«پـروژه درس اسـتنباط یـک»

# روش مــونـت کارلــو با زنجـیره‌ی مارکـف
# $MCMC$


اسـتاد : دکتر بـهاالدین خالـدی

محــی الـدین ایـزدی  عبداله حسـنی جـلیلیان


بهمن ۱۳۸۴

# فصل ۱

# روش مونت کارلو

## ۱.۱ مقدمه

**تعریف**

روش مونت کارلو راه حل تقریبی مسائل ریاضی ، آمار ، فیزیک و... است که برای شبیه سازی کمیت های تصادفی برای تقریب (به عبارت بهتر برآورد) کمیتی مجهول به کار می رود. در واقع هر روشی که سعی بر حل مسئله با تولید اعداد تصادفی دارد را روش مونت کارلو گویند.

**تاریخچه پیدایش روش مونت کارلو**

این روش در سال 1949 با انتشار مقاله‌ای تحت عنوان روش مونت کارلو از دو ریاضیدان آمریکایی به نام های J.NEYMAN و S.ULAM شناخته شد . پایه های نظری این روش از مدت ها قبل مشخص شده بود و در قرن نوزدهم و اوایل قرن بیستم گاهی اوقات مسائل آماری را با کمک شبیه سازی کمیت های تصادفی که همان روش مونت کارلو است حل می کردند.اما به طور قطع تا پیش از اختراع رایا نه های الکترونیکی از این روش به دلیل خسته کننده بودن شبیه سازی کمیت های تصادفی زیاد استفاده نمی شد . بنابراین می توان شروع استفاده از روش مونت کارلو به عنوان یک تکنیک عددی عمومی را همزمان با اختراع رایانه های الکترونیکی دانست . نام مونت کارلو از شهر مونت کارلو ، در موناکو که به خاطر قمارخانه‌هایش شهرت دارد گرفته شده است . مسائلی که می توان آن ها را به روش مونت کارلو حل کرد :

روش مونت کارلو ، شبیه سازی هر روندی که توسط عامل های تصادفی ایجاد شده اند را ممکن می سازد. اما این تنها مورد استفاده روش مونت کارلو نیست. برای بیشتر مسائل ریاضی که در آن ها شانس دخالت ندارد ما می توانیم به طور مصنوعی مدل احتمال آن را با تعداد زیادی نمونه تعیین کنیم ، به این دلیل روش مونت کارلو را یک روش عمومی حل مسائل ریاضی می توان دانست .

## ۲.۱ محاسبه انتگرال ها با روش مونت کارلو

فرض کنیم $g$ یک تابع حقیقی مقدار باشد که علاقمند به محاسبه انتگرال $g$ روی $[a,b]$ هستیم .

$$\theta = \int_a^b g(x)dx < \infty$$

برای محاسبه این انتگرال می توان به چند روش عمل کرد:

۱)روش تحلیلی : یافتن پادمشتق $g$ و با استفاده از قضیه اساسی حساب دیفرانسیل و انتگرال ،

$$\theta = \int_a^b g(x)dx = G(b) - G(a)$$

۲)روش های عددی : اگر قادر نباشیم مسئله ها را به روش تحلیلی حل کنیم از روش های عددی که برای تقریب مقدار $\theta$ وجود دارند ، استفاده می کنیم . در این روش عمدتا تابع $g$ با توابع ساده تری مانند توابع خطی تکه ای ، چند جمله ای و ...، تقریب زده شده و مقدار انتگرال برای این توابع خوش رفتار محاسبه شده ، به عنوان تقریب $\theta$ ارائه می شود.

۳)روش مونت کارلو : این روش نیز یک روش عددی است که درمواردی که روش های عددی در قسمت بالا با پیچیدگی های زیادی همراه هستند یا خطای این روش زیاد است و یا پیش فرض های روش های عددی برقرار نیستند ، به ویژه وقتی که $g$ یک بعدی نباشد کاربرد دارد.

### ۱.۲.۱ روش (برآورد) مونت کارلو:

هدف محاسبه انتگرال زیر است

$$\theta = \int_a^b g(x)dx$$

برای این منظور توزیع دلخواهی بر $[a,b]$ با چگالی $f(x)$ اختیار می کنیم . علاوه بر متغیر تصادفی $X$ ، که بر فاصله $[a,b]$ با چگالی $f(x)$ تعریف شده است احتیاج به متغیر تصادفی

$$h(x) = \frac{g(x)}{f(x)}$$

نیز نیاز دایم.دراین صورت

$$\theta = E(h(x)) = \int_a^b \frac{g(x)}{f(x)} f(x)dx$$

حال از چگالی $f$ نمونه تصادفی $X_1, X_2, ..., X_n$ را استخراج می کنیم. آنگاه $h(X_1), h(X_2), ..., h(X_n)$ نیز یک نمونه تصادفی بوده و داریم ،

$$E(\frac{1}{n}\sum_{i=1}^{n} h(X_i)) = \theta$$

طبق قانون قوی اعداد بزرگ

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{p} \theta$$

بنابراین $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$ برآورد معقولی برای $\theta$ می باشد . واین برآورد معادل تقریب انتگرال مورد نظر است .با این روش بسیاری ازانتگرال ها را که نمی توان با روش تحلیلی حل کرد تقریب زد .

**مثال ۱ :(برآورد $\pi$)**

فرض کنید $X, Y \sim U(-1, 1)$ آنگاه از احتمال می دانیم که:

$$\theta = p(X^2 + Y^2 \leq 1) = \frac{Area\ of\ c}{Area\ of\ \Omega} = \frac{\pi}{4}$$

با تعریف

$$g(x,y) = I_C(x,y) = 1 \qquad X^2 + Y^2 \leq 1$$

$$= 0 \qquad o.w$$

داریم :

$$\theta = E(g(X,Y)) = \frac{\pi}{4}$$

برای تقریب $\theta$ از روش مونت کارلو به شیوه زیر عمل می کنیم .

گام ۱: از توزیع $U(۰,۱)$ نمونه تصادفی $U_۱^۱,...,U_n^۱,U_۱^۲,...,U_n^۲$ را استخراج کرده و قرار می دهیم $X_i = ۲U_i^۱ - ۱$ و $Y_i = ۲U_i^۲ - ۱$ برای $i = ۱,۲,...,n$ در این صورت $X_i \sim U(-۱,۱)$ و $Y_i \sim U(-۱,۱)$.

گام ۲: $g(X_۱,Y_۱),...,g(X_n,Y_n)$ را محاسبه می کنیم .

گام ۳: $\theta$ را با $\hat{\theta} = \frac{۱}{n}\sum_{i=۱}^{n} g(X_i,Y_i)$ که نسبت تعداد $(X_i,Y_i)$ هایی است که در دایره $C$ قرار گرفته اند به $n$ می باشد ، تقریب می زنیم . و در نهایت اگر بخواهیم $\pi$ را تقریب بزنیم ، داریم

$$\pi \simeq ۴\hat{\theta}$$

اگر علا قمند به محاسبه مساحت یا حجم غیر معمول $D$ که معادله تحلیلی آن را در اختیار نداریم ، باشیم باتعریف $g(\cdot) = I_D(\cdot)$ مشابه فوق به تقریب (برآورد) $D$ می پردازیم .

به این منظور این ناحیه را در یک مستطیل ، مکعب یا حجره محصور کرده و از توزیع یکنواخت بر مستطیل ، مکعب یا حجره ، نمونه $\mathbf{U}_۱,...,\mathbf{U}_n$ را استخراج کرده و همچون مثال قبل از $\hat{\theta} = \frac{۱}{n}\sum_{i=۱}^{n} g(\mathbf{U}_i)$ به عنوان تقریب $D$ استفاده می کنیم .

**مثال ۲ :**

فرض کنیم علاقمند به محاسبه $\theta = \sum_{i=۱}^{N} \frac{a(i)}{N}$ که در آن $N$ بسیار بزرگ و $a(i)$ پیچیده بوده و محاسبه آن ساده نباشد ، باشیم . برای تقریب $\theta$ به روش مونت کارلو به قرار زیر عمل می کنیم .

گام ۱: از توزیع $U(۰,۱)$ ، نمونه تصادفی $U_۱^۱,...,U_n^۱,U_۱^۲,...,U_n^۲$ را استخراج کرده، $X_i = [NU_i] + ۱$ را محاسبه می کنیم ، برای $i = ۱,۲,...,n$ . در این صورت $X_i$ دارای توزیع یکنواخت بر $\{۱,۲,\ldots,N\}$ می باشد .

گام ۲: $a(X_۱),...,a(X_n)$ را محاسبه می کنیم .

گام ۳: $\hat{\theta} = \frac{۱}{n}\sum_{i=۱}^{n} a(i)$ را به عنوان برآورد $\theta$ در نظر می گیریم .

روشن است که روش مونت کارلو را برای توابع چند متغیره نیز می توان بکار برد. درواقع مزیت و برتری این روش به روش های عددی معمولی در حل مسائل با بعد بیشتر از یک نمایان می شود. زیرا روش های عددی معمولی را در حالت یک بعدی می توان به کار برد.

فرض کنیم $g : \Re \longrightarrow \Re$ و علاقمند به محاسبه انتگرال $g$ بر ناحیه $A = [a_۱,b_۱] \times [a_۲,b_۲] \times ... \times [a_d,b_d]$ باشیم .

$$\theta = \int_{a_۱}^{b_۱}\int_{a_۲}^{b_۲}...\int_{a_d}^{b_d} g(x_۱,...,x_d)dx_۱...dx_d$$

برای این منظور چگالی دلخواه $f$ بر $A$ را در نظر می گیریم و تابع $h(\cdot)$ را به صورت

$$h(x_1, ..., x_d) = \frac{g(x_1, ..., x_d)}{f(x_1, ..., x_d)}$$

تعریف می کنیم ، در این صورت

$$\theta = E(h(X_1, ..., X_d)) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_d}^{b_d} \frac{g(x_1, ..., x_d)}{f(x_1, ..., x_d)} f(x_1, ..., x_d) dx_1 ... dx_d$$

حال از چگالی $f(\cdot)$ نمونه تصادفی $\mathbf{x}_1, ..., \mathbf{x}_d$ که درآن $\mathbf{X}_i = (X_{1i}, ..., X_{di})$ را استخراج می کنیم ، آنگاه $h(\mathbf{X}_1), ..., h(\mathbf{X}_n)$ یک نمونه تصادفی بوده و داریم که :

$$\frac{1}{n} \sum_{i=1}^{n} h(\mathbf{X}_i) \xrightarrow{p} \theta$$

بنابراین $\frac{1}{n} \sum_{i=1}^{n} h(\mathbf{X}_i)$ یک برآورد معقول برای $\theta$ یعنی انتگرال مورد نظر می باشد .

## ۳.۱ چگونگی انتخاب چگالی

درانتخاب $f$ بدیهی است که باید $f$ی را انتخاب کرد که به سادگی قابل نمونه گیری باشد. زیرا اساس روش مونت کارلو استخراج نمونه و تقریب $\theta$ با استفاده از این نمونه می باشد. به هر حال برای هر $f$ قابل نمونه گیری داریم

$$\theta = E(\frac{g(X)}{f(X)}) = E(h(X))$$

که در آن

$$h(X) = \frac{g(X)}{f(X)}$$

با این حال واریانس و درنتیجه خطای استاندارد برآوردگر $\theta$ بسته به انتخاب $f$ می تواند کوچک یا بزرگ باشد که مطلوب آنست کمترین مقدار خطای استاندارد برآورد را داشته باشیم . یعنی

$$var(h(X)) = \int_a^b (\frac{g(x)}{f(x)})^2 f(x) dx - \theta^2$$

کمینه شود . می توان نشان داد که وقتی $\frac{|g(x)|}{f(x)}$ ثابت باشد این مقدار کمینه می شود. و

$$f(x) = \frac{| g(x) |}{\int_a^b | g(x) | dx}$$

چگالی مطلوب می باشد .

مشاهده می شود که انتخاب بهترین چگالی در عمل به خاطر محاسبه انتگرال $\int_a^b \mid g(x) \mid dx$ که آسانتر از محاسبه $\int_a^b g(x)dx$ نمی باشد ، غیر ممکن است پس $f$ی را انتخاب می کنیم که به سادگی قابل نمونه گیری بوده و نسبت $\frac{|g(x)|}{f(x)}$ کراندار باشد. این امر باعث می شود که واریانس برآوردگر $\frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i)}{f(X_i)}$ کاهش یابد.

