Olasılık Kuramı

Aynı Renk Top Beklentisi

(Yanlış kanıtla doğru yanıtı bulma sanatı!)

Evsel Atık

Geçenlerde bir dostum (Serdar Boztaş) üyesi olduğum bir e-posta grubuna şu soruyu yolladı: İçinde n tane siyah, n tane beyaz top bulunan bir kutudan teker teker rastgele top çekiyorsunuz. Kutuda aynı renk top kaldığında duruyorsunuz. n'nin çok çok büyük bir sayı olduğunu varsayarsak, kutuda ortalama aşağı yukarı kaç top kalır? Daha matematiksel deyişle, kutuda kalan top sayısı beklentisi kaçtır? Soruyu şöyle de dile getirebiliriz: Eğer bu oyunu $n = 100.000$ için, yani 200.000 topla 1 milyon defa oynasanız ve sürecin sonunda her oyunda kalan top sayısını toplasanız ve bu toplamı 1 milyona bölseniz, aşağı yukarı kaç elde edersiniz?

Seçenekler de sunulmuş:

a. Bir $\varepsilon > 0$ sayısı için aşağı yukarı εn kadar.
b. Aşağı yukarı $\sqrt{n}$ kadar.
c. Aşağı yukarı $\log n$ kadar.
d. Aşağı yukarı bir sabit.

Serdar Boztaş dostum, hesap kitap yapmayın, program yazmayın, internette cevabı aramayın, sadece sezgilerinizi kullanarak yanıtlayın diye de eklemiş. Zaten mesajının başlığı "sezginizi sınayın" idi.

İnsanın aklını başından alacak güzellikte bir soru. Hesap kitap yapmadan durmak da zor. Ama kendimi engelledim, sezgilerimle düşünmeye çalıştım. İlk önce b ve c seçeneklerinden birinin olması gerektiğinde karar kıldım. n ne kadar büyükse, kalan top sayısı da o kadar fazla olması gerektiğinden d şıkkını elimine ettim. a şıkkı da pek makul gelmedi.

Ama biraz daha düşününce bal gibi a ve d seçeneklerinin mümkün olabileceğine inanmaya başladım. Zaten sorunun yazılış ve sunuş biçiminden yanıtın şaşırtıcı olması gerektiği belli. Bu yönden bakınca d seçeneği doğru seçenekmiş gibi gelmeye başladı.

Sezgi bir yere kadar... Bir zaman sonra e-posta grubunda kim varsa kalem kâğıda sarılıp hesap yapmaya başladı.

$2n$ topla başlanan bir oyunda ortalama kalan top sayısına $f(n)$ diyelim. Oyunda en az 1 top kalacağından $f(n) \geq 1$ olur. Ama $f(n)$ her zaman bir tamsayı olmayabilir, ne de olsa deneylerin ortalamasını alıyoruz.

Önce küçük n sayıları için yanıtı bulmaya çalışalım.

Eğer $n = 1$ ise, yani 1 siyah ve 1 beyaz topla oyuna başlarsak, ilk çekişten sonra bir top kalır ve sonuç 1'dir. Bu durumda beklenti 1 olur. Demek ki $f(1) = 1$.

$n = 2$ durumunu ele alalım. Toplara

$$K_1, K_2, M_1, M_2$$

diyelim. Bu dört top arasından çekilen ilk iki top için 4'ün 2'lisi kadar, yani toplam 6 farklı seçenek var. Bu 6 seçeneğin olasılığı hep aynı, her biri 1/6. Ve bu seçeneklerin 2'si aynı renkten, geri kalan 4'ü iki farklı renkten. Demek ki 2/6 olasılıkla ilk iki çekişte aynı renk top çekeceğiz ve oyun sonuna 2 top kalacak ve 4/6 olasılıkla ilk iki çekişte 1 siyah ve 1 beyaz top çekeceğiz ve o zaman oyun sonuna sadece 1 top kalacak. Demek ki bu durumda beklenti

$$\frac{2}{6} \times 2 + \frac{4}{6} \times 1 = \frac{4}{3}$$

olur. Böylece $f(2) = 4/3$ değerini bulduk.

Eğer $n = 3$ ise, biraz daha karmaşık bir hesapla beklentinin 3/2 olduğunu gösterilebilir.

