به نام خداوند بخشنده مهربان

روش‌های آماری و اقتصادسنجی

در تحلیل و مدل‌سازی داده‌های حمل و نقلی

محمد مهدی بشارتی

besharati@iut.ac.ir

مقدمه‌ای بر اقتصادسنجی بیزین

فصل سیزدهم

Introduction to Bayesian Econometrics

- مقدمه
- برآوردگرهای بیزین
- برآورد نقطه‌ای و فاصله‌ای بیزین
- برآورد ضرایب رگرسیون با روش بیزین

برآورد رگرسیون به روش‌های کلاسیک؛

- OLS
- ML

- ویژگی مهم روش‌های کلاسیک؛
  - استفاده از اطلاعاتی که از نمونه‌گیری به دست آمده است.
  - قضاوت‌های ذهنی (تجربه قبلی) در مورد پدیده مورد مطالعه، هیچ نقشی در برآورد مدل ندارد.

❑ مثال؛

- ○ احتمال آمدن باران
- ○ براساس اطلاعات نمونه، احتمال آمدن باران در یک روز معین را ۴۰٪ برآورد کرده‌ایم.
- ○ اما بدیهی است که راجع به احتمال آمدن باران، صرفاً براساس اطلاعات نمونه قضاوت نمی‌کنیم، بلکه قسمت عمده‌ای از برآورد ما از احتمال آمدن باران، ناشی از قضاوت‌های ذهنی است.
- ○ در واقع، گاهی اوقات یک پارامتر را براساس احتمالات ذهنی (و تجربی) و نه فقط براساس مشاهدات نمونه برآورد می‌کنیم.
- ○ در رویکرد برآورد رگرسیون بیزین، هم از داده‌های نمونه و هم از قضاوت‌های ذهنی (غیر از نمونه) استفاده می‌شود.

- Prior Distribution
- Posterior Distribution

توزیع پیشین و توزیع پسین؛

- «احتمال ذهنی» در مورد یک پارامتر (مثلا میانگین جامعه) بیانگر باور و قضاوت ما در مورد آن پارامتر است.
- در واقع، چون این پارامتر مجهول است، ما آن را براساس قضاوت‌های ذهنی، برآورد می‌کنیم.
- یعنی قبل از آنکه وارد فرآیند نمونه‌گیری (و جمع‌آوری اطلاعات) شویم، راجع به پارامتر موردنظر ($\boldsymbol{\theta}$) دارای یک ذهنیت هستیم (با ذهن خالی وارد قضاوت نمی‌شویم).
- بنابراین، در این حالت، $\boldsymbol{\theta}$ یک پارامتر ثابت (و مجهول) نیست، که براساس داده‌های نمونه، آن را برآورد کنیم.
- بلکه، $\boldsymbol{\theta}$ مانند یک متغیر تصادفی خواهد بود که دارای توزیع معینی است و به آن، «توزیع پیشین» می‌گوییم.
- این توزیع پیشین براساس قضاوت‌های ذهنی شکل می‌گیرد.

مثال–۱:

- می‌خواهیم در مورد وضعیت هوا (بارانی/آفتابی) یک تحلیل آماری انجام دهیم.
- وضعیت هوا را با $\boldsymbol{\theta}$ نمایش می‌دهیم ($\boldsymbol{\theta_1}$ بارانی و $\boldsymbol{\theta_2}$ آفتابی)
- اگر فارغ از داده‌های نمونه، احتمال هوای بارانی را ۴۰٪ و احتمال هوای آفتابی را ۶۰٪ بدانیم، توزیع ذهنی یا توزیع پیشین پارامتر $\boldsymbol{\theta}$ (یعنی $\boldsymbol{P(\theta)}$) به صورت روبرو است:

$\boldsymbol{\theta}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$
$\boldsymbol{P(\theta)}$	0.4	0.6

مثال–۲:

- فرض کنید در پرتاب یک سکه، احتمال آمدن شیر برابر با $\boldsymbol{\theta}$ است.
- براساس قضاوت‌های ذهنی خود، حدس می‌زنیم توزیع پیشین $\boldsymbol{\theta}$ به صورت توزیع یکنواخت است:

$$\boldsymbol{P(\theta) = 1 \quad ; \quad 0 < \theta < 1}$$

- که دارای میانگین $\mathrm{E}(\boldsymbol{\theta}) = \frac{1}{2}$ می‌باشد.

❑ مثال–۳:

- می‌خواهیم در مورد میانگین نمرات درس آمار (μ) قضاوت کنیم.
- حدس قبلی یا ذهنی ما این است که پارامتر (μ) دارای توزیع نرمال با میانگین μ_0 و واریانس σ_0^2 است.
- بدین ترتیب، «توزیع پیشین» μ عبارتست از:

$$f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \quad ; \; -\infty < \mu < +\infty$$

مقدمه – برخی مفاهیم رویکرد بیزین

- **خلاصه:**
  - براساس قضاوت‌های ذهنی برای پارامتر θ یک توزیع در نظر گرفته می‌شود.
  - اما براساس داده‌های نمونه (مشاهدات تجربی) برای پارامتر θ یک برآوردگر تعریف می‌شود.

- در فصول قبل، برآوردگرهای کلاسیک را بررسی کردیم.

- **در روش کلاسیک،**
  - فرض می‌شود θ یک عدد ثابت است، که می‌خواهیم آن را برآورد کنیم.
  - برآوردگری مانند Y معرفی می‌شود که Y یک متغیر تصادفی با توزیع $f(Y|\theta)$ است و θ مقدار ثابتی است.
  - Y فقط بر مبنای داده‌های نمونه تعریف می‌شود؛ و بنابراین، $f(Y|\theta)$ فقط از اطلاعات موجود در نمونه استفاده می‌کند (قضاوت‌های ذهنی هیچ نقشی ندارند).

❑ مثال–۴– برآوردگر میانگین توزیع نرمال _ به روش کلاسیک؛

- برای برآورد میانگین توزیع نرمال (μ)، میانگین نمونه $(\bar{X})$ به عنوان برآوردگر معرفی می‌شود.
- $\bar{X}$ دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.
- توزیع $\bar{X}$ با فرض ثابت بودن μ به صورت روبرو است:

$$f(\bar{X}|\mu) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} e^{-\frac{(\bar{X}-\mu)^2}{2\sigma^2/n}} \quad ; -\infty < \bar{x} < +\infty$$

- توجه: در اینجا μ مجهول است و نه تصادفی

مقدمه – برخی مفاهیم رویکرد بیزین

توزیع پسین؛

- توزیع پسین از ترکیب قضاوت‌های ذهنی (توزیع پیشین) و داده‌های نمونه (توزیع برآوردگر) به دست می‌آید.
- بنابراین، $P(\theta)$ که توزیع پیشین است، با توجه به اطلاعات موجود در Y (که برآوردگر θ براساس داده‌های نمونه است) مورد تجدیدنظر قرار می‌گیرد و توزیع جدیدی برای θ به دست می‌آید.
- این توزیع جدید که با $P(\theta|Y)$ نشان داده می‌شود، موسوم به «توزیع پسین» است.

حالت ناپیوسته

$$P(\theta|Y) = \frac{P(\theta, Y)}{P(Y)} = \frac{\mathrm{P}(\theta)P(Y|\theta)}{P(Y)}$$

حالت پیوسته

$$f(\theta|Y) = \frac{f(\theta, Y)}{f(Y)} = \frac{f(\theta)f(Y|\theta)}{f(Y)}$$

مقدمه – برخی مفاهیم رویکرد بیزین

❑ مثال–۵:

- ○ می‌خواهیم در مورد وضعیت هوا (بارانی/آفتابی) یک تحلیل آماری انجام دهیم.
- ○ وضعیت هوا را با $\boldsymbol{\theta}$ نمایش می‌دهیم ($\boldsymbol{\theta_1}$ بارانی و $\boldsymbol{\theta_2}$ آفتابی)
- ○ توزیع پیشین پارامتر $\boldsymbol{\theta}$ (یعنی $\boldsymbol{P(\theta)}$) به صورت روبرو است:

$\boldsymbol{\theta}$	(بارانی) $\boldsymbol{\theta_1}$	(آفتابی) $\boldsymbol{\theta_2}$
$\boldsymbol{P(\theta)}$	$\mathrm{P}(\boldsymbol{\theta_1}) = 0.4$	$\mathrm{P}(\boldsymbol{\theta_2}) = 0.6$

- ○ داده‌های جدیدی از سازمان هواشناسی به دست آمده است
- ○ و سازمان هواشناسی براساس داده‌های جدید، فردا را بارانی پیش‌بینی کرده است.
- ○ سازمان هواشناسی هوای بارانی را با احتمال ۰.۹ و هوای آفتابی را با احتمال ۰.۸ درست پیش‌بینی می‌کند.
- ○ سوال: احتمال‌های پسین برای هوای بارانی و هوای آفتابی چقدر است؟

مقدمه – برخی مفاهیم رویکرد بیزین

$\boldsymbol{\theta}$	(بارانی) $\boldsymbol{\theta}_1$	(آفتابی) $\boldsymbol{\theta}_2$
$\boldsymbol{P(\theta)}$	$\mathrm{P}(\boldsymbol{\theta}_1) = 0.4$	$\mathrm{P}(\boldsymbol{\theta}_2) = 0.6$

- **مثال–۵؛**

- احتمال‌های (بارانی/آفتابی بودن) براساس داده‌های نمونه را محاسبه می‌کنیم:

$\boldsymbol{\theta}$ / پیش‌بینی وضع هوا (Y)	(بارانی) $\boldsymbol{\theta}_1$	(آفتابی) $\boldsymbol{\theta}_2$
Y_1 (بارانی)	$P(Y_1\|\boldsymbol{\theta}_1) = 0.9$	$P(Y_1\|\boldsymbol{\theta}_2) = 0.2$
Y_2 (آفتابی)	$P(Y_2\|\boldsymbol{\theta}_1) = 0.1$	$P(Y_2\|\boldsymbol{\theta}_2) = 0.8$
جمع	1	1

- سازمان هواشناسی هوای بارانی را با احتمال ۰.۹ و هوای آفتابی را با احتمال ۰.۸ درست پیش‌بینی می‌کند.
- با توجه به داده‌های به دست آمده (یعنی Y)از سازمان هواشناسی، می‌توانیم توزیع احتمالات پیشین را بروزرسانی کرده و توزیع احتمالات پسین را به دست آوریم.

