Kovaryans ve Korelasyon

Bugun “kovaryans gunu”, bu teknigi kullanarak nihayet bir toplamin varyansini bulabilecegiz, varyans lineer degildir (kiyasla beklenti -expectation- lineerdir). Bu lineer olmama durumu bizi korkutmayacak tabii, sadece yanlis bir sekilde lineerlik uygulamak yerine probleme farkli bir sekilde yaklasmayi ogrenecegiz.

Diger bir acidan, hatta bu ana kullanimlardan biri, kovaryans iki rasgele degiskeni beraber / ayni anda analiz etmemize yarayacak. Iki varyans olacak, ve onlarin alakasina bakiyor olacagiz, bu sebeple bu analize *ko*varyans deniyor zaten.

Tanim

$$ Cov(X, Y) = E((X - E(X))(Y - E(Y))) \tag{1} $$

Burada X, Y ayni uzayda tanimlanmis herhangi iki rasgele degisken. Ustteki diyor ki bu iki rasgele degisken X, Y'in kovaryansi, X'ten ortalamasi cikartilmis, Y'ten ortalamasi cikartilmis halinin carpilmasi ve tum bu carpimlarin ortalamasinin alinmasidir.

Tanim boyle. Simdi bu tanima biraz bakip onun hakkinda sezgi / anlayis gelistirmeye ugrasalim. Tanim niye bu sekilde yapilmis, baska bir sekilde degil?

Ilk once esitligin sag tarafindaki bir carpim, bir sey carpi bir baska sey. Bu seylerden biri X ile digeri Y ile alakali, onlari carparak ve carpimin bir ozelliginden faydalanarak sunu elde ettik; arti deger carpi arti yine arti degerdir, eski carpi arti eksidir, eksi carpi eksi artidir. Bu sekilde mesela “ayni anda arti” olmak gibi kuvvetli bir baglanti carpimin arti olmasi ile yakalanabilecektir. Ayni durum eksi,eksi de icin gecerli, bu sefer her iki rasgele degisken ayni sekilde negatiftir. Eksi carpim sonucu ise sifirdan az bir degerdir, “kotu korelasyon” olarak alinabilir ve hakikaten de eksi arti carpiminin isareti oldugu icin iki degiskenin ters yonlerde oldugunu gosterir. Demek ki bu arac / numara hakikaten faydali.

Unutmayalim, ustteki carpimlardan birisinin buyuklugu X'in ortalamasina bagli olan bir diger, Y ayni sekilde. Simdi X, Y'den bir orneklem (sample) aldigimizi dusunelim. Veri setinin her veri noktasi bagimsiz ve ayni sekilde dagilmis (i.i.d) durumda. Yani X, Y degiskenlerine “gelen” x_i, y_i ikilileri her i icin digerlerinden bagimsiz; fakat her ikilinin arasinda bir baglanti var, yani demek ki bu rasgele degiskenlerin baz aldigi dagilimlarin bir alakasi var, ya da bu iki degiskenin bir birlesik dagilimi (joint distribution) var.

Not: Eger X, Y bagimsiz olsaydi, o zaman

$$ Cov(X, Y) = E((X - E(X))E(Y - E(Y)) $$

olarak yazilabilirdi, yani iki beklentinin ayri ayri carpilabildigi durum... Ama biz bu derste bagimsizligin olmadigi durumla ilgileniyoruz..

Korelasyon kelimesinden bahsedelim hemen, bu kelime gunluk konusmada cok kullaniliyor, ama bu ders baglaminda korelasyon kelimesinin matematiksel bir anlami olacak, onu birazdan, kovaryans uzerinden tanimlayacagiz.

Bazi ilginc noktalar:

Ozellik 1

varyansi nasil tanimlamistik?

$$Var(X) = E((X - E(X))^2)$$

Bu denklem aslinda

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

denkleminde Y yerine X kullandigimizda elde ettigimiz seydir, yani

$$Cov(X, X) = E((X - E(X))(X - E(X)))$$

$$Cov(X, X) = E((X - E(X))^2)$$

$$= Var(X)$$

Yani varyans, bir degiskenin "kendisi ile kovaryansidir". Ilginc degil mi?

Ozellik 2

$$Cov(X, Y) = Cov(Y, X)$$

Ispati kolay herhalde, (1) formulunu uygulamak yeterli.

Teori

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Ispat

Bu ispat cok kolay, esitligin sol tarafindaki carpimi parantezler uzerinden acarsak, ve beklenti lineer bir operator oldugu icin toplamin terimleri uzerinde ayri ayri uygulanabilir,

$$E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y)$$

Carpimi uygularken mesela $E(-X \cdot E(Y))$ gibi bir durum ortaya cikti, burada $E(Y)$'nin bir sabit oldugunu unutmayalim, cunku beklenti rasgele degiskene uygulaninca tek bir sayi ortaya cikartir, ve vu $E(Y)$ uzerinde bir beklenti daha uygulaninca bu "icerideki" beklenti sabitmis gibi disari cikartilabilir, yani $-E(X)E(Y)$.

Devam edelim, $E(XY) - E(X)E(Y)$ ifadesini gosterdik, cunku cogu zaman bu ifade hesap acisindan (1)'den daha uygundur. Ama (1) ifadesi anlatim / sezgisel kavrayis acisindan daha uygun, cunku bu ifade X'in ve Y'nin kendi ortalamalarina izafi olarak belirtilmistir, ve akilda canlandirilmasi daha rahat olabilir. Fakat matematiksel olarak bu iki ifade de aynidir.

