

# Bayesci Yapay Öğrenme



Yapay Öğrenmenin Matematiksel Temelleri,  
Nesin Matematik Köyü  
A. Taylan Cemgil, Boğaziçi Üniversitesi

20.09.2016

# Özet

- Giriş
  - Bayes Teoremi,
  - Basit bir Örnek
  - Olasılık Kuramı hatırlatma, olasılık tabloları
  - Bayesci Öğrenme
- Zaman Dizileri
  - Hesaplama Problemleri
  - Saklı Markov Modelleri
- Yaklaşık Çıkarım (Variational Bayes)

# Bayes Kuralı



Thomas Bayes (1702-1761)

Bir  $\lambda$  parametresi hakkında,  $\mathcal{D}$  verisini gördükten **sonraki** bilgimiz veriyi görmeden **önceki** bilgimiz ve verinin bize söylediği bilginin birleşimidir.

$$p(\lambda|\mathcal{D}) = \frac{p(\mathcal{D}|\lambda)p(\lambda)}{p(\mathcal{D})}$$

$$\text{Sonsal Dağılım} = \frac{\text{Gözlem Modeli} \times \text{Önsel Dağılım}}{\text{Marjinal Olabilirlik}}$$

# İki Zar: ‘Kaynak Ayırıştırma’

1. zar  $\lambda$ , 2. zar  $y$

$$\mathcal{D} = \lambda + y$$

$$\mathcal{D} = 9 \text{ ise } \lambda = ?$$

# İki Zar

$$\mathcal{D} = \lambda + y = 9$$

$\mathcal{D} = \lambda + y$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	2	3	4	5	6	7
$\lambda = 2$	3	4	5	6	7	8
$\lambda = \mathbf{3}$	4	5	6	7	8	<b>9</b>
$\lambda = 4$	5	6	7	8	<b>9</b>	10
$\lambda = \mathbf{5}$	6	7	8	<b>9</b>	10	11
$\lambda = \mathbf{6}$	7	8	<b>9</b>	10	11	12

$$p(\lambda) \rightarrow p(\lambda|\mathcal{D}).$$

Gözlem modeli:  $p(\mathcal{D}|\lambda)$

## “Bürokratik” türetim

$$p(\lambda) = \mathcal{C}(\lambda; [1/6 \ 1/6 \ 1/6 \ 1/6 \ 1/6 \ 1/6])$$

$$p(y) = \mathcal{C}(y; [1/6 \ 1/6 \ 1/6 \ 1/6 \ 1/6 \ 1/6])$$

$$p(\mathcal{D}|\lambda, y) = \delta(\mathcal{D} - (\lambda + y))$$

$$p(\lambda, y|\mathcal{D}) = \frac{1}{p(\mathcal{D})} \times p(\mathcal{D}|\lambda, y) \times p(y)p(\lambda)$$

$$\text{Sonsal} = \frac{1}{\text{Kanıt}} \times \text{Olabilirlik} \times \text{Önsel}$$

$$\text{Kronecker delta } \delta(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$$

# Önsel Dağılım

$$p(y)p(\lambda)$$

$p(y) \times p(\lambda)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	1/36	1/36	1/36	1/36	1/36	1/36
$\lambda = 2$	1/36	1/36	1/36	1/36	1/36	1/36
$\lambda = 3$	1/36	1/36	1/36	1/36	1/36	1/36
$\lambda = 4$	1/36	1/36	1/36	1/36	1/36	1/36
$\lambda = 5$	1/36	1/36	1/36	1/36	1/36	1/36
$\lambda = 6$	1/36	1/36	1/36	1/36	1/36	1/36

- Olasılık  $p(\lambda, y)$

# Olabilirlik Fonksyonu - Gözlem modeli

$$p(\mathcal{D} = 9 | \lambda, y)$$

$p(\mathcal{D} = 9   \lambda, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0
$\lambda = 3$	0	0	0	0	0	<b>1</b>
$\lambda = 4$	0	0	0	0	<b>1</b>	0
$\lambda = 5$	0	0	0	<b>1</b>	0	0
$\lambda = 6$	0	0	<b>1</b>	0	0	0

- Olabilirlik  $\neq$  Olasılık. Sadece negatif olmayan bir fonksyon.



## Olabilirlik $\times$ Önsel

$$\phi_{\mathcal{D}}(\lambda, y) = p(\mathcal{D} = 9|\lambda, y)p(\lambda)p(y)$$

$p(\mathcal{D} = 9 \lambda, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0
$\lambda = 3$	0	0	0	0	0	<b>1/36</b>
$\lambda = 4$	0	0	0	0	<b>1/36</b>	0
$\lambda = 5$	0	0	0	<b>1/36</b>	0	0
$\lambda = 6$	0	0	<b>1/36</b>	0	0	0

## Marjinal Olabilirlik

$$\begin{aligned} p(\mathcal{D} = 9) &= \sum_{\lambda, y} p(\mathcal{D} = 9 | \lambda, y) p(\lambda) p(y) \\ &= 0 + 0 + \dots + 1/36 + 1/36 + 1/36 + 1/36 + 0 + \dots + 0 \\ &= 1/9 \end{aligned}$$

$p(\mathcal{D} = 9   \lambda, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0
$\lambda = 3$	0	0	0	0	0	<b>1/36</b>
$\lambda = 4$	0	0	0	0	<b>1/36</b>	0
$\lambda = 5$	0	0	0	<b>1/36</b>	0	0
$\lambda = 6$	0	0	<b>1/36</b>	0	0	0

## Sonsal Dağılım

$$p(\lambda, y | \mathcal{D} = 9) = \frac{1}{p(\mathcal{D})} p(\mathcal{D} = 9 | \lambda, y) p(\lambda) p(y)$$

$p(\mathcal{D} = 9   \lambda, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0
$\lambda = 3$	0	0	0	0	0	<b>1/4</b>
$\lambda = 4$	0	0	0	0	<b>1/4</b>	0
$\lambda = 5$	0	0	0	<b>1/4</b>	0	0
$\lambda = 6$	0	0	<b>1/4</b>	0	0	0

$$1/4 = (1/36)/(1/9)$$

# Marjinal Sonsal Dağılım

$$p(\lambda|\mathcal{D}) = \sum_y \frac{1}{p(\mathcal{D})} p(\mathcal{D}|\lambda, y) p(\lambda) p(y)$$

	$p(\lambda \mathcal{D} = 9)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0	0
$\lambda = 3$	<b>1/4</b>	0	0	0	0	0	1/4
$\lambda = 4$	<b>1/4</b>	0	0	0	0	1/4	0
$\lambda = 5$	<b>1/4</b>	0	0	0	1/4	0	0
$\lambda = 6$	<b>1/4</b>	0	0	1/4	0	0	0

## Orantılıdır $\propto$ notasyonu

$$p(\lambda|\mathcal{D} = 9) \propto p(\lambda, \mathcal{D} = 9) = \sum_y p(\mathcal{D} = 9|\lambda, y)p(\lambda)p(y)$$

	$p(\lambda, \mathcal{D} = 9)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$
$\lambda = 1$	0	0	0	0	0	0	0
$\lambda = 2$	0	0	0	0	0	0	0
$\lambda = 3$	1/36	0	0	0	0	0	<b>1/36</b>
$\lambda = 4$	1/36	0	0	0	0	<b>1/36</b>	0
$\lambda = 5$	1/36	0	0	0	<b>1/36</b>	0	0
$\lambda = 6$	1/36	0	0	<b>1/36</b>	0	0	0

## Model Seçim Örneği

Bilinmeyen sayıda zar atılıyor:  $\lambda_1, \lambda_2, \dots, \lambda_n$ ,

$$\mathcal{D} = \sum_{i=1}^n \lambda_i$$

$\mathcal{D} = 9$  ise kaç zar atıldı?

$$p(n) \propto 1$$

# Model Seçimi

$$p(n|\mathcal{D} = 9) = \frac{p(\mathcal{D} = 9|n)p(n)}{p(\mathcal{D})} \propto p(\mathcal{D} = 9|n)$$

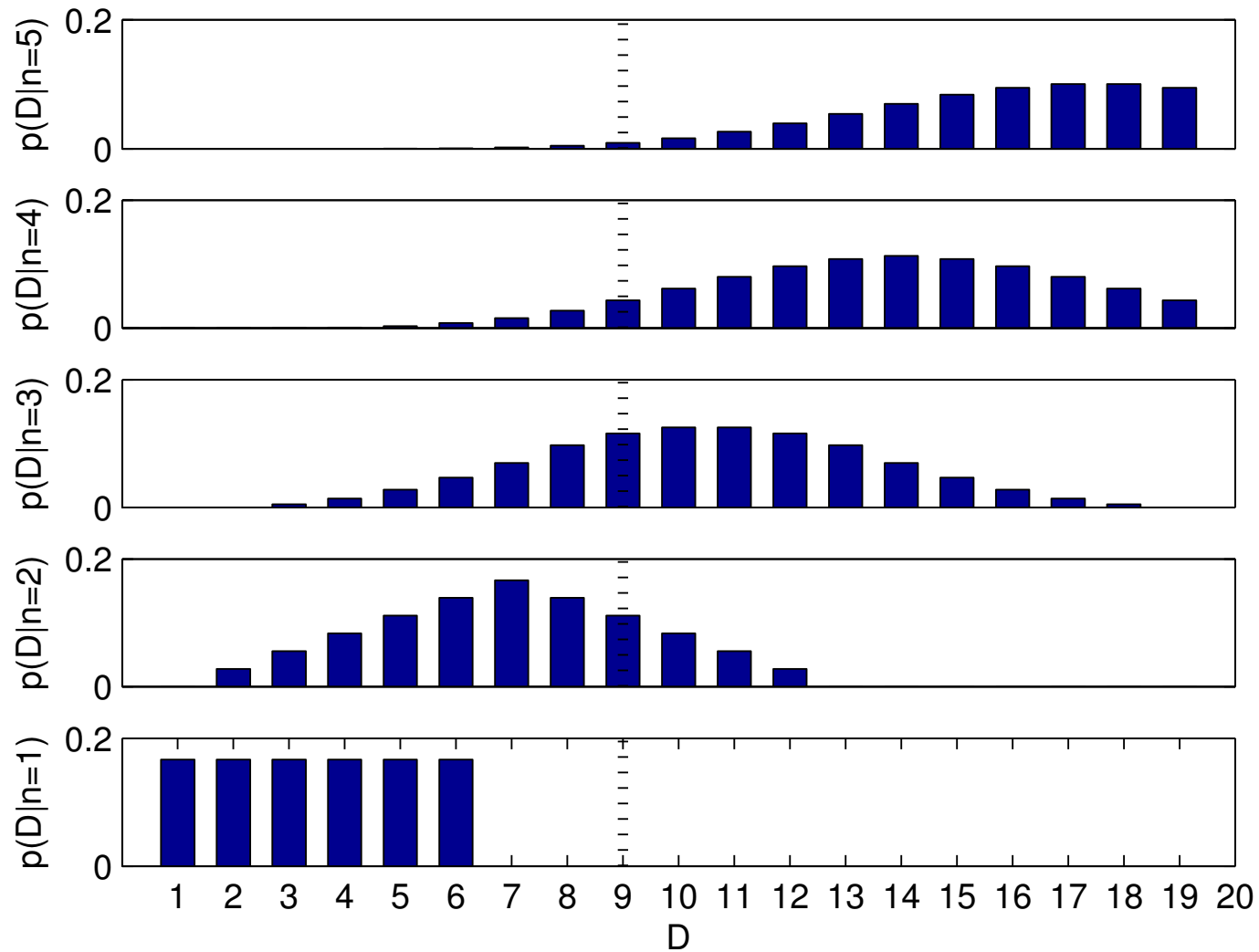
$$p(\mathcal{D}|n = 1) = \sum_{\lambda_1} p(\mathcal{D}|\lambda_1)p(\lambda_1)$$

$$p(\mathcal{D}|n = 2) = \sum_{\lambda_1} \sum_{\lambda_2} p(\mathcal{D}|\lambda_1, \lambda_2)p(\lambda_1)p(\lambda_2)$$

...

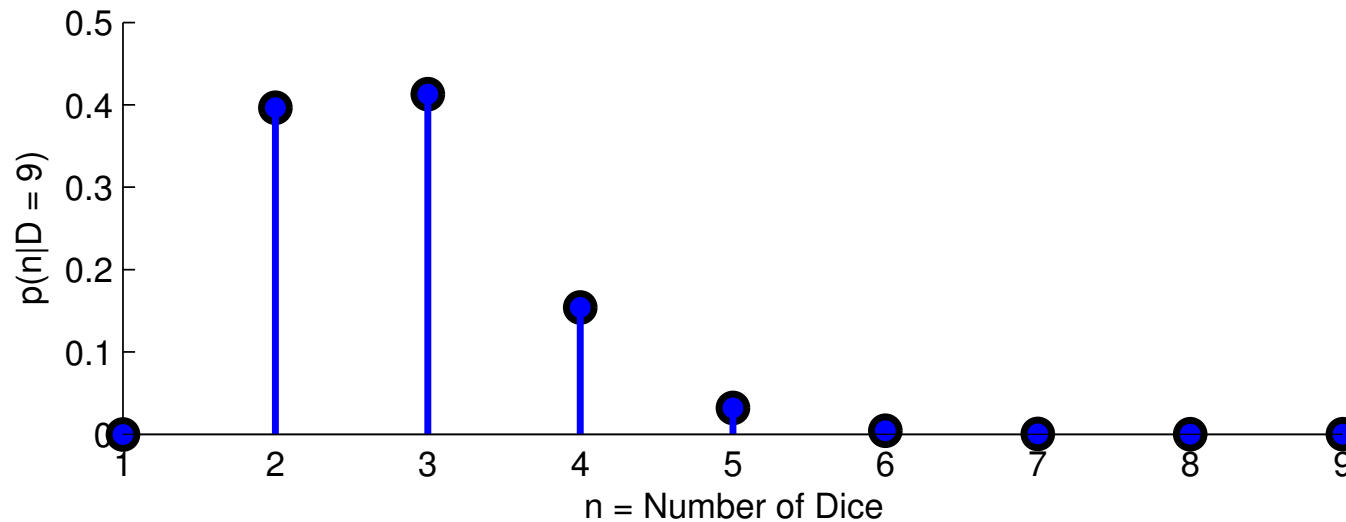
$$p(\mathcal{D}|n = n') = \sum_{\lambda_1, \dots, \lambda_{n'}} p(\mathcal{D}|\lambda_1, \dots, \lambda_{n'}) \prod_{i=1}^{n'} p(\lambda_i)$$

$$p(\mathcal{D}|n) = \sum_{\lambda} p(\mathcal{D}|\lambda, n)p(\lambda|n)$$





# Model Seçimi



- Sezgi: Karmaşık modellerde olasılık daha büyük bir alana yayılır, gözlemlenen tek bir olayın olabilirliği düşer.
- Bayesci çıkarım “basit modelleri” tercih eder – Occam’s razor
- Bütün parametreler üzerinden toplam (tümlev) hesabı

# Olasılıksal Yaklaşım

- Ne çözelim : Modelleme
  - Zanaat
- Nasıl çözelim : Çıkarım Algoritması
  - Mekanik-Otomatik (Teoride! Pratikte hep değil)
  - Genel

# Olasılık Kuramı

- Pascal ve Fermat arasındaki mektuplaşma (Soylu ve kumarbaz bey de Meré)
- 1930'lar Aksiyomatik gelişim (Reichenbach, Kolmogorov), Ölçüm (measure) Kuramı
- İstatistik: Ters olasılık – Olasılığın anlamı:
  - “Frequentist”: Tekrarlanabilir deneylerdeki frekanslar
    - \* Bu ilaç etkili.
  - “Bayesian”: Bilginin (inancın) derecesi
    - \* Yarın yüzde doksan yağmurlu.
- Brad Efron, *Modern science and the Bayesian-frequentist controversy* , 2005  
<http://www-stat.stanford.edu/~ckirby/brad/papers/2005NEWModernScience.pdf>
- Brad Efron, *Bayesians, frequentists, and scientists* , 2005  
<http://www-stat.stanford.edu/~ckirby/brad/papers/2005BayesFreqSci.pdf>

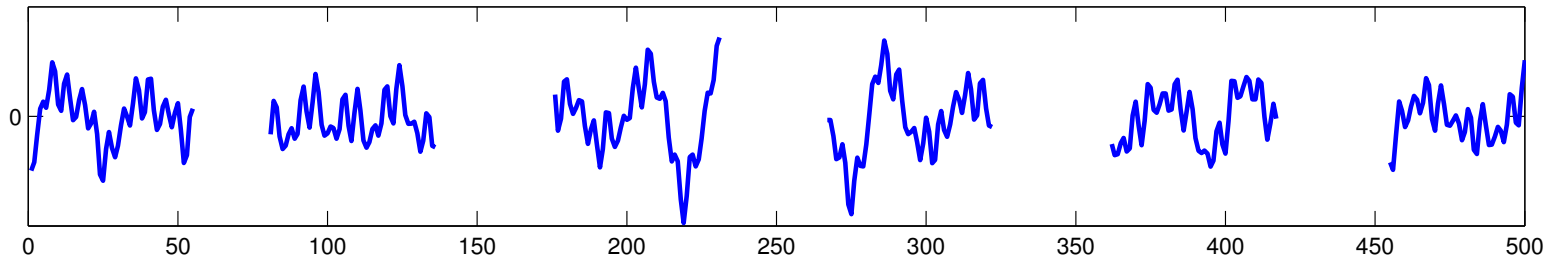
# Tümdengelim (Deduction) ve Tümevarım (Inductive)

- $a$ ,  $b$ , and  $c$  tam sayılar olmak üzere

$$a^n + b^n = c^n$$

denkleminin  $n > 2$  için çözümü yoktur.

- Aşağıda verilen ses dalgası içerisindeki kayıp örnekleri bulunuz



# Tümevarım'ın tehlikeleri

Borovik

$$\text{snc}(x) \equiv \sin(x)/x$$

$$\int_0^\infty \text{snc}(x) dx = \pi/2$$

$$\int_0^\infty \text{snc}(x) \text{snc}(x/3) dx = \pi/2$$

$$\int_0^\infty \text{snc}(x) \text{snc}(x/3) \text{snc}(x/5) dx = \pi/2$$

$$\int_0^\infty \text{snc}(x) \text{snc}(x/3) \text{snc}(x/5) \text{snc}(x/7) dx = \pi/2$$

$$\int_0^\infty \text{snc}(x) \text{snc}(x/3) \text{snc}(x/5) \text{snc}(x/7) \text{snc}(x/9) dx = \pi/2$$

$$\int_0^\infty \operatorname{snc}(x) \operatorname{snc}(x/3) \operatorname{snc}(x/5) \operatorname{snc}(x/7) \operatorname{snc}(x/9) \operatorname{snc}(x/11) dx = \pi/2$$

$$\int_0^\infty \operatorname{snc}(x) \operatorname{snc}(x/3) \operatorname{snc}(x/5) \operatorname{snc}(x/7) \operatorname{snc}(x/9) \operatorname{snc}(x/11) \operatorname{snc}(x/13) dx = \pi/2$$

$$\int_0^\infty \operatorname{snc}(x) \operatorname{snc}(x/3) \operatorname{snc}(x/5) \operatorname{snc}(x/7) \operatorname{snc}(x/9) \operatorname{snc}(x/11) \operatorname{snc}(x/13) \operatorname{snc}(x/15) dx = \frac{467807924713440738696537864469}{935615849440640907310521750000} \cdot \pi$$

# Uygulamalar

- Ön bilgi ve gözlemlenen verinin birleştirilmesi için doğal bir çerçeve  $\Rightarrow$  Öğrenme
  - Tıbbi tanı (Semptom/Hastalık)
  - Konuşma Tanıma (İşaret/Hece)
  - Bilgisayarla Görme (Görüntü/Nesne)
  - Robotik, Hedef Takibi (Algılayıcı/Pozisyon)
  - Finans (Geçmiş fiyatlar, Piyasa haberleri/Gelecek fiyat)

## Olasılık Tabloları

$p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.3	0.3
$x_1 = 2$	0.1	0.3

- Marjinal:  $p(x_1), p(x_2)$
- Şartlı:  $p(x_1|x_2), p(x_2|x_1)$
- Sonsal:  $p(x_1, x_2 = 2), p(x_1|x_2 = 2)$
- Marjinal olabilirlik:  $p(x_2 = 2)$
- En büyük:  $p(x_1^*) = \max_{x_1} p(x_1|x_2 = 1)$
- Mod:  $x_1^* = \arg \max_{x_1} p(x_1|x_2 = 1)$
- Max-marginal:  $\max_{x_1} p(x_1, x_2)$



# Yanıtlar

$p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.3	0.3
$x_1 = 2$	0.1	0.3

- Marjinaler:

$p(x_1)$	
$x_1 = 1$	0.6
$x_1 = 2$	0.4

$p(x_2)$	$x_2 = 1$	$x_2 = 2$
	0.4	0.6

- Şartlı:

$p(x_1 x_2)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.75	0.5
$x_1 = 2$	0.25	0.5

$p(x_2 x_1)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.5	0.5
$x_1 = 2$	0.25	0.75

## Yanıtlar

$p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.3	0.3
$x_1 = 2$	0.1	0.3

- Sonsal:

$p(x_1, x_2 = 2)$	$x_2 = 2$	$p(x_1 x_2 = 2)$	$x_2 = 2$
$x_1 = 1$	0.3	$x_1 = 1$	0.5
$x_1 = 2$	0.3	$x_1 = 2$	0.5

- Marjinal Olabilirlik (Kanit):

$$p(x_2 = 2) = \sum_{x_1} p(x_1, x_2 = 2) = 0.6$$

## Yanıt

$p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$
$x_1 = 1$	0.3	0.3
$x_1 = 2$	0.1	0.3

- En büyük: (değer)

$$\max_{x_1} p(x_1 | x_2 = 1) = 0.75$$

- Mod: (indis)

$$\operatorname{argmax}_{x_1} p(x_1 | x_2 = 1) = 1$$

- En büyük-marjinal: (silüet)  $\max_{x_1} p(x_1, x_2)$

$\max_{x_1} p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$
	0.3	0.3

# Öğrenme

- En büyük Olabilirlik (Maximum Likelihood),
- Cezalandırılmalı Olabilirlik - En büyük sonsal (Penalized Likelihood),
- Bayesci Öğrenme

# Çıkarım ve Öğrenme

- Veri kümesi

$$\mathcal{D} = \{x_1, \dots, x_N\}$$

- Model parametresi  $\lambda$

$$p(\mathcal{D}|\lambda)$$

- En büyük olabilirlik (ML)

$$\lambda^{\text{ML}} = \arg \max_{\lambda} \log p(\mathcal{D}|\lambda)$$

- Tahmin dağılımı (Predictive Distribution)

$$p(x_{N+1}|\mathcal{D}) \approx p(x_{N+1}|\lambda^{\text{ML}})$$

# Regularizasyon

- Önsel

$$p(\lambda)$$

- En büyük sonsal (Maximum a-posteriori (MAP))

$$\lambda^{\text{MAP}} = \arg \max_{\lambda} \log p(\mathcal{D}|\lambda)p(\lambda)$$

- Tahmin dağılımı

$$p(x_{N+1}|\mathcal{D}) \approx p(x_{N+1}|\lambda^{\text{MAP}})$$

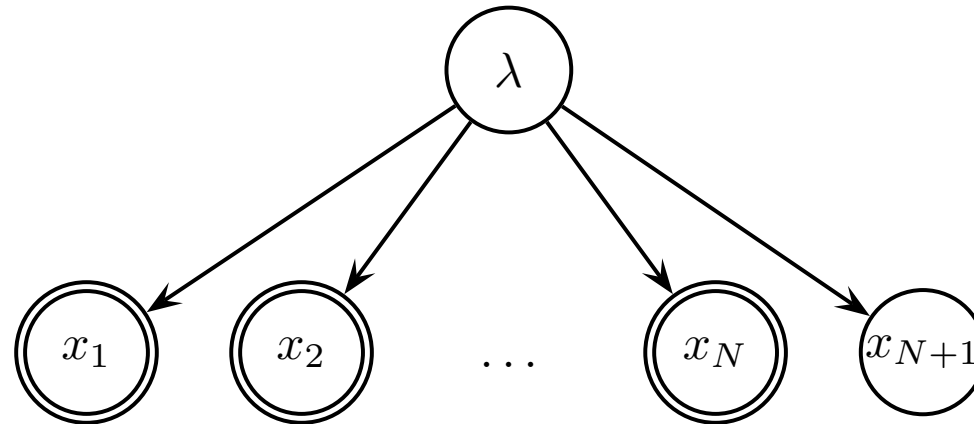
# Bayesci Öğrenme

- Bütün Parametreler rasgele değişkenler ( $\neq$  bilinmeyen deterministik değişken)
- Nokta kestirimleri yerine bilinmeyen parametreler üzerinden tümlev hesabı
  - Doğal olarak adaptasyon
  - Model seçimi

# Bayesci Öğrenme

- Tahmin dağılımı

$$p(x_{N+1}|\mathcal{D}) = \int d\lambda \ p(x_{N+1}|\lambda)p(\lambda|\mathcal{D})$$



- Bayesci öğrenme = Bayesci çıkarım



# Bayesci Öğrenme, $\lambda = p(x = \text{Tura})$

?

