محاسبات آماری پیشرفته

ترم اول سال تحصیلی ۹۳

جلسه هشتم: روش‌های مونت کارلو (ارزیابی، کنترل و سرعت دادن به همگرایی)

حسین باغیشنی

دانشگاه شاهرود

۲۳ آبان ۱۳۹۳

کاهش واریانس

برای دو برآوردگر $\hat{\theta}_۱$ و $\hat{\theta}_۲$، اگر $Var(\hat{\theta}_۱) < Var(\hat{\theta}_۲)$ آنگاه درصد کاهش واریانس توسط $\hat{\theta}_۱$ عبارتست از:

$$۱۰۰\left(\frac{Var(\hat{\theta}_۲) - Var(\hat{\theta}_۱)}{Var(\hat{\theta}_۱)}\right).$$

فرض کنید

$$X^{(j)} = (X_۱^{(j)}, \ldots, X_n^{(j)}),\ j = ۱, \ldots, m,$$

یک نمونه تصادفی ساده از توزیع X باشد. بنابراین

$$Y^{(j)} = h(X_۱^{(j)}, \ldots, X_n^{(j)}),\ j = ۱, \ldots, m,$$

یک نمونه تصادفی ساده از توزیع $Y = h(X_۱, \ldots, X_n)$ است و فرض کنید

$$\mathbb{E}(\bar{Y}) = \mathbb{E}\left(\frac{۱}{m}\sum_{j=۱}^{m} Y_j\right) = \theta.$$

بنابراین برآوردگر مونت کارلو، یعنی $\hat{\theta} = \bar{Y}$، نااریب است و واریانس آن برابر

$$Var(\hat{\theta}) = \frac{Var_f(h(X))}{m}$$

است.

از فرمول واریانس واضح است که هر چقدر حجم نمونه مونت کارلو، m، افزایش یابد واریانس برآوردگر کاهش می‌یابد.

اما برای کاهش میزان کوچکی در انحراف معیار، باید حجم نمونه بزرگی تولید شود. مثلا برای کاهش از ۰/۰۱ به ۰/۰۰۰۱ باید تقریباً ۱۰۰۰۰ تکرار بیشتر تولید شود.

بنابراین اگرچه با افزایش حجم نمونه مونت کارلویی واریانس برآوردگر کاهش می‌یابد، اما هزینه محاسبات نیز زیاد خواهد بود. پس باید بین میزان کاهش واریانس و کارایی روش (سرعت به دست آوردن برآوردگر) تعادلی ایجاد کرد.

روش‌های مختلفی برای کاهش واریانس معرفی شده‌اند در حالی که هزینه محاسباتی زیادی هم در بر ندارند که در این جلسه به معرفی برخی از آن‌ها می‌پردازیم.

متغیرهای ناهم‌سو

$$Var(\frac{U_۱ + U_۲}{۲}) = \frac{۱}{۴}\{Var(U_۱) + Var(U_۲)\} + \frac{۱}{۲}Cov(U_۱, U_۲).$$

واریانس میانگین کمتر از واریانس $U_۱$ و $U_۲$ خواهد بود هرگاه این دو متغیر دارای همبستگی منفی باشند. یا به عبارت دیگر ناهم‌سو باشند.

همین واقعیت ساده، روشی جهت کاهش واریانس معرفی می‌کند.

فرض کنید $(X_۱, \ldots, X_m)$ و $(Y_۱, \ldots, Y_m)$ دو نمونه از f برای برآورد

$$\mathcal{J} = \int_{\Re} h(t)f(t)dt,$$

به وسیله

$$\hat{\mathcal{J}}_۱ = \frac{۱}{m}\sum_{j=۱}^{m} h(X_j) \ \& \ \hat{\mathcal{J}}_۲ = \frac{۱}{m}\sum_{j=۱}^{m} h(Y_j),$$

باشند، به طوری که میانگین آن‌ها $\mathcal{J}$ و واریانسشان $\sigma^۲$ باشد.

