محاسبات آماری پیشرفته

ترم اول سال تحصیلی ۹۳

جلسه هفتم: روش‌های مونت کارلو (انتگرال‌گیری مونت کارلویی)

حسین باغیشنی

دانشگاه شاهرود

۱۶ آبان ۱۳۹۳

مقدمه

دو رده اصلی از مسایل استنباط‌های آماری عبارتند از:

- بهینه‌سازی: که معمولا با رهیافت‌های مبتنی بر درستنمایی همراه است
- انتگرال‌گیری: که معمولا با رهیافت بیزی همراه است

مقدمه

دو رده اصلی از مسایل استنباط‌های آماری عبارتند از:

- بهینه‌سازی: که معمولا با رهیافت‌های مبتنی بر درستنمایی همراه است
- انتگرال‌گیری: که معمولا با رهیافت بیزی همراه است

در استنباط‌های آماری، اغلب مواقع راه حل مساله منتهی به حل عددی یک انتگرال یا یک بهینه‌سازی می‌شود.

مثال: نظریه تصمیم بیزی

برآوردگرهای بیزی همیشه امید ریاضی توزیع پسین نیستند. اما جواب انتگرال زیر هستند:

$$\min_{\delta} \int_{\Theta} L(\theta, \delta)\pi(\theta)f(x|\theta)d\theta.$$

- زیان توان دوم خطا: برای $L(\theta, \delta) = (\theta - \delta)^{۲}$، برآوردگر بیزی میانگین توزیع پسین است
- زیان قدر مطلق خطا: برای $L(\theta, \delta) = |\theta - \delta|$، برآوردگر بیزی میانه توزیع پسین است
- بدون تابع زیان: از MAP استفاده می‌کنیم که عبارتست از $\arg\max_{\theta} \ell(\theta|x)\pi(\theta)$.

روش‌های انتگرال‌گیری عددی

انتگرالی مانند $\mathcal{J} = \int_{\chi} h(x)f(x)\,dx$ را می‌توان به کمک روش‌های عددی مانند قاعده سیمپسون یا ذوزنقه، محاسبه کرد.

روش‌های انتگرال‌گیری عددی

انتگرالی مانند $\mathcal{J} = \int_{\chi} h(x)f(x)\,dx$ را می‌توان به کمک روش‌های عددی مانند قاعده سیمپسون یا ذوزنقه، محاسبه کرد.

به عنوان مثال، R برای محاسبه انتگرال‌های یک بعدی دو تابع $area$ و $integrate$ را دارد. اما تابع $area$ برای انتگرال‌های با بعد بی‌کران قابل استفاده نیست.

روش‌های انتگرال‌گیری عددی

انتگرالی مانند $\mathcal{J} = \int_{\chi} h(x)f(x)\,dx$ را می‌توان به کمک روش‌های عددی مانند قاعده سیمپسون یا ذوزنقه، محاسبه کرد.

به عنوان مثال، R برای محاسبه انتگرال‌های یک بعدی دو تابع $area$ و $integrate$ را دارد. اما تابع $area$ برای انتگرال‌های با بعد بی‌کران قابل استفاده نیست.

تابع $integrate$ برای انتگرال‌های با بعد بی‌کران قابل استفاده است، اما جواب‌های قابل اعتمادی نتیجه نمی‌دهد.

روش‌های انتگرال‌گیری عددی

انتگرالی مانند $\mathcal{J} = \int_{\chi} h(x)f(x)\,dx$ را می‌توان به کمک روش‌های عددی مانند قاعده سیمپسون یا ذوزنقه، محاسبه کرد.

به عنوان مثال، R برای محاسبه انتگرال‌های یک بعدی دو تابع $area$ و $integrate$ را دارد. اما تابع $area$ برای انتگرال‌های با بعد بی‌کران قابل استفاده نیست.

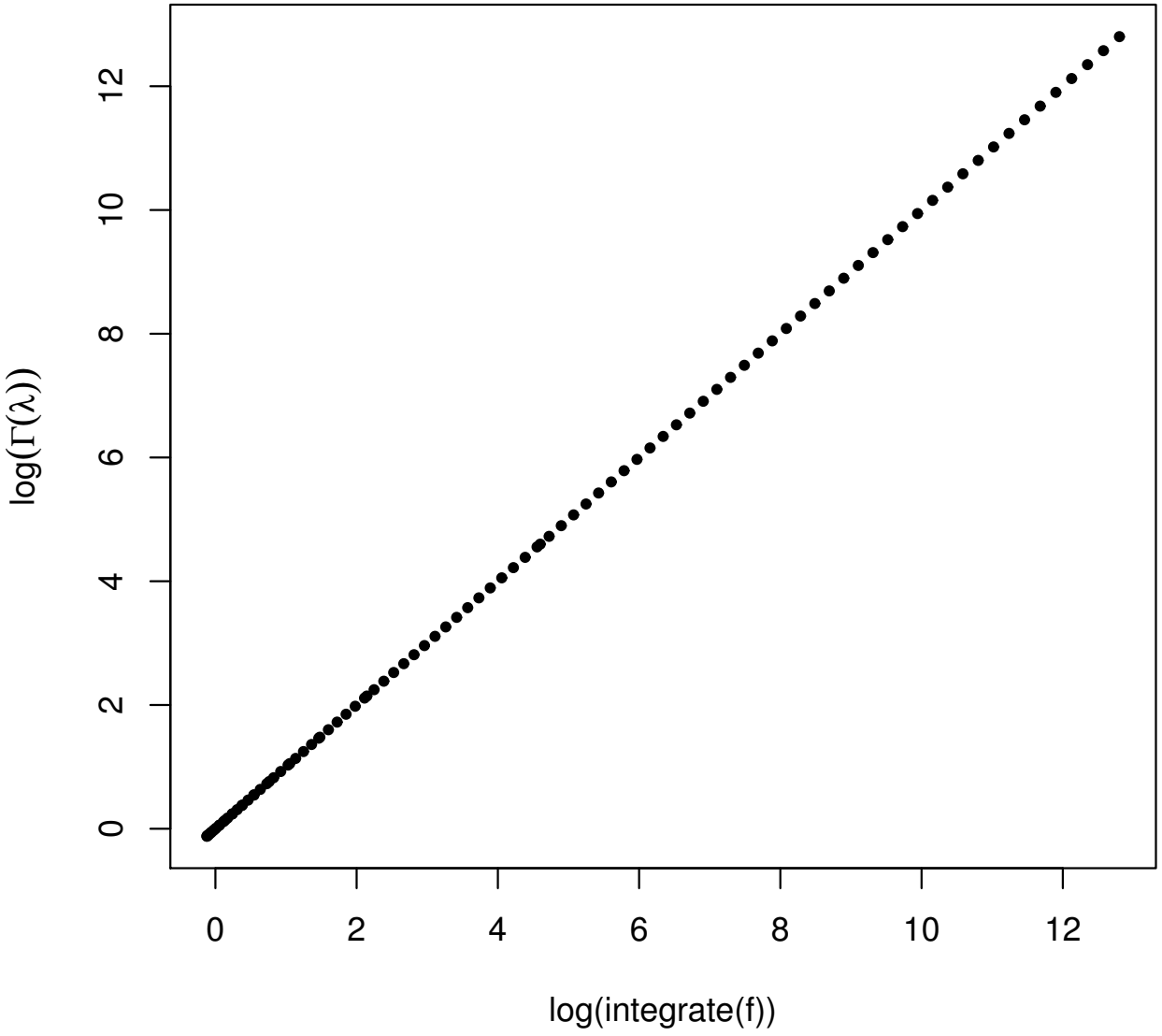
تابع $integrate$ برای انتگرال‌های با بعد بی‌کران قابل استفاده است، اما جواب‌های قابل اعتمادی نتیجه نمی‌دهد.