# Bayesci Öğrenme

$T, ?$

# Bayesci Öğrenme

T, T, ?

# Bayesci Öğrenme

T, T, T, ?

# Bayesci Öğrenme

T, T, T, T, ?

# Bayesci Öğrenme

T, T, T, T, T, ?

# Bayesci Öğrenme

T, T, T, T, T, Y, ?

# Bayesci Öğrenme

T, T, T, T, T, Y, T, ?

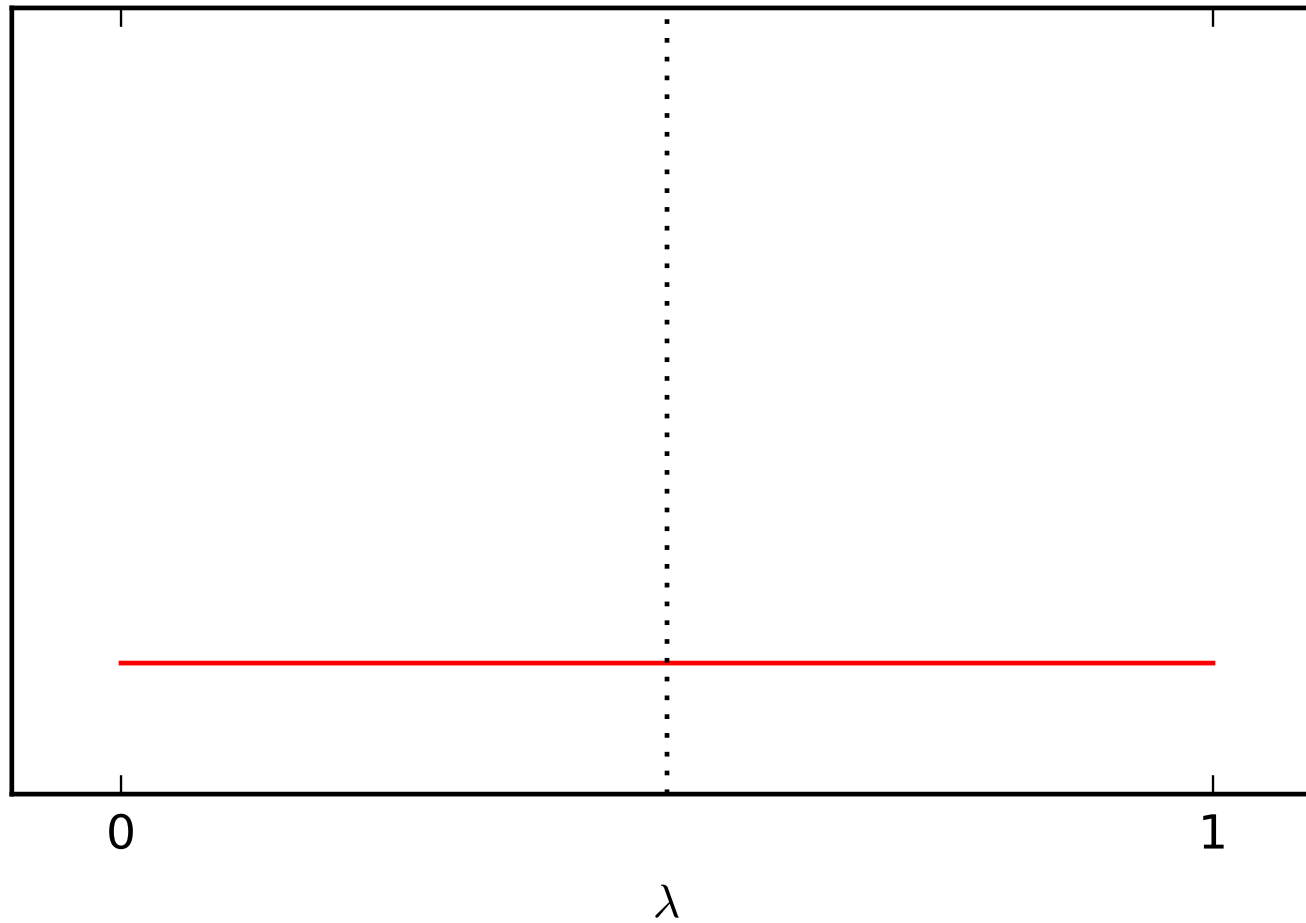


# Bayesci Öğrenme

T, T, T, T, T, Y, T, T, ?

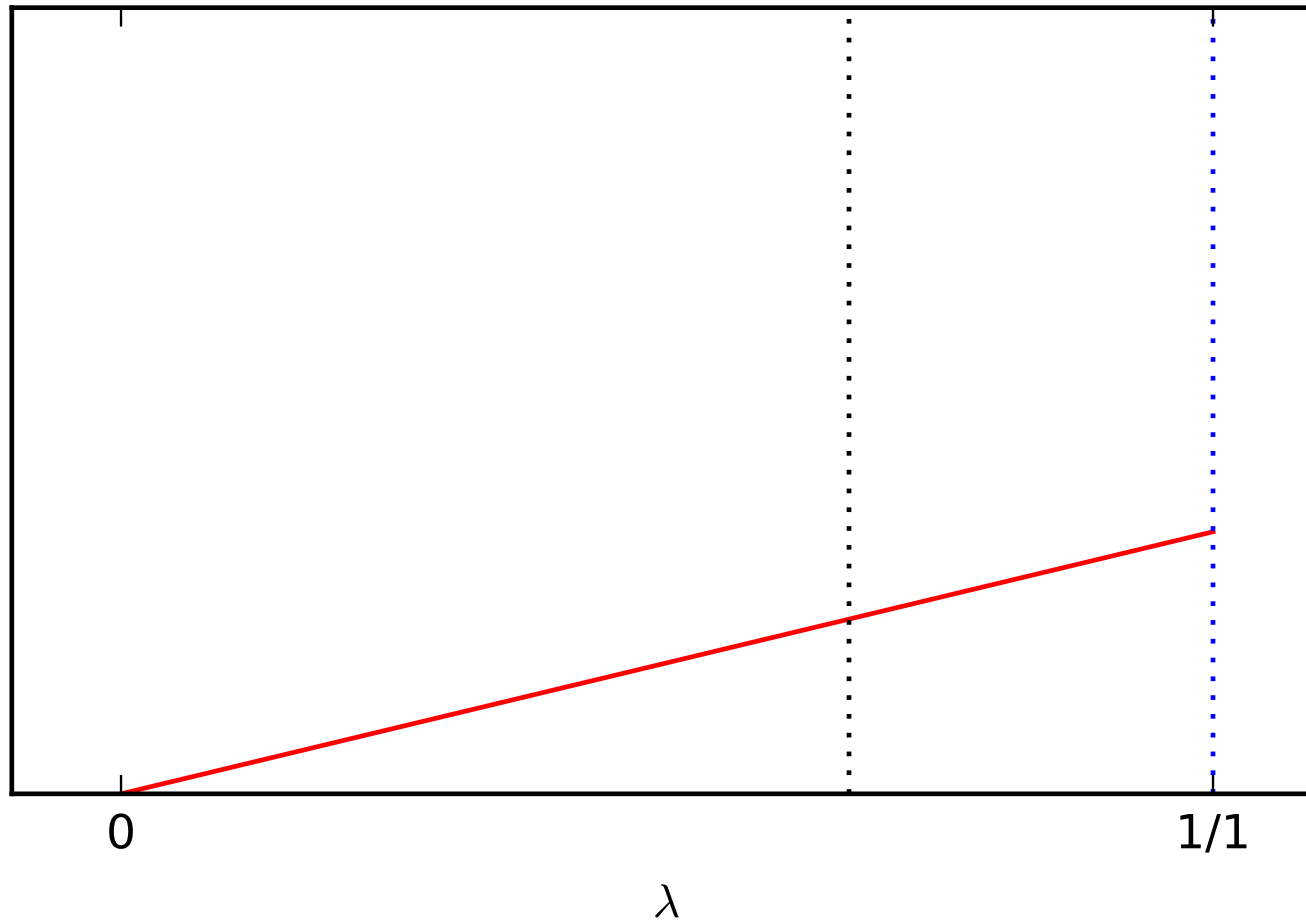
$$p(\lambda)$$

?



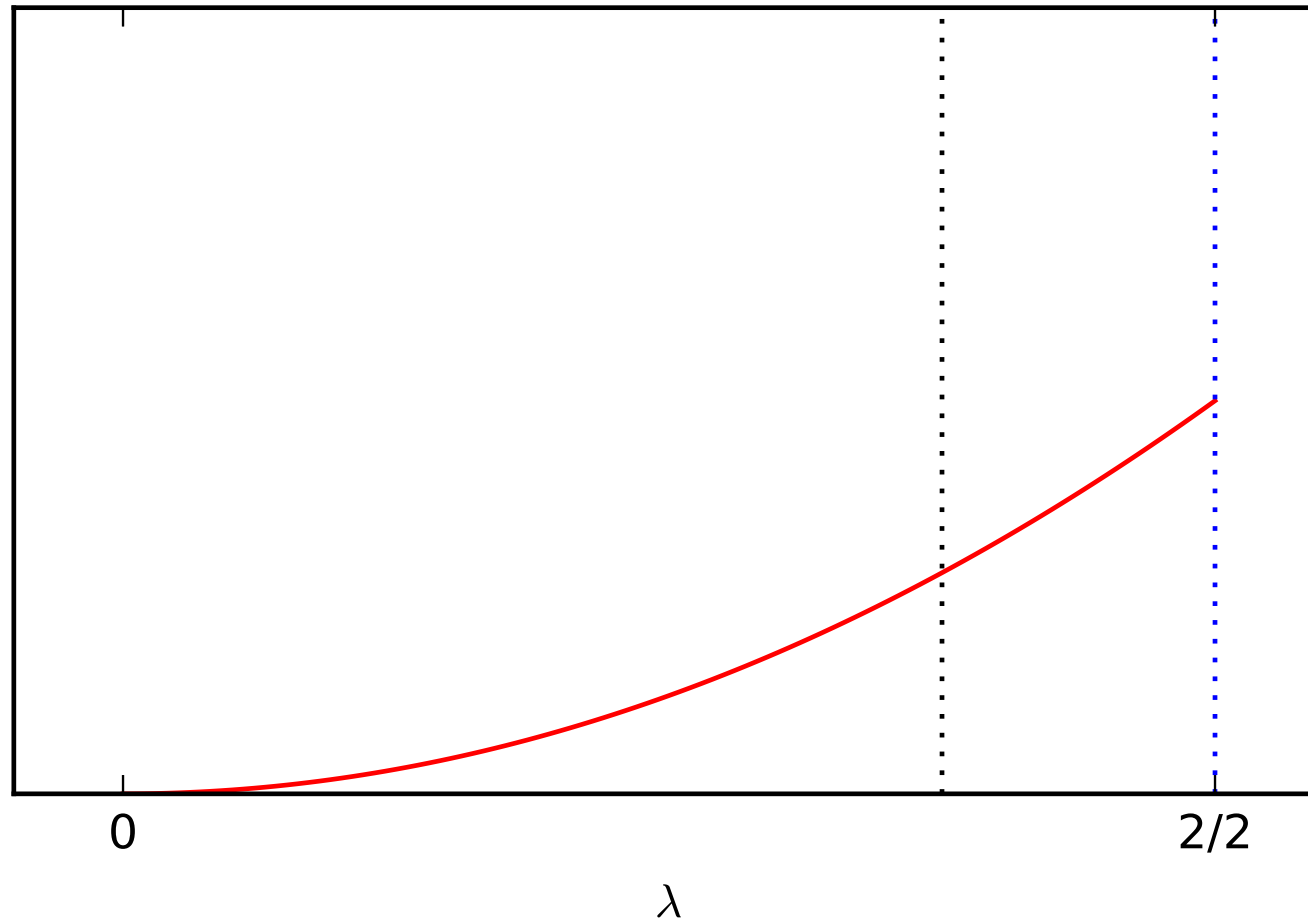
$$p(\lambda|x_1)$$

T, ?



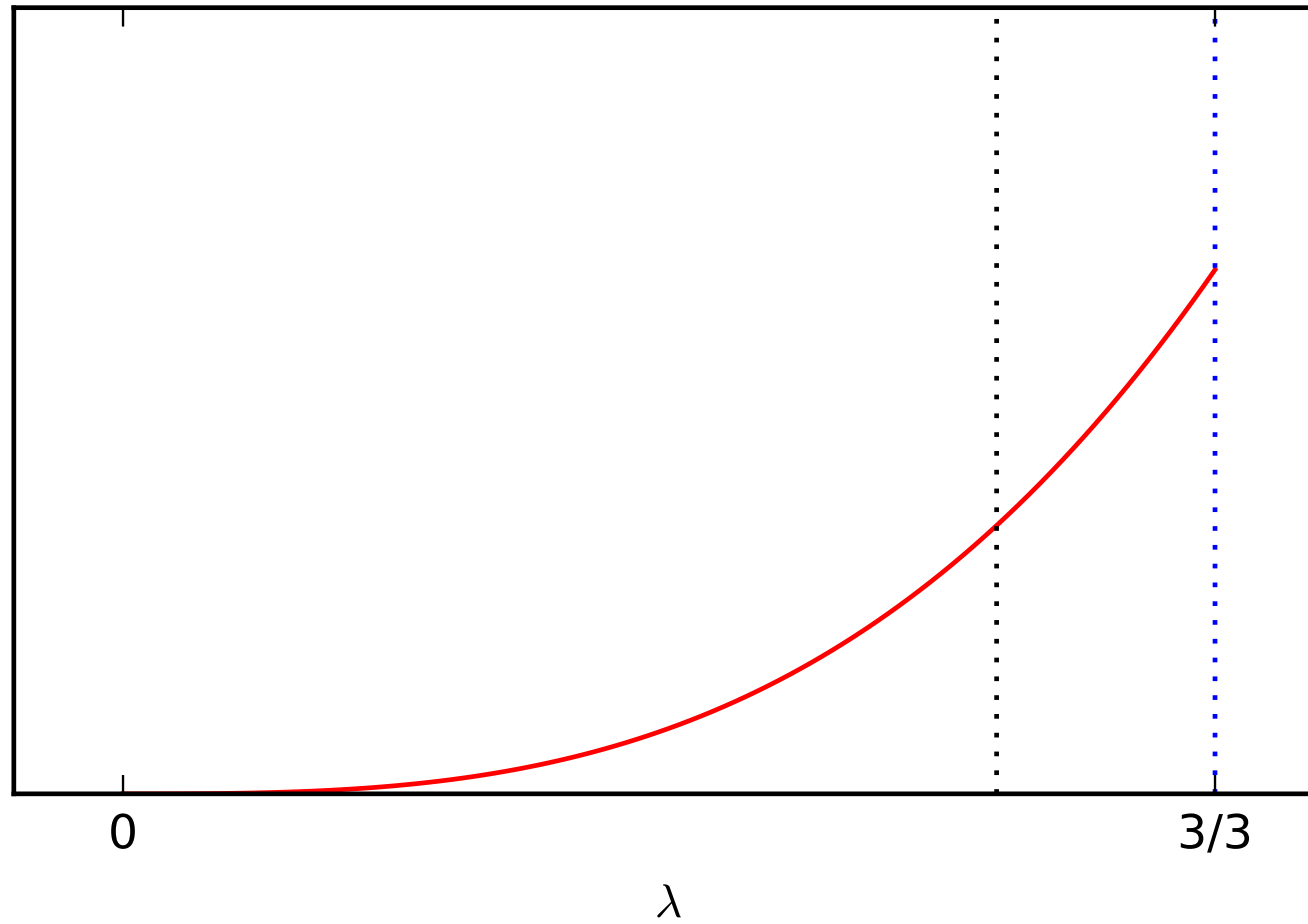
$$p(\lambda|x_1, x_2)$$

T, T, ?



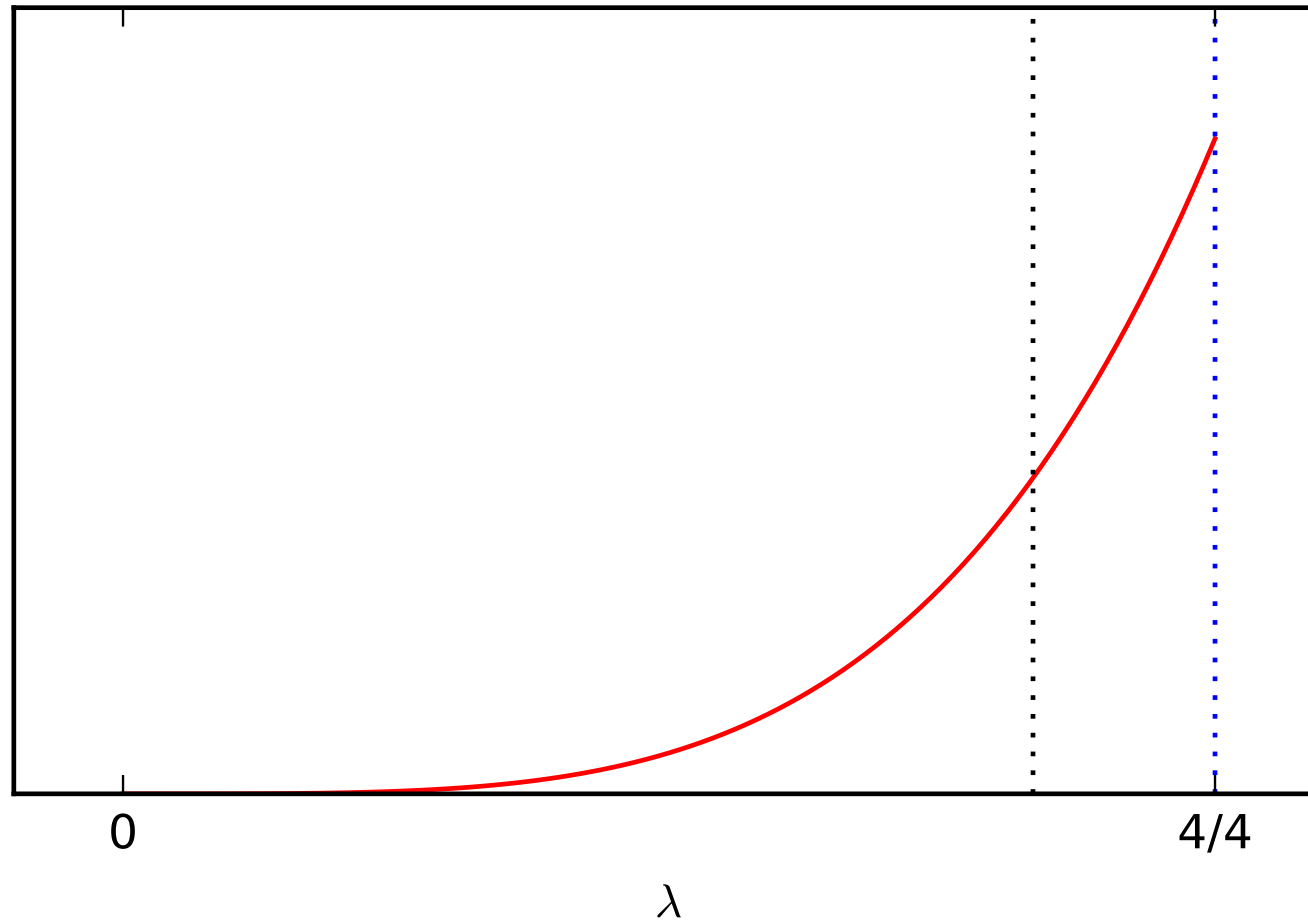
$$p(\lambda|x_{1:3})$$

T, T, T, ?



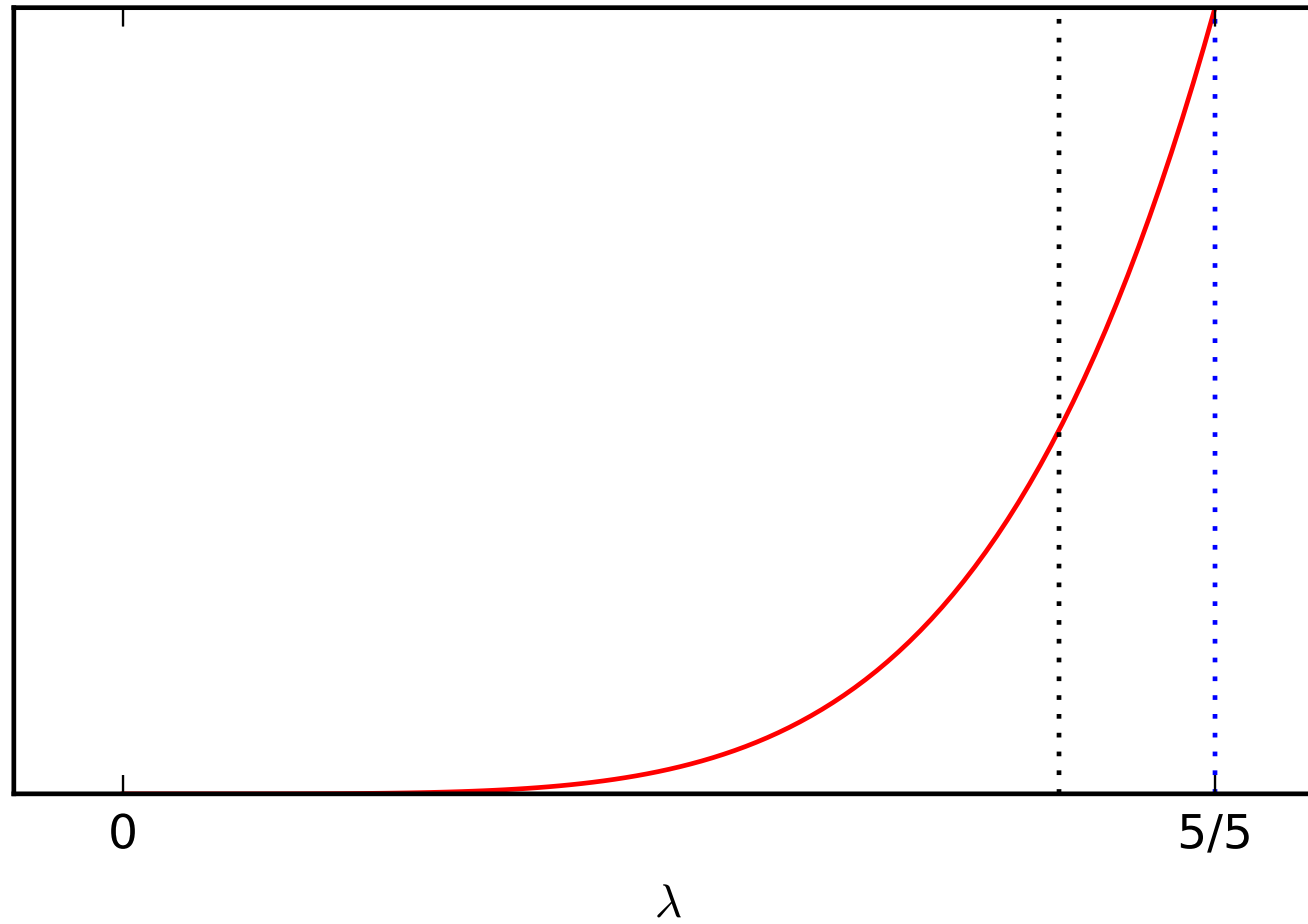
$$p(\lambda|x_{1:4})$$

T, T, T, T, ?



