

İkili bir değişkenin Bayesian modellenmesi

X ikili bir random değişken olsun, yani alabileceği değerler 0 ve 1. Ne değer alacağını önceden bilmiyoruz. Ancak bir eylem veya olay sonucunda 0 veya 1 değerini alıyor ve ancak o zaman gözlemleyebiliyoruz.

Frequentist yaklaşım ve Bayesian yaklaşım

Eğer bu olayı aynı şartlarda defalarca tekrarlayabilseydik şöyle bir dizi elde edecektik:

$$0, 0, 1, 0, 1, 1, 0, 0, 0, 1$$

Daha sonra dizideki 0 ve 1'lerin sıklığına göre $P(X = 0)$ ve $P(X = 1)$ olasılıklarını hesaplayabilirdik. Toplamı 1 olan bu iki olasılık değeri de $P(X)$ dağılımını verirdi. Buna “Frequentist” yaklaşım deniyor. Bu durumda varsayımlardan kaçınarak tamamen geçmiş gözlemlerden gelen bilgiyle davranıyoruz.

Ama birçok meselede bu mümkün olmuyor. Her seferinde farklı şartlarla karşılaşıyoruz ve fazla tekrarlama, deneme yanılma şansımız da olmuyor. Aynı nehirde tekrar yıkanamıyoruz. O zaman da sezgiye ihtiyaç duyuyoruz. Henüz olmamış şeylere, daha gerçekleştirilmemiş eylemlere dair önsel (prior) varsayımlar yapıyoruz. Buna da “Bayesian” yaklaşım deniyor.

Bernoulli dağılımı: 0 mı, 1 mi?

İkili değişken olan X 'i bir Bernoulli dağılımı olarak modelleyebiliriz:

$$X \mid \pi \sim \text{Bernoulli}(X ; \pi)$$

$$P(X \mid \pi) = \text{Bernoulli}(X ; \pi) = \pi^X (1 - \pi)^{(1-X)}$$

Modelde verilen $P(X|\pi)$ dağılımı, parametre (π) verildiğinde gözlemlerin (X) dağılımıdır. Bu yüzden “likelihood” olarak adlandırılır.

Bernoulli dağılımının tek parametresi olan π , $[0,1]$ aralığında bir reel sayıdır ve X 'in 1 olma olasılığını verir:

$$P(X = 1 \mid \pi) = \pi$$

$$P(X = 0 \mid \pi) = 1 - \pi$$

Beta dağılımı: π 'ye dair inancımız

“Full Bayesian treatment” yapmak, gözlemleri etkileyen parametreler için önsel dağılımlar tanımlamak demektir. Bu modeldeki tek parametre olan π 'yi Beta dağılımı olarak modelleyebiliriz:

$$\pi \sim \text{Beta}(\pi ; a, b)$$

$$P(\pi) = \text{Beta}(\pi ; a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \pi^{(a-1)} (1-\pi)^{(b-1)}$$

Modelde verilen $P(\pi)$ dağılımı, bir parametrenin bağımsız dağılımını gösterdiğinden “prior” olarak adlandırılır. Beta dağılımı, pozitif değerler alan iki parametresi a ve b ile $[0, 1]$ aralığında tepe, çukur, yamaç şekilleri alabilir. Bu şekilleri kullanarak, henüz gerçekleşmemiş bir olayın (X) sonucuna dair sezgilerimizi ifade edebiliriz. Bu şekilleri incelemeden önce a ve b 'nin ne anlama geldiğine bakalım.

Conjugate prior ve tecrübe birikimi

Beta dağılımını seçmemizin özel bir sebebi var. π 'nin “posterior” dağılımına bakalım:

$$P(\pi | X) = \frac{P(X|\pi) P(\pi)}{P(X)} \propto P(X | \pi) P(\pi)$$

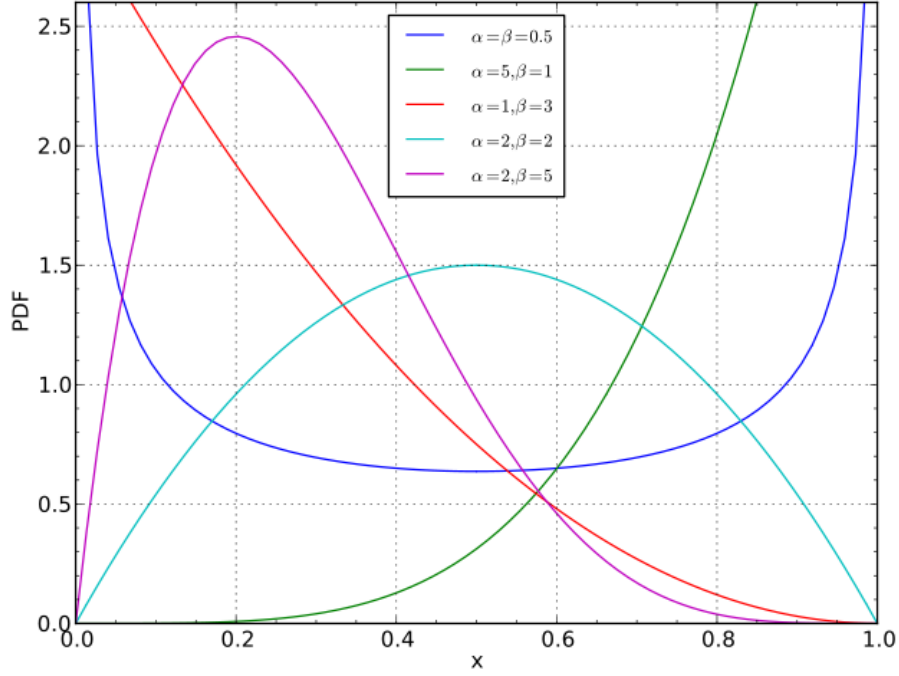
$$\begin{aligned} P(X | \pi) P(\pi) &= \text{Bernoulli}(X ; \pi) \text{Beta}(\pi ; a, b) \\ &= \pi^X (1-\pi)^{(1-X)} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \pi^{(a-1)} (1-\pi)^{(b-1)} \\ &= K \pi^{(a+X-1)} (1-\pi)^{(b+(1-X)-1)} \end{aligned}$$

$$P(\pi | X) = \text{Beta}(\pi ; a + X, b + (1 - X))$$

Bernoulli'nin parametresi olan π 'ye prior olarak Beta seçildiğinde, π 'nin posterior dağılımının da Beta olduğunu gösterdik. Bu yüzden Beta dağılımı Bernoulli'nin “conjugate prior”ıdır diyoruz ve işlem kolaylığı açısından bu prior'ları tercih ediyoruz.

Posterior'daki parametrelere baktığımızda, gözlem sonucunda eğer $X = 1$ ise $a \rightarrow a + 1$ olduğunu, $X = 0$ ise $b \rightarrow b + 1$ olduğunu görüyoruz. Demek ki yeni gözlemler sonunda yenileyeceğimiz posterior'daki a ve b her seferinde birer birer artacak. Yani a ve b parametreleri, $X = 1$ ve $X = 0$ olaylarına dair tecrübelerimizi biriktirirler. Bunlara “pseudocount” parametre deniyor.

Beta dağılımının aldığı şekiller

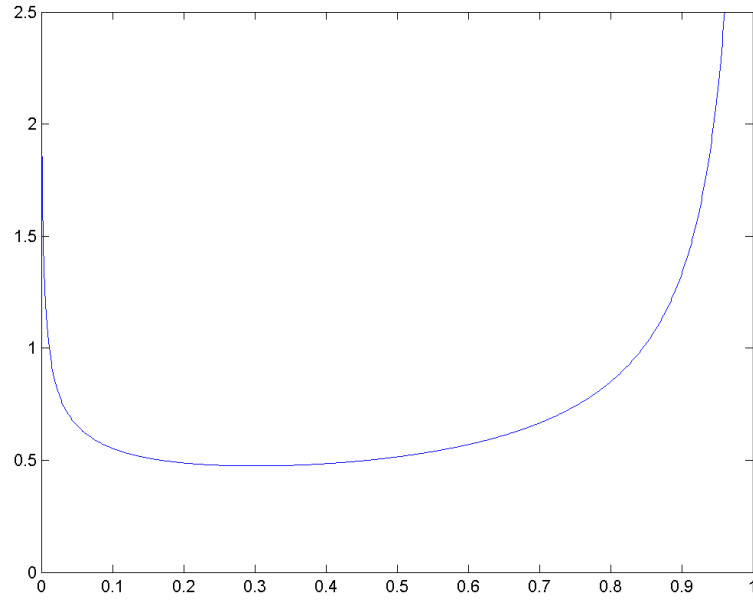


Eğer a ve b 'nin ikisi de 1'den küçükse, iki tarafı yamaç gibi yükselen bir çukur vardır. Yani ya hep $X = 0$ olacağına veya hep $X = 1$ olacağına inanırız. İkisi de 1 olduğunda dümdüz bir ova, yani her şeye eşit mesafede olduğumuz tam bir kararsızlık oluşur. Eğer $b = 1$ iken a değeri 1'i aşarsa, inancımız $\pi = 1$ etrafında yoğunlaşır. Yok eğer $a = 1$ iken b değeri 1'i aşarsa inancımız $\pi = 0$ etrafında yoğunlaşır. π 'nin $X = 1$ olma olasılığı olduğunu hatırlayalım.