مقدمه – برخی مفاهیم رویکرد بیزین

❑ مثال–۵؛

$\boldsymbol{\theta}$	(بارانی) $\boldsymbol{\theta_1}$	(آفتابی) $\boldsymbol{\theta_2}$
$\boldsymbol{P(\theta)}$	$\mathrm{P}(\boldsymbol{\theta_1}) = 0.4$	$\mathrm{P}(\boldsymbol{\theta_2}) = 0.6$

○ احتمال‌های (بارانی/آفتابی بودن) براساس داده‌های نمونه را محاسبه می‌کنیم:

$\boldsymbol{\theta}$ / پیش‌بینی وضع هوا (Y)	(بارانی) $\boldsymbol{\theta_1}$	(آفتابی) $\boldsymbol{\theta_2}$
Y_1 (بارانی)	$P(Y_1\|\boldsymbol{\theta_1}) = 0.9$	$P(Y_1\|\boldsymbol{\theta_2}) = 0.2$
Y_2 (آفتابی)	$P(Y_2\|\boldsymbol{\theta_1}) = 0.1$	$P(Y_2\|\boldsymbol{\theta_2}) = 0.8$
جمع	1	1

$$P(\theta_1|Y_1) = \frac{P(Y_1 \cap \theta_1)}{P(Y_1)} = \frac{0.36}{0.48} = 0.75$$

$$P(\theta_2|Y_1) = \frac{P(Y_1 \cap \theta_2)}{P(Y_1)} = \frac{0.12}{0.48} = 0.25$$

$$P(Y_1 \cap \theta_1) = P(\theta_1)P(Y_1|\boldsymbol{\theta_1}) = 0.4 * 0.9 = 0.36$$

$$P(Y_1 \cap \theta_2) = P(\theta_2)P(Y_1|\boldsymbol{\theta_2}) = 0.6 * 0.2 = 0.12$$

$$P(Y_1) = P(Y_1 \cap \theta_1) + P(Y_1 \cap \theta_2) = 0.36 + 0.12 = 0.48$$

$\boldsymbol{\theta}$	(بارانی) $\boldsymbol{\theta_1}$	(آفتابی) $\boldsymbol{\theta_2}$
$\boldsymbol{P(\theta)}$	$\mathrm{P}(\boldsymbol{\theta_1}) = 0.4$	$\mathrm{P}(\boldsymbol{\theta_2}) = 0.6$

- **مثال–۵:**
  - توزیع پسین احتمال‌های (بارانی/آفتابی بودن) عبارتست از:

$\boldsymbol{\theta}$	(بارانی) $\boldsymbol{\theta_1}$	(آفتابی) $\boldsymbol{\theta_2}$
$\boldsymbol{P(\theta \mid Y_1)}$	$P(\boldsymbol{\theta_1} \mid \boldsymbol{Y_1}) = 0.75$	$P(\boldsymbol{\theta}_2 \mid \boldsymbol{Y_1}) = 0.25$

مثال (از فصل احتمالات):

پلیس راهور در قطعه‌ای از یک راه، راننده‌ای را به جرم تخطئ از سرعت مجاز متوقف کرده و از او تست الکل می‌گیرد. اگر نتیجه تست مثبت بوده باشد، با توجه به مفرضات زیر می‌خواهیم بدانیم احتمال اینکه او واقعا مست بوده، چقدر است؟ فرض کنید احتمال اینکه دستگاه به اشتباه نتیجه آزمایش را مثبت نشان دهد ۵٪ و احتمال اینکه به اشتباه نتیجه آزمایش را منفی نشان دهد۱۰٪ می‌باشد. همچنین، فرض کنید از قبل می‌دانیم که به‌طور متوسط ۸٪ از رانندگانی که از این قطعه عبور می‌کنند مست هستند.

پیشامدهای زیر را درنظر می‌گیریم:

A_1: رانندگی درحال مستی

A_2: رانندگی درحال سلامت

B: مثبت بودن نتیجه آزمایش

مقدمه – برخی مفاهیم رویکرد بیزین

مثال–۶؛

- فرض کنید می‌خواهیم پارامتر **p** در توزیع دونقطه‌ای (تابع احتمال برنولی) را برآورد کنیم.

$$P(X=x)=p^x(1-p)^{1-x} \quad ; \quad X=0,1$$

- همچنین، فرض کنید توزیع پیشین **p** به صورت توزیع یکنواخت باشد

$$f(P)=1 \quad ; \quad 0<P<1$$

- اکنون برای کسب برآورد دقیق‌تری از پارامتر p، نمونه‌ای به حجم **n** برداشت می‌شود.
- مطلوبست توزیع پسین p ؟

❑ مثال–۶:

- اگر برآوردگر p را به صورت $\bar{X}=\frac{Y}{n}$ تعریف کنیم (که $Y=X_1+X_2+\ldots+X_n$ باشد) آنگاه،
- توزیع $\bar{X}$ با فرض ثابت بودن p به صورت دوجمله‌ای خواهد بود.

$$P(\bar{X}|p)=P(Y|p)=C_n^Y p^Y(1-p)^{n-Y} \quad ; \quad Y=0,1,\ldots,n$$

- که $C_n^Y=\frac{n!}{Y!(n-Y)!}$
- از طرف دیگر توزیع مشترک Y و p عبارتست از:

$$f(p,Y)=f(p)\,P(Y|p)=1*C_n^Y\,p^Y\,(1-p)^{n-Y}=C_n^Y\,p^Y\,(1-p)^{n-Y}$$

- بنابراین، $f(Y)$ برابر است با:

$$f(Y)=\int_0^1 f(p,Y)\,dp=\cdots=\frac{1}{n+1}$$

- **مثال–۶؛**

  - اکنون توزیع پسین را به صورت زیر می‌نویسیم؛

$$P(p|Y)=\frac{P(p,Y)}{P(Y)}=(n+1)\,C_n^Y\,p^Y\,(1-p)^{n-Y}\quad ;\quad 0<P<1$$

مقدمه‌ای بر اقتصادسنجی بیزین

فصل سیزدهم

Introduction to Bayesian Econometrics

- مقدمه
- برآوردگرهای بیزین
- برآورد نقطه‌ای و فاصله‌ای بیزین
- برآورد ضرایب رگرسیون با روش بیزین

برآوردگرهای بیزین

برآورد پارامترها در روش کلاسیک

- روش کلاسیک (مثلا ML) برای برآورد پارامترها از تابع درستنمایی استفاده می‌کند به گونه‌ای که پارامتر ثابت و مجهول θ (ولی نه تصادفی) با مشتق‌گیری از تابع درستنمایی به دست می‌آید.
- این تابع صرفاً براساس داده‌های نمونه به دست می‌آید.
- بنابراین، برآوردگر آن نیز فقط تابعی از متغیرهای نمونه است (و سایر اطلاعات مانند قضاوت‌های ذهنی در آن دخالتی ندارد)
- برای مثال، برای برآورد میانگین جامعه (μ)، برآوردگر ($\bar{X}$) به دست می‌آید.

برآوردگرهای بیزین

❑ تکرار مثال–۴– برآوردگر میانگین توزیع نرمال _ به روش کلاسیک؛

- برای برآورد میانگین توزیع نرمال (μ)، میانگین نمونه ($\bar{X}$) به عنوان برآوردگر معرفی می‌شود.
- $\bar{X}$ دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.
- توزیع $\bar{X}$ با فرض ثابت بودن μ به صورت روبرو است:

$$f(\bar{X}|\mu) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} e^{-\frac{(\bar{X}-\mu)^2}{2\sigma^2/n}} \quad ; -\infty < \bar{x} < +\infty$$

- توجه: در اینجا μ مجهول است و نه تصادفی

برآوردگرهای بیزین

○ برآورد پارامترها در روش بیزین

- ○ در روش بیزین برای برآورد پارامتر θ ، به جای تابع درستنمایی، از توزیع پسین استفاده می‌شود.
- ○ اگر در اینجا نیز بخواهیم از روش حداکثر درستنمایی استفاده کنیم، بایستی از توزیع پسین یعنی $f(\theta|Y)$ نسبت به θ مشتق بگیریم.
- ○ بنابراین، در روش بیزین، برآوردگر θ هم تابعی از داده‌های نمونه، و هم تابعی از قضاوت‌های ذهنی (اطلاعات غیرنمونه‌ای) است.
- ○ برآوردگر بیزین را به صورت $\tilde{\boldsymbol{\theta}}$ نمایش می‌دهیم.

$$\tilde{\theta} = g(\theta_{.}\ ,\ Y)$$

- ○ $\theta_{.}$ حدس قبلی در مورد θ و $\boldsymbol{Y}$ اطلاعات نمونه است.