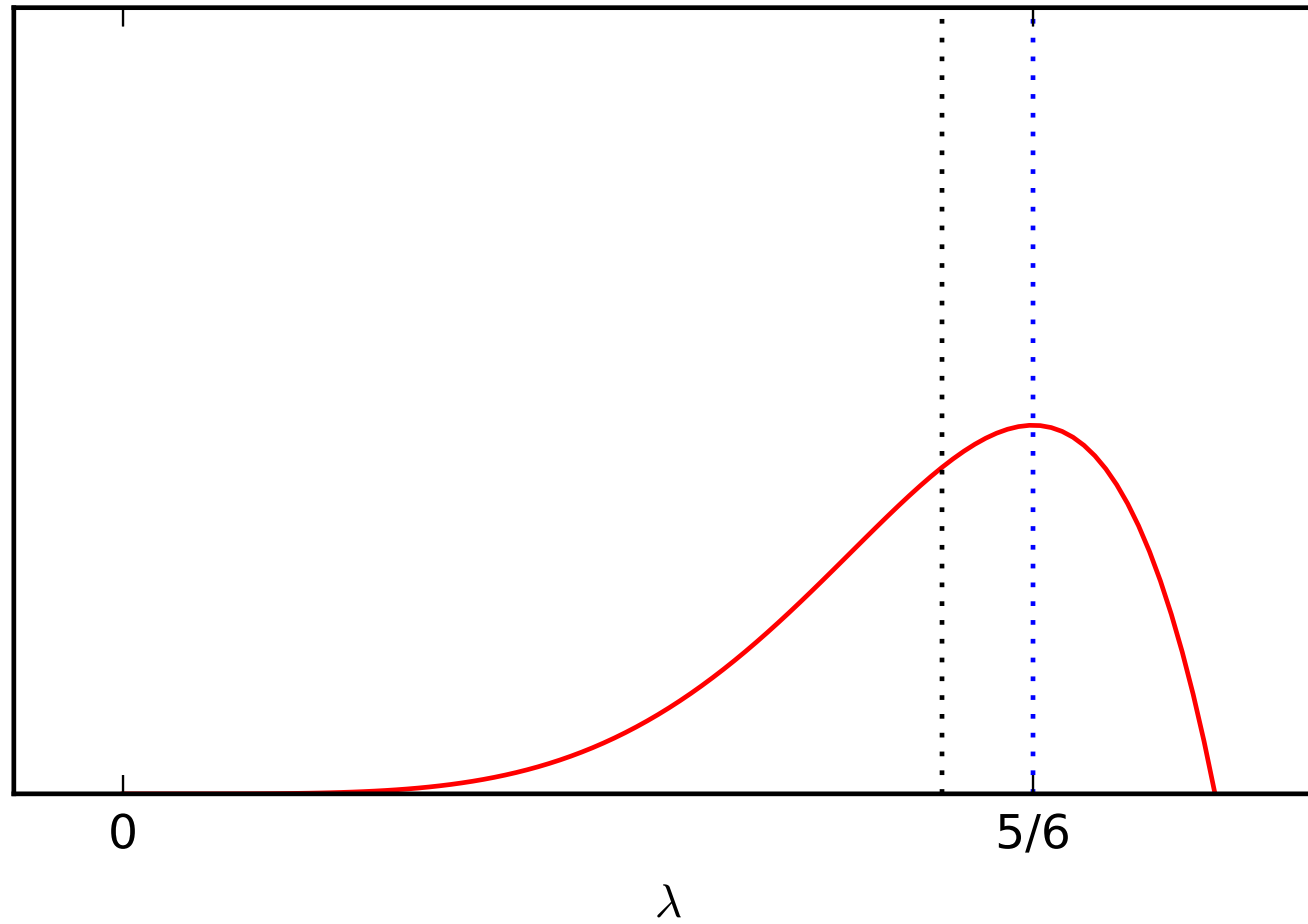
$$p(\lambda|x_{1:5})$$

T, T, T, T, T, ?



$$p(\lambda|x_{1:6})$$

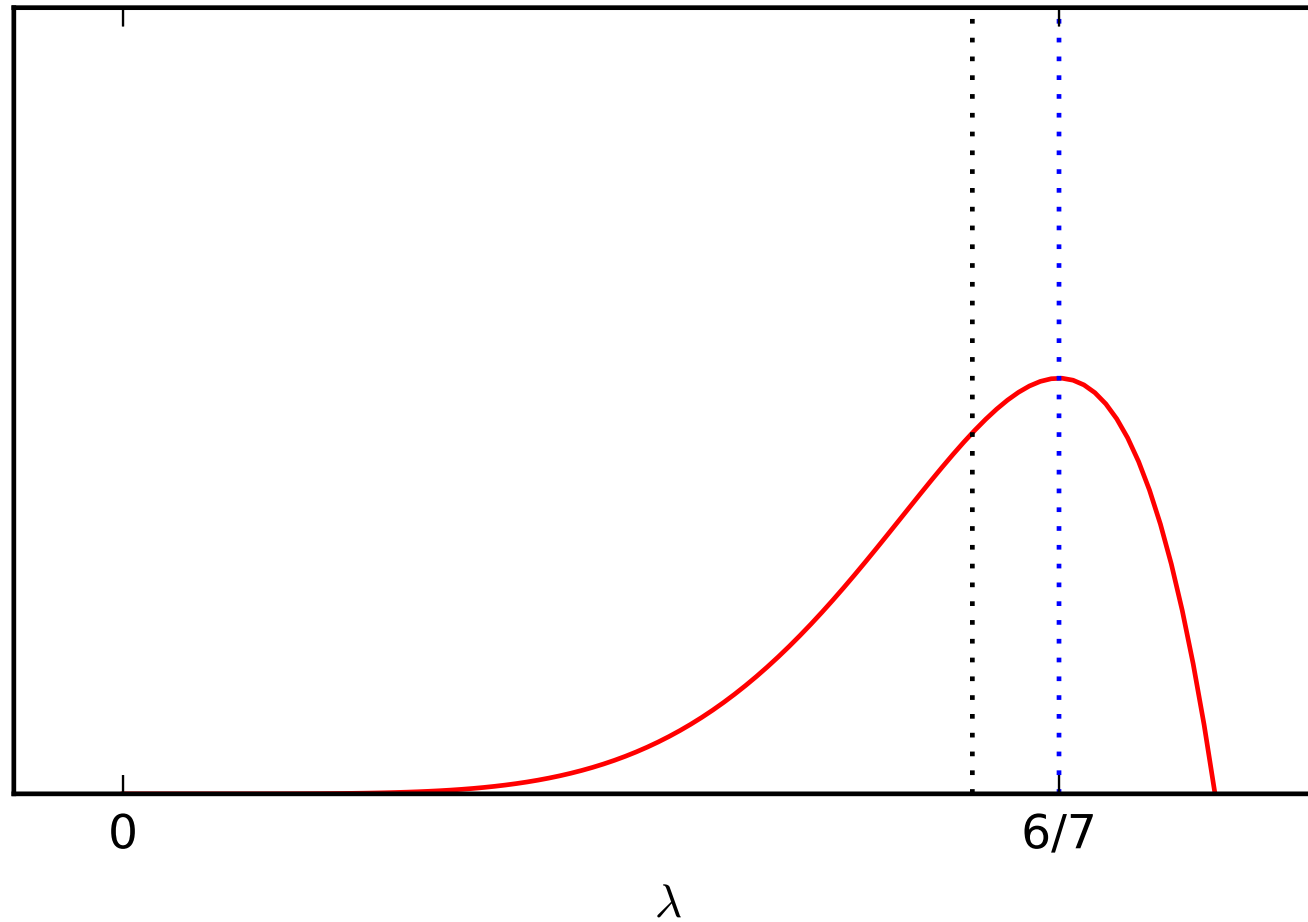
T, T, T, T, T, Y, ?





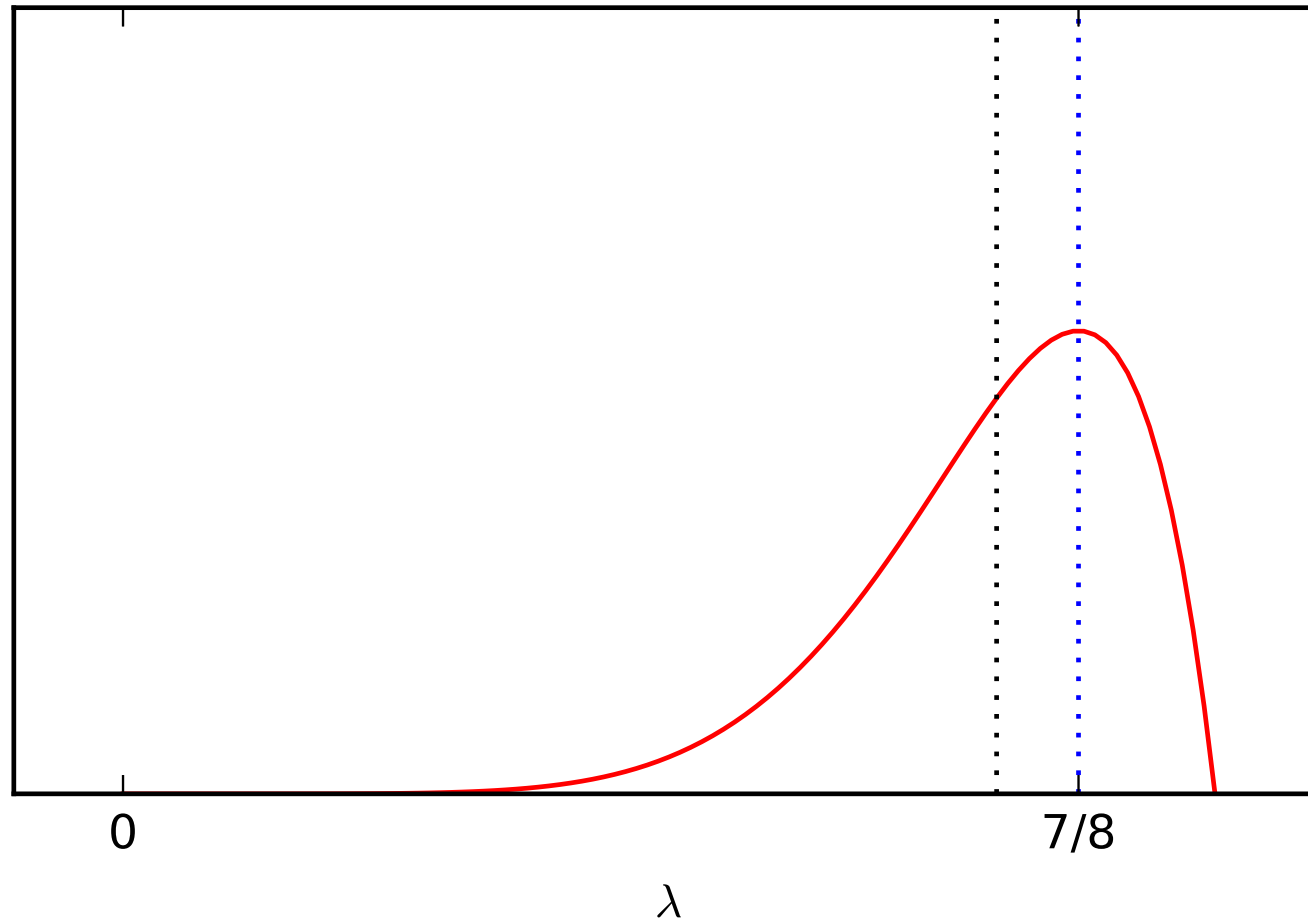
$$p(\lambda|x_{1:7})$$

T, T, T, T, T, Y, T, ?



$$p(\lambda|x_{1:8})$$

T, T, T, T, T, Y, T, T, ?



# Olasılık Modelleri



# Olasılık Dağılımları

- Olasılık modellerinin yapı taşları:
  - Kesikli
    - \* Kategorik, Bernoulli, Binomial, Multinomial, Poisson
  - Sürekli
    - \* Gaussian,
    - \* Beta, Dirichlet
    - \* Gamma, Ters Gamma, Üstel, Chi-kare, Wishart
    - \* Student-t, von-Mises

# Üstel aile (Exponential Family)

- Dağılımların genel yapısı

$$p(x|\theta) = h(x) \exp\{\theta^\top \psi(x) - A(\theta)\}$$

$$A(\theta) = \log \int_{\mathcal{X}^n} dx h(x) \exp(\theta^\top \psi(x))$$

$A(\theta)$  log-bölüntüleme (log-partition) fonksyonu

$\theta$

kanonik parametreler

$\psi(x)$

yeterli istatistik

$h(x)$

ağırlık fonksyonu

# En Büyük Entropi prensibi

İstenen beklenen değerlere sahip ama en çok belirsizlik içeren dağılım hangisidir?

$$H[p] = - \int_{\mathcal{X}} p(x) \log(p(x)) dx$$

maximize  $H[p]$

subject to

$$\int_{\mathcal{X}} p(x) dx = 1$$

Düzleştirme

$$\int_{\mathcal{X}} \psi(x) p(x) dx = s$$

Beklenti Eşleme

# Lagrange Fonksiyoneli

$$\Lambda(p; \lambda, \theta) = - \int_{\mathcal{X}} p(x) \log(p(x)) dx + \lambda(1 - \int_{\mathcal{X}} p(x) dx) + \theta(s - \int_{\mathcal{X}} \psi(x) p(x) dx)$$

$$\frac{\delta}{\delta p} \Lambda[p, \lambda, \theta] = -\log(p(x)) - 1 + \lambda + \theta \phi(x) = 0$$

$$p(x) = \exp(\theta \psi(x)) \exp(\lambda - 1)$$

Normalizasyon kısıtı

$$\int_{\mathcal{X}} p(x) dx = 1 = \exp(\lambda - 1) \int_{\mathcal{X}} \exp(\theta \psi(x)) dx$$

$$\exp(\lambda - 1) = \frac{1}{\int \exp(\theta \psi(x)) dx}$$

$\lambda$ 'dan kurtulalım

$$A(\theta) \equiv \log \int_{\mathcal{X}} \exp(\theta \psi(x)) dx$$

Çözüm: Üstel Aile formunda (Gibbs dağılımı)

$$p(x) = \exp(\theta \psi(x) - A(\theta)) \quad (1)$$



## Bernoulli. $\mathcal{BE}(c; w)$

Bernoulli rassal değişken  $c = \{0, 1\}$  başarı olasılığı  $w$

$$p(c = 1|w) = w \quad p(c = 0|w) = 1 - w$$

$$\begin{aligned} p(c|w) &= w^c(1 - w)^{1-c} \\ &= \exp(c \log w + (1 - c) \log(1 - w)) \\ &= \exp\left(\log\left(\frac{w}{1 - w}\right)c + \log(1 - w)\right) \\ &\equiv \mathcal{BE}(c; w) \end{aligned}$$

# Bernoulli Üstel Aileden mi?

$$\mathcal{BE}(c; w) = \exp \left( \log\left(\frac{w}{1-w}\right)c + \log(1-w) \right)$$

$$p(c|\theta) = h(c) \exp\{\theta^\top \psi(c) - A(\theta)\}$$

$$\theta = \log\left(\frac{w}{1-w}\right) \quad \text{canonical parameters}$$

$$A(\theta) = -\log(1 + e^\theta) \quad \text{log-partition function}$$

$$\psi(c) = c \quad \text{sufficient statistics}$$

$$h(c) = 1 \quad \text{weighting function}$$

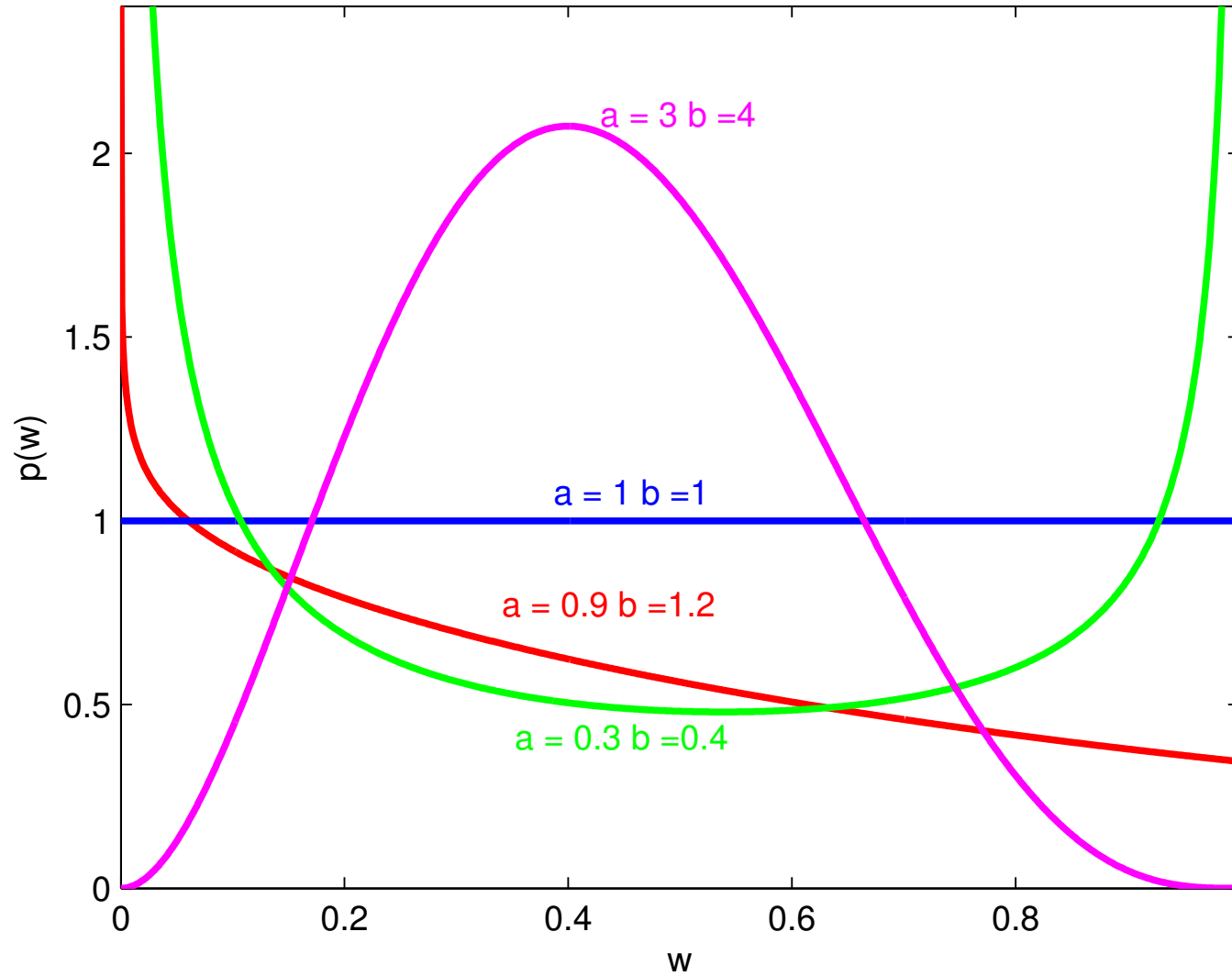
## Beta Dağılımı. $\mathcal{B}(w; a, b)$

$$\begin{aligned}\mathcal{B}(w; a, b) &\equiv \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1} \\ &= \exp((a-1)\log w + (b-1)\log(1-w) - A(a, b)) \\ &= \exp\left(\begin{pmatrix} a-1 & b-1 \end{pmatrix} \begin{pmatrix} \log w \\ \log(1-w) \end{pmatrix} - A(a, b)\right) \\ A(a, b) &= \log \Gamma(a) + \log \Gamma(b) - \log \Gamma(a+b)\end{aligned}$$

Mean :

$$\langle w \rangle_{\mathcal{B}} = a/(a+b)$$

## Beta Dağılımı. $\mathcal{B}(w; a, b)$



## Gauss. $\mathcal{N}(x; m, S)$

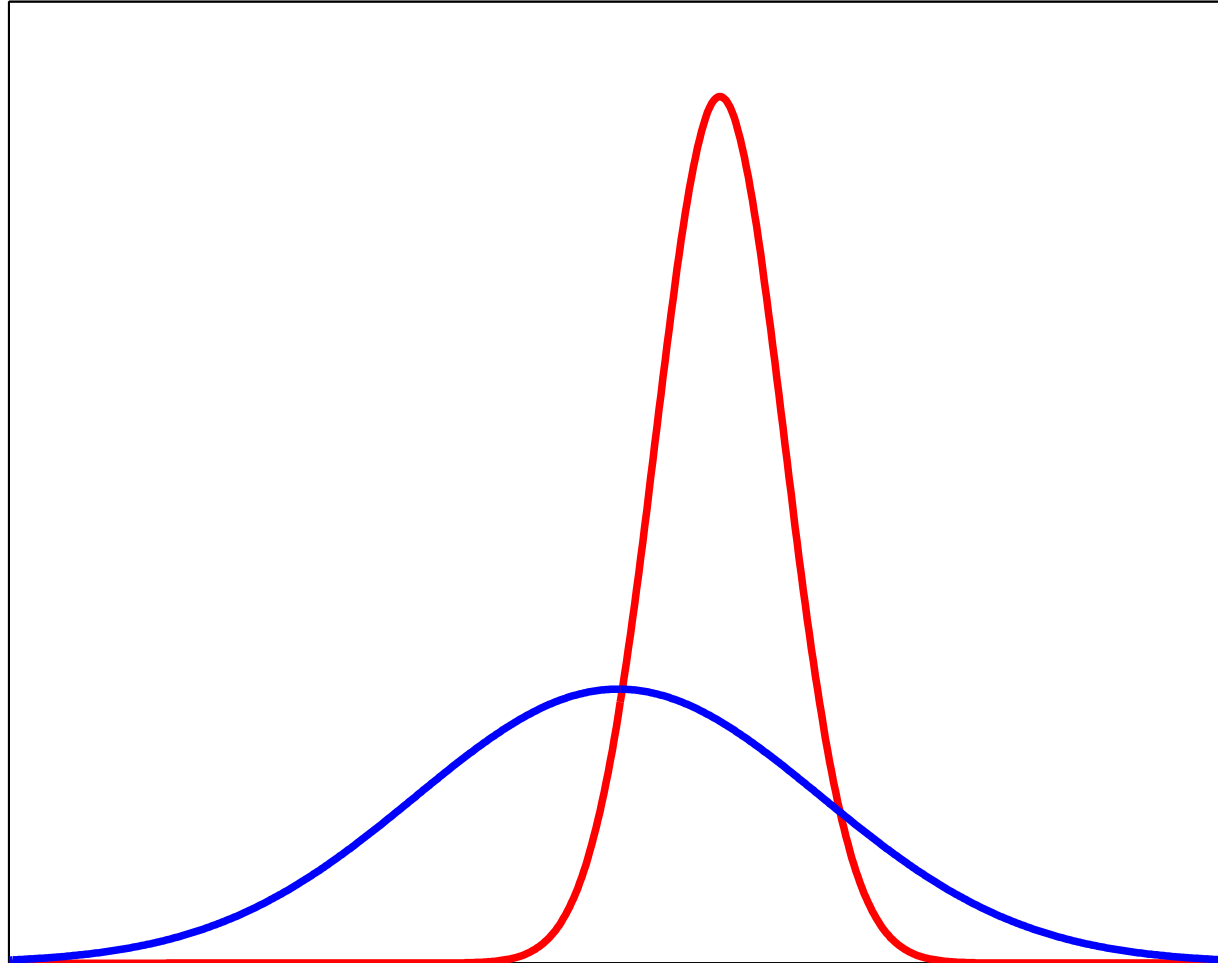
Gauss dağılımı, beklenen değer  $m$  ve varyans  $S$

$$\begin{aligned}\mathcal{N}(x; m, S) &= (2\pi S)^{-1/2} \exp\left\{-\frac{1}{2}(x - m)^2/S\right\} \\ &= \exp\left\{-\frac{1}{2}(x^2 + m^2 - 2xm)/S - \frac{1}{2}\log(2\pi S)\right\} \\ &= \exp\left\{\frac{m}{S}x - \frac{1}{2S}x^2 - \left(\frac{1}{2}\log(2\pi S) + \frac{1}{2S}m^2\right)\right\} \\ &= \exp\left\{\underbrace{\begin{pmatrix} m/S \\ -\frac{1}{2}/S \end{pmatrix}}_{\theta}^\top \underbrace{\begin{pmatrix} x \\ x^2 \end{pmatrix}}_{\psi(x)} - A(\theta)\right\}\end{aligned}$$

Katsayıları eşleştirme

$$\exp\left\{-\frac{1}{2}Kx^2 + hx + g\right\} \Leftrightarrow S = K^{-1} \quad m = K^{-1}h$$