İkisi de 1'den büyük olduğunda belirli bir π etrafında bir tepe oluşur. Yani tek merkezli düzenli bir inanç yerleşir. İki parametreden a daha büyükse tepe sağa yatıktır ve $X = 1$ 'i daha olası görürüz. b daha büyükse tepe sola yatıktır ve $X = 0$ 'ı daha olası görürüz. Eğer a ve b görece küçükse, tepe yayvan olur ve π 'ye dair inancımız henüz zayıftır, kararsızdır. a ve b büyüdükçe tepe sivrileşir ve inancımız belirli bir π değeri etrafında birikir ve netleşir, giderek daha kararlı olur.

Örnek problem

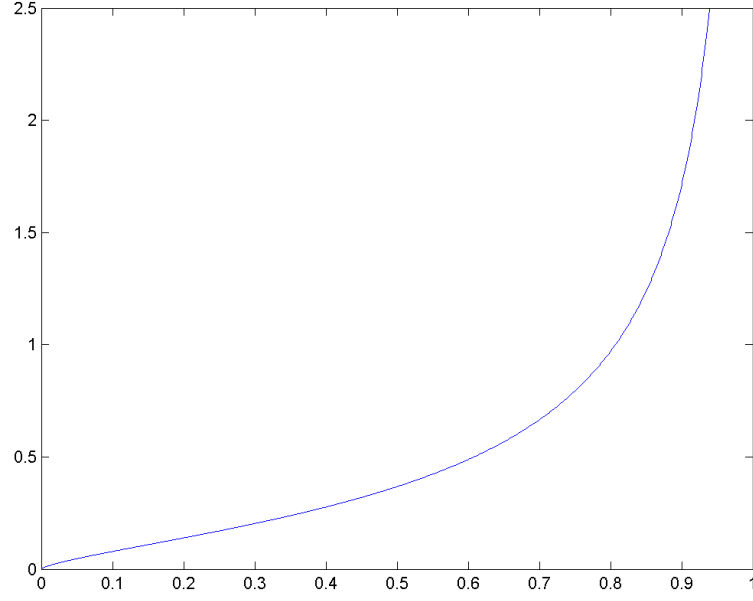
Beta dağılımının parametreleri olarak en başta $a = 0.7$ ve $b = 0.3$ seçtiğimizi düşünelim:



Dağılımın sol tarafında $X = 0$ korkusunu, sağ tarafında ise $X = 1$ umudunu gösteren, dikey olarak sonsuza giden asimtotları görüyoruz. Beklenen π değeri, $\mathbb{E}[\pi] = 0.7$ 'dir.

İlk gözlem yapıldı, sonuç olumlu: $X_1 = 1$.

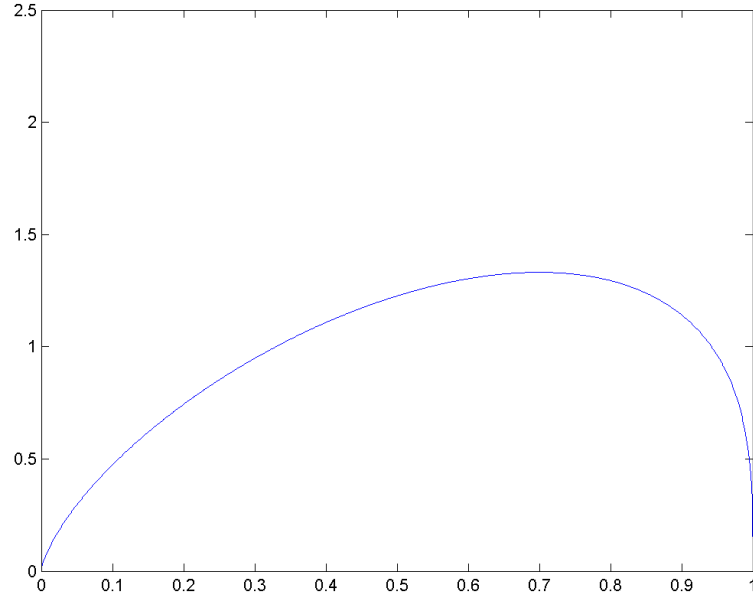
Posterior parametreleri: $a = 1.7$ ve $b = 0.3$.