به طور کلی، روش‌های عددی انتگرال‌گیری برای محاسبه انتگرال‌های با بعد بالا (که در روش‌های آمار نیز خیلی معمول هستند)، مناسب نیستند.

مثال

به عنوان مثال، محاسبه انتگرال $\int_0^\infty x^{\lambda-1} e^{-x} dx$ را با کمک تابع $integrate$ در نظر بگیرید و با تابع $gamma$ مقایسه کنید.

```
ch=function(la){
   integrate(function(x){x^(la-1)*exp(-x)},0,Inf)$val}
plot(lgamma(seq(.01,10,le=100)),log(apply(as.matrix(
seq(.01,10,le=100)),1,ch)),xlab="log(integrate(f))",
ylab=expression(log(Gamma(lambda))),pch=19,cex=.6)
```


log(Γ(λ))
0
2
4
6
8
10
12
0
2
4
6
8
10
12
log(integrate(f))

مشکل انتگرال‌گیری‌های عددی

مشکل اساسی با روش‌های انتگرال‌گیری عددی (مانند روش موجود در تابع $integrate$) آن است که در اکثر موارد، این روش‌ها قادر به شناسایی نواحی مهم (نواحی چگال‌تر) برای تابعی که باید انتگرال‌گیری شود، نیستند. اما روش‌های شبیه‌سازی با هدف قرار دادن این نواحی، به کمک تابع چگالی f، چنین مشکلی را ندارند.

پروژه: روش‌های معمول انتگرال‌گیری عددی مانند سیمپسون، ذوزنقه و غیره

انتگرال‌گیری مونت کارلویی

مساله:
مساله کلی، محاسبه انتگرال

$$\mathcal{J} = \mathbb{E}_f\left[h(X)\right] = \int_{\chi} h(x)f(x)\,dx,$$

است که در آن χ یک یا چندبعدی است. f یک تابع چگالی، دارای فرم بسته یا به طور جزیی بسته است. و h یک تابع می‌باشد.

انتگرال‌گیری مونت کارلویی: ادامه

جواب انتگرال به روش مونت کارلو:
از نمونه $(X_1, \ldots, X_m)$ حاصل از توزیع f برای تقریب انتگرال $\mathcal{J}$ به وسیله میانگین نمونه استفاده می‌کنیم:

$$\bar{h}_m = \frac{1}{m} \sum_{j=1}^{m} h(x_j),$$

به طوری که بنا بر قانون قوی اعداد بزرگ، به مقدار انتگرال میل می‌کند

$$\bar{h}_m \longrightarrow \mathbb{E}_f(h(X)).$$

دقت برآورد مونت کارلویی

واریانس برآوردگر را با کمیت

$$\nu_m = \frac{۱}{m(m-۱)} \sum_{j=۱}^{m} [h(x_j) - \bar{h}_m]^۲,$$

برآورد می‌کنیم و برای m بزرگ:

$$\frac{\bar{h}_m - \mathbb{E}_f(h(X))}{\sqrt{\nu_m}} \sim N(۰, ۱).$$

توجه: این نتیجه می‌تواند منتهی به آزمونی برای همگرایی برآورد و ساختن کران‌های اطمینان برای تقریب انتگرال شود

مثال

یک برآورد مونت کارلویی برای کمیت $\theta = \int_0^1 e^{-x} dx$ به دست آورید و با مقدار دقیق آن مقایسه کنید.

مثال

یک برآورد مونت کارلویی برای کمیت $\theta = \int_0^1 e^{-x} dx$ به دست آورید و با مقدار دقیق آن مقایسه کنید.

اگر در این مثال، f را یکنواخت استاندارد در نظر بگیریم، می‌توان با نوشتن کد زیر برآوردی برای θ به دست آورد.

```
m <- 10000
x <- runif(m)
theta.hat <- mean(exp(-x))
> print(theta.hat)
[1] 0.6327615
> print(1 - exp(-1))
[1] 0.6321206
```

کران‌های متفاوت

اگر انتگرال به شکل $\int_a^b h(x)dx$ باشد، می‌توان با تغییر متغیر حدود را به ۰ تا ۱ تغییر داد. تبدیل لازم به صورت زیر است:

$$\begin{aligned} y &= \frac{x-a}{b-a}, \\ dy &= (\frac{1}{b-a})dx, \end{aligned}$$

بنابراین:

$$\int_a^b h(x)dx = \int_۰^۱ h(y(b-a)+a)(b-a)dy.$$

البته به جای توزیع $U(۰,۱)$ برای f، می‌توان از توزیع $U(a,b)$ نیز استفاده کرد. در این صورت:

$$\int_a^b h(x)dx = (b-a)\int_a^b h(x)\frac{۱}{b-a}dx.$$

مثال

یک برآورد مونت کارلویی برای کمیت $\theta = \int_{۲}^{۴} e^{-x} dx$ به دست آورید و با مقدار دقیق آن مقایسه کنید.

```
m <- 10000
x <- runif(m, min=2, max=4)
theta.hat <- mean(exp(-x)) * 2
> print(theta.hat)
[1] 0.1180278
> print(exp(-2) - exp(-4))
[1] 0.1170196
```

انتگرال‌های بی‌کران

از روش مونت کارلو برای محاسبه تابع توزیع نرمال استاندارد استفاده کنید:

$$\Phi(z) = \int_{-\infty}^{z} \frac{۱}{\sqrt{۲\pi}} e^{-\frac{x^۲}{۲}}\, dx.$$

می‌توان با تغییر متغیر آن‌ها را به انتگرال‌های کراندار تبدیل کرد (تمرین: چگونه؟).

انتگرال‌های بی‌کران

از روش مونت کارلو برای محاسبه تابع توزیع نرمال استاندارد استفاده کنید:

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

می‌توان با تغییر متغیر آن‌ها را به انتگرال‌های کراندار تبدیل کرد (تمرین: چگونه؟).

نکته: با استفاده از تابع نشانگر، احتمال‌ها را نیز می‌توان به شکل امید ریاضی بیان کرد.

انتگرال‌های بی‌کران

از روش مونت کارلو برای محاسبه تابع توزیع نرمال استاندارد استفاده کنید:

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

می‌توان با تغییر متغیر آن‌ها را به انتگرال‌های کراندار تبدیل کرد (تمرین: چگونه؟).

نکته: با استفاده از تابع نشانگر، احتمال‌ها را نیز می‌توان به شکل امید ریاضی بیان کرد.

