دانشگاه پیام نور مرکز تهران

مرکزتهران شرق

پایان نامه

برای دریافت درجه کارشناسی ارشد

رشته آمار ریاضی

گروه آمار

عنوان پایان نامه:

برآورد پارامترهای توزیع نمایی، نمایی شده تعمیم یافته

استاد راهنما:

دکتر پرویز نصیری

استاد مشاور:

دکتر علی شادرخ

نگارنده:

حسنا اسکندری

تابستان

1391

چکیده

در این پایان نامه باتوجه به خانواده توزیع های نمایی، توزیع نمایی، نمایی شده و توزیع نمایی نمایی شده تعمیم یافته،ارایه می شود . برآورد پارامترهای توزیع نمایی، نمایی شده تعمیم یافته به روش های برآورد گشتاوری و برآورد ماکزیمم درستنمایی برآورد وبا استفاده از روش شبیه سازی مونت -کارلو مارکفی
مقایسه دو روش برآورد گشتاوری و برآورد ماکزیمم درستنمایی ارایه شده است . با توجه به نتایج بدست آمده ، برآورد کارا معرفی شده است .

فهرست

مقدمه

اخیراً با توجه به اهمیت توزیع های نمایی و کاربرد آن محققین توزیع نمایی، نمایی شده را به عنوان جایگزینی برای توزیع های گاما و وایبل پیشنهاد کرده اند و خواص مختلف آن را مورد مطالعه قرار داده اند که در این پایان نامه با توجه به خانواده توزیع های نمایی به معرفی توزی عهای نمایی شده پرداخته و با در نظر گرفتن این توزیع ها، توزیع نمایی، نمایی شده تعمیم یافته معرفی شده است .همچنین برآورد پارامترهای این توزیع مد نظر بوده است که به رو شهای برآورد گشتاوری و برآورد ماکزیمم درستنمایی در به دست آوردن آ نها در این پایان نامه اقدام شده است .از آ نجا که در حل معادلات غیرخطی نیاز به استفاده از روش های عددی و شبیه سازی می باشد لذا در فصل دوم به معرفی شبیه سازی گیبس و مونت -کارلو مارکفی پرداخته شده است. در فصل سوم ,چهارم با توجه به روش های برآورد کلاسیک پارامترها شامل برآورد گشتاوری و برآورد ماکزیمم درستنمایی به منظور برآورد پارامترهای توزیع نمایی، نمایی شده تعمیم یافته ارائه شده است .درفصل پنجم شبیه سازی Matlab v نتایج ب هدست آمده برای برآورد پارامترها به روش کلاسیک با استفاده از نرم افزار 7.1و ارائه گردیده است.

فصل اول

- خانواده توزیع های نمایی
- توزیع های نمایی شده
- برآورد گشتاوری و ماکزیموم درستنمایی
- روش شبیه سازی و مونت کارلو– مارکفی

۱-۱خانواده توزیع‌های نمایی

خواص خانواده توزیع‌های نمایی در سال‌های ۱۹۳۷ و ۱۹۳۸ به طور مستقل توسط سه آماردان به نام‌های کوپمن(منبع [۱])، پیتمن (منبع [۱]) و دارمویس(منبع [۱]) مطالعه و بررسی شده است. به همین دلیل بعضی از محققین این گونه توزیع‌ها را «خانواده توزیع‌های کوپمن، پیتمن و دارمویس» نیز می‌نامند.

خانواده توابع چگالی احتمال (یا توابع احتمال) $\{f_\theta(x):\theta\in\Theta\subseteq R\}$ را در نظر بگیرید. این خانواده، عضو خانواده نمایی (یا خانواده نمایی یک پارامتری) گفته می‌شود اگر بتوان $f_\theta(x)$ را به صورت زیر نوشت :

$$f_\theta(x)=a(\theta)b(x)e^{c(\theta)d(x)} \qquad (۱-۱)$$

که در آن $c(\theta)$و $a(\theta)$ توابعی دلخواه از θ، $b(x)$و $d(x)$ توابعی دلخواه از x می‌باشند، به طوری که $b(x)$و $a(\theta)$ هر دو توابعی مثبت هستند.