Sol taraftaki korkunun silindiğini ve sağ uçtaki sonsuz umudun kuvvetlendiğini görüyoruz. Olasılık dağılımı, sağ kenarda $\pi = 1$ noktasında sonsuza gider. $\mathbb{E}[\pi] = 0.85$ olmuştur.

İkinci gözlem yapıldı, sonuç olumsuz: $X_1 = 1, X_2 = 0$.

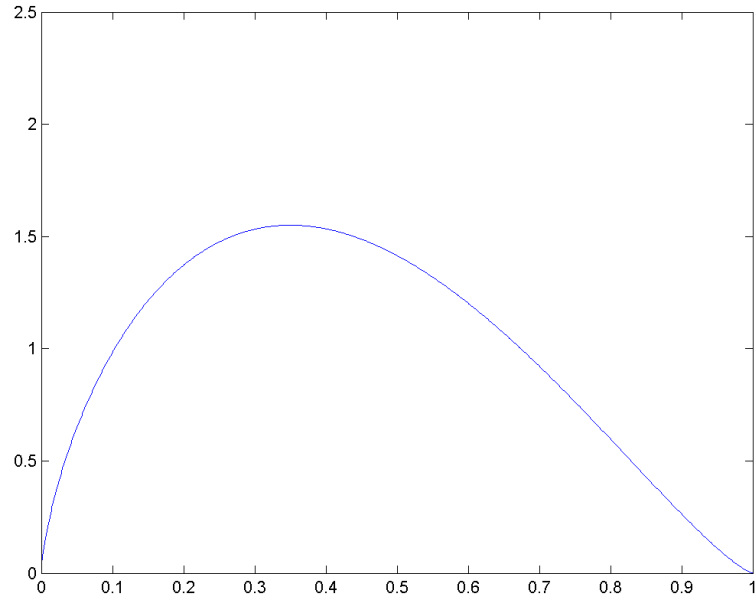
Posterior parametreleri: $a = 1.7$ ve $b = 1.3$.



Sağ uçtaki umut da ortadan kalkmış, dengeli, görece kararsız ama iyimser bir inanç oluşmuştur. $E[\pi] = 0.5667$

Üçüncü gözlem yapıldı, sonuç olumsuz: $X_1 = 1, X_2 = 0, X_3 = 0$.

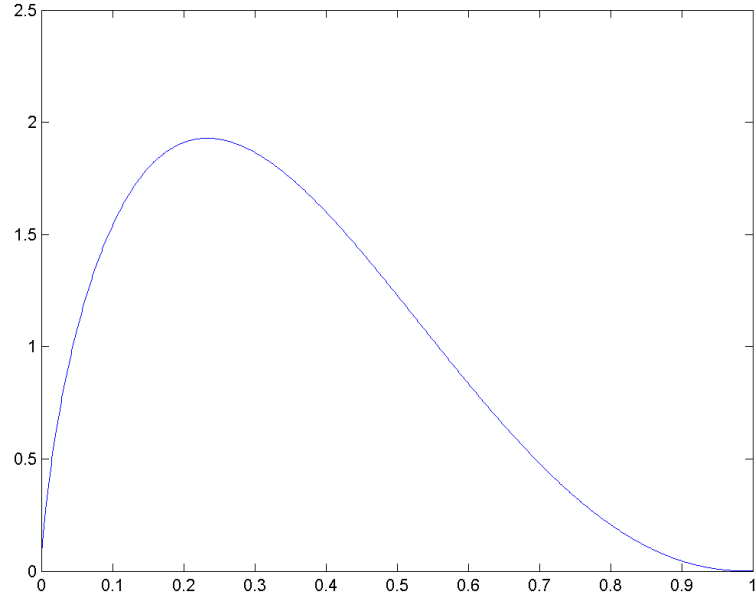
Posterior parametreleri: $a = 1.7$ ve $b = 2.3$.



İnanç kötümserliğe kaymış ve bir miktar yoğunlaşmıştır. $\mathbb{E}[\pi] = 0.4250$

Dördüncü gözlem yapıldı, sonuç olumsuz: $X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0$.

