

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 *kovaryans* deniyor zaten.

Tanim

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

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 alınmasidir.

Tanim boyle. Simdi bu tanima biraz bakip onun hakkında 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 faydalidir.

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 noktası 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

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

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?

$$\text{Var}(X) = E((X - E(X))^2)$$

Bu denklem aslinda

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

denkleminde Y yerine X kullandigimizda elde ettigimiz seydir, yani

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

$$\text{Cov}(X, X) = E((X - E(X))^2)$$

$$= \text{Var}(X)$$

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

Ozellik 2

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

Ispati kolay herhalde, (1) formuluunu uygulamak yeterli.

Teori

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

Ispat

Bu ispat cok kolay, esitligin sol tarafindaki carpimi parantezler uzerinden acarsak, ve beklenti lineer bir operator oldugu icin toplamın 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 $\forall E(Y)$ 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 ortalamalarına izafi olarak belirtilmistir, ve akilda canlandırılması daha rahat olabilir. Fakat matematiksel olarak bu iki ifade de aynidir.

İki ozellik bulduk bile. Bir ozellik daha,

Ozellik 3

$$\text{Cov}(X, c) = 0$$

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

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

Ispat icin alttaki formilde

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

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

Ozellik 5

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

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 ciktiği durumda oldugu gibi, bu ozellikte Y 'ye hicbir sey olmadı, o degismeden kaldı. Ayni sekilde 5. ozellikte X hic degismeden esitligin sagina aktarıldı sanki, sadece " Z durumu icin" yeni bir terim ekledik.

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

Ozellik 6

$$\text{Cov}(X + Y, Z + W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$$

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 $\text{Cov}(X, Y + Z)$ icin ama $\text{Cov}(Y + Z, X)$ yine ayni sonucu veriyor.

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

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

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 toplamın varyansini nasil hesaplariz?

Ozellik 7

$$\text{Var}(X_1 + X_2)$$

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

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

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)$$

Bu alanda bilinen tekerleme gibi bir baska deyiş, "eger kovaryans sifirsa toplamın varyansı varyansların toplamıdır". Hakikaten kovaryans sifir olunca ustteki denklemden dusecektir, geriye sadece varyansların toplamı kalacaktır. 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,

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

Sondaki toplamın indisinde bir numara yaptık, sadece 1 ile 2, 2 ile 3, vs. eslemek için, ve mesela 3 ile 1'i tekrar eslememek için. Tekrar dedik çünkü $\text{Cov}(X_1, X_3) = \text{Cov}(X_3, X_1)$. Eger indisleme numarasi kullanmasaydik, 2 ile carpimi cikartirdik (ona artik gerek olmazdi),

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

Simdi, korelasyon konusuna gelmeden once, bagimsizlik kavramini iyice anladigimizdan emin olalim. a

Teori

Eger X, Y bagimsiz ise bu degiskenler bagimsizdir, yani $\text{Cov}(X, Y) = 0$.

DIKKAT! Bu mantik cizgisinin tersi her zaman dogru olmayabilir, yani bagimsizlik kesinlikle $\text{Cov}(X, Y) = 0$ demektir, ama her $\text{Cov}(X, Y) = 0$ oldugu zaman ortada bir bagimsizlik var diyemeyiz. Bunu bir ornekle gorelim.

$$Z \sim N(0, 1), X = Z, Y = Z^2$$

Simdi X, Y kovaryansinin hesabi yapalim

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(Z^3) - E(Z)E(Z^2)$$

En sondaki terim sifirdir, çünkü hem $E(Z)$ ve $E(Z^3)$ sifirdir [hoca burada standart normalin tek sayili (odd) moment'leri hep sifirdir dedi]. O zaman su sonucu cikartiyoruz, X, Y arasinda korelasyon yok.

Ama bagimlilik var mi? Var. Çünkü hem X hem Y Z 'nin birer degiskeni, yani bu durumda X 'i bilmek bize Y 'yi tamamen bilmemizi sagliyor (sadece ek olarak bir kare aliyoruz). Tabii bagimlilik illa herseyin bilinmesi demek degildir, biraz bagimlilik ta olabilir, ama biraz bagimlilik bile varsa, bagimsizlik var diyemeyiz. Ayni sey ters yon için de gecerli, Y bilinince X 'in "buyuklugunu" bilebiliriz, karekok islemi oldugu için $-/+$ isareti bilemeyiz ama skalar bir buyuklugu elde edebiliriz. Yani ters yonde de bagimsizlik yoktur.

