Olasılık, İstatistik ve Deneysel Bilim

Arkadaşlar öncelikle hepinize merhaba. Bu bölümde benim yazılırımı okuyacaksınız o yüzden kendimi tanıtarak başlamak istiyorum. Ben Arda Ünal. Gaziantep Üniversitesi Elektrik-Elektronik Mühendisliği mezunuyum. Şuan CERN’de Uygulamalı Fizik alanında Teknik Öğrenci sıfatıyla staj yapıyorum. Buradaki çalışma konum Büyük Hadron Çarpıştırıcısındaki (LHC), Doğrusal Akım Işın Akımı Transformatörü’nün(DCCT ya da BCTDC) kalibrasyonu üzerine bir proje. Umarım yazacağım şeyler öğreticiği olduğu kadar eğlenceli de olur. Hoşgeldiniz ve iyi eğlenceler.

Bu yazımızda bilimde deneyin yeri, olasılık ve istatistiğin kullanım şekli ve bazı önemli terimleri tanıtıp, tartışacağız.Günümüzde bir çok bilim dalı(Fizik, Matematik, Kimya, Biyoloji, Meteoroloji, Biyofizik …) temeli deneysel süreçlerle ilerlemeye ve gelişmeye başlamıştır. Teknolojinin kendiside buna dahildir elbette. Peki biz neden deney yapmaya, neden İstatiksel analizler yapmak zorundayız? Bunun bir çok sebebi var elbette ama ben yazılarımda sorulara direkt olarak cevap vermek yerine siz okuyucuların bunu verdiğim örneklerden keşfetmesini isteyeceğim. Bu şekilde öğrendiğiniz şeyler çok daha uzun süre aklınızda kalıcaktır diye düşünüyorum.

Öncelikle şans kelimesinden başlayalım. Mesela “Benim 100 yaşıma kadar yaşama şansım var.” veya “ Tavla oynarken düşeş gelme şansı çok düşüktür” veya “ Bir fizikçi belirli bir bölgedeki parçacıkları Scintillator dedektörüyle tespit etmeye çalışırken şunu diyebilir: Bu bölgede 25 cm^2 alan başına, 10 saniye içinde 25 yüklü parçacık tespit etme şansımız yüksek.”. Aslında şans kelimesinin başka kullanım alanları olsada, yukarıdaki cümlelerde aslında hep aynı şeyin altını çizerler “olasılık”. Şans kelimesini yeterince bilgimiz olmayan olayları veya olguları tahmin etmek için kullanırız. Olasılık ise bu olayların olabilirliğini sayısal olarak resmeden, istatistik ise sayısal hale gelmiş verileri, olay sayısına bağlı olarak, tahmin etmeye yarayan araçtır.

Şimdi klasik olarak verilen yazı-tura deneyini bizde inceleyelim. Bir parayı attığımızda yazı veya tura gelme olasılığının yarı yarıya olduğunu hepimizin bildiğine şüphem yok. Lakin konu bu olasılığı nasıl belirlediğimizdir. Hadi birlikte bu deneyi yapalım ve olasılığın nasıl bir şey olduğunu bu deney ile anlayalım. Elimize bir para alalım ve atmaya başlayalım. 10 kere atalım ve sonuçlara bakalım. Benim yaptığım deneyde 7 yazı ve 3 tura sonucu çıktı. O zaman çok basit bir mantık kullanarak şu sonucu çıkarabiliriz : Yazı gelme olasığı 10 da 7(70%) iken, ture gelme olasılığı 10 da 3(30%)! Ama bizim bilgimiz 50% olmasını gerektiğini söylüyor. Fakat deneyimiz bizi yanılttı. Bir daha deneyelim belki bir yanlışlık oldu. 6 tura, 4 yazı. Bir daha 2 yazı, 8 tura. Bir daha 6 yazı, 4 tura. Ben bu deneylere bakarak nasıl yüzde 50 şans ile bir parayı attığımızda yazı veya tura gelecek diyebilirim. Diyemem. Peki ne diyebilirim. Belki para bozuktur, parayı değiştireyim. Tekrar deney yapalım başka para ile ve sonuçları yazalım. Sonuçta göreceğimiz şey sorunun parada olmamasıdır. Olasılık deneyimizin sonucunu sadece tahmin etmeye yarayan bir araçtır. Yüzde 50 şans ile yazı veya tura gelme ihtimali vardır. Peki bunu nasıl ispatlayabiliriz. Tabiki deney sayımızı arttırarak! Olasılık teorisi bize tahmin yapma gücü verir. Elimizde çok fazla değişen bir sistem olabilir ama sistemin kendisinin ortalama bir değeri olacaktır. Aslında deney sayımızı arttırarak olasılıkların dağılımını göreceğiz. Mesela 30 kere atalım bir parayı ve sonuçları yazalım. Bu 30 kere atma deneyini de 30 kere tekrarlayalım. Bunları bir yere yazalım ve sonuçları görelim. Bunu tabi ki elimiz ile yapmayacağız. Bilgisayarlar günümüzün modern deney araçlarıdır ve biz de bunu sonuna kadar kullanarak kendimizi geliştirebiliriz. Hadi bir küçük simülasyon yapalım.

