

MCMC, Degisim Nokta Hesabi, Gibbs Orneklemesi, Bayes Teorisi

Ingiltere’de 1851 ve 1962 yıllari arasında komur madenlerinde olan kazaların sayısı yıllık olarak kayıtlı (ekteki zip dosyasında `coal.txt` dosyasında). Acaba bu kazaların “oranının” değişimine bakarak, değişimin olduğu seneyi bulabilir miyiz? Böyle bir değişim anı neyi gösterir? Belki madenlerle alakalı regulasyonlarda, denetimlerde bir değişiklik olmuştur, ve kaza oranı azalmıştır.

Bu hesabi yapabilmek için “değişim noktası” hesabi (change-point analysis), ve Bayes kuralı ile Bayes formüllerini hesaplamamızı sağlayan Markov Chain Monte Carlo (MCMC) tekniğine bakacağız. Kazaların sayısının tümünü iki Poisson dağılımının birleşimi (joint distribution) üzerinden modelleyeceğiz, ve bu dağılımların birinci Poisson’dan ikincisine geçtiği anı hesaplamaya uğrasacağız.

Önce Bayes, dağılımlar konusuna bir bakalım:

Poisson dağılımı

$$p(y|\theta) = \frac{e^{-\theta}\theta^y}{y!}$$

Eldeki n tane veri noktası $y = y_0, y_1, \dots, y_n$ ’nin hep birlikte θ ile tanımlı bir Poisson dağılımından gelip gelmediğinin ne kadar mümkün olduğu (likelihood) hesabi şöyledir:

$$p(y|\theta) = \frac{e^{-n\theta}\theta^{\sum y_i}}{\prod y_i!}$$

Formülün bölünen kısmındaki tüm y noktaları toplanıyor, bölün kisminde ise tüm y değerleri teker teker faktöryel hesabi sonrası birbiri ile carpılıyor.

Şimdi yukarıdaki θ değişkeni de noktasal bir değer yerine bir “dağılıma”, mesela θ Gamma dağılımına sahip olabilir: $\text{Gamma}(\alpha, \beta)$. Formülde α , β sabit değerlerdir (fonksiyon değişkeni değil). Gamma olasılık formülü şöyledir:

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

O zaman $p(y|\theta)$ formülünü bulmak için Bayes teorisini kullanmamız gereke-

cekti. Bayes teorisi bilindigi gibi

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Ikinci formule dikkat, esitlik yerine orantili olma (proportional to) isaretini kullaniyor. Sebep: bolen kisimdeki $p(y)$ 'yi kaldirdik, sonuc olarak soldaki $p(\theta|y)$ degeri artik bir dagilim degil – bu bir bakimdan onemli ama ornekleme amaci icin bir fark yaratmiyor, basitlestirme amaciyla bunu yaptik, boylece $p(y)$ 'yi hesaplamamiz gerekmeyecek, ama ornekleme uzerinden diger tum hesaplari hala yapabiliriz. Tamam.

Simdi Bayes Teorisini Gamma oncul (apriori) ve Poisson mumkunlugu (likelihood) uzerinden kullanirsak,

$$p(\theta|y) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \times \frac{e^{-n\theta} \theta^{\sum y}}{\prod y!}$$

Benzer terimleri yanyana getirelim:

$$p(\theta|y) = \frac{\beta^\alpha}{\Gamma(\alpha) \prod y!} \theta^{\alpha-1} \theta^{\sum y} e^{-\beta\theta} e^{-n\theta}$$

Simdi sol taraftaki bolumu atalim; yine usttekine benzer numara, bu kisim gidince geri galan dagilim olamayacak, ama ona ”oranli” baska bir formül olacak.

$$p(\theta|y) \propto \theta^{\alpha-1} \theta^{\sum y} e^{-\beta\theta} e^{-n\theta} \\ \propto \theta^{\alpha-1+\sum y} e^{-(\beta+n)\theta}$$

Bu dagilim nedir? Formulun sag tarafi Gamma dagiliminin formulune benzemiyor mu? Evet, formulun sag tarafi $Gamma(\alpha + \sum y, \beta + n)$ dagilimi, yani ona orantili olan bir formül. Yani Bayes teorisi uzerinden sunu anlamis olduk; eger oncul dagilim Gamma ise, Poisson mumkunluk bizi tekrar Gamma sonuc dagilimina goturuyor. Gamma'dan baslayinca tekrar Gamma'ya ulasiyoruz. Bu bir rahatlik, bir kolaylik, bir matematiksel numara olarak kullanilabilir. Sonuc (posterior) dagilimlarin sekli, hesaplanma, cebirsel islemler acisinden onemli, eger temiz, kısa, oz olurlarsa hesap islerimiz kolaylasir.