# Gaussian.



## Ters Gamma Dağılımı. $\mathcal{IG}(r; a, b)$

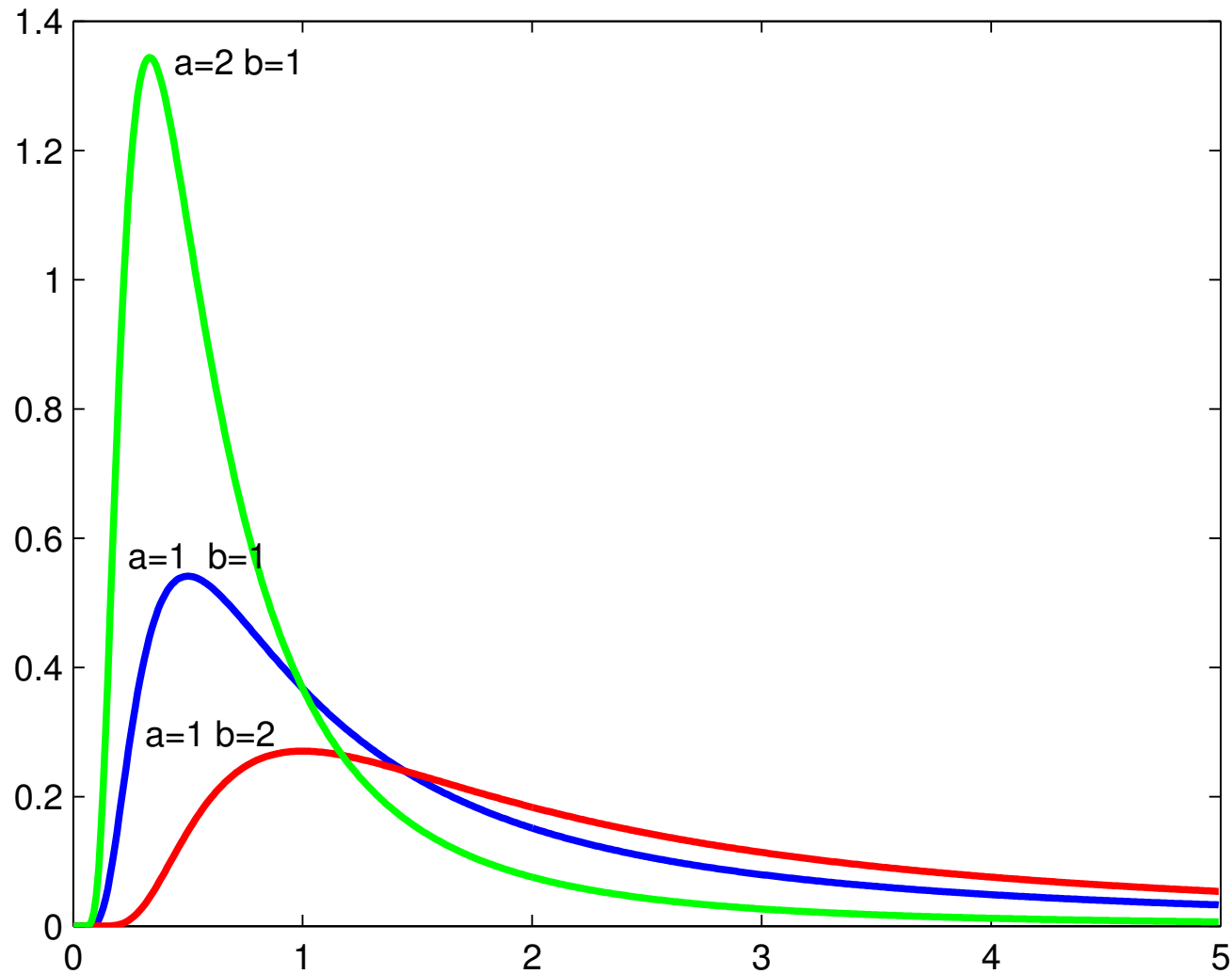
The inverse Gamma distribution with shape  $a$  and scale  $b$

$$\begin{aligned}\mathcal{IG}(r; a, b) &= \frac{1}{\Gamma(a)} \frac{r^{-(a+1)}}{b^{-a}} \exp\left(-\frac{b}{r}\right) \\ &= \exp\left(- (a+1) \log r - \frac{b}{r} - \log \Gamma(a) + a \log b\right) \\ &= \exp\left(\begin{pmatrix} -(a+1) \\ -b \end{pmatrix}^\top \begin{pmatrix} \log r \\ 1/r \end{pmatrix} - \log \Gamma(a) + a \log b\right)\end{aligned}$$

Katsayıları eşleştirme

$$\exp\left\{\alpha \log r + \beta \frac{1}{r} + c\right\} \Leftrightarrow a = -\alpha - 1 \quad b = -\beta$$

# Inverse Gamma





## Gamma Distribution. $\mathcal{G}(\lambda; a, b)$

The Gamma distribution with shape  $a$  and **inverse scale**  $b$

$$\begin{aligned}\mathcal{G}(\lambda; a, b) &= \frac{1}{\Gamma(a)} b^a \lambda^{(a-1)} \exp(-b\lambda) \\ &= \exp((a-1) \log \lambda - b\lambda - \log \Gamma(a) + a \log b) \\ &= \exp \left( \begin{pmatrix} (a-1) \\ -b \end{pmatrix}^\top \begin{pmatrix} \log \lambda \\ \lambda \end{pmatrix} - \log \Gamma(a) + a \log b \right)\end{aligned}$$

Hence by matching coefficients, we have

$$\exp \left\{ \alpha \log r + \beta \frac{1}{r} + c \right\} \Leftrightarrow a = \alpha + 1 \quad b = -\beta$$

# Rassal sayı üretme

- Bernoulli:  $\mathcal{BE}(x; p)$

```
x = double(rand < p) ;
```

- Binomial:  $\mathcal{BI}(x; p, N)$

```
x = sum(double(rand(N, 1) < p)) ;
```

Not efficient for large  $N$

- Poisson:  $\mathcal{PO}(x; \lambda)$

```
x = poissrnd(lambda) ;
```

- Beta:  $\mathcal{B}(x; a, b)$

```
x = betarnd(a, b) ;
```

- Gaussian:  $\mathcal{N}(x; \mu, S)$

```
x = sqrt(S) .* randn(size(S)) + mu;
```

- Gamma:  $x \sim \mathcal{G}(x; a, b)$

```
x = gamrnd(a, 1./b);
```

or more securely

```
x = gamrnd(a, 1) ./b;
```

which is also

```
x = gamrnd(a) ./b;
```

- Inverse Gamma  $x \sim \mathcal{IG}(x; a, b)$

```
x = b ./ gamrnd(a);
```

## Konjuge Önsel: Sonsal ve Önsel aynı ailedendir.

Örnek: Bir  $c$  rassal Bernoulli değişkenin parametresi  $w$  için sonsal dağılım çıkarımı

$$p(c|w) = \mathcal{BE}(c; w) = \exp(c \log w + (1 - c) \log(1 - w))$$

$$p(w) = \mathcal{B}(w; a, b)$$

$$p(w|c) \propto p(c|w)p(w)$$

$$\propto \exp(c \log w + (1 - c) \log(1 - w))$$

$$\times \exp((a - 1) \log w + (b - 1) \log(1 - w))$$

$$\propto \mathcal{B}(w; a + c, b + (1 - c))$$

$$p(w|c) = \begin{cases} \mathcal{B}(w; a + 1, b) & c = 1 \\ \mathcal{B}(w; a, b + 1) & c = 0 \end{cases}$$



# Konjuge Önsel dağılım

Example: Sıfır ortalamalı Gauss dağılımının varyansı

$$p(x|R) = \mathcal{N}(x; 0, R)$$

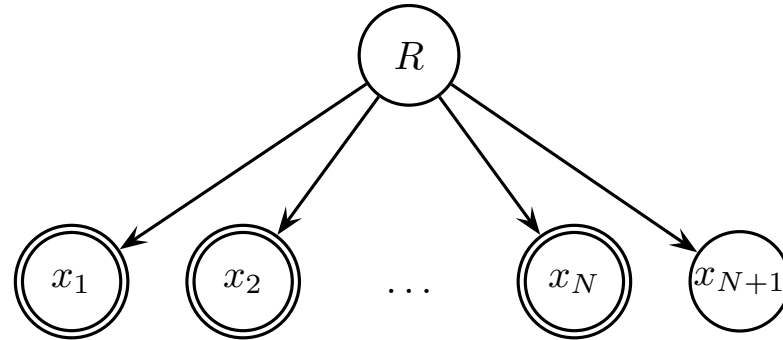
$$p(R) = \mathcal{IG}(R; a, b)$$

$$\begin{aligned} p(R|x) &\propto p(R)p(x|R) \\ &\propto \exp\left(-(a+1)\log R - b\frac{1}{R}\right) \exp\left(-(x^2/2)\frac{1}{R} - \frac{1}{2}\log R\right) \\ &= \exp\left(\begin{pmatrix} -(a+1+\frac{1}{2}) \\ -(b+x^2/2) \end{pmatrix}^\top \begin{pmatrix} \log R \\ 1/R \end{pmatrix}\right) \\ &\propto \mathcal{IG}(R; a + \frac{1}{2}, b + x^2/2) \end{aligned}$$

Like the prior, this is an inverse-Gamma distribution.

# Konjuge Önsel dağılım

Örnek: varyansın sonsal dağılımı  $R$  from  $x_1, \dots, x_N$ .

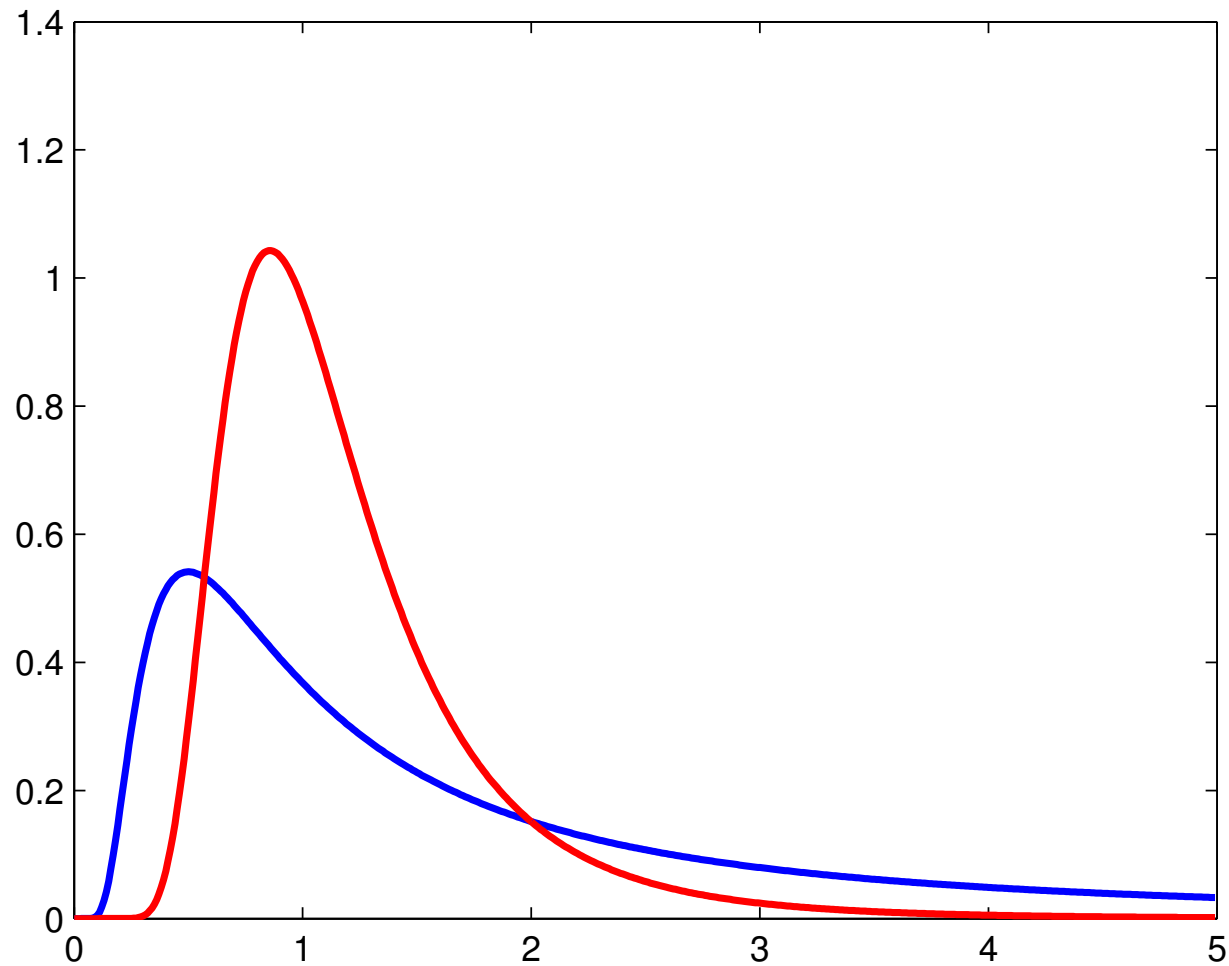


$$\begin{aligned} p(R|x) &\propto p(R) \prod_{i=1}^N p(x_i|R) \\ &\propto \exp \left( -(a+1) \log R - b \frac{1}{R} \right) \exp \left( - \left( \frac{1}{2} \sum_i x_i^2 \right) \frac{1}{R} - \frac{N}{2} \log R \right) \\ &= \exp \left( \begin{pmatrix} -(a+1 + \frac{N}{2}) \\ -(b + \frac{1}{2} \sum_i x_i^2) \end{pmatrix}^\top \begin{pmatrix} \log R \\ 1/R \end{pmatrix} \right) \propto \mathcal{IG}(R; a + \frac{N}{2}, b + \frac{1}{2} \sum_i x_i^2) \end{aligned}$$

Yeterli istatistikler **toplamsal**

# Ters Gamma, $\sum_i x_i^2 = 10$ $N = 10$

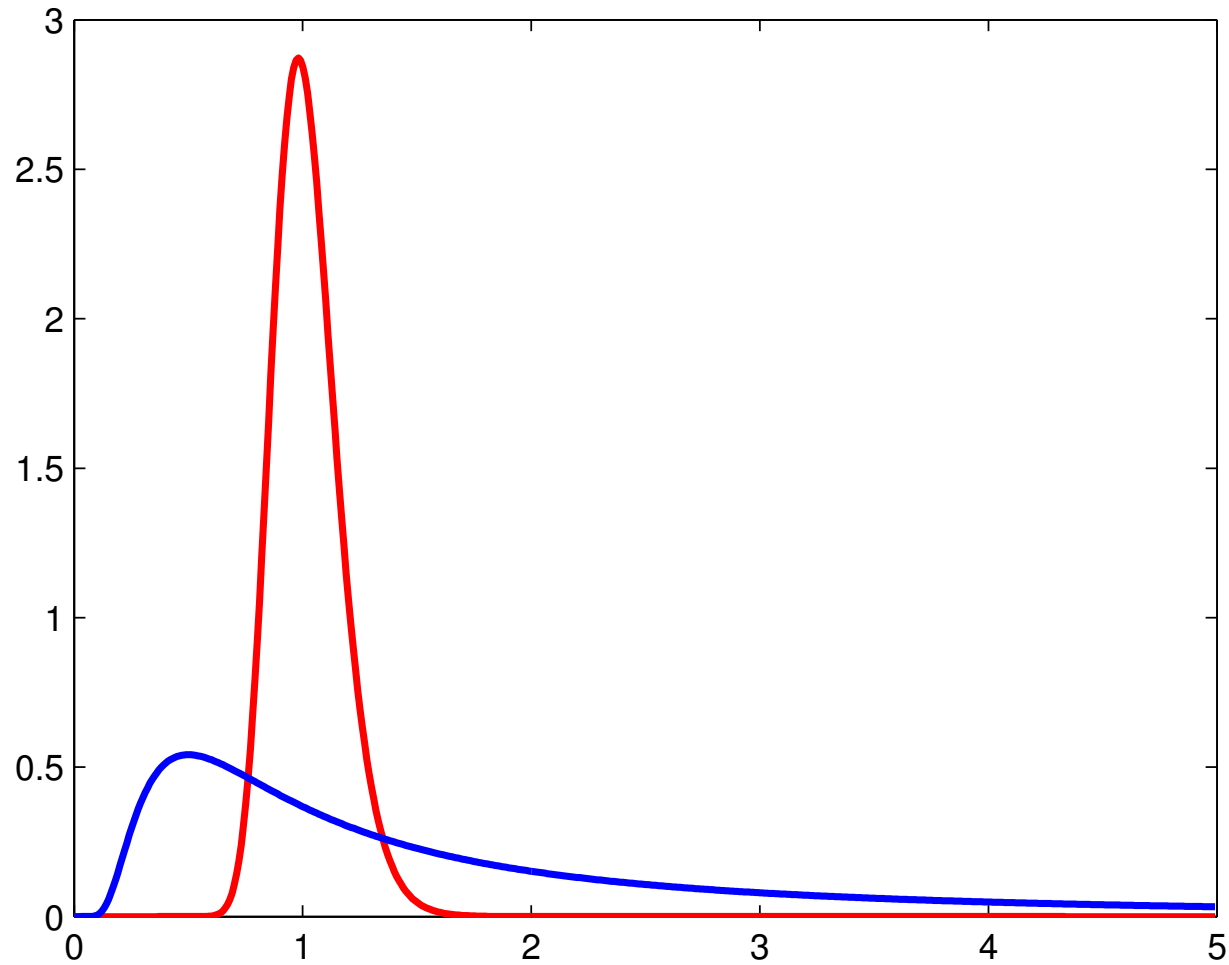
$$\sum_i x_i^2 = 10 \quad N = 10$$





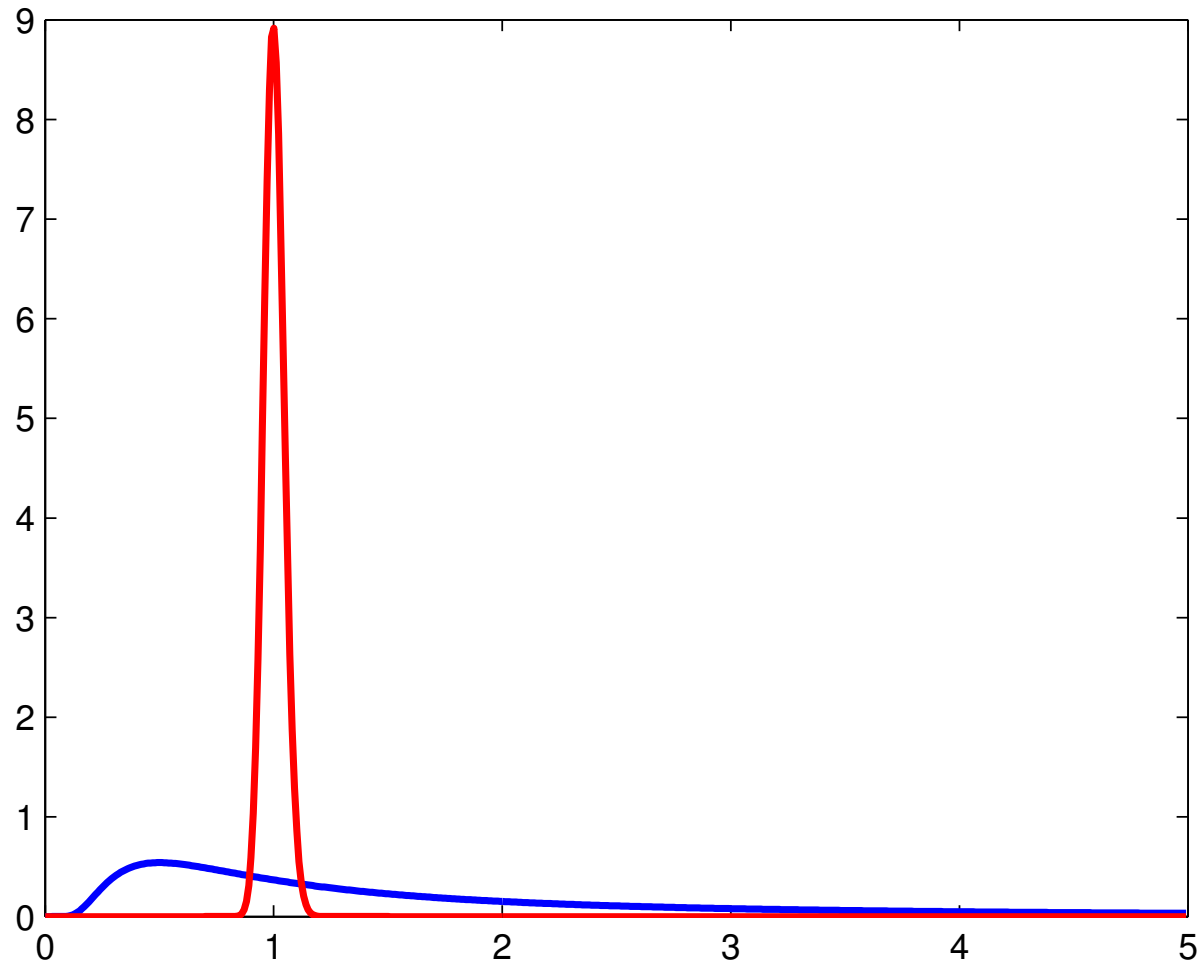
**Ters Gamma,  $\sum_i x_i^2 = 100$   $N = 100$**

$$\sum_i x_i^2 = 100 \quad N = 100$$

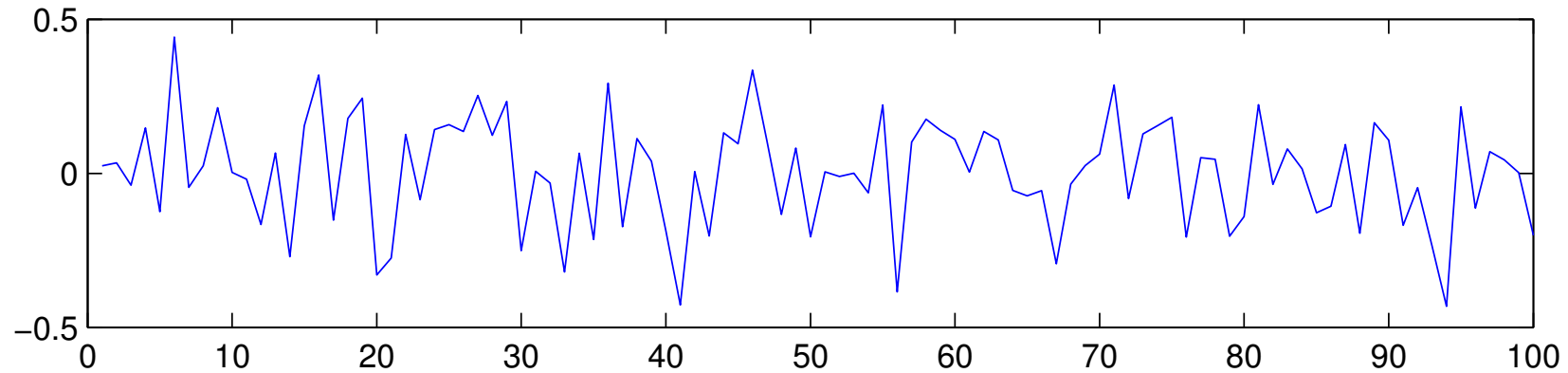


**Ters Gamma,  $\sum_i x_i^2 = 1000$   $N = 1000$**

$$\sum_i x_i^2 = 1000 \quad N = 1000$$



## Örnek: AR(1) modeli



$$x_k = Ax_{k-1} + \epsilon_k \quad k = 1 \dots K$$

$\epsilon_k$  is i.i.d., sıfır ortalamalı, varyans  $R$ .