Birinci Deneme. Önce a, b ve c yanıtlarının doğru olamayacağını sezgisel olarak görmeye çalışalım. Bunu görmek için $f(n)$'nin n ile birlikte çok artamayacağını göstermek yeter, çünkü ilk üç yanıtta da n sonsuza giderken $f(n)$ de sonsuza gidiyor. Diyelim $f(n)$ çok çok büyük bir sayı. Bir çelişki elde edeceğiz. $2n$ tane topla oyuna başladığımızı düşünelim. Muhtemelen oyunun sonunda $f(n)$'ye çok yakın sayıda top kalacak, çünkü böyle bir oyunda "standart sapma" çok büyük olamaz, yani oyunların sonuçları genellikle $f(n)$'ye yakın bir sayı olmalı. Oyun sonuna $f(n)$'ye çok yakın sayıda top kalmasa da $f(n)$'den büyük sayıda top kaldığı bayağı oyun oynamamız lazım. İşte bu oyunlardan birini oynadığımızı varsayalım. Bu oyunun son hamlesinden hemen önce kutuda aşağı yukarı $f(n) + 1$ tane top varmış ve bu topların biri bir renkten, diğerlerinin hepsi diğer renktenmiş. Ve biz son hamlemizde aşağı yukarı $f(n) + 1$ kadar olan top arasından seçe seçe bu topu seçmişiz. Bunun olasılığı da aşağı yukarı

$$\frac{1}{f(n)+1}$$

dir. Ama $f(n)$ çok büyük olduğundan, bu olasılık çok küçüktür... Demek ki son hamlenin olasılığı çok küçükmüş. Mümkün değil! Çok küçük olasılıklı bir olay arada bir gerçekleşebilir ama çok sık gerçekleşemez.

Her ne kadar ikna ediciyse de, bu akıl yürütmenin pek matematiksel olduğu söylenemez. Gene de bir gerçeği işaret ettiği kesin. Bir kanıt değil belki ama kesinlikle bir delil, ve güçlü bir delil.

İkinci Deneme. (Serdar Boztaş'ın denemesi). Bu sefer oyunu tersten oynayalım. Kutudan topları çekeceğimize topları rastgele teker teker seçip boş kutuya atalım ve en başından itibaren boş kutuya ardarda aynı renkten ortalama kaç top atacağımızı hesaplamaya çalışalım. Önümüzde (kutunun dışında) çok sayıda siyah ve beyaz top olsun. Oyunun ilk aşamalarında kutuya beyaz (ya da siyah) top atma olasılığı 1/2'ye çok yakındır, bu olasılığın tam 1/2 olduğunu varsayalım, en azından oyunun başlarında bu varsayımla fazla hata yapmayız. Birinci topu attık. İkinci topun birinci topla aynı renkten olma olasılığı 1/2'dir. İlk üç topun da aynı renkten olma olasılığı $1/2^2$'dir. İlk dört topun da aynı renkten olma olasılığı $1/2^3$'tür. Topların aynı renk olma olasılıkları giderek azalıyor, olasılık her seferinde yarıya iniyor. Mesela 10 tane aynı renkten top atma olasılığı 1/500'e yakın bir olasılık, çok düşük! Dolayısıyla çok çok şanslı bir günümüzde değilsek, kutuya çok sayıda aynı renk top atamayız... İşlemi tersine çevirirsek, yani sorudaki oyunu oynarsak, en sonda kalan aynı renkten top sayısının sınırlı olması gerektiğini görürüz.

Bu akıl yürütmenin de tam matematiksel olduğunu söyleyemeyiz. Ama bayağı ikna edici, hatta bir öncekinden daha fazla ikna edici sanki.

Üçüncü Deneme. Biraz daha matematiksel olalım. Bu sefer şöyle düşünelim: Topları oyundan önce numaralandıralım ve kutudan çıkardıkça bir masanın üstüne soldan sağa doğru sıraya dizelim. Kutu boşaldığında $2n$ tane top masanın üstüne soldan sağa doğru sıraya dizilmiş olacak. En sağda, kutuda en son tek başına kalan top olacak. Oyunun sonucu en sağda bulunan aynı renk top sayısıdır.

Toplamda $2n$ top var ve $2n$ top masaya $(2n)!$ farklı biçimde dizilir. Topları masaya rastgele dizdiğimizden her dizilimin boy gösterme olasılığı

$$1/(2n)!$$

dir.

