

Istatistik - Ders 1

Bu notlar makine öğrenimi, veri madenciliği gibi konularda gerekli olasılık ve istatistik bilgisini paylaşmak için hazırlanıyor. Notlarda olasılık ve istatistik aynı anda anlatılacak, ve uygulamalara ağırlık verilecek.

Orneklem Uzayı (Sample Space)

Orneklem uzayı Ω bir deneyin mümkün tüm olasılıksal sonuçların (outcome) kümesidir. Eğer deneyimiz ardi ardına iki kere yazı (T) tura (H) atıp sonucu kaydetmek ise, bu deneyin mümkün tüm sonuçları şöyledir

$$\Omega = \{HH, HT, TH, TT\}$$

Sonuçlar ve Olaylar (Outcomes and Events)

Ω içindeki her nokta bir sonuctur (outcome). Olaylar Ω 'nin herhangi bir alt kümesidir. Mesela üstteki yazı-tura deneyinde “iki atışın içinden ilk atışın her zaman H gelmesi olayı” böyle bir alt küme oluşturur, bu olaya A diyelim, ve bu küme $A = \{HH, HT\}$ olacaktır.

Ya da bir deneyin sonucu ω fiziksel bir ölçüm , diyelin ki sıcaklık ölçümü. Sıcaklık \pm , reel bir sayı olduğuna göre, $\Omega = (-\infty, +\infty)$, ve sıcaklık ölçümünün 10'dan büyük ama 23'ten küçük ya da esit olma “olayı” $A = (10, 23]$. Koseli parantez kullanıldı çünkü sınır değerini dahil ediyoruz.

Örnek

10 kere yazı-tura at. A = “en az bir tura gelme” olayı olsun. T_j ise j 'inci yazı-tura atışında yazı gelme olayı olsun. $P(A)$ nedir?

Bunun hesabi için en kolayı, hiç tura gelmeme, yani tamamen yazı gelme olasılığını, A^c 'yi hesaplamak, ve onu 1'den çıkartmaktır. c sembolü “tamamlayıcı (complement)” kelimesinden geliyor.

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P(\text{hepsi yazı}) \\ &= 1 - P(T_1)P(T_2)\dots P(T_{10}) \\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx .999 \end{aligned}$$

Rasgele Değişkenler (Random Variables)

Bir rasgele degisken X bir eslemedir, ki bu esleme $X : \Omega \rightarrow \Re$ her sonuc ile bir reel sayi arasindaki eslemedir.

Olasilik derslerinde bir noktadan sonra artik ornekleme uzayindan bahsedilmez, ama bu kavramin arkalarda bir yerde her zaman devrede oldugunu hic aklimizdan cikartmayalim.

Ornek

10 kere yazi-tura attik diyelim. VE yine diyelim ki $X(\omega)$ rasgele degiskeni ω siralamasindaki (sequence) tura sayisi. Iste bir esleme. Mesela eger $\omega = HHTHHTHHTT$ ise $X(\omega) = 6$. Tura sayisi eslemesi ω sonucunu 6 sayisina esledi.

Ornek

$\Omega = \{(x, y); x^2 + y^2 \leq 1\}$, yani kume birim cember ve icindeki reel sayilar (unit disc). Diyelim ki bu kumeden rasgele secim yapiyoruz. Tipik bir sonuc $\omega = (x, y)$ 'dir. Tipik rasgele degiskenler ise $X(\omega) = x$, $Y(\omega) = y$, $Z(\omega) = x + y$ olabilir. Goruldugu gibi bir sonuc ile reel sayi arasinda esleme var. X rasgele degiskeni mesela sonucu x 'e esliyor, yani (x, y) icinden sadece x 'i cekip cikartiyor. Benzer sekilde Y, Z degiskenleri var.