در بسیاری از مسائل چگالی $f$ داده شده است ، به عنوان مثال در استنباط بیزی ، علاقمند به محاسبه انتگرال زیر هستیم

$$\gamma = E(h(\theta) \mid \mathbf{x}) = \int_{\Theta} h(\theta)\pi(\theta \mid \mathbf{x})d\theta$$

و برای تقریب این انتگرال به روش مونت کارلو کافی است که نمونه‌ای از چگالی $\pi(\theta \mid \mathbf{x})$ استخراج کرده و از $\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n} h(\theta_i)$ برای تقریب $\gamma$ استفاده می کنیم. بنابراین از این به بعد فرض می کنیم در پی محاسبه امید ریاضی تابعی از متغیر تصادفی $X$ با توزیع $f$ به روش مونت کارلو هستیم . یعنی

$$\gamma = E(h(X)) = \int_{\chi} h(x)f(x)dx$$

## ۴.۱ استخراج نمونه از چگالی $f$

حال اگر متغیر تصادفی $X$ را با چگالی $f$ در نظر بگیریم برای استخراج نمونه از این چگالی می توان از دو روش زیر استفاده کرد.

**روش تبدیل معکوس تابع توزیع** : می دانیم که برای متغیر تصادفی $X$ داریم :

$$F(X) \sim U(0, 1)$$

حال اگر از توزیع $U(0, 1)$ نمونه تصادفی $U_1, ..., U_n$ را استخراج کنیم آنگاه با تبدیل

$$X_i = F^{-1}(U_i) = \inf\{x : F(x) \geq U_i\} \quad i = 1, ..., n$$

$X_1, \ldots X_n$ یک نمونه تصادفی از چگالی $f$ می باشد .

**مثال ۱:**

فرض کنیم بخواهیم از چگالی $f(x) = \lambda e^{-\lambda x}$ نمونه‌ای تصادفی استخراج کنیم . بدین منظور :

گام ۱: $U_۱ \dots U_n$ را از $U(۰, ۱)$ استخراج می کنیم.
گام ۲: قرار می دهیم

$$X_i = F^{-۱}(U_i) = \frac{-\log(۱ - U_i)}{\lambda}$$

آنگاه $X_۱, \dots, X_n$ یک نمونه تصادفی از چگالی $f$ می باشد .

**مثال ۲:**

برای استخراج نمونه از چگالی

$$f(x) = \frac{\lambda^m e^{-\lambda x} x^{m-۱}}{\Gamma(m)}$$

که در آن $m \in \mathbb{N}$ و $\lambda > ۰$ معلوم باشند . چون فرم بسته‌ای برای تابع توزیع این چگالی موجود نیست لذا از روش معکوس تابع توزیع نمی توان استفاده کرد . اما از روابط بین توزیعی توزیع گاما و توزیع نمایی ، می توان نمونه ای به قرار زیر بدست آورد . اگر $Y_۱, \dots, Y_n$ متغیر های مستقل و هم توزیع با توزیع $E(\lambda)$ باشند آنگاه

$$X = \sum_{i=۱}^{m} Y_i \sim \Gamma(m, \lambda)$$

گام ۱: نمونه تصادفی $U_۱^۱, \dots U_۱^m, U_۲^۱, \dots, U_۲^m, \dots, U_n^۱, \dots U_n^m$ را از $U(۰, ۱)$ استخراج می کنیم .
گام ۲:

$$Y_i = \frac{-\log(۱ - U_j^i)}{\lambda} \qquad i = ۱, \dots, m \qquad j = ۱, \dots, n$$

بدست می آوریم .
گام ۳: قرار می دهیم :

$$X_j^i = \sum_{i=۱}^{m} Y_j^i$$

در این صورت $X_j \sim \Gamma(m, \lambda)$ ، بنابراین $X_۱, \dots, X_n$ یک نمونه تصادفی از چگالی

$$f(x) = \frac{\lambda^m e^{-\lambda x} x^{m-۱}}{\Gamma(m)}$$

می باشد. مشابه این مثال موارد دیگر نیز پیش می آید که مستقما نمی توان از روش معکوس تابع توزیع استفاده کرد اما با استفاده از روابط بین توزیعی می توان این مشکل را

برطرف کرد. مثلا برای توزیع نرمال $N(\mu, \sigma^2)$ قادر به استفاده از روش تبدیل معکوس تابع توزیع نیستیم. اما می دانیم اگر $U_1, U_2$ متغیرهای تصادفی مستقل و هم توزیع از $U(0, 1)$ باشند و قرار دهیم

$$X_1 = \mu + \sigma\sqrt{-2\log U_1}\cos(2\pi U_2)$$

$$X_2 = \mu + \sigma\sqrt{-2\log U_1}\sin(2\pi U_2)$$

آنگاه $X_1, X_2$ متغیرهای تصادفی مستقل وهم توزیع از توزیع $N(\mu, \sigma^2)$ خواهند بود.

### ۱.۴.۱ روش پذیرش — رد

فرض کنیم بخواهیم از چگالی $f$ نمونه ای استخراج کنیم و این امر با روش تبدیل معکوس تابع توزیع و روابط بین توزیعی مقدور نباشد. اگر چگالی $g$ با تکیه گاهی شامل تکیه گاه $f$ چنان موجود باشد که

۱: به سادگی قابل نمونه گیری باشد .

۲: نسبت $\frac{f(x)}{g(x)}$ کراندار باشد یعنی یک $M \geq 1$ چنان وجود داشته باشد به طوری که $\forall x \in \chi \quad \frac{f(x)}{g(x)} \leq M$ آن گاه می توان با روش زیر نمونه مورد نظر را گرفت :

قرار می دهیم $\forall x \quad e(x) = g(x)M$ . بنابراین $\forall x \quad e(x) \geq f(x)$ ، که تابع $e(x)$ را پوششی برای $f$ می نامند. برای نمونه گیری به طریق زیر عمل می کنیم .

گام ۱: $Y$ را از چگالی $g$ استخراج می کنیم .

گام ۲: $U$ را از $U(0, 1)$ استخراج می کنیم .

گام ۳: اگر $U > \frac{f(Y)}{e(Y)}$ آنگاه $Y$ را رد می کنیم . یعنی مقدار $Y$ را به عنوان یک مشاهده از چگالی $g$ در نظر نمی گیریم .

گام ۴: در غیر این صورت مقدار $Y$ را پذیرفته و قرار میدهیم $X = Y$ ، یعنی $X$ را به عنوان یک مشاهده از چگالی $f$ در نظر می گیریم و برای بدست آوردن مشاهدات بعدی گام های ۱ تا ۴ را تکرار می کنیم تا به نمونه مورد نظر دست یابیم .

لازم به ذکر است که نمونه بدست آمده با این روش دارای چگالی $f$ می باشند. زیرا

$$P(X \leq y) = P(Y \leq y \mid U \leq \frac{f(Y)}{e(Y)})$$

$$= \frac{P(Y \leq y, U \leq \frac{f(Y)}{e(Y)})}{P(U \leq \frac{f(Y)}{e(Y)})}$$

$$= \frac{\int_{-\infty}^{y} \int_{0}^{\frac{f(z)}{e(z)}} g(z)dudz}{\int_{-\infty}^{\infty} \int_{0}^{\frac{f(z)}{e(z)}} g(z)dudz}$$

$$= \int_{-\infty}^{y} f(z)dz$$

چند تذکر :

۱: شرط کرانداری $\frac{f}{g}$ به $M \geq 1$ شرط حیاتی است .

۲: احتمال پذیرش $Y$ ی استخراج شده از $g$ ، $\frac{1}{M}$ می باشد . و این نشان می دهد که اگر $M$ بزرگ باشد نرخ پذیرش $(\frac{1}{M})$ کوچک بوده و الگوریتم کارا نخواهد بود .

۳: چگونه می توان $M$ را کوچک ساخت ؟ بدیهی است که مقدار $M$ به $g$ انتخاب شده بستگی دارد . اگر $g$ شبیه $f$ باشد آنگاه احتیاج به $M$ بزرگ نخواهد بود. پس بنابراین $g$ ای انتخاب می کنیم که به سادگی قابل نمونه گیری بوده و هر چه بیشتر مشابه $f$ باشد .

۴: اگر دم های چگالی $g$ کوتاهتر از $f$ باشد ، آنگاه $M$ مناسبی نمی توان پیدا کرد .

**مثال ۱:**

فرض کنیم می خواهیم از توزیع نرمال دم بریده در نقطه $c$ ، نمونه ای به حجم $n$ استخراج کنیم. با استفاده از روش پذیرش — رد

$$g(x) = \lambda e^{-\lambda x} \quad x > 0 \quad \lambda > 0$$

را پیشنهاد می کنیم . می دانیم که استخراج نمونه از چگالی نمایی با استفاده از روش تبدیل معکوس تابع توزیع براحتی امکان پذیر بوده و نیز دم توزیع نمایی ضخیم تر از توزیع نرمال است . بنابراین توزیع نمایی انتخاب مناسبی است . در این جا

$$f(x) = \frac{\phi(x+c)}{1 - \Phi(c)}$$

که در آن

$$\Phi(c) = P(Z \leq c) \quad Z \sim N(0, 1)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

حال می خواهیم $M \geq 1$ ای را پیدا کنیم که $\frac{f}{g}$ به آن کراندار بوده یعنی

$$Mg(x) \geq f(x) \quad \forall x > c$$

و همچنین $M$ کمترین مقدار را اختیار کند . بدین منظور قرار می دهیم

$$M = \sup_{\{x:\ x>c\}} \frac{f(x)}{g(x)}$$

$$= \sup_{x>c} \frac{\phi(x+c)}{\lambda e^{-\lambda x}(۱ - \Phi(c))}$$

$$= \frac{۱}{\lambda\sqrt{۲\pi}(۱ - \Phi(c))} \sup_{x>c} \exp\{\lambda x - \frac{(x+c)^۲}{۲}\}$$

سوپریمم در نقطه $x = \lambda - c$ رخ می دهد . و بنابراین :

$$M = \frac{e^{(\frac{\lambda^۲ - ۲\lambda c}{۲})}}{\sqrt{۲\pi}\ \lambda(۱ - \Phi(c))}$$

حال می توان $\lambda$ را طوری انتخاب کرد که $M$ کمینه و در نتیجه $\frac{۱}{M}$ بیشینه گردد .
می توان نشان داد که

$$\lambda = \frac{c + \sqrt{c^۲ + ۴}}{۲}$$

مقدار مطلوب برای $\lambda$ است .

### ۲.۴.۱ محدودیت های روش پذیرش ـرد

دیدیم که روش پذیرش ـرد ابراز قدرتمندی برای استخراج نمونه از توزیع های نچندان خوش رفتار می باشد . اما مواردی نیز پیش می آید که قادر به استفاده از این روش نیستیم . از این موارد به دو مورد زیر می توان اشاره کرد .

۱: شایعترین مورد زمانی اتفاق می افتد که چگالی هدف ، $f(x)$ به طور کامل معلوم نباشد و تنها می دانیم که $f(x) = c\tilde{f}(x)$ . که در آن $c$ مجهول می باشد. دراستنباط بیزی به کرات با این مورد برخورد می کنیم .

۲: پیدا کردن تابع چگالی $g$ که به سادگی قابل نمونه گیری باشد و نسبت $\frac{f}{g}$ کراندار بماند، همیشه کار چندان ساده ای هم نیست . به علاوه ممکن است که چنین $g$ ای را پیدا کنیم اما کران $M$ عدد نسبتا بزرگی باشد و در نتیجه احتمال پذیرش $\frac{۱}{M}$ ، ناچیز بوده و این کارایی روش پذیرش ـرد را به شدت کاهش می دهد

# فصل ۲

# زنـجیـره‌های مارکـف

## ۱.۲ تعاریـــف

دنباله‌ی متغیرهای تصادفی $\{X^{(t)} : t \geq ۰\}$، که همه‌ی $X^{(t)}$ ها در مجموعه‌ی شمارای $S$ مقدار می گیرند، را « زنجیره‌ی مارکف همگن » با فضای وضعیت $S$ گوییم هرگاه در شرط زیر صدق کند:

$$Pr(X^{(t+۱)} = j \mid X^{(t)} = i, X^{(t-۱)} = i_{t-۱}, \ldots, X^{(۱)} = i_۱, X^{(۰)} = i_۰)$$
$$= Pr(X^{(t+۱)} = j \mid X^{(t)} = i) = p_{ij}$$

$p_{ij}$ را احتمال انتقال یک مرحله‌ای از وضعیت $i$ به وضعیت $j$ گفته، و قرار می دهیم : $P = [p_{ij}]_{i,j \in S}$ در اینصورت $P$ ماتریس احتمال انتقالات یک مرحله‌ای زنجیره‌ی $\{X^{(t)}\}$ است. بدیهی است که : $\forall i, j \in S : p_{ij} \geq ۰; \sum_{j \in S} p_{ij} = ۱$ یعنی هر سطر ماتریس $P$ یک احتمال شرطی بر فضای وضعیت $S$ تعریف می کند.

مشابهاً اگر تعریف کنیم : $P^{(m)} = [p_{ij}^{(m)}]_{i,j \in S}; p_{ij}^{(m)} = Pr(X^{t+m} = j \mid X^t = i)$ آنگاه $p_{ij}^{(m)}$ را احتمال انتقال $m$ مرحله‌ای از وضعیت $i$ به وضعیت $j$، و $P^{(m)}$ را ماتریس احتمال انتقال $m$ مرحله ای زنجیره‌ی $\{X^{(t)}\}$ می گوییم.

