

ASIA: EXPERT SYSTEM

Gibbs Örneklemesi

Gibbs örneklemesi ve Metropolis-Hastings Örneklemesi Monte Carlo Markov Zincirlerine (MCMC) dayanmaktadır. Gibbs örneklemesi, geçiş dağılımının ya da başka bir deyişle geçiş olasılığının tam koşullu dağılımlar ile tanımlandığı bir MCMC yöntemidir.

Başlangıçta Gibbs dağılımlarını analiz etmek için uygulanmasına rağmen uygulanabilirliği Gibbs dağılımlarıyla sınırlandırılmamıştır. Temel olarak Gibbs örneklemesi, aday noktaların kabul edilmesi olasılığı 'nın 1 olduğu durumda Metropolis-Hastings örneklemesinin özel bir durumudur.

Gibbs dağılımı şekilde tanımlanmıştır:

$$f(\mathbf{x}_1, \dots, \mathbf{x}_d) \propto \left[\frac{1}{kS} E(\mathbf{x}_1, \dots, \mathbf{x}_d) \right].$$

Burada, k , pozitif bir sabit; S , sistemin sıcaklığı; E , sistemin enerjisi olan pozitif bir fonksiyon ve x_i , sistemin i . bileşeni için ilgili olunan bir özelliktir ($i=1, 2, \dots, d$).

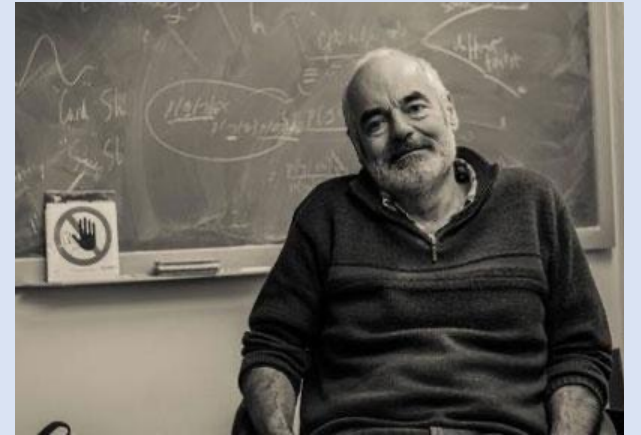
Temelde Gibbs örneklemesi ya da herhangi bir MCMC örneklemesi, yeterli uzunlukta çalıştırıldığı zaman, başlangıç değerleri unutulur ve bundan dolayı başlangıç değerlerinin seçimi önemli değildir.

Asya: Basit Bir Uzman Sistem

- Hayali Asya uzman sistemi klasik bir bayesyen ağıları örneğidir. Lauritzen ve Spiegelhalter tarafından ortaya konmuştur. “Dispne yani nefes alma zorluğu durumu tüberküloz, akciğer kanseri veya bronşite bağlı olabilir veya birden fazlasıyla alakalı olabilir ya da hiçbiri ile ilgili de olmayabilir. Tek bir göğüs röntgeni sonucu akciğer kanseri ve tüberküloz arasında ayırım yapamadığı gibi dispnenin varlığı veya yokluğu gibi kesin bir sonuç da çıkaramayabilir.
- Yakın zamanda Asya'ya yapılan bir ziyaret tüberküloz olasılığını artırırken, sigaranın hem akciğer kanseri hem de bronşit için bir risk faktörü olduğu biliniyor. Lauritzen ve Spiegelhalter (1988), bir göğüs kliniğine başvuran, Asya gezisinden yeni dönmüş ve dispne (nefes darlığı) gösteren bir hastanın teşhisini temsil eden kurgusal bir "uzman sistemi" sunmuştur.

