14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 21

Konrad Menzel

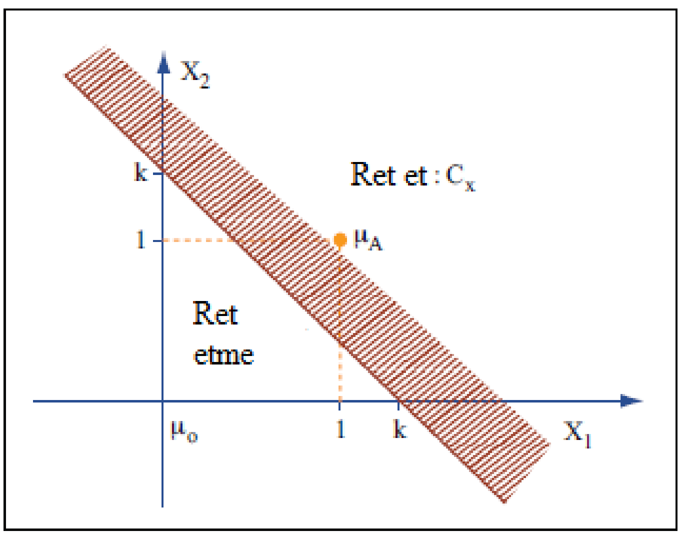
5 Mayıs 2009

Hipotez Testi Oluşturmak

Eğer X_i S_X desteğine sahipse, o zaman örneklem $X = (X_1, \ldots, X_n)$ S_X^n'i destekler. Bir testin kritik bölgesi kendisi için boş hipotezi ret edeceğimiz örneklemin desteğinin $C_X \subset S_X^n$ bölgesidir.

İzleyen örnek bir standart kurulumun en önemli unsurlarını göstermektedir, bu nedenle ona çok dikkatli bakmanız ve aynı adımların benzer problemlere nasıl uygulanacağını öğrenmeniz gerekiyor.

Örnek 1. *Varsayalım ki $X_1, \ldots, X_n$ bir i.i.d. örneklemdir, $X_i \sim N(\mu,4)$'tür ve $H_A : \mu = 1$'e karşı $H_0 : \mu = 0$'ı test etmekle ilgileniyoruz. Önce $n = 2$ durumuna bakalım:*



Kaynak: MIT OpenCourseWare

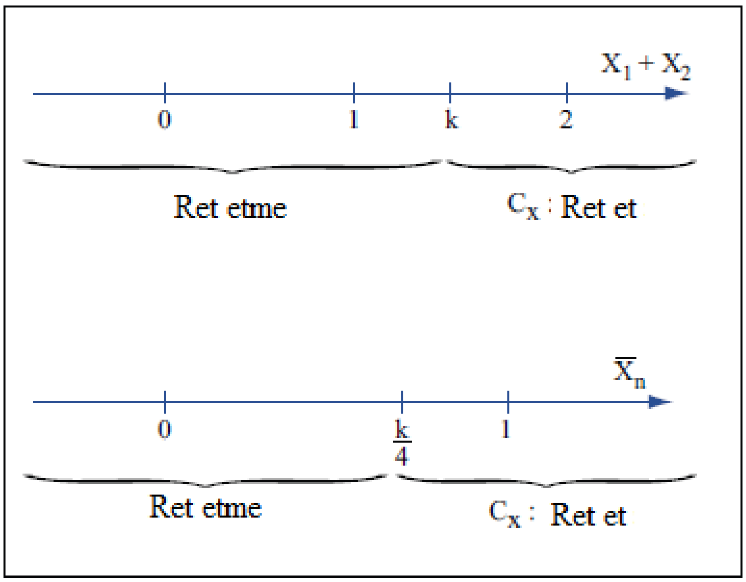
$\mu = 0$ ile uyumlu olmayacak kadar "çok büyük" $X_1 + X_2$'nin değerlerini ret edecek bir test tasarlayabiliriz. Bu ret bölgesini aynı zamanda bir doğrunun üst tarafında da gösterebiliriz.