از آنـجایـی کـه :

$$p_{ij}^{(۲)} = Pr(X^{(۲)} = j \mid X^{(۰)} = i) = \sum_{k \in S} Pr(X^{(۲)} = j, X^{(۱)} = k \mid X^{(۰)} = i)$$
$$= \sum_{k \in S} Pr(X^{(۲)} = j \mid X^{(۱)} = k) Pr(X^{(۱)} = k \mid X^{(۰)} = i) = \sum_{k \in S} p_{ik} p_{kj}$$

لـذا : $P^{(۲)} = P \times P$.

به همین ترتیب می توان نشان داد :

$$P^{(m)} = \underbrace{P \times \ldots \times P}_{m}$$
$$P^{(m)} \times P^{(l)} = P^{(m+l)}$$

که به برابری های « چپمن–کلموگروف » مشهورند و بیان می کنند که با داشتن ماتریس احتمال انتقالات یک مرحله‌ای، ماتریس احتمال انتقالات چند مرحله‌ای را می توان بدست آورد. لذا ماتریس احتمال انتقالات یک مرحله‌ای $(P)$ کاملاً ساختار انتقالات زنجیره ی مارکف $\{X^{(t)}\}$ را مشخص می کند. به عبارت دیگر: هر زنجیره مارکف با ماتریس احتمال انتقالات یک مرحله‌ای‌اش قابل بیان است(با داشتن $P$ زنجیره $\{X^{(t)}\}$ را داریم).

## ۲.۲ ساختار احتمالی زنجیره

$X^{(0)}$ را وضعیت اولیه زنجیره می گویند و با داشتن توزیع $X^{(0)}$، توزیع تمام $X^{(t)}$ ها را خواهیم داشت. زیرا :

فرض کنیم $X^{(0)}$ دارای توزیع $\Pi^{(0)}$ باشد. این توزیع را – برای سادگی نمادگذاری – بصورت زیر نمایش می دهیم :

$$\pi_i^{(0)} = Pr(X^{(0)} = i)$$
$$\Pi^{(0)} = [\pi_i^{(0)}]_{i \in S}$$

حال اگر توزیع $X^{(1)}$ را با $\Pi^{(1)}$ نشان دهیم در اینصورت :

$$\Pi^{(1)} = [\pi_j^{(1)}]_{j \in S}$$
$$\pi_j^{(1)} = Pr(X^{(1)} = j) = \sum_{i \in S} Pr(X^{(1)} = j \mid X^{(0)} = i) Pr(X^{(0)} = i)$$
$$= \sum_{i \in S} \pi_i P_{ij}$$

با نماد گذاری ماتریسی یعنی :

$$(\Pi^{(1)})^T = (\Pi^{(0)})^T \times P$$

مشابهاً اگر توزیع $X^{(t)}$ را با $\Pi^{(t)}$ نشان دهیم داریم :

$$\Pi^{(t)} = [\pi_j^{(t)}]_{j \in S}$$
$$(\Pi^{(t)})^T = (\Pi^{(t)})^T \times P$$

که نشان می دهد توزیع $X^{(t)}$ تنها به ماتریس $P$ و توزیع $\Pi^{(0)}$ بستگی دارد. به عبارت دیگر $P$ و $\Pi^{(0)}$ کل ساختار احتمالی زنجیره را بدست می دهند.

## ۳.۲ تـوزیع حـدی زنـجیـره

دیدیم که برای زنجیره مارکف با ماتریس احتمال انتقالات یک مرحله ای $P$، با انتخاب یک توزیع احتمال برای وضعیت اولیه زنجیره ($X^{(۰)}$)، دنباله‌ی توزیع های احتمال $\{\Pi^{(t)}\}_{t\geq ۰}$ برای $X^{(t)}$ ها بدست آمد. برای این دنباله از توزیع ها چنانچه داشته باشیم :

$$\lim_{t\to\infty} \Pi^{(t)} = \Pi$$

و $\Pi$ یک توزیع احتمال بر$S$ باشد، آنگاه $\Pi$ را تـوزیع حـدی زنـجیـره گوییم.

روشـن است کـه بـــرای انـتخاب های مـخـتـلـف $\Pi^{(۰)}$، تـوزیع های حـدی —در صورت وجود— ممکن است متفاوت باشند. آیا زنجیره‌هایی هستند که برای هر انتخاب $\Pi^{(۰)}$، توزیع حدی همواره وجـود داشته و یـــکتا باشد ؟

در تعریف زیر این نوع زنجیره ها را نامگذاری کرده و به ویژگی مطلوب آنها برای استخراج نمونه اشاره می کنیم. در بخش های بعد با ارائه‌ی مقدمات و سپس قضیه ای (قضیه ارگودیک)، شرطهای کافی برای این نوع زنجیره‌های مارکف را می آوریم.

زنجیره‌ی $\{X^{(t)}\}$ را «ارگـودیـــک» گوییم هر گاه برای هر توزیع ابتدایی $\Pi^{(۰)}$، دنباله توزیع های احتمال بدست آمده $\{\Pi^{(t)}\}_{t\geq ۰}$ به توزیع احتمال $\Pi$ همگرا باشد. یعنی :

$$\forall\ \Pi^{(۰)} \quad : \quad \lim_{t\to\infty} \Pi^{(t)} = \Pi$$

به عـبارت دیـگـر برای زنـــجیـره ارگـودیـک، صرف نظر از توزیع اولیه‌ی زنـجیـره($\Pi^{(۰)}$)، تـوزیع $X^{(t)}$ ها ($\Pi^{(t)}$) به توزیع حدی یکتای $\Pi$ میل می کند.

همین مسئله یک ایـــــده برای استخراج نمونه به مامی دهد : فرض کنیم بخواهیم از چگالی $f(x)$ با تکیه گاه شمارای $S$ نمونه‌ای استخراج کنیم اما با روشهای معمول[۱] قادر به این کار نباشیم. اگر زنجیره‌ی مارکف ارگودیکی داشته باشیم که توزیع حدی یکتای آن $f(x)$ باشد و از هر سطر ماتریس احتمال انتقالات یک مرحله‌ای آن ($P$) بتوان نمونه تولید کرد[۲] به روش زیر عمل می کنیم:

۱) از تـوزیع دلخواه $\Pi^{(۰)}$، $X^{(۰)}$ را تولید می کنیم.

۲) با داشتن $X^{(t)} = x^{(t)}$ —بدست آمده از مرحله‌ی قبل— از سطر $x^{(t)}$ ام ماتریس $P$، $X^{(t+۱)}$ را تولید می کنیم.

۳) این کار را تا بدست آمدن تعداد مشاهدات مورد نـــیاز ادامه می دهیم.

در اینصورت با توجه به ارگودیک بودن زنجیره، توزیع $X^{(t)}$ ها برای $t$ های به قدر کافی بزرگ به $f(x)$ میل می کند. پس می توان نتیجه گرفت که نمونه بدست آمده از الگوریتم

---

[۱] روش معکوس تابع توزیع و روش پذیرش—رد

[۲] هر سطر ماتریس احتمال انتقالات، یک توزیع شرطی بر فضای وضعیت $S$ تعریف می کند.

فوق یک نمونه تقریبی از $f(x)$ است. برای بهبود کیفیت این نمونه (یعنی نزدیکی بیشتر توزیع $X^{(t)}$ به $f(x)$ ) و برای کاهش تأثیر انتخاب توزیع اولیه نمونه بسیار بزرگی به روش فوق استخراج کرده، تنها مشاهدات انتهایی بدست آمده را در نظر بگیریم و از مشاهدات ابتدایی صرف نظر کنیم.

نتایج فوق تنها بر پایه‌ی ارگودیک بودن زنجیره بناشده بود. حال باید ببینیم شرایط ارگودیک بودن یک زنجیره مارکف چیست؟

## ۴.۲ توزیع ایستای زنجیره

اگر برای وضعیت اولیه‌ی زنجیره مارکف، $X^{(0)}$)، توزیع $\Lambda = [\lambda_i]_{i \in S}$ را که برای آن داریم:

$$(\Lambda)^T = (\Lambda)^T \times P$$

در نظر بگیریم، آنگاه برای هر $t \geq 0$ توزیع $X^{(t)}$ نیز $\Lambda = [\lambda_i]_{i \in S}$ خواهد بود. در این حالت $\Lambda$ را توزیع ایستای زنجیره $\{X^{(t)}\}$ (یا متناظراً توزیع ایستای ماتریس $P$) می گوییم. واضح است که با انتخاب $\Lambda$ به عنوان توزیع اولیه زنجیره $\Lambda$ توزیع حدی زنجیره نیز می باشد.

## ۵.۲ زنجیره‌های تحویل ناپذیر و نامتناوب

برای هر $i, j \in S$، گوییم وضعیت $j$ از وضعیت $i$ قابل دسترسی است اگر :

$$\exists k \in N \quad : \quad p_{ij}^{(k)} = P(X^{(t+k)} = j \mid X^{(t)} = i) > 0$$

به عبارت دیگر از وضعیت $i$ پس از مدت زمان متناهی بتوان به وضعیت $j$ رفت. دو وضعیت $i$ و $j$ را مرتبط گوییم هرگاه $j$ از $i$ قابل دسترسی و $i$ از $j$ قابل دسترسی باشد. زنجیره‌ی مارکف $\{X^{(t)}\}$ را تحویل ناپذیر گوییم هرگاه تمام وضعیت های آن مرتبط باشند.

برای $i \in S$ دوره‌ی تناوب وضعیت $i$ را با $d(i)$ نشان داده و آنرا به صورت زیر تعریف می کنیم :

$$d(i) = \gcd\{k \in N \quad : \quad p_{ii}^{(k)} > 0\}$$

که در آن gcd نشان دهنده‌ی بزرگترین مقسوم علیه مشترک یک مجموعه از اعداد می باشد. اگر $d(i) = 1$ وضعیت $i$ را نامتناوب می گوییم. به عبارت دیگر وضعیت $i$ را

نامتناوب است هرگاه از $i$ در هر زمانی بتوان دوباره به $i$ برگشت. بطور کلی زنجیره را نا متناوب گوییم هرگاه هر وضعیت آن نامتناوب باشد.

قضیه‌ی ارگودیک : برای زنجیره‌ی مارکف تحویل ناپذیر و نامتناوب $\{X^{(t)}\}$، اگر بتوان یک جواب غیر صفر برای معادله‌ی

$$(Z)^T \times P = (Z)^T \quad ; \quad Z = [z_i]_{i \in S}$$

بدست آورد طوری که : $\sum_{i \in S} z_i < \infty$، آنگاه زنجیره $\{X^{(t)}\}$ ارگودیک است.

نتیجه : زنجیره‌ی مارکف تحویل ناپذیر و نامتناوب $\{X^{(t)}\}$ اگر توزیع ایستا داشته باشد آنگاه ارگودیک است. در این حالت این توزیع ایستا، توزیع حدی یکتای زنجیره‌ی مارکف $\{X^{(t)}\}$ می باشد.

پس اگر یک زنجیره‌ی مارکف تحویل ناپذیر و نامتناوب باشد، برای ارگودیک بودن آن کافی است نشان دهیم که توزیع ایستا دارد و با تعیین آن، توزیع ایستای حدی یکتای آن بدست می آید. اما نشان دادن این مطلب مستلزم پیدا کردن توزیع $\Pi = [\pi_i]_{i \in S}$ است که برای آن داشته باشیم : $(\Pi)^T = (\Pi)^T \times P$ که جز در موارد خیلی خاص کار مشکل، و عمدتاً غیر ممکنی است. ویژگی زمان برگشت پذیری کار را ساده‌تر می کند.

## ٦.٢ زنجیره‌ی زمان برگشت پذیر

زنجیره‌ی مارکف $\{X^{(t)}\}$ را زمان برگشت پذیر گویند هرگاه یک توزیع $\Pi = [\pi_i]_{i \in S}$ بر $S$ یافت شود که برای آن داشته باشیم :

$$\forall i, j \in S \quad : \quad \pi_i p_{ij} = \pi_j p_{ji}$$

اگر زنجیره‌ی مارکف $\{X^{(t)}\}$ به معنی فوق زمان برگشت پذیر باشد آنگاه $\Pi$ توزیع ایستای زنجیره بوده ویکتاست. زیرا، با جمع بر روی $i$ دوطرف تساوی فوق داریم :

$$\forall j \in S \quad : \quad \sum_{i \in S} \pi_i p_{ij} = \pi_j$$

و این یعنی : $(\Pi)^T = (\Pi)^T \times P$ .

نتیجه : زنجیره مارکف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر ارگودیک است.

## ٧.٢ زنجیره‌ی مارکف با فضای وضعیت ناشمارا

دیدیم که با ساختن یک زنجیره‌ی مارکف ارگودیک که توزیع ایستای حدی یکتایش چگالی هدف $f(x)$ باشد، می توان یک نمونه تقریبی از $f(x)$ استخراج کرد. بدین منظور در این

بخش فضای وضعیت زنجیره را از حالت شمارا به حالت ناشمارا تعمیم داده، و تعاریف و نتایج بدست آمده در بخش‌های قبل را در این حالت بررسی می کنیم.