**Kestirim problemi:**

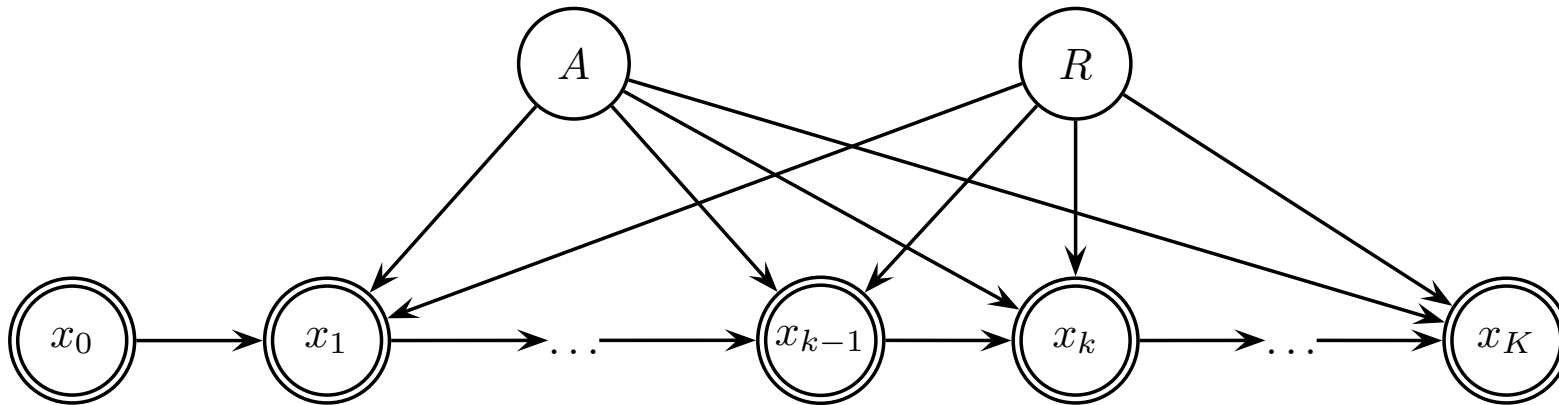
Gözlemlenen  $x_0, \dots, x_K$  verilerden,  $A$  ve  $R$  katsayılarını çıkaralım.

# AR(1) model, Üreten model notasyonu

$$A \sim \mathcal{N}(A; 0, P)$$

$$R \sim \mathcal{IG}(R; \nu, \beta/\nu)$$

$$x_k | x_{k-1}, A, R \sim \mathcal{N}(x_k; Ax_{k-1}, R) \quad x_0 = \hat{x}_0$$



Gözlemlenen değişkenler çift dairelerle gösterilmiştir

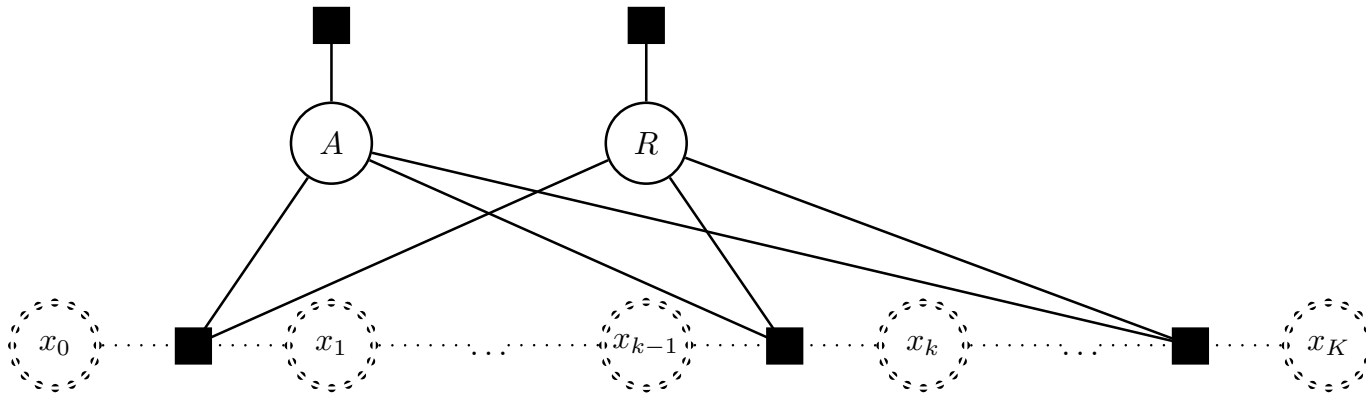
# AR(1) Modeli. Bayesci Sonsal Dağılımın Çıkarımı

$$p(A, R|x_0, x_1, \dots, x_K) \propto p(x_1, \dots, x_K|x_0, A, R)p(A, R)$$

$$\text{Sonsal} \propto \text{Olabilirlik} \times \text{Önsel}$$

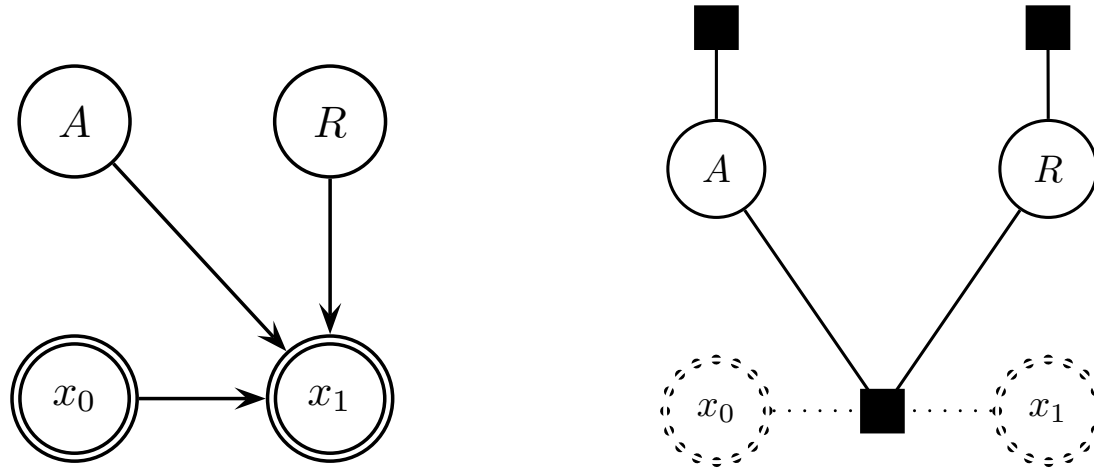
Markov şartlı bağımsızlık özelliğini kullanarak

$$p(A, R|x_0, x_1, \dots, x_K) \propto \left( \prod_{k=1}^K p(x_k|x_{k-1}, A, R) \right) p(A)p(R)$$



# Numerik Örnek

Suppose  $K = 1$ ,



Bayes kuralı ve AR modelinin yapısı:

$$\begin{aligned} p(A, R|x_0, x_1) &\propto p(x_1|x_0, A, R)p(A)p(R) \\ &= \mathcal{N}(x_1; Ax_0, R)\mathcal{N}(A; 0, P)\mathcal{IG}(R; \nu, \beta/\nu) \end{aligned}$$

## Numerik Örnek

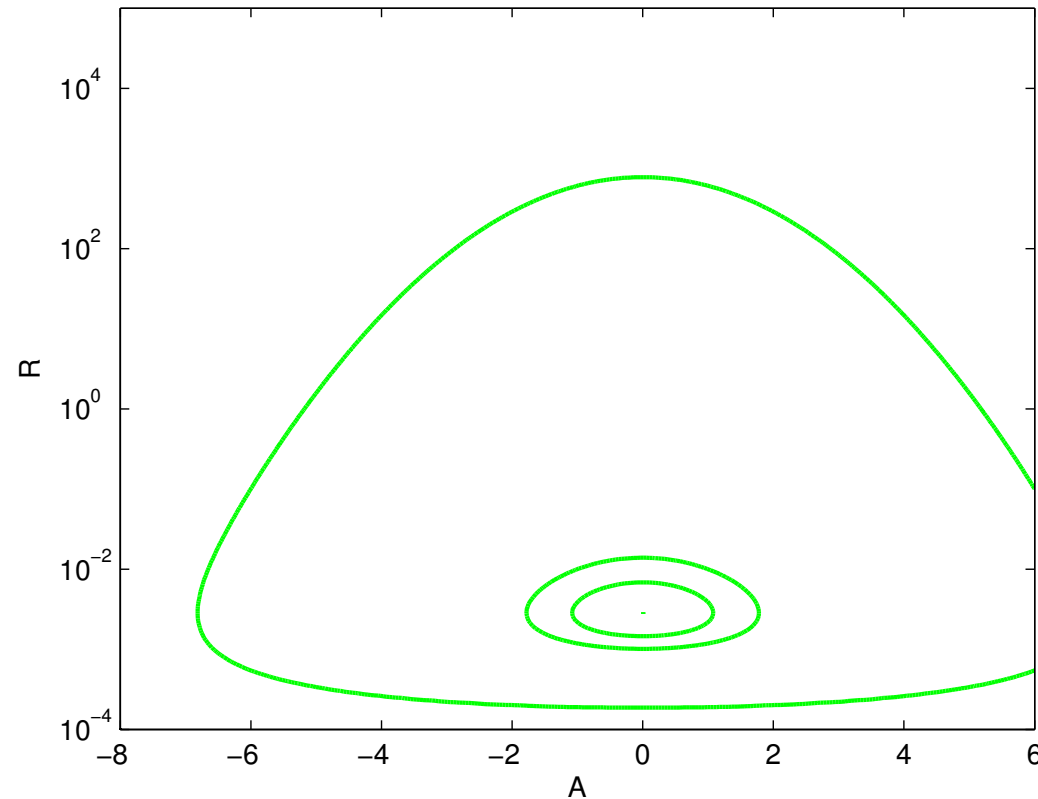
$$\begin{aligned} p(A, R|x_0, x_1) &\propto p(x_1|x_0, A, R)p(A)p(R) \\ &= \mathcal{N}(x_1; Ax_0, R)\mathcal{N}(A; 0, P)\mathcal{IG}(R; \nu, \beta/\nu) \\ &\propto \exp\left(-\frac{1}{2}\frac{x_1^2}{R} + x_0x_1\frac{A}{R} - \frac{1}{2}\frac{x_0^2A^2}{R} - \frac{1}{2}\log 2\pi R\right) \\ &\quad \exp\left(-\frac{1}{2}\frac{A^2}{P}\right) \exp\left(-(\nu + 1)\log R - \frac{\nu}{\beta}\frac{1}{R}\right) \end{aligned}$$

Sonsal dağılım standard bir forma sahip değil

$$\exp\left(\alpha_1\frac{1}{R} + \alpha_2\frac{A}{R} + \alpha_3\frac{A^2}{R} + \alpha_4\log R + \alpha_5A^2\right)$$

# Nümerik örnek $p(A, R)$

Eşit olasılık konturu  $p(A)p(R)$

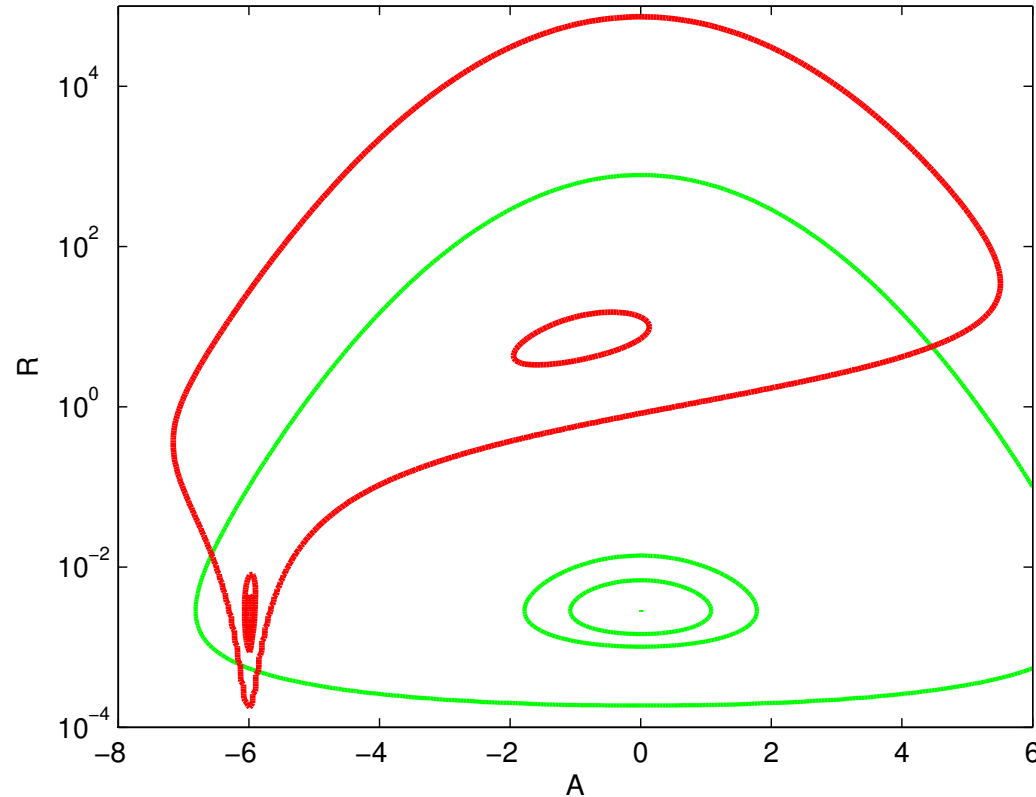


$$A \sim \mathcal{N}(A; 0, 1.2) \quad R \sim \mathcal{IG}(R; 0.4, 250)$$

$$\text{Suppose: } x_0 = 1 \quad x_1 = -6 \quad x_1 \sim \mathcal{N}(x_1; Ax_0, R)$$



## Numerik Örnek, sonsal dağılım $p(A, R|x)$



Note the bimodal posterior with  $x_0 = 1, x_1 = -6$

- $A \approx -6 \Leftrightarrow$  low noise variance  $R$ .
- $A \approx 0 \Leftrightarrow$  high noise variance  $R$ .

# Notlar

- ML veya MAP gibi noktasal kestirimler her zaman çok belirleyici olmayabilir, sonsal dağılım her zaman daha detaylı bilgi içerir
- Sonsal dağılımı tam olarak hesaplamak bazı özel durumlar dışında çok zor olabilir
- Çok basit gözüken modeller bile karmaşık sonsal dağılımlar oluşturabilir
- Veri birbirleri ile çelişen değişik şekillerde açıklanabiliyorsa bu birden fazla kipi olan bir sonsal dağılım ile sonuçlanabilir

# Notla

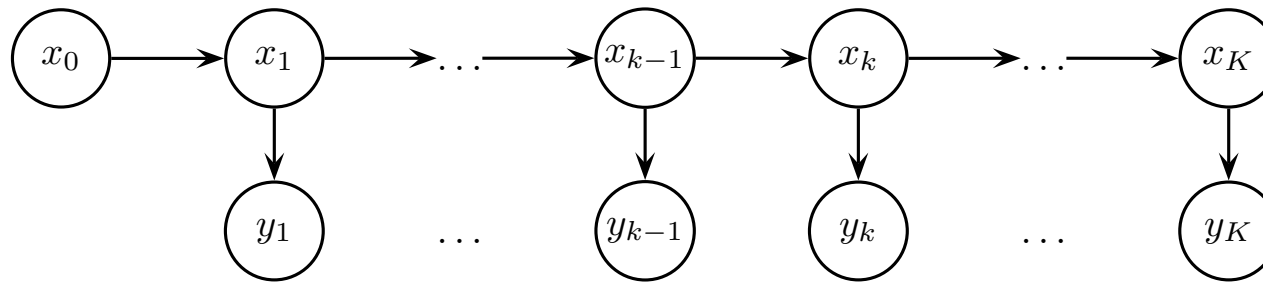
- Öncül (*a-priori*) olarak birbirinden bağımsız değişkenler sonsal olarak birbirleri ile bağlanabilirler (“Explaining away”)
- Bir çıkarım probleminin zorluğu verilen verinin ve kestirilmeye çalışılan parametrelerin ilişkisine (parametre rejimine) doğrudan bağlıdır. Genelde veri kıtlığı (minik veri) veya çok veri (büyük veri) olduğu durumlarda çözümlere ulaşmak daha kolay olabilir. Ara durumlar (ne çok ne az) daha zor olabilir.

# Ders

- Sıralı veriler, Terminoloji
- Saklı Markov Modelleri
- İleri-Geri algoritması
- MAP gezingesinin bulunması: Viterbi algoritması

# Sıralı veriler: Modeller, Çıkarım, Terminoloji

İşaret işleme, yapay öğrenme, robotik, istatistik alanında bir çok veri dinamik modeller ile modellenenbilir



$$x_k \sim p(x_k | x_{k-1})$$

Durum Geçiş Modeli

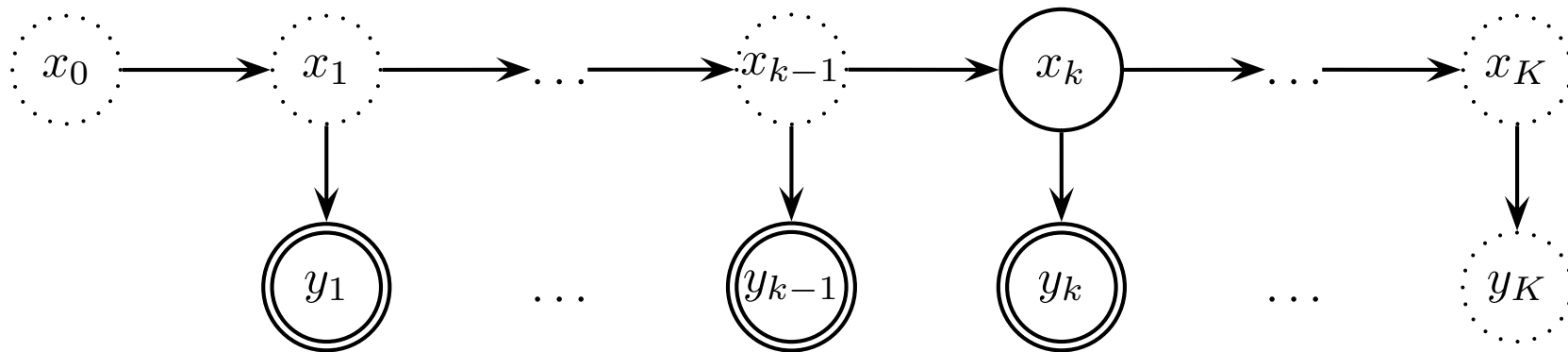
$$y_k \sim p(y_k | x_k)$$

Gözlem Modeli

- $x$  gizli durum (hız, sınıf etiketi, bilinmeyen model parametreler )
- $y$  gözlemler (örnekler, algılayıcılardan okuması, öznelilikler, ... )

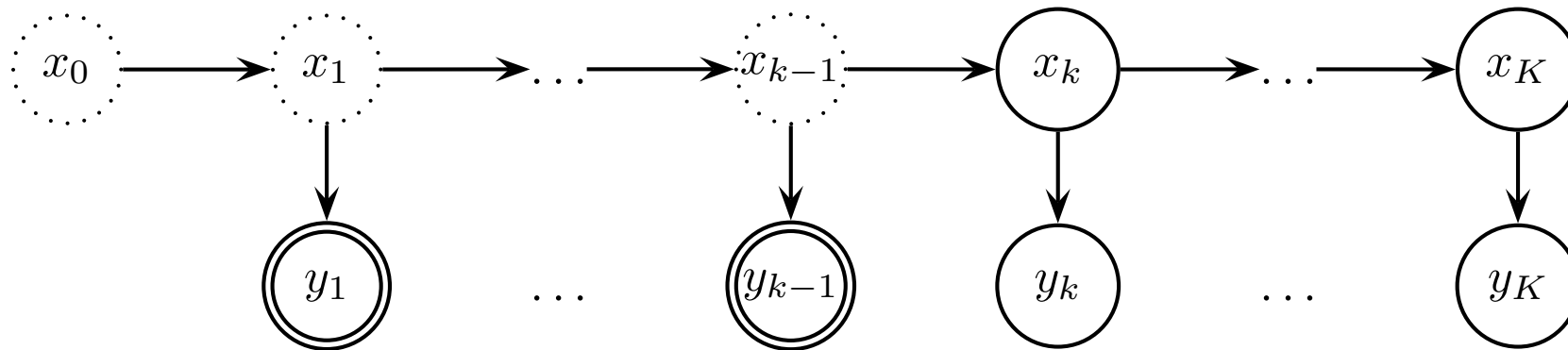
# Sıralı (Sequential/Online) Çıkarım, Terminoloji

- **Süzgeçleme (Filtering):**  $p(x_k | y_{1:k})$ 
  - Geçmiş gözlemler verildiğinde şu andaki durum
  - Gerçek zamanda sıralı olarak çıkarım



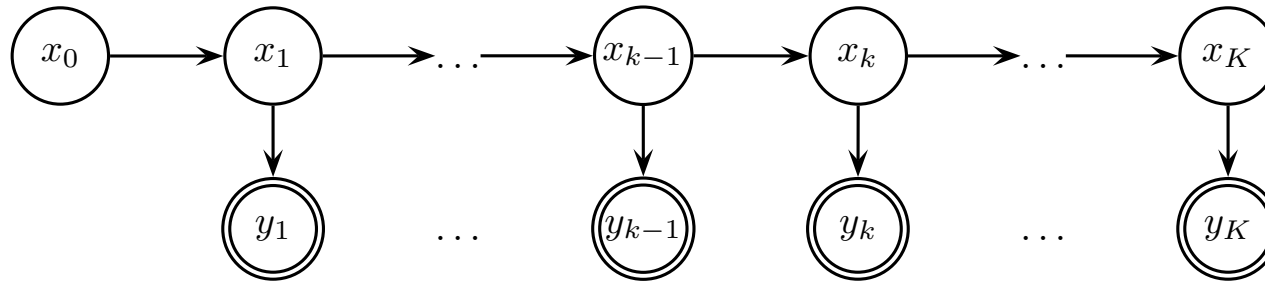
# Sıralı (Sequential/Online) Çıkarım, Terminoloji

- **Tahminleme**  $p(y_{k:K}, x_{k:K} | y_{1:k-1})$ 
  - Gelecekte olabilecek durumların ve gözlemlerin tahmini; veri olmadan tahminleme

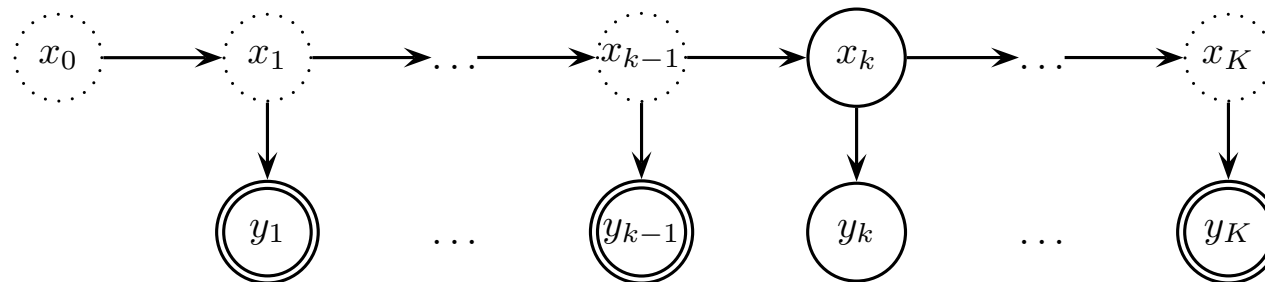


# Offline Inference, Terminology

- **Smoothing**  $p(x_{0:K}|y_{1:K})$ ,  
**Most likely trajectory – Viterbi path**  $\arg \max_{x_{0:K}} p(x_{0:K}|y_{1:K})$   
better estimate of past states, essential for learning



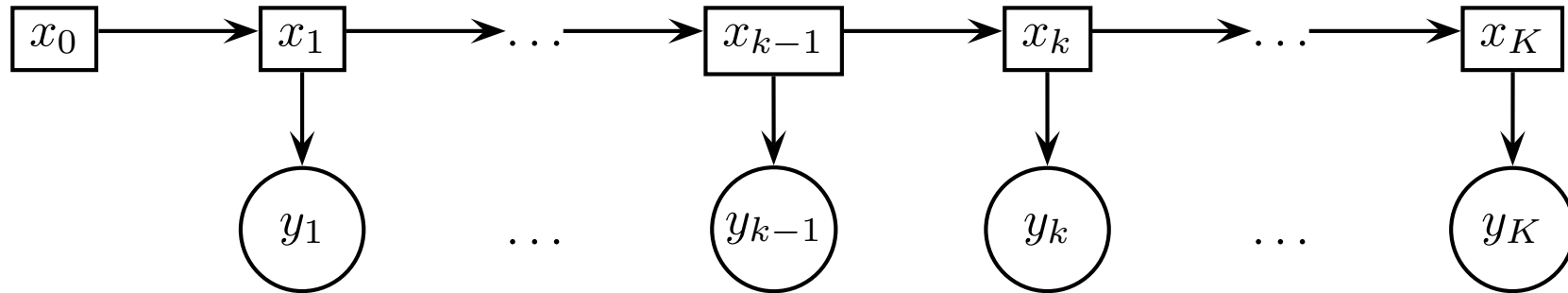
- **Interpolation**  $p(y_k, x_k|y_{1:k-1}, y_{k+1:K})$   
fill in lost observations given past and future