اگر در رابطه (۱-۱)، $b(x)=\exp\{R(x)\}$ و $a(\theta)=\exp\{Q(\theta)\}$ باشند آنگاه رابطهٔ (۱) را می‌توان به صورت زیر نوشت :

$$f_\theta(x)=\exp\{c(\theta)d(x)+Q(\theta)+R(x)\} \qquad (۲-۱)$$

حال، خانوادهٔ توابع چگالی احتمال (یا توابع احتمال) $\{f_{\underline{\theta}}(x):\underline{\theta}\in\Theta\subseteq R^k\}$ را در نظر بگیرید. این خانواده، عضو خانوادهٔ توزیع‌های نمایی (یا خانوادهٔ توزیع‌های نمایی k - پارامتری) گفته می‌شود اگر بتوان $f_{\underline{\theta}}(x)$ را به صورت زیر نوشت :

$$f_{\underline{\theta}}(x)=a(\underline{\theta})b(x)\exp\left\{\sum_{i=1}^{k}c_i(\underline{\theta})d_i(x)\right\} \qquad (۳-۱)$$

که در آن $a(\underline{\theta})$ و $c_i(\underline{\theta})$، $i=1,\dots,k$، توابعی از $\underline{\theta}=(\theta_1,\dots,\theta_k)$و $b(x)$و $d_i(x)$، $i=1,\dots,k$، توابعی از x می‌باشند، به طوری که $b(x)$ و $a(\underline{\theta})$ توابعی مثبت هستند.

اگر در رابطهٔ (۱-۳) $a(\underline{\theta}) = \exp\{Q(\underline{\theta})\}$ و $b(x) = \exp\{R(x)\}$ انتخاب شوند، آنگاه رابطهٔ (۱-۳) به صورت زیر خواهد شد :

$$f_{\underline{\theta}}(x) = \exp\left\{\sum_{i=1}^{k} c_i(\underline{\theta}) d_i(x) + Q(\underline{\theta}) + R(x)\right\} \quad (۱-۴)$$

خانواده توزیع‌های نمایی در حالت پیوسته و گسسته وجود دارد که بصورت زیر توضیح داده خواهد شد.

۱-۱ -۱ خانواده توزیع نمایی حالت پیوسته

۱ . خانوادهٔ چگالی‌های یک پارامتری $\{f_\theta(x): \theta \epsilon \Theta \subseteq R\}$ متعلق به خانوادهٔ توزیعهای نمایی (یک پارامتری) با شرایط مطلوب گفته می‌شود، اگر بتوان $f_\theta(x)$ را به صورت رابطهٔ (۱-۲) باتکیه‌گاه $S_x = (a, b)$ نوشت؛ که در آن :

الف : هیچ یک از مقادیر a و b به θ بستگی ندارند.

ب : $c(\theta)$ تابعی غیر صفر و پیوسته از θ، $\theta \in \Theta$، است.

ج : هر یک از توابع $d'(x) \not\equiv 0$ ($d'(x)$ مشتق تابع $d(x)$ می‌باشد) و $R(x)$، توابعی پیوسته از x روی S_X می‌باشند.

۲. خانواده چگالی‌های k - پارامتری $\{f_{\underline{\theta}}(x): \underline{\theta} \in \Theta \subseteq R^k\}$ متعلق به خانوادهٔ توزیع‌های نمایی (k - پارامتری) با شرایط مطلوب گفته می‌شود، اگر بتوان $f_{\underline{\theta}}(x)$ را به صورت رابطهٔ (۴) با تکیه‌گاه $S_X = (a,b)$ نوشت؛ که در آن :

الف : هیچ یک از مقادیر a و b به $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ بستگی ندارند.

ب : $c_i(\underline{\theta})$، $i = 1, \ldots, k$، توابع غیر صفر، به طور تابعی مستقل و پیوسته از θ_j ها، $j = 1, \ldots, k$، هستند.

ج : $d_i'(x)$، $i = 1, \ldots, k$، توابعی پیوسته از $x \in (a,b)$ می‌باشند و هیچ یک از آن‌ها تابع خطی همگن از بقیه نیستند.

د : $R(x)$ تابعی پیوسته از $x \in (a,b)$ می‌باشد.