دنباله ی متغیرهای تصادفی $\{X^{(t)} : t \geq 0\}$، که همه‌ی $X^{(t)}$ ها[3] در مجموعه‌ی ناشمارای $S$ مقدار می گیرند، را « زنجیره‌ی مارکف همگن » با فضای وضعیت $S$ گوییم هرگاه در شرط زیر صدق کند:

$$\forall A \in \mathcal{F}, \forall t \geq 0, \forall x_0, x_1, \ldots, x_{t-1}, x \in S \quad :$$
$$Pr(X^{(t+1)} \in A \mid X^{(t)} = x, X^{(t-1)} = x_{t-1}, \ldots, X^{(1)} = x_1, X^{(0)} = x_0)$$
$$= Pr(X^{(t+1)} \in A \mid X^{(t)} = x) = \int_A k(x, y)dy$$

که در آن $\mathcal{F}$ یک $\sigma$—میدان بر $S$، و $k(x, y)$ برای $x \in S$ ثابت، چگالی توزیع شرطی $K$ بر $S$ است(یعنی $(Y \mid X = x \sim K)$. به عبارت دیگر $\{k(x, .) : x \in S\}$ خانواده ای از چگالیهای شرطی بر $S$ می باشد. در این حالت $k(x, y)$ را هسته‌ی احتمال انتقالات[4] یک مرحله‌ای زنجیره گوییم و همان نقش ماتریس احتمال انتقالات یک مرحله‌ای را در اینجا ایفا می کند.

اگر برای وضعیت اولیه‌ی $X^{(0)}$، توزیع $\Pi^{(0)}$ با چگالی $\pi^{(0)}(x)$ را در نظر بگیریم، آنگاه مانند زنجیره‌ی مارکف با فضای وضعیت شمارا، $k(x, y)$ و $\pi^{(0)}(x)$ کل ساختار احتمالی زنجیره را مشخص می کنند. زیرا برای هر $t \geq 0$ داریم :

اگر $X^{(t)}$ داری توزیع $\Pi^{(t)}$ با چگالی $\pi^{(t)}(x)$ باشد و $X^{(t+1)}$ دارای توزیع $\Pi^{(t+1)}$ با چگالی $\pi^{(t+1)}(x)$، آنگاه :

$$P(X^{(t+1)} \in A) = \int_A \pi^{(t+1)}(y)dy = \int_A \int_S \pi^{(t)}(x)k(x, y)dxdy$$
$$\pi^{(t+1)}(y) = \int_S \pi^{(t)}(x)k(x, y)dx$$

و این یعنی توزیع $X^{(t+1)}$ با $\pi^{(t)}(x)$ و $k(x, y)$ قابل بیان است. پس برای زنجیره مارکف با هسته‌ی احتمال انتقالات یک مرحله‌ای $k(x, y)$، با انتخاب یک توزیع احتمال $\Pi^{(0)}$ برای وضعیت اولیه زنجیره $(X^{(0)})$، دنباله‌ی توزیع های احتمال $\{\Pi^{(t)}\}_{t\geq 0}$ با دنباله‌ی چگالیهای متناظر $\{\pi^{(t)}(x)\}_{t\geq 0}$ برای $X^{(t)}$ ها بدست می آید. در اینصورت اگر دنباله‌ی توابع چگالی $\{\pi^{(t)}\}$ بر $S$ به تابع $\pi$ همگرای یکنواخت باشد، که در آن $\pi$ یک تابع چگالی بر $S$ است، آنگاه $\pi(x)$ را چگالی حدی زنجیره، و توزیع $\Pi$ متناظر با چگالی $\pi$ را توزیع حدی زنجیره گوییم. در اینجا نیز برای انتخاب های مختلف $\Pi^{(0)}$ توزیع های حدی—در صورت وجود— ممکن است متفاوت باشند و تنها برای زنجیره‌های ارگودیک، صرف نظر از توزیع اولیه‌ی زنجیره، توزیع $X^{(t)}$ ها به توزیع حدی یکتای $\Pi$ میل می کند.

---

[3] فرض بر این است که همه‌ی $X^{(t)}$ ها پیوسته‌ی مطلق اند، یعنی توزیع آنها چگالی دارد.

[4] Transition kernel: با ماتریس احتمال انتقالات مقایسه کنید.

اگـــر بـرای تـوزیع $\Lambda$ با چـگالی $\lambda(x)$ داشتـه باشیـم :

$$\forall y \in S \quad : \quad \lambda(y) = \int_S \lambda(x) k(x,y) dx$$

در اینصورت $\Lambda$ را توزیع ایستا، و $\lambda(x)$ را چگالی ایستای زنجیره می گوییم. واضح است که با انتخاب $\Lambda$ به عنوان توزیع اولیه زنجیره، $\Lambda$ توزیع حدی زنجیره نیز می باشد.

تمام تعاریف و نتایج ارگودیک بودن، تحویل ناپذیری، نامتناوب بودن و زمان بازگشتی بودن مشابه با آنچه در حالت زنجیره با فضای وضعیت شمارا آمده است می باشد، تفاوت تنها در استفاده از $k(x,y)$ به جای $p_{ij}$، $\pi^{(t)}(x)$ به جای $\pi_i^{(t)}$، و انـــتـگـرال به جای ســیـگـما می باشد.

# فصل ۳

# روش مونت کارلو با زنجیره‌های مارکف ($MCMC$)

دیدیم که برای تقریب کمیت

$$\theta = E(X) = \int_S h(x)f(x)dx$$

به روش مونت کارلو، با استخراج مشاهدات مستقل و هم توزیع، از چگالی $f(x)$ و استفاده از $\hat{\theta}_{MC} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$ کمیت $\theta$ را تقریب می زدیم. اما استخراج نمونه از چگالی $f(x)$ مشکل اساسی روش مونت کارلو می باشد. برای این کار دو روش معکوس تابع توزیع و روش پذیرش و رد را معرفی کرده، و محدودیتهای آنها را برشمردیم. اکنون به دنبال روش قدرتمندتری برای استخراج نمونه هستیم که محدودیتهای کمتری داشته باشد.

در بحث زنجیره‌های مارکف ارگودیک، الگوریتمی را معرفی کردیم که به کمک آن می توان از چگالی هدف $f(x)$ با تکیه گاه $S$ نمونه استخراج کرد. برای این کار، لازم بود که یک زنجیره‌ی مارکف ارگودیک با توزیع ایستای حدی یکتای $f(x)$ داشته باشیم بطوریکه از هر سطر ماتریس احتمال انتقالات در حالت $S$ شمارا و یا از هر چگالی $k(x,y)$، برای $x \in S$ ثابت، درحالت $S$ ناشمارا بتوان نمونه تولید کرد. به علاوه دیدیم که زنجیره‌ی مارکف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر ارگودیک است. حال مسئله این است که چگونه می توان یک زنجیره ی مارکف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر باتوزیع ایستای حدی یکتای $f(x)$ بر $S$ ساخت طوریکه از سطرهای ماتریس احتمال انتقالات و یا چگالیهای هسته احتمال انتقالات بتوان نمونه تولید کرد.

در واقع روش مونت کارلو با زنجیره‌های مارکف ($MCMC$)، چنین زنجیره‌ای را برای استخراج یک نمونه‌ی تقریبی از $f(x)$ ساخته (بخش زنجیره‌ی مارکف ِروش) و با استخراج

نمونه، روش مونت کارلو (بخش مونت کارلو روش) را برای تقریب کمیتها را به کار برد. مارکف نشان داد که قانون اعداد بزرگ در حالت وابستگی مارکفی نمونه، باز هم برقرار است. بنابراین در بخش مونت کارلو روش، چنانچه از تقریبی بودن توزیع کناری مشاهدات بدست آمده چشم پوشی، و فرض کنیم که مشاهدات ما وابسته‌ی مارکفی‌اند و توزیع کناری آنها $f(x)$ است آنگاه با پشتوانه‌ی قانون اعداد بزرگ می توان روش مونت کارلو را برای تقریب کمیت ها به کار برد. لذا بخش مونت کارلو روش با پذیرش یک تقریب سر راست است. اما مسئله‌ی اصلی ساختن زنجیره‌ی مارکف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر با چگالی ایستای حدی یکتای $f(x)$، یعنی بخش زنجیره‌ی مارکف $MCMC$ می باشد.

برای تولید چنین زنجیره‌ای دو الگوریتم ارائه شده است : الگوریتم متروپلیس–هستینگ[۱] و نمونه گیری گیبز[۲].

## ۱.۳ الگوریتم متروپلیس–هستینگ

این الگوریتم با انتخاب یک هسته‌ی احتمال انتقالات $g(y \mid x)$ بر $S$، با داشتن مشاهده $X^{(t)}$، پس از سه گام، مشاهدی $X^{(t+1)}$ را به ما می دهد. بدین منظور به یک مقدار اولیه برای بدست آوردن سایر مشاهدات نیاز داریم :

در زمان $t = 0$، $X^{(0)} = x^{(0)}$ را از توزیع ابتدایی دلخواه $\Pi^{(0)}$ با چگالی $\pi^{(0)}(x)$،که بر $S$ تعریف شده است، با این شرط که $f(x^{(0)}) > 0$، تولید کند. البته می توان $x^{(0)} \in S$ را با شرط $f(x^{(0)}) > 0$، به دلخواه انتخاب کرد که معادل است با انتخاب یک توزیع بر $S$ با چگالی $\pi^{(0)}(y) = 0 \quad ; \quad \forall y \in S, y \neq x^{(0)} \quad : \quad \pi^{(0)}(x^{(0)}) = 1$.

حال با داشتن $X^{(t)} = x^{(t)}$، الگوریتم مقدار $X^{(t+1)}$ را بصورت زیر تولید می کند:

۱) مقدار پیشنهادی $X^*$ را از چگالی $g(. \mid x^{(t)})$ تولید می کنیم.

۲) نسبت متروپلیس–هستینگ را که بصورت $R(u,v) = \frac{f(v)g(u|v)}{f(u)g(v|u)}$ تعریف شده است را برای $(x^{(t)}, X^*)$ محاسبه می کنیم($R(x^{(t)}, X^*)$).

۳) $X^{(t+1)}$ را بصورت زیر انتخاب می کنیم :

- با احتمال $\min\{1, R(x^{(t)}, X^*)\}$ ; $X^{(t+1)} = X^*$
- با احتمال $1 - \min\{1, R(x^{(t)}, X^*)\}$ ; $X^{(t+1)} = x^{(t)}$

[۱] Metropolis-Hasting
[۲] Gibbs Sampling

این اگوریتم را تا زمانی که تعداد مشاهدات مورد نیاز را بدست آمد، تکرار می کنیم. لذا این الگوریتم با انتخاب هسته‌ی احتمال انتقالات $g(y \mid x)$ بر $S$، و مقدار اولیه‌ی $x^{(0)}$ یک زنجیره ما می دهد.

## تـــوجـــــــه

- مقدار $X^*$ را که در گام یک تولید می شود، مقدار پیشنهادی می گوییم چون در گام سوم پذیرفته یا رد می شود.
- چگالی شرطی $g(y \mid x)$ را که با آن مقادیر پیشنهادی $X^*$ را با داشتن $x^{)t)}$ تولید می کنیم، چگالی پیشنهادی می گوییم.
- نسبت مترو پلیس–هستینگ $R(x^{(t)}, X^*)$ همواره تعریف می شود. زیرا مقدار پیشنهادی $X^* = x^*$ تنها وقتی تولید می شود که $f(x^{(t)}) > 0$ و $g(x^{(*)} \mid x^{(t)}) > 0$.
- مارکف بودن زنجیره بدست آمده با این روش بدیهی است زیرا $X^{(t+1)}$ تنها به $X^{(t)}$ بستگی دارد.
- تحویل ناپذیری و نامتناوب بودن زنجیره به انتخاب چگالی پیشنهادی $g(y \mid x)$ بستگی دارد. بنابراین باید $g(y \mid x)$ طوری انتخاب شود که در نهایت زنجیره‌ی بدست آمده از الگوریتم تحویل ناپذیر و نامتناوب باشد.
- زمان برگشت پذیری زنجیره همواره برقرار است.

