14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

Problem Seti 7 - Çözümler
14.30 Ekonomide İstatistiksel Yöntemlere Giriş

Konrad Menzel

Son Gün: 14 Nisan 2009

Soru Bir

PDF ∅(z) ve CDF Φ(z)'li standart Normal Dağılımı daha yeni öğrendik. Daha sonra çok kullanacağımız için, kendimizi bu dağılıma alıştıralım.

1. z = {-3, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3} için u = Φ(z)'yi bulunuz.

- 1'in Çözümü:

```
Q1_1 =
   -3.0000    0.0013
   -2.5000    0.0062
   -2.0000    0.0228
   -1.5000    0.0668
   -1.0000    0.1587
   -0.5000    0.3085
         0    0.5000
    0.5000    0.6915
    1.0000    0.8413
    1.5000    0.9332
    2.0000    0.9772
    2.5000    0.9938
    3.0000    0.9987
```

2. Bölüm (1)'deki cevabınızı kullanarak z = {0, 0.5, 1, 1.5, 2, 2.5, 3} için u = 1 – Pr(|Z| ≤ z)'yi bulunuz. Cevabınızı nasıl elde ettiğinizi açıklayınız(matematiksel ifadeler ile). Bu *u*'nun değerlerini en uçta gerçekleşen sonucun "olasılık değerleri" anlamına gelen "p-değerleri" olarak adlandıracağız. Bu yedi değeri ezberleyin-daha sonra kesinlikle kullanacaksınız.

- 2'nin Çözümü:

```
Q1_2 =
         0    1.0000
    0.5000    0.6171
    1.0000    0.3173
    1.5000    0.1336
    2.0000    0.0455
    2.5000    0.0124
    3.0000    0.0027
```

3. z = {0.001, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20} için z = $\Phi^{-1}(u)$'yu bulunuz.

- 3'ün Çözümü:

```
Q1_3 =
    0.0010   -3.0902
    0.0050   -2.5758
    0.0100   -2.3263
    0.0200   -2.0537
    0.0500   -1.6449
    0.1000   -1.2816
    0.2000   -0.8416
```

4. Bölüm (3)'teki sonuçları kullanarak her bir u için z = $\Phi^{-1}(1 - u)$'yi elde ediniz.

- 4'ün Çözümü:

```
Q1_4 =
    0.0010    3.0902
    0.0050    2.5758
    0.0100    2.3263
    0.0200    2.0537
    0.0500    1.6449
    0.1000    1.2816
    0.2000    0.8416
```

5. Bölüm (3) ve (4)'teki sonuç ve/veya yöntemleri kullanarak z'yi elde ediniz. (3)'teki her bir u için 1 – Pr(|Z| $\leq$ z) = u'dır. z'nin bu değerleri "iki-taraflı α-düzey kritik değerleri" olarak adlandırılır. Buradaki örnekte α = u'dur. Örneğin, şöyle deriz, "standart normal dağılımın (iki-taraflı) %10 kritik değerleri z =…'dir". Daha sonra bu çok faydalı olacaktır, bu yedi değeri ezberleyin.

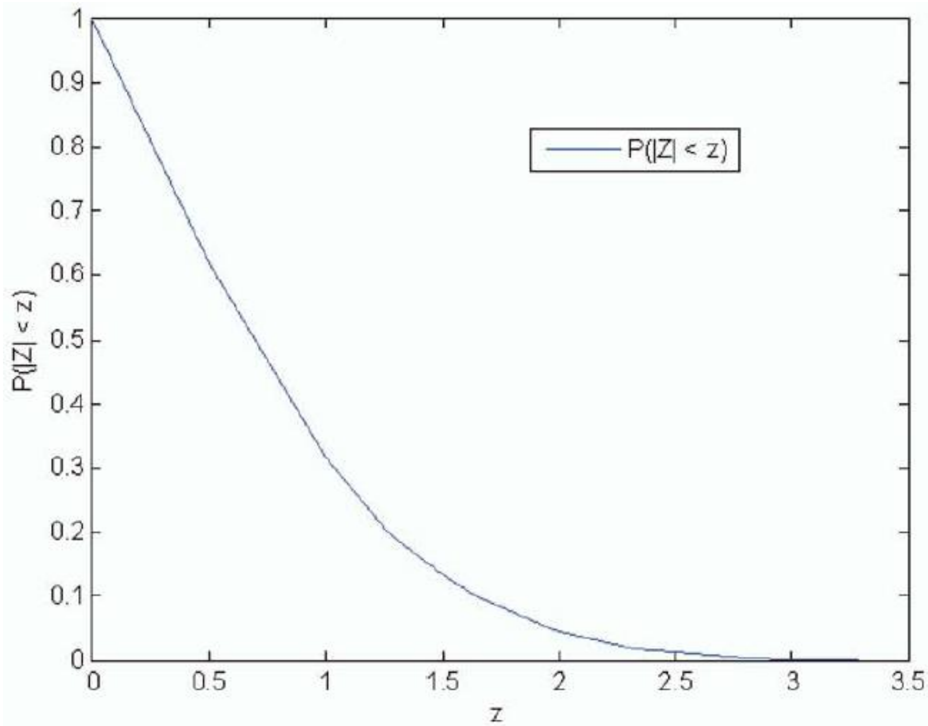
- 5'in Çözümü:

```
Q1_5 =
    0.0010    3.2905
    0.0050    2.8070
    0.0100    2.5758
    0.0200    2.3263
    0.0500    1.9600
    0.1000    1.6449
    0.2000    1.2816
```