فرض کنید $I(\cdot)$ تابع نشانگر باشد و $X \sim N(۰, ۱)$. بنابراین برای هر ثابت z، داریم

$$\mathbb{E}[I(X \leq z)] = \mathbb{P}(X \leq z) = \Phi(z).$$

پس با در نظر گرفتن توزیع نرمال استاندارد برای f، داریم:

$$\hat{\Phi}(z) = \frac{۱}{m} \sum_{j=۱}^{m} I(x_i \leq z).$$

```
z <- seq(.1, 2.5, length = 10)
m <- 10000
x <- rnorm(m)
dim(z) <- length(z)
p <- apply(z, MARGIN = 1,
     FUN = function(z, x) {mean(x < z)}, x = x)
Phi <- pnorm(z)
> print(round(rbind(z, p, Phi), 3))
     [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
z   0.100 0.367 0.633 0.900 1.167 1.433 1.700 1.967 2.233 2.500
p   0.525 0.627 0.718 0.803 0.871 0.919 0.952 0.973 0.988 0.994
Phi 0.540 0.643 0.737 0.816 0.878 0.924 0.955 0.975 0.987 0.994
```

کران‌های خطا برای انتگرال‌گیری مونت کارلویی

برای مثال بالا، واریانس برآوردگر را محاسبه و کران‌های اطمینان ٪۹۵ را برای برآورد $\Phi(۲)$ به دست می‌آوریم.

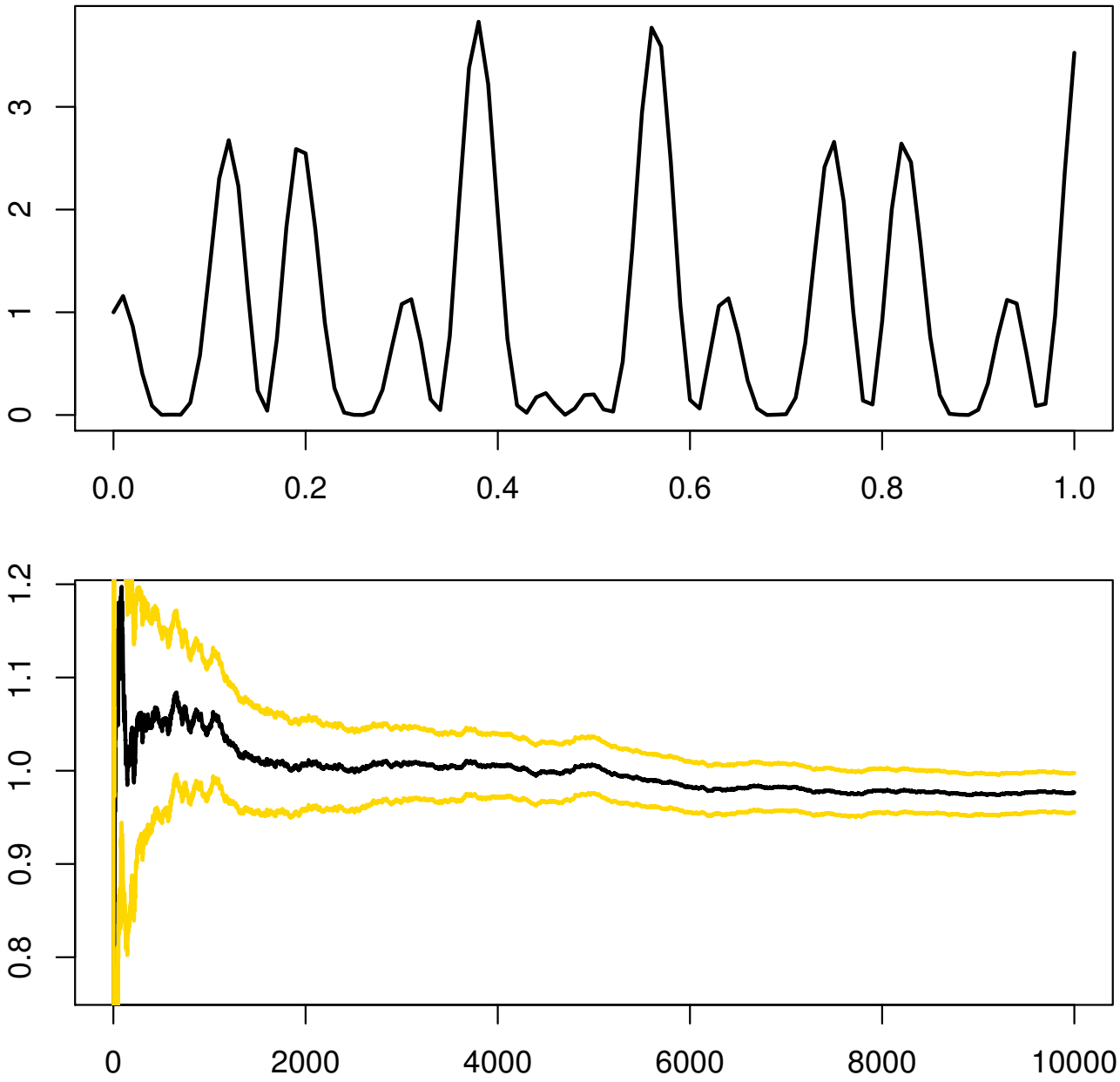
```
z <- 2
m <- 10000
x <- rnorm(m)
g <- (x < z)  #the indicator function
cdf <- mean(g)
v <- mean((g - mean(g))^2) / m
> c(cdf, v)
[1] 9.769000e-01 2.256639e-06
> c(cdf - 1.96 * sqrt(v), cdf + 1.96 * sqrt(v))
[1] 0.9739557 0.9798443
```

دقت کنید در اینجا تابع $h(x) = I(x \leq z)$ یک متغیر برنولی است و بنابراین واریانس آن برابر $\Phi(۲)(۱ - \Phi(۲))/m$ است که برای این کد برابر مقدار $۲/۲۲۳e - ۰۶$ است.

مثال

برای تابع $h(x) = [\cos ۵۰x + \sin ۲۰x]^۲$، مقدار انتگرال آن را در بازه $[۰, ۱]$ محاسبه کنید.

```
h=function(x){(cos(50*x)+sin(20*x))^2}
par(mar=c(2,2,2,1),mfrow=c(2,1))
curve(h,xlab="Function",ylab="",lwd=2)
> integrate(h,0,1)
0.9652009 with absolute error < 1.9e-10
x=h(runif(10^4))
estint=cumsum(x)/(1:10^4)
esterr=sqrt(cumsum((x-estint)^2))/(1:10^4)
plot(estint, xlab="Mean and error range",type="l",lwd=2,
ylim=mean(x)+20*c(-esterr[10^4],esterr[10^4]),ylab="")
lines(estint+2*esterr,col="gold",lwd=2)
lines(estint-2*esterr,col="gold",lwd=2)
```


0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
0.8
0.9
1.0
1.1
1.2
0
2000
4000
6000
8000
10000

چند نکته:

باید دقت داشت که برآورد مونت کارلویی مادامی که ν_m برآوردی مناسب برای واریانس $\bar{h}_m$ است، قابل اتکا است.