زیرا با انتخاب $f(x)$ به عنوان توزیع اولیه برای تولید $X^{(0)}$ داریم : اگر $w(x^{(0)}, x^{(1)})$ چگالی توام $(X^{(0)}, X^{(1)})$ باشد آنگاه :

$$w(x^{(0)}, x^{(1)}) = Pr(X^{(1)} = x^{(1)} \mid X^{(0)} = x^{(0)}) f(x^{(0)})$$

ولی، برای احتمال شرطی داریم :

$$\begin{gathered} Pr(X^{(1)} = x^{(1)} \mid X^{(0)} = x^{(0)}) \\ = Pr(X^* = x^{(1)}, U \le \min\{1, R(x^{(0)}, X^*)\} \mid X^{(0)} = x^{(0)}) \\ = Pr(U \le \min\{1, R(x^{(0)}, X^*)\} \mid X^* = x^{(1)}, X^{(0)} = x^{(0)}) \\ Pr(X^* = x^{(1)} \mid X^{(0)} = x^{(0)}) \\ = Pr(U \le \min\{1, R(x^{(0)}, x^{(1)})\} \times g(x^{(1)} \mid x^{(0)}) \end{gathered}$$

اگر $R(x^{(0)}, x^{(1)}) \le 1$ آنگاه : $f(x^{(1)}) g(x^{(0)} \mid x^{(1)}) \le f(x^{(0)}) g(x^{(1)} \mid x^{(0)})$

$$\begin{gathered} Pr(X^{(1)} = x^{(1)} \mid X^{(0)} = x^{(0)}) = R(x^{(0)}, x^{(1)}) g(x^{(1)} \mid x^{(0)}) \\ = \frac{f(x^{(1)}) g(x^{(0)} \mid x^{(1)})}{f(x^{(0)}) g(x^{(1)} \mid x^{(0)})} g(x^{(1)} \mid x^{(0)}) = \frac{f(x^{(1)}) g(x^{(0)} \mid x^{(1)})}{f(x^{(0)})} \end{gathered}$$

و اگر $f(x^{(0)})g(x^{(1)} \mid x^{(0)}) > f(x^{(1)})g(x^{(0)} \mid x^{(1)})$ آنگاه : $R(x^{(0)}, x^{(1)}) > 1$

$$Pr(X^{(1)} = x^{(1)} \mid X^{(0)} = x^{(0)}) = g(x^{(1)} \mid x^{(0)})$$

پـــس :

$$\begin{aligned} w(x^{(0)}, x^{(1)}) &= f(x^{(1)})g(x^{(0)} \mid x^{(1)}) \;\; ; \;\; R(x^{(0)}, x^{(1)}) \leq 1 \\ &= f(x^{(0)})g(x^{(1)} \mid x^{(0)}) \;\; ; \;\; R(x^{(0)}, x^{(1)}) \;\; > 1 \end{aligned}$$

مـــشابهاً مـی تـــوان نـــشان داد:

$$\begin{aligned} w(x^{(1)}, x^{(0)}) &= f(x^{(1)})g(x^{(0)} \mid x^{(1)}) \;\; ; \;\; R(x^{(0)}, x^{(1)}) \leq 1 \\ &= f(x^{(0)})g(x^{(1)} \mid x^{(0)}) \;\; ; \;\; R(x^{(0)}, x^{(1)}) \;\; > 1 \end{aligned}$$

یعنی $w(x^{(0)}, x^{(1)}) = w(x^{(1)}, x^{(0)})$، به عبارت دیگر چگالی توام $(X^{(0)}, X^{(1)})$ متقارن است. بنابراین توزیع کـناری $X^{(0)}$ و $X^{(1)}$ یکسان و همان $f(x)$ می باشد. این مطلب نه تنها برای $(X^{(0)}, X^{(1)})$ بلکه برای هر $(X^{(t)}, X^{(t+1)})$ صادق است. لـــذا :

$$\begin{aligned} \forall x, y \in S \;\; : \;\; & Pr(X^{(t)} = x)Pr(X^{(t+1)} = y \mid X^{(t)} = x) = \\ & Pr(X^{(t+1)} = y)Pr(X^{(t)} = x \mid X^{(t+1)} = y) \end{aligned}$$

یعـــنـــی :

$$\forall x, y \in S \;\; : \;\; f(x)g(y \mid x) = f(y)g(x \mid y)$$

و این یعنی زنجیره‌ی بدست آمده زمان برگشت پذیر است، بنابراین $f(x)$ توزیع ایستای یکتای زنجیره می باشد[۳].

در اگوریتم متروپلیس–هستینگ مسئلـه‌ی اصلی پیدا کردن $g(y \mid x)$–ایی است که اولاً برای هر $x \in S$ ثابت قابل نمونه گیری باشد و ثانیاً باعث شود زنجیره‌ی حاصل از الگوریتم ارگودیک باشد. علاوه بر اینها، ملاکهای دیگری را نیز برای انتخاب $g(y \mid x)$ مناسب باید مد نظر قرار داد. واضح است که مشاهدات حاصل از الگوریتم باید نمونه‌ی خوبی از جامعه‌ی $f(x)$ باشند. به عنوان مثال زنجیره‌ای که بیشتر مقادیر پیشنهادی را رد می کند، منجر به مشاهداتی از $f(x)$ می شود که خیلی از آنها تکراری و ثابت هستند و بنابراین این نمونه نمی تواند پراکندگی جامعه‌ی $f(x)$ را به خوبی نشان دهد. همچنین زنجیره‌ای که همه‌ی مقادیر پیشنهادی را بپذیرد، منجر به مشاهداتی از $f(x)$ می شود که بسیار پراکنده بوده و این در سرعت همگرایی توزیع مشاهدات به توزیع حدی $f(x)$ تأثیر

[۳] اثبات برای حالت پیوسته مشابه همین است.

گذاشته، باعث می شود برای بدست آورد یک نمونه‌ی قابل قبول، مشاهدات بسیار زیادی را با الگوریتم تولید کنیم. لذا انتخاب $g(y \mid x)$ مناسب در الگوریتم متروپلیس–هستینگ حیاتی است.اجراهای عملی این الگوریتم نشان داده که با انتخاب $g(y \mid x)$ مناسب، نتایج بسیار رضایت بخش‌اند.

اما متأسفانه در حالت کلی نمی توان قاعده‌ای برای انتخاب $g(y \mid x)$ مناسب ارائه کرد. بلکه باید با توجه به ساختار چگالی $f(x)$ و تکیه گاه آن $S$، و با در نظر گرفتن شرطها و ملاکهای انتخاب $g(y \mid x)$ مناسب، $g(y \mid x)$ را انتخاب کرد. بهترین کار دسته‌بندی مسائلی است که با روش $MCMC$ در پی حل آنها هستیم. در آمار بیشترین کاربرد این روش در استنباط بیزی است. وقتی که ما می خواهیم میانگین پسین، میانه‌ی پسین، احتمالهای پسین و ...را محاسبه کنیم، اما در حل معمول انتگرال (بعضاً سیگما) حاصل از تعریف این کمیتها با مشکل روبروییم. استفاده از زنجیره‌های مستقل حداقل در این مورد خاص[۴] بسیار کارساز است.

## ۱.۱.۳ الگوریتم متروپلیس–هستینگ و زنجیره‌های مستقل

اگر در الگوریتم متروپلیس–هستینگ، $g(y \mid x)$ طوری انتخاب شود که داشته باشیم : $g(y \mid x) = g(y)$، که در آن $g(x)$ چگالی یک توزیع بر $S$ است، در اینصورت هر مقدار پیشنهادی $X^*$ مستقل از مشاهده‌ی قبلی $x^{(t)}$ از $g(x)$ تولید می شود. لذا زنجیره‌ی حاصل از الگوریتم مستقل می باشد.

با این انتخاب، نسبت متروپلیس–هستینگ به فرم زیر است :

$$R(x^{(t)}, X^*) = \frac{f(X^*)g(x^{(t)})}{f(x^{(t)})g(X^*)}$$

به علاوه اگر برای هر $x \in S$ که $f(x) > 0$، داشته باشیم : $g(x) > 0$ آنگاه زنجیره حتماً تحویل ناپذیر و نامتناوب بوده، لذا ارگودیک با چگالی ایستای حدی یکتای $f(x)$ می باشد. بنابراین در این حالت دیگر نیازی به چک کردن تحویل ناپذیری و نامتناوب بودن زنجیره‌ی حاصل از الگوریتم نبوده، بلکه کافیست شرط فوق در مورد $g(x)$ صدق کند.

مثال : استنباط بیزی

در استنباط بیزی با مسئله‌ی زیر مواجه‌ایم :

مشاهدات $y_1, \ldots, y_k$ را با تابع درستنمایی $L(\theta \mid y)$ داریم. به علاوه، برای پارامتر $\theta$ چگالی پیشین $\pi(\theta)$ را در نظر گرفته ایم. استنباط بیزی (برآورد نقطه‌ای یا فاصله‌ای و آزمون فرض) در مورد $\theta$ تماماً برپایه‌ی تابع چگالی پسین زیر می باشد :

[۴] البته این مورد آنقدرها هم خاص نیست!

$$\forall \theta \in \Theta \quad : \quad \pi(\theta \mid y) \propto \pi(\theta) L(\theta \mid y)$$
$$\forall \theta \in \Theta \quad : \quad \pi(\theta \mid y) = c\pi(\theta) L(\theta \mid y)$$

که ضریب تناسب بصورت زیر محاسبه می شود :

$$c = (\int_{\Theta} \pi(\theta) L(\theta \mid y) d\theta)^{-1}$$

جز در موارد استاندارد، معمولاً محاسبه‌ی انتگرالی که $c$ را تعیین می کند پیچیده است. بنابراین در استنباط بیز معمولاً $c$ را نداریم.

در مسئله‌ی برآورد بیزی $\theta$، چنانچه تابع زیان مسئله را مربع خطا در نظر بگیریم، برآورد زیر را برای پارامتر مجهول $\theta$ داریم:

$$\hat{\theta}_B = \int_{\Theta} \theta \pi(\theta \mid y) d\theta = \int_{\Theta} \theta c \pi(\theta) L(\theta \mid y) d\theta$$

از آنجا که محاسبه‌ی $c$ را دشوار فرض کردیم، دلیلی وجود ندارد که محاسبه‌ی $\hat{\theta}_B$ را ساده‌تر از محاسبه‌ی $c$ فرض کنیم. بنابراین به نظر می رسد برای داشتن برآورد بیزی باید به تقریب $\hat{\theta}_B$ قانع شد. نیز فرض کنیم که $\pi(\theta)$ قابل نمونه گیری بوده و می خواهیم با روش مونت کارلو این کار را انجام دهیم.

برای این کار ممکن است این راه حل به ذهن مان برسد : ابتدا با استخراج نمونه از $\pi(\theta)$، روش مونت کارلو را برای تقریب $c$ به کار ببریم. سپس با استخراج نمونه‌ای دیگر از $\pi(\theta)$، $\hat{\theta}_B$ را با $\frac{1}{n} \sum_{i=1}^{n} c\theta_i L(\theta_i \mid y)$ تقریب بزنیم. به این راه حل دو ایراد وارد است : اولاً ما برای تقریب $\hat{\theta}_B$ از دوبار تقریب، تقریب $c$ و بعد تقریب $\hat{\theta}_B$ با روش مونت کارلو، استفاده کرده ایم که مسلماً دقت تقریب $\hat{\theta}_B$ را کاهش میدهد. دوماً معمولاً محاسبه‌ی $L(\theta \mid y)$ به دلیل فرم حاصل ضربی آن ساده نبوده و اغلب با تقریب و خطاهای گرد کردن همراه است. بنابراین این تقریب نیز بر تقریب های اشاره شده قبلی اضافه شده، دقت برآورد را کاهش می دهد.

راه حل منطقی استفاده از روش مونت کارلو با زنجیره‌های مارکف $(MCMC)$ می باشد : با انتخاب $g(\theta^* \mid \theta^{(t)}) = \pi(\theta^{(t)})$ و استفاده از الگوریتم متروپلیس–هستینگ، یک زنجیره مستقل با توزیع ایستای حدی یکتای $\pi(\theta \mid y)$ بدست می آوریم (زیرا برای هر $\theta \in \Theta$ که $\pi(\theta \mid y) > 0$، به دلیل متناسب بودن، داریم : $\pi(\theta) > 0$). آنگاه با داشتن مشاهدات کافی $(n)$ از زنجیره و حذف مشاهدات اولیه $(m$ تای اول) برای بهبود کیفیت نمونه و کاهش تاثیر مقدار اولیه، $\hat{\theta}_B$ را با $\frac{1}{n-m} \sum_{t=m+1}^{n} \theta^{(t)}$ تقریب می زنیم.