مثال–۷– برآوردگر میانگین توزیع نرمال ـ به روش بیزین؛

- در مثال–۳، برای برآورد میانگین جامعه (μ)، توزیع پیشین زیر را به دست آوردیم؛

$$f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \quad ; \ -\infty < \mu < +\infty$$

- از طرف دیگر، بر مبنای داده‌های نمونه، توزیع برآوردگر μ یعنی($\bar{X}$) با فرض معلوم بودن σ^2 به صورت روبرو است (مثال–۴):

$$f(\bar{X}|\mu) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} e^{-\frac{(\bar{X}-\mu)^2}{2\sigma^2/n}} \quad ; -\infty < \bar{x} < +\infty$$

- براساس توزیع پیشین μ و توزیع $\bar{X}$ می‌توان توزیع پسین μ را به دست آورد:

$$f(\mu|\bar{X}) = \frac{f(\mu,\bar{X})}{f(\bar{X})} = \frac{f(\mu)f(\bar{X}|\mu)}{f(\bar{X})}$$

❑ مثال–۷– برآوردگر میانگین توزیع نرمال ـ به روش بیزین؛

- در نهایت، توزیع پسین μ به صورت زیر به دست می‌آید (توزیع نرمال با میانگین μ_1 و واریانس σ_1^2)؛

$$f(\mu|\bar{X}) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}}$$

- امیدریاضی توزیع پسین μ برابر با μ_1 است؛

$$\mu_1 = \frac{n\bar{X}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

- واریانس توزیع پسین μ برابر با σ_1^2 است؛

$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

- بنابراین، امیدریاضی توزیع پسین μ برابر با μ_1 است.
- پس می‌توان از μ_1 به عنوان برآوردگر بیزین μ استفاده کرد.

❑ مثال–۷– برآوردگر میانگین توزیع نرمال ـ به روش بیزین؛

- اکنون می‌توان μ_1 را به صورت روبرو بازنویسی کرد؛

$$\mu_1 = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\bar{X} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0$$

$$w = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \qquad \mu_1 = w\bar{X} + (1 - w)\mu_0$$

- بنابراین، برآورد بیزین پارامتر μ برابر با متوسط وزنیِ دو برآورد زیر است؛
  - برآورد به دست آمده از داده‌های نمونه $(\bar{X})$
  - برآورد به دست آمده از حدس‌های قبلی (μ_0)

- در برآورد به دست آمده از روش کلاسیک، $w = 1$ است یعنی فقط از $\bar{X}$ استفاده می‌شود.

برآوردگرهای بیزین

❑ مثال–۷– برآوردگر میانگین توزیع نرمال _ به روش بیزین؛

- توجه: w را می‌توان به صورت زیر نیز نوشت:

$$w = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \quad \rightarrow \quad w = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}}$$

- بدین ترتیب، w بستگی به **واریانس توزیع پیشین** (یعنی σ_0^2) و **واریانس توزیع $\bar{X}$** (یعنی $\frac{\sigma^2}{n}$) دارد.
- هرچه پراکندگی مشاهدات نمونه بیشتر باشد (یعنی $\frac{\sigma^2}{n}$ بزرگتر باشد)، سهم $\bar{X}$ از برآوردگر $\boldsymbol{\mu}$ کمتر است.
- به عبارت دیگر، سهم $\bar{X}$ از برآوردگر $\boldsymbol{\mu}$ با واریانس توزیع $\bar{X}$ رابطه معکوس دارد.
- سهم μ_0 (حدس‌های قبلی) از برآوردگر $\boldsymbol{\mu}$ نیز با پراکندگی این حدس‌ها (σ_0^2) رابطه معکوس دارد.

❑ مثال–۸:

- فرض کنید که در مورد میانگین نمرات درس آمار حدس می‌زنیم که میانگین نمرات برابر با $\mu_0 = 12$ با واریانس $\sigma_0^2 = 9$ باشد.
- نمونه‌ای (n=16) از این دانشجویان انتخاب می‌شود. با فرض اینکه پراکندگی نمرات در جامعه آماری $\boldsymbol{\sigma = 8}$ باشد و میانگین نمونه برابر با $\bar{X} = 14$ به دست آید.
- مطلوبست برآورد بیزین از میانگین جامعه؟

$$\mu_1 = w\bar{X} + (1-w)\mu_0$$

$$w = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} = \frac{9}{9 + \frac{64}{16}} = 0.692$$

$$\mu_1 = w\bar{X} + (1-w)\mu_0$$
$$= (0.692 * 14) + (1 - 0.692) * 12 = 13.85$$

- برآورد بیزین از میانگین جامعه= **13.85**
- برآورد کلاسیک از میانگین جامعه= $\bar{X} = 14$

❑ مثال–۹: برآورد پارامتر توزیع دونقطه‌ای (p) با استفاده از توزیع نمونه‌ای $\bar{X}$؛

- می‌خواهیم پارامتر p در توزیع دونقطه‌ای (تابع احتمال برنولی) را به روش کلاسیک برآورد کنیم.
- اگر نمونه‌ای به حجم n از جامعه‌ای با توزیع دونقطه‌ای برداشت شود و $\bar{X} = \frac{Y}{n}$ به عنوان برآوردگر p انتخاب شود؛ این برآوردگر دارای توزیع دوجمله‌ای (به صورت زیر) خواهد بود:

$$P(\bar{X}|p) = P(Y|p) = C_n^Y p^Y (1-p)^{n-Y} \quad ; \quad Y = 0,1,\ldots,n$$

$$E(\bar{X}) = p \quad ; \quad Var(\bar{X}) = \frac{pq}{n}$$

- مثال–۹: برآورد پارامتر توزیع دونقطه‌ای (p) با استفاده از توزیع نمونه‌ای $\bar{X}$؛

  - می‌خواهیم پارامتر p در توزیع دونقطه‌ای (تابع احتمال برنولی) را به روش بیزین برآورد کنیم.
  - فرض کنید توزیع پیشین p به صورت توزیع یکنواخت باشد:

$$f(p) = 1 \quad ; \quad 0 < P < 1$$

  - امیدریاضی توزیع پیشین:

$$E(p) = \frac{1}{2}$$

  - قبلاً توزیع پسین p نیز به صورت زیر به دست آمد:

$$P(p|Y) = (n+1)\, C_n^Y\, p^Y\, (1-p)^{n-Y} \quad ; \quad 0 < P < 1$$

  - امیدریاضی توزیع پسین بیانگر برآورد بیزین برای پارامتر p است:

$$E(p|Y) = \int_0^1 p\, f(p|Y) dp = (n+1)\, C_n^Y \int_0^1 p^{Y+1}\, (1-p)^{n-Y} dp = \frac{Y+1}{n+2}$$

- **مثال–۹: برآورد پارامتر توزیع دونقطه‌ای (p) با استفاده از توزیع نمونه‌ای $\bar{X}$؛**

  - این برآورد بیزین برای پارامتر p را به صورت زیر نیز می‌توان بازنویسی کرد:

$$E(p|Y) = E(p|\bar{X}) = \frac{Y+1}{n+2} = \frac{Y}{n+2} + \frac{1}{n+2}$$

$$= \frac{n}{n+2} \times \frac{Y}{n} + \frac{1}{n+2} \times \frac{2}{2} = \frac{n}{n+2} \times \bar{X} + \frac{2}{n+2} \times \frac{1}{2}$$

$$= w\bar{X} + (1-w)\frac{1}{2} = w\bar{X} + (1-w)E(p)$$

  - بدین ترتیب؛ $w = \frac{n}{n+2}$ بیانگر وزن داده‌های نمونه‌ای ($\bar{X}$) و $1-w$ بیانگر وزن داده‌های ذهنی (یعنی $E(p)$) است.

مقدمه‌ای بر اقتصادسنجی بیزین

فصل
سیزدهم

Introduction to Bayesian Econometrics

- مقدمه
- برآوردگرهای بیزین
- **برآورد نقطه‌ای و فاصله‌ای بیزین**
- برآورد ضرایب رگرسیون با روش بیزین

برآورد بیزین (نقطه‌ای و فاصله‌ای)

❖ **یادآوری؛**

○ **در روش برآورد کلاسیک،**

- پارامتر θ یک مقدار ثابت و مجهول است.
- بدون هیچ ذهنیت قبلی، پارامتر θ را با استفاده از داده‌های نمونه برآورد می‌کنیم.

○ **در روش برآورد بیزین،**

- پارامتر θ یک متغیر تصادفی و مجهول است.
- راجع به پارامتر θ یک ذهنیت قبلی داریم (توزیع پیشین)، که آن اطلاعات را با اطلاعات داده‌های نمونه ترکیب می‌کنیم و از هر دو منبع اطلاعاتی برای برآورد پارامتر θ استفاده می‌کنیم.

