

## برآورد مقدار $e$ به کمک متغیرهای تصادفی زمان توقف

مهران نقی‌زاده قمی<sup>۱</sup> آزاده کیاپور<sup>۲</sup>

چکیده:

در این مقاله، به کمک متغیرهای تصادفی زمان توقف و انجام شبیه‌سازی، برآوردگرهای ناریب برای عدد  $e$ ، به دست می‌آوریم. **واژه‌های کلیدی:** توزیع یکنواخت، متغیرهای تصادفی زمان توقف، مقدار  $e$ .

### ۱ مقدمه

متغیر تصادفی بررسی شده به وسیله‌ی راسل در حالت کلی‌تر و یک متغیر تصادفی دیگر مورد بررسی قرار می‌گیرد. در بخش ۴ با انجام شبیه‌سازی، مقدار  $e$  برآورد می‌شود. برنامه‌های  $R$  در بخش ۵ آمده است.

اولین مطالعه در مورد عدد  $e$ ، در سال ۱۶۱۸ در جدول ضمیمه کاری در مورد لگاریتم‌ها به وسیله‌ی جان نپر منتشر شد. البته کشف این عدد به یاکوب برنولی منتسب می‌شود که در تلاش برای یافتن بسط  $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$  بود. یکی از مسأله‌های جالب در احتمال، مسأله‌ی جور کردن کلاه‌ها است:  $n$  نفر به مهمانی دعوت می‌شوند. در ورود به اتاق کلاه‌ها شماره‌گذاری شده و به هوا پرت می‌شوند. احتمال اینکه هیچ‌یک از مهمان‌ها کلاه خود را نیابد چقدر است؟ پاسخ این سؤال به صورت

### ۲ متغیرهای تصادفی زمان توقف

متغیرهای تصادفی هندسی و دوجمله‌ای منفی از جمله متغیرهای تصادفی زمان توقف هستند. برای مثال سکه‌ای که احتمال شیر آمدن با آن  $p$  باشد، را  $n$  بار پرتاب می‌کنیم. تعداد آزمایش‌های لازم برای رسیدن به اولین شیر، یک متغیر تصادفی زمان توقف و دارای توزیع هندسی با پارامتر  $p$  است. برای بررسی بیشتر، فرض کنید  $X_1, X_2, \dots \stackrel{iid}{\sim} Bin(1, p)$  باشند. متغیر تصادفی زیر

$$Y = \min \{n \geq 1 : X_n = 1\} \quad (1)$$

را در نظر بگیرید. فرض کنید آزمایش (پرتاب سکه) را آنقدر ادامه دهیم تا در پرتاب ششم شیر بیاید، یعنی در ۵ پرتاب اول خط و در ششمی شیر آمده است و در نتیجه آزمایش متوقف می‌شود و بنابراین  $Y = 6$ . واضح است که متغیر تصادفی  $Y$  دارای توزیع هندسی با پارامتر  $p$  است، زیرا:

$$\begin{aligned} P(Y = n) &= P(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdots P(X_{n-1} = 0)P(X_n = 1) \\ &= p(1-p)^{n-1}, \quad n = 1, 2, \dots \end{aligned}$$

حال متغیر تصادفی

$$Z = \min \left\{ n \geq 5 : \sum_{i=1}^n X_i = 5 \right\}$$

$$p_n = 1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + (-1)^n \frac{1}{n!}$$

است که برای  $n$  بزرگ به مقدار  $e^{-1}$  نزدیک می‌شود (راس (۱۹۹۷)، صفحه‌ی ۱۱۶). همچنین اگر شخصی که احتمال موفقیت او در آزمونی  $\frac{1}{n}$  باشد و به طور مستقل  $n$  بار در آزمونی شرکت کند، به شرط بزرگ بودن  $n$ ، احتمال اینکه در هیچ آزمونی موفق نباشد برابر  $e^{-1}$  است. از کاربردهای دیگر  $e$  می‌توان به فرمول استرلینگ در تقریب مقدار  $n!$  به فرم

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

اشاره کرد.

راسل (۱۹۹۱) در یک مقاله‌ی آموزشی، با معرفی یک متغیر تصادفی زمان توقف با انجام شبیه‌سازی به برآورد مقدار  $e$  پرداخت. مقاله حاضر به این شکل ادامه می‌یابد. در بخش ۲، متغیرهای تصادفی زمان توقف معرفی می‌شوند. در بخش ۳، ویژگی‌های