Eğer n büyük ise bu gösterimin kullanımı çok kolaydır, çünkü ret bölgesini doğrudan $X_1, \ldots, X_n$ ile hayal etmek çok zordur. Ancak, resmi n'den tek boyuta indirgeyerek, kritik *bölgelerin garip şekillerini belirme yetimizi kaybedebiliriz, ancak onlar pratik uygulama açısından çokta yararlı değildirler zaten.*

Böylece bu örnekte, test sürecini bir test istatistiğine dayandıracağız, $T_n(X_1, \ldots, X_n) = \overline{X}_n$ *ve* T_n*'nin büyük değerleri için ret edeceğiz.*

k'yi nasıl seçeceğiz? iki hata türü arasında değiş-tokuş ile kaşı karşıya gelmek zorunda kalacağız. Varsayalım ki şimdi $n = 25$*'tir.* $X_i \sim N(\mu,4)$ *olduğu için,*

$$T_n := \bar{X}_n \sim \begin{cases} N\left(0, \frac{4}{25}\right) & \text{H0 altında} \\ N\left(1, \frac{4}{25}\right) & \text{HA altında} \end{cases}$$



Kaynak: MIT OpenCourseWare

Şimdi 1. Tip ve 2. Tip hataların olasılıklarını hesaplayabiliriz.

$$\begin{aligned} \alpha &= P(\bar{X} > k|\mu = 0) = 1 - \Phi\left(\frac{k-0}{2/5}\right) = \Phi\left(-\frac{k}{2/5}\right) \\ \beta &= P(\bar{X} \leq k|\mu = 1) = \Phi\left(\frac{k-1}{2/5}\right) \end{aligned}$$

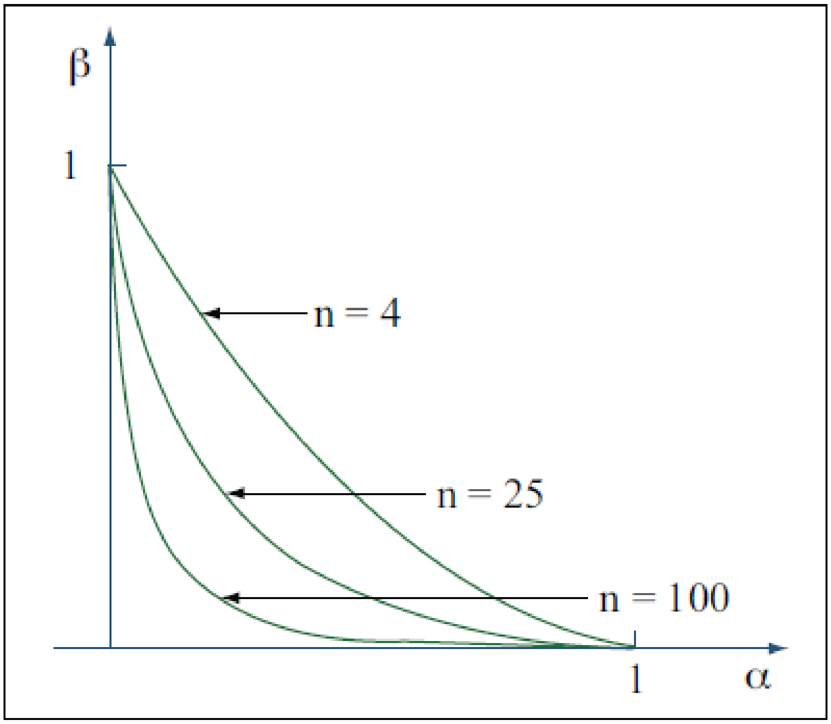
Bu durumda, α*,* β*, k'den herhangi birini belirlemek diğer ikisini belirler, ve o seçim 1.nci ve 2.nci Tip hataların olasılıkları arasında belirli bir değiş-tokuş içerir – eğer k'yi yükseltirsek, güvenirlik düzeyi* α *düşer, aynı şekilde* $1 - \beta$ *gücüde düşer. Spesifik olmak gerekirse, eğer* $k = 3/5$ *olarak seçersek,* $\alpha \approx \%6.7$ *ve* $\beta \approx \%15.87$ *olur.*