6. Son olarak, Bölüm (2) ve (5)’te elde ettiğiniz 14 z ve u değerini kısa bir tabloda sıralayınız. z = 0 ve p = 1 ile başlayınız ve z’leri ilgili olasılıkları ile beraber sağ tarafta sıralayınız.

- 6’nın Çözümü:

```
Q1_6 =
         0    1.0000
    0.5000    0.6171
    1.0000    0.3173
    1.2816    0.2000
    1.5000    0.1336
    1.6449    0.1000
    1.9600    0.0500
    2.0000    0.0455
    2.3263    0.0200
    2.5000    0.0124
    2.5758    0.0100
    2.8070    0.0050
    3.0000    0.0027
    3.2905    0.0010
```



Soru İki

1. Bir rasgele değişken X verilmişken, X'in standardizasyonu Z'yi tanımlayınız ve varyansını türetiniz.

- 1'in Çözümü: X'in Z standardizasyonu sonlu bir dönüştürmedir. Burada ortalamayı çıkarıp standart sapmaya (varyansın karekökü) bölüyoruz.

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{Var(X)}} = \frac{X - \mu_x}{\sigma}$$

Varyansı 1'dir:

$$Var(Z) = Var\left(\frac{X-\mu_x}{\sigma}\right) = \frac{1}{\sigma^2} Var(X-\mu_x) = \frac{1}{\sigma^2} Var(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

2. Varsayalım ki $\overline{X}_{20}$ $X_i \sim N(1,2)$ dağılımlı n = 20 gözlemli bir i.i.d. örneklemin ortalamasıdır. Bu ortalamanın beklenen değeri ve varyansı nedir? Dağılımı nedir (İpucu: Önerme 24 civarındaki ders notlarına bakınız)? Örneklem ortalaması $\overline{X}_{20}$'in 0.75 ile 1.25 arasında olma olasılığı nedir?

- 2'nin Çözümü: k kadar i.i.d. rasgele değişkenin toplamının beklenen değeri sadece tek çekilişin beklenen değeridir:

$$\mathbb{E}[\frac{1}{k}(X_1 + ... + X_k)] = \mathbb{E}[X]$$

Bu durum, ortalamanın 1 olmasını ve varyansın şimdiye kadar türetmediğimiz aynı formülü izlemesini beklememememiz gerektiği anlamına gelir:

$$\begin{aligned} Var\left[\frac{1}{k}(X_1 + ... + X_k)\right] &= \frac{1}{k^2}Var(X_1 + ... + X_k) \\ &= \frac{1}{k^2}(Var(X_1) + ... + Var(X_k)) \\ &= \frac{1}{k^2}k \cdot Var(X) \\ Var\left[\frac{1}{k}(X_1 + ... + X_k)\right] &= \frac{1}{k}Var(X) \end{aligned}$$

