3. Gün
3. Ders:
R'ye Giriş

Bulaşıcı hastalık dinamiklerinin R'de modellenmesi üzerine kısa kurs

Ankara, Türkiye, Eylül 2025

Dr Juan F Vesga

Oturumun amaçları

- Belirsizlik kaynaklarını anlamak
- Belirsizliği bildirmenin önemini anlamak
- Model belirsizliğini ele almaya yönelik temel yöntemleri öğrenmek

Problemin doğası

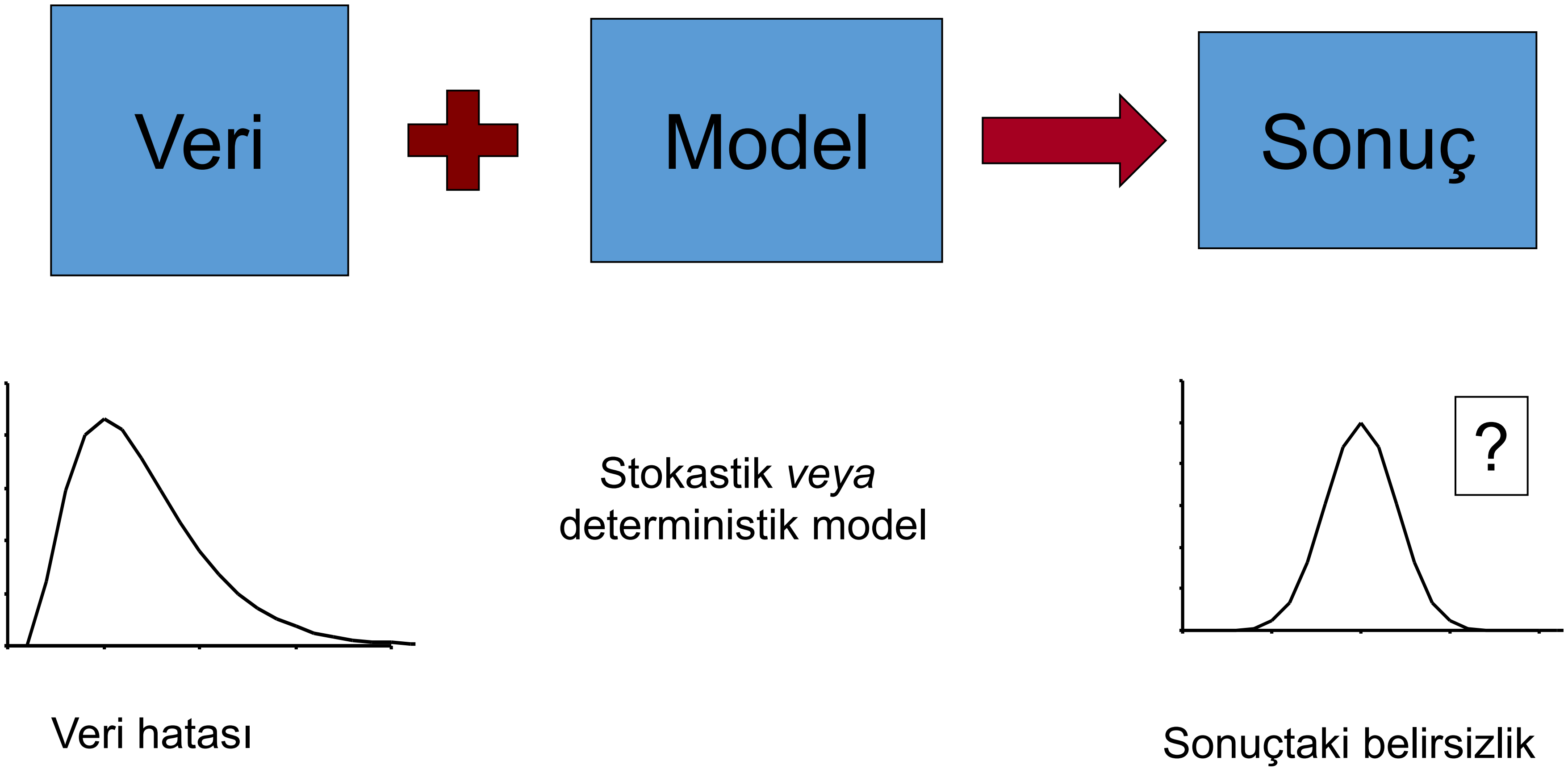
Veri + **Model** → **Sonuç**

Ham veri: diğer çalışmalardan alınan insidans,
serolojik veriler vb.
veya parametre değerleri

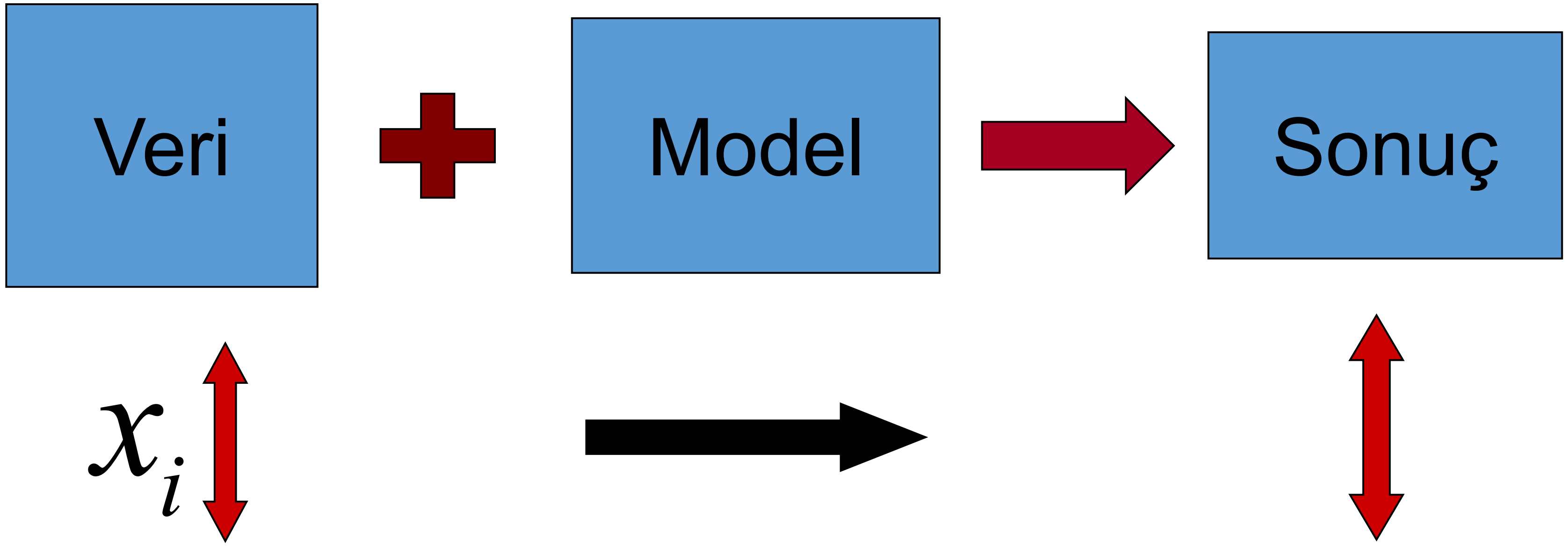
Deterministik veya
Stokastik

Sistemin özelliği veya istatistiği (Temas oranı, R_0 vb.)

Belirsizlik



Duyarlılık



Deterministik model için, iki tür:
•Yerel duyarlılık: parametrede bilinen değişiklik için sonuçta meydana gelen değişiklik.
•Global duyarlılık: sonuçtaki belirsizliğin herhangi bir girdi parametresinin belirsizliğiyle olan ilişkisi.

Parametre duyarlılığını gösterme modeli

Sonlu popülasyonda SIR tipi salgın, ör. İnfluenza, kızamık, kızamıkçık, Hepatit A.

