Maximum Likelihood, Likelihood, log-likelihood, Joint Prob

* Joint Probability ⇒

ÖR Bir para 2 kez atılıyor } 2 adet Rastgele değişken vardır.
Bir zar 1 kez atılıyor

X = Paranın tura gelmelerinin sayısı = {0, 1, 2}
Y = Zarın 3 gelmelerinin sayısı = {0, 1}

Sonuç olarak ⇒ $P(X=0, Y=0) = 1/2 \times 1/2 \times 5/6$
$P(X=0, Y=1) = 1/2 \times 1/2 \times 1/6$

Yani 2 farklı rastgele değişkenin olasılıklarına ait probabilitydir.

(ortak olasılık Dağılımı)
Joint Probability Distribution → 2 rastgele Değişken
(2 rastgele değişkenin aynı anda gerçekleşmelerinin olasılığı)
→ Continiuos
→ Discrete

Probability Distribution → 1 rastgele Değişken
→ Contiuous
→ Discrete

İstatistik, any value that is calculated from a given **Sample**.

"Max Likelihood" tüm dağılımlarda model parametrelerini optimize etmeye çalışır.

ÖR: Exponential bir dağılımda λ'yı
Normal bir dağılımda σ veya/ve μ'ni gibi.

Örnek ile gidelim

Amacımız, elimizde bulunan verilerin ait olduğu yığını (popülasyon) tahmin etmektir. Çünkü genelde popülasyona ait parametrenin gerçek değerini bilmeyiz bir bayli zordur. Maximum likelihood uygulanabilecek çözümlerden bir tanesidir.

"Olabilirlik fonksiyonu" genelde Kesikli ve Sürekli olasılık dağılımları için farklı şekilde tanımlanır.

— Mesela; Discrete Durumda

$$P(HH \mid p_H = 0.5) = 0.5^2 = 0.25$$

Yani adil bir para olduğunda Tura gelme olasılığımız = 0.5 olacak. 2 kez para atıldığında "HH" gelme olasılığımız 0.25 olacak.

Likelihood ise;

$\mathcal{L}(p_H = 0.5 \mid HH) = 0.25$ >>> Given the observed data "HH", the likelihood that the model parameter $p_H = 0.5$ is 0.25

Yani; for each value of p_H, we can calculate the corresponding likelihood.

Discrete Durumda: Likelihood, parametrenin gerçek değeri θ olduğunda x outcome'unu gözleme olasılığıdır (θ = p_H oldu bu case'de)

$X = x$ } Random Variable

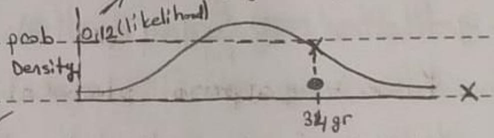
Yani; $P(HH \mid p_H = 0.3) = 0.3^2 = 0.09$ Unfair Para

$\mathcal{L}(p_H = 0.3 \mid HH) = 0.09$

Continious Durumda: Likelihood, parametrenin gerçek değeri θ olduğunda x outcome'unun olasılık yoğunluğuna eşittir.

$X = x$ } Random Variable $\quad \mathcal{L}(\theta \mid x)$

Bu durumda θ = μ, σ olabilir



Assume mean = 32, sd = 2.5

$\Rightarrow \mathcal{L}$(mean = 32 and sd = 2.5 | mouse weight 34 gr)

The likelihood of distribution with mean = 32, s.d = 2.5 given mouse with weight 34 gr is = 0.12.

*Amaç ne demiştik? Popülasyon parametrelerini tahmin etmek.

Reel bir örnekle Gidelim

Popülasyondan çekilmiş $X_1, X_2, \ldots, X_n$ örnekleri olsun elimizde. Varsayımımız şu olmalı: Örnekler birbirinden bağımsız.

Popülasyona ait dağılım tipini biliyoruz (Normal olsunlar)

$\underbrace{\mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n)}_{\text{likelihood Function}} = \mathcal{L}(\theta \mid x_1) \times \ldots \times \mathcal{L}(\theta \mid x_n)$

Bilinen x'lere karşı bilinmeyen θ parametresinin vurgulanması $= \mathcal{L}(\theta \mid x_1, x_2 \ldots, x_n)$

Bu fonksiyonu maksimum yapmak için $\mathcal{L}(\theta)$'nın θ parametresine göre türevi alınmalı.

$\mathcal{L}(\theta)$'nın e tabanına göre logaritmasını alalım.

$$\ln \mathcal{L}(\theta) = \ln\left(p(x_1 \mid \theta) \cdot p(x_2 \mid \theta) \cdot \ldots \cdot p(x_n \mid \theta)\right)$$
$$= \ln p(x_1 \mid \theta) + \ln p(x_2 \mid \theta) + \ldots + \ln p(x_n \mid \theta)$$

Log-likelihood $\Longrightarrow \sum_{k=1}^{n} \ln p(x_k \mid \theta)$

Kısaca çarpma işlemini toplamaya çevirdik, böylece türev aşamasında işler kolay olsun.

Daha sonra $\sum_{k=1}^{n} \ln p(x_k \mid \theta)$'nın θ'ya göre türevi alınacak $\ln \mathcal{L}(\theta)$'yı maksimum yapan θ değeri bulunur.

Kitap Açıklaması: Maksimum Olabilirlik Yöntemi, Bayes Çıkarımının aksine, parametreleri sabit bir nokta olarak görür. Yani yapılan işlem sonucunda çıkan sonuç sabit bir sayıdır.

$\mathcal{X} = \{x^t\}_{t=1}^{N}$, x^t are instances (gözlem/örnek) drawn from same known probability density family ($p(x \mid \theta)$).

$x^t \sim p(x \mid \theta)$

Likelihood func $\rightarrow \mathcal{L}(\theta \mid \mathcal{X}) \equiv p(\mathcal{X} \mid \theta) = \prod_{t=1}^{N} p(x^t \mid \theta)$

Log likelihood func $\rightarrow \mathcal{L}(\theta \mid \mathcal{X}) \equiv \log l(\theta \mid \mathcal{X}) = \sum_{t=1}^{N} \log p(x^t \mid \theta)$

../..../....

Not: We want to find θ that makes sampling x^t from $p(x|\theta)$ as likely as possible. Because x^t are independent, the likelihood of parameter θ; given sample X is the product of the likelihoods of individual points:

$$\mathcal{L}(\theta|X) \equiv \underbrace{p(X|\theta)}_{\text{prob density}} = \prod_{t=1}^{N} p(x^t|\theta)$$

prob density for continuous dist.;

prob Mass for Discrete Dist.