برآورد بیزین (نقطه‌ای و فاصله‌ای)

❖ **تفاوت این دو روش (در قالب مثال برآورد میانگین جامعه نرمال (μ) با استفاده از توزیع نمونه‌ای $\bar{X}$)؛**

- فرض کنید نمونه‌ای به حجم n=10 انتخاب شده و براساس داده‌های آن، $\bar{X} = 20$ به دست آمده است.
- فرض کنید واریانس جامعه معلوم بوده و برابر با $\sigma^2 = 10$ است.
- بر این مبنا، برآورد نقطه‌ای و فاصله‌ای در روش کلاسیک به صورت زیر خواهد بود؛

$$\hat{\theta} = \bar{X} = 20$$

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \longrightarrow 20 \pm 1.96$$

برآورد بیزین (نقطه‌ای و فاصله‌ای)

❖ **تفاوت این دو روش (در قالب مثال برآورد میانگین جامعه نرمال (μ) با استفاده از توزیع نمونه‌ای $\bar{X}$)؛**

- فرض کنید نمونه‌ای به حجم n=10 انتخاب شده و براساس داده‌های آن، $\bar{X} = 20$ به دست آمده است.
- فرض کنید واریانس جامعه معلوم بوده و برابر با $\sigma^2 = 10$ است.
- برآورد نقطه‌ای و فاصله‌ای در **روش بیزین** به صورت زیر خواهد بود؛
- فرض کنید توزیع پیشین μ به صورت روبرو باشد $f(\mu) = N(\mu_0 = 25 \ , \sigma_0^2 = 4)$
- فرض کنید توزیع $\bar{X}$ با فرض ثابت بودن μ عبارتست از: $f(\bar{X}|\mu) = N\left(\mu \, , \frac{\sigma^2}{n} = 1\right)$
- قبلاً نشان دادیم که توزیع پسین μ به صورت روبرو خواهد بود، $f(\mu|\bar{X}) = N(\mu_1 \ , \sigma_1^2)$
- رابطه محاسبه μ_1:

$$w = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} \qquad \mu_1 = w\bar{X} + (1-w)\mu_0$$

- رابطه محاسبه σ_1^2:

$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

برآورد بیزین (نقطه‌ای و فاصله‌ای)

❖ **تفاوت این دو روش (در قالب مثال برآورد میانگین جامعه نرمال (μ) با استفاده از توزیع نمونه‌ای $\bar{X}$)؛**

- فرض کنید نمونه‌ای به حجم n=10 انتخاب شده و براساس داده‌های آن، $\bar{X} = 20$ به دست آمده است.
- فرض کنید واریانس جامعه معلوم بوده و برابر با $\sigma^2 = 10$ است.
- برآورد نقطه‌ای و فاصله‌ای در روش بیزین به صورت زیر خواهد بود؛

$$w = \frac{4}{4+1} = 0.8$$

$$\tilde{\mu} = \mu_1 = 0.8 * 20 + 0.2 * 25 = 21$$

$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} = \frac{10}{10} + \frac{1}{4} = 1.25 \qquad \sigma_1^2 = \frac{1}{1.25} = 0.8$$

- **Bayesian Credible Intervals** :

$$\tilde{\mu} = \mu_1 = 21$$

$$\mu_1 \pm Z_{\alpha/2}\frac{\sigma_1}{\sqrt{n}} \longrightarrow 21 \pm 1.96\sqrt{0.8} \longrightarrow 21 \pm 1.753$$

مقدمه‌ای بر اقتصادسنجی بیزین

فصل سیزدهم

Introduction to Bayesian Econometrics

مقدمه

برآوردگرهای بیزین

برآورد نقطه‌ای و فاصله‌ای بیزین

برآورد ضرایب رگرسیون با روش بیزین

برآورد ضرایب رگرسیون با روش بیزین

- **یادآوری؛**
  - **در روش برآورد کلاسیک،**
    - پارامتر θ یک مقدار **ثابت** و مجهول است.
    - برآورد پارامترها از طریق روش‌هایی مانند OLS و ML انجام می‌شود.
  - **در روش برآورد بیزین،**
    - پارامتر θ یک **متغیر تصادفی** و مجهول است.
    - به همین دلیل، در روش بیزین به دنبال یافتن توزیع پسین یعنی $f(\theta|Y)$ هستیم، نه به دست آوردن برآورد نقطه‌ای.
    - سپس از توزیع پسین برای آزمون فرضیه و ارائه فاصله اطمینان استفاده می‌کنیم.

برآورد ضرایب رگرسیون با روش بیزین

❖ **تابع درستنمایی؛**

- **تحلیل بیزین برای رگرسیون را می‌توان براساس روش درستنمایی بررسی نمود.**

- **قبلاً دیدیم که؛**
  - **روش کلاسیک از «تابع درستنمایی» استفاده می‌کند.**
  - **روش بیزین از «تابع پسین» استفاده می‌کند.**

- **پس برای مقایسه رگرسیون‌های کلاسیک و بیزین باید از تابع درستنمایی شروع کنیم.**

برآورد ضرایب رگرسیون با روش بیزین

❖ تابع درستنمایی؛

- یک رگرسیون ساده که فقط یک ضریب β دارد را در نظر بگیرید.

$$y_i = \beta x_i + u_i$$

- دو پارامتر باید برآورد شود که عبارتست از پارامترهای β و σ^2
- با این فرض که u_i دارای توزیع نرمال با میانگین صفر و واریانس σ^2 است؛ y_i نیز از توزیع نرمال با مشخصات روبرو پیروی می‌کند؛

$$y_i|x_i,\beta,\sigma^2 \sim N(\beta x_i\,,\sigma^2)$$

برآورد ضرایب رگرسیون با روش بیزین

❖ ثابت می‌شود که؛

- تابع درستنمایی معادل است با **حاصلضرب** «توزیع نرمال برای پارامتر β» ضربدر «توزیع معکوس گاما برای σ^2 »
- **توجه:** توزیع نرمال برای پارامتر β با فرض معین بودن σ^2 تعریف شده است، پس بیانگر یک تابع چگالی شرطی است (که با $f(\beta|\sigma^2)$ نشان می‌دهیم)
- تابع چگالی توزیع معکوس گاما برای σ^2 را نیز با $f(\sigma^2)$ نشان می‌دهیم.
- پس تابع درستنمایی معادل است با:

$$L(y|\beta,\sigma^2) \propto f(\beta|\sigma^2)f(\sigma^2)$$

- از طرف دیگر، توزیع مشترک پیشین $\boldsymbol{\beta}$ و $\boldsymbol{\sigma^2}$ عبارتست از؛

$$f(\beta,\sigma^2) = f(\beta|\sigma^2)f(\sigma^2)$$

- بنابراین، توزیع مشترک $\boldsymbol{\beta}$ و $\boldsymbol{\sigma^2}$ شامل **حاصلضرب** «توزیع نرمال شرطی برای پارامتر β» و «توزیع معکوس گاما برای σ^2 » است.

برآورد ضرایب رگرسیون با روش بیزین

- **ثابت شد که برای این مدل؛**
  - **تابع درستنمایی (که روش برآورد حداکثر درستنمایی بر پایه آن استوار است) معادل است با؛**
  - **توزیع مشترک «نرمال-معکوس گاما» برای پارامترهای β و σ^2**
  - یعنی تابع درستنمایی معادل است با **حاصلضرب** «توزیع نرمال برای پارامتر β» ضربدر «توزیع معکوس گاما برای σ^2 »

- **به عبارت دیگر، (برای رگرسیون تک متغیره) اگر توزیع‌های پیشین β و σ^2 را به صورت زیر تعریف کنیم؛**
  - **برای β توزیع پیشین نرمال با میانگین $\hat{\beta}$ و واریانس $\frac{\sigma^2}{\sum x_i^2}$**
  - **برای σ^2 توزیع پیشین معکوس گاما با پارامترهای $m = \frac{\nu}{2} + 1$ و $\lambda = \frac{\nu s^2}{2}$**

- **آنگاه توزیع پیشین مشترک β و σ^2 به صورت «نرمال-معکوس گاما» بوده که همان تابع درستنمایی است.**
- **$\nu = n - 2$: درجه آزادی و s^2 انحراف معیار نمونه است.**

برآورد ضرایب رگرسیون با روش بیزین

- ❖ **ثابت شد که برای این مدل؛**
  - ○ **همچنین، برای رگرسیون k-متغیره اگر؛**
    - ➤ **برای بردار $\boldsymbol{\beta}$ توزیع پیشین نرمال شرطی k-متغیره با میانگین $\widehat{\boldsymbol{\beta}}$ و واریانس $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ تعریف کنیم،**
    - ➤ **برای σ^2 توزیع پیشین معکوس گاما با پارامترهای $m = \frac{\nu}{2} + 1$ و $\lambda = \frac{\nu s^2}{2}$ تعریف کنیم،**
  - ○ **آنگاه توزیع پیشین مشترک $\boldsymbol{\beta}$ و $\boldsymbol{\sigma^2}$ به صورت «نرمال-معکوس گاما» بوده که همان تابع درستنمایی است.**
  - ○ **یعنی تابع درستنمایی معادل با حاصلضرب** «توزیع نرمال شرطی k-متغیره برای بردار پارامترهای $\boldsymbol{\beta}$» ضربدر «توزیع معکوس گاما برای σ^2 »

توزیع پسین ضرایب رگرسیون و برآورد بیزین

- در **تحلیل کلاسیک** فرض می‌شود که؛
  - پارامترهای معادله رگرسیون (θ ها)، ثابت هستند.
  - و Y_i چون تابعی از متغیر تصادفی u_i است، تصادفی است.
  - توزیع مشترک Y_i که با $L(\mathbf{y}|\boldsymbol{\beta},\sigma^2)$ نمایش داده می‌شود، معروف به تابع درستنمایی است (در اسلایدهای قبل معرفی شد)
  - این تابع فقط شامل اطلاعات جدید (نمونه) است و اطلاعات قبلی هیچ نقشی در برآورد پارامترها ندارد.

  - برای ادامه، دوباره یک رگرسیون ساده که فقط یک ضریب β دارد را در نظر بگیرید.
  - دو پارامتر باید برآورد شود که عبارتست از پارامترهای β و σ^2

$$y_i = \beta x_i + u_i$$

توزیع پسین ضرایب رگرسیون و برآورد بیزین

❖ در **تحلیل بیزین**؛

- حدس‌های قبلی در مورد دو پارامتر β و σ^2 را به کار می‌گیریم.
- این حدس‌های قبلی (توزیع پیشین) را با توزیع مشترک به صورت $f(\beta,\sigma^2)$ نشان می‌دهیم.
- با اضافه شدن اطلاعات جدید (در قالب مشاهدات نمونه)، حدس‌های قبلی را به روزرسانی می‌کنیم.
- این بروزرسانی را با $f(\beta,\sigma^2|y)$ نشان می‌دهیم که همان تابع چگالی پسین است.
- با داشتن تابع چگالی پسین می‌توان از امیدریاضی پسین به عنوان «برآوردهای بروزشده» استفاده کرد.
- بنابراین، مسأله اصلی، عبارتست از **تعیین تابع چگالی پسین ($f(\beta,\sigma^2|y)$)**.