Burada ikinci satırdaki k rasgele değişkenin bağımsızlıklarını ve üçüncü satırdaki eş dağılımı kullandık. Bu özel probleme özgü olarak varyans:

$$Var(\bar{X}_{20}) = \frac{1}{20}4 = \frac{1}{5}$$

Bu, ortalaması 1 ve varyansı 1/5 olan bir normal dağılımdır(24. Önermeye göre). Örneklem ortalamasının 0.75 ile 1.25 arasında olma olasılığı, P(0.75 $\leq \overline{X}_{20} \leq$ 1.25), aşağıdaki gibidir. Örneklem ortalamasını standardize edip rasgele değişken Z = $\frac{\bar{X}_{20}-1}{\sqrt{\frac{1}{5}}}$'i elde ederiz:

$$
\begin{aligned}
P(0.75 \leq \bar{X}_{20} \leq 1.25) &= P\left(\frac{-0.25}{\sqrt{\frac{1}{5}}} \leq \frac{\bar{X}_{20}-1}{\sqrt{\frac{1}{5}}} \leq \frac{0.25}{\sqrt{\frac{1}{5}}}\right) \\
&= P\left(-0.25\sqrt{5} \leq z \leq 0.25\sqrt{5}\right) \\
&= 1-2(0.2881) \\
&= 0.4238
\end{aligned}
$$

3. Problem Seti 5'te, k i.i.d. üstel rasgele değişkenin ortalamasının (ve toplamının) sonlu dağılımını belirlemek için büklüm formülünü kullandık ve ortalamanın dağılıma bağlı olmadığını da belirttik. k i.i.d. rasgele değişkenin toplamının varyansı nedir? Ortalamalarının varyansı nedir? Üstel RVlerin üzerindeki ortalamaların normal dağılıma yaklaştığını (problem set 5'in cevap setinde) fark ettik. Bu özelliğin üstel dağılıma özgü olduğunu düşünür müsünüz?

- 3'ün Çözümü: Bu sorunun cevabın 1'de türetmiştik:

$$
\begin{aligned}
Var\left[X_1+\ldots+X_k\right] &= kVar(X) \\
Var\left[\frac{1}{k}(X_1+\ldots+X_k)\right] &= \frac{1}{k}Var(X)
\end{aligned}
$$

Üstel RV'lerin normal dağılıma yaklaşan ortalaması üstel dağılıma özgü değildir. O, PDF'si Merkezi Limit Teoreminin doğruluğu için gerekli olan genel koşulu sağlayan rasgele değişkenin özelliğidir.

Soru Üç

Varsayalım ki ortalaması μ ve varyansı 3 olan bir normal dağılımdan *n* büyüklüğünde bir rasgele örneklem seçildi ve ortalaması, $\overline{X}_n$, hesaplandı.

1. $\overline{X}_n$'in ortalama μ'nun 0.1 kadar etrafında olma olasılığının en az %90 olması için n ne olmalıdır?[1]

[1] Bunlar bizim güç hesapları olarak adlandırdığımız şeylerdir ve çok para harcamadan önce eğer varsa ekonomik olarak anlamlı etkileri belirlemek için bize garantili bir şekilde anket çalışması ve deney oluşturma konusunda yardımcı olurlar. Örneğin genellikle log ücretlerin normal dağılımlı olduğunu düşünürüz. Bazen, log ücretlerin ortalamasının gerçek değerine yeterince yakın olması için bir anket çalışmasında kaç tane gözleme ihtiyacımız olduğunu bilmemiz gerekir. Dahası, Merkezi Limit Teorisi hakkında öğreneceğimiz gibi, ilgilenebildiğimiz herhangi

- 1'in Çözümü: Güç hesabı şöyledir:

$$\begin{aligned} P\left(\left|\bar{X}_n - \mu\right| \leq 0.1\right) &\geq 0.90 \\ P\left(\left|\frac{\bar{X}_n - \mu}{\sqrt{3/n}}\right| \leq \frac{0.1}{\sqrt{3/n}}\right) &\geq 0.90 \\ P\left(|Z| \leq \frac{0.1}{\sqrt{3/n}}\right) &\geq 0.90 \end{aligned}$$

bunun sağlanacağı minimum n'i bulmak maksadıyla eşitlik için çözebiliriz (n +1 için olasılığın yükselişini veya düşüşünü kontrol ederek gerçekten bunun bir minimum olup olmadığını kontrol edebilirsiniz). Olasılık ifadesinin sağlanması için önce Z için kritik değere ihtiyacımız var:

$$\begin{aligned} P\left(|Z| \leq z\right) &= 0.90 \\ z &= 1.6449 \end{aligned}$$

Ortalama için şimdi bunu kritik değere eşitleyeceğiz:

$$\begin{aligned} 1.6449 &= z = \frac{0.1}{\sqrt{3/n}} \\ n &= \left\lceil \left(\frac{1.6449}{0.1}\sqrt{3}\right)^2 \right\rceil \end{aligned}$$

Bu bize n = 812 verir.

2. $\overline{X}_n$'in ortalama μ'nun 0.01 kadar etrafında olma olasılığının en az %90 olması için n ne olmalıdır?