چند نکته:

باید دقت داشت که برآورد مونت کارلویی مادامی که ν_m برآوردی مناسب برای واریانس $\bar{h}_m$ است، قابل اتکا است.

در موارد بحرانی که ν_m همگرا نیست یا به کندی همگرا می‌شود، نمی‌توان از قضیه حد مرکزی و توزیع تقریبی نرمال بهره برد و در نتیجه محاسبه نواحی اطمینان، قابل اتکا نیستند.

چند نکته:

باید دقت داشت که برآورد مونت کارلویی مادامی که ν_m برآوردی مناسب برای واریانس $\bar{h}_m$ است، قابل اتکا است.

در موارد بحرانی که ν_m همگرا نیست یا به کندی همگرا می‌شود، نمی‌توان از قضیه حد مرکزی و توزیع تقریبی نرمال بهره برد و در نتیجه محاسبه نواحی اطمینان، قابل اتکا نیستند.

در مثال محاسبه تابع توزیع نرمال استاندارد در یک نقطه،

$$\hat{\Phi}(z) = \frac{1}{m}\sum_{j=1}^{m} I(x_j \leq z) \longrightarrow \Phi(z),$$

واریانس دقیق این برآوردگر برابر $\Phi(z)(1-\Phi(z))/m$ است که تقریبا برابر $Var \approx \frac{1}{4m}$ می‌باشد. در این حالت برای داشتن برآوردی با دقت چهار رقم اعشار، $2 \times \sqrt{\frac{1}{4m}} \leq 10^{-4}$، نیاز به تکرار مونت کارلوی $m = (10^4)^2 = 10^8$ داریم.

مثال بیزی: توزیع نرمال–پیشین کوشی

برای برآورد میانگین توزیع نرمال، یک توزیع پیشین تنومند، توزیع کوشی است:

$$X \sim N(\theta, ۱), \quad \theta \sim C(۰, ۱).$$

تحت تابع زیان توان دوم خطا، میانگین پسین عبارتست از:

$$\delta^{\pi} = \frac{\int_{-\infty}^{\infty} \frac{\theta}{۱+\theta^{۲}} e^{-(x-\theta)^{۲}/۲} d\theta}{\int_{-\infty}^{\infty} \frac{۱}{۱+\theta^{۲}} e^{-(x-\theta)^{۲}/۲} d\theta}.$$

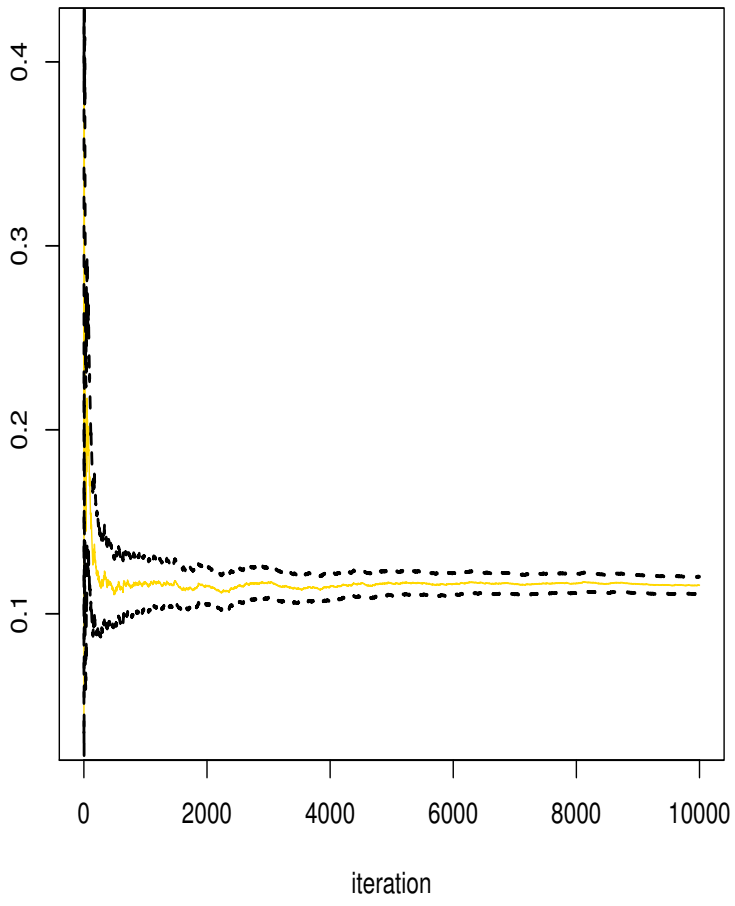
شکل δ^{π}، شبیه‌سازی از توزیع $N(x, ۱)$ را پیشنهاد می‌کند:

$$\theta_{۱}, \ldots, \theta_{m} \sim N(x, ۱),$$

و بنابراین

$$\hat{\delta}_{m}^{\pi} = \sum_{i=۱}^{m} \frac{\theta_i}{۱ + \theta_i^{۲}} / \sum_{i=۱}^{m} \frac{۱}{۱ + \theta_i^{۲}} \longrightarrow \delta^{\pi}.$$

```
set.seed(21351)
x=2
Niter=10^4
co=rnorm(Niter,mean=x)
x1=co/(1+co^2)
x2=1/(1+co^2)
>  mean(x1)/mean(x2)
[1] 1.2729
th=rcauchy(Niter)
d1=th*dnorm(th,mean=x)
d2=dnorm(th,mean=x)
> mean(d1)/mean(d2)
[1] 1.27617
estint1=cumsum(d1)/(1:Niter)
esterr1=sqrt(cumsum((d1-estint1)^2))/(1:Niter)
plot(estint1,type="l",xlab="iteration",ylab="",
col="gold")
lines(estint1-2*esterr1,lty=2,lwd=2)
lines(estint1+2*esterr1,lty=2,lwd=2)
```



نمونه‌گیری نقاط مهم

این روش را به این دلیل نمونه‌گیری نقاط مهم گویند که تمرکز نمونه‌گیری بر نواحی چگال‌تر (و در نتیجه انتخاب نقاط مهم‌تر) قرار دارد.

نمونه‌گیری نقاط مهم

این روش را به این دلیل نمونه‌گیری نقاط مهم گویند که تمرکز نمونه‌گیری بر نواحی چگال‌تر (و در نتیجه انتخاب نقاط مهم‌تر) قرار دارد.

انتگرال $\mathcal{J}$ را به خاطر آورید.

تضاد:
تولید مستقیم نمونه از تابع چگالی f لزوماً بهینه نیست. (چرا؟)

نمونه‌گیری نقاط مهم

این روش را به این دلیل نمونه‌گیری نقاط مهم گویند که تمرکز نمونه‌گیری بر نواحی چگال‌تر (و در نتیجه انتخاب نقاط مهم‌تر) قرار دارد.

انتگرال $\mathcal{J}$ را به خاطر آورید.