Iki ozellik bulduk bile. Bir ozellik daha,

Ozellik 3

$$Cov(X, c) = 0$$

Bu nereden geldi? (1)'e bakalim, Y yerine c koymus olduk, yani bir sabit. Bu durumda (1)'in $(Y - E(Y))$ kismi $c - E(c) = c - c = 0$ olur [aslinda bayagi absurt bir durum], ve bu durumda (1) tamamen sifira donusur, sonuc sifir.

Ozellik 4

$Cov(cX, Y) = c \cdot Cov(X, Y)$

Ispat icin alttaki formulde

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

X yerine cX koymak yeterli, c her iki terimde de disari cikacaktir, ve grubun disina alincan bu ozelligi elde ederiz.

Ozellik 5

$$Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$$

Ispat icin bir ustteki ozellikte yaptigimizin benzerini yapariz.

En son iki ozellik oldukca faydalidir bu arada, onlara ikili-lineerlik (bilinearity) ismi veriliyor. Isim biraz renkli / sukseli bir isim, soylemek istedigi su aslinda, bu son iki ozellikte sanki bir kordinati sabit tutup digeri ile islem yapmis gibi oluyoruz, yani bir kordinat sabit olunca digeri "lineermis gibi" oluyor; Mesela c'nin disari ciktigi durumda oldugu gibi, bu ozellikte Y'ye hicbir sey olmadi, o degismeden kaldi. Ayni sekilde 5. ozellikte X hic degismeden esitligin sagina aktarildi sanki, sadece "Z durumu icin" yeni bir terim ekledik.

4. ve 5. ozellik cok onemlidir, bunlari bilirseniz bir ton hesabi yapmadan hizlica tureterek hesaplarinizi kolaylastirabilirsiniz.

Ozellik 6

$$ Cov(X+Y, Z+W) = Cov(X,Z) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W) $$

Simdi 5. ozelligi hatirlayalim, orada gosterilen sanki bir nevi basit cebirdeki dagitimsal (distributive) kuralin uygulanmasi gibiydi sanki, yani $(a+b)(c+d)$'i actigimiz gibi, 5. ozellik te sanki kovaryansi carpip topluyormus gibi "aciyordu". En temelde gercekten olan bu degil ama nihai sonuc benzer gozuktugu icin akilda tutmasi kolay bir metot elde etmis oluyoruz. Her neyse, 6. ozellik icin aslinda 5. ozelligi tekrar tekrar uygulamak yeterli. Bu arada 5. ozellik $Cov(X, Y+Z)$ icin ama $Cov(Y+Z, X)$ yine ayni sonucu veriyor.

Bu arada 6. ozellik cok cetrefil toplamlar uzerinde de uygulanabilir, mesela

$$ Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_i Y_i \right) $$

Bu son derece karmasik gozukuyor, fakat cozumu icin aynen 6. ozellikte oldugu gibi 5. ozelligi yine tekrar tekrar uygulamak yeterli (4. ozellik ile de sabiti disari cikaririz, vs).

Cogu zaman ustteki gibi pur kovaryans iceren bir acilimla calismak, icinde beklentiler olan formullerle ugrasmaktan daha kolaydir.

Simdi toplamlara donelim; kovaryanslara girmemizin bir sebebi toplamlarla is yapabilmemizi saglamasi. Mesela, bir toplamin varyansini nasil hesaplariz?

Ozellik 7

$$ Var(X_1 + X_2) $$

Simdilik iki degisken, ama onu genellestirip daha fazla degiskeni kullanabiliriz.

Cozelim. 1. ozellik der ki varyans degiskenin kendisi ile kovaryansidir, yani $Var(X) = Cov(X,X)$. O zaman $Var(X_1+X_2) = Cov(X_1+X_2, X_1+X_2)$. Boylece icinde toplamlar iceren bir kovaryans elde ettik ama bunu cozmeyi biliyoruz artik. "Dagitimsal" islemleri yaparken $Cov(X_1, X_1)$ gibi ifadeler cikacak, bunlar hemen varyansa donusecek. Diger taraftan $Cov(X_1, X_2)$ iki kere gelecek, yani

$$ Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) $$

Bu alanda bilinen tekerleme gibi bir baska deyis, "eger kovaryans sifirsa toplamin varyansi varyanslarin toplamidir". Hakikaten kovaryans sifir olunca ustteki denklemden dusecektir, geriye sadece varyanslarin toplami kalacaktir. Kovaryans ne zaman sifirdir? Eger X_1, X_2 birbirinden bagimsiz ise. Tabii bu bagimsizlik her zaman ortaya cikmaz.

Ikiden fazla degisken olunca? Yine tum varyanslarin ayri ayri toplami, ve kovaryanslar da sonda toplanacak,

$$Var(X_1 + .. + X_n) = Var(X_1) + .. + Var(X_n) + 2 \sum_{i<j} Cov(X_i, X_j)$$

Sondaki toplamin indisinde bir numara yaptik, sadece 1 ile 2, 2 ile 3, vs. eslemek icin, ve mesela 3 ile 1'i tekrar eslememek icin. Tekrar dedik cunku $Cov(X_1, X_3) = Cov(X_3, X_1)$. Eger indisleme numarasi kullanmasaydik, 2 ile carpimi cikartirdik (ona artik gerek olmazdi),

$$.. + \sum_{i \neq j} Cov(X_i, X_j)$$