import random # Python random modulü

import matplotlib.pyplot as plt # Grafik Çizim modulü

# İlk önce 30x30 tane yazı, tura atalım ve bu sonuçları gözlemleyelim.

olaylar = ["yazı", "tura"] # İki farklı sonuç olabilir. Yazı veya tura

sonuc\_listesi = [] # 30 deney için ayrı bir liste oluşturduk

for j in range(30):

yazi = 0

tura = 0

for i in range(30):

rastgele\_secim = random.choice(olaylar) # Olaylar içinden rastgele bir olay seçiyoruz

if rastgele\_secim == "yazı":

yazi += 1

else:

tura += 1

sonuc = [yazi, tura]

sonuc\_listesi.append(sonuc) # Liste içinde oluşan listeye sonuçlarımızı ekliyoruz

# Şimdi görselleştirme vakti

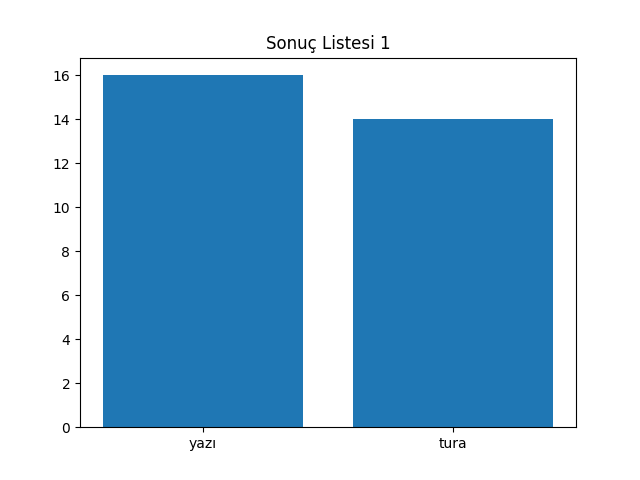
for i in range(30):

plt.bar(olaylar, sonuc\_listesi[i])

plt.title("Sonuç Listesi " + str(i + 1))

plt.show()

# 30 tane grafik ve her grafik yazı ve tura sonuçlarından kaç tane olduğunu gösteriyor. Kaç kere 15 e 15 sonucunu gördünüz?



Haydi bir de 30 deneyin sonuçlarındaki yazıların sayısının dağılımına bakalım.

import random

import matplotlib.pyplot as plt

olaylar = ["yazı", "tura"]

sonuc\_listesi = []

for j in range(30):

yazi = 0

tura = 0

for i in range(30):

rastgele\_secim = random.choice(olaylar)

if rastgele\_secim == "yazı":

yazi += 1

else:

tura += 1

sonuc = yazi

sonuc\_listesi.append(sonuc)

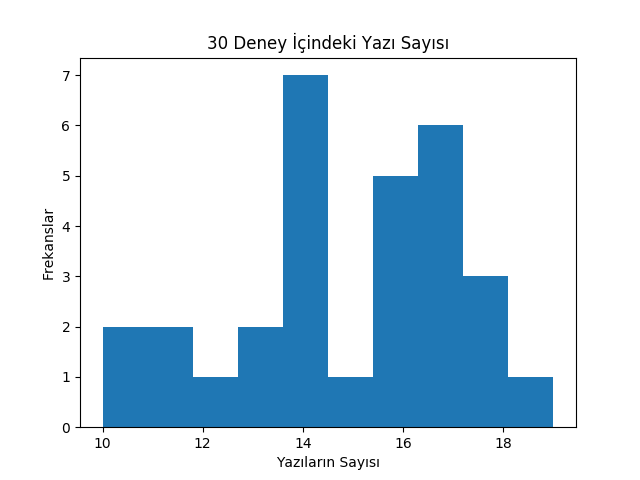
plt.hist(sonuc\_listesi)

plt.title("30 Deney İçindeki Yazı Sayısı")

plt.xlabel("Yazıların Sayısı")

plt.ylabel("Frekanslar")

plt.show()