# Saklı Markov Modelleri

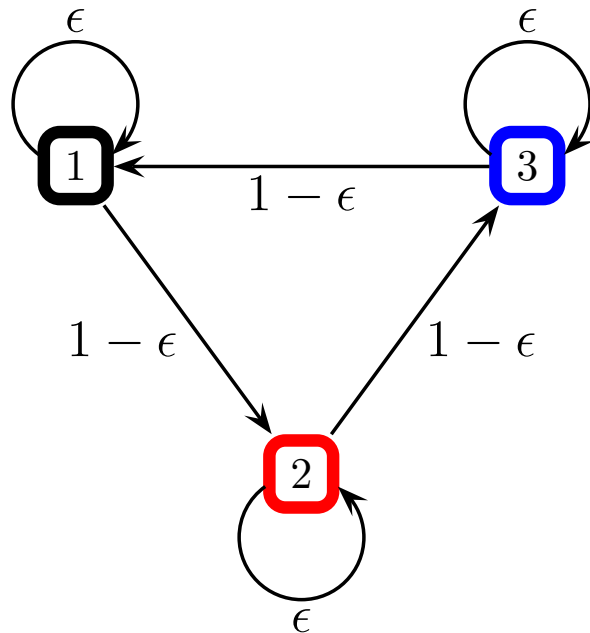
- Zaman içerisinde ilerleyen karışım modeli



- Gözlemler  $y_k$  sonlu veya sonsuz bir kümeden
- Gizli durumlar  $x_k$  sonlu bir kümeden
  - Sürecin sönmömlenen hafızası
- Tam çıkarım ancak  $x_k$  az sayıda duruma sahipse mümkündür

## Örnek: Saklı Markov Modelleri

- Durum geçiş modelleri (a  $N$  by  $N$  matrix)

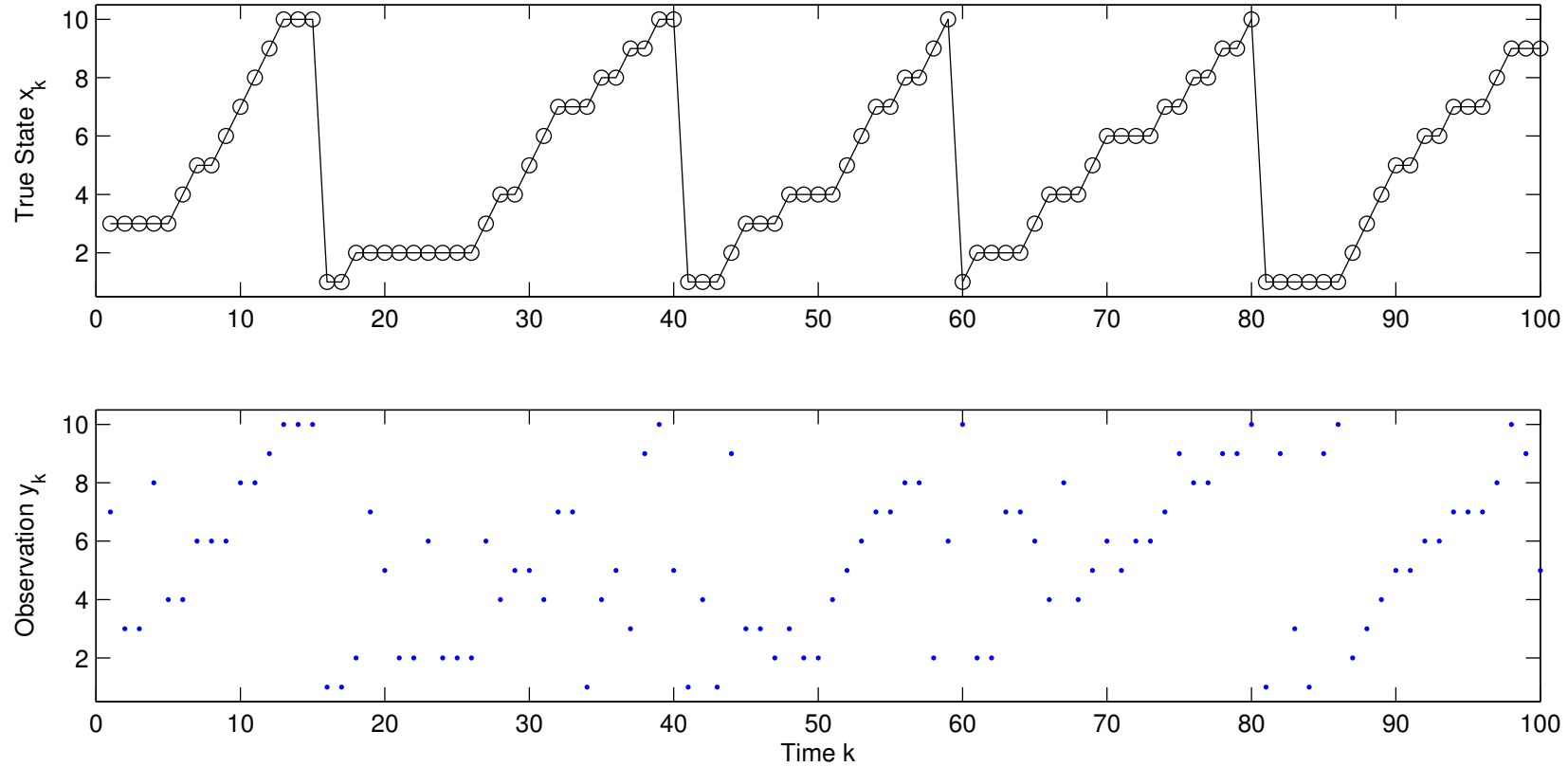


$$(1 - \epsilon) \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \epsilon \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

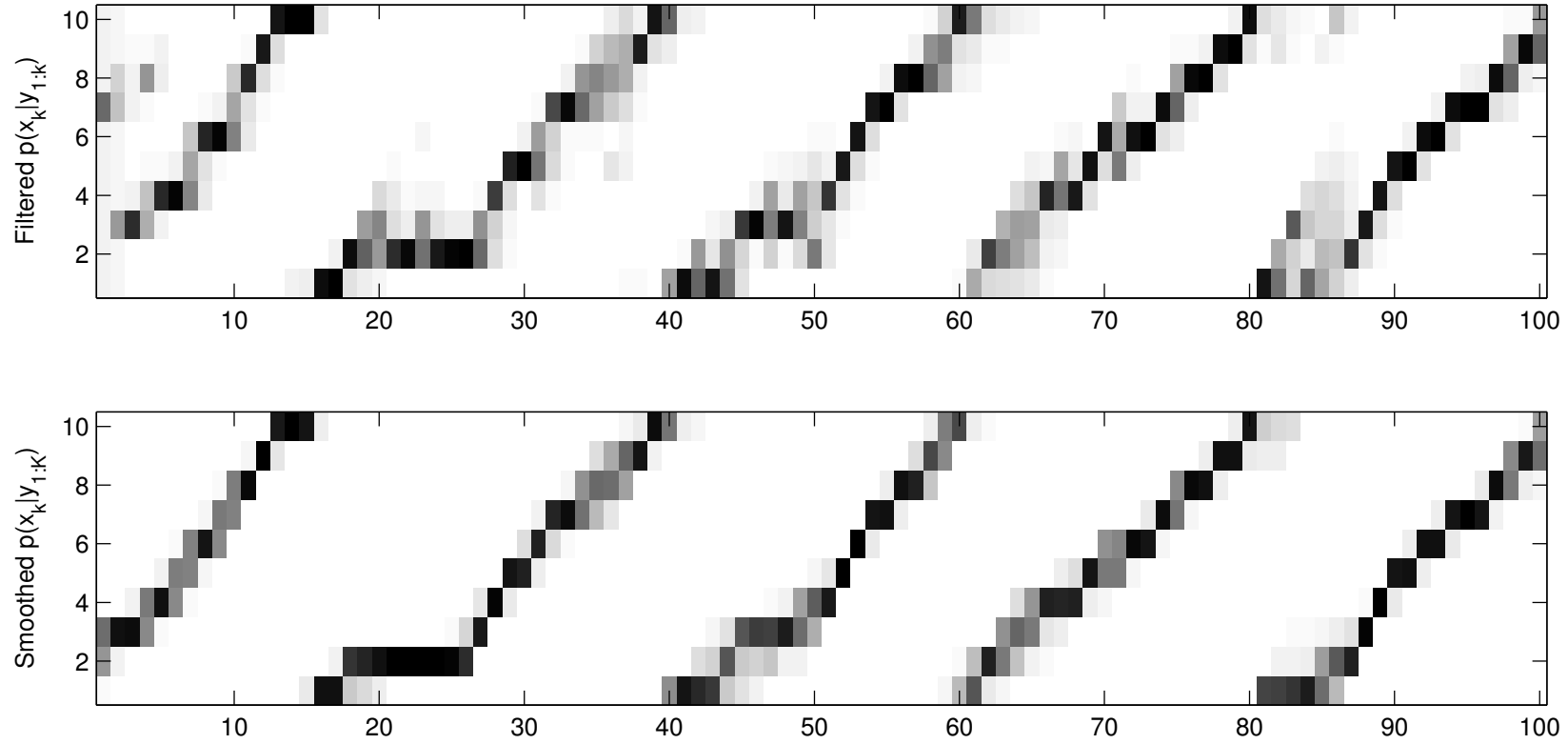
- Gözlem modeli  $p(y_k | x_k)$

$$y_k \sim w\delta(y_k - x_k) + (1 - w)u(1, N)$$

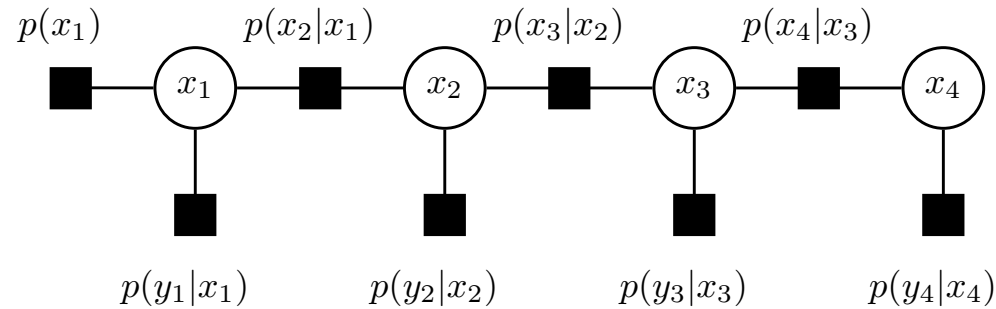
# Örnek: Saklı Markov Modelleri



# Örnek: Saklı Markov Modelleri



# İleri/Geri Algoritması



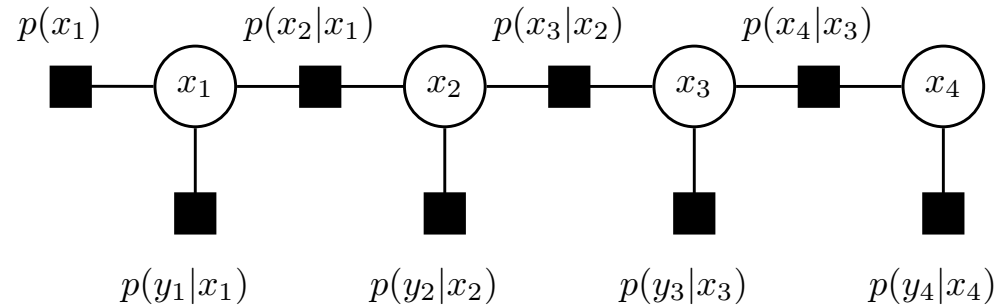
## • İleri

$$\begin{aligned}
 p(y_{1:K}) &= \sum_{x_{1:K}} p(y_{1:K}|x_{1:K}) p(x_{1:K}) \\
 &= \underbrace{\sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \sum_{x_2} p(x_3|x_2) p(y_2|x_2)}_{\alpha_K} \underbrace{\sum_{x_1} p(x_2|x_1) p(y_1|x_1)}_{\alpha_2} \underbrace{p(x_1)}_{\alpha_1|0}
 \end{aligned}$$

## • Geri

$$p(y_{1:K}) = \sum_{x_1} p(x_1) p(y_1|x_1) \cdots \underbrace{\sum_{x_{K-1}} p(x_{K-1}|x_{K-2}) p(y_{K-1}|x_{K-1})}_{\beta_{K-2}} \underbrace{\sum_{x_K} p(x_K|x_{K-1}) p(y_K|x_K)}_{\beta_{K-1}} \underbrace{1}_{\beta_K}$$

# Viterbi Algoritması



- $+$   $\rightarrow$  max,  $\times$   $\rightarrow$  +, dinamik programlama
- "ler'"

$$\begin{aligned}
 p(y_{1:K}|x_{1:K}^*) &= \max_{x_{1:K}} p(y_{1:K}|x_{1:K})p(x_{1:K}) \\
 &= \underbrace{\max_{x_K} p(y_K|x_K) \max_{x_{K-1}} p(x_K|x_{K-1}) \dots \max_{x_2} p(x_3|x_2)}_{\alpha_K} \underbrace{p(y_2|x_2) \max_{x_1} p(x_2|x_1)}_{\alpha_2} \underbrace{p(y_1|x_1) p(x_1)}_{\alpha_1}
 \end{aligned}$$

- Ger'

$$p(y_{1:K}|x_{1:K}^*) = \max_{x_1} p(x_1)p(y_1|x_1) \dots \underbrace{\max_{x_{K-1}} p(x_{K-1}|x_{K-2})p(y_{K-1}|x_{K-1})}_{\beta_{K-2}} \underbrace{\max_{x_K} p(x_K|x_{K-1})p(y_K|x_K)}_{\beta_{K-1}} \underbrace{1}_{\beta_K}$$

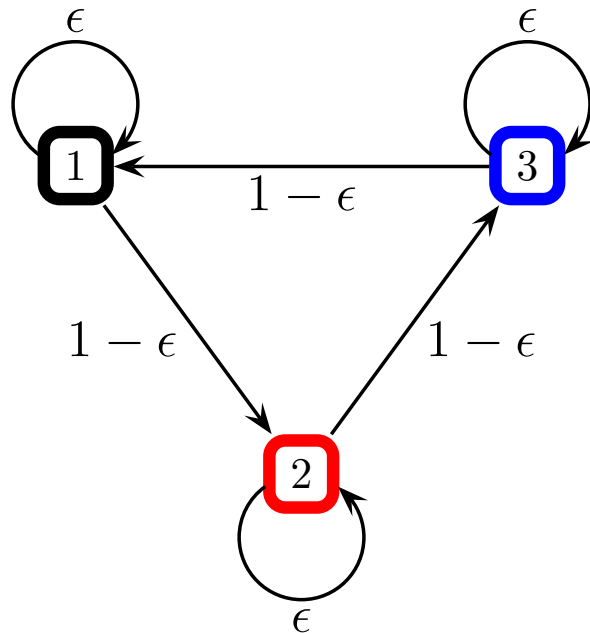
# İleri-Geri algoritmasının gerçekleştirilmesi

## Adımlar

1. Parametrelerin belirlenmesi
2. Gerçek modelden sentetik veri üretimi
3. Gerçek parametreler kullanarak çıkarım
4. Test ve görselleştirme

## Örnek: Saklı Markov modeli

- Durum geçiş modeli ( $N \times N$  matris)



$$(1 - \epsilon) \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \epsilon \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Gözlem Modeli  $p(y_k | x_k)$

$$y_k \sim w\delta(y_k - x_k) + (1 - w)u(1, N)$$



# 1. Parametreler

```
N = 50;      % Number of states

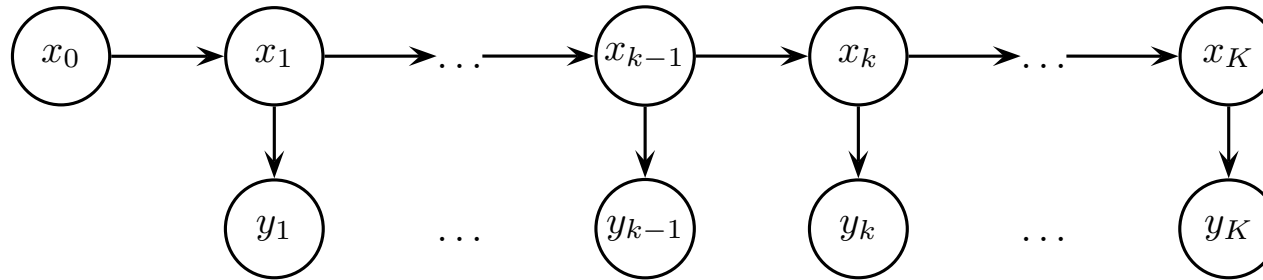
% Transition model;
ep = 0.5;    % Probability of not-moving
E = eye(N);
A = ep*E + (1-ep)*E(:, [2:N 1]); % Transition Matrix

% Observation model
w = 0.3; % Probability of observing true state
C = w*E + (1-w)*ones(N)/N; % Observation matrix

% Prior p(x_1)
pri = ones(N, 1)/N;

% Create a parameter structure
hm = struct('A', A, 'C', C, 'p_x1', pri);
```

## 2. Sentetik veri oluşturma



$$x_k | x_{k-1} \sim p(x_k | x_{k-1})$$

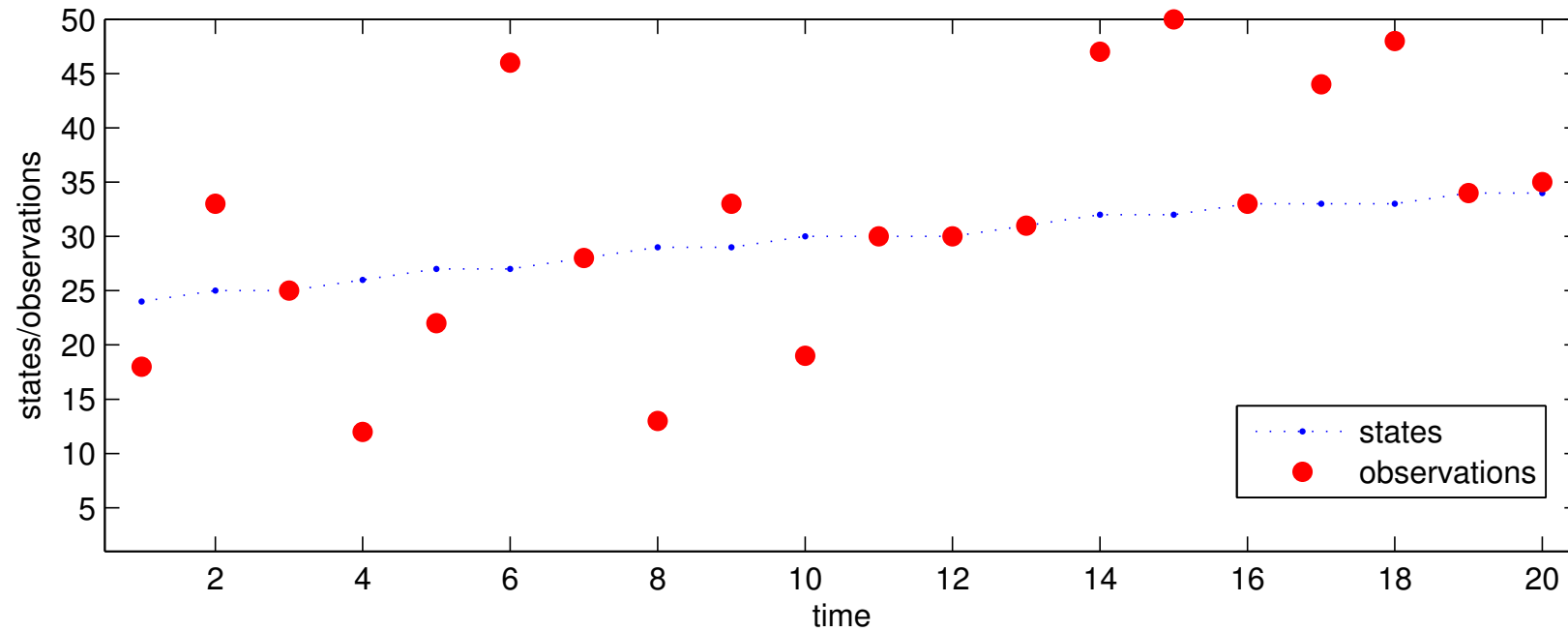
$$y_k | x_k \sim p(y_k | x_k)$$

## 2. Sentetik veri oluşturma

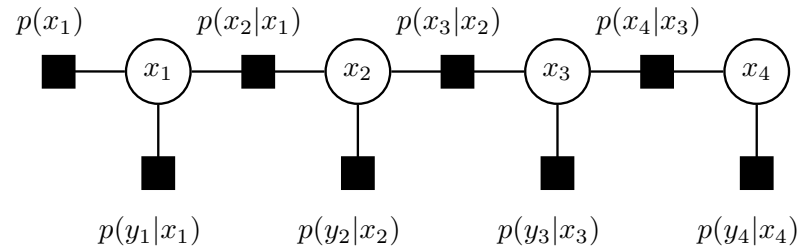
```
function [obs, state] = hmm_generate_data(hm, K)
% Inputs :
%         hm : A HMM parameter structure
%         K : Number of time slices to simulate
% Outputs :
%         obs, state : Observations and the state trajectory

state = zeros(1, K);
obs = zeros(1, K);
for k=1:K,
    if k==1,
        state(k) = randgen(hm.p_x1);
    else
        state(k) = randgen(hm.A(:, state(k-1)));
    end;
    obs(k) = randgen(hm.C(:, state(k)));
end;
```

## 2. Sentetik veri oluşturma



### 3. Çıkarım: İleri faz



- Predict

$$\begin{aligned}\alpha_{k|k-1}(x_k) &= p(y_{1:k-1}, x_k) = \sum_{x_{k-1}} p(x_k|x_{k-1})p(y_{1:k-1}, x_{k-1}) \\ &= \sum_{x_{k-1}} p(x_k|x_{k-1})\alpha_{k-1|k-1}(x_{k-1})\end{aligned}$$

- Update

$$\begin{aligned}\alpha_{k|k}(x_k) &= p(y_{1:k}, x_k) = p(y_k|x_k)p(y_{1:k-1}, x_k) \\ &= p(y_k|x_k)\alpha_{k|k-1}(x_k)\end{aligned}$$

$$\begin{aligned}
p(y_{1:K}) &= \sum_{x_{1:K}} p(y_{1:K}|x_{1:K})p(x_{1:K}) \\
&= \sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \sum_{x_2} p(x_3|x_2)p(y_2|x_2) \sum_{x_1} p(x_2|x_1) \underbrace{p(y_1|x_1)p(x_1)}_{\alpha_{1|1}}^{\alpha_{1|0}} \\
&= \sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \sum_{x_2} p(x_3|x_2)p(y_2|x_2) \sum_{x_1} p(x_2|x_1) \alpha_{1|1}(x_1) \\
&= \sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \sum_{x_2} p(x_3|x_2)p(y_2|x_2) \alpha_{2|1}(x_2) \\
&= \sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \sum_{x_2} p(x_3|x_2) \alpha_{2|2}(x_2) \\
&= \sum_{x_K} p(y_K|x_K) \sum_{x_{K-1}} p(x_K|x_{K-1}) \cdots \alpha_{3|2}(x_3)
\end{aligned}$$

### 3. Çıkarım: Geri faz

```
log_alpha = zeros(N, K);
log_alpha_predict = zeros(N, K);
for k=1:K,
    if k==1,
        log_alpha_predict(:,k) = log(hm.p_x1);
    else
        log_alpha_predict(:,k) ...
            = state_predict(hm.A, log_alpha(:, k-1));
    end;
    log_alpha(:, k) ...
        = state_update(hm.C(y(k), :), log_alpha_predict(:,k));
end;
```

### 3. Çıkarım: Tahminleme

```
function [lpp] = state_predict(A, log_p)
% STATE_PREDICT Computes A*p in log domain
%
% [lpp] = state_predict(A, log_p)
%
% Inputs :
% A : State transition matrix
% log_p : log p(x_{k-1}, y_{1:k-1}) Filtered potential
%
% Outputs :
% lpp : log p(x_{k}, y_{1:k-1}); Predicted potential

mx = max(log_p(:)); % Stable computation
p = exp(log_p - mx);
lpp = log(A*p) + mx;
```