Farklı örneklem büyüklükleri için 1.nci ve 2.nci Tip hataların olasılıkları arasındaki değiş-tokuşu çeşitli k seçimleri için aşağıdaki gibi gösterebiliriz:

Düşük bir k değeri daha büyük güç ve aynı zamanda daha büyük güvenirlik düzeyi verir, böylece k'yi yükseltmek bizi grafik üzerinde sola doğru kaydırır.

k'yi nasıl seçmemiz gerekir? Normal kurulumda önceliğin yanlış ret etme olasılığı α'nın kontrol edilmesine verildiğini hatırlayınız, bu nedenle k'yi 1. Tip hata olasılığını kabul edilebilir bir düzeyde tutacak şekilde seçeceğiz, genellikle bu %5 veya %1'dir.

Elbette, $n \to \infty$ iken, sabit α için testin gücü, $1 - \beta$, 1'e doğru gider. Bir teamül olarak, genellikle α = %5 düzeyindeki bir ret ediş "anlamlılık "tır, benzer şekilde α = %1'de ret etme "yüksek anlamlılık "tır.



Kaynak: MIT OpenCourseWare

Örnek 2. *Bir önceki örnekte, sabit hipotez $\mu \in \{0, 1\}$ idi, fakat bu çok yapay bir varsayımdı ve genellikle bunun böyle bir durum olduğuna inanmak için bir nedenimiz yoktur.*

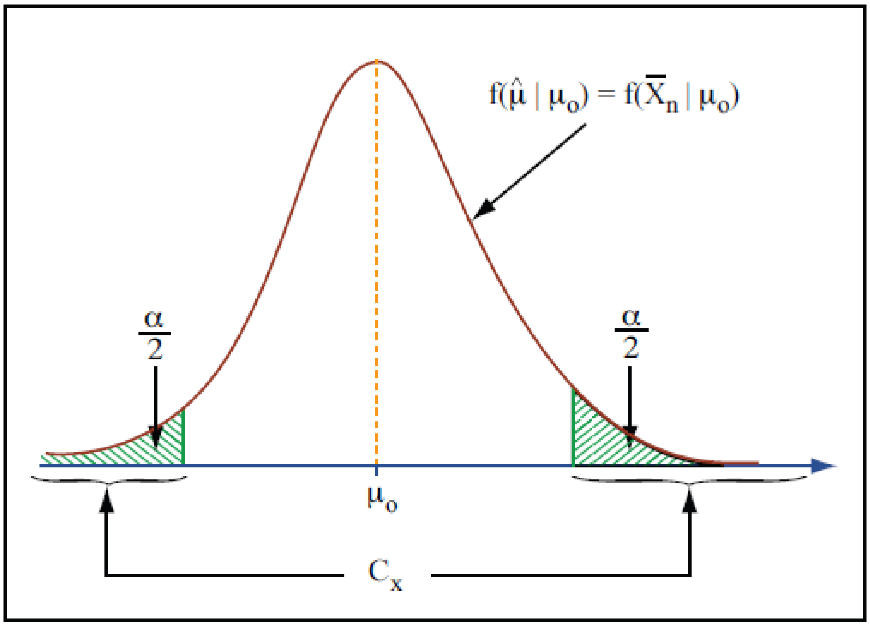
Daha önce olduğu gibi, varsayalım ki X_1, …, X_n bir i.i.d. örneklemdir, ancak şimdi aşağıdakini test edeceğiz:

$$H_0 : \mu = 0$$

$$H_A : \mu \neq 0$$

Şimdi H_A iki yönlü bir bileşik hipotezdir (yani alternatif altında, μ birkaç değer alabilir, bazıları μ_0'ın sollunda bazılar sağındadır). Yine sadece örneklem ortalamasına, $\overline{X}_{25}$, dayalı bir teste bakacağız – Kritik bölge şimdi nasıl görünür?

Sezgisel olarak, $\overline{X}$'in hem büyük ve hem küçük değerleri için H_0'ı ret etmek anlamalıdır. Yani eğer boş hipotez doğru ise, büyük ihtimalle her iki kuyrukta da değerler görmeyeceğiz. Alternatif hipotez, μ'nün 0'dan ya büyük ya da küçük olduğuna dair kanıtlar ile ilgilendiğimizi ifade etmektedir.



