14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 4

Konrad Menzel

12 Şubat 2009

1. Bayes Teoremi

Son derste koşullu olasılığı açıkladık ve Toplam Olasılık Kanunun A olayın koşulsuz olasılığını P(A) koşulu olasılığa ($P(A|B_i)$) bağlayan bir yol olduğunu gördük. Koşulu olasılıklar arasındaki bir diğer önemli ilişki P(A|B) koşullu olasılığını P(B|A) koşulu olasılıkla ilişkilendiren Bayes kanunudur, yani koşul sırasını tersine çevirme yoludur. Bu sonuç, istatistiğin ve olasılığın birçok alanında önemli bir rol oynar, en önemlisi de B "verisi"ni gözlemlerken" "dünya hali" A'yı "öğrenme" durumlarında oynadığı roldür.

Örnek 1. *Antik Yunanlılar (belli ki henüz istatistik hakkında çok şey bilmiyorlardı) her bir gemi batışından sonra, kurtulan bütün deniz adamlarının deniz tanrısı Poseidan'a dua ettiklerini fark ederler. Bu gözlemden hareketle, onların gerçekten kurtulmasının nedeni dua etmiş olmaları olduğu yorumu yapılır. Bu örnek gerçekte 16ncı yüzyıl İngiliz filozofu Francsi Bacon tarafından gündeme getirilmiştir. İstatistiki terimlerle, "kurtulanlar"ı olay A ve "dua etme"yi olay B olarak tanımlayalım. Böylece, soru dua etmenin kurtulma olasılığını yükseltip yükselmediği olur, yani (P(A|B) > P(A) ≡ p ilişkisi doğru mu, değil mi? Bütün kurtulan denizcilerin dua etmiş olduğu gözlemi P(B|A) = 1'e dönüşür. Bu bilgi gerçekten de dua etmenin kurtulma şansını önemli ölçüde artırıp arıtmadığı sorusunu cevaplandırmak için yeterli midir? P(B|A) ile ilgili bilgiyi kullanarak P(A|B) hakkındaki bilgiyi nasıl öğreniriz?*

Koşullu olasılığın tanımdan aşağıdaki ilişkiyi elde ederiz,

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

İkinci denkliği yeniden düzenleyince

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

ilişkisini elde ederiz. Aşağıdaki olayı bölüntüleyebileceğimizi daha önce görmüştük,

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

Böylece şu sonuca varırız:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

İzleyen teoremde özetlendiği gibi bu sonucu S'nin herhangi bir bölüntüsüne genelleştirebiliriz.

Teorem 1. (Bayes Teoremi) Eğer A_1, A_2, …, S'nin bir bölüntüsü ise, P(B) > 0'lı herhangi bir B olayı için aşağıdakini yazabiliriz

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j\geq 1} P(B|A_j)P(A_j)}$$

- $P(A_i)$ olay A_i'nın önceki olasılığıdır (yani deneyden önceki olasılıktır)
- $P(A_i|B)$ A_i' nin sonraki olasılığıdır (yani deneyden ve B hakkında –Bayes teorimden elde edildiği gibi- bilgi elde edildikten sonraki olasılıktır)

Optimal karar vermenin istatistiksel teorisinin tümü bu basit fikir üzerine kuruludur: Bayes Karar Teorisi

Örnek 2. *Önceki, batan gemiden kurtulan deniz adamları örneği için, P(B|A) = 1'i ve (koşulsuz) deniz adamlarının kurtulma oranını P(A) gözlemleyebiliyorduk. Ancak, $P(B|A^C)$'yi (boğulanlar arasında dua edenlerin oranı) gözlemleyemediğimiz için, dua etmenin kesinlikle kurtulma şansını artırıp artırmadığı ile ilgili soruyu cevaplandırmak için yeterli bilgiye sahip olmadığımızı da görebiliyoruz. Onların da ölüm korkusundan dua ettiğini rahatlıkla varsayabiliriz (yani $P(B|A^C) = 1$). Böylece aşağıdaki ilişkiyi elde edebiliriz,*

$$P(A|B) = \frac{1 \cdot p}{1 \cdot p + 1 \cdot (1-p)} = p = P(A)$$

Antik Yunanlıların muhakemesi, bir bakıma, "kurtulan sapma" sına bir örnektir (hoş! en gerçekçi ifadesiyle): Bayes teoremi bize, eğer sadece kurtulanları gözlemlersek, kurtulamayanlar hakkında çok şey bilmedikçe kurtulan alt-nüfusun neden kurtulduğu konusunda bir yargıya varmayacağımızı gösterir.