تضاد:
تولید مستقیم نمونه از تابع چگالی f لزوماً بهینه نیست. (چرا؟)

یک روش جانشین، استفاده از نمونه‌گیری از نقاط مهم ($Importance\ Sampling$) است که بر اساس نمایش جانشینی برای $\mathcal{J}$ عمل می‌کند:

$$\mathbb{E}_f(h(X)) = \int_{\chi} \left[h(x) \frac{f(x)}{g(x)} \right] g(x)\, dx = \mathbb{E}_g \left[h(X) \frac{f(X)}{g(X)} \right].$$

نمونه‌گیری نقاط مهم

این روش را به این دلیل نمونه‌گیری نقاط مهم گویند که تمرکز نمونه‌گیری بر نواحی چگال‌تر (و در نتیجه انتخاب نقاط مهم‌تر) قرار دارد.

انتگرال $\mathcal{J}$ را به خاطر آورید.

تضاد:
تولید مستقیم نمونه از تابع چگالی f لزوماً بهینه نیست. (چرا؟)

یک روش جانشین، استفاده از نمونه‌گیری از نقاط مهم ($Importance\ Sampling$) است که بر اساس نمایش جانشینی برای $\mathcal{J}$ عمل می‌کند:

$$\mathbb{E}_f(h(X)) = \int_{\chi} \left[h(x) \frac{f(x)}{g(x)} \right] g(x)\, dx = \mathbb{E}_g \left[h(X) \frac{f(X)}{g(X)} \right].$$

این صورت نمایش به ما اجازه استفاده از توزیع‌های دیگری به جز f را می‌دهد.

نمونه‌گیری نقاط مهم

این روش را به این دلیل نمونه‌گیری نقاط مهم گویند که تمرکز نمونه‌گیری بر نواحی چگال‌تر (و در نتیجه انتخاب نقاط مهم‌تر) قرار دارد.

انتگرال $\mathcal{J}$ را به خاطر آورید.

تضاد:
تولید مستقیم نمونه از تابع چگالی f لزوماً بهینه نیست. (چرا؟)

یک روش جانشین، استفاده از نمونه‌گیری از نقاط مهم (*Importance Sampling*) است که بر اساس نمایش جانشینی برای $\mathcal{J}$ عمل می‌کند:

$$\mathbb{E}_f(h(X)) = \int_{\chi} \left[h(x) \frac{f(x)}{g(x)} \right] g(x)\, dx = \mathbb{E}_g \left[h(X) \frac{f(X)}{g(X)} \right].$$

این صورت نمایش به ما اجازه استفاده از توزیع‌های دیگری به جز f را می‌دهد.

به f توزیع اصلی و به g توزیع ابزاری (*Instrumental Distribution*) نیز می‌گویند.

الگوریتم نمونه‌گیری نقاط مهم

برای محاسبه انتگرال $\mathcal{J}$ می‌توان به صورت زیر عمل کرد:

1. نمونه $X_1, \ldots, X_m$ را از توزیع g تولید کن

الگوریتم نمونه‌گیری نقاط مهم

برای محاسبه انتگرال $\mathcal{J}$ می‌توان به صورت زیر عمل کرد:

1. نمونه $X_1, \ldots, X_m$ را از توزیع g تولید کن
2. از تقریب زیر استفاده کن:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j).$$

استفاده از سایر توزیع‌ها به جای f می‌تواند باعث کاهش واریانس برآوردگرهای حاصل شود.

مشابه قبل

همگرایی:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j) \longrightarrow \int_{\chi} h(x)f(x)\,dx.$$

مشابه قبل

همگرایی:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j) \longrightarrow \int_{\chi} h(x)f(x)\,dx.$$

این همگرایی به ازای هر توزیع g برقرار است، مشروط بر آن که $supp(f) \subset supp(g)$.

مشابه قبل

همگرایی:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j) \longrightarrow \int_{\chi} h(x)f(x)\,dx.$$

این همگرایی به ازای هر توزیع g برقرار است، مشروط بر آن که $supp(f) \subset supp(g)$.

دو نکته مهم:

- توزیع g باید به گونه‌ای انتخاب شود که تولید نمونه از آن ساده باشد

مشابه قبل

همگرایی:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j) \longrightarrow \int_{\chi} h(x)f(x)\,dx.$$

این همگرایی به ازای هر توزیع g برقرار است، مشروط بر آن که $supp(f) \subset supp(g)$.

دو نکته مهم:

- توزیع g باید به گونه‌ای انتخاب شود که تولید نمونه از آن ساده باشد
- از نمونه تولیدشده (توسط g) می‌توان به دفعات استفاده کرد. نه تنها برای توابع h مختلف بلکه برای توزیع‌های f متفاوت هم نمونه مشابه قابل استفاده است

انتخاب توزیع ابزاری

اگرچه g هر تابع چگالی می‌تواند باشد، بعضی از انتخاب‌ها بهتر از بقیه هستند.

- برآوردگر فقط وقتی دارای واریانس متناهی است که

$$\mathbb{E}_f\left[h^{۲}(X)\frac{f(X)}{g(X)}\right] = \int_{\chi} h^{۲}(x)\frac{f(x)}{g(x)}\,dx < \infty.$$

انتخاب توزیع ابزاری

اگرچه g هر تابع چگالی می‌تواند باشد، بعضی از انتخاب‌ها بهتر از بقیه هستند.

- برآوردگر فقط وقتی دارای واریانس متناهی است که

$$\mathbb{E}_f\left[h^{۲}(X)\frac{f(X)}{g(X)}\right] = \int_{\chi} h^{۲}(x)\frac{f(x)}{g(x)}\,dx < \infty.$$

- توزیع‌های ابزاری با دم‌های سبک‌تر از دم‌های f (یعنی زمانی که $\sup \frac{f}{g} = \infty$) مناسب نیستند.

انتخاب توزیع ابزاری

اگرچه g هر تابع چگالی می‌تواند باشد، بعضی از انتخاب‌ها بهتر از بقیه هستند.

- برآوردگر فقط وقتی دارای واریانس متناهی است که

$$\mathbb{E}_f\left[h^2(X)\frac{f(X)}{g(X)}\right] = \int_\chi h^2(x)\frac{f(x)}{g(x)}dx < \infty.$$

- توزیع‌های ابزاری با دم‌های سبک‌تر از دم‌های f (یعنی زمانی که $\sup \frac{f}{g} = \infty$) مناسب نیستند.
- اگر $\sup \frac{f}{g} = \infty$، وزن‌های $\frac{f(x_j)}{g(x_j)}$ به شدت متغیر خواهند بود و این به آن معنی است که به تعداد اندکی از x_jها اهمیت بیش از اندازه داده می‌شود

انتخاب توزیع ابزاری

اگرچه g هر تابع چگالی می‌تواند باشد، بعضی از انتخاب‌ها بهتر از بقیه هستند.