Not: Hatta uzerinde calistigimiz problem sebebiyle eger Poisson mumkunluk olacagini biliyorsak, sadece bu sebeple bile oncul dagilimi, ustteki ko-

laylik bilindigi icin, ozellikle Gamma secebiliriz, cunku biliriz ki Gamma ile baslarsak elimize tekrar Gamma gececektir.

Simdi komur madeni verisine gelelim. Bu madendeki kazalarin sayisinin Poisson dagilimindan geldigini one suruyoruz, ve kazalarin "iki turlu" oldugunu bildigimizden hareketle, birinci tur kazalarin ikinci tur kazalardan degisik Poisson parametresi kullandigini one surecegiz.

O zaman degisim anini, degisim senesini nasil hesaplariz?

Kazalarin ilk k senede ortalama θ ile, ve k ve n arasindaki senelerde ortalama λ Poisson ile dagildigini soyleyelim: Yani

$$Y_i = \text{Poisson}(\theta) \quad i = 1, \dots, k$$

$$Y_i = \text{Poisson}(\lambda) \quad i = k + 1, \dots, n$$

Burada Y_i sene i sirasinda olan kazalarin sayisini belirtiyor. Bayes kuralini hatirlarsak θ ve λ parametrelerine oncul dagilim atayacagiz. Bu dagilim Gamma olacak. Yani $\theta \sim \text{Gamma}(a_1, b_1)$ ve $\lambda \sim \text{Gamma}(a_2, b_2)$.

Ayrica k degerini de bilmiyoruz, k degeri yani "degisim noktası" Poisson dagilimlarin birinden otekine gectigi andir. Bu seneyi bulmaya calisiyoruz. Simdi tum verinin, tum seneleri kapsayacak sekilde modelini kurmaya baslayalim. k parametresinin aynen oteki parametreler gibi bir oncul dagilimi olacak (ki sonradan elimize k icin de bir sonuc dagilimi gecek), ama bu parametre elimizdeki 112 senenin herhangi birinde "esit olasilikta" olabilecegi icin onun oncul dagilimi Gamma degil $k \sim \text{Unif}(1, 112)$ olacak. Yani ilk basta her senenin olasiligi birbiriyle esit, her sene $\frac{1}{112}$ olasilik degeri tasiyor.

Bu modelin tamaminin mumkunlugu nedir?

$$L(\theta, \lambda, k|y) = \frac{1}{112} \times \prod_{i=1}^k \frac{e^{-\theta} \theta^{y_i}}{y_i!} \times \prod_{i=k+1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

Eger sonuc (posterior) gecisini yapinca yukarida oldugu gibi Gamma dagilimlarini elde ederiz:

$$L(\theta, \lambda, k|y) \propto \theta^{a_1-1+\sum_{i=1}^k y_i} e^{-(b_1+k)\theta} \lambda^{a_2-1+\sum_{i=k+1}^n y_i} e^{-(b_2+n-k)\lambda}$$

$\frac{1}{112}$ 'yi bir sabit oldugu icin formulden attik, bu durum orantili hali etkilemiyor. Ustteki formül icindeki Gamma dagilimlarini gorebiliyoruz, hemen

yerlerine koyalım:

$$L(\theta, \lambda, k|y) \propto \text{Gamma}(a_1 + \sum_{i=1}^k y_i, b_1 + k) \text{Gamma}(a_2 + \sum_{i=k+1}^n y_i, b_2 + n - k)$$

Gibbs orneklemeye geelim. Bu orneklemeye gore sartasal dagilim (conditional distribution) formulu bulunmaya ugrasilir, hangi degiskenlerin verili olduguna gore, o degiskenler sabit kabul edilebilir, ve orantisal formulden atilabilir. Bu her degisken icin teker teker yapilir.

Sorna hesap sirasinda her sartasal dagilima teker teker zar attirilir, ve elde edilen deger, bu sefer diger sartasal dagilimlara deger olarak gecilir. Bu islem sonuca erisilinceye kadar yineli (iterative) olarak tekrar edilir (mesela 1000 kere). O zaman,

$$\theta|Y_1, \dots, Y_n, k \sim \text{Gamma}(a_1 + \sum_{i=1}^k y_i, b_1 + k)$$

$$\lambda|Y_1, \dots, Y_n, k \sim \text{Gamma}(a_2 + \sum_{i=k+1}^n y_i, b_2 + n - k)$$

$$p(k|Y_1, \dots, Y_n) \propto \theta^{\sum_{i=1}^k y_i} e^{-k\theta} \lambda^{\sum_{i=k+1}^n y_i} e^{k\lambda}$$