Kaynak: MIT OpenCourseWare

Bu nedenle iki değer yani k_1 ve k_2'yi belirleyeceğiz ki böylece

$$\alpha = P(\bar{X}_{25} > k_2|\mu = 0) + P(\bar{X}_{25} < k_1|\mu = 0) = \left[1 - \Phi\left(\frac{k_2}{2/5}\right)\right] + \Phi\left(\frac{k_1}{2/5}\right)$$

β nedir? Alternatif tek bir olasılık kanunu belirlemediği ve onun yerine onların sürekliliğini verdiği için, β çok iyi tanımlanmamıştır, yani sabit bir μ için şu yazılır:

$$\beta(\mu) = P(\bar{X}_{25} > k_2|\mu) + P(\bar{X}_{25} < k_1|\mu) = \left[1 - \Phi\left(\frac{k_2 - \mu}{2/5}\right)\right] + \Phi\left(\frac{k_1 - \mu}{2/5}\right)$$

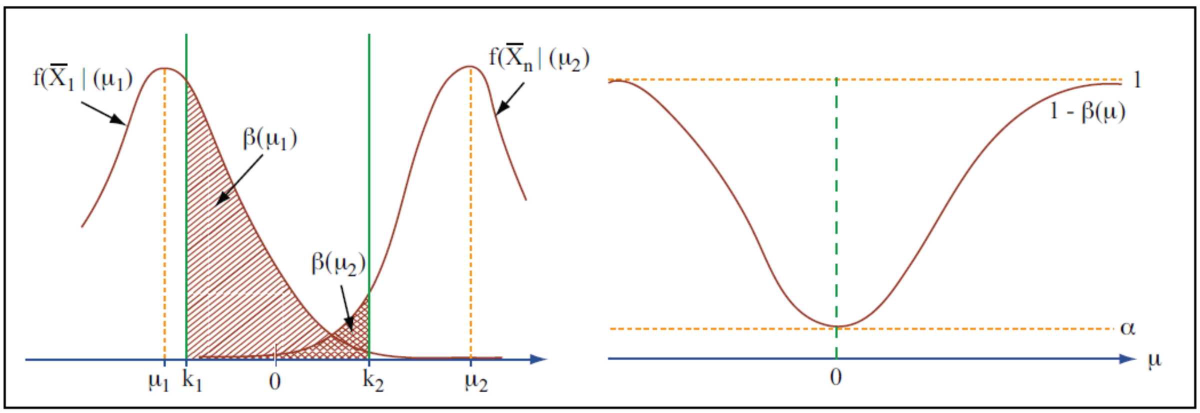
Genellikle arzulanan bir güvenirlik düzeyi α için, k_1, k_2'yi boş hipotezi tarafından varsayılan değer etrafında simetrik olarak seçeriz (normal dağılımın bir tek tepe noktası ve simetrik olduğu için, bu da kritik bölgeyi olabildiğince büyük yapar).

Son örnek varyansı bilinen bir normal kitleden elde edilen μ için güven aralığını oluşturma yolunu hatırlatmalı: Yukarıdaki süreç aslında aşağıdakine benzerdir:

1. μ için bir $1 - \alpha$'lık güven aralığı [A(X), B(X)] oluştur (Durum 1, son dersteki notlara bakınız)
2. Eğer $\mu_0 = 0 \notin$ [A(X), B(X)] ise, H_0'ı ret et

Dolaylı olarak boş hipotez altında $P_\theta(A(X) < \theta < B(X) = 1 - \propto$ olacak şekilde [A(X), B(X)] aralığını oluşturduğumuz için,

$$P(\theta_0 \notin [A(X), B(X)] | H_0 : \theta = \theta_0) = \alpha$$



Kaynak: MIT OpenCourseWare

1. Testlerin Değerlendirilmesi ve Oluşturulması

Tahmin ile ilgili tartışmamızda olduğu gibi, önce genel fikri tanıttık sonra birkaç örnek gördük. Şimdi ise testler arasında nasıl seçim yapacağımızı ve onları baştan itibaren nasıl oluşturacağımızı göreceğiz.