۱-۱-۲حالت گسسته

۱. خانوادهٔ توابع احتمال یک پارامتری $\{f_\theta(x):\theta\in\Theta\subseteq R\}$ متعلق به خانواده توزیع‌های نمایی (یک پارامتری) با شرایط مطلوب گفته می‌شود اگر بتوان $f_\theta(x)$ را به صورت (۱-۲) با تکیه‌گاه $S_X=\{x:x=a_1,a_2,...\}$ نوشت؛ که در آن :

الف : مجموعهٔ S_X به θ بستگی ندارد.

ب : $c(\theta)$ یک تابع غیر صفر پیوسته از θ، $\theta\in\Theta$، است.

ج : $d(x)$ یک تابع غیر صفر از x روی مجموعهٔ S_X می‌باشد.

۲. خانواده چگالی‌های k - پارامتری $\{f_{\underline{\theta}}(x):\underline{\theta}\in\Theta\subseteq R^k\}$ متعلق به خانوادهٔ توزیع‌های نمایی (k - پارامتری) با شرایط مطلوب گفته می‌شود، اگر بتوان $f_{\underline{\theta}}(x)$ را به صورت رابطهٔ (۴) با تکیه‌گاه $S_X=\{x:x=a_1,a_2,...\}$ نوشت؛ که در آن :

الف : مجموعهٔ S_X به $\underline{\theta}=(\theta_1,...,\theta_k)$ بستگی ندارد.

ب : $c_i(\underline{\theta})$، $i=1,...,k$، توابعی غیر صفر، به طور تابعی مستقل و پیوسته از θ_j ها، $j=1,...,k$، هستند.

ج : $d_i(x)$، $i=1,...,k$، توابعی غیر صفر از x روی مجموعهٔ S_X است و هیچ یک از آن‌ها تابعی خطی از بقیه نیست.

با توجه به آنچه که بیان شد بسیاری از توزیع‌های استاندارد در حالت پیوسته نظیر : توزیع نرمال، توزیع گاما، توزیع بتا، توزیع پاراتو، توزیع وایبل و توزیع گوسین معکوس و در حالت گسسته از قبیل : توزیع برنولی، توزیع دوجمله‌ای، توزیع دوجمله‌ای منفی (پاسکال)، توزیع پواسن و توزیع سری‌های لگاریتمی متعلق به خانوادهٔ توزیع‌های نمایی (یک پارامتری و k - پارامتری)می‌باشند.

در پایان لازم است به برخی از ویژگی‌ها و خواص خانوادهٔ توزیع‌های نمایی اشاره نماییم :

۱- اگر X یک متغیر تصادفی با شرایط مطلوب از خانوادهٔ توزیع‌های نمایی به صورت (۱-۲) باشد، آنگاه :

$$E_\theta\left[d(X)\right]=-\frac{Q'(\theta)}{c'(\theta)}$$

که در آن $Q'(\theta)$ و $c'(\theta)$ به ترتیب مشتق توابع $Q(\theta)$ و $c(\theta)$ می‌باشند.

۲- اگر X یک متغیر تصادفی با شرایط مطلوب از خانوادهٔ توزیع‌های نمایی k - پارامتری به صورت (۱-۴) باشد، آنگاه با انتخاب $c_i(\underline{\theta}) = \theta_i$،

الف :

$$E_{\underline{\theta}}\left[d_i(X)\right] = -\frac{\partial Q(\underline{\theta})}{\partial \theta_i}$$

ب :

$$Cov_{\underline{\theta}}\left[d_i(X), d_j(X)\right] = -\frac{\partial^2 Q(\underline{\theta})}{\partial \theta_i \partial \theta_j} \quad ; \qquad \mathrm{i} \neq \mathrm{j}$$

۳- برای یک نمونه تصادفی مانند $X_1, \ldots, X_n$ از خانوادهٔ توزیع‌های نمایی k - پارامتری به صورت (۱-۳)، $\left(\sum_{j=1}^{n} d_1(X_j), \ldots, \sum_{j=1}^{n} d_k(X_j)\right)$ یک آماره بسنده برای $\underline{\theta}$ می‌باشد. در واقع $\left(\sum_{j=1}^{n} d_1(X_j), \ldots, \sum_{j=1}^{n} d_k(X_j)\right)$ آماره‌های بسنده مینیمال می‌باشند.