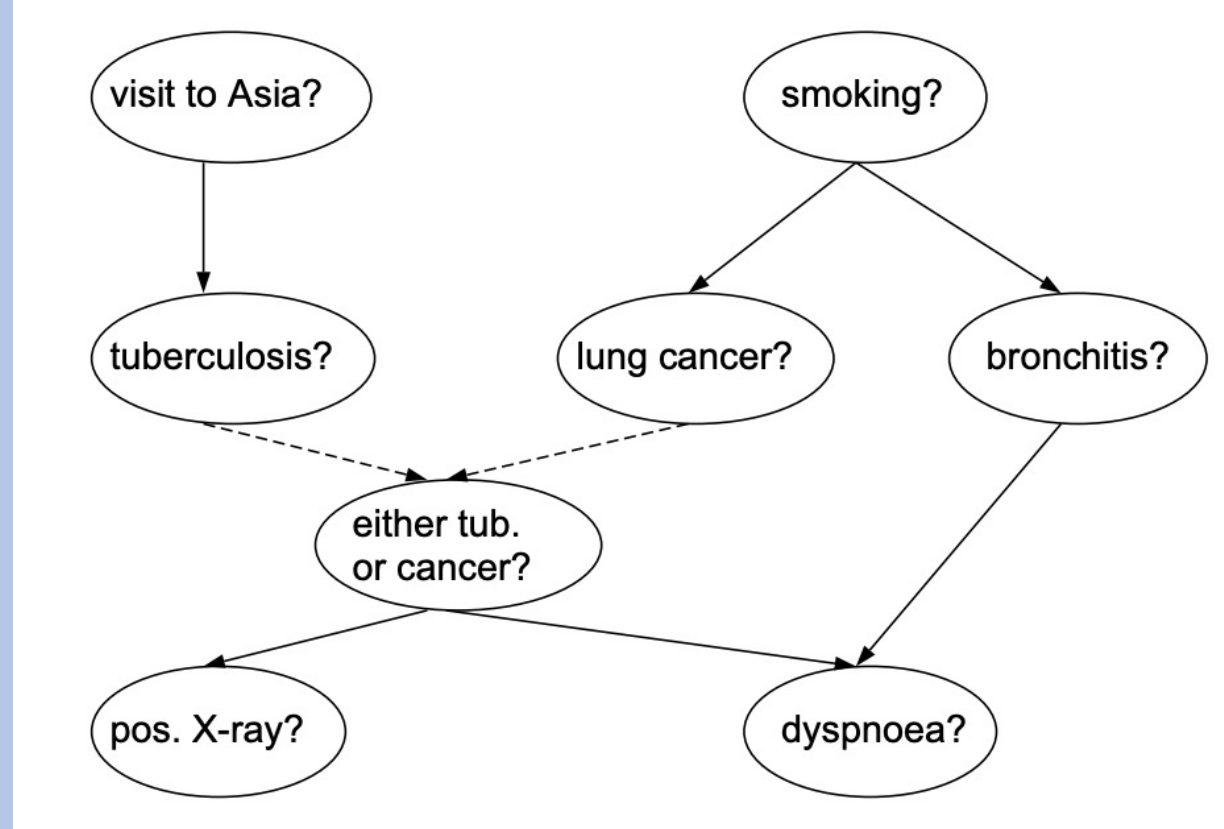


Steffen Lauritzen



David Spiegelhalter

Şekil 1, Asya örneğinde temel süreç için bir grafik model gösterilmektedir. Koşullu olasılıklar Lauritzen ve Spiegelhalter'de (1988) verilmiştir. Bu basit ağ kolayca izlenebilir. Yakın zamanda Asya'yı ziyaret eden ve nefes darlığı çeken bir hastanın herhangi bir hastalığa sahip olma olasılığı (tüberküloz, akciğer kanseri, bronşit) gibi ağ üzerinden ilerlenebilir.