مثال : برآورد پارامتر آمیخته

فرض کنیم $y_1, \ldots, y_{100}$ را بطور مستقل و همتوزیع از توزیع آمیخته‌ی

$$\delta N(7, \frac{1}{4}) + (1 - \delta) N(10, \frac{1}{4}) \quad ; \quad \delta \in (0, 1)$$

مشاهده کرده‌ایم. به عنوان مثال این مشاهدات ممکن است مربوط به دو جامعه‌ی $N(۷, \frac{۱}{۴})$ و $N(۱۰, \frac{۱}{۴})$ باشند که با احتمال مجهول $\delta$ از $N(۷, \frac{۱}{۴})$ و احتمال $۱ - \delta$ از $N(۱۰, \frac{۱}{۴})$ می باشند. با فرض چگالی پیشین $\pi(\delta) = ۱ \quad ; \quad \delta \in (۰, ۱)$ برای $\delta$، هدف برآورد بیز پارامتر مجهول $\delta$ می باشد. تابع درستنمایی مشاهدات عبارتست از :

$$L(\delta \mid y) = \prod_{j=۱}^{۱۰۰} \frac{۲}{\sqrt{۲\pi}} (\delta \exp(-۲(y_i - ۷)^۲) + (۱ - \delta) \exp(-۲(y_i - ۱۰)^۲))$$

لذا چگالی پسین $\delta$ بصورت $\pi(\delta \mid y) = cL(\delta \mid y)$ می باشد که در آن :

$$c^{-۱} = \int_۰^۱ L(\delta \mid y) d\delta$$
$$= \int_۰^۱ \prod_{j=۱}^{۱۰۰} \frac{۲}{\sqrt{۲\pi}} (\delta \exp(-۲(y_i - ۷)^۲) + (۱ - \delta) \exp(-۲(y_i - ۱۰)^۲))$$

با اینکه بعد $\delta$ یک است و مجموعه‌ای که انتگرال بر آن گرفته می شود خوشرفتار است ($[۰, ۱]$ در $\Re$ فشرده است) با این حال محاسبه‌ی $c$ کار چندان راحتی به نظر نمی رسد! و اگر برآورد بیز $\delta$ را بخواهیم با انتگرال زیر مواجه‌ایم :

$$\int_۰^۱ \delta \pi(\delta \mid y) d\delta$$

اما $\pi(\delta)$ به سادگی قابل نمونه گیری بوده و با توجه به آنچه در مثال قبل گفته شد استفاده از روش مونت کارلو با زنجیره‌های مارکف سر راست است. با انتخاب مقدار اولیه‌ی $\delta^{(۰)}$ از توزیع یکنواخت بر $(۰, ۱)$، و انتخاب هسته‌ی احتمال انتقالات $g(y \mid x) = \pi(x)$ بر $(۰, ۱)$، برای این مثال، الگوریتم متروپلیس—هستینگ با داشتن $\delta^{(t)}$، مشاهده‌ی $\delta^{(t+۱)}$ را بصورت زیر تولید می کند :

۱) $\delta^*$ را از یکنواخت بر $(۰, ۱)$ تولید می کنیم.

۲) نسبت متروپلیس—هستینگ $R(\delta^{(t)}, \delta*) = \frac{\pi(\delta^* | y)}{\pi(\delta^{(t)} | y)}$ را محاسبه می کنیم.

۳) $\delta^{(t+۱)}$ را بصورت زیر انتخاب می کنیم :

$$\delta^{(t+۱)} = \delta^* \quad ; \quad \min\{۱, R(\delta^{(t)}, \delta*)\} \text{ با احتمال}$$
$$= \delta^{(t)} \quad ; \quad ۱ - \min\{۱, R(\delta^{(t)}, \delta*)\} \text{ با احتمال}$$

همانطور که ملاحظه شد، استفاده از زنجیره‌های مستقل تا حدی ساده بوده، به ویژه در استنباط بیزی بسیار راه گشاست. البته در استنباط بیزی اگر از توزیع پیشین نتوان به سادگی نمونه گرفت این روش کارساز نیست و باید به دنبال زنجیره‌های دیگری بود.

معمولاً زنجیره‌های غیر مستقل سرعت همگرایی بیشتری نسبت به زنجیره‌های مستقل داشته، درنتیجه نمونه‌هایی با کیفیت تری تولید می کنند. بنابراین در مسائلی که می توان از آنها استفاده، بر زنجیره‌های مستقل ارجحیت دارند. علاوه بر زنجیره‌های مستقل زنجیره‌های ساده‌ی دیگری (مانند زنجیره‌ی گام تصادفی و ...) نیز برای انتخاب توزیع پیشنهادی دسته بندی شده‌اند. برای اطلاع از زنجیره‌های دیگر، [۱] منابعی را ذکر کرده است. حال دو مثال را می آوریم که در آنها از زنجیره‌های مستقل استفاده نشده است.

مثال : از جانسون و آلبرت ۱۹۹۹

کارخانه‌ای یکی از محصولاتش را بهبود داده، می خواهد نسبت مشتریانی که محصول جدید را به محصول قدیم ترجیح می دهند برآورد کند. این کارخانه مطمئن است که این نسبت از $\frac{۱}{۲}$ بیشتر است، یعنی $p \in [\frac{۱}{۲}, ۱]$. با انتخاب توزیع یکنواخت بر $[\frac{۱}{۲}, ۱]$ به عنوان توزیع پیشین پارامتر $p$، برآورد بیز $p$ مد نظر است. یک بررسی نمونه‌ای نشان داده، از ۲۰ مشتری متقاضی محصول، ۱۲ نفر محصول جدید را ترجیح داده‌اند. چگالی پیشین، تابع درستنمایی و چگالی پسین برای این مسئله بصورت زیر است :

$$\pi(p) = ۲ \quad ; \quad p \in [\tfrac{۱}{۲}, ۱]$$
$$L(p \mid y) = p^{۱۲}(۱-p)^{۸}$$
$$\pi(p \mid y) = cp^{۱۲}(۱-p)^{۸} \quad ; \quad p \in [\tfrac{۱}{۲}, ۱] \ , \quad c^{-۱} = \int_{\frac{۱}{۲}}^{۱} p^{۱۲}(۱-p)^{۸}dp$$

براورد بیز $p$ از حل $\int_{\frac{۱}{۲}}^{۱} cp^{۱۳}(۱-p)^{۸}dp$ بدست می آید. علاقمندیم که این انتگرال را به روش مونت کارلو با زنجیره‌های مارکف، منتها نه با زنجیره‌ی مستقل حل کنیم. باید $g(p^* \mid p^{(t)})$ مناسبی را که در شرطها و ملاکهای گفته شده صدق کند پیدا کنیم. در مورد $p \in [\frac{۱}{۲}, ۱]$ این کار بسیار سخت است. راه حل جانسون و آلبرت بدین صورت است :

با دوره پارامتر کردن مسئله هم فضای پارامتر را می توان طوری تغییر داد که انتخابهای بیشتری برای توزیع پیشنهادی داشته باشیم و هم فرم توزیع پسین انعطاف پذیرتر می شود[۵]. بدین منظور تبدیل $\theta = \ln(\frac{p-\frac{۱}{۲}}{۱-p}) \quad ; \quad \theta \in \Re$ که عکس این تبدیل $p = \frac{\frac{۱}{۲}+\exp(\theta)}{۱+\exp(\theta)}$ است، را برای پارامتر $p$ در نظر می گیریم. در اینصورت چگالی پسین به فرم زیر است :

$$\pi(\theta \mid y) \propto \frac{(\frac{۱}{۲}+\exp(\theta))^{۱۲}\exp(\theta)}{(۱+\exp(\theta))^{۲۲}} \quad ; \quad \theta \in \Re$$

حال برای $\theta$ انتخاب $g(\theta^*, \theta^{(t)})$ مناسب ساده‌تر است. جانسون و آلبرت چگالی پیشنهادی زیر را در نظر گرفته‌اند :

[۵] علاوه بر این موارد، دوباره پارامتره کردن مناسب مدل مزایای دیگری مانند : کاهش واریانس براوردگر، افزایش سرعت اجرای الگوریتم و ...، نیز دارد. به مرجع [۱] رجوع شود.

$$g(\theta^*, \theta^{(t)}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(\theta^* - \theta^{(t)})^2)$$

که چگالی توزیع $N(\theta^{(t)}, \sigma^2)$ است و $\sigma^2$ پیش از اجرای الگوریتم تعیین می شود. $\sigma^2$ در مناسب بودن توزیع پیشنهادی، کیفیت مشاهدات زنجیره و سرعت همگرایی به توزیع حدی نقش مهمی دارد. با این انتخاب و انتخاب مقدار اولیه‌ی زنجیره $\theta^{(0)}$ از توزیع $N(0, 1)$، الگوریتم متروپلیس—هستینگ با داشتن $\theta(t)$، $\theta^{(t+1)}$ را بصورت زیر تولید می کند:

۱) $\theta^*$ را از توزیع $N(\theta^{(t)}, \sigma^2)$ تولید می کنیم.

۲) نسبت متروپلیس—هستینگ $R(\theta^{(t)}, \theta^*) = \frac{\pi(\theta^*|y)g(\theta^{(t)}|\theta^*)}{\pi(\theta^{(t)}|y)g(\theta^*|\theta^{(t)})}$ را محاسبه می کنیم.

۳) $\theta^{(t+1)}$ را بصورت زیر انتخاب می کنیم :

$$\theta^{(t+1)} = \theta^* \quad ; \quad \min\{1, R(\theta^{(t)}, \theta^*)\} \text{ با احتمال}$$
$$= \theta^* \quad ; \quad 1 - \min\{1, R(\theta^{(t)}, \theta^*)\}$$

مثال : برآورد بیز پارامترهای توزیع وایبل
توزیع وایبل با پارامترهای $\alpha$ و $\lambda$، دارای چگالی به فرم زیر است :

$$f(x \mid \alpha, \lambda) = \alpha\lambda x^{\alpha-1} \exp(-\lambda x^\alpha) \quad ; \quad x > 0 \quad , \alpha, \lambda > 0$$

این توزیع عضو خانواده‌ی توزیعهای نمایی نمی باشد و با انتخاب یک توزیع پیشین برای پارامترهای $(\alpha, \lambda)$،توزیعهای پسین دقیقی برای $(\alpha, \lambda)$ بدست نمی دهد. لذا برآورد بیز $(\alpha, \lambda)$ با مشکل مواجه است. سعی می کنیم با روش مونت کارلو با زنجیره‌های مارکف، برآورد بیزی برای $(\alpha, \lambda)$ پیدا کنیم.

فرض کنیم مشاهدات $(x_1, x_2, x_3) = (\frac{2}{10}, \frac{1}{10}, \frac{25}{100})$ را از توزیع وایبل مشاهده کرده‌ایم و توزیع پیشین $\pi(\alpha, \lambda) \propto \exp(-\alpha)\lambda^{\beta-1} \exp(-\beta\lambda)$ را برای $(\alpha, \lambda)$ در نظر گرفته ایم که $\beta$ ثابت و معلوم است. تابع درستنمایی و چگالی پسین برای این مسئله بصورت زیر است :

$$L(\alpha, \lambda \mid x) = (\frac{5}{1000})^{\alpha-1}\alpha^3\lambda^3 \exp(-\lambda((\frac{1}{10})^\alpha + (\frac{2}{10})^\alpha + (\frac{25}{100})^\alpha))$$
$$\pi(\alpha, \lambda \mid x) \propto (\frac{5}{1000})^{\alpha-1}\alpha^3 \exp(-\alpha)\lambda^{\beta-2} \exp(-\lambda((\frac{1}{10})^\alpha + (\frac{2}{10})^\alpha + (\frac{25}{100})^\alpha + \beta))$$
$$c^{-1} =$$
$$\int_0^\infty \int_0^\infty (\frac{5}{1000})^{\alpha-1}\alpha^3 \exp(-\alpha)\lambda^{\beta-2} \exp(-\lambda((\frac{1}{10})^\alpha + (\frac{2}{10})^\alpha + (\frac{25}{100})^\alpha + \beta))d\alpha d\lambda$$

حل دو انتگرال زیر برآورد بیز $(\alpha, \lambda)$ را به ما می دهد :

$$\int_0^\infty \int_0^\infty (\frac{5}{1000})^{\alpha-1}\alpha^4 \exp(-\alpha)\lambda^{\beta-2} \exp(-\lambda((\frac{1}{10})^\alpha + (\frac{2}{10})^\alpha + (\frac{25}{100})^\alpha + \beta))d\alpha d\lambda$$
$$\int_0^\infty \int_0^\infty (\frac{5}{1000})^{\alpha-1}\alpha^3 \exp(-\alpha)\lambda^{\beta-1} \exp(-\lambda((\frac{1}{10})^\alpha + (\frac{2}{10})^\alpha + (\frac{25}{100})^\alpha + \beta))d\alpha d\lambda$$

روشن است که این انتگرالها را نمی توان به روشهای معمول حل کرد لذا حل به روش مونت کارلو با زنجیره‌های مارکف گزینه‌ی مناسبی است. با انتخاب چگالی پیشنهادی $g(\alpha^*, \lambda^* \mid \alpha^{(t)}, \lambda^{(t)}) = \frac{1}{\alpha^{(t)}\lambda^{(t)}} \exp(-\frac{\alpha^*}{\alpha^{(t)}} - \frac{\lambda^*}{\lambda^{(t)}})$ که حاصلضرب چگالی دو توزیع نمایی با میانگین های $\alpha^{(t)}$ و $\lambda^{(t)}$ است، و انتخاب مقدارهای اولیه‌ی $\alpha^{(0)}$ و $\lambda^{(0)}$ از توزیع نمایی با میانگین ۱، الگوریتم متروپلیس–هستینگ با داشتن $(\alpha(t), \lambda^{(t)})$، $(\alpha^{(t+1)}, \lambda^{(t+1)})$ را بصورت زیر تولید می کند :

۱) $\alpha*$ را از توزیع نمایی با میانگین $\alpha^{(t)}$ و $\lambda^*$ را از توزیع نمایی با میانگین $\lambda^{(t)}$ تولید کرده، تا $(\alpha^*, \lambda^*)$ را بدست آوریم.