Örnek 3. *Bayes kuralının önemli bir uygulaması da tıbbi testlerin nasıl yorumlanacağı ile ilgilidir. Bir doktorun çok nahoş bir hastalık için bir hastaya test uyguladığını varsayalım ve hastanın gerçekte hasta olduğu olayına A diyelim. Test pozitif bir sonuç verebilir, ki biz buna B olayı diyeceğiz, veya negatif sonuç verebilir, bu da B^C'dir.*

Hastanın hastalıklı olup olmadığını belirleme konusunda test tam olarak güvenli değildir, fakat pozitif test sonucunun olasılığı,

$$P(B|A) = \%\,99\ ,\quad P(B|A^C) = \%5$$

Sonuç olarak, hastalığın nispetten ender görüldüğünü ve hastaların yaşı, cinsiyeti ve diğer özelliklerine bağlı olarak nüfusun %5'ni etkilediğin biliyoruz. Diyelim ki test pozitif sonuç verdi. Hastanın gerçekten hastalıklı olmasının (koşullu) olasılığı nedir? Bayes kuralı aşağıdaki ilişkiyi verir

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.05 \cdot 0.995} \approx 0.0905$$

Hastalığın genel yaygınlığı, P(A), oldukça düşük olduğu için, pozitif test sonucu bile hastalığın nispeten zayıf kanıtı oluyor.

Örnek 4. *Romeo ile Juliet bir süredir görüşüyorlar ve bir gün sevgililer günü (hatırlatmak için: bu Cumartesi'dir) gelir. Romeo Juliet'e ya mücevher hediye edebilir, J, ya da ona bir serenat yapabilir, S. Juliet mücevher ister. Eğer Romeo onu gerçekten sevdiyse, onun isteğini gözlerinden okuyabilirdi. Doğrusu, Juliet ona bu isteğini iki hafta önce, Amerikan futbol ligi finalinin son yarım saatinde söylemişti. Juliet aynı zamanda Romeo'nun kendisini hala sevip sevmediği konusunda ilk kez şüpheye düşer. Buna L olayı diyelim. Spesifik olmak gerekirse,*

$$P(L) = 0.95$$

Juliet, Rome kendisini seviyorsa, ona P(J|L) = 0.80 olasılıkla mücevher vereceğini veya

$$P(S|L) = 0.20$$

olasılıkla serenat yapacağını da biliyor (Bu sadece Juliet''in düşündüğüdür, unutmayın ki Romeo Amerikan futbolunu da çok seviyor). Eğer Romeo onu artık sevmiyorsa,

$$P(S|L^C) = 0.80$$

olasılığıyla Juliet'in ne sevdiği konusunda bir fikri olmayacak ya da ona bir serenat yapacaktır (veya daha gerçekçi bir şekilde, önceki sene Juliet'in istediği gülleri verebilir ya da sevgililer gününü tamamen unutabilir). (Not: Serenat yapmak Romeo için çok utanç verici olabilir ama aynı zamanda daha da ucuzdur). Sonuçta Romeo Juliet'e serenat yapar. Juliet onu hemen terk etmeli mi? Bayes teoremine göre, Juliet'in Romeo'nun eğilimi hakkında sonraki inancı aşağıdaki ilişki ile verilir

$$P(L|S) = \frac{P(S|L)P(L)}{P(S|L)P(L) + P(S|L^C)P(L^C)} = \frac{(1-0.8)0.95}{(1-0.8)0.95 + 0.8(1-0.95)} = \frac{\frac{2}{10}\cdot\frac{19}{20}}{\frac{2}{10}\cdot\frac{19}{20} + \frac{8}{10}\frac{1}{20}} = \frac{38}{46} \approx 0.826$$

Bunun Juliet için iyi olup olmayacağına kendisinin karar vermesini bekleyeceğiz.