<sup>۱</sup> گروه آمار، دانشگاه مازندران، بابلسر

<sup>۲</sup> گروه آمار، دانشگاه آزاد، واحد بابل، بابل

$$M = \min\{n \geq 2 : \sum_{i=1}^n U_i > 1\}$$

در نظر بگیرید. نشان می‌دهیم  $E(N) = E(M) = e$ .  
 $N$ ، شاخص اولین متغیر تصادفی یکنواخت است که از مقدار  
 بلافاصله ماقبل خود بزرگتر است. برای توجیه متغیر تصادفی  $N$   
 فرض کنید اعداد تصادفی فرضی زیر از توزیع  $U(0, 1)$  تولید شده  
 باشند

$$u_1 = 0/64, u_2 = 0/38, u_3 = 0/41$$

اولین مقداری که از مقدار بلافاصله ماقبل خود بزرگتر است  $u_3$   
 است. بنابراین آزمایش متوقف می‌شود و  $N = 3$  است.  
 $M$ ، کمترین مقداری از  $n$  است که برای آن  $\sum_{i=1}^n U_i > 1$   
 (راسل (۱۹۹۱)، صفحه ۶۶). دوباره اعداد تصادفی تولیدشده  
 را در نظر بگیرید. ملاحظه می‌کنید که  $u_1 + u_2 = 1/02 > 1$   
 بنابراین  $M = 2$  است.

ابتدا توزیع  $N$  را می‌یابیم. تعداد  $n!$  ترتیب ممکن یکسان از  
 $U_1, \dots, U_n$  وجود دارد. بنابراین داریم:

$$P(N > n) = P(U_1 > \dots > U_n) = \frac{1}{n!}.$$

حال نشان می‌دهیم  $P(M > n) = \frac{1}{n!}$ . برای این کار در حالت  
 کلی‌تر نشان می‌دهیم

$$P(M(x) > n) = \frac{x^n}{n!}, \quad n \geq 1 \quad (2)$$

که در آن  $U_1, U_2, \dots \stackrel{iid}{\sim} U(0, 1)$  فرض کنید. فرض کنید متغیر تصادفی  
 باشند. می‌توان نشان داد که متغیر تصادفی زیر

$$N = \min\{n \geq 1 : \prod_{i=1}^n U_i < e^{-\lambda}\} - 1$$

دارای توزیع پواسون با پارامتر  $\lambda$  است (روهاتگی و صالح (۲۰۰۱)،  
 صفحه ۲۱۷).

### ۳ تقریب مقدار $e$

فرض کنید  $U_1, U_2, \dots$  دنباله‌ای از متغیرهای تصادفی مستقل از  
 توزیع  $U(0, 1)$  باشند. متغیرهای تصادفی زمان توقف  $N$  و  $M$  را  
 به صورت

$$N = \min\{n \geq 2 : U_n > U_{n-1}\},$$

$$M(x) = \min\left\{n \geq 1 : \sum_{i=1}^n U_i > x\right\}.$$

رابطه‌ی (۲) برای  $n = 1$  درست است، زیرا

$$P(M(x) > 1) = P(U_1 \leq x) = x.$$

فرض می‌کنیم برای هر  $0 < x \leq 1$ ، فرض استقراء یعنی  
 $P(M(x) > n) = \frac{x^n}{n!}$  برقرار باشد. در این صورت داریم

$$\begin{aligned} P(M(x) > n+1) &= \int_0^1 P(M(x) > n+1 | U_1 = y) dy \\ &= \int_0^x P(M(x) > n+1 | U_1 = y) dy \\ &= \int_0^x P(M(x-y) > n) dy \\ &= \int_0^x \frac{(x-y)^n}{n!} dy \\ &= \int_0^x \frac{u^n}{n!} du = \frac{x^{n+1}}{(n+1)!}, \end{aligned}$$

و در نتیجه

در نتیجه استقراء کامل و رابطه‌ی (۲) اثبات می‌شود. با قرار دادن

$x = 1$  در رابطه‌ی (۲) داریم

$$P(M > n) = \frac{1}{n!}$$

$$\begin{aligned} \sigma_N &= \sqrt{E(N^2) - E^2(N)} \\ &= \sqrt{3e - e^2} \end{aligned} \quad (۶)$$

بنابراین متغیرهای تصادفی  $M$  و  $N$  هم‌توزیع هستند. حال به محاسبه امید ریاضی متغیرهای تصادفی  $N$  و  $M$  می‌پردازیم. داریم

$$E(M) = E(N) = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty} \frac{1}{n!} = e,$$

فرض کنید  $M_1, M_2, \dots, M_n$  و  $N_1, N_2, \dots, N_n$  نمونه‌هایی تصادفی به ترتیب از توزیع‌های  $M$  و  $N$  باشند. در این صورت طبق قانون قوی اعداد بزرگ (کاسلا و برگر، صفحه‌ی ۲۳۵) برای  $n$  بزرگ داریم:

و بنابراین:

$$E(M) = E(N) = e. \quad (۳)$$

$$E(N) \approx \bar{N}, \quad E(M) \approx \bar{M}$$

تابع جرم احتمال را به صورت زیر

که در آن

$$\begin{aligned} P(N = n) &= P(N > n - 1) - P(N > n) \\ &= \frac{1}{(n - 1)!} - \frac{1}{n!} = \frac{n - 1}{n!}, \quad n > 1 \end{aligned}$$

$$\bar{N} = \frac{1}{n} \sum_{i=1}^n N_i, \quad \bar{M} = \frac{1}{n} \sum_{i=1}^n M_i$$