- برآوردگر فقط وقتی دارای واریانس متناهی است که

$$\mathbb{E}_f\left[h^2(X)\frac{f(X)}{g(X)}\right] = \int_\chi h^2(x)\frac{f(x)}{g(x)}dx < \infty.$$

- توزیع‌های ابزاری با دم‌های سبک‌تر از دم‌های f (یعنی زمانی که $\sup \frac{f}{g} = \infty$) مناسب نیستند.
- اگر $\sup \frac{f}{g} = \infty$، وزن‌های $\frac{f(x_j)}{g(x_j)}$ به شدت متغیر خواهند بود و این به آن معنی است که به تعداد اندکی از x_jها اهمیت بیش از اندازه داده می‌شود
- اگر $\sup \frac{f}{g} = M < \infty$، برای تولید نمونه از f به طور مستقیم، از روش پذیرش ـ رد هم می‌توان استفاده کرد

مثال: توزیع کوشی

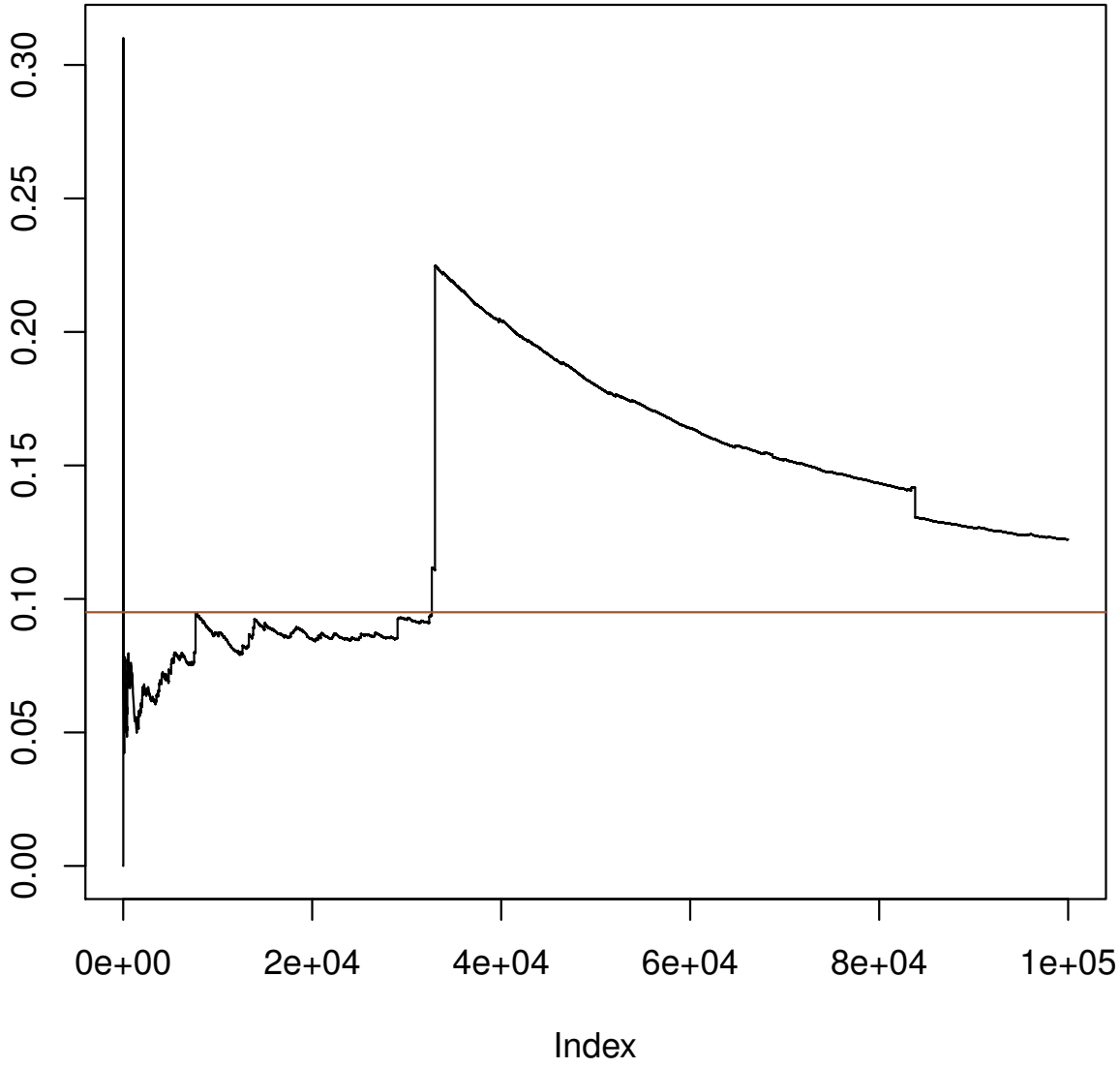
اگر توزیع کوشی استاندارد را به عنوان تابع اصلی و نرمال استاندارد را به عنوان تابع ابزاری در نظر بگیریم، داریم

$$\rho(x) = \frac{f(x)}{g(x)} = \sqrt{۲\pi}\frac{\exp x^{۲}/۲}{\pi(۱ + x^{۲})}.$$

این نسبت به شدت بدرفتار است، یعنی

$$\int_{-\infty}^{\infty} \rho^{۲}(x)g(x)dx = \infty \quad \text{(تمرین)}$$

و در نتیجه عملکرد ضعیفی برای برآوردگر حاصل از نمونه‌گیری نقاط مهم می‌توان متصور بود.


0.00
0.05
0.10
0.15
0.20
0.25
0.30
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
Index

تابع ابزاری بهینه

تابع g که واریانس برآوردگر نمونه‌گیری نقاط مهم را می‌نیمم می‌کند، عبارتست از

$$g^*(x) = \frac{|h(x)|f(x)}{\int |h(z)|f(z)dz}.$$

دستیابی به این انتخاب بهینه میسر نیست.

تابع ابزاری بهینه

تابع g که واریانس برآوردگر نمونه‌گیری نقاط مهم را می‌نیمم می‌کند، عبارتست از

$$g^*(x) = \frac{|h(x)|f(x)}{\int |h(z)|f(z)dz}.$$

دستیابی به این انتخاب بهینه میسر نیست.

زیرا اطلاع از آن نیازمند دانستن مقدار انتگرالی است که به دنبال محاسبه آن هستیم.

نسخه عملی برآوردگر نمونه‌گیری نقاط مهم

یک روش جانشین برای برآوردگر معرفی‌شده به عنوان برآوردگر نمونه‌گیری نقاط مهم (در چند اسلاید قبل)، استفاده از

$$\frac{\sum_{j=1}^{m} h(x_j)\frac{f(x_j)}{g(x_j)}}{\sum_{j=1}^{m} \frac{f(x_j)}{g(x_j)}},$$

است که هم مشکل متناهی بودن واریانس برآوردگر را مرتفع می‌کند و هم به طور کلی برآوردگر پایدارتری را نتیجه می‌دهد.

در این نسخه، به جای m از مجموع وزن‌ها، $w_j = \frac{f(x_j)}{g(x_j)}$، استفاده می‌شود.

دقت کنید که چون $\frac{1}{m}\sum_{j=1}^{m} \frac{f(x_j)}{g(x_j)} \longrightarrow 1$، زمانی که $m \longrightarrow \infty$، این برآوردگر هم بنابر قانون قوی اعداد بزرگ به $\mathbb{E}_f(h(X))$ میل می‌کند.

با این که این برآوردگر اریب است، اما اریبی کوچک آن و واریانس کمتر، آن را به برآوردگر ارجح نسبت به نسخه قبلی تبدیل کرده است.

