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

$$var(X) = \frac{50}{3} + 200$$

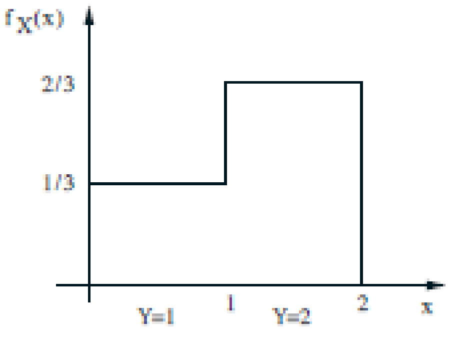
This is like (average variability **within** each section + variability **between** each section) = total variability of the quiz score

Another numerical example

Somebody came and ask you the variance of X?

Complicated, can we divide and conquer?

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$



$$E[X|Y=1] = \frac{1}{2}, E[X|Y=2] = \frac{3}{2}$$

$$var(X|Y=1) = \frac{1}{12}, var(X|Y=2) = \frac{1}{12}$$

- With that we can find the overall $E[X]$

$$E[X] = E\big[E[X|Y]\big] = \frac{1}{3} * \frac{1}{2} + \frac{2}{3} * \frac{3}{2} = \frac{7}{6}$$

- Now, let's try

$$E[X|Y] = \begin{cases} \frac{1}{2,} \text{ with } p = 1/3 \\ \frac{3}{2}, \text{with } p = 2/3 \end{cases}$$

$$var(E[X|Y]) = E\big[(E[X|Y] - E\big[E[X|Y]\big])^2\big]$$

$$= \frac{1}{3} * \left(\frac{1}{2} - \frac{7}{6}\right)^2 + \frac{2}{3} * \left(\frac{3}{2} - \frac{7}{6}\right)^2$$

Now, let's compute $E[var(X|Y)]$:

$$E[var(X|Y)] = \frac{1}{12}$$

Now we have everything to compute the total variance of X

$$var(X|Y = 1) = \frac{1}{12}, var(X|Y = 2) = \frac{1}{12}$$

Sum of random number of independent random variables

Over the weekend, you are going to visit a random number of bookstores, at each store, you are going to spend a random amount of money

Let N be number of stores that you are visiting, n is an integer (non-negative)

Each time you walk into a store, your mind is refreshed, and you just buy a random number of books that has nothing to do with what you have done for the day, each time you enter a book store as a brand new person, buys a random number of books, and spend a random amount of money

- Now let X_i be the money spent in store i
  - X_i assume i.i.d.
  - Independent of N

- Now let's set Y be the total money spent on book
  - $Y = X_1 + X_2 + X_3 \ldots + X_N$
  - We are dealing with sum of random variable except the N itself is also a random variable
  - First let's compute $E[Y]$?
  - Let's work in the conditional universe
    - Say if we are given $N = n$

$$E[Y|N = n] = E[X_1 + X_2 + X_3 \ldots + X_n|N = n]$$
$$= E[X_1 + X_2 + X_3 \ldots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$
$$= nE[X]$$

If we don’t know N before hand,

$$E[Y|N] = NE[X]$$

- This is a random variable, and if you are given N to a specific value, then you get a number!

- Now we can invoke the iterated expectation law

$$E[Y] = E\big[E[Y|N]\big] = E\big[NE[X]\big] = E[X]E[N]$$

  - $E[X]$ is a number

  - This should also be intuitively easy!

- What if I want to know the variance in this case?

$$var(Y) = E[var(Y|N)] + var(E[Y|N])$$

  - $var(E[Y|N]) = var(NE[X]) = E[X]^2 var(N)$
    - Recall $var(aX) = a^2 var(X)$
    - Variability in how much money you are spending as the randomness exists in how many stores you visit
  - $var(Y|N = n) = nvar(X)$
  - $var(Y|N) = Nvar(X)$
  - $E[var(Y|N)] = E[Nvar(X)] = var(X)E[N]$
    - Randomness exists inside each store

So the total variability exists in how much you are going to spend

$$var(Y) = E[N]var(X) + (E[X])^2 var(N)$$

New topic

- Transforms of random variable
  - r.v.'s are functions, transform are a different representation of a functions, imagine, Fourier transform
  - Intuition around transforms in probability is kinda abstract, but often quite useful for mathematical manipulation

- Definition:
  - For a random variable X, the transform (or something called moment generating function) is defined below:
$$M_X(s) = E[e^{sX}]$$
  - s is a scalar parameter

- Let's write out the actual formula:

$$M(s) = \sum_x e^{sX} p_X(x)$$

$$M(s) = \int_{-\infty}^{\infty} e^{sX} f_X(x) dx$$

- Note that these transforms are not numbers, they are still a function of parameter s
- Linear function of X (e.g., $Y = aX + b$)

$$M_Y(s) = E\left[e^{s(aX+b)}\right] = e^{sb} E[e^{saX}] = e^{sb} M_X(sa)$$

Sample transforms

- Poisson random variable with parameter λ:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,3 \dots$$

Now, let's try to transform it:

$$M(s) = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!}$$

Now, we can set $a = e^s \lambda$

$$M(s) = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^{-\lambda} e^{a} = e^{a-\lambda}$$

$$= e^{\lambda(e^s - 1)}$$