Aynı deneyi olay sayısını arttırarak bir daha yapalım.

import random

import matplotlib.pyplot as plt

olaylar = ["yazı", "tura"]

sonuc\_listesi = []

for j in range(1000):

yazi = 0

tura = 0

for i in range(30):

rastgele\_secim = random.choice(olaylar)

if rastgele\_secim == "yazı":

yazi += 1

else:

tura += 1

sonuc = yazi

sonuc\_listesi.append(sonuc)

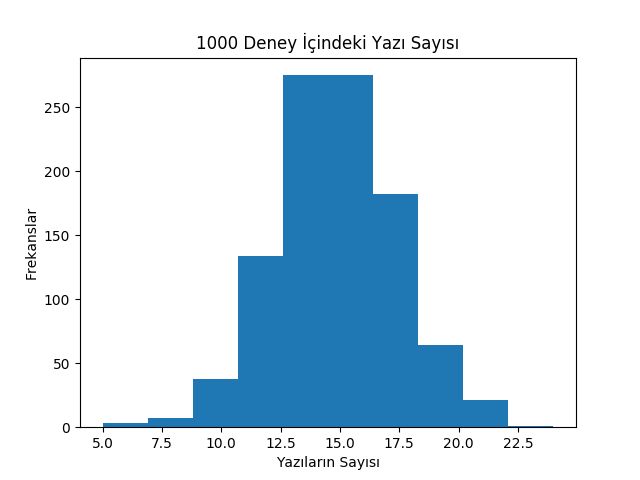
plt.hist(sonuc\_listesi)

plt.title("1000 Deney İçindeki Yazı Sayısı")

plt.xlabel("Yazıların Sayısı")

plt.ylabel("Frekanslar")

plt.show()



Deneyin sonuçlarını incelediğinizi varsayarak gördüğünüz sonuçlar üzerine düşünüp buradan nasıl yazı veya tura sonucunun %50 olasılık ile gelebileceğini keşfetmenizi tavsiye ediyorum ve işte size düşünmek için bir ipucu: Olaylarımız birbirinden bağımsız olduğu için ayrık olay olarak adlandırılır. Önceki attığımız para bir sonraki paraya etki etmeyecektir. Bu demek oluyorki her para atışımızda aslında olasılık en baştaki para atışımızla aynı olasılık sonucuna sahip. Aslında cevabı vermiştik. Deney sayımızı arttırmamız gerekti. Bir şeyin olasılığından bahsediyorsak eğer, bu olayın ya da olgunun tekrarlanabilir ve gözlemlenebiliyor olması gerektir. Tekrarlı olayların sonuçlarını gözlemleyerek bir olasılık değerine karar vermemiz gerekiyor. Bunu ise yukarıda verdiğimiz histogram dağılımının ortalaması olarak ele alabiliriz. Gördüğünüz üzere yazıların sayısı 14, 15 ve 16 kere olacak şekilde dağılmıştır. Bu durumda ortalama olarak 15’i seçebiliriz. Bu bir Gaussian dağılımdır(Dağılımlara sonraki bölümlerimizde değineceğiz). Bir önceki simülasyonumuzda yaptığımız deneyi bu sefer farklı bir şekilde gerçekleştireceğiz.

import random

import matplotlib.pyplot as plt

olaylar = ["yazı", "tura"]

yazi = 0

tura = 0

for i in range(1000):

rastgele\_secim = random.choice(olaylar)

if rastgele\_secim == "yazı":

yazi += 1

else:

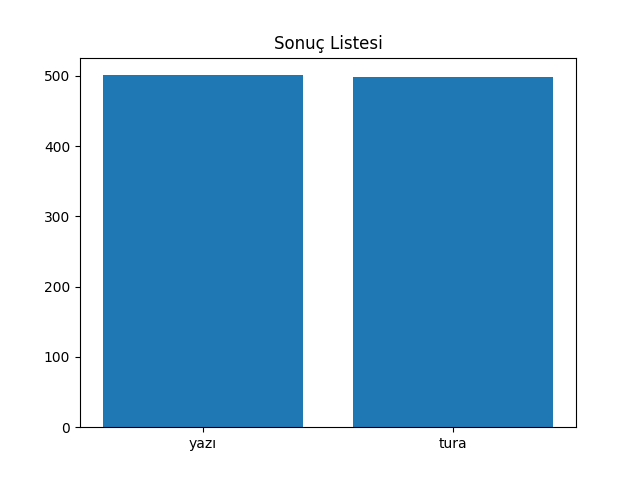
tura += 1

sonuclar = [yazi, tura]

plt.bar(olaylar, sonuclar)

plt.title("Sonuç Listesi")

plt.show()