میانگین هارمونیک

با در نظر گرفتن تابع بهینه، برای این نسخه از برآوردگر نمونه‌گیری نقاط مهم داریم:

$$\frac{\sum_{j=1}^{m} h(x_j)\frac{f(x_j)}{g(x_j)}}{\sum_{j=1}^{m} \frac{f(x_j)}{g(x_j)}} = \frac{\sum_{j=1}^{m} h(x_j)|h(x_j)|^{-1}}{\sum_{j=1}^{m} |h(x_j)|^{-1}},$$

به طوری که $x_j \sim g \propto |h|f$.

دقت کنید که صورت عبارت بالا عبارتست از تعداد دفعاتی که $h(x_j)$ مثبت است منهای تعداد دفعاتی که منفی است.

اگر h مثبت باشد، این برآوردگر به میانگین هارمونیک تبدیل می‌شود.

علی رغم آنچه که تصور می‌شود بهینگی مطرح‌شده، برای این برآوردگر برقرار نیست و ثابت شده است که می‌تواند به برآوردگری اریب و به شدت ناپایدار منجر شود.

در متون مختلفی در مورد غیرقابل اعتماد بودن برآوردگر میانگین هارمونیک بحث شده است.

رهنمود کلی

پروژه: نمایش عملکرد برآوردگر میانگین هارمونیک

از نقطه نظر کاربردی، توزیعی برای g بهینه است که $|h|f/g$ تقریبا ثابت باشد و واریانس متناهی داشته باشد

مثال

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

در واقع در این مثال،

$$g(x) = \begin{cases} \frac{e^{-x}}{۱+x^{۲}} & ۰ < x < ۱ \\ ۰ & o.w. \end{cases}$$

برای حل آن از توابع ابزاری زیر استفاده می‌کنیم:

$$\begin{aligned}
f_0(x) &= 1, \quad 0 < x < 1, \\
f_1(x) &= e^{-x}, \quad 0 < x < \infty, \\
f_2(x) &= \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty, \\
f_3(x) &= \frac{e^{-x}}{1-e^{-1}}, \quad 0 < x < 1, \\
f_4(x) &= \frac{4}{\pi(1+x^2)}, \quad 0 < x < 1.
\end{aligned}$$

دو تابع f_1 و f_2 دارای تکیه‌گاه بزرگتری هستند و بسیاری از مقادیر تولید شده از این توزیع‌ها، در مجموع سهمی برابر صفر خواهند داشت. و این یعنی ناکارایی.

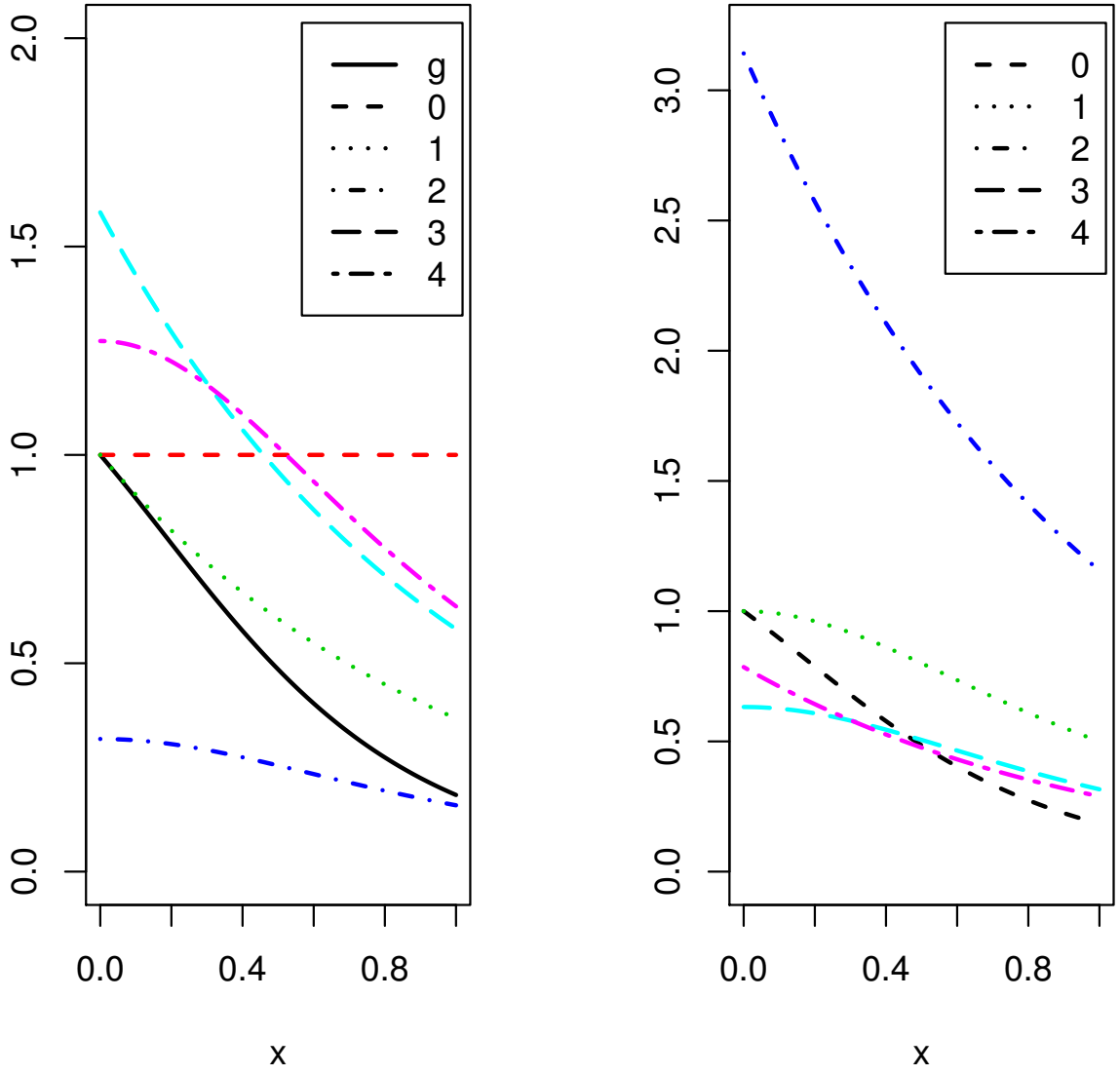
از تمام این توزیع‌ها به راحتی می‌توان نمونه تولید کرد.