En son formilde icinde k olan terimleri tuttuk, gerisini attik. Formül e terimleri birlestirilerek biraz daha basitlestirilebilir:

$$p(k|Y_1, \dots, Y_n) \propto \theta^{\sum_{i=1}^k y_i} \lambda^{\sum_{i=k+1}^n y_i} e^{(\lambda-\theta)k}$$

Bir basitlestirme daha soyle olabilir

$$K = \sum_{i=1}^k y_i$$

$$\lambda^{\sum_{i=k+1}^n y_i} = \lambda^{\sum_{i=1}^n y_i - \sum_{i=1}^k y_i}$$

Ustel islemlerde eksi isareti, ustel degisken ayrilince bolum islemine donusur:

$$\begin{aligned} &= \frac{\lambda^{\sum_{i=1}^n y_i}}{\lambda^{\sum_{i=1}^k y_i}} \\ &= \frac{\lambda^{\sum_{i=1}^n y_i}}{\lambda^K} \end{aligned}$$

$$p(k|Y_1, \dots, Y_n) \propto \theta^K \frac{\lambda^{\sum_{i=1}^n y_i}}{\lambda^K} e^{(\lambda-\theta)k}$$

$$= \left(\frac{\theta}{\lambda}\right)^K \lambda^{\sum_{i=1}^n y_i} e^{(\lambda-\theta)k}$$

$\lambda^{\sum_{i=1}^n y_i}$ terimi k 'ye değil n 'ye bağlı olduğu için o da final formülde atılabilir

$$p(k|Y_1, \dots, Y_n) \propto \left(\frac{\theta}{\lambda}\right)^K e^{(\lambda-\theta)k}$$

$p(k)$ için ortaya çıkan bu formüle bakarsak, elimizde verilen her k değeri için bir olasılık döndürecek bir formül var. Daha önceki Gamma örneğinde formüle bakarak elimizde hemen bir Gamma dağılımı olduğunu söyleyebilmistik. Bu kodlama sırasında isimize yarayacak bir şeydi, hesaplama için bir dağılıma “zar attırmamız” gerekiyor, ve Gamma örneğinde hemen Python Numpy kütüphanesindeki `random.gamma` çağrısına Gamma’dan gelen rasgele sayılar ürettirebiliriz. Üstteki formüle bakarsak, hangi dağılıma zar attıracağız?

Cevap şöyle: $p(k|..)$ pdf fonksiyonundaki k değişkeni 1, ..., 119 arasındaki tam sayı değerleri alabilir, o zaman ortada bir ayrık (discrete) dağılım var demektir. Ve her k noktası için olabilecek olasılık değerini üstteki $p(k|..)$ formülüne hesaplattırabiliyorsak, ayrık bir dağılımı her nokta için üstteki çağrı, ve bu sonuçları normalize ederek (vektörün her elemanını vektörün toplamına bölerek) bir dağılım sekline dönüştürebiliriz. Daha sonra bu “vektörel dağılım” üzerinden zar attırırız. Python kodundaki `w_choice` ya da R dilindeki `sample` çağrısı bu işi yapar.

Kodları işletince elimize $k = 41$ değeri geçecek, yani değişim anı $1851+41 = 1892$ senesidir.

```
import numpy as np
import math
import random
```

```
# samples indexes from a sequence of probability table
# based on those probabilities
def w_choice(lst):
    n = random.uniform(0, 1)
    for item, weight in enumerate(lst):
        if n < weight:
            break
```

```

        n = n - weight
    return item

#
# hyperparameters: a1, a2, b1, b2
#
def coal(n,x,init ,a1,a2,b1,b2):
    nn=len(x)
    theta=init[0]
    lam=init[1]
    k = init[2]
    z=np.zeros((nn,))
    for i in range(n):
        ca = a1 + sum(x[0:k])
        theta = np.random.gamma(ca, 1/float(k + b1), 1)
        ca = a2 + sum(x[(k+1):nn])
        lam = np.random.gamma(ca, 1/float(nn-k + b2), 1)
        for j in range(nn):
            z[j]=math.exp((lam-theta)*(j+1)) * (theta/lam)**sum(x[0:j])
        # sample
        zz = z / sum(z)
        k = w_choice(zz)
        print float(theta), float(lam), float(k)

if __name__ == "__main__":

    data = np.loadtxt("coal.txt")
    coal(1100, data, init=[1,1,30], a1=1,a2=1,b1=1,b2=1)

```

Kaynaklar:

Ioana A. Cosma and Ludger Evers, Markov Chain Monte Carlo Methods (Lecture)

Charles H. Franklin, Bayesian Models for Social Science Analysis (Lecture)