Korelasyon

Tanim

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)} \quad (2)$$

Bu arada hatirlarsak ustte SD ile gosterilen standart sapma, varyansin karesidir.

Bu tanim genelde kullanılan tanimdir. Fakat ben daha farkli bir tanimi tercih ediyorum. Standardize etmeyi hatirliyoruz degil mi? Bir rasgele degiskenden ortalamasini cikartip standart sapmaya bolunce standardize ediyorduk. Bunu kullanarak aslinda korelasyonu alttaki gibi tanimlayabiliriz,

$$\text{Corr}(X, Y) = \text{Cov}\left(\frac{X - E(X)}{\text{SD}(X)}, \frac{Y - E(Y)}{\text{SD}(Y)}\right) \quad (3)$$

Yani korelasyonun anlami aslinda sudur: X, Y degiskenlerini standardize et, ondan sonra kovaryanslarini al.

Niye kovaryans iceren ifadeyi tercih ediyoruz? Cunku, diyelim ki X, Y degiskenleri bir uzaklik olcusunu temsil ediyor, ve birimleri mesela nanometre. Fakat bir baskasi gelip ayni olcumu, atiyorum, isik yili olarak kullanmaya baslarsa problem cikabilir. Yani eger birim yoksa ve ben " X, Y korelasyonum 42" dersem, bunun ne oldugunu anlamak zordur. 42 onumuzdeki veriye gore kucuk mudur, buyuk mudur? Bilemeyiz. Yani 42 sayisi tabii ki evrendeki tum sorularin cevabidir [hoca bir filme atfen espri yapiyor, orada 42 sayisinin ozel bir anlami vardi], ama onumuzdeki problem icin, nedir?

Fakat ustteki formül olcu birimsiz (dimensionless) bir sonuc verir, yani bir olcu biriminden bahsetmeden birine rahatca aktarabilecegimiz bir bilgidir. Niye birimsiz oldu? Cunku X 'in birimi cm olsa, $X - E(X)$ yine cm, $\text{SD}(X)$ varyansin karekoku oldugu icin cm^2 'nin karekoku yine cm, cm bolu cm birim ortadan kalkar.

Bu arada (3) niye (2) ile aynidir? Eger bir rasgele degiskenden bir sabiti cikartirsam onun baska bir degisken ile kovaryansini degistirmis olmam. Ki standardize etme islemi bunu yapar. O zaman niye bu cikartma islemini yaptim? Cunku standardize etme islemini ozellikle kullanmak istedim - standardizasyon bilinen ve rahatca kullanilabilen bir islem. Standart sapmayi bolmeye gelirse, simdiye kadar gordugumuz ozelliklerden biri, bolumu disari alabilecegimizi gosteriyor, boyle olunca (2) ifadesini aynen elde ediyorum.

[Not: cok ilginç bir noktaya isaret etmek isterim. Hoca X, Y 'yi standardize etti, ama bu demek degil ki X, Y "standard normal" dagilimi haline geldi, X, Y 'in eger kendileri normal olsaydi o zaman standart normal elde ederdik. Diger durumlarda X, Y 'yi ortalamasi 0, varyansi 1 olan bir sey haline getirmis olduk sadece. Ama yine bu dusunceden devam edersek, eger X, Y normal olsaydi, standardize bize iki tane Z verirdi, ve $\text{Cov}(Z, Z) = 1$ olurdu. Yani hangi sekilde olursa olsun eger X, Y 'nin ikisi de normal ise korelasyon 1 demektir].

Onemli bir nokta daha: korelasyon her zaman -1 ve $+1$ arasindadir.

Teori

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Yani olcu biriminden bagimsiz olmasi avantajina ek olarak hep ayni skalada olan bir degerin rapor edilmesi de faydalidir. Eger korelasyon 0.99 bulursam bunun hemen yuksek bir korelasyon oldugunu bilirim.

Bu arada, Cauchy-Schwarz esitsizliginden bahsedeyim -ki bu esitsizlik tanimi tum matematikteki en onemli esitsizliklerden biridir- eger korelasyon formülünü lineer cebirsel sekilde ifade etseydim direk Cauchy-Schwarz esitsizligini elde ederdim.