۴- اگر $X_1, \ldots, X_n$ نمونه‌ای تصادفی از خانوادهٔ توزیع‌های نمایی یک پارامتری باشد به صورت (۱-۱) باشد. آنگاه $\sum_{i=1}^{n} d(X_i)$ یک آمارهٔ بسنده مینیمال کامل است.

۵- هر خانواده چگالی‌ها که برای آن دامنه مقادیری که در آن چگالی نامنفی به پارامتر θ وابسته باشد، متعلق به خانوادهٔ نمایی نمی‌باشد.

۶- اگر $X_1, \ldots, X_n$ یک نمونهٔ تصادفی از خانوادهٔ توزیع‌های نمایی به صورت (۱-۲) باشد، تابع چگالی احتمال توأم $X_1, \ldots, X_n$ برابر است با

$$f_\theta(x_1, \ldots, x_n) = \exp\left\{c(\theta)\sum_{i=1}^{n} d(x_i) + nQ(\theta) + \sum_{i=1}^{n} R(x_i)\right\}$$

آنگاه واریانس هر تابع خطی از $\sum_{i=1}^{n} d(X_i)$ در برآورد پارامتر $\gamma(\theta) = E_\theta\left[\sum_{i=1}^{n} d(X_i)\right]$ کران پایین کرامر – رائو را به دست می‌دهد.

۲-۱ توزیع‌های نمایی شده

تابع $H(x)=[G(x)]^{\alpha}$ توزیع نمایی شده نامیده می‌شود اگر تابع توزیع تجمعی (CDF)(Cumulative Distribution Function)، $G(x)$ توسط پارامتر α که یک عدد حقیقی مثبت می‌باشد، نمایی شده باشد. همچنین مدل نرخ مخاطره (زیان) معکوس شده (PRHRM) (Proportional Reversed Hazard Rate Model) نیز نامیده می‌شود، زمانی که تابع نرخ مخاطره معکوس شده (RHRF) (Reversed Hazard Rate Function)، H به صورت زیر تعریف شده باشد :

$$\lambda_H^*(x)=\frac{d}{dx}\left[\ln H(x)\right]=\frac{h(x)}{H(x)}$$

که در آن $h(x)$ تابع چگالی احتمال (PDF)(Probability Density Function) مطابق با تابع توزیع $H(x)$ می‌باشد؛ بنابراین :

$$\lambda_H^*(x)=\frac{\alpha\{G(x)\}^{\alpha-1}g(x)}{\{G(x)\}^{\alpha}}=\alpha\lambda_G^*(x)$$

بنابراین تابع نرخ مخاطره معکوس شده H با نسبت ثابت α با تابع نرخ مخاطره معکوس شده G متناسب است.

می‌توان نشان داد که تابع توزیع تجمعی $H(x)$ را می‌توان به صورت جملات تابع نرخ مخاطره $\lambda_H(x)$ و تابع نرخ مخاطره معکوس شده به صورت زیر نوشت :

$$H(x)=\frac{\lambda_H(x)}{\lambda_H(x)+\lambda_H^*(x)}$$

اگر :

۱. $\alpha=1$، آنگاه به طور خطی $H(x)=G(x)$.
۲. اگر α یک عدد صحیح مثبت باشد، $\alpha=N$، $N=1,2,\ldots$، آنگاه $H(x)=[G(x)]^{N}$ تابع توزیع تجمعی ماکزیمم نمونه تصادفی به حجم N گرفته شده از تابع توزیع تجمعی G می‌باشد.

افزودن یک یا چند پارامتر به یک توزیع برای مدل سازی داده‌ها آن را قوی‌تر و منعطف‌ترمی‌سازد. برای مثال، اضافه کردن پارامتر α به صورت نمایی به تابع توزیع تجمعی G تابع توزیع تجمعی G^{α} را به وجود می‌آورد که برای مدل سازی داده‌ها تواناتر و منعطف‌ترمی‌باشد و تابع توزیع نمایی شده G^{α} کاملاً با تابع توزیع G متفاوت است.