Gerçek hayatta, bir çok insanın bu tür yargıları çok iyi değildir ve, son iki örnekte olduğu gibi, testlerin güvenirliğini gereğinden fazla önemseme eğilimindedirler. Bilişsel psikoloji literatüründe, bu durum "Temel-Oran Yanılgısı" olarak bilinir. Örneğimizde "temel-oran" etkilenen veya sağlıklı insanların, sırasıyla, P(A) ve P(A^C) oranları ile Romeo'nun Juliet'i sevip, P(L), veya sevmeme, P(LC), ön bilgisidir. Eğer bu olasılıklar çok farklı ise, sezgilerdeki muhakemenin yanılgısı çok fena olabilir.

Örnek 5. *Monty Hall paradoksu*[1]*: Bir zamanlar TV'de bir yarışmacıdan A, B, ve C' gibi üç kapıdan birisini seçmesi istenilen bir programı vardı. Kapıların birinin arkasında bir ödül (son model bir çamaşır makinesi gibi) ve diğer iki kapının arkasında ise birer keçi olurdu. Eğer yarışmacı arkasında ödül olan kapıyı seçerse, ödül kendisinin olurdu. Eğer arkasında keçi olan bir kapıyı açarsa, hiçbir şey kazanmazdı. Oyunu biraz daha ilginç hale getirmek için, yarışmacı ilk seçimini yaptıktan sonra sunucu her zaman diğer iki kapıdan arkasında keçi olan birini açardı. Bu durumda, yarışmacının istemesi halinde seçtiği kapıyı diğer kapalı kapıyla değiştirmesine izin verilirdi. Kapıyı değiştirmek iyi bir fikir olur muydu?*

Genelleştirmeyi bozmadan, varsayalım ki ben A kapısını seçtim. Ödülün A kapısının arkasında olmasının koşulsuz olasılığı 1/3'tür. Eğer ödül gerçekten A kapısının arkasında olsaydı, sunucu, eşit olasılıkla her ikisinin arkasında keçi olan B kapısını ya da C kapısını açacaktı. Eğer ilk tahmin yanlış olsaydı, hem yarışmacı tarafından seçilmeyen hem de arkasında ödül olmayan bir tane kapı kalmış olacaktı. O halde, Benim A'yı seçmem durumunda, sunucunun o esnada C'yi açma olasılığı

$$P(\text{Ödül A'nın arkasında}|\text{C açıldı}) = \frac{P(C\ açıldı|\text{Ödül }A'nın\ arkasında)P(\text{Ödül }A'nın\ arkasında)}{P(C\ açıldı)} = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

Diğer taraftan

$$P(\text{Ödül B'nın arkasında}|\text{C açıldı}) = \frac{P(C\ açıldı|\text{Ödül }B'nın\ arkasında)P(\text{Ödül }B'nın\ arkasında)}{P(C\ açıldı)} = \frac{1\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Bu durumda, kapıları değiştirerek ödülü kazanma şansımı yükseltirdim.

Sezgisel olarak, yeni açılan kapı ödülün A kapısının arkasında olma ihtimali konusunda hiç bilgi içermemektedir, çünkü sunucu onu hiçbir şekilde açmayacaktır. Gerçekte A'yı seçmemiz halinde, "A'da ödül var" ile "C açıldı" olayları bağımsızdır. Ancak, sunucunun B kapısını açmaması iki nedenden kaynaklanabilirdi: (1) ödül A kapısının arkasındaydı, sunucu C'yi tamamen rasgele açtı, (2) ödül B kapısının arkasındaydı, sunucu başka

[1] Bu konudaki tartışmayı şu adreste okuyabilirsiniz: http://people.csail.mit.edu/carroll/probSem/Documents/Monty-NYTimes.pdf

seçeneği olmadığı için C'yi açtı. Dolaysıyla, C olasılığını ortadan kaldırmak sadece B'ye "avantaj" sağlar.