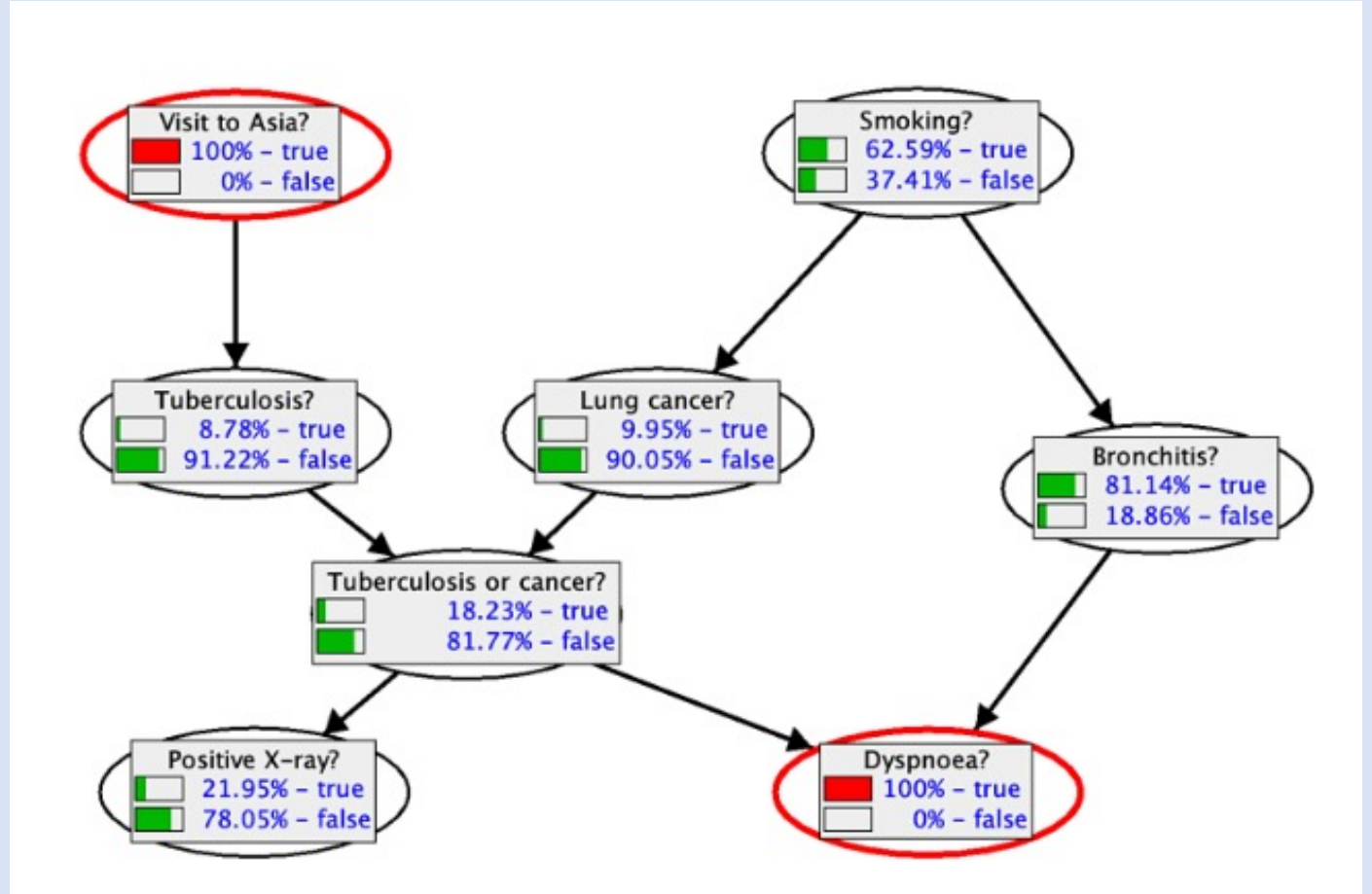


Şekil 1. Asya örneği temel süreç grafiği

Asya'yı ziyaret eden ve nefes darlığı çeken bir hasta göz önüne alındığında, her hastalığa sahip olma şansı nedir? Asya'yı ziyaret ettiği varsayılarak tüberküloz için %8,78, akciğer kanseri için %9.95 ve bronşit için %81,14 gibi.

Gözlenen düğümler (durumun bilindiği yer) kırmızı olarak işaretlenir ve hasta hakkındaki mevcut bilgileri temsil eder.

Yeni bilgiler edinilirse model güncellenebilir. Örneğin yukarıdaki aynı hastanın röntgeni pozitif çıktıysa ve sigara içmediğini belirttiyse her hastalığa sahip olma olasılığı değişir: %63,2 tüberküloz, akciğer kanseri için %12.7 ve bronşit için %45.9 gibi. Bu, bir BN'(Bayesian Networks)deki olasılıkların, mevcut bilgi olarak bayesci anlamda yorumlanması gerektiğini vurgular.



Şekil 2: Samlam'da uygulanan Asya uzman sistemi.