۱-۲-۱) خانوادۀ توزیع‌های نمایی، نمایی شده

خانوادۀ چگالی‌های $f\left(.,\theta_1,\ldots,\theta_k\right)$ را که بتوان برای یک انتخاب مناسب توابع $a\left(.,\ldots,.\right)$، $b\left(.\right)$، $c_j\left(.,\ldots,.\right)$ و $d_j\left(.\right)$، $j=1,\ldots,k$، به صورت زیر نوشت :

$$f\left(x;\theta_1,\ldots,\theta_k\right)=a\left(\theta_1,\ldots,\theta_k\right)b\left(x\right)\exp\left\{\sum_{j=1}^{k}c_j\left(\theta_1,\ldots,\theta_k\right)d_j\left(x\right)\right\}$$

متعلق به خانواده نمایی k - پارامتری تعریف می‌شود. برای $k=1$، خانوادۀ نمایی یک پارامتری (یک بُعدی) حاصل می‌شود.

اگر متغیر تصادفی X دارای تابع چگالی احتمال به صورت :

$$f_{\lambda}\left(x\right)=\lambda e^{-\lambda x} \qquad ; \quad x\geq 0\ ,\lambda>0$$

و تابع توزیع تجمعی به صورت:

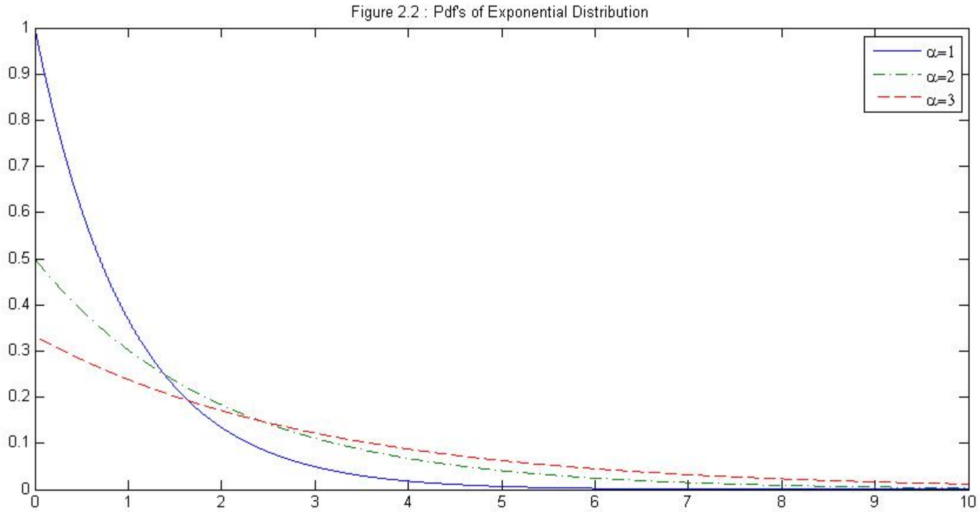
$$F_X\left(x\right)=1-e^{-\lambda x} \qquad ; \qquad \mathrm{x}\geq 0\ ,\lambda>0$$

باشد، آنگاه می‌گوییم X دارای توزیع نمایی با پارامتر λ می‌باشد و به صورت $X\sim\exp\left(\lambda\right)$ نمایش داده می‌شود.

اگر در تابع چگالی گاما با پارامترهای α و λ، یعنی $\mathrm{G}(\alpha,\lambda)$، $\alpha=1$، تابع چگالی گاما به تابع چگالی نمایی (منفی) تبدیل می‌شود که تابع چگالی آن به صورت زیر می‌باشد

$$f_{\lambda}\left(x\right)=\lambda e^{-\lambda x} \qquad ; \qquad x\geq 0\ ,\ \lambda>0$$

و نمودار آن برای $\lambda=1,2,3$ به صورت زیر خواهد بود :



بنابراین اول این که تابع چگالی نمایی حالت خاصی از تابع چگالی گاما می‌باشد و دوم اینکه مجموع متغیرهای تصادفی مستقل همتوزیع با توزیع نمایی دارای توزیع گاما می‌باشد.

از آنجا که توزیع نمایی به عنوان الگویی برای طول عمر فرایندهای گوناگون مورد استفاده قرار می‌گیردمی‌توان نشان داد که طول فاصله زمانی بین وقایع متوالی، مشروط بر این که تعداد وقایع در فاصله زمانی ثابت دارای توزیع پواسن باشد، دارای توزیع نمایی است.