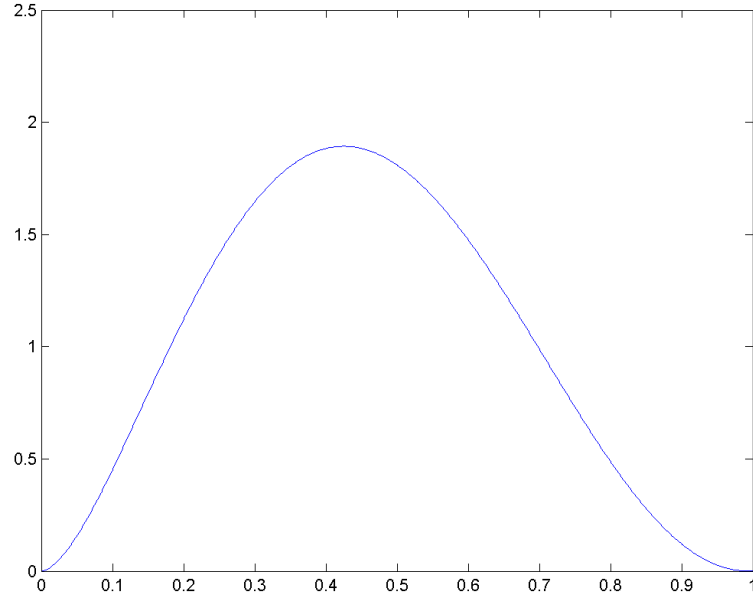
Posterior parametreleri: $a = 1.7$ ve $b = 3.3$.



Kötümser tarafa doğru daha da yoğunlaşmıştır. $\mathbb{E}[\pi] = 0.3400$

Beşinci gözlem yapıldı, sonuç olumlu: $X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1$.

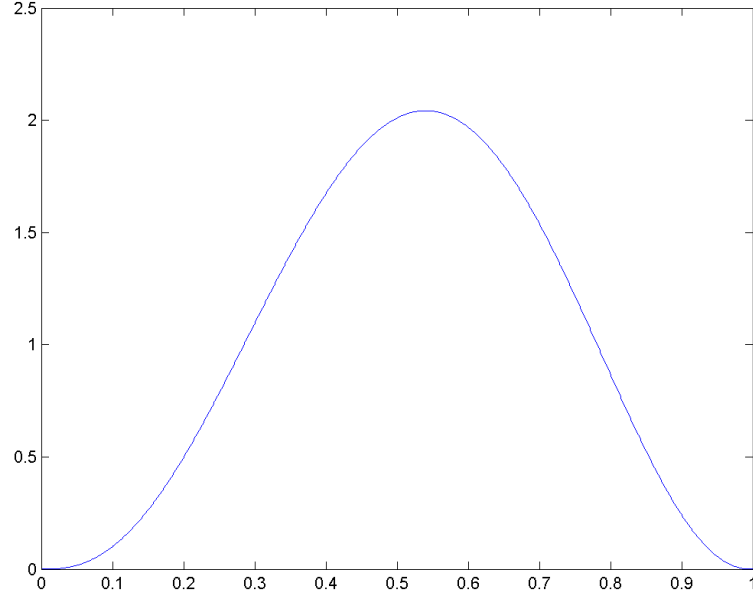
Posterior parametreleri: $a = 2.7$ ve $b = 3.3$.



Kötümserliği azalmış ve kararlılığı güçlenmiştir. $\mathbb{E}[\pi] = 0.4500$

Altıncı gözlem yapıldı, sonuç olumlu: $X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1$.

Posterior parametreleri: $a = 3.7$ ve $b = 3.3$.



Daha iyimser ve daha kararlı olmuştur. $\mathbb{E}[\pi] = 0.5286$

Bu verileri oluşturan Matlab kodu:

```
p = [0:0.001:1];  
  
a = 0.7;  
b = 0.3;  
x = [1 0 0 0 1 1];  
bp=betapdf(p,a,b);  
a/(a+b)  
plot(p,bp);  
axis([0 1 0 2.5]);  
print('-dpng', 'beta0.png');  
for i=1:size(x,2)  
    a = a+x(i);  
    b = b+(1-x(i));  
    bp=betapdf(p,a,b);  
    a/(a+b)  
    plot(p,bp);  
    axis([0 1 0 2.5]);  
    print('-dpng', sprintf('beta%d.png',i));  
end
```