توزیع پسین ضرایب رگرسیون و برآورد بیزین

❖ **تعیین تابع چگالی پسین؛**

- **فرض کنید توزیع مشترک پارامترها و مشاهدات به صورت $f(y,\beta,\sigma^2)$ باشد.**
- **در این صورت، تابع چگالی پسین پارامترها عبارتست از؛**

$$f(\beta,\sigma^2|y)=\frac{f(y,\beta,\sigma^2)}{f(y)}=\frac{f(y|\beta,\sigma^2)f(\beta,\sigma^2)}{f(y)}=\frac{(\text{توزیع پیشین})(\text{تابع درستنمایی})}{f(y)}$$

- **چون $f(y|\beta,\sigma^2)$ تابع چگالی مشترک y_i ها است، پس همان تابع درستنمایی است (اسلایدهای قبل).**
- **$f(y)$ شامل پارامترها نیست، پس کنار گذاشتن آن هیچ تاثیری در تعیین و برآورد پارامترها ندارد.**
- **پس تابع چگالی پسین پارامترها معادل است با؛**

$$f(\beta,\sigma^2|y)\ \propto\ L(y|\beta,\sigma^2)f(\beta,\sigma^2)\ \propto\ (\text{توزیع پیشین})(\text{تابع درستنمایی})$$

- **$L(y|\beta,\sigma^2)$: تابع چگالی مشترک y_i ها است، پس همان تابع درستنمایی (اطلاعات جدید)**
- **$f(\beta,\sigma^2)$: توابع چگالی پشین پارامترها (اطلاعات قبلی)**

توزیع پسین ضرایب رگرسیون و برآورد بیزین

❖ **تعیین تابع چگالی پسین؛**

- اگر از رابطه زیر استفاده کنیم (توابع چگالی توزیع پیشین پارامترها)؛

$$f(\beta,\sigma^2)=f(\beta|\sigma^2)f(\sigma^2)$$

- تابع چگالی پسین پارامترها **معادل است با**؛

$f(\beta,\sigma^2|y)\propto L(y|\beta,\sigma^2)f(\beta|\sigma^2)f(\sigma^2)\propto$ (تابع درستنمایی)(توزیع پیشین بتاها) (توزیع پیشین واریانس)

- $f(\beta|\sigma^2)$ تابع چگالی پیشین β با فرض معلوم بودن σ^2
- $f(\sigma^2)$ تابع چگالی پیشین σ^2

- **خلاصه؛**
- روش کلاسیک تنها از اطلاعات جدید استفاده می‌کند: $L(y|\beta,\sigma^2)$
- روش بیزین از حاصلضرب اطلاعات جدید و اطلاعات قدیمی استفاده می‌کند: $L(y|\beta,\sigma^2)f(\beta,\sigma^2)$
- پس تفاوت روش کلاسیک و بیزین، در استفاده از حدس‌ها و اطلاعات قبلی ($f(\beta,\sigma^2)$) است.

توزیع‌های پیشین ضرایب رگرسیون

- گفتیم، تفاوت روش کلاسیک و بیزین، در استفاده از حدس‌ها و اطلاعات قبلی ($f(\beta,\sigma^2)$) است.

- سوال: چه تابعی را برای **توزیع پیشین** تعریف کنیم؟
  - ✓ اینکه چه تابعی را برای **توزیع پیشین** تعریف کنیم، می‌تواند اختیاری باشد.

- انواع توابع **توزیع پیشین**،
  - ◁ توزیع‌های پیشین با اطلاعات غیرمفید
  - ◁ توزیع‌های پیشین با اطلاعات مفید

توزیع‌های پیشین با اطلاعات غیرمفید

- اگر بخواهیم از روش بیزین استفاده کنیم، و از سوی دیگر، هیچ اطلاعات قبلی نداشته باشیم، آنگاه،
  - ✓ یک حدس قبلی داریم که حاوی هیچ اطلاعات مفیدی نیست.
- باید راهی برای اظهار فقدان اطلاعات درمورد پارامترها پیدا کنیم.

- در این شرایط می‌توانیم توزیع پیشین پارامترها را ثابت در نظر بگیریم. این بدان معناست که،
  - احتمال هر یک از مقادیر پارامترها در دامنه موردنظر، یکسان است ($\beta \sim U(-\infty, +\infty)$).

- نوع دیگری از توزیع پیشین با اطلاعات غیرمفید، توزیع نرمال با میانگین صفر و واریانس بسیار بزرگ است ($\beta \sim N(0,10000)$).

توزیع پیشین با اطلاعات غیرمفید:

➢ noninformative or vague, diffuse , flat

مکانیسم MCMC برای برآورد مدل‌های بیزین

- به صورت خلاصه، مدل‌های بیزین در واقع تجمیع نظریه بیز با مدل‌های آماری کلاسیک است.
- هر مدل آماری کلاسیکی را می‌توان از طریق یک مدل معادل بیزین برآورد نمود.
- رواج استفاده از مدل‌های بیزین نه صرفاً بابت مزایای نظریه بیز، بلکه بیشتر به دلیل در دسترس بودن مدل‌های بیزین از طریق روش‌های زنجیره مارکوفِ مونت‌کارلو (MCMC) است.
- MCMC یک رویکرد مبتنی بر نمونه‌گیری برای برآورد پارامترهای مدل است که به خوبی برای مدل‌های بیزین قابل استفاده است.
- علاوه بر این، روش‌های MCMC، برآورد مدل‌های با شکل‌های تابعی پیچیده را امکان‌پذیر می‌کنند.
- ((شکل‌های تابعی که برآورد آن‌ها با روش حداکثر درستنمایی غالباً بسیار دشوار است))
- به طور خلاصه، MCMC تکنیکی برای نمونه‌برداری از یک توزیع پیچیده است.
- از سوی دیگر، اکثر کاربردهای مدل بیزین در مطالعات حوزه حمل‌ونقل، شامل «توزیع‌های پیشین غیرمفید» و البته «تابعی با شکل پیچیده» است.
- در ادامه، ابتدا به توضیح مختصر روش MCMC پرداخته و سپس چارچوب استفاده از الگوریتم‌های MCMC در برآورد مدل‌های بیزین را بررسی می‌کنیم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

- **معرفی روش MCMC**
- درصورتیکه یک مدل احتمالاتی دارای ویژگی‌های زیر باشد نمی‌توان به راحتی از آن نمونه‌های تصادفی استخراج کرد؛
  - تابع توزیع مدل احتمالاتی پیچیده باشد (به خصوص دارای احتمالات شرطی باشد)
  - یا تابع توزیع دارای ابعاد بالا باشد (چند متغیره) باشد.
- در اکثر مطالعات پژوهشی چنین چالش‌هایی وجود دارد.
- روش MCMC به پژوهشگر کمک می‌کند که از توابع توزیع احتمالاتی که چنین ویژگی‌هایی دارند، نمونه‌های تصادفی تهیه کنند.
- بنابراین، به طور خلاصه،
  - MCMC تکنیکی برای نمونه‌برداری از یک توزیع پیچیده است.
  - به خصوص برای کار با داده‌های با ابعاد زیاد، مفید است.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **معرفی روش MCMC**

❖ **روش MCMC در واقع از ترکیب دو روش «مونت‌کارلو» و «زنجیره مارکوف» تشکیل شده است.**

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش نمونه‌گیری «مونت‌کارلو»**

❖ خیلی از مدل‌ها وجود دارند که ما نمی‌توانیم آن‌ها را به صورت مستقیم (صریح) حل کنیم.

❖ Closed-form solution ندارند.

❖ در این شرایط برای حل چنین مدل‌هایی از روش‌های مبتنی بر نمونه‌گیری کمک می‌گیریم.

❖ در واقع سعی می‌کنیم با تولید نمونه‌های تصادفی از مدل، جواب مسأله را حدس بزنیم.

❖ به عبارت دیگر، روش مونت کارلو یک روش محاسباتی است که برای محاسبه نتایج، به نمونه‌های تصادفی تولیدشده خودش اکتفا می‌کند.

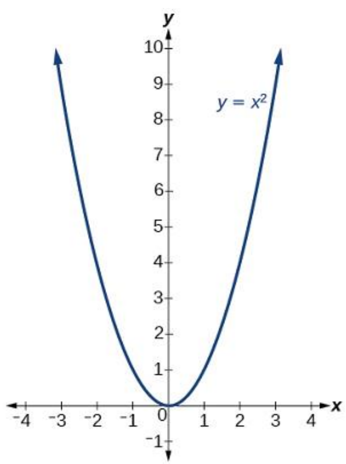
❖ سه گام اصلی روش نمونه‌گیری «مونت‌کارلو»:

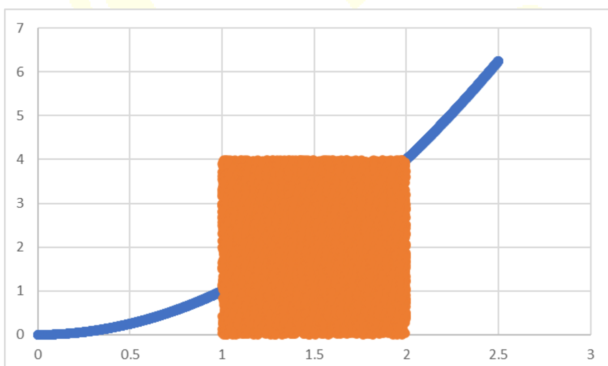
1. برای متغیرهای تصادفی موجود در مسأله، نمونه تصادفی به تعداد کافی تولید کنیم (n نمونه تصادفی).
2. شرطی را بر این نمونه‌های تصادفی اِعمال کنیم، و نمونه‌هایی که دارای این شرط هستند را انتخاب کنیم (m نمونه).
3. محاسبه پارامتر موردنظر مسأله از طریق m و n (پارامتر موردنظر می‌تواند مقدار احتمال باشد، مساحت باشد یا ...)