Ispat

Once “WLOG cercevesinde” X, Y ’nin onceden standardize edilmis oldugunu kabul edelim. [WLOG ne demek? Matematikciler ispatlar sirasinda bunu bazen kullanirlar, genelleme kuvvetinde bir kayip olmadan (without loss of generality) takip eden sey kullanabiliriz demektir, yani “bir baska sey kullaniyorum, ama teori bu cercevede de hala gecerli” demek isterler].

Onceden standardize edildigini kabul etmek niye fark yaratmiyor? Cunku bunu gorduk, standart olmayan degiskenleri standardize edince yine ayni sonucu elde ediyorum, yani bir sey farketmiyor.

$\text{Var}(X + Y)$ ’i hesaplayalim.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \quad (4)$$

Simdi sembol olarak $\rho = \text{Corr}(X, Y)$ kullanalim,

Standardize ettigimizi kabul etmistik, o zaman $\text{Var}(X) = 1, \text{Var}(Y) = 1$. Ayrica (3)’te gordugumuz uzere, standardize durumda kovaryans korelasyona esittir, o zaman $\text{Cov}(X, Y) = \rho$, yani $2\text{Cov}(X, Y) = 2\rho$. Tum ifade,

$$\text{Var}(X + Y) = 1 + 1 + 2\rho = 2 + 2\rho$$

Peki farklari varyansi, $\text{Var}(X - Y)$ nedir? Bir numara kullanalim, $\text{Var}(X - Y)$ ’i $\text{Var}(X + (-Y))$ olarak gorelim,

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = 2 - 2\rho$$

Aslinda bu son ifade ispati tamamlamis oldu, cunku varyans negatif olmayam bir seydir, yani

$$0 \leq \text{Var}(X + Y) = 2 + 2\rho$$

$$0 \leq \text{Var}(X - Y) = 2 - 2\rho$$

Bu iki esitsizligi kullanarak

$$-2 \leq 2\rho$$

$$-2 \leq -2\rho$$

ve

$$-1 \leq \rho$$

$$\rho \leq 1$$

Multinom Dagilimin Kovaryansi

Kovaryansi multinom dagilimi baglaminda ele alalim, bildigimiz gibi multinom dagilimi bir vektordur [ve binom dagiliminin daha yuksek boyuttaki halidir, binom dagilimi bildigimiz gibi n deney icinde kac tane basari sayisi oldugunu verir], ve vektorun her hucresinde “vs. kategorisinde kac tane vs var” gibi bir deger tasinir, ki bu her hucre baglaminda “o kategori icin zar atilrsa kac tane basari elde edilir” gibi okunabilir.

Biz ise bu hucrelerden iki tanesini alip aralarindaki kovaryasyona bakmak istiyoruz. Gayet dogal bir istek.

Notasyon

Elimizde k tane obje var,

$$(X_1, \dots, X_k) \sim \text{Mult}(n, \vec{p})$$

Dikkat, p bir vektor, tabii ki, cunku binom durumunda p tek sayi idi, simdi “pek cok p ”ye ihtiyac var.

Her i, j icin $\text{Cov}(X_i, X_j)$ ’yi hesapla.

Eger $i = j$ ise $\text{Cov}(X_i, X_i) = \text{Var}(X_i) = np_i(1 - p_i)$.

ki son ifade binom dagiliminin varyansidir. Bu basit durum tabii ki, ilginç olan $i \neq j$ olmadigi zaman.

Tek ornek secelim, mesela $\text{Cov}(X_1, X_2)$, buradan gelen sonuc gayet kolayca genellestirilebilir.

Hesaba baslamadan once kabaca bir akil yurutelim; $\text{Cov}(X_1, X_2)$ icin arti mi eksi mi bir deger elde ederdik acaba? Multinom dagilimi hatirlayalim, belli sayida "sey" yine belli sayida kategori arasinda "kapisiliyor", yani bu kategoriler arasinda bir yaris var. O zaman herhangi iki kategorinin kovaryansinin negatif olmasini bekleriz.