همانگونه که در بخش قبل بیان شد اگر به تابع توزیع تجمعی یک متغیر تصادفی، پارامتر α را به صورت نمایی اضافه نمود توزیعی به دست می‌آید که به توزیع نمایی شده معروف می‌باشد.(منبع [۲۰]) حال اگر به تابع توزیع تجمعی نمایی پارامتر α را اضافه کرد، توزیع نمایی، نمایی شده حاصل می‌گردد که تابع چگالی احتمال و تابع توزیع تجمعی آن به ترتیب به صورت زیر می‌باشد :

$$f_E\left(x;\alpha,\lambda\right)=\alpha\lambda e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{\alpha-1} \qquad ; \qquad x>0\ ,\ \alpha>0\ ,\ \lambda>0$$

$$F_E\left(x;\alpha,\lambda\right)=\left(1-e^{-\lambda x}\right)^{\alpha} \qquad ; \qquad x>0\ ,\ \alpha>0\ ,\ \lambda>0$$

پارامترهای تابع توزیع نمایی، نمایی شده همانند پارامترهای مقیاس و مکان توزیع‌های گاما و وایبل می‌باشد. توابع چگالی مختلف توزیع نمایی، نمایی شده به طور معنی داری وابسته به پارامتر مقیاس می‌باشند. همچنین بسیاری از ویژگی‌هایآن‌ها کاملاً مشابه ویژگی‌های توزیع گاما می‌باشد.

تابع بقا و تابع مخاطره توزیع نمایی، نمایی شده به ترتیب به صورت زیر می‌باشد :

$$S_E(x;\alpha,\lambda)=1-\left(1-e^{-\lambda x}\right)^{\alpha} \qquad ; \qquad x>0\ ,\alpha>0\ ,\lambda>0$$

$$h_E(x;\alpha,\lambda)=\frac{\alpha\lambda e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{\alpha-1}}{1-\left(1-e^{-\lambda x}\right)^{\alpha}} \qquad ; \qquad x>0\ ,\alpha>0\ ,\lambda>0$$

که در آن α و λ به ترتیب پارامترهای مقیاس و مکان می‌باشند. برای $\alpha=1$ توزیع نمایی حاصل می‌گردد؛ بنابراین خانواده توزیع‌های گاما، وایبل و نمایی، نمایی شده تعمیم یافته توزیع نمایی می‌باشند. برای هر λ، اگر $\alpha>1$،آنگاه تابع مخاطره غیر نزولی و اگر $\alpha<1$، آنگاه تابع مخاطره، تابعی غیر صعودی و برای $\alpha=1$ تابعی ثابت می‌باشد.

۱–۳ برآوردگرهای گشتاوری و ماکزیمم درستنمایی

چندین روش برای به دست آوردن برآوردگرهای نقطه‌ای وجود دارد که از میان این روش‌هامی‌توان به :

۱– روش گشتاوری.

۲– روش ماکزیمم درستنمایی.

۳– روش بیز.

اشاره کرد که در قسمت مقدمه به اختصار به معرفی روش‌های گشتاوری و ماکزیمم درستنمایی پرداخته و در فصل‌های دومو سوم به ترتیب به بررسی روش‌های گشتاوری و ماکزیمم درستنمایی تحت عنوان برآوردگرهای کلاسیک پارامترها خواهیم پرداخت.

۱–۳–۱) روشهای کلاسیک برآوردیابی

روش‌های برآورد گشتاوری (Method of Moment Estimator) ، که مبتنی بر به کارگیری گشتاورهای نمونه‌ایکه وابسته به صحیح بودن یا نبودن پارامترهای توزیع می‌باشد، و برآورد ماکزیمم درستنمایی که عبارت است از دستورالعملی برای به دست آوردن برآوردگری به نام «برآوردگر ماکزیمم درستنمایی»که از آن به اختصار MLE (Maximum Likelihood Estimator) یاد می‌شود و مبتنی بر یک تابع آماری مهم به نام «تابع درستنمایی» است و تابعی از پارامتر می‌باشد، را روش‌های کلاسیک برآوردیابی پارامترهای توزیع می‌گویند. در روش برآورد ماکزیمم درستنمایی برای تداعی اینکه تابع