Senaryo: I_0 enfektelerinin N-I_0 sayıda duyarlı bireyden oluşan popülasyona karışması.

$$S \rightarrow I \quad \text{: oran} \ \lambda = \frac{\beta I}{N} \ \text{/gün,}$$

$$I \rightarrow R \quad \text{: oran } \mu \text{ /gün}$$

$$R_0 = \beta / \mu$$

Parametre duyarlılığını gösterme modeli

Model aşağıdaki denklemler sistemiyle tanımlanabilir:

$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$

$I_0 = 100$

Parametre duyarlılığını gösterme modeli

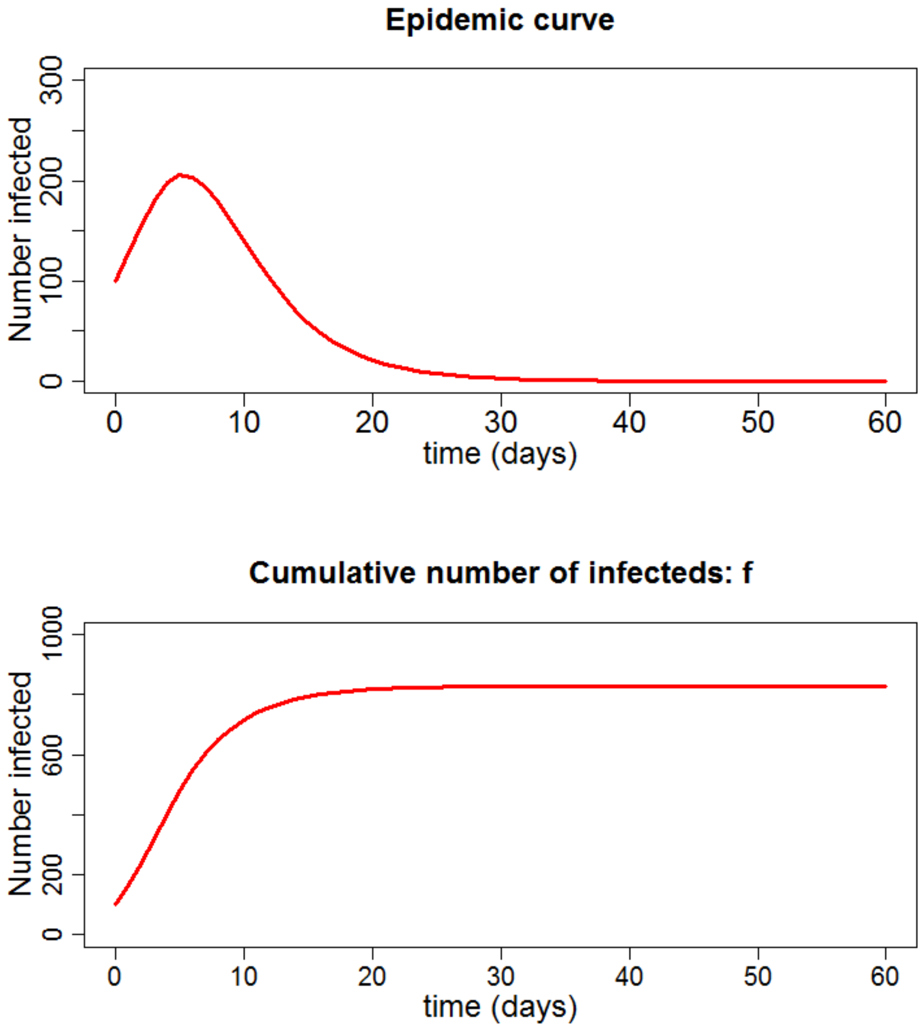
$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$

$I_0 = 100$



Parametre duyarlılığını gösterme modeli

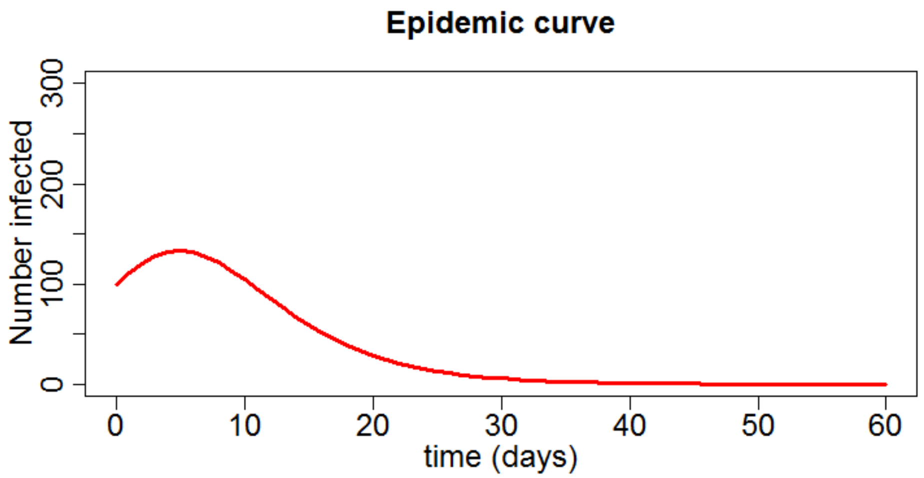
$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

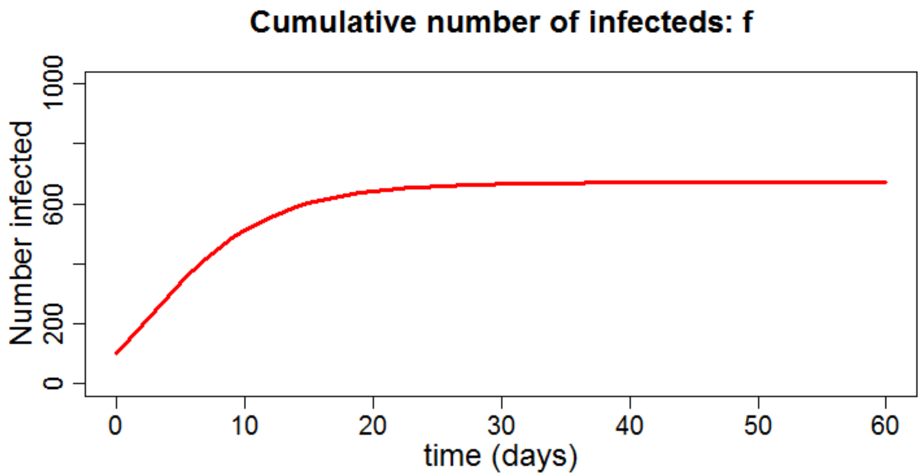
$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

R_0 = **1,5**

$I_0 = 100$





Parametre duyarlılığını gösterme modeli

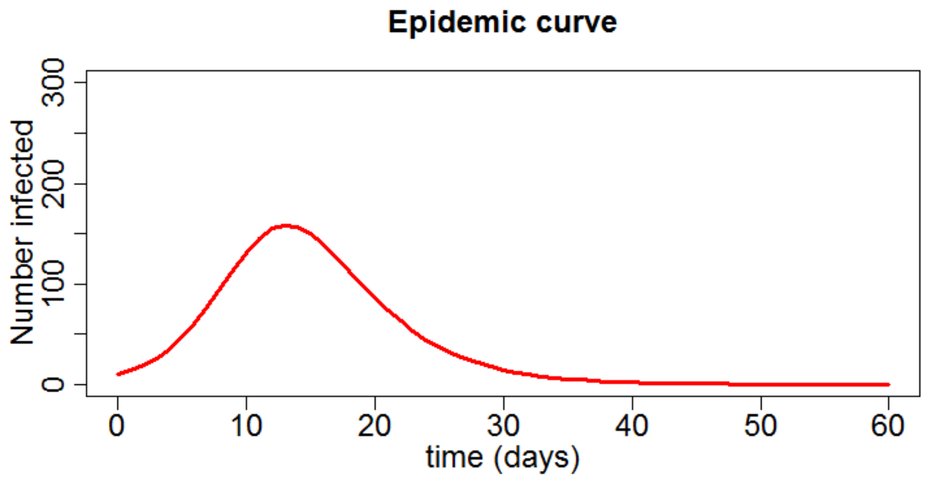
$$\frac{dS}{dt} = -\beta \frac{I}{N} S$$