کاهش واریانس

$$Var(\frac{\hat{\mathcal{J}}_۱ + \hat{\mathcal{J}}_۲}{۲}) = \frac{\sigma^۲}{۲} + \frac{۱}{۲} Cov(\hat{\mathcal{J}}_۱, \hat{\mathcal{J}}_۲).$$

اگر دو نمونه همبسته منفی باشند، آنگاه

$$Cov(\hat{\mathcal{J}}_۱, \hat{\mathcal{J}}_۲) \leq ۰$$

و در نتیجه دو نمونه مستقل و هم‌حجم برآوردگر را بهبود می‌دهند.

فرض کنید $X_1, \ldots, X_n$ با روش تبدیل انتگرال احتمال تولید شده باشند. یعنی به ازای هر تکرار m،

1. U_j از توزیع $U(0,1)$ تولید شده است
2. $X^{(j)} = F_X^{-1}(U_j),\ j = 1, \ldots, n$

می‌دانیم U و $1 - U$ هم‌توزیع هستند و دارای همبستگی منفی می‌باشند. بنابرین

$$Y_j = h(F_X^{-1}(U_1^{(j)}), \ldots, F_X^{-1}(U_n^{(j)})),$$

و

$$Y_j' = h(F_X^{-1}(1 - U_1^{(j)}), \ldots, F_X^{-1}(1 - U_n^{(j)})),$$

هم‌توزیع هستند.

اگر تابع h یکنوا باشد، متغیرهای Y_j و Y_j' همبسته منفی هستند

اثبات آن در صفحه ۱۲۹ کتاب قرار دارد

دستور اجرای روش چگونه است؟

اگر m نمونه مونت کارلو لازم باشد، $m/۲$ آن‌ها را برابر

$$Y_j = h(F_X^{-۱}(U_۱^{(j)}), \ldots, F_X^{-۱}(U_n^{(j)})),$$

و بقیه $m/۲$ را برابر

$$Y_j' = h(F_X^{-۱}(۱ - U_۱^{(j)}), \ldots, F_X^{-۱}(۱ - U_n^{(j)})),$$

قرار می‌دهیم، به طوری که $U_i^{(j)},\ i = ۱, \ldots, n; j = ۱, \ldots, m/۲,$ نمونه‌های تصادفی از $U(۰, ۱)$ هستند.

در نتیجه برآوردگر حاصل از روش متغیرهای ناهم‌سو عبارت است از:

$$\hat{\theta} = \frac{۱}{m/۲} \sum_{j=۱}^{m/۲} \left(\frac{Y_j + Y_j'}{۲} \right).$$

با این روش به جای تولید nm نمونه، نمونه‌ای به حجم $nm/۲$ تولید می‌شود و برآوردگری با واریانس کمتر خواهیم داشت.

مثال

$$\Phi(x) = \int_{-\infty}^{x} \frac{۱}{\sqrt{۲\pi}} e^{-t^{۲}/۲} dt.$$

با تبدیل متغیر، محاسبه این انتگرال تبدیل می‌شود به محاسبه $\theta = \mathbb{E}_U \left[x e^{-(xU)^{۲}/۲} \right]$ که در آن $U \sim U(۰, ۱)$.

با در نظر گرفتن مقادیر مثبت x، تابع h یکنوا خواهد بود و در نتیجه با تولید $u_۱, \ldots, u_{m/۲} \sim U(۰, ۱)$ داریم:

$$Y_j = h^{(j)}(u) = x e^{-(x u_j)^{۲}/۲}, \; j = ۱, \ldots, m/۲.$$

بقیه $m/۲$ نمونه را به صورت زیر تولید می‌کنیم:

$$Y_j' = x e^{-(x(۱-u_j))^{۲}/۲}, \; j = ۱, \ldots, m/۲.$$

بنابراین

$$\hat{\theta} = \frac{۱}{m/۲} \sum_{j=۱}^{m/۲} \frac{xe^{-(xu_j)^۲/۲} + xe^{-(x(۱-u_j))^۲/۲}}{۲} \longrightarrow \theta$$

```
MC.Phi <- function(x, R = 10000, antithetic = TRUE) {
        u <- runif(R/2)
        if (!antithetic) v <- runif(R/2) else
            v <- 1 - u
        u <- c(u, v)
        cdf <- numeric(length(x))
        for (i in 1:length(x)) {
            g <- x[i] * exp(-(u * x[i])^2 / 2)
            cdf[i] <- mean(g) / sqrt(2 * pi) + 0.5
        }
        cdf
}
```

```
x <- seq(.1, 2.5, length=5)
Phi <- pnorm(x)
set.seed(123)
MC1 <- MC.Phi(x, anti = FALSE)
set.seed(123)
MC2 <- MC.Phi(x)
> print(round(rbind(x, MC1, MC2, Phi), 5))
       [,1]    [,2]    [,3]    [,4]    [,5]
x   0.10000 0.70000 1.30000 1.90000 2.50000
MC1 0.53983 0.75825 0.90418 0.97311 0.99594
MC2 0.53983 0.75805 0.90325 0.97132 0.99370
Phi 0.53983 0.75804 0.90320 0.97128 0.99379
```

کاهش تقریبی واریانس را می‌توان برای یک مقدار معلوم x توسط روش متغیرهای ناهم‌سو در مقابل روش مونت کارلوی استاندارد، محاسبه کرد:

```
m <- 1000
MC1 <- MC2 <- numeric(m)
x <- 1.95
for (i in 1:m) {
     MC1[i] <- MC.Phi(x, R = 1000, anti = FALSE)
     MC2[i] <- MC.Phi(x, R = 1000)
}
> print(sd(MC1))
[1] 0.006874616
> print(sd(MC2))
[1] 0.0004392972
> print((var(MC1) - var(MC2))/var(MC1))
[1] 0.9959166
```

مشاهده می‌کنید که با روش متغیرهای ناهم‌سو، برآوردگر جدید ٪۹۹/۵ کاهش واریانس در نقطه $x = ۱/۹۵$ ایجاد کرده است.

چند نکته

- این روش همیشه قابل اجرا نیست ولی اگر امکان اجرای آن باشد، روشی کارا محسوب می‌شود
- اگر f متقارن باشد، قرار دهید $Y_j = ۲\mu - X_j$ که در آن μ میانگین f می‌باشد.
- اگر $(A_i)_i$ تکیه‌گاه متغیر X را افراز کند، با نمونه‌گیری لایه‌ای، یعنی نمونه‌گیری X_jها در هر A_i، می‌توان واریانس را کاهش داد

متغیرهای کنترلی

رهیافتی دیگر برای کاهش واریانس برآوردگرهای مونت کارلویی $\mathbb{E}(h(X))$، استفاده از متغیرهای کنترلی است.

فرض کنید

$$\mathcal{J} = \int h(x)f(x)\,dx,$$

نامعلوم و

$$\mathcal{J}_{\circ} = \int h_{\circ}(x)f(x)\,dx,$$

معلوم باشد. $\mathcal{J}$ با $\hat{\mathcal{J}}$ برآورد می‌شود و $\mathcal{J}_{\circ}$ با $\hat{\mathcal{J}}_{\circ}$.

متغیرهای کنترلی: ادامه

با ترکیب دو برآوردگر داریم:

$$\hat{\mathcal{J}}^* = \hat{\mathcal{J}} + \beta(\hat{\mathcal{J}}_۰ - \mathcal{J}_۰)$$

در این حالت $\hat{\mathcal{J}}^*$ برای $\mathcal{J}$ نااریب است و

$$Var(\hat{\mathcal{J}}^*) = Var(\hat{\mathcal{J}}) + \beta^۲ Var(\hat{\mathcal{J}}_۰) + ۲\beta Cov(\hat{\mathcal{J}}, \hat{\mathcal{J}}_۰).$$