Samlam: (Sensitivity analysis, modeling, Inference and more –free software tool)

WinBugs'da Asya Uzman Sistemin Uygulaması

- **Bayesian Using Gibbs Sampling (BUGS)** programı Cambridge Üniversitesinde 1996 yılında geliştirilmiştir. Program, MCMC kullanarak karmaşık modellerin Bayesyen yaklaşımla analizini gerçekleştirmek üzere oluşturulmuştur.
- BUGS'un Kategorik Dağılım için, **dcat(p[])**: p, ayırık dağılıma sahip bir vektör olan dcat kullanılır.
- Ön bilgi olasılıkları ile mevcut bilginin birleştirilip son bilgi dağılışı oluşturulmasında Gibbs Örneklemesi yöntemini kullanmaktadır.
- İlk olarak Asya Uzman Sistem örneğini modellemek için Winbugs'a Şekil 3'teki gibi kodlarımızı yazıyoruz.
- Gözlenen özelliklere veri dosyasında 2 değeri verilmiştir Asia = 2 ve dyspnoea = 2 olarak.

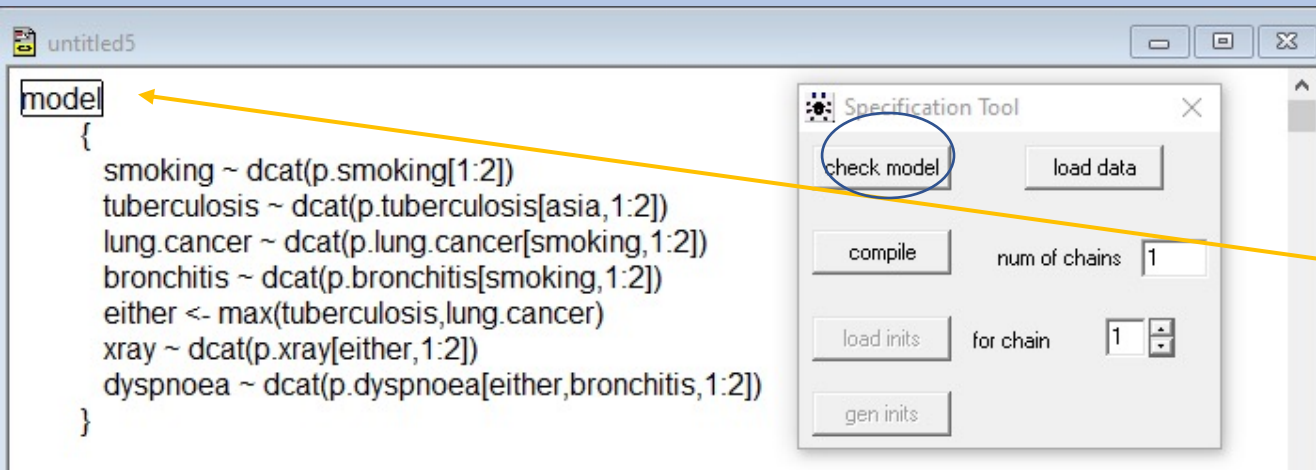
```
model
{
  smoking ~ dcat(p.smoking[1:2]) # 1: no, 2: yes
  tuberculosis ~ dcat(p.tuberculosis[asia,1:2])
  lung.cancer ~ dcat(p.lung.cancer[smoking,1:2])