1.1. Testlerin Özelikleri

Herhangi bir testin güvenirlik düzeyi α = *P(1. Tip)* ve onun gücü $1 - \beta = 1 - P(2.\ Tip)$ ile ilgileniriz. Eğer her iki H_0 ve H_A basit hipotez ise, α ve β verili bir α için iyi tanımlanmıştır ve en basitinden $1 - \beta$' sı en yüksek, yani en güçlü, testi seçeriz.

Eğer H_A bileşik ve H_0 basit ise, verili bir α büyüklüğünde $1 - \beta(\theta) = 1 - P(2.\text{Tip Hata}|\theta)$ güç fonksiyonlarını karşılaştırmak için bir metriğe ihtiyacımız vardır. Bir test en azından her $\theta \in H_A$ noktasında diğer herhangi benzer büyüklükteki bir test kadar güçlü olduğu zaman, o test *uniform(tekdüze) olarak en güçlüdür* (UMP). Genel olarak, bir UMP'nin var olması gerekmez.

Örnek 3. *Bazen uniform olarak güçlü bir test bulmak olasıdır: Varsayalım ki $X_i \sim N(\mu,4)$ ve aşağıdaki hipotezi test etmek ile ilgileniyoruz:*

$$H_0 : \mu = 0$$

$$H_A : \mu > 0$$

$H_A : \mu = 1$'e karşı $H_0 : \mu = 0$ için en güçlü test eğer $\overline{X} > k$ ise ret et formunu alan testtir. $\mu_A > \mu_0$ olduğu sürece μ_A ne olursa olsun testin genel formu değişmez. Bundan ötürü, $H_A : \mu = 1$'e karşı $H_0 : \mu = 0$ için en güçlü test te aynı zamanda "ret et eğer $\overline{X} > k$ ise" formunu alır.

Aşağıdaki önemli sonuçta, $H_0 : \mu = \mu_0$ boş hipotezi altında $X_1, \ldots, X_n$ örneklemin bileşik p.d.f.si $f_0(x) = f_0(x_1, \ldots, x_n)$ ve $H_A : \mu = \mu_A$ altında örneklemin bileşik p.d.f.si $f_A(x)$ olarak ifade edilir.

Önerme 1 (Neyman-Pearson Lemma). *f_A'ya karşı f_0'ın testinde (her iki H_0 ve H_A basit hipotezdir), kritik bölge*

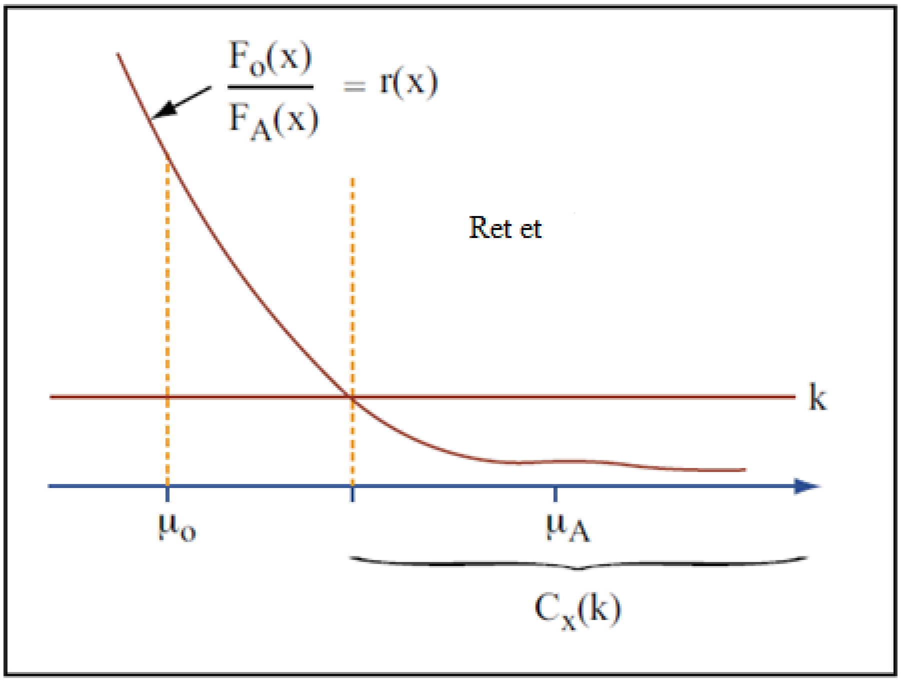
$$C(k) = \left\{ \mathbf{x} : \frac{f_0(\mathbf{x})}{f_A(\mathbf{x})} < k \right\}$$

herhangi bir $k \geq 0$ tercihinde en güçlüdür.