f_2 توزیع کوشی استاندارد است

تابعی که برای آن نسبت g/f، دقت کنید در این مثال $h = 1$ است، از بقیه به مقداری ثابت نزدیکتر است، تابع f_3 است. از روی شکل هم دیده می‌شود.

```
x <- seq(0, 1, .01)
w <- 2
f1 <- exp(-x)
f2 <- (1 / pi) / (1 + x^2)
f3 <- exp(-x) / (1 - exp(-1))
f4 <- 4 / ((1 + x^2) * pi)
g <- exp(-x) / (1 + x^2)
```

```
par(mfrow=c(1,2))
## figure (a)
plot(x, g, type = "l", main = "", ylab = "",
     ylim = c(0,2), lwd = w)
lines(x, g/g, lty = 2, col = 2, lwd = w)
lines(x, f1, lty = 3, col = 3, lwd = w)
lines(x, f2, lty = 4, col = 4, lwd = w)
lines(x, f3, lty = 5, col = 5, lwd = w)
lines(x, f4, lty = 6, col = 6, lwd = w)
legend("topright", legend = c("g", 0:4),
      lty = 1:6, lwd = w, inset = 0.02)
# figure (b)
plot(x, g, type = "l", main = "", ylab = "",
     ylim = c(0,3.2), lwd = w, lty = 2)
lines(x, g/f1, lty = 3, col = 3, lwd = w)
lines(x, g/f2, lty = 4, col = 4, lwd = w)
lines(x, g/f3, lty = 5, col = 5, lwd = w)
lines(x, g/f4, lty = 6, col = 6, lwd = w)
legend("topright", legend = c(0:4),
       lty = 2:6, lwd = w, inset = 0.02)
```


g
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
0.0
0.4
0.8
x
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.4
0.8
x

```
m <- 10000
theta.hat <- se <- numeric(5)
g <- function(x) {
     exp(-x - log(1+x^2)) * (x > 0) * (x < 1)
}
x <- runif(m)     #using f0
fg <- g(x)
theta.hat[1] <- mean(fg)
se[1] <- sd(fg)
x <- rexp(m, 1)   #using f1
fg <- g(x) / exp(-x)
theta.hat[2] <- mean(fg)
se[2] <- sd(fg)
x <- rcauchy(m)   #using f2
i <- c(which(x > 1), which(x < 0))
x[i] <- 2  #to catch overflow errors in g(x)
fg <- g(x) / dcauchy(x)
theta.hat[3] <- mean(fg)
se[3] <- sd(fg)
u <- runif(m)     #f3, inverse transform method
x <- - log(1 - u * (1 - exp(-1)))
fg <- g(x) / (exp(-x) / (1 - exp(-1)))
theta.hat[4] <- mean(fg)
se[4] <- sd(fg)
u <- runif(m)    #f4, inverse transform method
x <- tan(pi * u / 4)
fg <- g(x) / (4 / ((1 + x^2) * pi))
theta.hat[5] <- mean(fg)
se[5] <- sd(fg)
### Results
> rbind(theta.hat, se)
               [,1]      [,2]      [,3]       [,4]      [,5]
theta.hat 0.5268425 0.5206398 0.5236278 0.52512375 0.5257661
se        0.2467741 0.4191347 0.9496426 0.09542621 0.1416929
```

مثال

استفاده از روش معمول مونت کارلو برای تقریب احتمال‌های دمی مانند $P(X > a)$ وقتی که a بزرگ باشد، می‌تواند بسیار پرهزینه و نادقیق باشد.

به عنوان مثال اگر فرض کنید $Z \sim N(۰, ۱)$، و مایل باشیم $P(Z > ۴/۵)$ را محاسبه کنیم، که مقدار آن خیلی کوچک است،

```
> pnorm(-4.5)
[1] 3.397673e-06
```

تولید نمونه‌ای از $Z^{(i)} \sim N(۰, ۱)$ به طوری که بزرگتر از ۴/۵ قرار گیرد، هر سه میلیون بار یک بار اتفاق می‌افتد!!!

از آنجا که ما علاقه‌مند به محاسبه احتمال یک پیشامد خیلی نادر هستیم، استفاده از روش معمول برای تولید نمونه از f و رسیدن به یک جواب مناسب (با دقت مناسب)، نیازمند تولید بسیار زیادی نمونه است.

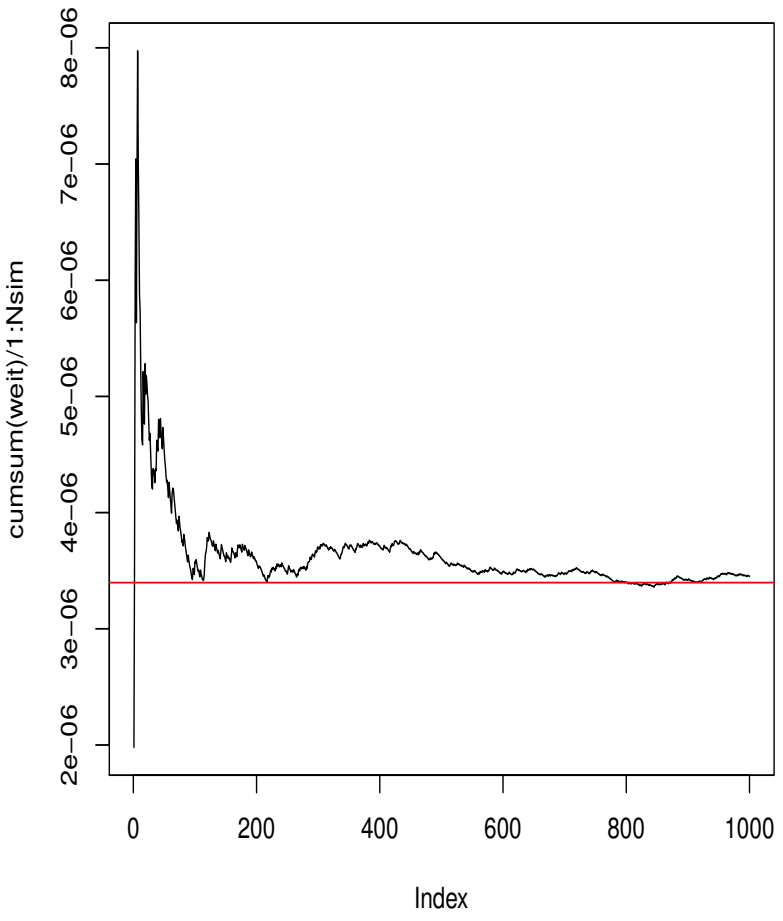
اما نمونه‌گیری نقاط مهم به طور چشمگیری دقت و کارایی محاسبه چنین احتمالی را افزایش می‌دهد.

اگر توزیعی با تکیه‌گاه محدود به $(۴/۵, \infty)$ در نظر بگیریم، قسمتی از واریانس اضافی و غیرضروری برآوردگر مونت کارلو که ناشی از تولید نمونه‌های با وزن صفر است (وقتی که $x < ۴/۵$)، حذف خواهد شد.
یک انتخاب مناسب برای g یعنی تابع ابزاری، نمایی بریده شده در ۴/۵ است:

$$g(y) = \frac{e^{-y}}{\int_{۴/۵}^{\infty} e^{-x}\,dx} = e^{-(y-۴/۵)}.$$

بنابراین برآوردگر IS عبارتست از:

$$\frac{۱}{m}\sum_{j=۱}^{m}\frac{f(y_j)}{g(y_j)} = \frac{۱}{m}\sum_{j=۱}^{m}\frac{e^{-y_j^۲/۲+y_j-۴/۵}}{\sqrt{۲\pi}},$$

که در آن Y_j ها از g تولید شده‌اند.

```
Nsim=1000
y=rexp(Nsim)+4.5
weit=dnorm(y)/dexp(y-4.5)
plot(cumsum(weit)/1:Nsim,type="l")
abline(a=pnorm(-4.5),b=0,col="red")
## Estimated Value = 3.452e-06 vs.
## True value = 3.398e-06
```