  bronchitis ~ dcat(p.bronchitis[smoking,1:2])
  either <- max(tuberculosis,lung.cancer)
  xray ~ dcat(p.xray[either,1:2])
  dyspnoea ~ dcat(p.dyspnoea[either,bronchitis,1:2])
}

list(asia = 2, dyspnoea = 2,
     p.tuberculosis = structure(.Data = c(0.99,0.01,0.95,0.05), .Dim = c(2,2)),
     p.bronchitis = structure(.Data = c(0.70,0.30,0.40,0.60), .Dim = c(2,2)),
     p.smoking = c(0.50,0.50),
     p.lung.cancer = structure(.Data = c(0.99,0.01,0.90,0.10), .Dim = c(2,2)),
     p.xray = structure(.Data = c(0.95,0.05,0.02,0.98), .Dim = c(2,2)),
     p.dyspnoea = structure(.Data = c(0.9,0.1,
                                         0.2,0.8,
                                         0.3,0.7,
                                         0.1,0.9), .Dim = c(2,2,2)))

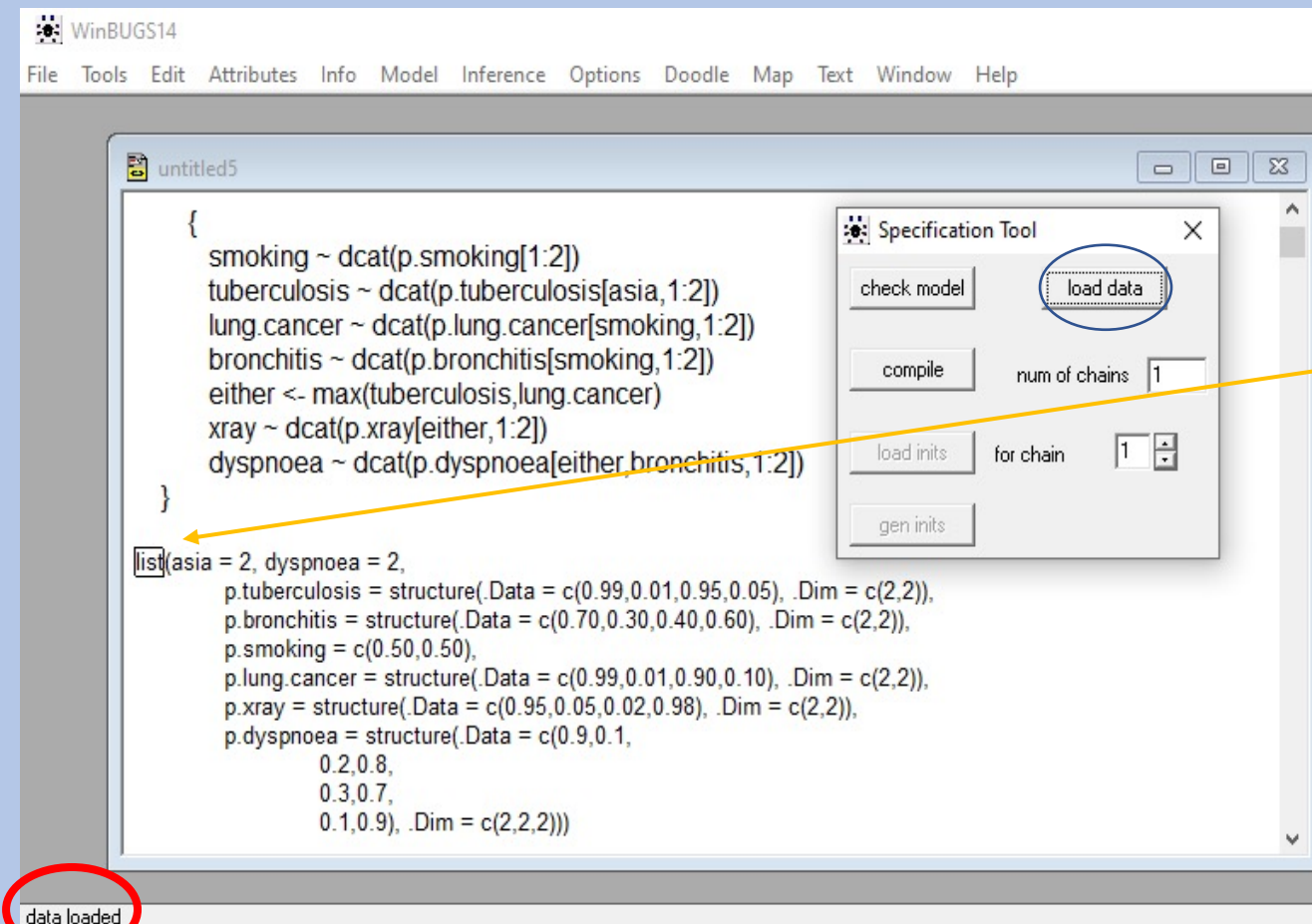
list(smoking = 1, tuberculosis = 1, lung.cancer = 1, bronchitis = 1, xray = 1)
```

(b) WinBUGS

Şekil 3: Winbugs uygulaması



WinBUGS 1.4.3'te
Model üzerine gelinerek specification tool'da
model checklenir.