مکانیسم MCMC برای برآورد مدل‌های بیزین

$$y = x^2 \qquad \int_1^2 (x^2)dx$$

- **مثال روش نمونه‌گیری «مونت‌کارلو»**
- مطلوبست محاسبه انتگرال روبرو؛
- به جز روش حل مستقیم، می‌توان از طریق نمونه‌گیری از مقادیر تابع y در نقاط مختلف از x، مقدار این انتگرال را تخمین زد.
- پس نیازمند یک نمونه‌گیری تصادفی هستیم (تولید یک مجموعه نمونه تصادفی).
- برای استخراج نمونه‌ها به صورت تصادفی، نیازمند این هستیم که؛
  - تعداد نمونه موردنظر را تعیین کنیم (مثلا n=1000)،
  - یک تابع توزیع دلخواه تعیین کنیم (مثلا تابع توزیع یکنواخت) که اعداد تصادفی را در بازه موردنظر تولید کند.
- بدین‌ترتیب، نمونه‌های تصادفی از اعداد داخل فضای x و y را تولید کرده‌ایم.





- مساحت مستطیل = 1*4=4
- تعداد اعدادتصادفی تولید شده: n=1000
- سهم هر نقطه از مساحت: $\frac{4}{10000}$

مکانیسم MCMC برای برآورد مدل‌های بیزین

$$y = x^2 \qquad \int_1^2 (x^2)dx$$

X	Y	x^2	
1.327506	2.842656	1.762272	
1.69407	2.724685	2.869873	
1.403956	2.895353	1.971093	
1.761928	3.36034	3.104389	
1.587015	2.537926	2.518617	
1.227633	1.297558	1.507084	
1.735012	3.476611	3.010267	
1.363341	1.297257	1.858698	
1.741825	1.188273	3.033953	
1.065856	1.476818	1.136048	
1.31545	3.603216	1.730409	
1.050623	1.016939	1.10381	
1.350805	1.436975	1.824673	
1.079555	1.610489	1.165439	
1.961523	2.476023	3.847572	

- **مثال روش نمونه‌گیری «مونت‌کارلو»**
- مطلوبست محاسبه انتگرال روبرو؛
  - مساحت مستطیل = **1*4=4**
  - تعداد اعدادتصادفی تولید شده: **n=1000**
  - سهم هر نقطه از مساحت: $\frac{4}{10000}$
  - سوال: چه تعداد از نقاط در زیر منحنی x^2 قرار می‌گیرند (**m** تا).
  - نیاز است مرز x^2 را مشخص کرده و نقاطی که زیر این مرز قرار دارند را انتخاب کنیم (یعنی مواردی که $y \leq x^2$).
  - بدین ترتیب، مساحت زیر نمودار برابر است با: $m * \frac{4}{10000}$
  - سوال: اگر محاسبه انتگرال سخت باشد باید چکار کرد؟

مکانیسم MCMC برای برآورد مدل‌های بیزین

- ❖ **مثال روش نمونه‌گیری «مونت‌کارلو»**
- ❖ مطلوبست استخراج اعداد تصادفی از یک توزیع نرمال به صورت زیر

$$\theta_t \sim N(0.5\ ,\sigma)$$

- ❖ در واقع ما در حال نمونه‌گیری از یک متغیر تصادفی هستیم که از توزیع نرمال با ویژگی بالا پیروی می‌کند.
- ❖ اگر مثلا پس از ۱۰۰۰ مرتبه نمونه‌گیری از این متغیر تصادفی، توزیع فراوانی مقادیر به دست آمده را ترسیم کنیم، نموداری شبیه به همان $N(0.5\ ,\sigma)$ خواهیم داشت.

مکانیسم MCMC برای برآورد مدل‌های بیزین

- ❖ **روش نمونه‌گیری «مونت‌کارلو»**
- ❖ **یادآوری**
- ❖ **سه گام اصلی روش نمونه‌گیری «مونت‌کارلو»،**

1. **برای متغیرهای تصادفی موجود در مسأله، نمونه تصادفی به تعداد کافی تولید کنیم (n نمونه تصادفی).**
2. **شرطی را بر این نمونه‌های تصادفی اِعمال کنیم، و نمونه‌هایی که دارای این شرط هستند را انتخاب کنیم (m نمونه).**
3. **محاسبه پارامتر موردنظر مسأله از طریق m و n (پارامتر موردنظر می‌تواند مقدار احتمال باشد، مساحت باشد یا ...)**

سوال: در گام اول، تولید نمونه‌های تصادفی قرار است چگونه تولید شود؟ از چه تابع توزیعی پیروی کند؟

در مثال قبلی (انتگرال x^2)، از تابع توزیع یکنواخت پیروی می‌کردند.

پس می‌خواهیم تابع توزیع ما، از یک تابع توزیع موردنظر (هدف) پیروی کند.

- ❖ **در اینجا اگر تابع توزیع هدف، پیچیده باشد، نمی‌توانیم به راحتی از آن نمونه تصادفی تولید کنیم.**

- ❖ در اینجاست که روش زنجیره مارکوف به کمک ما می‌آید و این توانایی را به ما می‌دهد که حتی اگر تابع هدف پیچیده باشد (حتی اگر دارای توزیع شرطی بود، حتی اگر ابعاد بالا داشت)، ما بتوانیم از آن نمونه‌های تصادفی تولید کنیم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

تابع حالت گذار

- روش «**زنجیره مارکوف**»؛
- فرض کنید یک حالت اولیه داریم ($x^{(1)}$) و می‌خواهیم به حالت دوم $x^{(2)}$ منتقل شویم.
- فرض کنید برای انتقال از حالت اولیه به حالت بعدی، یک تابع داریم (با نام تابع حالت گذار). $x^{(1)} \to x^{(2)}$
- تابع حالت گذار $\boldsymbol{f}(x^{(t)}|x^{(t-1)})$

- اگر این کار در یک توالی انجام شود، عملاً زنجیره‌ای از مقادیر به صورت زیر خواهیم داشت

$$x^{(1)} \to x^{(2)} \to x^{(3)} \to \ldots \to x^{(t-1)} \to x^{(t)} \to \ldots$$

- یک ویژگی این توالی این است که هر حالت دلخواه آینده $x^{(t)}$ تنها به حالت فعلی $x^{(t-1)}$ وابسته است.
- بنابراین، زنجیره مارکوف یک فرایند تصادفی **بدون حافظه** است، بدین معنی که توزیع احتمال شرطی حالت بعد تنها به حالت فعلی بستگی دارد و مستقل از گذشته‌ی آن است.

- برای مثال اگر از متغیر تصادفی θ_t به صورت روبرو نمونه‌گیری کنیم، در واقع زنجیره‌ای از مقادیر θ_t خواهیم داشت که هر مقدار بعدی تنها به حالت فعلی بستگی دارد .

$$\theta_t \sim N(\theta_{t-1}, \sigma)$$

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ روش «**زنجیره مارکوف**»؛

❖ یک ویژگی زنجیره مارکوف: اگر توالی زنجیره مارکوف به اندازه کافی زیاد باشد، فارغ از اینکه زنجیره از چه حالت اولیه‌ای شروع شده باشد، پس از تولید **تعداد کافی از توالی‌ها**، به یک حالت پایدار (steady state) می‌رسد

به اندازه کافی زیاد

$x^{(1)} \rightarrow \ldots\ldots\ldots\ldots\ldots\ldots \rightarrow x^{(n)}$

$x'^{(1)} \rightarrow \ldots\ldots\ldots\ldots\ldots\ldots \rightarrow x^{(n)}$

Initial state

steady state

❖ پس ما از هر x رندمی که شروع کنیم، زنجیره مارکوف ما را به تابع توزیع هدف می‌رساند.

❖ یعنی زنجیره مارکوف ابزاری است که x رندم ما را به x رندمی که از تابع توزیع هدف پیروی می‌کند تبدیل می‌نماید.

یک نمونه تصادفی داریم ← آن را به زنجیره مارکوف می‌سپاریم ← یک نمونه تصادفی براساس توزیع هدف تحویل می‌گیریم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

- روش MCMC در واقع از ترکیب دو روش «مونت‌کارلو» و «زنجیره مارکوف» تشکیل شده است.
- **ترکیب دو روش «مونت‌کارلو» و «زنجیره مارکوف»؛**
- بنابراین، با استفاده از روش زنجیره مارکوف، مشکل موجود در (گام ۱) از روش مونت کارلو را حل کردیم.
- یعنی می‌توانیم از هر تابع توزیع هدفی، نمونه‌تصادفی تهیه کنیم و سپس به روش مونت کارلو بسپاریم.

- یادآوری: به طور کلی، روش MCMC یک ابزار برای تولید اعداد تصادفی است.

- سه الگوریتم معروف برای روش MCMC؛
  - متروپلیس
  - متروپلیس-هستینگز
  - گیبز

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس؛**

- فرض کنید یک تابع توزیع $p(\theta)$ داریم که بر روی یک مجموعه متغیر تصادفی θ تعریف شده است.
- این همان تابع توزیعی است که می‌خواهیم براساس آن، عدد تصادفی تولید کنیم (تابع توزیع هدف)
- متغیرهای تصادفی θ به هر حال دارای یک بازه‌ای از مقادیر هستند.
- برای شروعِ تولید اعداد تصادفی از این تابع توزیع، باید از یک مقدار اولیه شروع کنیم و سپس تابع توزیع حاکم بر آن مقدار اولیه را به سمت تابع توزیع هدف ($p(\theta)$) همگرا کنیم.