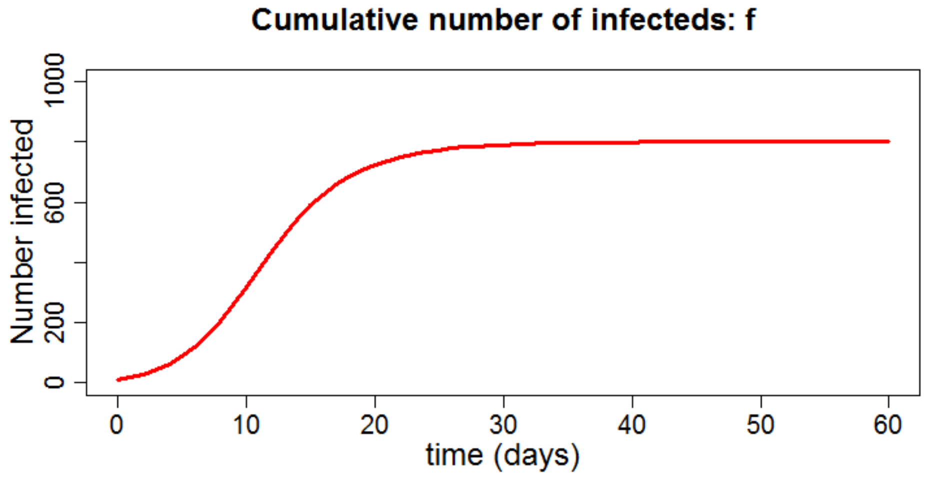
$$\frac{dI}{dt} = \beta \frac{I}{N} S - \mu I$$

$$\frac{dR}{dt} = \mu I$$

$R_0 = 2$

$I_0 = \mathbf{10}$





Parametre duyarlılığını gösterme modeli

I_0 ve R_0'a bağlı olduğundan salgın sırasında enfekte
olan kişi sayısıyla (yani nihai boyut (f) veya atak hızı) ilgileniyoruz.

Deterministik model için,

$$1 - f = (1 - I_0 / N)\exp(-R_0 f)$$

Parametre duyarlılığını gösterme modeli

I_0 ve R_0'a bağlı olduğundan salgın sırasında enfekte
olan kişi sayısıyla (yani nihai boyut (f) veya atak hızı) ilgileniyoruz.
Deterministik model için,

$$1-f=(1-I_0/N)\exp(-R_0 f)$$

***R_0*'daki varyasyona duyarlılık:**

I_0	R_0	f
100	1,25	544
100	1,5	671
100	**2**	**828**
100	3	948
100	4	982

Parametre duyarlılığını gösterme modeli

I_0 ve R_0'a bağlı olduğundan salgın sırasında enfekte
olan kişi sayısıyla (yani nihai boyut (f) veya atak hızı) ilgileniyoruz.
Deterministik model için,

$$1 - f = (1 - I_0 / N)\exp(-R_0 f)$$

I_0'daki varyasyona duyarlılık:

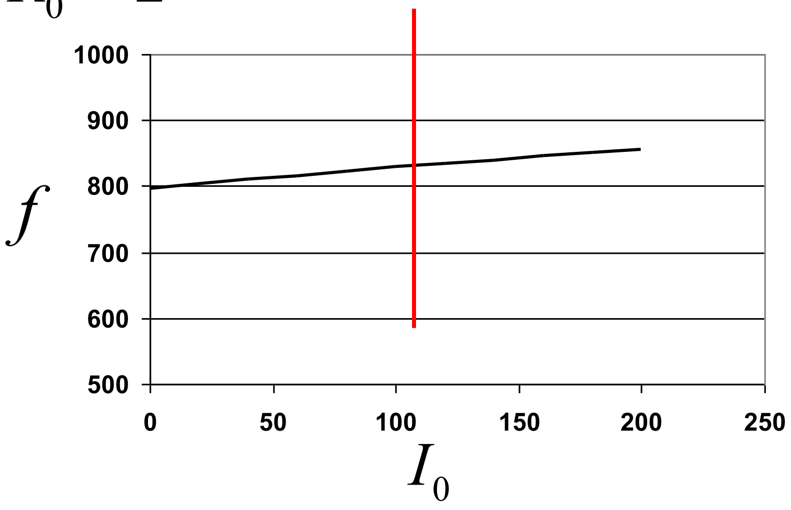
I_0	R_0	f
1	2	797
10	2	800
100	**2**	**828**
200	2	855
300	2	879

Yerel belirsizlik

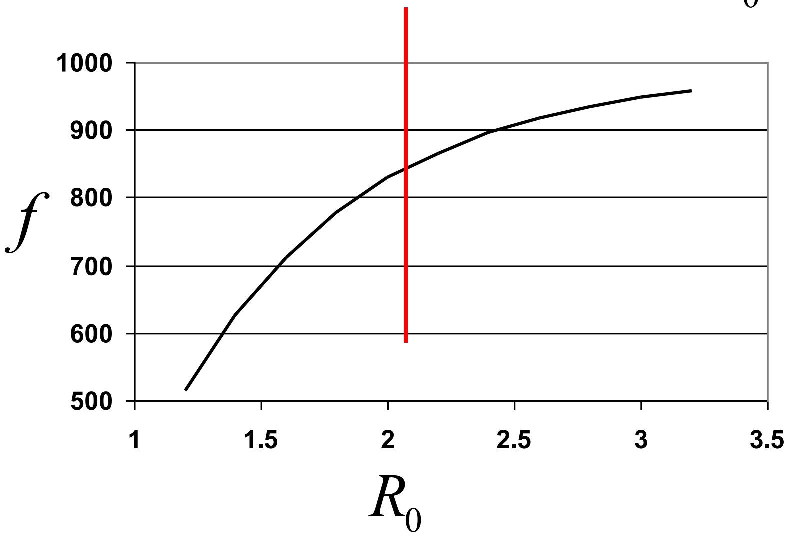
- Parametreler tek tek değiştirildiğinden yerel duyarlılık ilgili sonuçtaki değişime bakar.
- Modelimizde nihai boyut üreme sayısına (R_0) ve başlangıçtaki enfekte bireylere (I_0) bağlıdır.

R_0=2, I_0=100 olduğundaki duyarlılığa bakabiliriz.

$R_0 = 2$



$I_0 = 100$



Farklı parametrelerin *birleşerek* sonucu nasıl etkilediğini bize göstermez.

Latin Hiperküp Örneklemesi

•Değerlendirme masraflıysa tüm değerleri gözden geçirmeyi istemeyin.

•LHC, bölgeyi 'kapsayan' noktaların alt kümesini seçme yöntemidir.

Önceki örnekten:

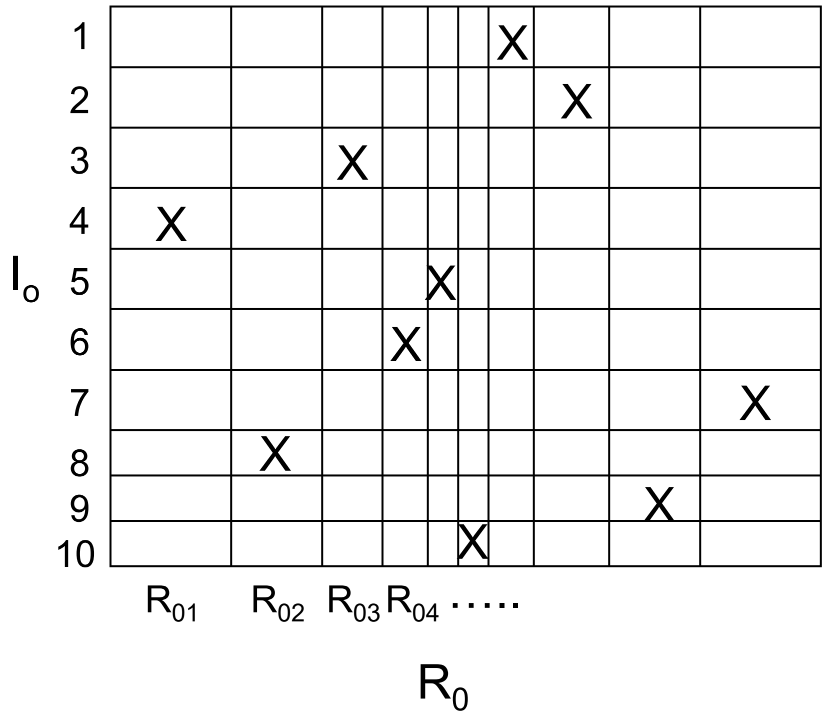
Boyut artıyor

$$(I_{01})$$
$$(I_{02})$$
$$(I_{03})$$
$$(I_{04})$$
$$(I_{05})$$
$$\ldots$$

Bir R_0 seçin

Her bir I_0 için

$$(I_{01}, R_{07})$$
$$(I_{02}, R_{08})$$
$$(I_{03}, R_{03})$$
$$(I_{04}, R_{01})$$
$$(I_{05}, R_{05})$$
$$\ldots$$



10X10 küp: ortalama = 780, ss. = 86

•Her bir parametre değeri yalnızca bir kez görünür.
•Daha fazla set elde etmek için işlemi tekrarlayabilir.