Datayı yüklemek için list üzerine gelinerek
load data tıklanır. Data loaded yazısını
gördükten sonra başarıyla datamız yüklenmiş
olur.

untitled5

```
tuberculosis ~ dcat(p.tuberculosis[asia,1:2])
lung.cancer ~ dcat(p.lung.cancer[smoking,1:2])
bronchitis ~ dcat(p.bronchitis[smoking,1:2])
either <- max(tuberculosis,lung.cancer)
xray ~ dcat(p.xray[either,1:2])
dyspnoea ~ dcat(p.dyspnoea[either,bronchitis,1:2])
}
```

```
list(asia = 2, dyspnoea = 2,
     p.tuberculosis = structure(.Data = c(0.99,0.01,0.95,0.05), .Dim = c(2,2)),
     p.bronchitis = structure(.Data = c(0.70,0.30,0.40,0.60), .Dim = c(2,2)),
     p.smoking = c(0.50,0.50),
     p.lung.cancer = structure(.Data = c(0.99,0.01,0.90,0.10), .Dim = c(2,2)),
     p.xray = structure(.Data = c(0.95,0.05,0.02,0.98), .Dim = c(2,2)),
     p.dyspnoea = structure(.Data = c(0.9,0.1,
                                         0.2,0.8,
                                         0.3,0.7,
                                         0.1,0.9), .Dim = c(2,2,2)))
list(smoking = 1, tuberculosis = 1, lung.cancer = 1, bronchitis = 1, xray = 1)
```

Specification Tool

check model load data

compile num of chains 1

load inits for chain 2

gen inits

Son olarak bir sonraki list'e gelerek load inits tıklanarak model tamamlanmış olur.

model compiled

Sample Monitor Tool

node **smoking** chains 1 to 1 percentiles

beg 1 end 1000000 thin 1

clear set trace history density

stats coda quantiles bgr diag auto cor

2.5
5
10
25
median
75
90
95
97.5

Sonrasında inference'a girerek sampleları "sample monitor tool"da düzenleriz. Smoking, tuberculosis, bronchitis, lung.cancer, xray ve either için nodeları kurarız.

WinBUGS14

File Tools Edit Attributes Info Model Inference Options Doodle Map Text Window Help

| Node statistics | | | | | | | | |
|-----------------|-------|--------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| bronchitis | 1.807 | 0.3949 | 0.004395 | 1.0 | 2.0 | 2.0 | 1 | 10000 |

| Node statistics | | | | | | | | |
|-----------------|------|--------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| either | 1.18 | 0.3843 | 0.003966 | 1.0 | 1.0 | 2.0 | 1 | 10000 |

| Node statistics | | | | | | | | |
|-----------------|-------|--------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| lung.cancer | 1.102 | 0.3032 | 0.002944 | 1.0 | 1.0 | 2.0 | 1 | 10000 |

| Node statistics | | | | | | | | |
|-----------------|-------|--------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| smoking | 1.623 | 0.4845 | 0.005351 | 1.0 | 2.0 | 2.0 | 1 | 10000 |