۲) نسبت متروپلیس–هستینگ

$$R((\alpha^{(t)}, \lambda^{(t)}), (\alpha^*, \lambda^*)) = \frac{\pi((\alpha^*,\lambda^*)|x)g((\alpha^{(t)},\lambda^{(t)})|(\alpha^*,\lambda^*))}{\pi((\alpha^{(t)},\lambda^{(t)})|x)g((\alpha^*,\lambda^*)|(\alpha^{(t)},\lambda^{(t)}))}$$

را محاسبه می کنیم.

۳) $(\alpha^{(t+1)}, \lambda^{(t+1)})$ را بصورت زیر انتخاب می کنیم :

$$(\alpha^{(t+1)}, \lambda^{(t+1)}) = (\alpha^*, \lambda^*) \quad ; \quad \text{با احتمال } \min\{1, R((\alpha^{(t)}, \lambda^{(t)}), (\alpha^*, \lambda^*))\}$$
$$= (\alpha^{(t)}, \lambda^{(t)}) \quad ; \quad \text{با احتمال } 1 - \min\{1, R((\alpha^{(t)}, \lambda^{(t)}), (\alpha^*, \lambda^*))\}$$

دو نمودار که معمولاً برای بررسی زنجیره‌های تولید شده با الگوریتم متروپلیس–هستینگ مورد استفاده قرار می گیرد، مسیر نمونه و هیستوگرام مشاهدات زنجیره می باشند. در نمودار مسیر نمونه زنجیره، مشاهدات $X^{(t)}$ در مقابل زمان $t$ در یک محور مختصات رسم می شوند. با این نمودار سرعت و وضعیت همگرایی زنجیره، تأثیر مقدار اولیه و نحوه‌ی گردش و تغییرات مشاهدات زنجیره در $S$ را می توان دید و در مورد کارایی زنجیره‌ی بدست آمده، و در نتیجه الگوریتم قضاوت کرد. همچنین هیستوگرام مشاهدات زنجیره نیز می تواند ملاکی برای کارایی زنجیره تولید شده با الگوریتم باشد. مسلماً زنجیره‌ای کاراتر است که هیستوگرام مشاهدات آن به نمودار چگالی هدف $f(x)$ شبیه‌تر باشد.

مانند سایر الگوریتمهای عددی، چند بار اجرا کردن الگوریتم و بررسی نتایج این اجراها نیز ملاک دیگری برای کارایی الگوریتم می باشد. اگر در چند اجرای مختلف، الگوریتم نتایج تقریباً یکسانی به ما داد، نتیجه می گیریم الگوریتم کاراست. در غیر اینصورت، نمی توان به نتایج الگوریتم زیاد اعتماد کرد.

## ۲.۳ نمونه گیری گیبز

برای تقریب $\theta = E(X) = \int_S h(x)f(x)dx$ به روش مونت کارلو وقتی که با روشهای معمول نتوان نمونه‌ای از $f(x)$ استخراج کرد، الگوریتم متروپلیس—هستینگ روشی عمومی برای تولید یک زنجیره ارگودیک با چگالی ایستای حدی یکتای $f(x)$ می باشد که تحقق های این زنجیره را می توان به عنوان نمونه‌ای تقریبی از $f(x)$ فرض کرد. اما اجرای این الگوریتم زمانی امکان پذیر است که هسته‌ی احتمال انتقالات (چگالی پیشنهادی) مناسبی را برای تولید یک مشاهده به شرط داشتن مشاهده قبلی، داشته باشیم. اشاره شد که بطور کلی نمی توان قاعده‌ای برای انتخاب توزیع پیشنهادی ارائه کرد.

بنابراین بهتر است براساس مسائلی که به روش $MCMC$ می خواهیم حل کنیم، الگوریتمهای مخصوصی را طرح کرده و زنجیره‌های تولید شده با این الگوریتمها را دسته بندی کنیم. مثلاً در الگوریتم متروپلیس—هستینگ، یک دسته بندی بر اساس انواع زنجیره‌هایی که می توان با این الگوریتم تولید کرد، زنجیره‌های مستقل است.

حال با طرح مسئله‌ای، الگوریتم دیگری را برای حل این مسئله طرح می کنیم :

فرض کنیم $\mathbf{X}= (X_1, \ldots, X_p) \sim f(\mathbf{x})$. $\mathbf{X}_{-i}$ را تعریف می کنیم :

$$\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p) \;\; ; \;\; i = 1, \ldots, p$$

(بردار $\mathbf{X}= (X_1, \ldots, X_p)$ با حذف مولفه‌ی $i$ ام آن). همچنین برای $i = 1, \ldots, p$، توزیع شرطی $X_i \mid \mathbf{X}_{-i} = \mathbf{x}_{-i}$ با چگالی $f(X_i \mid \mathbf{x}_{-i})$ را در نظر می گیریم.

می خواهیم $\theta = E(\mathbf{X}) = \int_S h(\mathbf{x})f(\mathbf{x})d\mathbf{x}$ را به روش مونت کارلو تقریب بزنیم، اما به روشهای معمول نمی توانیم از $f(\mathbf{x})$ نمونه استخراج کنیم. چنانچه برای $i = 1, \ldots, p$، $f(x_i \mid \mathbf{x}_{-i})$ طوری باشد که با داشتن $\mathbf{x}_{-i}$ بتوان از آن نمونه تولید کرد، آنگاه با انتخاب وضعیت اولیه‌ی $\mathbf{x}^{(0)}$، الگوریتم زیر با داشتن $\mathbf{X}^{(t)}$، $\mathbf{X}^{(t+1)}$ را تولید می کند :

۱) یک ترتیب از مولفه‌های $\mathbf{x}^{(t)}$ را انتخاب می کنیم.

۲) برای $i = 1, \ldots, p$، $X_i^*$ را از $f(x_i \mid \mathbf{x}_{-i}^{(t)})$ تولید کرده، تا $\mathbf{X}^* = (X_1^*, \ldots, X_p^*)$ بدست آید.

۳) قرار می دهیم : $\mathbf{X}^{(t+1)} = \mathbf{X}^*$

### توجه

- این الگوریتم را گیبز برای تقریب یک انتگرال در فیزیک پیشنهاد کرده است و به همین خاطر به نمونه گیری گیبز یا الگوریتم گیبز مشهور است.

- زنجیره‌ای که با این الگوریتم بدست می آید، $\{\mathbf{X}^{(t)}\}_{t\geq 0}$، مارکف است زیرا $\mathbf{X}^{(t+1)}$ تنها به $\mathbf{X}^{(t)}$ وابسته است.

- در گام دوم به جای اینکه $\mathbf{X}^*$ از $f(\mathbf{x}|\ \mathbf{x}^{(t)})$ تولید شود، با تولید $p$ مولفه‌ی $X_i^*$ بردار $\mathbf{X}^*$ بدست می آید. برای اینکه ترتیب تولید مولفه‌ها بر تولید $\mathbf{X}^*$ تأثیر نداشته باشد، در گام اول از بین $p!$ جایگشت ممکن از مولفه‌های $\mathbf{X}^{(t)}$ یکی را انتخاب و گام دوم را براساس این ترتیب انتخاب شده اجرا می کنیم. به همین دلیل در برخی از موارد ابتدا یک توزیع را احتمال بر $p!$ جایگشت مولفه‌های $\mathbf{X}^{(t)}$ در نظر گرفته، گام اول را چنین بیان می کنند : یک جایگشت را از توزیع تعریف شده بر جایگشتهای مولفه‌های $\mathbf{X}^{(t)}$ تولید می کنیم. سپس گام دوم را با این ترتیب تولید شده اجرا می کنند.

- اجرای گام دوم بر اساس ترتیب انتخاب شده در گام اول را یک سیکل می گوییم.

- برای یک سیکل خاص، در $i$ امین مرحله‌ی سیکل الگوریتم با داشتن $\mathbf{X}^{(t)}$، مقدار پیشنهادی $g_i(\mathbf{x}^*\ |\ \mathbf{x}^{(t)})$ $\mathbf{X}^* = (x_1^{(t)}, \ldots, X_i^*, \ldots, x_p^{(t)})$ را از چگالی پیشنهادی تولید می کند. که در آن :

$$g_i(\mathbf{x}^*\ |\ \mathbf{x}^{(t)}) = f(x_i^*\ |\ \mathbf{x}_{-i}^{(t)}) \quad ; \quad \mathbf{x}_{-i}^* = \mathbf{x}_{-i}^{(t)}$$
$$= 0 \qquad ; \quad \text{سایر نقاط}$$

در این حالت، $R(\mathbf{x}^{(t)}, \mathbf{x}^*) = \frac{f(x^*)g_i(x^{(t)}|x^*)}{f(x^{(t)})g_i(x^*|x^{(t)})} \equiv 1$، یعنی مقدار پیشنهادی $\mathbf{X}^*$ پذیرفته می شود. بنابراین هر سیکل از $p$ الگوریتم متروپلیس—هستینگ تشکیل شده است و می توان الگوریتم فوق را حالت خاصی از متروپلیس—هستینگ انگاشت.

- می دانیم که $f(x_i\ |\ \mathbf{x}_{-i}) = c_i f(\mathbf{x})$، که در آن $c_i^{-1} = \int f(\mathbf{x}_{-i})d\mathbf{x}_{-i}$. چنانچه در این الگوریتم $f(x_i\ |\ \mathbf{x}_{-i})$ برای برخی از $i$ ها نرمال نشده باشد[6] یعنی : $c_i$ مجهول باشد، چون هر سیکل در الگوریتم متشکل از $p$ الگوریتم متروپلیس—هستینگ است، مشکلی برای اجرای الگوریتم نخواهیم داشت.

- برای بهبود عملکرد این الگوریتم، در یک سیکل وقتی که مولفه‌ی اول پیشنهادی $X_1^*$ با داشتن $\mathbf{X}^{(t)}$ تولید شد، مولفه‌ی دوم پیشنهادی $X_2^*$ را از $g_i(\mathbf{x}^*\ |\ x_1^{(t)}, x_2^{(t)}, \ldots, x_p^{(p)})$ به جای $g_i(\mathbf{x}^*\ |\ x_1^{(t+1)}, x_2^{(t)}, \ldots, x_p^{(t)})$ تولید کرده و به همین ترتیب برای سایر مولفه‌ها تا کامل شدن سیکل از مقدار جدید $x_i^{(t+1)}$ به جای $x_i^{(t)}$ استفاده می کنیم.

[6] تابع $f : \Re \to \Re$ تابع چگالی است اگر ۱) $\forall x : f(x) \geq 0$، ۲) $\int_{\Re} f(x)dx = 1$. اگر $f$ در شرط اول صدق کند و بدانیم $\int_{\Re} f(x)dx < \infty$، آنگاه $f$ را یک چگالی نرمال نشده می گوییم.

- زنجیره‌ی مارکفی که با این الگوریتم تولید می شود تحویل ناپذیر، نامتناوب و زمان برگشت پذیر و در نتیجه ارگودیک بوده، بنابراین این الگوریتم نیز یکی از بزارهای روش $MCMC$ است.

مثال : استنباط بیزی

فرض کنیم $\theta= (\theta_1, \ldots, \theta_p)$ پارامترهای یک مدل آماری باشند. همانطور که در بخش الگوریتم متروپلیس—هستینگ اشاره شد، برای استنباط بیزی در مورد $\theta$ با توزیع پیشین $\pi(\theta)$، باید کمیتهای پسین را محاسبه کرد و یک روش مناسب استفاده از روش مونت کارلو با زنجیره‌های مارکف است. دیدیم که اگر $\pi(\theta)$ قابل نمونه گیری باشد، با انتخاب $\pi(\theta)$ به عنوان توزیع پیشنهادی یک زنجیره‌ی مستقل برای روش مونت کارلو با زنجیره‌های مارکف به دست می آید. اما اگر $\pi(\theta)$ قابل نمونه گیری نباشد چه؟

فرض کنیم $\pi(\theta)$ قابل نمونه گیری نبوده اما برای $i = 1, \ldots, p$، $\pi(\theta_i \mid \theta)$ قابل نمونه گیری باشد، از الگوریتم گیبز برای داشتن یک نمونه‌ی تقریبی از $\pi(\theta \mid y)$ استفاده کرد.