Monte Carlo Markov Zinciri

•MCMC dağılımdan gelmiş gibi görünen parametre setleri dizisi üretmenin bir yoludur.
•Kural gereği adı üzerinde 'zincir' olduğundan her eni set önceki setten üretilir.

$$(x_1, x_2)_n \Rightarrow (x_1, x_2)_{n+1}$$

<u>Metropolis Algoritması</u>

Öncelikle, rastgele bir şekilde eskisinden yeni bir parametre seti seçmenin yolunu oluşturur. (simetrik olmalıdır. Ör. her birine normal sapma eklenir).

•Son noktadan (x_1,x_2) yeni bir nokta oluşturur (x_1^*, x_2^*)

•A'yı hesaplar, burada $A = p(x_1^*, x_2^*) / p(x_2, x_2)$

•[0,1] arasında rastgele bir sayı (u) alır.

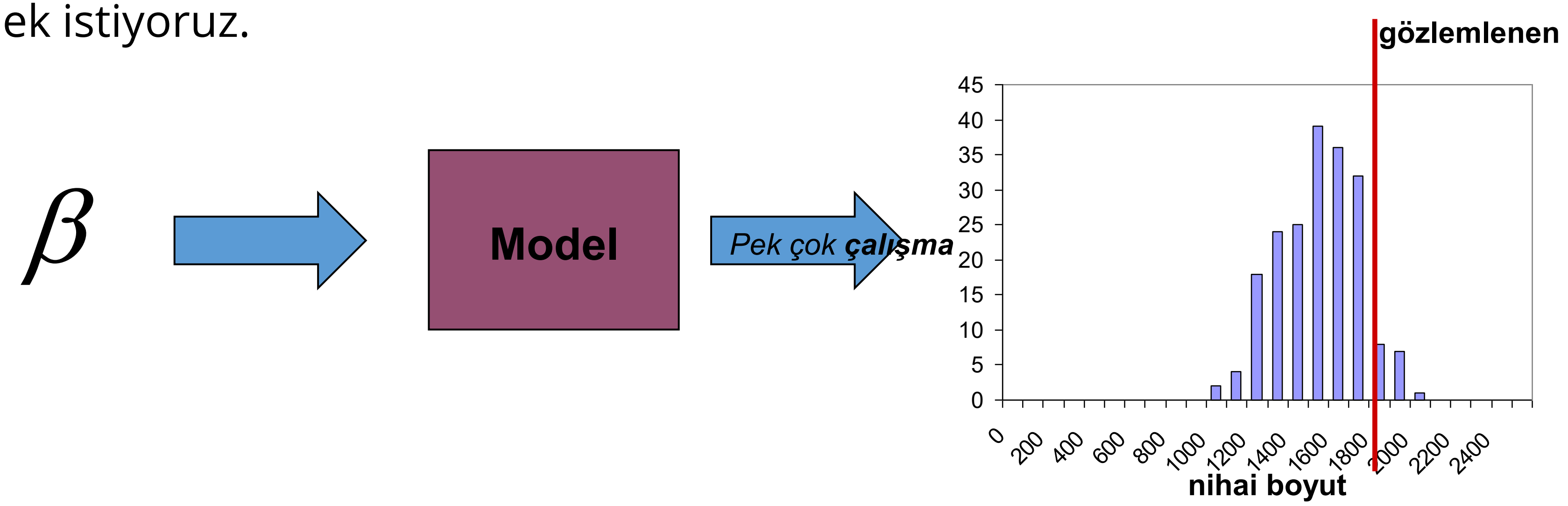
•$u<A$ ise yeni noktayı kabul eder. Değilse eskisini yeniden ku

Not: ardışık noktalar bağımsız *<u>değildir</u>*!

Stokastik ve olasılık modelleri: olabilirlik

•Stokastik modeller, temsil ettikleri süreçlerin rastgeleliğini içerir,
ör. enfeksiyon, iyileşme, ölüm, vb.
•Bu yüzden modelin her çalışması aynı parametrelerden farklı sonuçlar verecektir.

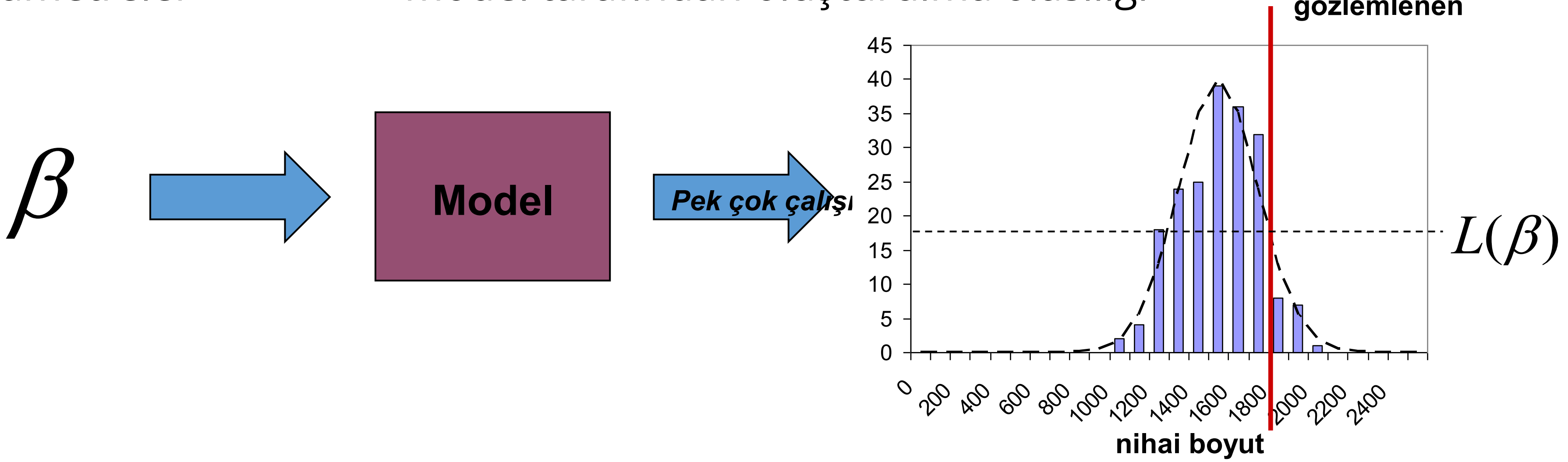
Bir salgın modelini düşünün: β değerlerinin gözlemlenen nihai salgın boyutunu verdiğini bilmek istiyoruz.