- 2'nin Çözümü: Aşağıdaki ifadeyi elde etmek için uzaklık için sadece farklı değerler koyacağız:

---

bir ortalama için (n ≥ 30 durumunda) örneklem elde ettiğimiz dağılımın genele olarak n'in hesaplamasını etkilemediğini göreceğiz.

$$n = \left\lceil \left(\frac{1.6449}{0.01} \sqrt{3} \right)^2 \right\rceil$$

bu durumda n = 81171'dir. Dolaysıyla, örneklemde 100 katlık artış, doğrulukta 10 katlık artışı, O(δ), sağlar.

3. Yahoo! tecrübeme dayanarak, online reklam gösterimlerinin etkinliğini belirlemek için denek ve kontrol farklılıklarına bakıyordum. Aylık satışların standart sapması (ölçekli) 15.00 dolardır ve ortalama haftalık alışlar kontrol grubu için 1 dolar ve denek grubu için 1 dolar + δ'dır. Eğer kontrol grubundaki gözlemlerin yalnızca %25'ine ve denek grubun %75'ine bakmakla sınırlandırılmış olsaydım, $\overline{X}_T$ - $\overline{X}_C$ arasındaki farkın varyansı ne olurdu? Burada $\overline{X}_T$ tüm denek kişilerinin örneklem ortalaması ve $\overline{X}_C$ bütün kontrol kişilerinin ortalamasıdır. $\overline{X}_T$ - $\overline{X}_C$ > 0 olasılığının en az %95 olabilmesi için, N ne kadar büyük olmak zorundadır? Denek ve kontrol grupları için örneklem ortalamasının normal dağılımlı olduğunu varsayınız. Bunu genel δ için yapınız ve δ = 0.5 dolar ile δ = 0.10 dolar için hesaplayınız. N'nin bu değerleri büyük mü? Sonuçları yorumlayınız.

- 3'ün Çözümü: Δ gibi bir rasgele değişken ile ilgileniyoruz. Δ = . $\overline{X}_T$ - $\overline{X}_C$'dir, beklenen değeri $\mathbb{E}[\Delta] = \mathbb{E}[\overline{X}_T - \overline{X}_C] = \mathbb{E}[\overline{X}_T] - \mathbb{E}[\overline{X}_C] = \delta$ ve varyansı Var(Δ) = $\frac{1}{\frac{3}{4}N}$ Var(X_T) + $\frac{1}{\frac{1}{4}N}$ Var(X_C) = $\frac{4+12}{3N}$Var(X) = (16/N)225. $\overline{X}_T$ - $\overline{X}_C$ > 0 olasılığının en az %95 olabilmesi için, N'nin ne kadar büyük olmak zorunda olduğunu bilmek istiyoruz:

$$\begin{aligned} P(\Delta \leq \delta) &\geq 0.95 \\ P\left(\frac{\Delta}{\sqrt{\frac{16}{3N} \cdot 225}} \leq \frac{\delta}{\sqrt{\frac{16}{3N} \cdot 225}} \right) &\geq 0.95 \\ P(Z \leq z^*) &\geq 0.95 \end{aligned}$$

burada z* = $\frac{\delta}{\sqrt{\frac{16}{3N}.225}}$'tir. Dolaysıyla sadece %0.95'lik kritik değeri bulmamız gerekiyor: 1,6449 (Bu son problemdekinin aynısıdır, burada tek yanlı olasılık elde ettik, o da önceki problemin %90'nın yarısıdır). Böylece,

$$1.6449 = z^* = \frac{\delta}{\sqrt{\frac{16}{3N} \cdot 225}}$$

$$N = \left\lceil \left(\frac{1.6449}{\delta} \sqrt{\frac{16}{3} \cdot 225} \right)^2 \right\rceil$$

δ = 0.5 dolar için, N = 1.298.735'in %95'lik keşif olasılığı yaratır. Diğer taraftan %95'lik keşif olasılığı yaratmak için δ = 0.10 dolar için sadece toplam gözlemin 4'te birine ihtiyacımız var, yani N = 324.684. Eğer denek ve kontrol gruplarının her birisini %50 olacak şekilde dengelersek, 4/(16/3) = 0.75 çarpımı kadar daha az N'ye ihtiyaç duyacaktık. Fakat, bu durumda, reklam yapanlar reklamlarını çok fazla kişiye göstermeyeceklerdi. Öğrenmeyi kazanmaya değiş tokuş ederken, ekonomik düşünce bizi daha düşük güce sahip deney yapmaya zorlamaktadır. Dahası eğer kontrol edilebilir özelikleri kullanarak varyansı 15^2'den düşürebilmenin bir yolu olsaydı, sadece hata varyansını hesaplamalarda kullanırdık, ve belki de gözlem sayısını biraz daha düşürürdük (ya da arzuladığımız δ düzeyine düşürerek büyük olasılıklı küçük etkileri keşfe ederdik).