2. Özet Bölüm 1: Olasılık

Dersin ikinci ünitesine geçmeden önce şimdiye kadar yaptıklarımızı, aşina olduklarınızı ve kendinizi rahata hissedeceklerinizi özetleyelim:

2.1 Sayma Kuralları

- *N* arasından *n*'i yere *koyarak* çekmek: N^n olanak
- *N* arasından *n*'i yere *koymadan* çekmek: $\frac{N!}{(N-n)!}$ kadar olanak
- Permutasyonlar: *N!* olanak
- *N*'den *n* kombinasyonu: $\binom{N}{n}$ olanak

2.2 Olasılıklar

- Bağımsızlık: *P(AB) = P(A)P(B)*
- Koşullu olasılık: Eğer P(B) > 0 ise, P(A|B) = $\frac{P(AB)}{P(B)}$
- Sadece ve sadece A ve B bağımsız ise P(A|B) = P(A),
- Toplam Olasılık Kanunu: P(Bi) > 0 için S'nin $B_1, \ldots, B_n$ bölüntüsü aşağıdaki gibidir

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

- Bayes Teoremi

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Olasılık muhakemesi için gördüğümüz birkaç genel şey daha var

- manipülasyonları kullanarak ilgilenilen olayı yeniden formüle ederek olasılıkları kolay hesaplanabilir bir şeye dönüştürme (örneğin tümleyen, bölüntüler vs.)
- temel oranın koşullu olasılığı marjinal/koşulsuz olasılığa dönüştürme önemi (Örneğin Bayes Teoreminde olduğu gibi veya kalp cerrahı örneğindeki kompozisyon etkisi gibi).

- bazen A ve B bağımlı olaylarını C’ye koşullayarak bağımsızlaştırabilme (veya bağımsız olayları bağımlı yapmak).

3 Rasgele Değişkenler

Şimdi dersimizin ikinci büyük konusu olan rasgele değişkenler ile devam edelim.

Örnek 6. *Yazı Tura atma, versiyon I: Tura, H, geldiğinde 1, yazı, T, geldiğinde 0 değerini alan bir X değişkeni tanımlayalım. Bu rasgele deney için örneklem uzayı S = {H, T} ve rasgele değişkenin genişliği {0, 1}’dir.*

Tanım 1. *Reel-değerli bir rasgele X değişkeni bir deneyin sonuçlarını reel sayılar ile eşleştiren aşağıdaki gibi herhangi bir fonksiyondur.*

$$X : \begin{cases} S \to \mathbb{R} \\ s \mapsto X(s) \end{cases}$$

Tarihsel olarak, 1800 yıllarda rasgele değişken fikri geliştirildiği zaman matematikçiler ile diğer bilim adamlarının düşüncesinde “hakiki” rassalığın işlevi yoktu. Onun yerine, şans, analiz ettiğimiz durumun bütün parametreleri hakkında tam bilgi sahibi olmayışımız ve bir deneyin sonucunu tahmin etmede doğanın kanunlarından (güya tam belirleyici) faydalanma konusundaki yeteneksizliğimizin bir sonucu olarak görülüyordu. Bunların tümünü yapabilmek “Laplace Şeytanı” olarak bilinir ve ünlü matematikçi Pierre Simon de Laplace aşağıdaki gibi açıklar:

> Zamanın herhangi bir anında bir akıl doğaya canlılık veren bütün güçleri ve onu oluşturan bütün varlıkların karşılıklı duruşunu bilmiş olsaydı, eğer bu akıl verisini analiz etmek için verecek kadar engin olsaydı, kâinatın büyük bedenlerinin ve en hafif atomların hareketini bir tek formül ile özetleyebilseydi: böyle bir akıl için hiçbir şey belirsiz olmazdı ve gelecek tıpkı geçmiş gibi gözlerinin önünde olurdu.[2]