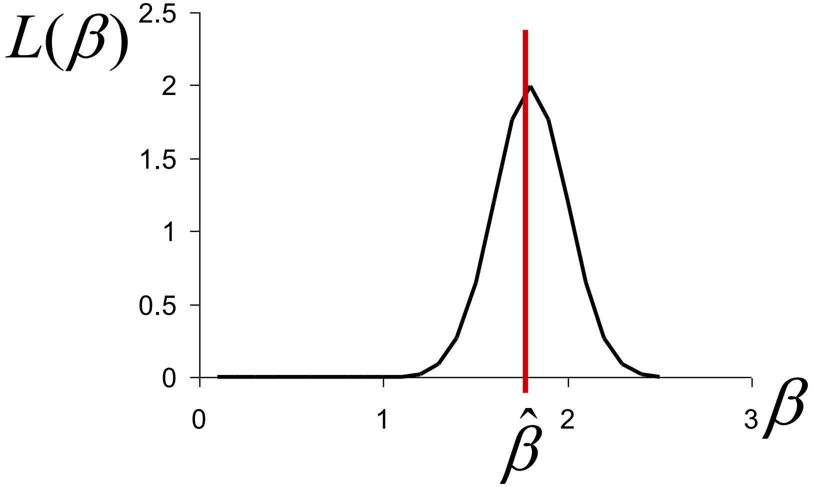


Sorular:
•β için 'en iyi' değer nedir?
•Hangi β değeri aralığı kabul edilebilir?

Olabilirlik

Verilerin olabilirliği, verilen parametreler = Gözlemlenen verilerin parametreler göz önünde bulundurularak model tarafından oluşturulma olasılığı





β'lerin aralığı için tekrar ederek,
β'ye karşılık olasılık
grafiği oluşturabiliriz.

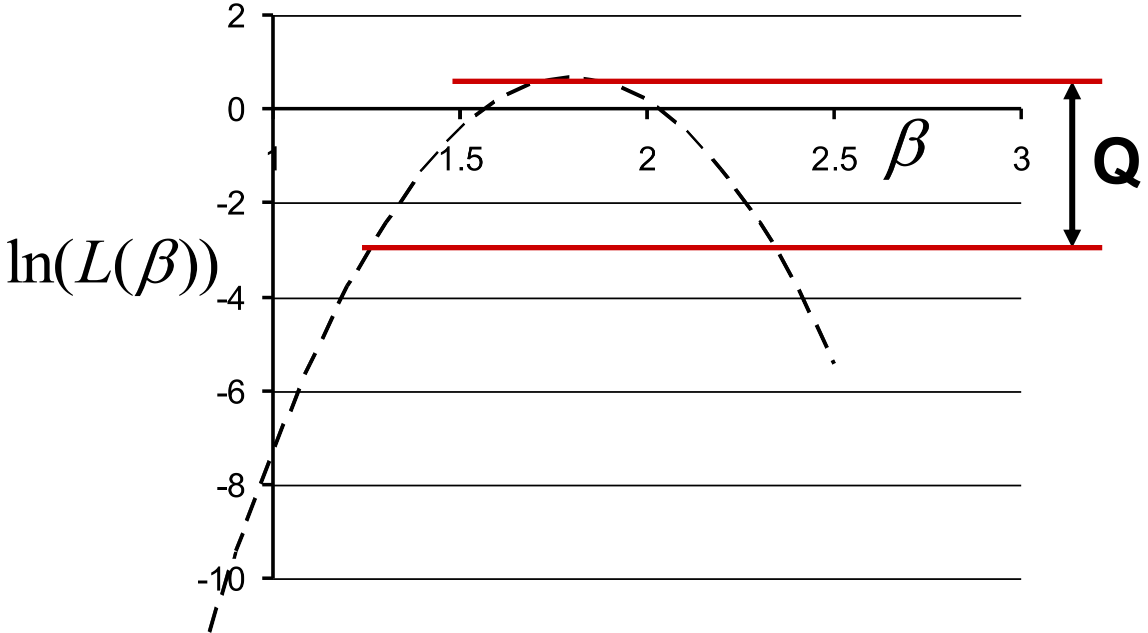
Pik değer 'en iyi' değeri gösterir, buna maksimum olabilirlik tahmini (MLE) denir.

Log-Olabilirlik ve güven aralıkları

Olabilirlik eğrisinin log'unun şekli bize uygun b değerlerinin aralığını verir.
Eğrinin tepesinden Q'ya kadar ölçersek

$$Q = \frac{1}{2}\chi^2_{0.05}(n)$$

Bu, parametre değerleri için %95 güven değerini tanımlar.



Verilerden kaynaklanan belirsizlik

Sıtma antijeni için seropozitif oranını tahmin etmek için bir popülasyondan örnek almayı düşünün. Şunu varsayın:

Örneklenen bireyler = **N**

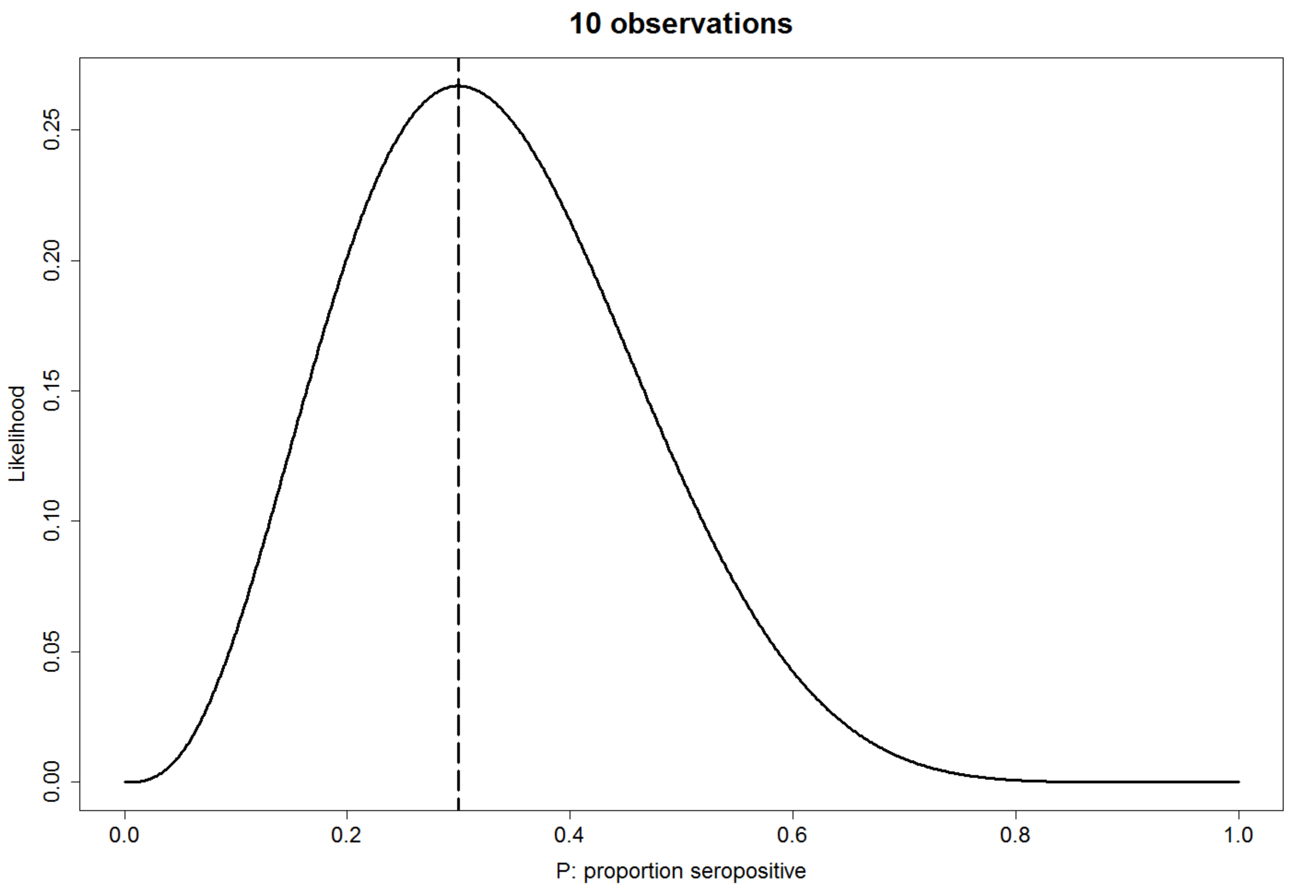
Seropozitif bireyler = **k**

p seropozitif oranını tahmin etme ihtiyacı.