مثال : مدل نرمال هایپر پارامتر[۷]

فرض کنیم $y_1, \ldots, y_m$ را از $N(\mu, \frac{1}{\tau})$ مشاهده کرده‌ایم و معتقدیم که $\mu \sim N(0, 1)$ و $\tau \sim G(2, 1)$. در اینصورت :

$$f(\mathbf{y} \mid \mu, \tau) = (2\pi)^{-\frac{m}{2}} \tau^{\frac{m}{2}} \exp(-\frac{\tau}{2} \sum_{j=1}^{m} (y_j - \mu)^2)$$
$$\pi(\tau) = \tau \exp(-\tau) \text{ و } \pi(\mu) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\mu^2}{2})$$

بنابراین :

$$f(\mathbf{y}, \mu, \tau) = (2\pi)^{-\frac{m+1}{2}} \tau^{\frac{m+2}{2}} \exp(-\frac{\tau}{2} \sum_{j=1}^{m} (y_j - \mu)^2 - \frac{\mu^2}{2} - \tau)$$

داریم : $\pi(\mu, \tau \mid \mathbf{y}) \propto f(\mathbf{y}, \mu, \tau)$. برای داشتن نمونه از $\pi(\mu, \tau \mid \mathbf{y})$ با روشهای معمول نمی توان عمل کرد. اما $\pi(\mu \mid \tau, \mathbf{y})$ و $\pi(\tau \mid \mu, \mathbf{y})$ قابل نمونه گیری‌اند :

$$\pi(\mu \mid \tau, \mathbf{y}) = \frac{f(\mathbf{y}, \mu, \tau)}{\pi(\tau, \mathbf{y})} = c f(\mathbf{y}, \mu, \tau) \quad ; \quad c^{-1} = \int \int f(\mathbf{y}, \mu, \tau) d\tau d\mathbf{y}$$

---

[۷] $Hyperparameter Model$ :

در مدلهای بیز معمول، فرض می شود که پارامترهای اصلی مدل متغیر تصادفی‌اند و دارای توزیع احتمال با پارامترهای ثابت می باشند. این پارامترها را اصطلاحاً پارامترهای فرعی مدل می گویند. برای بیان تجربیات و اعتقادات خود در قالب یک توزیع برای پارامترهای اصلی مدل، تعیین پارامترهای فرعی نقش بسیار مهمی دارد. چنانچه این پارامترهای فرعی را نیز متغیر تصادفی با توزیع احتمال مشخصی فرض کنیم، مدل را هایپر پارامتر می گوییم. مدلهای هایپر پارامتر انعطاف بیشتری داشته ولی محاسبات عددی در آنها با پیچیدگیهای زیادی روبرو است. استفاده از روش مونت کارلو با زنجیره‌های مارکف به ویژه الگوریتم گیبز برای انجام محاسبات این مدلها، بسیار رایج است.

یعنی :

$$\pi(\mu \mid \tau, \mathbf{y}) \propto \exp(-\frac{1+m\tau}{2}(\mu - \frac{\tau \sum_{j=1}^{m} y_j}{1+m\tau})^2)$$

و یا $\mu \mid \tau, \mathbf{y} \sim N(\frac{\tau \sum_{j=1}^{m} y_j}{1+m\tau}, \frac{1}{1+m\tau})$. مشا بهاً داریم :

$$\pi(\tau \mid \mu, \mathbf{y}) \propto \tau^{\frac{m}{2}+1} \exp(-\tau(1 + \frac{1}{2}\sum_{j=1}^{m}(y_j - \mu)^2))$$

و یا $\tau \mid \mu, \mathbf{y} \sim G(\frac{m}{2} + 2, 1 + \frac{1}{2}\sum_{j=1}^{m}(y_j - \mu)^2)$.

با انتخاب یک مقدار اولیه‌ی $(\mu^{(0)}, \tau^{(0)})$، اگوریتم گیبز با داشتن $(\mu^{(t)}, \tau^{(t)})$ در سه گام $(\mu^{(t+1)}, \tau^{(t+1)})$ را بصورت زیر تولید می کند :

۱) یکی از دو ترتیب $(\mu, \tau)$ یا $(\tau, \mu)$ را انتخاب می کنیم.

۲) با توجه به ترتیب انتخاب شده در گام اول، $\mu^*$ را از $N(\frac{\tau^{(t)} \sum_{j=1}^{m} y_j}{1+m\tau^{(t)}}, \frac{1}{1+m\tau^{(t)}})$، و $\tau^*$ را از $G(\frac{m}{2} + 2, 1 + \frac{1}{2}\sum_{j=1}^{m}(y_j - \mu^{(t)})^2)$ تولید می کنیم.

۳) قرار می دهیم : $(\mu^{(t+1)}, \tau^{(t+1)}) = (\mu^*, \tau^*)$

مثال : مدل مولفه‌های واریانس

فرض کنیم برای $i = 1, \ldots, k$ و $j = 1, \ldots, m$، $Y_{ij}$ ها مستقل و بوده و $Y_{ij} \sim N(\theta_i, \sigma_e^2)$. برای پارامترهای اصلی مدل، توزیع های پیشین زیر را در نظر می گیریم :

$\theta_1, \ldots, \theta_k \sim^{iid} N(\mu, \sigma_\theta^2)$ و $\sigma_e^2 \sim IG(a_1, b_1)$، که در آن : $a_1, b_1$ ثابت بوده اما $\mu \sim N(\mu_0, \sigma_0^2)$ و $\mu_0 \in \Re$ و $\sigma_\theta^2 \sim IG(a_2, b_2)$. $a_2, b_2, \sigma_0^2 \in \Re^+$ همگی معلوم‌اند.

$$f(\mathbf{y} \mid \theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{km}{2}} \exp(-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{m}(y_{ij} - \theta_i)^2)$$

$$\pi(\theta_1, \ldots, \theta_k \mid \mu, \sigma_\theta^2, \sigma_e^2) = (2\pi\sigma_\theta^2)^{-\frac{k}{2}} \exp(-\frac{1}{2\sigma_\theta^2}\sum_{i=1}^{k}(\theta_i - \mu)^2)$$

$$\pi(\mu \mid \theta_1, \ldots, \theta_k, \sigma_\theta^2, \sigma_e^2) = (2\pi\sigma_0^2)^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2)$$

$$\pi(\sigma_\theta^2 \mid \theta_1, \ldots, \theta_k, \mu, \sigma_e^2) = \frac{b_2^{a_2}}{\Gamma(a_2)}(\sigma_\theta^2)^{-a_2-1} \exp(-\frac{b_2}{\sigma_\theta^2})$$

$$\pi(\sigma_e^2 \mid \theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2) = \frac{b_1^{a_1}}{\Gamma(a_1)}(\sigma_e^2)^{-a_1-1} \exp(-\frac{b_1}{\sigma_e^2})$$

لذا چگالی پیشین و پسین پارامترهای مدل عبارتست از :

$$\pi(\theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \sigma_e^2)$$
$$= (2\pi\sigma_\theta^2)^{-\frac{k}{2}}(2\pi\sigma_0^2)^{-\frac{1}{2}} \frac{b_1^{a_1} B_2^{a_2}}{\Gamma(a_1)\Gamma(a_2)}$$
$$\exp(-\frac{1}{2\sigma_\theta^2}\sum_{i=1}^{k}(\theta_i - \mu)^2 - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 - \frac{b_1}{\sigma_e^2} - \frac{b_2}{\sigma_\theta^2})$$
$$\pi(theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \sigma_e^2 \mid \mathbf{y}) \propto$$
$$f(\mathbf{y} \mid \theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \sigma_e^2)\pi(\theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \sigma_e^2)$$

با توجه به فرم پیچیده‌ی توزیع پسین، برای محاسبات عددی استنباط بیز نمی توان از آن با روشهای معمول نمونه‌ای استخراج کرد. اما محاسبات ساده نشان می دهد که :

$$\theta_i \mid \mu, \sigma_\theta^2, \sigma_e^2, \mathbf{y} \sim N(\frac{m\sigma_\theta^2 \bar{y}_{i.} + \sigma_e^2 \mu}{m\sigma_\theta^2 + \sigma_e^2}, \frac{\sigma_\theta^2 \sigma_e^2}{m\sigma_\theta^2 + \sigma_e^2})$$
$$\mu \mid \theta_1, \ldots, \theta_k, \sigma_\theta^2, \sigma_e^2, \mathbf{y} \sim N(\frac{\sigma_\theta^2 \mu_0 + \sigma_0^2 \sum_{i=1}^k \theta_i}{\sigma_\theta^2 + k\sigma_0^2}, \frac{\sigma_\theta^2 \sigma_0^2}{\sigma_\theta^2 + k\sigma_0^2})$$
$$\sigma_\theta^2 \mid \theta_1, \ldots, \theta_k, \mu, \sigma_e^2, \mathbf{y} \sim IG(a_2 + \frac{k}{2}, b_2 + \frac{1}{2}\sum_{i=1}^k (\theta_i - \mu)^2)$$
$$\sigma_e^2 \mid \theta_1, \ldots, \theta_k, \mu, \sigma_\theta^2, \mathbf{y} \sim IG(a_1 + \frac{k}{2}, b_1 + \frac{1}{2}\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \theta_i)^2)$$

وچون می توان از توزیعهای نرمال و معکوس گاما نمونه گرفت، لذا اگوریتم گیبز را برای تولید یک نمونه‌ی تقریبی از توزیع پسین بکار می بریم. با انتخاب مقدار اولیه‌ی $(\theta_1^{(0)}, \ldots, \theta_k^{(0)}, \mu^{(0)}, (\sigma_\theta^2)^{(0)}, (\sigma_e^2)^{(0)})$، الگریتم گیبز با داشتن $(\theta_1^{(t)}, \ldots, \theta_k^{(t)}, \mu^{(t)}, (\sigma_\theta^2)^{(t)}, (\sigma_e^2)^{(t)})$، $(\theta_1^{(t+1)}, \ldots, \theta_k^{(t+1)}, \mu^{(t+1)}, (\sigma_\theta^2)^{(t+1)}, (\sigma_e^2)^{(t+1)})$ را بصورت زیر تولید می کند :

۱) یک ترتیب از مولفه‌های $(\theta_1^{(t)}, \ldots, \theta_k^{(t)}, \mu^{(t)}, (\sigma_\theta^2)^{(t)}, (\sigma_e^2)^{(t)})$ انتخاب می کنیم.

۲) با توجه به ترتیب انتخاب شده در گام یک، $\theta_i^*$ را از توزیع

$$N(\frac{m(\sigma_\theta^2)^{(t)} \bar{y}_{i.} + (\sigma_e^2)^{(t)} \mu^{(t)}}{m(\sigma_\theta^2)^{(t)} + (\sigma_e^2)^{(t)}}, \frac{(\sigma_\theta^2)^{(t)} (\sigma_e^2)^{(t)}}{m(\sigma_\theta^2)^{(t)} + (\sigma_e^2)^{(t)}})،$$

$\mu^*$ را از توزیع

$$N(\frac{(\sigma_\theta^2)^{(t)} \mu_0 + \sigma_0^2 \sum_{i=1}^k \theta_i^{(t)}}{(\sigma_\theta^2)^{(t)} + k\sigma_0^2}, \frac{(\sigma_\theta^2)^{(t)} \sigma_0^2}{(\sigma_\theta^2)^{(t)} + k\sigma_0^2})،$$

$(\sigma_\theta^2)^*$ را از توزیع

$$IG(a_2 + \frac{k}{2}, b_2 + \frac{1}{2}\sum_{i=1}^k (\theta_i^{(t)} - \mu^{(t)})^2)،$$

و $(\sigma_e^2)^*$ را از توزیع

$$IG(a_1 + \frac{k}{2}, b_1 + \frac{1}{2}\sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \theta_i^{(t)})^2)$$

تولید می کنیم.

۳) قرار می دهیم :

$$(\theta_1^{(t+1)}, \ldots, \theta_k^{(t+1)}, \mu^{(t+1)}, (\sigma_\theta^2)^{(t+1)}, (\sigma_e^2)^{(t+1)}) = (\theta_1^*, \ldots, \theta_k^*, \mu^*, (\sigma_\theta^2)^*, (\sigma_e^2)^*)$$

# « مـــــــــــنابع »


[۱] $Computationalstatistics; G.Givens, J.Hoeting; Wiley(۲۰۰۵)$

[۲] $MCMCMethodology; BraniVidakovic; Lecture$

[۳] $NumericalMethodsOfStatistics; J.F.Monahan; CambridgeUniversityPress(۲۰۰۱)$

[۴] $Simulation; S.Ross; ۳rdEdition; (۱۹۹۹)$

[۵] $AShortCourseOnComputaionalStatistics; ShojaedinChenouri;$
$ShahidBeheshtiUniversity(Spring۲۰۰۵)$

[۶] $GeneralStateSpaceMarkovChainsAndMCMCalgorithms;$
$G.O.Roberts, J.S.Rosenthal; ProbabilitySurveys(۲۰۰۴)$

[۷] نخستین درس در فرآیندهای تصادفی؛ ترجمه‌ی : علی اکبر عالم زاده، عین اله پاشا.

[۸] روش مونت کارلو؛ ترجمه‌ی : یداله دُرج.