Dizinin en sağında ardarda tam (ne eksik ne fazla!) k tane beyaz top olma olasılığını hesaplayalım. Bu olasılık elbette kutuda en sonda tam k tane beyaz top kalma olasılığıdır. Aynı hesap beyaz yerine siyah için de geçerli olduğundan, oyunun sonunda k puan kazanma olasılığımız $2p(k)$ olur ve böylece beklentiye $2kp(k)$ eklenir. Eğer bu $p(k)$ sayısını hesaplayabilirsek, bulmak istediğimiz beklenti,

$$f(n) = 2\sum_{k=1}^{n} k \cdot p(k) \qquad (1)$$

şeklini alır. Şimdi iş bu $p(k)$ sayısını bulmaya geldi. Bunun için de en sonda k beyaz top gelecek biçimde olası kaç farklı diziliminin olduğunu bulalım.

En sağdaki k tane beyaz toptan hemen önce siyah bir top olmalıdır elbette, ne de olsa en sonda (ne eksik ne fazla!) tam k tane beyaz top kalmasını istiyoruz. En sağdaki k tane beyaz top n tane beyaz top arasından seçiliyor. Bu seçimi,

$$\binom{n}{k}$$

farklı biçimde yapabiliriz. En sağa gelecek k tane beyaz topu seçtikten sonra bu k tane beyaz topu $k!$ farklı biçimde sıraya dizebiliriz. Demek ki en sağdaki k beyaz top için,

$$k!\binom{n}{k}$$

farklı diziliş var.

En sağda bulunan bu k tane beyaz topun hemen solunda siyah bir top olmalı. Bu siyah top için de n tane seçeneğimiz var.

En sağdaki k tane beyaz topu ve bu k tane beyaz topun hemen soluna siyah bir top yerleştirdik. Yani toplam $k + 1$ tane top yerleştirdik. Geriye

$$2n - (k + 1)$$

tane top kaldı. Bu toplar dizinin en soluna yerleştirilmeli. Bunun için de

$$(2n-(k+1))!$$

tane seçenek var.

Böylece, yarısı siyah yarısı beyaz $2n$ adet topu, en sağa tam k tane beyaz top gelecek biçimde

$$\binom{n}{k}k!n(2n-k-1)!$$

farklı dizilebilir. Bu dizilimlerin her birinin sonucu k ve olasılığı $1/(2n)!$. Aynı şey siyah toplar için de geçerli. Demek ki (1) formülünden dolayı, yukarıdaki sayıyı $(2n)!$ sayısına bölüp, k ile çarpıp, tüm k'lar için toplarsak oyunumuzun beklentisini buluruz:

$$f(n)=\frac{2}{(2n)!}\sum_{k=1}^{n}k\binom{n}{k}k!n(2n-k-1)! \qquad (2)$$

Bu sayıyı hesaplamalıyız. İfadeyi sadeleştirelim.

$$\begin{aligned}
f(n) &= \frac{2}{(2n)!}\sum_{k=1}^{n}k\binom{n}{k}k!n(2n-k-1)!\\
&= \frac{2}{(2n)!}\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}k!n(2n-k-1)!\\
&= 2n\frac{n!}{(2n)!}\sum_{k=1}^{n}k\frac{(2n-k-1)!}{(n-k)!}\\
&= 2n\frac{n!n!}{(2n)!}\sum_{k=1}^{n}k\frac{(2n-k-1)!}{n!(n-k)!}\\
&= 2\frac{n!n!}{(2n)!}\sum_{k=1}^{n}k\frac{(2n-k-1)!}{(n-1)!(n-k)!}\\
&= \frac{2}{\binom{2n}{n}}\sum_{k=1}^{n}k\binom{2n-k-1}{n-1}.
\end{aligned}$$

Böylece,

$$f(n)=\frac{2}{\binom{2n}{n}}\sum_{k=1}^{n}k\binom{2n-k-1}{n-1} \qquad (3)$$

buluruz. Tabii hesapların doğruluğunu kontrol etmek için, yukarıdaki formülde $n = 1, 2, 3$ aldığımızda sırasıyla 1, 4/3 ve 3/2 bulacağımızdan emin olmalıyız. Bu kontrolü müsveddede yaptık.

Şimdi sıra bu toplamı hesaplamaya geldi, hesaplayamasak da n çok büyükken alacağı değerleri kestirmemiz lazım. n çok büyükken, yazının ilk iki adımında bu sayının bir sabite yakın olduğu tahmininde bulunmuştuk.

Dördüncü Deneme. Problemi daha da genel olarak kavramak amacıyla, siyah ve beyaz top sayısını eşit almayalım, siyah top sayısı s, beyaz top sayısı b olsun, çünkü ne de olsa her hamleden sonra kutudaki siyah ve beyaz top sayısı değişiyor.