Cozum icin (4) formulunu kullanacagim, ama secici bir sekilde,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

icinde $\text{Var}(X+Y)$, $\text{Var}(X)$, $\text{Var}(Y)$ 'i biliyorsam, geriye bilinmeyen $\text{Cov}(X, Y)$ kalir. Kisaltma amaciyla $c = \text{Cov}(X, Y)$ diyelim,

$$\text{Var}(X_1 + X_2) = np_1(1 - p_1) + np_2(1 - p_2) + 2c$$

Simdi $X_1 + X_2$ 'nin ne oldugunu dusunelim, bu yeni rasgele degisken "ya kategori 1 ya da 2" sonucunu tasiyan bir degiskendir, ki bu da yeni bir "birlesik" binom degiskenidir. Bu degiskenin p 'si toplami oldugu iki kategorinin p 'sinin toplamidir, yani $p_1 + p_2$. O zaman bu yeni degiskenin varyansi,

$$\text{Var}(X_1 + X_2) = n(p_1 + p_2)(1 - (p_1 + p_2))$$

Eh artik denklemdaki her seyi biliyoruz, sadece c 'yi bilmiyoruz, ona gore herseyi duzenleyelim,

$$n(p_1 + p_2)(1 - (p_1 + p_2)) = np_1(1 - p_1) + np_2(1 - p_2) + 2c$$

Burada biraz haldir huldur islem lazim [bu kısmi okuyucu isterse yapabilir], sonuc

$$\text{Cov}(X_1, X_2) = -np_1p_2$$

Genel olarak

$$\text{Cov}(X_i, X_j) = -np_i p_j, \forall i \neq j$$

Dikkat edin, bu sonuc her zaman negatiftir (cunku p degerleri olasilik degerleridirler, yani pozitif olmak zorundadirlar)

Ornek

Binom degiskenin varyansini hesaplayalim simdi. Bunu daha once yapmistik ama gostergec (indicator) rasgele degiskenleri kullanarak yapmistik bunu, simdi elimizde yeni bir arac var, onu kullanalim. Varacagimiz sonuc $\text{Var}(X) = npq$ olacak. Tanimlar,

$$X \sim \text{Bin}(n, p), X = X_1 + \dots + X_n$$

ki X_i degiskenleri i.i.d. Bernoulli.

Aslinda her X_i degiskeni bir gostergec degiskeni gibi gorulebilir. Diyelim ki bir A olayi icin gostergec degisken I_A olsun. Bu durumda

$$I_A^2 = I_A$$

$$I_A^3 = I_A$$

Degil mi? Gostergec sadece 1/0 olabiliyorsa onun karesi, kupu ayni sekilde olur. Bunu vurguluyorum, cunku bazen atlaniyor.

Peki $I_A I_B$? Ki A, B ayri ayri olaylar. Gayet basit,

$$I_A I_B = I_{A \cap B}$$

Bu normal degil mi? Esitligin solundaki carpim sadece her iki degisken de 1 ise 1 sonucunu verir, bu ise sadece A, B olaylari ayni anda oldugu zaman mumkundur, ki bu ayni anda olmak kume kesismesinin tanimidir.

Bernoulli durumuna donelim, her Bernoulli icin

$$\text{Var}(X_i) = EX_j^2 - E(X_j)^2$$

$X_j^2 = X_j$ 'dir, bunu biraz once gorduk, ve Binom degiskenleri gostergec gibi goruyoruz, o zaman $EX_j^2 = E(X_j) = p$.

$$\text{Var}(X_i) = p - p^2 = p(1 - p) = pq$$

Verisel Kovaryans (Empirical Covariance)

Eger verinin kolonlari arasindaki iliskiyi gormek istersek, en hizli yontem matristeki her kolonun (degiskenin) ortalamasini kendisinden cikartmak, yani onu "sifirda ortalamak" ve bu matrisin devrigini alarak kendisi ile carpma. Bu islem her kolonu kendisi ve diger kolonlar ile noktasal carpimdan gecirecektir ve carpim, toplama sonucunu nihai matrise yazacaktır. Carpimlari bildigimiz ozelligine gore, arti deger arti degerle carpilince arti, eksi ile eksi arti, eksi ile arti eksi

verir, ve bu bilgi bize ilinti bulma hakkında güzel bir ipucu sunar. Pozitif sonucun pozitif korelasyon, negatif ise tersi şekilde ilinti olduğu sonucuna böylece kolayca erisebiliriz.