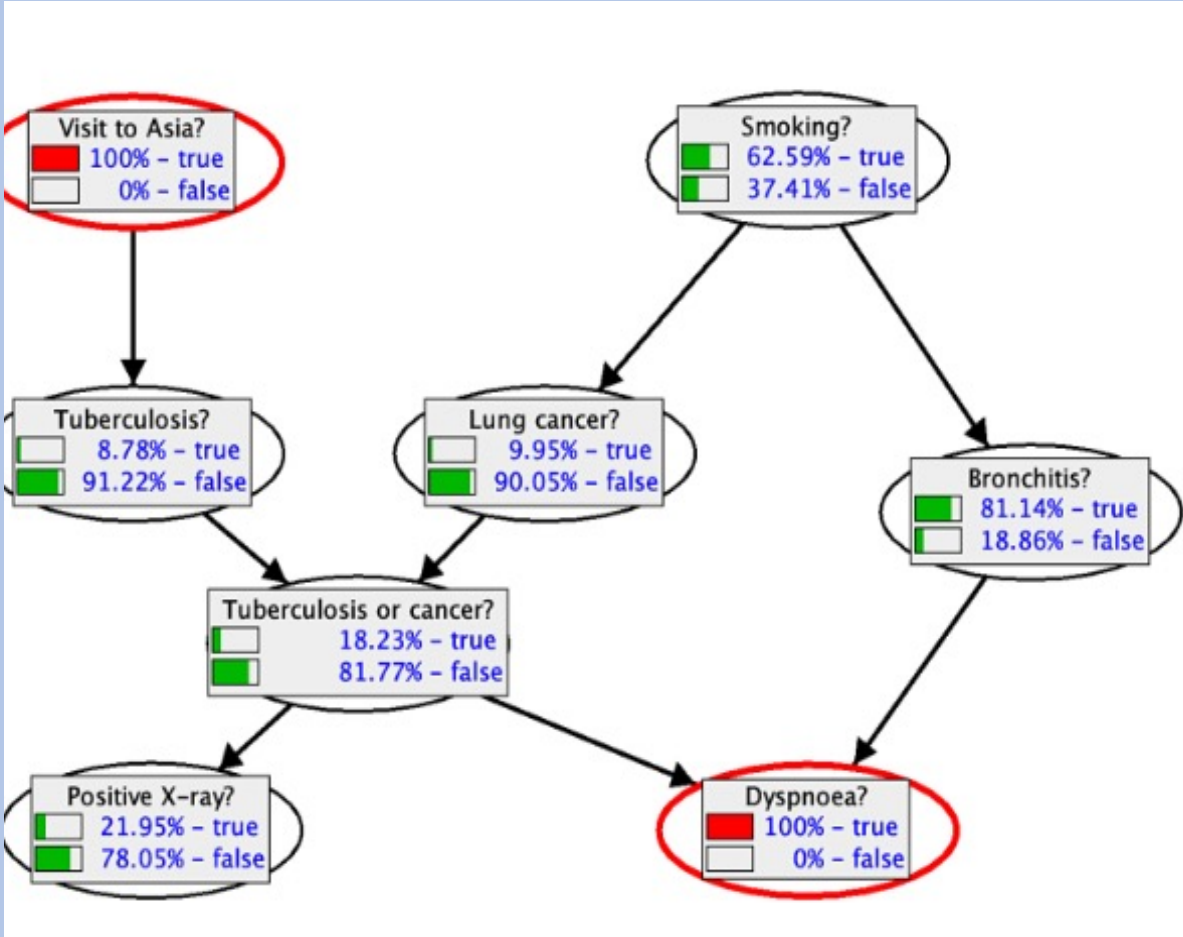
| Node statistics | | | | | | | | |
|-----------------|-------|--------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| tuberculosis | 1.084 | 0.2771 | 0.003064 | 1.0 | 1.0 | 2.0 | 1 | 10000 |

| Node statistics | | | | | | | | |
|-----------------|-------|-------|----------|------|--------|-------|-------|--------|
| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
| xray | 1.217 | 0.412 | 0.004122 | 1.0 | 1.0 | 2.0 | 1 | 10000 |

10000 yineleme yanda görüleceği gibi istatistikleri verir. Veriler yanıtlar olduğu ve bu yanıtlar da 1-No; 2-Yes şeklinde kategorik değişkenler olduğu için grafikler üzerinden yorumlamak daha doğru olacaktır.

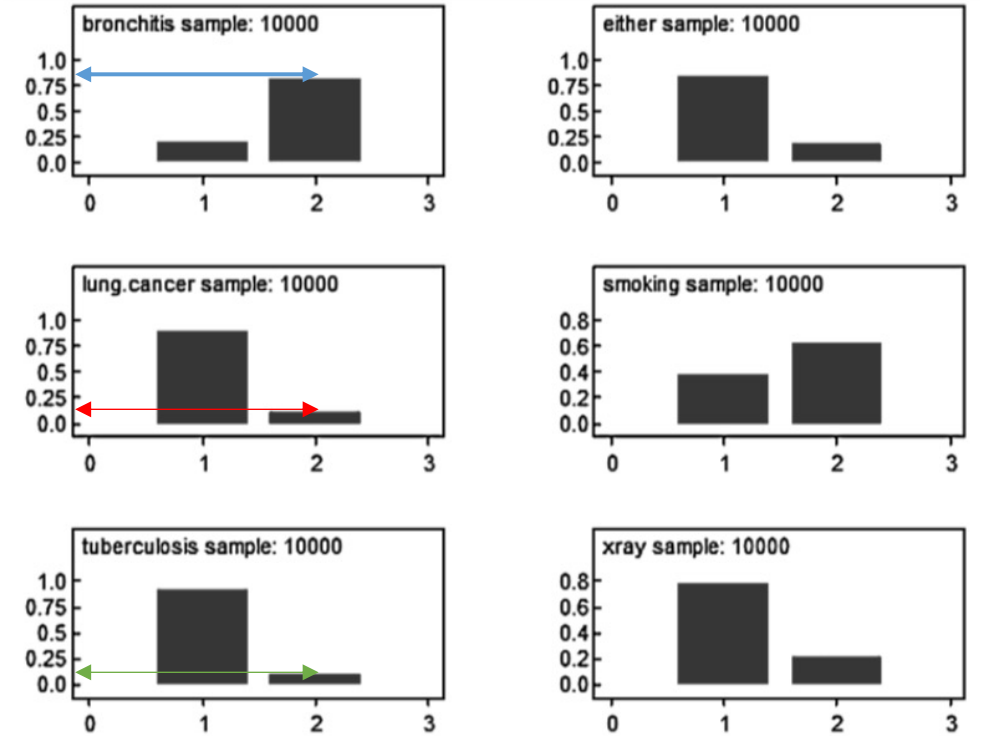
- MCMC örnekleyici için bir dizi başlangıç değeri belirlenir. Bu durumda, 1.000 burn-in örneğini 10.000 yinleme takip eder ve her bir koşul değişkeni için simüle edilmiş koşullu olasılık yoğunlukları Şekil 4b'de normalleştirilmiş, ikili değerli (1-No, 2-Yes) çubuk grafikler şeklinde gösterilir.