Önce s tane siyah topu belli aralıklarla soldan sağa doğru sıraya dizelim. s tane siyah top $s + 1$ tane aralık belirler: en soldaki topun solu, ilk iki topun arası, ikinci ve üçüncü topların arası vs ve en sağdaki topun sağı. Bu $s + 1$ tane aralığa b tane beyaz topu rastgele bir biçimde yerleştirsek, en sağdaki aralığa ortalama kaç top düşer? $s + 1$ tane aralığın birinin diğerinden farkı olmadığı için her aralığa ortalama

$$\frac{b}{s+1}$$

tane beyaz top düşer. Dolayısıyla en sağdaki aralığa da bu kadar beyaz top düşer. Yani kutudan rastgele toplar çektiğimizde bu sayı kadar beyaz topla oyunu bitiririz... Aynı şey beyazla siyahı değiştirirsek de geçerlidir. Demek ki, bu oyunun beklentisi

$$f(s,b)=\frac{b}{s+1}+\frac{s}{b+1} \qquad (4)$$

olur.

Bu akıl yürütme de pek ikna edici değil doğrusu... Topları kutudan çıkarıp sıraya dizerek problemimizin bir modelini çıkarmaya çalıştık ama modelin doğru olup olmadığı hakkında bir fikrimiz yok. Yukarıdaki paragrafta kullandığımız olasılık dağılımının asıl problemdeki olasılık dağılımıyla bir ilgisinin olup olmadığı kuşkulu.

Belki yanlış, ama gene de bir yanıt bulduk!

Şimdi asıl problemde olduğu gibi $s = b = n$ alalım. O zaman yukarıda bulduğumuz yanıt,

$$f(n)=\frac{2n}{n+1} \qquad (5)$$

şeklini alır.

Bakalım bu yanıt $n = 1$, 2 ve 3 için ilk sayfada bulduğumuz $f(1) = 1$, $f(2) = 4/3$, $f(3) = 3/2$ yanıtlarından ne kadar uzak. Ufak bir hesap yaptığınızda göreceksiniz ki $n = 1, 2, 3$ iken

$$\frac{2n}{n+1}$$

ifadesi, aynen $f(n)$ gibi, sırasıyla 1, 4/3 ve 3/2 değerlerini alıyor!

Çok tuhaf! Muhtemelen f'nin bir sonraki değerinde başka başka yanıt bulunur... Hesaplamadan bilinmez!

Beşinci Deneme. Şimdi gerçekten matematiksel olalım. Gene s tane siyah topla ve b tane beyaz topla oynanan oyunu düşünelim. Bu oyunda en sona kalan aynı renk top sayısına

$$f(s, b)$$

diyelim. Sorumuz n çok çok büyükken $f(n, n)$ sayısının aşağı yukarı kaç olacağı. Elbette,

$$f(s, 0) = s \text{ ve } f(0, b) = b$$

olur.

$f(s, b)$ sayıları arasında tümevarımsal bir ilişki bulalım: Diyelim s tane siyah ve b tane beyaz topla oyuna başladık. Eğer $s = 0$ ya da $b = 0$ ise yanıtı bildiğimizden, s ve $b > 0$ varsayımını yapalım. İlk topumuzu seçmek üzereyiz.

$$\frac{b}{b+s}$$

olasılıkla beyaz top seçeceğiz ve kutuda $b - 1$ tane beyaz, s tane siyah top kalacak; bu durumda beklentimizin $f(s, b - 1)$ olduğunu biliyoruz. Öte yandan,

$$\frac{s}{b+s}$$

olasılıkla beyaz top seçeceğiz ve kutuda $s - 1$ tane siyah, b tane beyaz top kalacak; bu durumda da beklentimizin $f(s - 1, b)$ olduğunu biliyoruz. Demek ki,

$$f(s,b) = \frac{b}{b+s} f(s,b-1) + \frac{s}{b+s} f(s-1,b). \qquad (6)$$

İşte bu tümevarımsal ilişkimiz. Bu formülü kullanarak $f(s, b)$ sayılarını teker teker bulabiliriz. Mesela

$$f(2,1) = \frac{1}{1+2} f(2,0) + \frac{2}{1+2} f(1,1) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

olur. Problemin simetrisinden dolayı $f(s, b) = f(b, s)$ olduğundan, yukarıdaki tümevarım ilişkisinde $s = b = n$ alarak,

$$f(n, n) = f(n, n - 1)$$

buluruz. Demek ki,

$$f(2, 2) = f(2, 1) = 4/3,$$

aynen yazının başlarında bulduğumuz gibi. $f(3, 3)$ sayısını bulmak için, önce $f(3, 1)$'i, sonra $f(3,2)$'yi hesaplayalım:

$$f(3,1) = \frac{1}{1+3} f(3,1-1) + \frac{3}{1+3} f(3-1,1)$$
$$= \frac{1}{4} 3 + \frac{3}{4}\frac{4}{3} = \frac{7}{4},$$
$$f(3,2) = \frac{2}{2+3} f(3,2-1) + \frac{3}{2+3} f(2,2)$$
$$= \frac{2}{5}\frac{7}{4} + \frac{3}{5}\frac{4}{3} = \frac{42+48}{60} = \frac{3}{2}.$$