کنترل بهینه

انتخاب بهینه برای β عبارتست از:

$$\beta^* = -\frac{Cov(\hat{\mathcal{J}}, \hat{\mathcal{J}}_۰)}{Var(\hat{\mathcal{J}}_۰)}$$

همچنین

$$Var(\hat{\mathcal{J}}^*) = (۱ - \rho^۲)\, Var(\hat{\mathcal{J}}),$$

که در آن ρ ضریب همبستگی بین $\hat{\mathcal{J}}$ و $\hat{\mathcal{J}}_۰$ است. بنابراین میزان کاهش واریانس برابر است با

$$۱۰۰\rho^۲$$

راه حل معمول برای به دست آوردن β^*، استفاده از رگرسیون است. ضریب رگرسیون $h(x_i)$ روی $h_۰(x_i)$ همان β^* بهینه را به دست خواهد داد.

هر چه میزان همبستگی بین $\hat{\mathcal{J}}$ و $\hat{\mathcal{J}}_۰$ بیشتر باشد، میزان کاهش واریانس بیشتر خواهد بود.

مثال

$$\theta = \int_{۰}^{۱} \frac{e^{-x}}{۱ + x^{۲}} dx,$$

به طوری که $X \sim U(۰, ۱)$.
اگر قرار دهیم $h_{۰}(x) = \frac{e^{-۰/۵}}{۱+x^{۲}}$، آنگاه در فاصله $(۰, ۱)$ به $h(x)$ نزدیک است، یعنی همبستگی بالایی دارند، و

$$\mathbb{E}(h_{۰}(x)) = e^{-۰/۵} \int_{۰}^{۱} \frac{۱}{۱ + t^{۲}} dt = e^{-۰/۵} \arctan(۱) = e^{-۰/۵} \frac{\pi}{۴}.$$

```
f <- function(u)
     exp(-.5)/(1+u^2)
g <- function(u)
     exp(-u)/(1+u^2)
set.seed(510) #needed later
u <- runif(10000)
B <- f(u)
A <- g(u)
## beta^*
cor(A, B)
a = -cov(A,B) / var(B)    #est of beta*
> a
[1] -2.436228
## Control variate estimate
m <- 100000
u <- runif(m)
T1 <- g(u)
T2 <- T1 + a * (f(u) - exp(-.5)*pi/4)
> c(mean(T1), mean(T2))
[1] 0.5253543 0.5250021
> c(var(T1), var(T2))
[1] 0.060231423 0.003124814
# Percent of reduced variance
> (var(T1) - var(T2)) / var(T1)
[1] 0.9481199
```

محاسبات با استفاده از رگرسیون

```
set.seed(510)
u <- runif(10000)
f <- exp(-.5)/(1+u^2)
g <- exp(-u)/(1+u^2)
c.star <-  - lm(g ~ f)$coeff[2]   # beta[1]
mu <- exp(-.5)*pi/4
> c.star
        f
-2.436228
u <- runif(10000)
f <- exp(-.5)/(1+u^2)
g <- exp(-u)/(1+u^2)
L <- lm(g ~ f)
theta.hat <- sum(L$coeff * c(1, mu))  #pred. value at mu
> theta.hat
[1] 0.5253113
> summary(L)$sigma^2
[1] 0.003117644
> summary(L)$r.squared
[1] 0.9484514
```

چند نکته

- می‌توان به جای یک متغیر از چند متغیر کنترلی استفاده کرد. در این صورت از رگرسیون چندگانه برای برآورد ضرایب بهینه و میزان کاهش واریانس می‌توان بهره برد. همچنین برآورد جدید مقدار پیشگوی حاصل از تابع رگرسیون در نقاط میانگین متغیرهای کنترلی است.
- میزان کاهش واریانس، همان ضریب تعیین رگرسیون است.
- واریانس برآوردگر جدید با روش متغیرهای کنترلی، در مدل رگرسیونی همان MSE/n است
- بین روش‌های متغیرهای ناهم‌سو و کنترلی رابطه مستقیم وجود دارد