## Numerik hesaplama $\log(\sum_i \exp(l_i))$

- Türetim

$$\begin{aligned} L &= \log\left(\sum_i \exp(l_i)\right) \\ &= \log\left(\sum_i \exp(l_i) \frac{\exp(l^*)}{\exp(l^*)}\right) \\ &= \log\left(\exp(l^*) \sum_i \exp(l_i - l^*)\right) \\ &= l^* + \log\left(\sum_i \exp(l_i - l^*)\right) \end{aligned}$$

- $l^* = \max_i l_i$

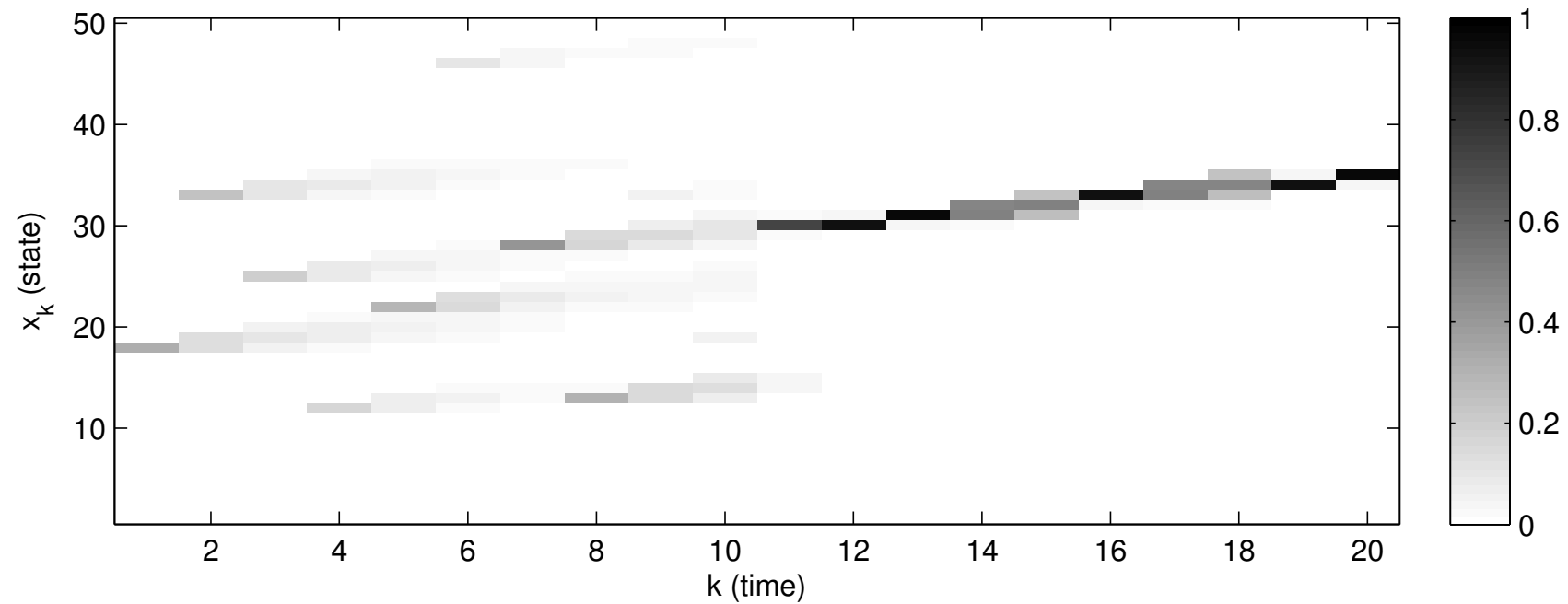
### 3. Çıkarım: Güncelleme adımı

```
function [lup] = state_update(obs, log_p)
% STATE_UPDATE State update in log domain
%
% [lup] = state_update(obs, log_p)
%
% Inputs :
%         obs :  $p(y_k | x_k)$ 
%         log_p :  $\log p(x_k, y_{\{1, k-1\}})$ 
%
% Outputs :
% lup :  $\log p(x_k, y_{\{1, k-1\}}) + p(y_k | x_k)$ 

lup = log(obs(:)) + log_p;
```

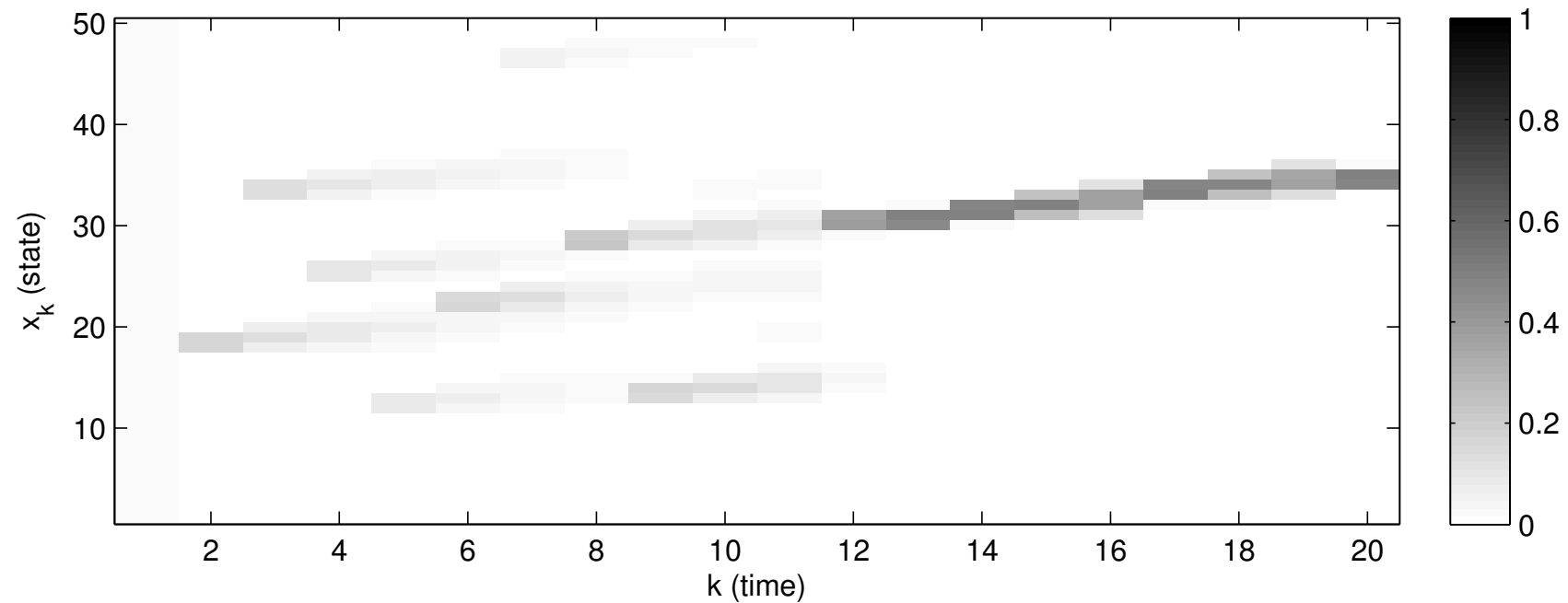
### 3. Çıkarım: İleriye doğru

$$\alpha_{k|k} \equiv p(y_{1:k}, x_k)$$

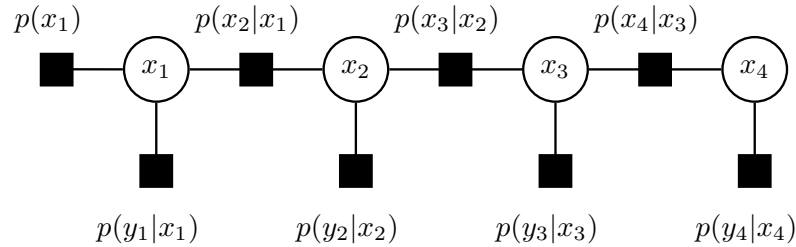


### 3. Çıkarım: İleriye doğru

$$\alpha_{k|k-1} \equiv p(y_{1:k-1}, x_k)$$



### 3. Çıkarım: Geriye doğru



- “Postdict”

$$\begin{aligned}\beta_{k|k+1}(x_k) &= p(y_{k+1:K}|x_k) = \sum_{x_{k+1}} p(x_{k+1}|x_k)p(y_{k+1:K}|x_{k+1}) \\ &= \sum_{x_{k+1}} p(x_{k+1}|x_k)\beta_{k+1|k+1}(x_{k+1})\end{aligned}$$

- Güncelleme

$$\begin{aligned}\beta_{k|k}(x_k) &= p(y_{k:K}|x_k) = p(y_k|x_k)p(y_{k+1:K}|x_k) \\ &= p(y_k|x_k)\beta_{k|k+1}(x_k)\end{aligned}$$

$$\begin{aligned}
p(y_{1:K}) &= \sum_{x_1} p(x_1)p(y_1|x_1) \cdots \sum_{x_{K-1}} p(x_{K-1}|x_{K-2})p(y_{K-1}|x_{K-1}) \sum_{x_K} p(x_K|x_{K-1})p(y_K|x_K) \underbrace{1}_{\beta_{K|K+1}} \\
&= \sum_{x_1} p(x_1)p(y_1|x_1) \cdots \sum_{x_{K-1}} p(x_{K-1}|x_{K-2})p(y_{K-1}|x_{K-1}) \sum_{x_K} p(x_K|x_{K-1}) \beta_{K|K} \\
&= \sum_{x_1} p(x_1)p(y_1|x_1) \cdots \sum_{x_{K-1}} p(x_{K-1}|x_{K-2})p(y_{K-1}|x_{K-1}) \beta_{K-1|K} \\
&= \sum_{x_1} p(x_1)p(y_1|x_1) \cdots \sum_{x_{K-1}} p(x_{K-1}|x_{K-2}) \beta_{K-1|K-1} \\
&= \sum_{x_1} p(x_1)p(y_1|x_1) \cdots \beta_{K-2|K-1}
\end{aligned}$$

### 3. Çıkarım: Geriye doğru

```
log_beta = zeros(N, T);
log_beta_postdict = zeros(N, T);
for t=T:-1:1,
    if t==T,
        log_beta_postdict(:,t) = zeros(N,1);
    else
        log_beta_postdict(:,t) ...
            = state_postdict(hm.A, log_beta(:, t+1));
    end;
    log_beta(:, t) ...
        = state_update(hm.C(y(t), :), log_beta_postdict(:,t));
end;
```

### 3. Çıkarım: Geriye doğru

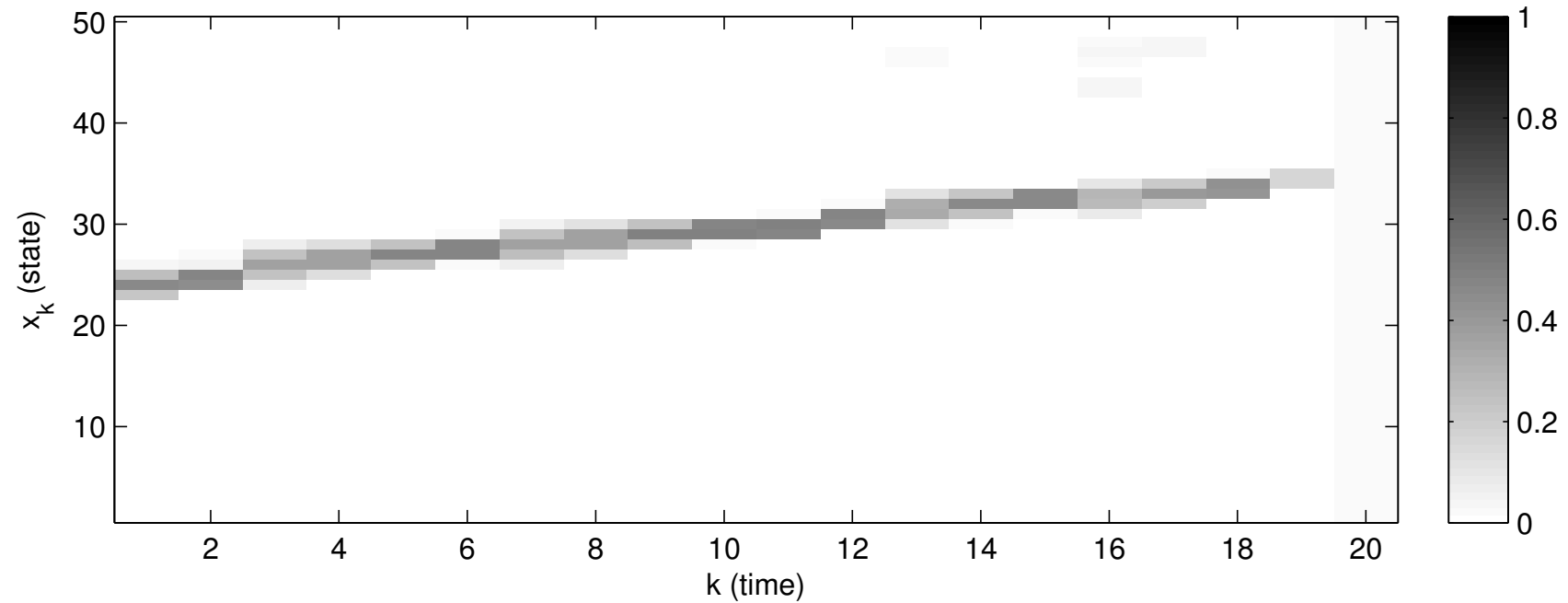
```
function [lpp] = state_postdict(A, log_p)
% STATE_POSTDICT Computes  $A' * p$  in log domain
%
% [lpp] = state_postdict(A, log_p)
%
% Inputs :
% A : State transition matrix
%          log_p :  $\log p(y_{k+1:K} | x_{k+1})$  Updated potential
%
% Outputs :
% lpp :  $\log p(y_{k+1:K} | x_k)$  Postdicted potential

mx = max(log_p(:)); % Stable computation
p = exp(log_p - mx);
lpp = log(A' * p) + mx;
```



### 3. Çıkarım: Geriye doğru

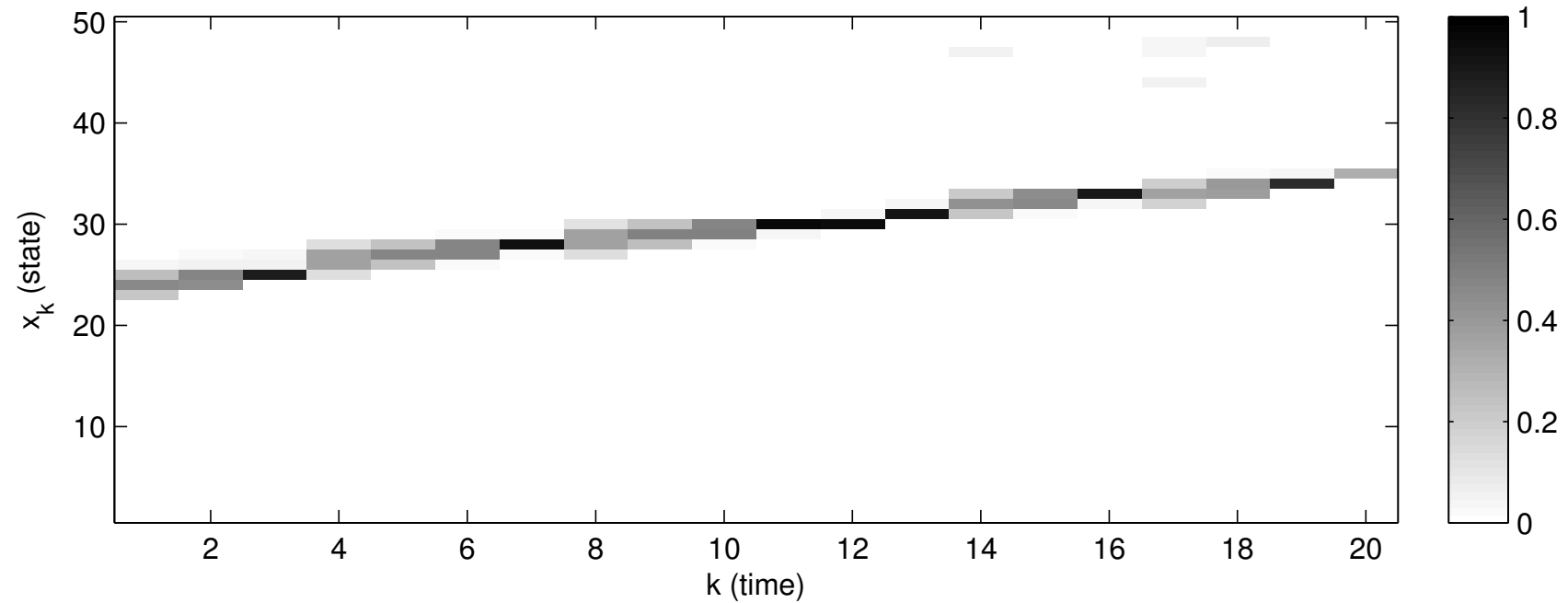
$$\beta_{k|k+1}(x_k) = p(y_{k+1:K}|x_k)$$



We visualise  $\hat{\beta} \propto \beta_{k|k+1}(x_k)u(x_k)$

### 3. Çıkarım: Geriye doğru

$$\beta_{k|k}(x_k) = p(y_{k:K}|x_k)$$



### 3. Çıkarım: Smoothing

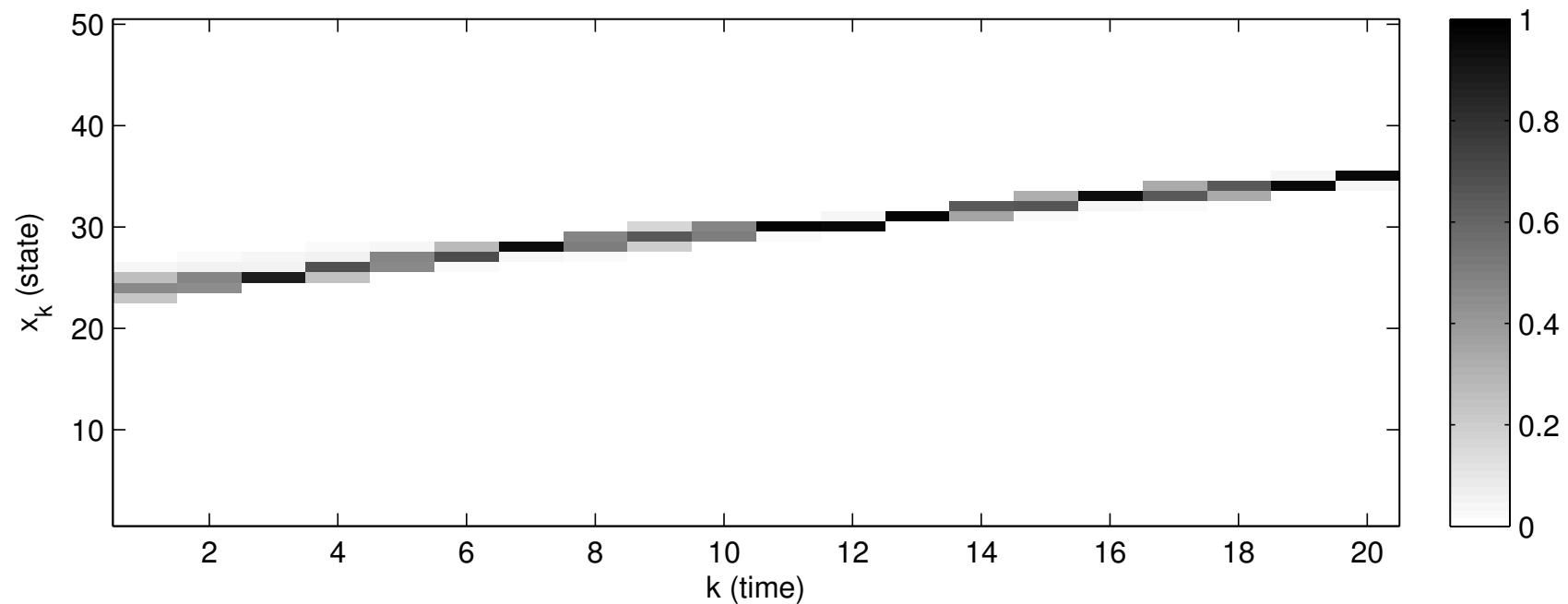
$$\begin{aligned} p(y_{1:K}, x_k) &= p(y_{1:k}, x_k) p(y_{k+1:K} | x_k) \\ &= \alpha_{k|k}(x_k) \beta_{k|k+1}(x_k) \\ &\equiv \gamma_k(x_k) \end{aligned}$$

Alternatif

$$\begin{aligned} \gamma_k(x_k) &= \alpha_{k|k-1}(x_k) \beta_{k|k}(x_k) \\ &= \alpha_{k|k-1}(x_k) p(y_k | x_k) \beta_{k|k+1}(x_k) \end{aligned}$$

### 3. Çıkarım: Smoothing

$$p(x_k|y_{1:K}) \propto p(y_{1:K}, x_k) = \alpha_{k|k}(x_k)\beta_{k|k+1}(x_k) \equiv \gamma_k(x_k)$$



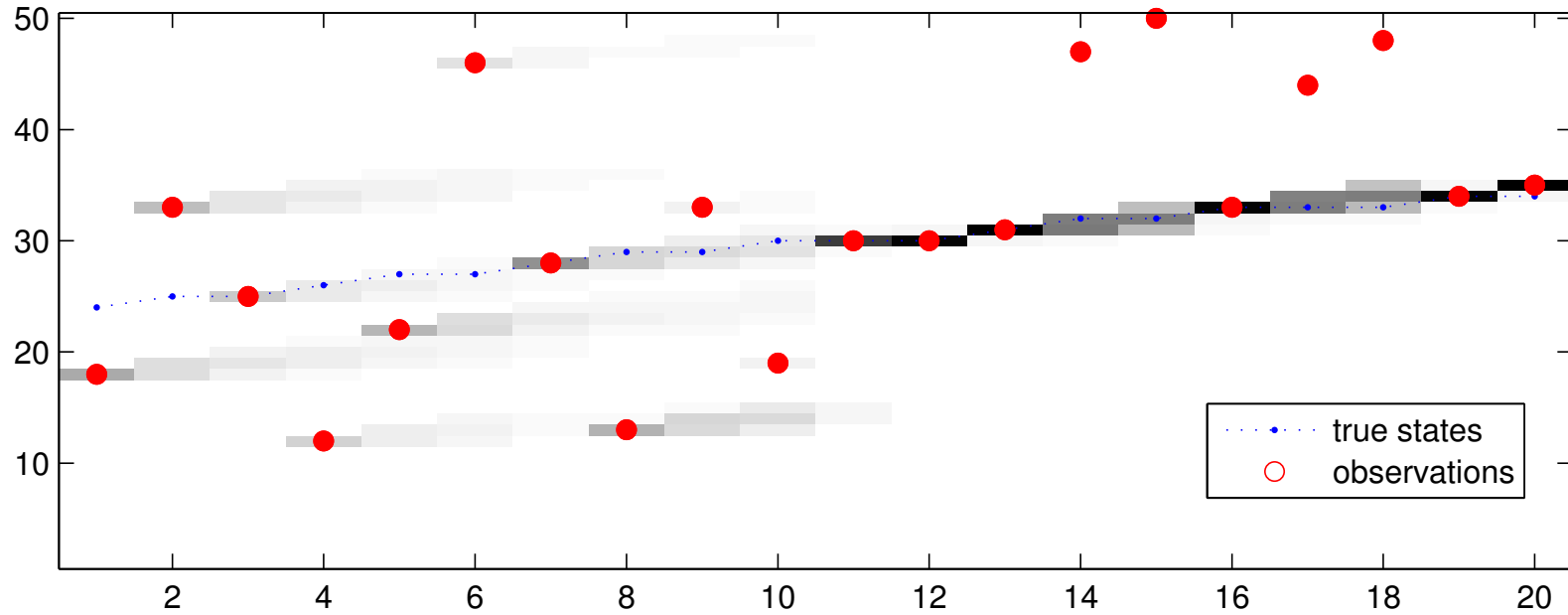
### 3. Çıkarım: Smoothing

`log_gamma = log_alpha + log_beta_postdict`

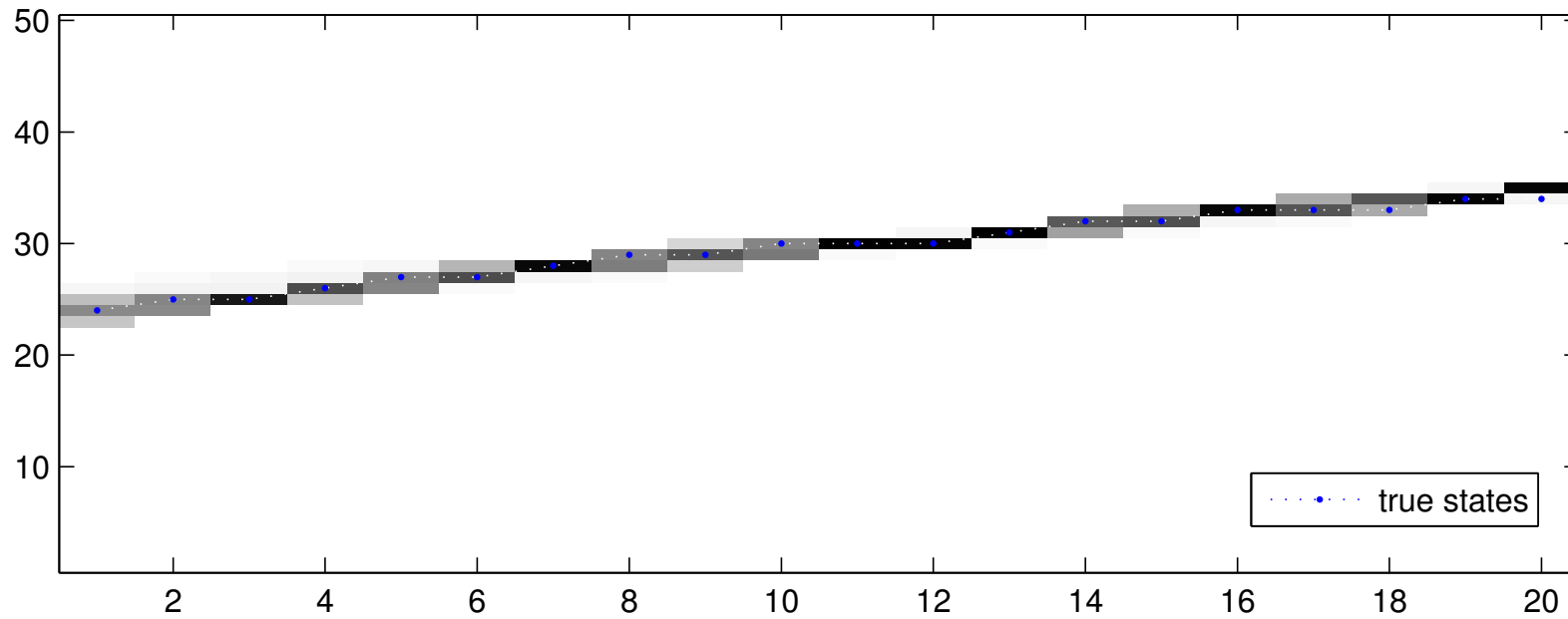
## 4. Test ve Görselleştirme

```
imagesc(normalize_exp(log_gamma, 1));  
set(gca, 'ydir', 'n');  
colormap(flipud(gray));  
xlabel('k (time)'); ylabel('x_k (state)');  
caxis([0 1]);  
colorbar  
  
% This has to be constant !! (why)  
plot(log_sum_exp(log_gamma, 1));
```