Tanim

$$S = \frac{1}{n}(X - E(X))^T(X - E(X))$$

Pandas ile `cov` cagrisi bu hesabi hizli bir sekilde yapar,

```
print df.cov()
```

	Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	0.685694	-0.039268	1.273682	0.516904
Sepal Width	-0.039268	0.188004	-0.321713	-0.117981
Petal Length	1.273682	-0.321713	3.113179	1.296387
Petal Width	0.516904	-0.117981	1.296387	0.582414

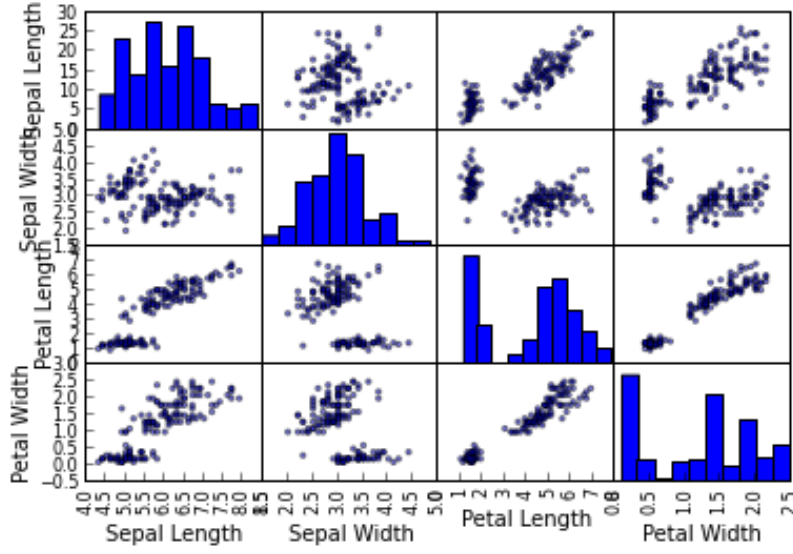
Eger kendimiz bu hesabi yapmak istersek,

```
means = df.mean()
n = df.shape[0]
df2 = df.apply(lambda x: x - means, axis=1)
print np.dot(df2.T, df2) / n
```

```
[[ 0.68112222 -0.03900667  1.26519111  0.51345778]
 [-0.03900667  0.18675067 -0.319568   -0.11719467]
 [ 1.26519111 -0.319568    3.09242489  1.28774489]
 [ 0.51345778 -0.11719467  1.28774489  0.57853156]]
```

Verisel kovaryansin sayisal gosterdigini grafiklemek istersek, yani iki veya daha fazla boyutun arasindaki iliskileri grafiklemek icin yontemlerden birisi verideki mumkun her ikili iliskiye grafiksel olarak gosterme'dir. Pandas `scatter_matrix` bunu yapabilir. Iris veri seti uzerinde gorelim, her boyut hem y-ekseni hem x-ekseninde verilmiş, iliskiye gormek icin eksende o boyutu bulup kesisme noktalarindaki grafige bakmak lazim.

```
import pandas as pd
df = pd.read_csv('iris.csv')
df = df.ix[:, 0:4]
pd.scatter_matrix(df)
plt.savefig('stat_summary_01.png')
```



İliski olduğu zaman o ilişkiye tekabül eden grafikte “düz çizgiye benzer” bir görüntü olur, demek ki değişkenlerden biri artınca diğeri de artıyor (eger çizgi soldan sağa yukarı doğru gidiyorsa), azalınca diğeri de azalıyor demektir (eger çizgi aşağı doğru iniyorsa). Eger ilinti yok ise bol gürültülü, ya da yuvarlak kuryeye benzer bir şekil çıkar. Üstteki grafiğe göre yaprak genişliği (petal width) ile yaprak boyu (petal length) arasında bir ilişki var.

Tanım

X, Y rasgele değişkenlerin arasındaki kovaryans,

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

Yani hem X hem Y’nin beklentilerinden ne kadar saptıklarını her veri ikilisi için, çıkartarak tespit ediyoruz, daha sonra bu farkları birbiriyle çarpıyoruz, ve beklentisini alıyoruz (yani tüm olasılık üzerinden ne olacağını hesaplıyoruz).

Ayrı ayrı X, Y değişkenleri yerine çok boyutlu X kullanırsak, ki boyutları m, n olsun yani m veri noktası ve n boyut (özellik, öge) var, tanımı şöyle ifade edebiliriz,

$$\Sigma = \text{Cov}(X) = E((X - E(X))^T (X - E(X)))$$