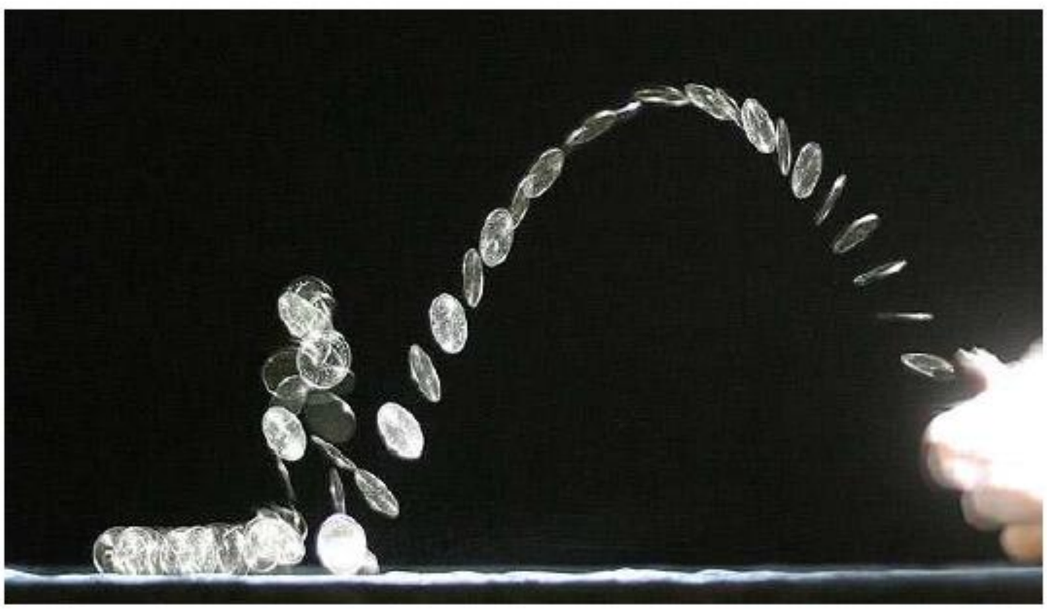
Ardından gelen Fizik’teki (örneğin quantum fiziğindeki gerçek belirsizlik) veya hesap teorisindeki (örneğin Gödel teoremi: Bir akıl kendisinden daha karmaşık olmalı çünkü tahminleri tahmin etmeye çalıştığı kâinatın bir parçasıdır) gelişmeler bu dünya görüşünü korumadı ama hala bizim olasılık kavramının temel altyapısını oluşturmaya devam etmektedir: etrafımızdaki dünyanın rassalığı esas itibariyle onun hakkındaki yetersiz bilgimizi yansıtmaktır.

[2] Laplace, P. (1814): A Philosophical Essay on Probabilities.

Örnek 7. *Bir örnek olarak, Yazı Tura atmanın İkinci Versiyonu: Laplace'nin fikrini göstermek için, örneklem uzayının yukarıda ilk yapılanlardan daha karmaşık tanımını düşünebiliriz: Klasik mekanikte herhangi bir zaman noktasında paranın durumunun (katı cisim) tam bir açıklamasını verebiliriz (en azından prensipte). Ve o zaman klasik mekaniğin kanunlarını kullanarak onun tam yönünü ve özelike Tura (H) veya Yazı (T) ile sonuçlanıp sonuçlanmayacağını tahmin ederiz. Daha spesifik olarak, paranın havaya atıldığı andaki mekanik sistemin durumunu örneklem uzayı olarak açıklayabilirdik. Bir sistemin durumunun tam açıklaması (çok ideal!) (1) konumu, (2) paranın kütlesinin merkezinin ivmesi ile (3) yönü ve (4) belli bir* t_0 *zamanındaki açısal momenti ile verilir. Bunların her birisinin üç koordinatı vardır, yani* $S = \mathbb{R}^{12}$. *Her nokta* $s \in S$ *açıkça iki olaydan {H, T} birine aittir. Turanın gelmesi olayı,* $H \subset S$, *için* $X = 1$ *ve yazı için* $X = 0$ *değerlerini verirsek, bu eşleştirme aşağıdaki gibi ifade edilen rasgele değişkendir.*

$$X := \mathbb{R}^{12} \longrightarrow \{0, 1\},$$

$$X: \begin{cases} s \longrightarrow 1 & Eğer\ s \in H \\ s \longrightarrow 0 & diğer\ durumlarda\ yani\ eğer\ s \in T \end{cases}$$



Şekil 1: Metalik para atmanın stroboskopik resmi (Andrew Davidhazy ve Rozhester Institute of Technology, School of Photo Arts and Sciences'in müsaadesiyle. İzin alınarak kullanılmıştır. © Andrew Davidhazy, 2007)