Şekil 4a'daki BN grafik yapısından hastanın %8,78 tüberküloz, %9,95 akciğer kanseri ve %81,14 bronşit olasılığı olduğu hemen teşhis edilebilir.



Şekil 4a : **Samlam'da** uygulanan Asya uzman sistemi kesin çıkarımları

Şekil 4b'deki simüle edilmiş koşullu olasılık yoğunlukları hastanın, %80 bronşit, %10 akciğer kanseri ve %8 tüberküloz olasılığı olduğu hemen teşhis edilebilir.



(b) approximate inference with MCMC (WinBUGS kernel density_b)

Figure 4. Conditional probability distributions for the Asia expert system

Şekil 4b : Asya uzman sistemi için **winbugs** koşullu olasılık dağılımları grafikleri

- Tablo 3'te karşılaştırıldığında, Samlam (LBP/EDBP, yazılım varsayılan ayarları) ve WinBUGS (MCMC) tarafından sağlanan yaklaşık çözümler, kesin çözümlerle yakından eşleşir. MCMC sonuçları daha doğru olma eğilimindedir (akciğer kanseri ve bronşit teşhisi için), oysa varyasyonel LBP veya EDBP yöntemi çok daha hızlı çalışır (altı yinelemede yakınsar).
- Göreceli hata: $\frac{\text{approximate solution} - \text{exact solution}}{\text{exact solution}} \times 100\%$.

Table 3. Asia expert system: exact vs. approximate solutions.

| Likelihood for... | Exact solution | LBP/EDBP solution | LBP/EDBP relative error | MCMC solution | MCMC relative error |
|---------------------|----------------|-------------------|-------------------------|---------------|---------------------|
| <i>Tuberculosis</i> | 0.088 | 0.086 | -2.1% | 0.085 | -3.8% |
| <i>Lung cancer</i> | 0.100 | 0.111 | 11.5% | 0.103 | 3.2% |
| <i>Bronchitis</i> | 0.811 | 0.803 | -1.1% | 0.809 | -0.3% |

Tablo1. Asya uzman sistem: kesin vs. yaklaşık sonuçlar

Kaynakça

- [1] Graphical Models Bayesian Networks and Bayesian Graphical Models. Available from: https://www.researchgate.net/publication/236867715_Directed_Graphical_Models_Bayesian_Networks_and_Bayesian_Graphical_Models [accessed Jun 02/2021].
- [2] Lauritzen, S. L.; Spiegelhalter, D. J. (1988). "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems". *Journal of the Royal Statistical Society*. **50** (2): 157–224.
- [3] Asia: Expert System, Evidence propagation. Available from: <https://www.multibugs.org/examples/latest/Asia.html> [accessed Jun 01/2021].
- [4] BUGS supports many distributions. Available from: http://www.di.fc.ul.pt/~jpn/r/bugs/bugs_tutorial.html [accessed Jun 02/2021].
- [5] Akar, M.; Gündoğsu S. (2013). "Bayes Teorisinin Su Ürünlerinde Kullanım Olanakları". *Journal of FisheriesSciences com* 8(1):8-16