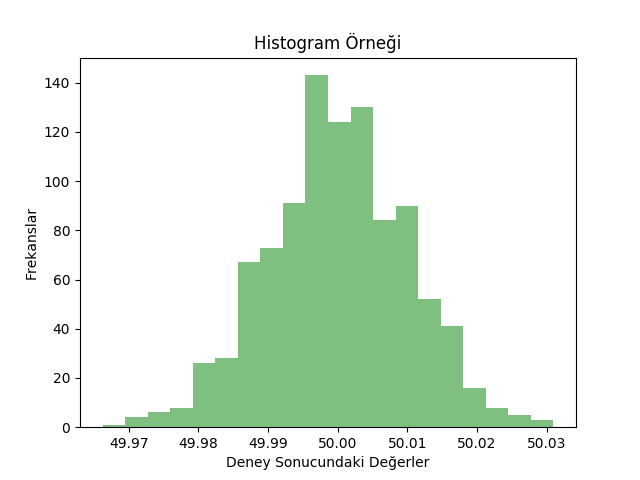
Eminim artık anlamaya başladık. Aslında deney sayısını arttırdıkça olayın sonucunun belirli bir dağılımı (bu durumda düzgün “uniform” dağılım) olduğunu keşfediyoruz. Heyecan verici tarafı ise iki ardışık deneyin birbirinden bağımsız olmasıyla birlikte iki sonucunda gelme olasılığının deney sonucu arttıkça eşit olmaya başladğını keşfediyoruz. İşte olasılığın tahmin gücü!

Peki bunu nerede kullanıyoruz? Olasılık bütün dünyadaki yaşama ait bir kuramdır. Atom altı parçacıkların doğasından ekonomiye kadar geniş bir kullanımı vardır ve bilimin en önemli çarklarından biridir. Olasılığın doğasını iyi anlamak demek aslında doğanın kendisini daha iyi anlamak demektir!

Bu yazımız bu kadar bir sonraki yazıda görüşmek üzere hoşçakalın!

Burası diğer yazıya aittir.

Ama simülasyona geçmeden önce size bir kaç tane data görselleştirme aracından ve dağılımlardan bahsedeceğim “Histogram, düzgün dağılım, üssel dağılım ve ”. Histogram bir datanın sayısına göre dağılımını temsil eder. Veri analizi yapan kimselerin kullandığı çok güçlü bir araçtır. Elimizdeki verilerin nasıl bir dağılıma sahip olduğu hakkında bilgi verir. Bir çok farklı dağılım şekli vardır lakin onlara sonra geleceğiz. Şimdi bir histogram örneği göstermek istiyorum sizlere.



Burada gördüğün çizim bir histograma aittir. Bu örnekte bir Gaussian dağılım görüyorsunuz. Gauss dağılımı görebileceğiniz en geniş dağılım örneğidir. Mesela bir elimizde bir direnç olsun ve bu direncin fabrikasyonunu yapan şirket bu direncin 50 ohm olduğunu iddia ediyor olsun. Biz bunu ispatlamak istiyoruz diyelim. Elimize bir dijital multi metre(dmm) alıp bir dizi ölçümden sonra şu sonucları elde ediyoruz (Not : dmm’izin mükemmel bir dmm olduğunu varsayıyoruz):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 50.03 ohm | 49.98  ohm | 49.99  ohm | 50.00  ohm | 50.01  ohm | 49.96  ohm | 50.03  ohm | 50.02  ohm |

Burada 8 tane deney sonucu olmasına rağmen biz eney sayımızı arttırırsak eğer bunu okumamız ve karşılaştırmamız zorlaşacaktır. Bu ölçümleri yukarıda gördüğümüz histogram ile görselleştirerek çok daha kolay okuyabilir ve çok açık bir şekilde en yüksek sayıdaki verinin 50 olduğunu görürüz.