k seçiminin testin belirtilen güvenirlik düzeyi α'ya bağlı olduğunu not ediniz. Bu, eğer $X_1, \ldots, X_n$ örneklemi için aşağıdaki *olabilirlik oranı* düşük ise, en güçlü test ret eder anlamına gelir.

$$r(X_1, \ldots, X_n) = \frac{f_0(X_1, \ldots, X_n)}{f_A(X_1, \ldots, X_n)}$$

Yani, veri büyük bir olasılıkla H_A altında oluşmuştur.



Kaynak: MIT OpenCourseWare

Neyman-Pearson Lemma'da doğrudan verilen en güçlü test, örnek uzayda her x noktasında (burada integraller çok boyutludur, yani tipik olarak $x \in \mathbb{R}^n$) büyüklük

$$\alpha = P(\text{ ret } |H_0) = \int_{C(k)} f_0(\mathbf{x})d\mathbf{x}$$

ve güç

$$1 - \beta = P(\text{ ret } |H_A) = \int_{C(k)} f_A(\mathbf{x})d\mathbf{x}$$

arasındaki değiş-tokuşu çözer. α ve 1- β ifadelerinden olabilirlik oranı $\frac{f_0(x)}{f_A(x)}$'in kritik bölgeye x eklemenin "fiyatı"nı C_X bölgesinden bir nokta eklemenin güçteki kazanımına nispetten α cinsinden ne kadar "ödeyeceğimiz"i verdiğini görebiliriz.

Dolaysıyla, "en ucuz" –yani küçük olabilirlik oranına sahip- x noktalarını ekleyerek kritik bölgeler oluşturmaya başlamalıyız. O zaman, olabilirlik oranına göre sıralanmış x değerlerinden aşağı doğru gideriz ve α'nın büyüklüğü istenilen düzeye ininceye kadar nokta seçmeye devam ederiz.

Örnek 4. *Bir sanık (D) kapkaççılıktan mahkemededir. Mahkûm etmek için, jüri %95 olasılıkla kararın doğru olduğuna inanmak zorundadır.*

Savcının ortaya koyabileceği veya koyamayacağı üç parça potansiyel kanıt vardır. Bir duruşmada jüri kendisine gösterilen üç parçadan sadece birine dayanarak mahkûmiyet kararı verir. Aşağıdakiler potansiyel kanıt parçalarıdır, karşılıklı bağımsızlık varsayılmıştır. Tabloda ayrıca sanığın suçlu olduğu veri iken her parçanın incelenmesi olasılığı ile sanığın suçlu olmadığı veri iken her parçanın incelenmesi olasılığı da verilmiştir.

		suçlu	suçsuz	olabilirlik oranı
1.	D polisin geldiğini görünce kaçar	0.6	0.3	1/2
2.	D suç işlendiğinde başka yerde olduğunu kanıtlayamaz (mazeret)	0.9	0.3	1/3
3.	D'nin evininin yakınında boş çanta bulunur	0.4	0.1	1/4

Neyman-Pearson gösterimine göre, x kanıt parçalarının 2^3 olası kombinasyonlarından herhangi birisi olabilir. Bağımsızlık varsayımını kullanarak, bütün ipuçlarının kombinasyonlarını, her bir ipucunun her hipotez altındaki ilgili olabilirliklerini ve ilgili olabilirlik oranlarını listeleyebiliriz. Listeyi üçüncü kolondaki olabilirlik oranına göre sıraladım. Son kolona sıralanmış x kombinasyon listesine göre birikimli toplamı ekledim:

$$\alpha(k) = \sum_{r(\mathbf{x}) \leq k} f_0(\mathbf{x})$$

		Suçlu $f_A(x)$	Suçsuz($f_0(x)$	Olabilirlik Oranı $r(x) = \frac{f_0(x)}{f_A(x)}$	$\alpha(k)$
1.	her üç ipucu	216/1000	9/1000	0.0417	9/1000
2.	mazeret, bulunan çanta	144/1000	21/1000	0.1458	30/1000
3.	kaçış, mazeret	324/1000	81/1000	0.25	111/1000
4.	mazeret	216/1000	189/1000	0.875	300/1000
5.	kaçış, bulunan çanta	24/1000	21/1000	0.875	321/1000
6.	bulunan çanta	16/1000	49/1000	3.0625	370/1000
7.	kaçış	36/1000	189/1000	5.25	559/1000
8.	hiçbirisi	24/1000	441/1000	18.375	1

Jüri, %5'ten daha düşük olasılıklı yanlış mahkûmiyete karşın (yani sanık masum ise), eğer doğru ise en az %95 güvenilirlikle mahkûmiyet kararı verir. Hipotez testi terminolojisine göre, mahkûmiyet kararı, α = %5 büyüklüğündeki en güçlü testi kullanarak sanığın masum olduğu boş hipotezinin ret edilmesi ile ilintilidir.

Son kolondaki α(k) değerlerine bakınca, ilk iki kanıttın kombinasyondan fazla kanıt eklemek yanlış mahkûmiyet olasılığı α'yı %5'ten fazla artırdığını okuyabiliyoruz. Dolaysıyla, jüri suçlunun polisi gördüğünde kaçıp kaçmamasına bakmadan, başka *yerde olduğunu ispatlamamasından ve evinin yakınında bulunan çantadan ötürü sanığı mahkûm etmelidir. Prensip olarak, jüri, ilaveten, sanığın kaçışını, bulunduğu yeri ispatlamamayı ve çantanın bulunmamasını rasgele belirleyebilir (3. durum): Eğer bu durumda jüri sanığı* $\frac{50+30}{81} \approx \frac{1}{4}$ *olasılıkla mahkum etmiş olsaydı, yanlış mahkumiyetin olasılığı tam olarak %5 olurdu, fakat bu muhtemelen hukuk sistemi tarafında kabul edilebilir bir uygulama olmazdı.*

Örnek 5. *Ortalamaya dayalı bir testin normal durumda en güçlü test olduğunu şimdi gösterebiliriz. Varsayalım ki* $X_i \sim N(\mu, 4)$ *ve* $H_A : \mu = 1$*'e karşı* $H_0 : \mu = 0$*'ı test edeceğiz. Burada 25 gözlemli i.i.d. olan bir* $X_1, \ldots, X_{25}$ *örneklemimiz var.*

Gözlemler i.i.d. normal oldukları için, gözlemlenen örnekleme göre hesaplanan olabilirlik oranı aşağıdaki ile verilir:

$$
\begin{aligned}
r(\mathbf{X}) &= \frac{f(\mathbf{X}|\mu=0)}{f(\mathbf{X}|\mu=1)} = \prod_{i=1}^{25} \frac{\frac{1}{\sqrt{2\pi 2}} e^{-\frac{(X_i-0)^2}{2\cdot 4}}}{\frac{1}{\sqrt{2\pi 2}} e^{-\frac{(X_i-1)^2}{2\cdot 4}}} \\
&= \prod_{i=1}^{25} e^{\frac{1}{8}\left[(X_i^2-2X_i+1)-(X_i^2)\right]} \\
&= e^{\frac{25}{8}-\frac{1}{4}\sum_{i=1}^{25} X_i} = e^{\frac{25}{8}-\frac{1}{100}\bar{X}_{25}}
\end{aligned}
$$

r(X)'in örneklem ortalaması $\overline{X}_{25}$ *aracılığıyla örnekleme bağlı olduğunu ve* $\overline{X}_{25}$*'te kesin azalan olduğunu görebiliriz. Bu nedenle, en güçlü testin kritik bölgesi aşağıdaki formu alır:*

$$C_X(k) = \{\mathbf{x} : r(\mathbf{x}) \le k\}$$