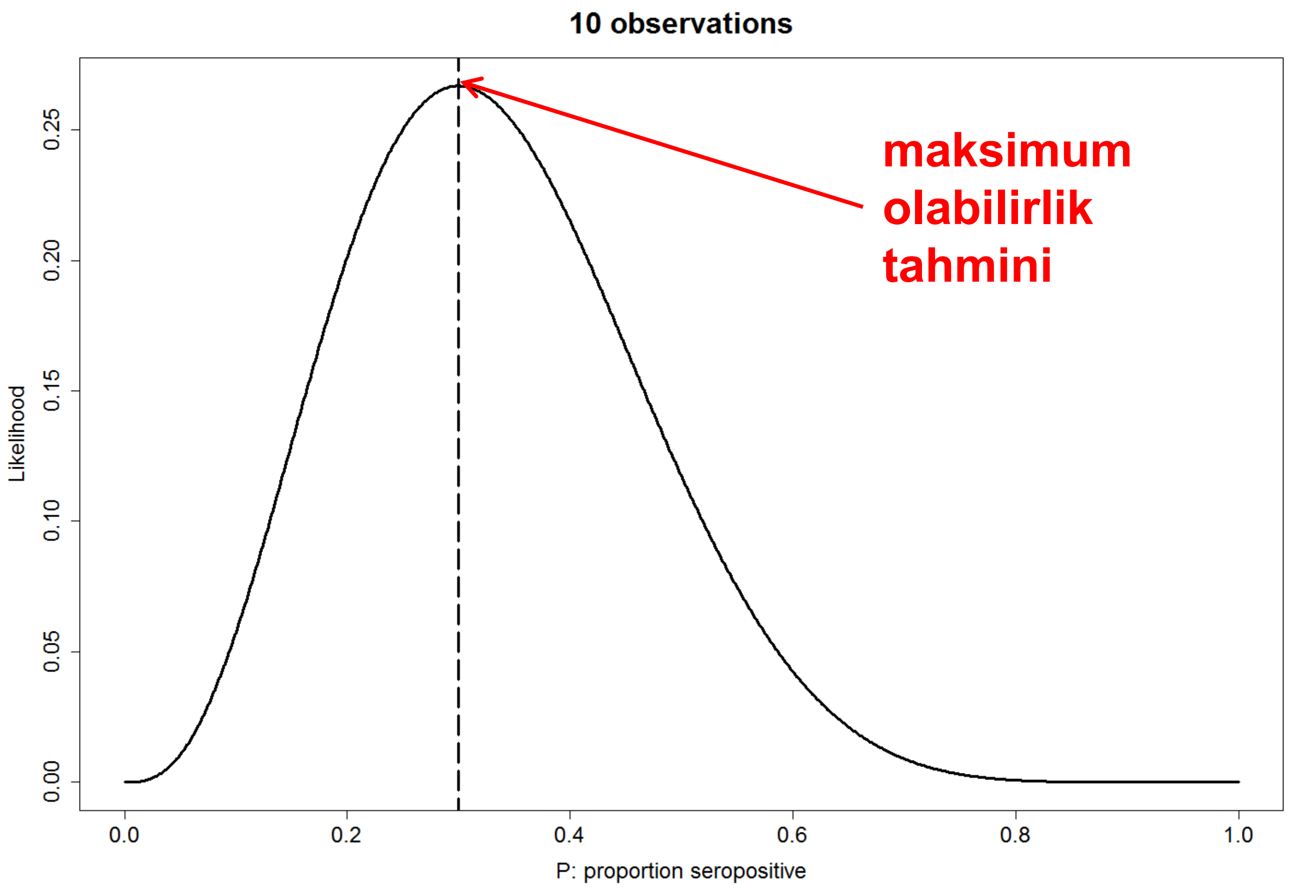
Binom olabilirlik fonksiyonu kullanırız:

$$L = \binom{N}{k} p^k \left(1 - p\right)^{N-k}$$

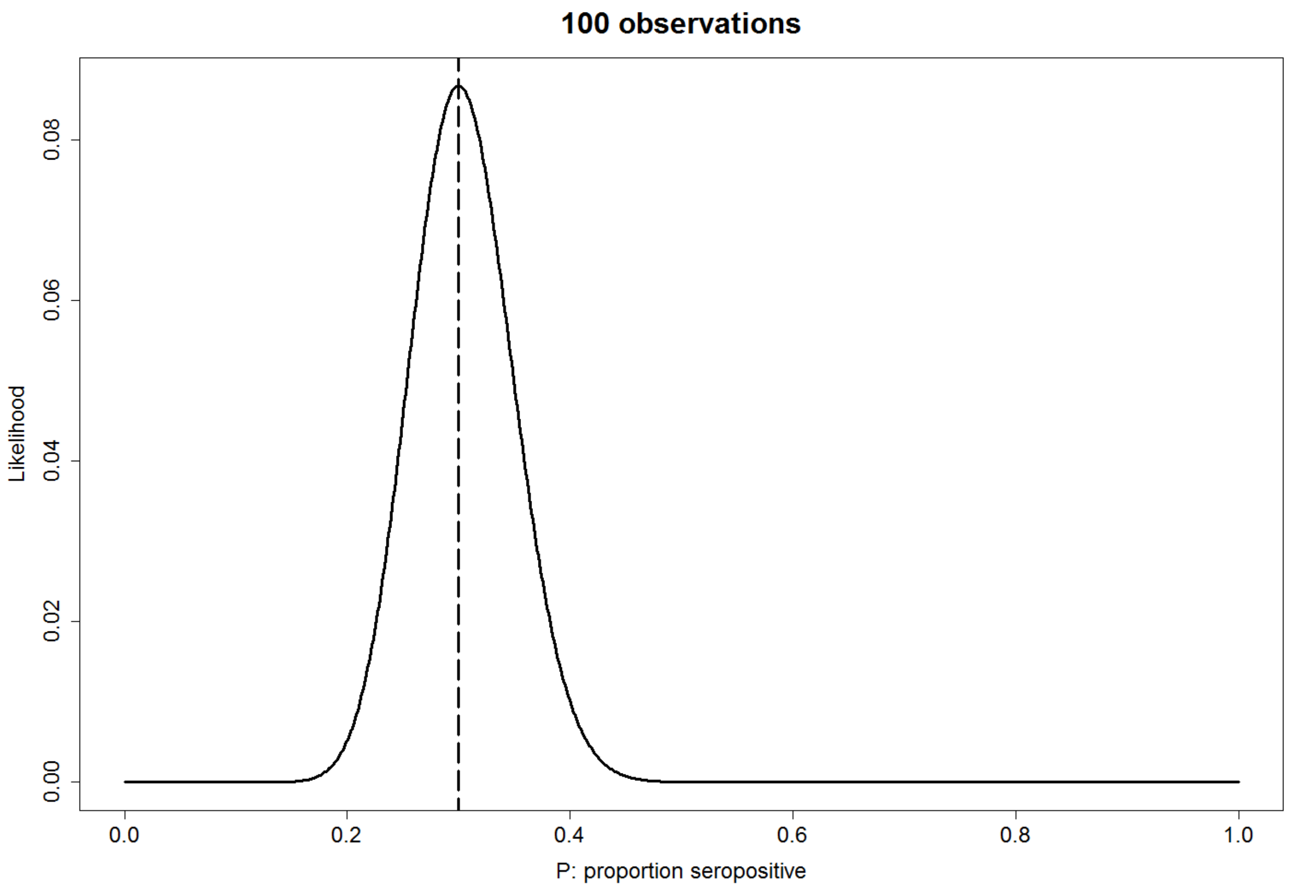
Verilerden kaynaklanan belirsizlik



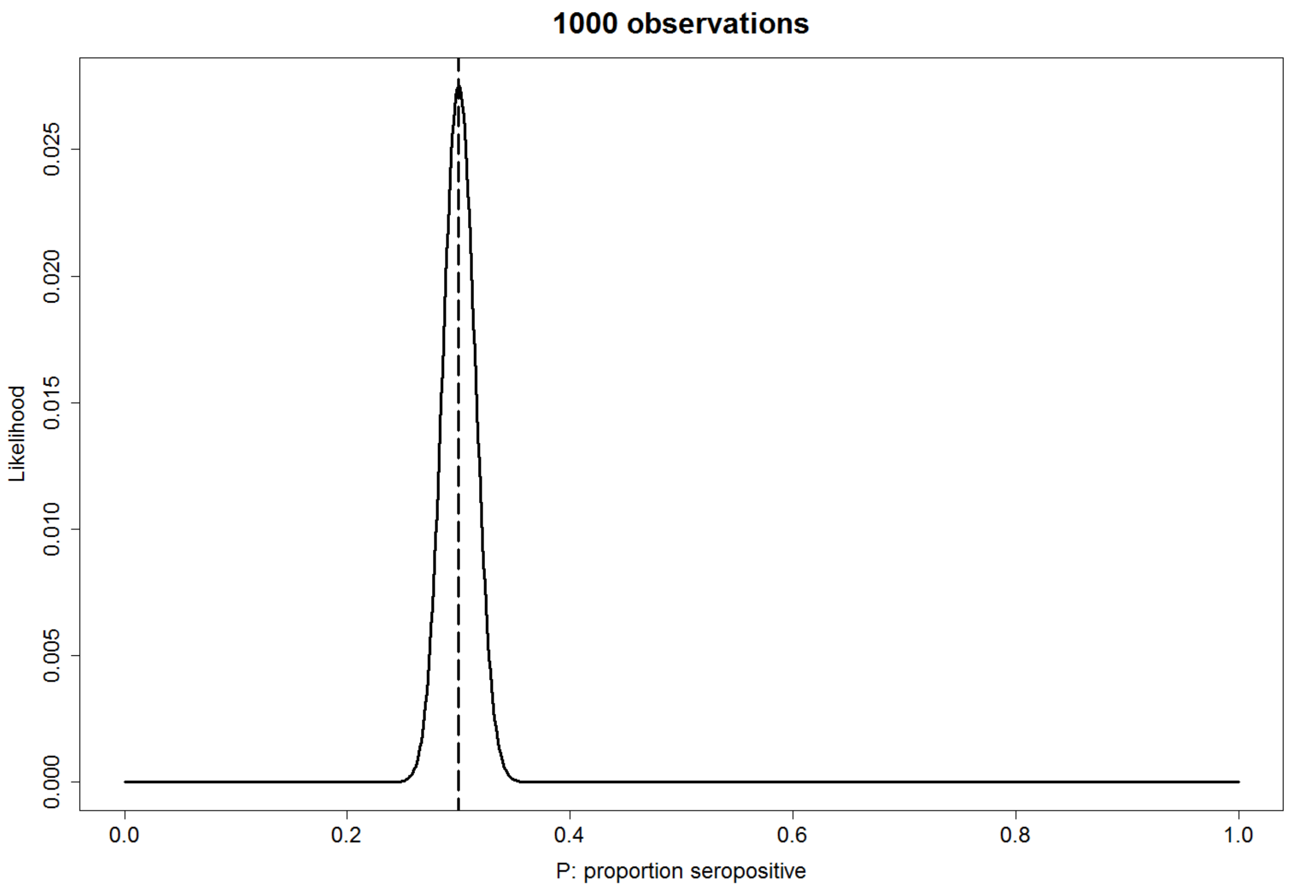
Verilerden kaynaklanan belirsizlik



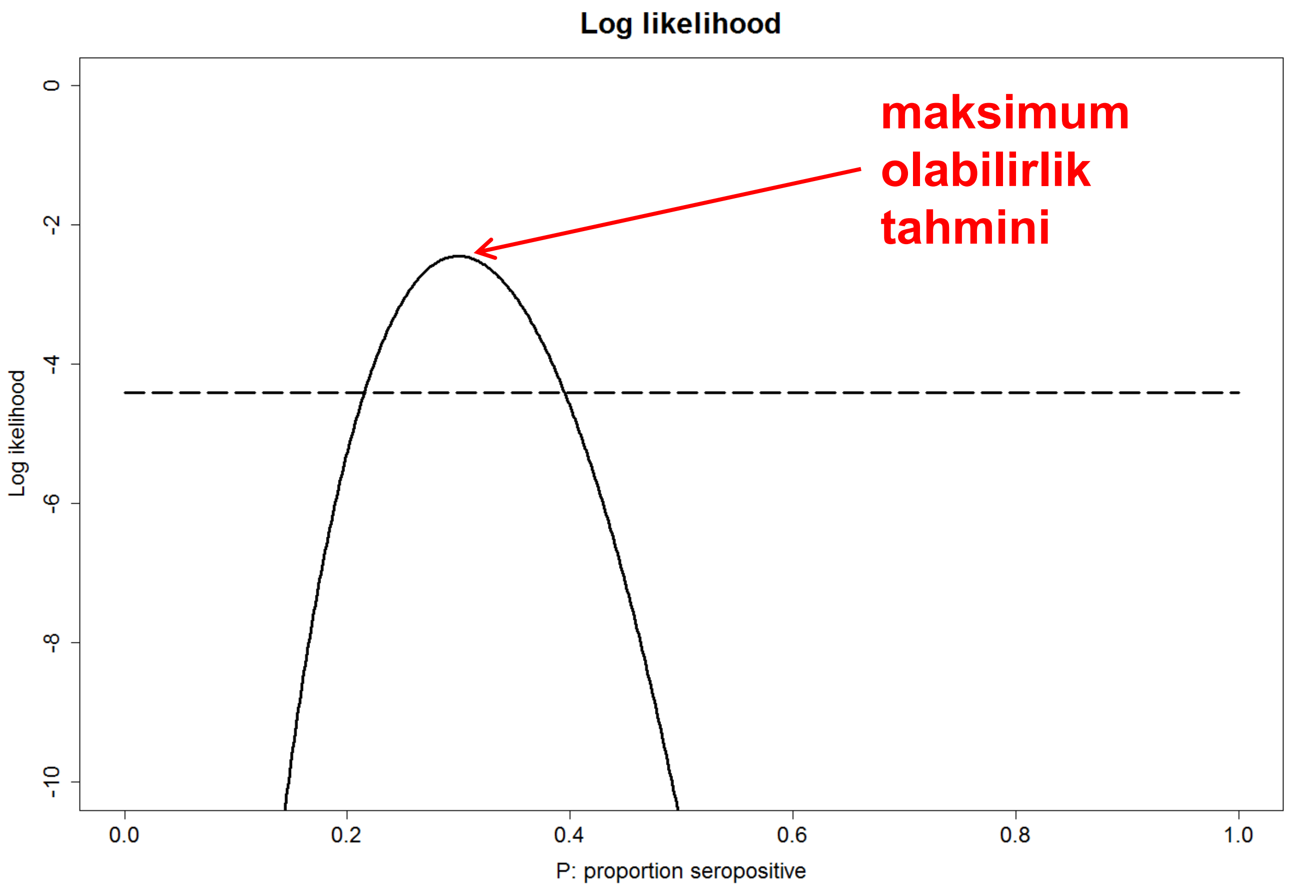
Verilerden kaynaklanan belirsizlik



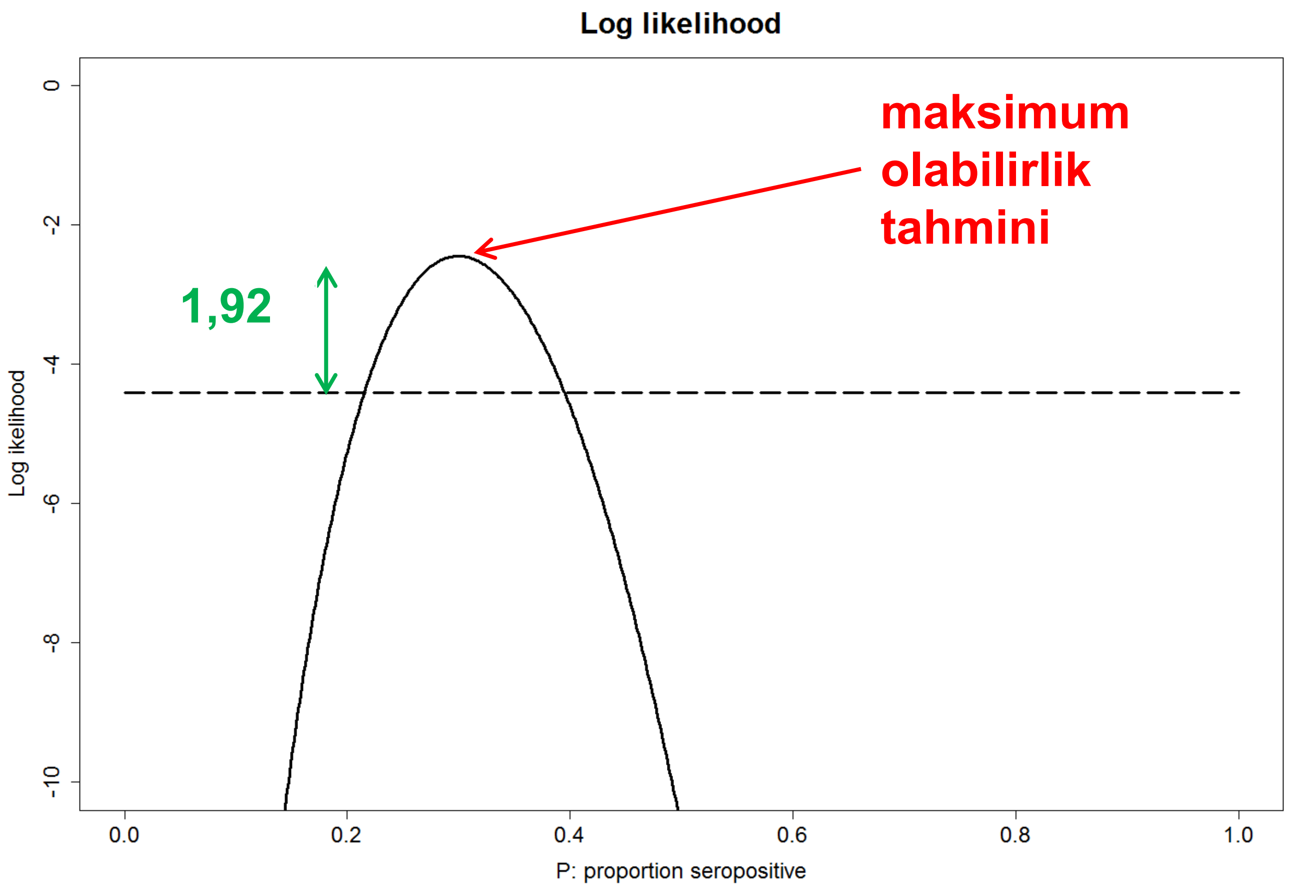
Verilerden kaynaklanan belirsizlik



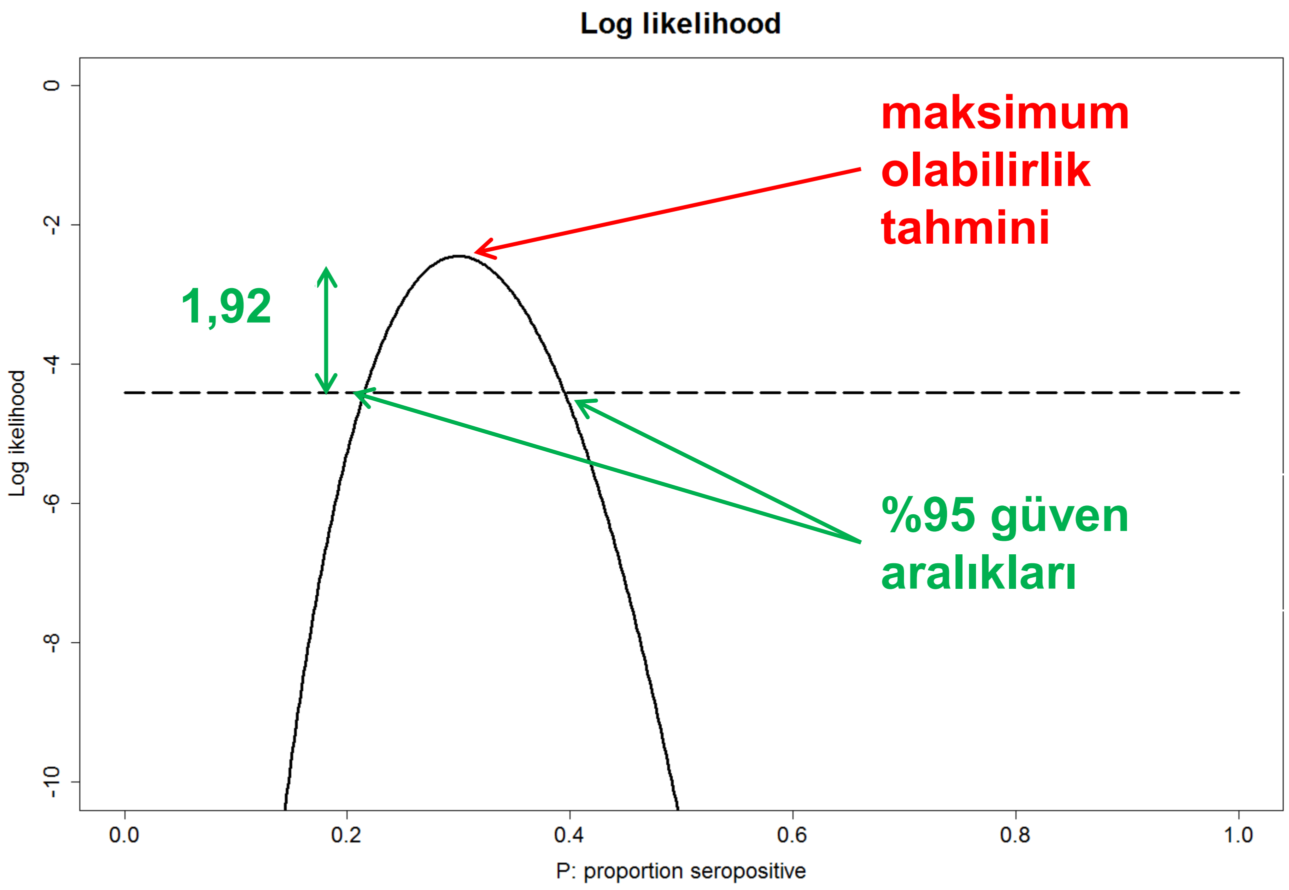
Verilerden kaynaklanan belirsizlik



Verilerden kaynaklanan belirsizlik



Verilerden kaynaklanan belirsizlik



Verilerden kaynaklanan belirsizlik

Genel olarak:

- daha fazla veri => daha az belirsizlik
- daha fazla veri => daha dar güven aralıkları

Model seçimine bağlı olabilir, örneğin model kötü seçilmişse verilerin çok olması da bir işe yaramayacaktır.

Özet

- Olası parametre değerlerini bulmanın yolları incelendi
- Sonuç değişkenlerinin dağılım ve duyarlılığı tahmin edildi.
- Analiz için parametre seti oluşturmak amacıyla kullanılan bir diğer yöntem
- Monte Carlo Markov Zincir algoritmalarıdır. Çok basit ve yaygın kullanılan sağlam bir algoritmadır.
- Olabilirlik yöntemleri, gözlemlenen verilere karşılık gelen parametre aralıklarını bulmak için oldukça faydalıdır.
- Veri setleri karmaşık hale geldikçe (ör. insidans eğrileri) olabilirliğin hesaplanması da zorlaşmaktadır.
- Elinizde ne kadar çok veri olursa parametre tahmininiz de o kadar doğru olur.
- Model iyi değilse en iyi uyumu elde etmeniz mümkündür bu yüzden saçma olduğunu düşünmeyin.