با استفاده از متروپلیس

$$\theta^{(1)} \to \underbrace{\theta^{(2)} \to \ldots \to \theta^{(t)} \to \ldots \to \theta^{(n)}}_{\text{به اندازه کافی زیاد}}$$

به اندازه کافی زیاد

Burn-in

- پس از اینکه تعداد کافی از توالی‌ها را پشت سرگذاشتیم، امیدواریم که تابع توزیع حاکم بر $\theta^{(n)}$ به سمت تابع توزیع هدف ($p(\theta)$) همگرا شده باشد.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس؛**

- اکنون سوال اینست که چطور از حالت فعلی $\theta^{(t-1)}$ به حالت بعدی $\theta^{(t)}$ حرکت می‌کنیم.
- در الگوریتم متروپلیس یک θ^* پیشنهادی تولید می‌کنیم و سپس تصمیم می‌گیریم که آیا این θ^* پیشنهادی را به عنوان $\theta^{(t)}$ انتخاب کنیم یا خیر؟
- برای تولید θ^* پیشنهادی از یک توزیع پیشنهادی استفاده می‌کنیم؛ که معمولاً یک توزیع شناخته شده است (مثلا توزیع یکنواخت یا توزیع نرمال)
- یعنی با استفاده از مقدار فعلی $\theta^{(t-1)}$ و از طریق یک توزیع پیشنهادی، یک مقدار پیشنهادی تصادفی برای حالت بعدی $\theta^{(t)}$ پیشنهاد می‌دهیم.
- اکنون باید تصمیم بگیریم که این مقدار پیشنهاد شده θ^* را قبول کنیم یا رد کنیم؟

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس؛**

- اکنون باید تصمیم بگیریم که این مقدار پیشنهاد شده θ^* را قبول کنیم یا رد کنیم؟

- این تصمیم‌گیری را به صورت احتمالاتی انجام می‌دهیم (یعنی براساس یک احتمالی این مقدار را قبول یا رد می‌کنیم)
- برای این کار، دو شاخص روبرو را محاسبه می‌کنیم:

احتمال پذیرش : $\alpha = min\left\{1, \dfrac{p(\theta^*)}{p(\theta^{(t-1)})}\right\}$

u: یک عدد تصادفی (بین ۰ و ۱) استخراج شده از توزیع یکنواخت $u \sim U(0,1)$

- اگر $\alpha \geq u$ باشد، مقدار پیشنهاد شده θ^* را به عنوان $\theta^{(t)}$ قبول می‌کنیم.
- اگر $\alpha < u$ باشد، مقدار پیشنهاد شده θ^* را قبول نمی‌کنیم و همان مقدار فعلی $\theta^{(t-1)}$ را به عنوان $\theta^{(t)}$ در نظر می‌گیریم.
- این فرآیند زنجیره‌ای را تکرار می‌کنیم (معمولاً به یک تعداد از پیش تعیین شده **Burn-in**).

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **گام‌های روش MCMC – با الگوریتم متروپلیس؛**

الف) یک مقدار اولیه برای متغیر(های) تصادفی موردنظر انتخاب می‌کنیم $\theta^{(1)}$.

ب) گام‌های زیر را تکرار می‌کنیم تا زمانی که شرط خاتمه محقق شود؛

1. یک مقدار پیشنهادی θ^{*} به صورت تصادفی تولید می‌کنیم (با استفاده از مقدار فعلی $\theta^{(t-1)}$ و از طریق یک توزیع پیشنهادی)
2. مقدار احتمال پذیرش را محاسبه می‌کنیم (α)
3. یک عدد تصادفی یکنواخت (U) بین صفر و یک تولید می‌کنیم.
4. تصمیم می‌گیریم که مقدار پیشنهادی θ^{*} را به عنوان مقدار جدید $\theta^{(t)}$ قبول کنیم یا کار را دوباره با همان مقدار فعلی $\theta^{(t-1)}$ ادامه دهیم؟
   - اگر $u \leq \alpha$ باشد، مقدار پیشنهاد شده θ^{*} را به عنوان $\theta^{(t)}$ قبول می‌کنیم.
   - اگر $u > \alpha$ باشد، مقدار پیشنهاد شده θ^{*} را قبول نمی‌کنیم و همان مقدار فعلی $\theta^{(t-1)}$ را به عنوان $\theta^{(t)}$ در نظر می‌گیریم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس–هستینگز؛**

- گام‌های دو الگوریتم **متروپلیس** و **متروپلیس–هستینگز** یکسان است.
- **تفاوت:** توزیع پیشنهادی (در الگوریتم متروپلیس) باید متقارن باشد.
- اما در الگوریتم متروپلیس–هستینگز، توزیع پیشنهادی لازم نیست حتما متقارن باشد. این تفاوت اصلی این دو الگوریتم هست.
- این ایده را از طریق دستکاری در احتمال پذیرش مقدار پیشنهادی انجام می‌دهیم.

احتمال پذیرش در الگوریتم متروپلیس :

$$\alpha = min\left\{1, \frac{p(\theta^*)}{p(\theta^{(t-1)})}\right\}$$

احتمال پذیرش در الگوریتم متروپلیس–هستینگز :

$$\alpha = min\left\{1, \frac{p(\theta^*)}{p(\theta^{(t-1)})} \times \frac{q(\theta^{(t-1)}|\theta^*)}{q(\theta^*|\theta^{(t-1)})}\right\}$$

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس–هستینگز؛**

- حالتی که با یک **تابع توزیع چند متغیره** روبرو باشیم. یعنی تابع توزیع هدف، چند متغیره است.
- قبلا دیدیم که برای پیدا کردن مقدار جدید برای متغیر موردنظر، از توزیع پیشنهادی استفاده کردیم.
- اکنون هم که تابع توزیع هدف ما چندمتغیره شده است، باید تابع توزیع پیشنهادی هم چندمتغیره (هم بُعد با تابع توزیع هدف) باشد.
- یعنی مثلا وقتی تابع توزیع هدف یک تابع با ۵ متغیر تصادفی است، تابع توزیع پیشنهادی هم به همین شکل خواهد بود.
- پس اینجا θ^* در واقع یک بردار با ۵ مولفه خواهد بود.

$$\boldsymbol{\theta}^* = ({\theta^*}_1, {\theta^*}_2, \dots, {\theta^*}_n)$$

- سوالی که اینجا مطرح می‌شود اینست که قضاوت پذیرش یا رد $\boldsymbol{\theta}^*$ را به صورت تک‌تک انجام دهیم یا برای کل متغیرها (پارامترها) به صورت همزمان پذیرش/رد کنیم؟
- بر مبنای پاسخ به این سوال، دو روش Block-wise updating و Component-wise updating قابل استفاده است.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **گام‌های روش MCMC – با الگوریتم متروپلیس–هستینگز (به صورت Block-wise updating)؛**

الف) یک مجموعه مقدار اولیه برای بردار متغیرهای تصادفی $\boldsymbol{\theta}^{(1)}$ انتخاب می‌کنیم.

ب) گام‌های زیر را تکرار می‌کنیم تا زمانی که شرط خاتمه محقق شود؛

1. یک مقدار پیشنهادی برای هر یک از مولفه‌های بردار $\boldsymbol{\theta}^*$ به صورت تصادفی تولید می‌کنیم (با استفاده از بردار مقادیر فعلی $\boldsymbol{\theta}^{(t-1)}$ و از طریق یک توزیع پیشنهادی) $\boldsymbol{\theta}^* = ({\theta^*}_1, {\theta^*}_2, \dots, {\theta^*}_n)$
2. مقدار احتمال پذیرش را محاسبه می‌کنیم (α در اینجا یک عدد است یا برداری nتایی از اعداد؟).
3. یک عدد تصادفی یکنواخت (u) بین صفر و یک تولید می‌کنیم.
4. تصمیم می‌گیریم که بردار مقادیر پیشنهادی θ^* را به عنوان بردار مقادیر جدید $\theta^{(t)}$ قبول کنیم یا کار را دوباره با همان بردار مقادیر فعلی $\theta^{(t-1)}$ ادامه دهیم؟

توجه: پذیرش/رد مقادیر پیشنهادی θ^* را به صورت بلوکی انجام می‌دهیم (یعنی همه مولفه‌های بردار یا پذیرش و یا رد می‌شوند)

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم متروپلیس–هستینگز (به صورت** Component-wise updating**)؛**

❖ مشکلات روش Block-wise updating:

○ پیدا کردن تابع توزیع پیشنهادی با ابعاد بالا، پیچیدگی زیادی خواهد داشت.

○ در هر مرتبه بروزرسانی، مجبوریم کل بردار مقادیر پیشنهادی را پذیرش/رد کنیم. باعث می‌شود نرخ رد داده‌ها زیاد شود.