به دست می‌آوریم. برای یافتن انحراف معیار، از تابع مولد احتمال بهره می‌گیریم. تابع مولد احتمال متغیر تصادفی  $N$  برابر مقدار زیر است:

با توجه به رابطه‌های (۳) و (۶) برآوردهای  $e$  به صورت

$$\hat{e}_N = \bar{N}, \quad \hat{e}_M = \bar{M},$$

$$\begin{aligned} \phi_N(s) &= E(s^N) = \sum_{n=2}^{\infty} s^n P(N = n) \\ &= \sum_{n=2}^{\infty} s^n \frac{n - 1}{n!} \\ &= \sum_{n=2}^{\infty} s^n \frac{1}{(n - 1)!} - \sum_{n=2}^{\infty} s^n \frac{1}{n!} \\ &= s \sum_{n=1}^{\infty} \frac{s^n}{n!} - \sum_{n=2}^{\infty} \frac{s^n}{n!} \\ &= s[e^s - 1] - [e^s - s - 1] \\ &= 1 + e^s(s - 1) \end{aligned}$$

و برآوردهای  $\sigma$  به صورت

$$\hat{\sigma}_N = \sqrt{\frac{1}{n - 1} \sum_{i=1}^n (N_i - \bar{N})^2}, \quad \hat{\sigma}_M = \sqrt{\frac{1}{n - 1} \sum_{i=1}^n (M_i - \bar{M})^2}$$

هستند.

## ۴ شبیه‌سازی

در این بخش، با انجام شبیه‌سازی از توزیع  $N$  و  $M$ ، به برآورد مقدار  $e$  می‌پردازیم. مقدار  $e$  در نرم‌افزار  $R$  با عبارت  $\exp(1)$  برابر  $2/718282$  به دست می‌آید. با توجه به رابطه‌ی (۶) مقدار  $\hat{e} \pm \frac{\hat{\sigma}}{\sqrt{n}}$  به دست می‌آید. برآوردهای  $\hat{e}$  و بازه‌های  $\hat{e} \pm \frac{\hat{\sigma}}{\sqrt{n}}$  برای تکرارهای مختلف در جدول (۱) آمده است. نتایج نشان می‌دهد که با استفاده از متغیرهای تصادفی زمان توقف  $N$  و  $M$  به خوبی می‌توان مقدار  $e$  را برآورد نمود.

با توجه به ویژگی‌های تابع مولد احتمال داریم

$$\phi_N''(1) = E(N^2) - E(N). \quad (۴)$$

مشتق دوم تابع مولد احتمال عبارت از

$$\phi_N''(s) = e^s + se^s \quad (۵)$$

است. با استفاده از رابطه‌های (۳)، (۴) و (۵) داریم

$$E(N^2) = \phi_N''(1) + E(N) = 3e.$$

## ۵ برنامه‌های R

```

### approximation of e based on M ###
M<-c()
for(j in 1:10000){
i<-0 ;s<-0
repeat{
u<-runif(1) ; s<-s+u
if (s>1) break
i<-i+1
}
M[j]<-i+1
mean(M)
sd(M)/10000

### approximation of e based on N ###
N<-c()
for(j in 1:10000){
i<-1 ; u<-runif(1)
repeat{
u2<-runif(1)
if (u2>u) break
i<-i+1 ; u<-u2
}
N[j]<-i+1
}
mean(N)
sd(N)/10000

```

جدول ۱. میانگین و انحراف استاندارد  $e$ 

$\hat{e}_M \pm \frac{\hat{\sigma}_M}{\sqrt{n}}$	$\hat{e}_N \pm \frac{\hat{\sigma}_N}{\sqrt{n}}$	$n$
$2/716 \pm 0/88 \times 10^{-3}$	$2/7231 \pm 0/83 \times 10^{-3}$	۱۰۰۰
$2/7006 \pm 0/86 \times 10^{-4}$	$2/7275 \pm 0/89 \times 10^{-4}$	۱۰۰۰۰
$2/7194 \pm 0/174 \times 10^{-4}$	$2/7197 \pm 0/175 \times 10^{-4}$	۵۰۰۰۰
$2/7184 \pm 0/8735 \times 10^{-5}$	$2/7275 \pm 0/8734 \times 10^{-5}$	۱۰۰۰۰۰

## مراجع

- [1] Cassela, G. and Berger, R. (2002), *Statistical Inference*, 2nd Edn., Thomson Learning.
- [2] Rohatgi, V. and Saleh A. (2001), *An Introduction to Probability and Statistics*, 2nd Edn, John Wiley and Sons.
- [3] Ross, S. M. (1997), *Introduction to Probability Models*, Academic.
- [4] Russell, K. G. (1991), *Estimating Value of e by Simulation*, the American Statistician, 45(1), 66-68.