Demek ki,

$$f(3, 3) = f(3, 2) = 3/2,$$

daha önce bulduğumuzu söylediğimiz gibi.

Sonuç olarak,

$$\left.\begin{array}{l} f(s,0) = s,\ f(0,b) = b. \\ f(s,b) = \frac{b}{b+s} f(s,b-1) + \frac{s}{b+s} f(s-1,b) \end{array}\right\} \qquad (7)$$

eşitliklerini sağlayacak iki değişkenli bir $f(s, b)$ fonksiyonu bulmalıyız. Fikri olan beri gelsin!

Öldürücü Darbe: Yukarıdaki (7) eşitliklerini sağlayacak bir $f(s, b)$ fonksiyonu arıyoruz. Ya da aynı şey, (3)'teki toplamın kapalı bir biçimini arıyoruz.

Tabii (3)'teki toplamı bulmak için yaptığımız akıl yürütmeyi s siyah ve b beyaz topla yaparsak (7) eşitliklerini sağlayan bir ifade buluruz ama, zaten (3)'teki toplam hoş bir ifade değil ki, bu yöntemle bulunacak bir ifade hoş olsun! Mümkünse (7)'yi sağlayan ve kapalı bir formülle tanımlanmış bir fonksiyon bulmak istiyoruz.

Şimdi arkanıza yaslanın ve emniyet kemerlerinizi bağlayın: Dördüncü denememizde saçmasapan bir akıl yürütmeyle bulduğumuz (4)'teki fonksiyon (7) eşitliklerini sağlıyor. (Okur lütfen hesapları yapsın.) Demek ki (4)'teki fonksiyon doğru yanıtmış. Yani yanıt,

$$f(n) = \frac{2n}{n+1}$$

imiş. n çok büyükken bu sayı 2'ye yakındır. Demek ki çok sayıda topla oynandığında, en sonda ortalama 2 top kalır, yani Serdar Boztaş'ın sorusunda d şıkkı doğru yanıttır.

Bu arada, (4)'teki toplamın ne olduğunu bulduk: $2n/(n+1)$. Bu eşitliği düzenlersek,

$$\sum_{k=1}^{n} k\binom{2n-k-1}{n-1} = \binom{2n}{n+1}$$

formülünü elde etmiş oluruz. İlginç bir formül.

Bu formülün çok daha basit bir kanıtını Refail Alizade verdi: $2n$ elemanlı $\{0, 1, ..., 2n-1\}$ kümesinin $n+1$ elemanlı altkümelerini $i_0 < ... < i_n$ olmak üzere $A = \{i_0, ..., i_n\}$ olarak yazalım ve bu altkümeye n elemanlı $B = \{i_1, ..., i_n\}$ altkümesine karşılık getirelim. Formülle uyumlu olması için $k = i_1$ olsun. Tam $k = i_1$ tane $n+1$ elemanlı A altkümesi aynı $B = \{i_1, ..., i_n\}$ altkümesine karşılık gelecektir, ne de olsa tam k tane i_0 seçebiliriz. En küçük elemanı $k = i_1$ olan kaç farklı B seçilebilir? $2n - 1 - k$ elemandan $n - 1$ eleman seçmeliyiz ve bu elemanlara k'yı eklemeliyiz. Buradan formül çıkıyor.

Muamma. 4'üncü denemedeki saçmasapan akıl yürütme nasıl oluyor da bize doğru yanıt veriyor? Yanlış bir akıl yürütmenin doğru yanıt verdiği vakidir, ama bu durumda bu kadar şansın olamayacağını sanıyorum. Galiba o denemedeki akıl yürütme doğru da, neden doğru olduğunu göremedim.

Bu yazıda yazılanların hepsi ya da bir kısmı Serdar Boztaş, Gregory Cherlin ve Doğan Bilge tarafından da bulunmuştur. Bazılarıyla ortak çalışılmıştır. Her birine teşekkür ederim. ♥