Soru Dört

Bir tablo, hesap makinesi, internet, simülasyon, veya başka herhangi bir yöntemi (kopya hariç) kullanarak aşağıdaki kritik değerleri belirleyiniz:

1. T ~ t dağılımı: $1 - P(|T| \leq t; dof) = 0.05$, dof = 5, 10, 20, 30, 50. Bunlar Normal'in 0.05 kritik değerleri ile nasıl karşılaştırılır? (dof: serbestlik derecesidir, ÇN)

- 1'in Çözümü: Serbestlik derecesi ve t-dağılımı kritik değerler aşağıdaki tablodadır:

```
Q4_1 =
    5.0000    2.5706
   10.0000    2.2281
   20.0000    2.0860
   30.0000    2.0423
   50.0000    2.0086
```

Bunların hepsi 0.05'lik Normal'in kritik değerinden, yani 1.96'ten, büyüktür. Hakkında istatistik hesapladığınız bilimsel sorunuzun çıkarımını etkileyeceği için, ortalamayı küçük sayıdaki gözlem ($N \leq 100$ gibi) ile hesaplamaya çalıştığımızda bunu hatırlamak önemli olacaktır.

2. X ~ χ^2_k dağılımı: P(X ≥ x;k) = 0.05, k = 1,2,3,4,5,100. Bunlar Normal'in 0.05 kritik değerleri ile nasıl karşılaştırılır? (İpucu: k ile böl ve ortalama gibi bir şeyin kara kökünü al. k =1'e yoğunlaş ve sadece yüksek serbestlik derecelerini N(k,2.k)'nin kritik değerleri ile karşılaştır.)

- 2'nin Çözümü:

```
Q4_2 =
    1.0000    3.8415
    2.0000    5.9915
    3.0000    7.8147
    4.0000    9.4877
    5.0000   11.0705
  100.0000  124.3421
```

k = 1 için, elde ettiğimiz kritik değer sadece karesi alınmış Normal kritik değerdir: 1.96^2 = 3.84. Daha büyük k'ler için, ortalamanın karekökü daha düşüktür. Bu sadece ortalamanın yoğunlaşmasına neden olan bir diğer Büyük Sayılar Kanunu örneğidir.

Soru Beş

$Y_1 \sim \chi^2_{k1}$ ve $Y_2 \sim \chi^2_{k2}$iken, $\frac{Y_1/k_1}{Y_2/k_2}$'nin dağılımı nedir? $k_2 \rightarrow \infty$ iken, payın değerine ve varyansına ne olur? (İpucu: İki i.i.d. χ^2_1 rasgele değişkenin toplamı ne olur? O zaman, onların ortalamasının varyansı nedir? Peki onların k_2'lerinin toplamına ve ortalamasına ne demeli?) Oranın yakınsadığı ortalamanın hangi dağılıma ait olduğunu düşünürsünüz (k_1'i sabit tutarak).

- Çözüm: $\frac{Y_1/k_1}{Y_2/k_2}$'nin dağılımı F(k_1,k_2)'dir. Paydanın değeri $k_2\chi^2_1$ rasgele değişkenin ortalamasına yakınsar. Her bir rasgele değişkenin beklenen değeri 1'dir. Böylece, payda 1'e yakınsar ve varyans k2 → ∞ iken, $\frac{1}{k_2}$Var(χ^2_1) = $\frac{1}{k_2}$.2→ 0. Bunun anlamı $Y_1 \sim \chi^2_{k1}$ iken $\frac{Y_1/k_1}{Y_2/k_2} \rightarrow \frac{Y_1}{k_1}$ 'dir. Bundan ötürü, oran χ^2_{k1}'in ölçeklenmiş versiyonuna yakınsar, ki o, büyük k_1 için yaklaşık olarak N(1, $\frac{2}{k_1}$)'dir. Dahası, Eğer şimdi Merkezi Limit Teoremi hakkında düşünecek olursak, $k_1\chi^2_1$'nin toplamı, ortalaması k_1 ve varyansı k_1Var(χ^2_1) = k_1.2'ye eşit olan bir Normal'e yakınsayacaktır.