## 4. Test ve Görselleştirme. Süzgeç



## 4. Test ve Görselleştirme. Smoother.





# Özet

- Bayesci Çıkarım
- Ortalama alan (Mean Field), Varyasyonel Bayes

# Varyasyonel Formülasyon

Basit ama kuvvetli bir fikir:

- Çözümü bir eniyileme probleminin çözümü olarak göster
- Örnek: Doğrusal denklem çözümü  $p \in \mathcal{X}$

$$Ap = b$$

- Varyasyonel formülasyon

$$p = \operatorname{argmin}_q \underbrace{\left\{ \frac{1}{2} (b - Aq)^\top (b - Aq) \right\}}_{\mathcal{F}(q)}$$

# Varyasyonel Formülasyon

- Yaklaşık çözümler de üretilebilir
- Kısıtları arttırırsak

$$q \in \mathcal{X}_q \subset \mathcal{X}$$

- Basit bir gözlem

$$\mathcal{F}(p) = \min_{q \in \mathcal{X}} \{\mathcal{F}(q)\} \leq \min_{q \in \mathcal{X}_q} \{\mathcal{F}(q)\}$$

## Örnek: Marjinal hesabı

- Birleşik dağılım  $i, j \in \{0, 1\}$

$$p(x_1 = i, x_2 = j) = \pi_{i,j}$$

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	$\pi_{0,0}$	$\pi_{0,1}$
$x_1 = 1$	$\pi_{1,0}$	$\pi_{1,1}$

- Marjinaller

$p(x_1)$	
$x_1 = 0$	$\pi_{0,0} + \pi_{0,1}$
$x_1 = 1$	$\pi_{1,0} + \pi_{1,1}$

$p(x_2)$	$x_2 = 0$	$x_2 = 1$
	$\pi_{0,0} + \pi_{1,0}$	$\pi_{0,1} + \pi_{1,1}$

- Bir dağılımın marjinallerini varyasyonel olarak nasıl gösterebiliriz?

# Marjinal hesabı

- Faktörize olmuş bir dağılım alalım

$$q(x_1 = i, x_2 = j) = q(x_1 = i)q(x_2 = j)$$

$$q(x_1 = 1) = q_1$$

$$q(x_2 = 1) = q_2$$

$q(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	$(1 - q_1)(1 - q_2)$	$(1 - q_1)q_2$
$x_1 = 1$	$q_1(1 - q_2)$	$q_1q_2$

- $p$  ve  $q$  arasındaki uzaklığı **Kullback-Leibler (KL) İraksayı** ile hesaplayalım

# Kullback-Leibler (KL) İraksayı

- $\mathcal{P} = p(x)$  ve  $\mathcal{Q} = q(x)$  olan iki olasılık dağılımı arasında “quasi-distance” (uzaklıkımsı - çakma mesafe).

$$KL(\mathcal{P}||\mathcal{Q}) \equiv \int_{\mathcal{X}} dx p(x) \log \frac{p(x)}{q(x)} = \langle \log \mathcal{P} \rangle_{\mathcal{P}} - \langle \log \mathcal{Q} \rangle_{\mathcal{P}}$$

- Simetrik değil

$$KL(\mathcal{P}||\mathcal{Q}) \neq KL(\mathcal{Q}||\mathcal{P})$$

- Ama pozitif (by Jensen's Inequality)

$$\begin{aligned} KL(\mathcal{P}||\mathcal{Q}) &= - \int_{\mathcal{X}} dx p(x) \log \frac{q(x)}{p(x)} \\ &\geq - \log \int_{\mathcal{X}} dx p(x) \frac{q(x)}{p(x)} = - \log \int_{\mathcal{X}} dx q(x) = - \log 1 = 0 \end{aligned}$$

## Kullback-Leibler (KL) İraksayı

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$q(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	$\pi_{0,0}$	$\pi_{0,1}$	$x_1 = 0$	$(1 - q_1)(1 - q_2)$	$(1 - q_1)q_2$
$x_1 = 1$	$\pi_{1,0}$	$\pi_{1,1}$	$x_1 = 1$	$q_1(1 - q_2)$	$q_1q_2$

$$\begin{aligned}
 KL(p||q) &= \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log \left( \frac{p(x_1, x_2)}{q(x_1, x_2)} \right) \\
 &= \sum_i \sum_j \pi_{i,j} \log \left( \frac{\pi_{i,j}}{q(x_1 = i, x_2 = j)} \right) \\
 &= \pi_{0,0} \log \left( \frac{\pi_{0,0}}{(1 - q_1)(1 - q_2)} \right) + \pi_{1,0} \log \left( \frac{\pi_{1,0}}{q_1(1 - q_2)} \right) \\
 &\quad + \pi_{0,1} \log \left( \frac{\pi_{0,1}}{(1 - q_1)q_2} \right) + \pi_{1,1} \log \left( \frac{\pi_{1,1}}{q_1q_2} \right)
 \end{aligned}$$

# Kullback-Leibler (KL) İraksayı

- KL ıraksayısını  $q_1$ 'e göre en küçült

$$\begin{aligned} KL(p||q) &= -\pi_{0,0}(\log(1 - q_1) + \log(1 - q_2)) - \pi_{1,0}(\log q_1 + \log(1 - q_2)) \\ &\quad - \pi_{0,1}(\log(1 - q_1) + \log q_2) - \pi_{1,1}(\log q_1 + \log q_2) \\ &\quad + \sum_i \sum_j \pi_{i,j} \log \pi_{i,j} \end{aligned}$$

- Türevi alıp sıfıra eşitleyelim

$$\frac{\partial KL(p||q)}{\partial q_1} = \frac{\partial}{\partial q_1} (-\pi_{0,0} \log(1 - q_1) - \pi_{1,0} \log q_1 - \pi_{0,1} \log(1 - q_1) - \pi_{1,1} \log q_1)$$



## Marjinal dağılım $KL(p||q)$

$$\begin{aligned} 0 &= \pi_{0,0} \frac{1}{(1-q_1)} - \pi_{1,0} \frac{1}{q_1} + \pi_{0,1} \frac{1}{(1-q_1)} - \pi_{1,1} \frac{1}{q_1} \\ &= (\pi_{0,0} + \pi_{0,1}) \frac{1}{(1-q_1)} - (\pi_{1,0} + \pi_{1,1}) \frac{1}{q_1} \end{aligned}$$

$$q_1 = \frac{(\pi_{1,0} + \pi_{1,1})}{(\pi_{0,0} + \pi_{0,1} + \pi_{1,0} + \pi_{1,1})} = \pi_{1,0} + \pi_{1,1} = p(x_1 = 1)$$

$$1 - q_1 = 1 - (\pi_{1,0} + \pi_{1,1}) = \pi_{0,0} + \pi_{0,1} = 1 - q_1 = p(x_1 = 0)$$

The derivation for  $q_2$  is identical.

## Öteki: $KL(q||p)$

$$\begin{aligned} KL(q||p) &= \sum_{x_1} \sum_{x_2} q(x_1, x_2) \log \left( \frac{q(x_1, x_2)}{p(x_1, x_2)} \right) \\ &= \sum_i \sum_j q(x_1 = i, x_2 = j) \log \left( \frac{q(x_1 = i, x_2 = j)}{\pi_{i,j}} \right) \\ &= (1 - q_1)(1 - q_2) \log \left( \frac{(1 - q_1)(1 - q_2)}{\pi_{0,0}} \right) + q_1(1 - q_2) \log \left( \frac{q_1(1 - q_2)}{\pi_{1,0}} \right) \\ &\quad + (1 - q_1)q_2 \log \left( \frac{(1 - q_1)q_2}{\pi_{0,1}} \right) + q_1q_2 \log \left( \frac{q_1q_2}{\pi_{1,1}} \right) \end{aligned}$$

## Öteki: $KL(q||p)$

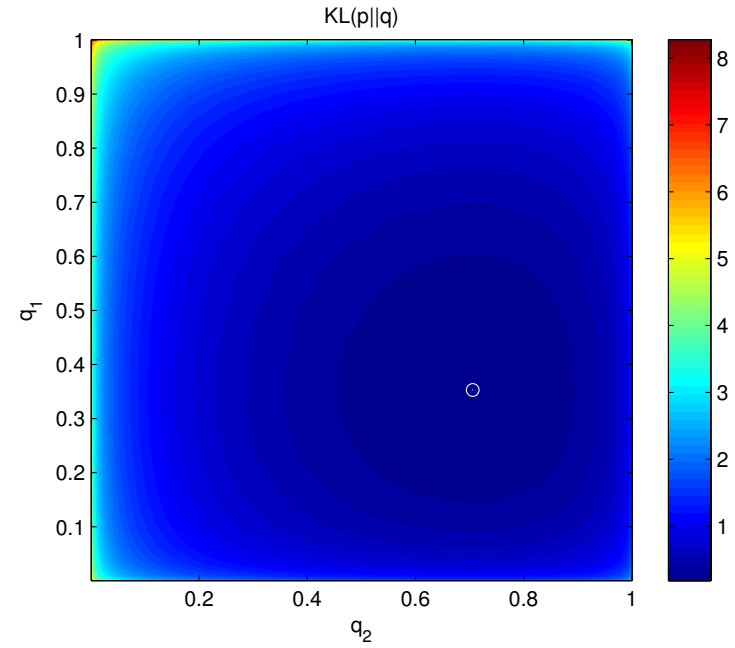
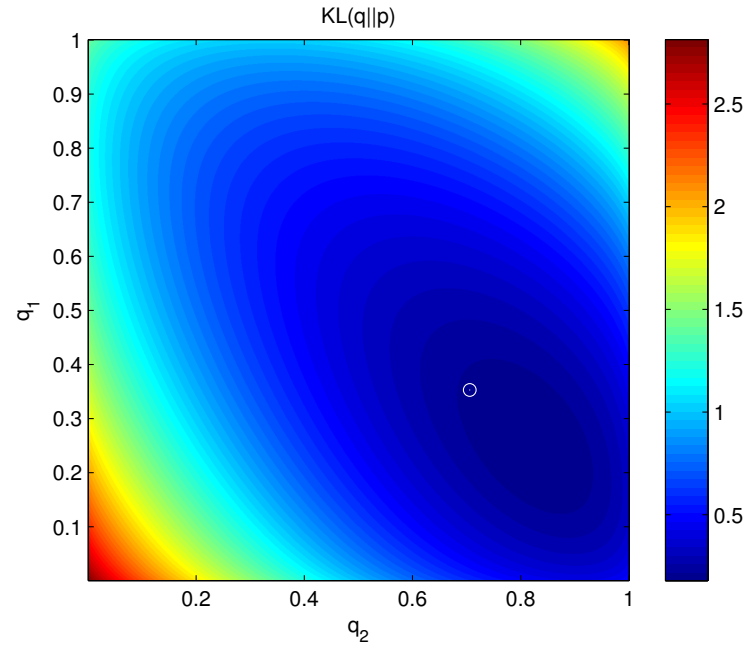
$$\frac{\partial KL(q||p)}{\partial q_1} = (-\log(1 - q_1) + \log \pi_{0,0} + \log q_1 - \log \pi_{1,0}) \\ q_2 (-\log \pi_{0,0} + \log \pi_{1,0} + \log \pi_{0,1} - \log \pi_{1,1})$$

## Öteki: $KL(q||p)$

$$\begin{aligned} \mathcal{Q}_1 &= \begin{pmatrix} 1 - q_1 \\ q_1 \end{pmatrix} = \frac{1}{Z_1} \begin{pmatrix} \pi_{0,0}^{(1-q_2)} \pi_{0,1}^{q_2} \\ \pi_{1,0}^{(1-q_2)} \pi_{1,1}^{q_2} \end{pmatrix} \\ &\propto \begin{pmatrix} \exp((1 - q_2) \log \pi_{0,0} + q_2 \log \pi_{0,1}) \\ \exp((1 - q_2) \log \pi_{1,0} + q_2 \log \pi_{1,1}) \end{pmatrix} \\ &= \begin{pmatrix} \exp((1 - q_2) \log \pi_{0,0} + q_2 \log \pi_{0,1}) \\ \exp((1 - q_2) \log \pi_{1,0} + q_2 \log \pi_{1,1}) \end{pmatrix} \\ &\equiv \exp(\langle \log \pi \rangle_{\mathcal{Q}_2}) \end{aligned}$$

$$\mathcal{Q}_2 \propto \exp(\langle \log \pi \rangle_{\mathcal{Q}_1})$$

## $KL(q||p)$ ve $KL(p||q)$ karşılaştırma



# Varyasyonel Bayes (VB), Ortalama alan (mean field)

We will approximate the posterior  $\mathcal{P}$  with a simpler distribution  $\mathcal{Q}$ .

$$\begin{aligned}\mathcal{P} &= \frac{1}{Z_x} p(x = \hat{x} | s_1, s_2) p(s_1) p(s_2) \\ \mathcal{Q} &= q(s_1) q(s_2)\end{aligned}$$

Here, we choose

$$q(s_1) = \mathcal{N}(s_1; m_1, S_1) \quad q(s_2) = \mathcal{N}(s_2; m_2, S_2)$$

A “measure of fit” between distributions is the KL divergence

# Kullback-Leibler (KL) İraksayı

- A “quasi-distance” between two distributions  $\mathcal{P} = p(x)$  and  $\mathcal{Q} = q(x)$ .

$$KL(\mathcal{P}||\mathcal{Q}) \equiv \int_{\mathcal{X}} dx p(x) \log \frac{p(x)}{q(x)} = \langle \log \mathcal{P} \rangle_{\mathcal{P}} - \langle \log \mathcal{Q} \rangle_{\mathcal{P}}$$

- Unlike a metric, (in general) it is not symmetric,

$$KL(\mathcal{P}||\mathcal{Q}) \neq KL(\mathcal{Q}||\mathcal{P})$$

- But it is non-negative (by Jensen’s Inequality)

$$\begin{aligned} KL(\mathcal{P}||\mathcal{Q}) &= - \int_{\mathcal{X}} dx p(x) \log \frac{q(x)}{p(x)} \\ &\geq - \log \int_{\mathcal{X}} dx p(x) \frac{q(x)}{p(x)} = - \log \int_{\mathcal{X}} dx q(x) = - \log 1 = 0 \end{aligned}$$

## Ortalama Alan Çözümü

$$\begin{aligned} 0 &\leq \langle \log q(s_1)q(s_2) \rangle_{q(s_1)q(s_2)} + \log Z_x - \langle \log \phi(s_1, s_2) \rangle_{q(s_1)q(s_2)} \\ \log Z_x &\geq \langle \log \phi(s_1, s_2) \rangle_{q(s_1)q(s_2)} - \langle \log q(s_1)q(s_2) \rangle_{q(s_1)q(s_2)} \\ &\equiv -F(p; q) + H(q) \end{aligned} \tag{2}$$

$F$  enerji ve  $H$  entropi. Sağ tarafı en büyükle.

$$\text{Evidence} \geq -\text{Enerji} + \text{Entropi}$$

Sağ taraf bir **alt sınır** [?]. Ortalama alan eşitlikleri **monotonik** olarak bu sınırı arttırıyorlar.



# Detaylar

- Lagrang fonksyonu

$$\begin{aligned}\Lambda = & \int ds_1 q(s_1) \log q(s_1) + \int ds_2 q(s_2) \log q(s_2) + \log Z_x - \int ds_1 ds_2 q(s_1) q(s_2) \log \phi(s_1, s_2) \\ & + \lambda_1(1 - \int ds_1 q(s_1)) + \lambda_2(1 - \int ds_2 q(s_2))\end{aligned}\quad (3)$$

- $q(s_1)$ 'e göre fonksiyonel türevleri hesapla ve sıfır'a eşitle

$$\frac{\delta}{\delta q(s_1)} \Lambda = \log q(s_1) + 1 - \langle \log \phi(s_1, s_2) \rangle_{q(s_2)} - \lambda_1$$

- $q(s_1)$  için çöz,

$$\begin{aligned}\log q(s_1) &= \lambda_1 - 1 + \langle \log \phi(s_1, s_2) \rangle_{q(s_2)} \\ q(s_1) &= \exp(\lambda_1 - 1) \exp(\langle \log \phi(s_1, s_2) \rangle_{q(s_2)})\end{aligned}\quad (4)$$

- Kısıt'ı kullan

$$\begin{aligned}1 &= \int ds_1 q(s_1) = \exp(\lambda_1 - 1) \int ds_1 \exp(\langle \log \phi(s_1, s_2) \rangle_{q(s_2)}) \\ \lambda_1 &= 1 - \log \int ds_1 \exp(\langle \log \phi(s_1, s_2) \rangle_{q(s_2)})\end{aligned}$$

# Çözümün yapısı

- Analitik çözüm yok
- Sabit nokta denklemleri

$$q(s_1) \propto \exp(\langle \log \phi(s_1, s_2) \rangle_{q(s_2)})$$

$$q(s_2) \propto \exp(\langle \log \phi(s_1, s_2) \rangle_{q(s_1)})$$

Simetriye dikkat

# Bayesci çıkarım - Olabilirlik Enbüyültme

Algoritmalar benzer olmasına rağmen bulunan sonuçların yapısı çok farklı olabilir

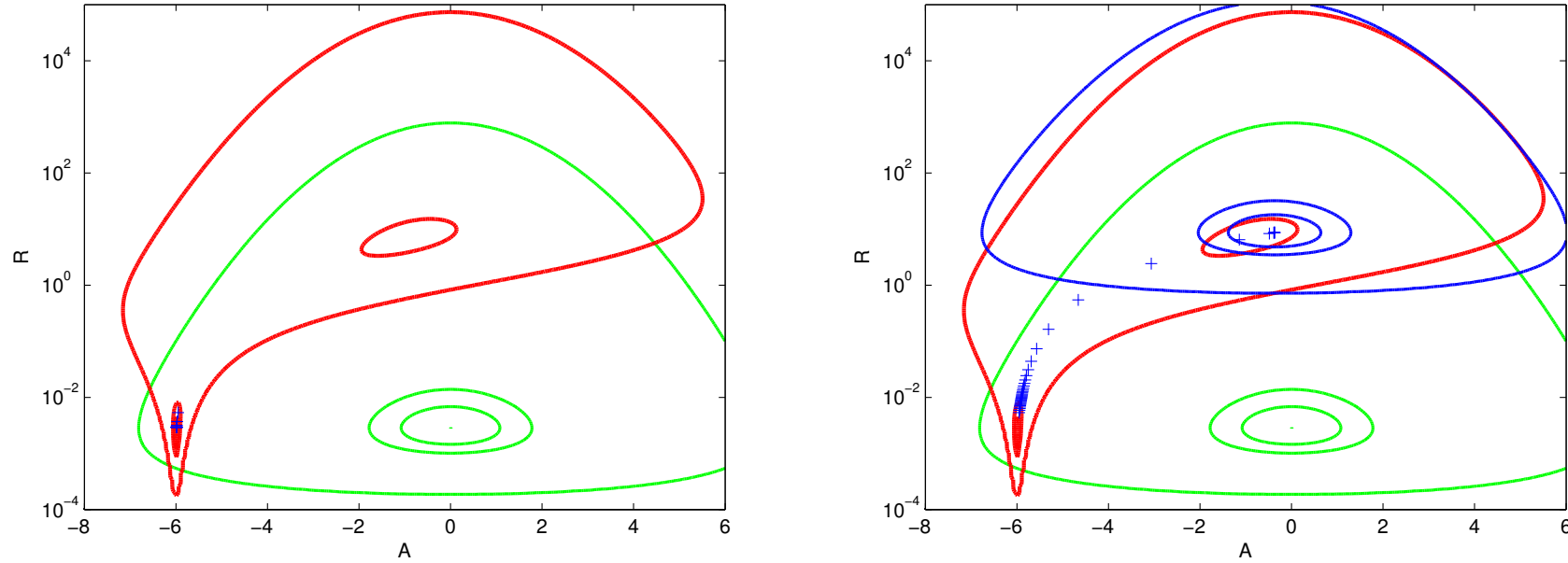


Figure 1: Sol: ML, Sağ: VB.

# Referanslar

## Text Books:

- Bayesian Reasoning and Machine Learning, David Barber, 2012, CUP Online
- Pattern Recognition and Machine Learning, Christopher Bishop, 2006 Springer
- Machine Learning, A Probabilistic Perspective, Kevin P. Murphy, 2012 MIT Press

# Referanslar

## Bayesci Zaman Serileri, Monte Carlo

- A. T. Cemgil, A Tutorial Introduction to Monte Carlo methods, Markov Chain Monte Carlo and Particle Filtering, 2012. (<https://dl.dropboxusercontent.com/u/9787379/cmpe58n/cmpe58n-lecture-notes.pdf>)
- D. Barber, A. T. Cemgil and S. Chiappa, Bayesian Time Series Models. Cambridge University Press, 2011.
- D Barber and A. T. Cemgil, Graphical Models for Time Series, IEEE Signal Processing Magazine, Special issue on graphical models, vol. 27, no. 6, pp. 18-28, October 2010.

## Referanslar

M. J. Wainwright and M. I. Jordan, 2008, Graphical Models, Exponential Families, and Variational Inference, Foundations and Trends in Machine Learning, DOI: 10.1561/22000000001

# Referanslar

## Recent Trends

- Z. Ghahramani, Probabilistic machine learning and artificial intelligence, Nature, 2015, doi:10.1038/nature14541
  - Olasılıksal programlama,
  - Bayesci eniyileme,
  - Veri sıkıştırma
  - Otomatik model keşfetme