Problem –neredeyse tam olarak- çok belirsiz olduğu için (knife-edged), sonuç, başlangıç durumundaki $s \in S$*'deki ufacık değişikliklere karşı çok duyarlıdır (örneğin, diyelim ki, yandan geçen bir arabanın yarattığı çekim gibi dışsal etkilemeler hesaba bile* *katılmadan). Parayı nasıl fırlattığımıza bakmadan, istenilen sonucu kesin olarak verecek başlangıç durumunu, ivmeyi vs. kontrol altında tutmak tamamen imkânsızdır. Ayrıca, tipik olarak bir sistemi açıklayan diferansiyel denklemleri kesin doğrulukla çözmek te imkânsızdır. Bu nedenle, sadece S'nin bir parçası olma olasılıklarını verebiliriz, ki bu H ve T sonuçlarının olasılıklarını eşleştirir. Dolayısıyla, bu durumda "hakiki" bir rassallığa ihtiyaç olmazsa bile, bu pratikte bizim için nasıl çalıştığını gösterir.*

Bu tanım rasgele değişken ve olasılık hakkında ne düşündüğümüz konusunda felsefi noktayı öne çıkarırken (belli ki çok kullanışlı değil), uygulama amacıyla, problemin ilk açıklama şekline bağlı kalmayı tercih edeceğiz.

Açıklama 1. *X genişliğindeki herhangi bir $A \subset \mathbb{R}$ olayı için, örneklem uzay S'nin olasılık fonksiyonu aşağıdaki ifade aracılığıyla X için bir olasılık dağılım fonksiyonu meydana getirir.*

$$P(X \in A) = P(\{s \in S : X(s) \in A\})$$

X biçimsel olarak örneklem uzayını reel sayılara dönüştüren bir fonksiyon olmasına rağmen, genellikle onu değişken olarak ele alırız. Yani argümanları belirtmeden onun olasılıkla ilintili çeşitli değerler "aldığını" söyleriz. Başka bir ifadeyle, uygulamaların çoğu için, ilgili herhangi bir örneklem uzayı S ve S'nin olasılıklarını referans göstermeden, sadece $P(X \in A)$'i belirtiriz. Örneğin yazı tura örneğinde – yukarıda açıklandığı gibi- S'deki koordinatlar(metal paranın başlangıç durumu, ivme, yönü gibi) arasındaki doğru ilişkiyi, sonuçlarını (sayısal olarak imkansızdır) ve koordinatların bir olasılık dağılımını anlamaya çalışmayacağız, sadece (P(X = 1) = P(X = 0) = $\frac{1}{2}$'yi bilmemiz yeterlidir.

Örnek 8. *Eğer 10 tane metalik parayı birbirinden bağımsız olarak atarsak, Bir rasgele değişken tanımlayabiliriz, X = (Toplam Yazı Sayısı). Bu tür rasgele değişkenlerin dağılımlarını aşağıda detaylı bir şekilde analiz edeceğiz.*

Örnek 9. *Bir seçimin sonuçları ile ilgileniyor olabiliriz. Diyelim ki 100 milyon seçmen ve iki aday var. Her seçmen adaylardan sadece birisi için oy kullanabilir, hangi seçmenin hangi aday için oy kullandığına bağlı olarak $2^{100\ 000\ 000}$ kadar farklı sonuç vardır. Şimdi, Aday A için (ve esas ilgilendiğimiz seçim için) kullanılan toplam oylar ile ilgili olarak bir rasgele X değişkeni tanımlayabiliriz. Oyların sayısının her bir değeri için, esas sonuçlarla ilintili bir sayı vardır, yani aday A'nın bütün sonuçları almasının tek bir yolu vardır. Her bir sonucun olasılığını basit olasılık cinsinden formüle edebiliriz ve oradan hareketle benzer değerler üzerinde toplulaştırma yaparak verilen toplam oy sayısının olasılığını elde edebiliriz.*

Açıklama 2. *Rasgele olayların tümü, ilgilendiğimiz gibi, sayısal bir özeliğe sahip değildir (örneğin, eğer olay "yarın yağmur yağacak" ise, ne kadar yağacağı ile ilgilenmeyebiliriz).*

O zaman rasgele değişkene ilişmemek gerekir, sadece olaylarla daha önceki gibi ilgilenebiliriz. Başka bir seçenek olarak ta, olayın geçekleşmesi halinde 1 diğer durumlarda 0 alan bir rasgele değişken tanımlayabiliriz (gelecekte bazen "hile" ye başvuracağız).