○ خصوصیات روش Component-wise updating:

○ در اینجا، تابع توزیع پیشنهادی تک‌متغیره خواهد بود (فارغ از اینکه تابع توزیع هدف چند متغیره است)

○ و هر بار یکی از مولفه‌های بردار θ^* را پیشنهاد می‌دهیم و وقتی که این مقدار، پذیرش/رد شد، به سراغ مولفه بعدی می‌رویم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **گام‌های روش MCMC – با الگوریتم متروپلیس–هستینگز (به صورت Component-wise updating)؛**

الف) یک مجموعه مقدار اولیه برای بردار متغیرهای تصادفی $\boldsymbol{\theta}^{(1)}$ انتخاب می‌کنیم.

$$\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_n^{(1)})$$

ب) گام‌های زیر را تکرار می‌کنیم تا زمانی که شرط خاتمه محقق شود؛

1. یک مقدار پیشنهادی برای مولفه اول از بردار متغیرهای تصادفی $\boldsymbol{\theta}$ یعنی θ_1^* به صورت تصادفی تولید می‌کنیم (با استفاده از مقدار فعلی $\boldsymbol{\theta}_1^{(t-1)}$ و از طریق یک توزیع پیشنهادی (که تک‌متغیره است))

2. مقدار احتمال پذیرش را محاسبه می‌کنیم (α).

$$\alpha = min\left\{1, \frac{p(\theta_1^*, \theta_2^{(t-1)}, \dots, \theta_n^{(t-1)})}{p(\theta_1^{(t-1)}, \theta_2^{(t-1)}, \dots, \theta_n^{(t-1)})} \times \frac{q(\theta_1^{(t-1)}|\theta_1^*)}{q(\theta_1^*|\theta_1^{(t-1)})}\right\}$$

3. یک عدد تصادفی یکنواخت (U) بین صفر و یک تولید می‌کنیم.

4. تصمیم می‌گیریم که آیا مقدار پیشنهادی برای θ_1^* را به عنوان مقدار جدید $\theta_1^{(t)}$ قبول کنیم یا کار را دوباره با همان مقدار فعلی $\theta_1^{(t-1)}$ ادامه دهیم؟

5. گام‌های ۱ تا ۴ را برای تک‌تک مولفه‌های بردار متغیرهای تصادفی $\boldsymbol{\theta}$ تکرار می‌کنیم تا مرحله اول پرش به صورت کامل انجام شود.

مکانیسم MCMC برای برآورد مدل‌های بیزین

- ❖ **روش MCMC – با الگوریتم گیبز؛**
- ❖ مشکلات روش‌های متروپلیس و متروپلیس-هستینگ؛
  - ○ به تابع توزیع پیشنهادی نیاز داریم که ساختن آن گاهی اوقات سخت است.
  - ○ به هرحال برخی از مقادیر پیشنهادی، رد می‌شوند.
  - ○ الگوریتم گیبز سعی می‌کند این مشکلات را حل کند.
    - ✓ در این الگوریتم تابع توزیع پیشنهادی نداریم.
    - ✓ همه مقادیر تولید شده، پذیرش می‌شوند.
  - ○ اما برای استفاده از الگوریتم گیبز باید حتماً شرط زیر برقرار باشد؛
    - ➤ تابع توزیع احتمال شرطی بین هر کدام از متغیرهای تصادفی باید معلوم (شناخته‌شده) باشد.
    - ➤ مثلاً توابع چگالی نرمال، لگ‌نرمال، دوجمله‌ای، پواسون، گاما، بتا، و غیره.

❖ **روش MCMC – با الگوریتم گیبز؛**

- فرض کنید یک تابع توزیع هدف با دو پارامتر (θ_1,θ_2) داریم:

$$(\theta_1,\theta_2) \sim f(\theta_1,\theta_2)$$ تابع توزیع هدف

- شرط استفاده از الگوریتم گیبز: تابع توزیع احتمال شرطی بین هر کدام از متغیرهای تصادفی باید معلوم (شناخته‌شده) باشد. بدین معناست که؛ مثلا تابع توزیع احتمال شرطی θ_1 به شرط معلوم بودنِ θ_2، برای ما مشخص باشد.

- و نه تنها معلوم باشد، بلکه بتوانیم از آن عدد تصادفی انتخاب کنیم.

- پس باید بتوانیم از این توابع توزیع احتمال شرطی (روبرو) نمونه تصادفی انتخاب کنیم.

$$\begin{cases} f(\theta_1|\theta_2 = \theta_2^{(t-1)}) \\ f(\theta_2|\theta_1 = \theta_1^{(t-1)}) \end{cases}$$

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **روش MCMC – با الگوریتم گیبز؛**

○ فرض کنید برای دو پارامتر (θ_1,θ_2) مقدار اولیه انتخاب کرده‌ایم: $(\theta_1^{(1)},\theta_2^{(1)})$

○ به صورت زیر، مقادیر بعدی برای دو پارامتر (θ_1,θ_2) انتخاب می‌شود؛

$$(\theta_1^{(1)},\theta_2^{(1)}) \longrightarrow f(\theta_1|\theta_2=\theta_2^{(1)}) \longrightarrow \theta_1^{(2)}$$

$$(\theta_1^{(2)},\theta_2^{(1)}) \longrightarrow f(\theta_2|\theta_1=\theta_1^{(2)}) \longrightarrow \theta_2^{(2)} \longrightarrow (\theta_1^{(2)},\theta_2^{(2)})$$

○ بدین‌ترتیب، با نمونه‌گیری شرطی، از بردار $(\theta_1^{(1)},\theta_2^{(1)})$ به بردار بعدی $(\theta_1^{(2)},\theta_2^{(2)})$ منتقل شدیم.

○ در واقع الگوریتم گیبز یک الگوریتم نمونه‌گیری شرطی تکراری (Iterative conditional sampling) است.

مکانیسم MCMC برای برآورد مدل‌های بیزین

❖ **گام‌های روش MCMC – با الگوریتم گیبز؛**

الف) یک مجموعه مقدار اولیه برای بردار متغیرهای تصادفی $\boldsymbol{\theta}^{(1)}$ انتخاب می‌کنیم.

$$\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_n^{(1)})$$

ب) گام‌های زیر را تکرار می‌کنیم تا زمانی که شرط خاتمه محقق شود؛

1. یک مقدار تصادفی برای مولفه اول از بردار متغیرهای تصادفی $\boldsymbol{\theta}$ یعنی $\boldsymbol{\theta}_1^{(t)}$ به صورت تصادفی تولید می‌کنیم (با استفاده از مقدار فعلی سایر متغیرها و از طریق توزیع احتمال شرطی)

$$f(\boldsymbol{\theta}_1^{(t)} | \theta_2 = \theta_2^{(t-1)}, \dots, \theta_n = \theta_n^{(t-1)})$$

2. همین کار را برای تمام n متغیر تصادفی انجام می‌دهیم.

3. تا در نهایت، یک مقدار تصادفی برای مولفه n-ام از بردار متغیرهای تصادفی $\boldsymbol{\theta}$ یعنی $\boldsymbol{\theta}_n^{(t)}$ به صورت تصادفی تولید می‌کنیم (با استفاده از مقدار فعلی سایر متغیرها و از طریق توزیع احتمال شرطی)

$$f(\boldsymbol{\theta}_n^{(t)} | \theta_1 = \boldsymbol{\theta}_1^{(t)}, \theta_2 = \theta_2^{(t)}, \dots, \theta_{n-1} = \theta_{n-1}^{(t)})$$

4. بدین‌ترتیب مرحله اول پرش به صورت کامل انجام شده است.

❑ سوال: آیا الگوریتم نمونه‌گیر گیبز حالت خاصی از الگوریتم متروپلیس-هستینگ است؟

مکانیسم MCMC برای برآورد مدل‌های بیزین

1. تعیین شکل مدل
2. تعیین تابع درستنمایی برای مدل موردنظر
3. تعیین توزیع پیشین برای پارامترهای مدل
4. فرموله کردن توزیع پسین
5. اجرای الگوریتم MCMC برای نمونه‌گیری از توزیع پسین
6. ذخیره مقادیر به دست آمده از نمونه‌گیری از توزیع پسین

مکانیسم MCMC برای برآورد مدل‌های بیزین

1. تعیین شکل مدل

- فرض کنید Y ها دارای یک رابطه خطی به صورت روبرو و یک خطای نرمال هستند؛ $Y = aX + b$
- می‌خواهیم یک برآورد بیزین از پارامترهای مدل (a ، b و خطای استاندارد .s.d) به دست بیاوریم.

2. تعیین تابع درستنمایی برای مدل موردنظر

- باید تابع درستنمایی برای مدلی بسازیم که می‌خواهیم پارامترهای آن را برآورد کنیم.
- احتمال مشاهده داده‌های نمونه تحت این مدل (با پارامترهای موردنظر) را محاسبه می‌کنیم.

3. تعیین تابع توزیع پیشین برای پارامترهای مدل

- برای هر یک از پارامترهای مدل یک توزیع پیشین در نظر می‌گیریم.

مکانیسم MCMC برای برآورد مدل‌های بیزین

4. **تعیین تابع توزیع پسین**

- توزیع پسین به صورت حاصل‌ضرب تابع درستنمایی و توزیع پیشین است.
- این توزیع، همان فضایی است که MCMC بر روی آن کار می‌کند (توزیع هدف).
- توجه: چون با لگاریتم درستنمایی کار می‌کنیم، به جای حاصل‌ضرب، از حاصل‌جمع استفاده می‌کنیم.

5. **اجرای الگوریتم MCMC برای نمونه‌گیری از توزیع پسین**

- با اجرای یکی از الگوریتم‌های MCMC در واقع، توزیع پارامترهای موردنظر به سمت توزیع هدف (که در رویکرد بیزین همان توزیع پسین است) همگرا می‌شود.
- ایده حاکم بر این الگوریتم چیست؟
  - در فضای جواب پارامتر موردنظر پرش کن، به گونه‌ای که احتمال بودن در یک نقطه متناسب با تابع هدف (توزیع پسین) باشد این ایده باعث می‌شود که به تدریج با اجرای این الگوریتم، توزیع مقادیر نمونه‌گیری شده از پارامترهای موردنظر به سمت توزیع هدف این پارامترها (که در رویکرد بیزین همان پسین است) همگرا شود.

مکانیسم MCMC برای برآورد مدل‌های بیزین

لینک یک مثال ساده از کد نویسی یک مدل در محیط R

https://theoreticalecology.wordpress.com/2010/09/17/metropolis-hastings-mcmc-in-r/