درستنمایی تابعی از θ است، نماد $L\left(\theta,x_1,...,x_n\right)$ یا $L\left(.;x_1,...,x_n\right)$ را برای تابع درستنمایی به کار می‌بریم. تابع درستنمایی $L\left(\theta,x_1,...,x_n\right)$، درستنمایی‌ای که متغیرهای تصادفی مقدار ویژهٔ $x_1,...,x_n$ را اختیار کنند می‌دهد. درستنمایی مقدار یک تابع چگالی است؛ بنابراین برای متغیرهای تصادفی گسسته، برابر با احتمال یک است. فرض کنید θ معلوم باشد و مقدار آن را با θ_0 نشان دهید، مقدار ویژهٔ متغیرهای تصادفی که «بیشتر محتمل است رخ دهد» آن مقدار $x_1',...,x_n'$ است به طوری که $f_{X_1,...,X_n}\left(x_1,...,x_n;\theta_0\right)$ ماکزیمم باشد.

همچنین در برآورد ماکزیمم درستنمایی هدف پیدا کردن آن مقدار از θ در Θ، که با $\hat{\theta}$ نشان داده می‌شود، است که تابع درستنمایی $L\left(\theta,x_1',...,x_n'\right)$ را ماکزیمم کند. معمولاً مقدار $\hat{\theta}$ که تابع درستنمایی را ماکزیمم می‌کند، تابعی از $x_1,...,x_n$، مانند $\hat{\theta}=\hat{\vartheta}\left(x_1,...,x_n\right)$ می‌باشد. وقتی چنین حالتی باشد، متغیر تصادفی $\hat{\Theta}=\hat{\vartheta}\left(x_1,...,x_n\right)$ را برآوردگر ماکزیمم درستنمایی θ می‌نامند. [فرض کنید ماکزیمم تابع درستنمایی همواره وجود دارد].

۱–۴ روش شبیه سازی مونت – کارلو مارکفی

در برآورد بیزی پارامترها، معمولاً با انتگرالهایی به صورت $E_\pi\left[\phi(\theta)\right]=\int\phi(\theta)\pi\left(\theta|x\right)d\theta$ روبرو می‌شویم که در آن $\pi\left(\theta|x\right)$ تابع چگالی احتمال پسین θمی‌باشد. θ ممکن است برداری با بُعد بالا باشد که منجر به بروز مشکلاتی در حل انتگرال ذکر شده گردد. بالا بودن بُعد بردار θ حتی با استفاده از تقریبهاپیشنهاد شده، به عنوان مثال، تقریب لیندلی(منبع [۹]) و تقریب تیرنی و کادان (منبع [۱۸]) حل انتگرال فوق را می‌تواند با مشکل مواجه سازد و این مشکل ممکن است به وضوح زمانی که توزیع مورد نظر، توزیع آمیخته باشد ایجاد شود.

الگوریتم زنجیره مارکف – مونت – کارلو(Markov chain Monte Carlo) به عنوان الگوریتم‌های متروپلیس – هاستینگ(Metropolis – Hasting algorithm) (و در حالت خاص نمونه‌گیر گیبس(Gibbs sampler) در آمار بسیار کاربرد دارد و علت آن است که این نامگذاری پس از متروپلیس و همکاران (منبع [۱۱]) و هاستینگ (منبع [۸])صورت گرفته است و در اوایل قرن نوزدهم کاربرد آن در آمار آغاز شده است. چون این تقریبها از مقادیر پیشین نمونه برای تولید تصادفی نمونه بعدی استفاده می‌کنند و یک زنجیره مارکف تولید می‌کنند به این نام شناخته می‌شوند.

برای استفاده از روش مونت – کارلو مارکفی در محاسبه برآوردهای بیزی از تابع $\phi(\theta)$(تابع توزیع بردار پارامترهای θ)، ابتدا باید توجه داشت که مشکل اساسی در حل انتگرال $E_{\pi}\left[\phi(\theta)\right]=\int\phi(\theta)\pi(\theta|x)d\theta$ فرض آن است که $\int\left|\phi(\theta)\right|\pi(\theta|x)d\theta<\infty$. اگر بتوان نمونه‌های $\theta^{(1)},...,\theta^{(N)}$ را از تابع چگالی احتمال پسین $\pi(\theta|x)$ کشید آنگاه با استفاده از مجموع مونت – کارلو می‌توان به صورت میانگین برآورد مورد نظر را به دست آورد :

$$\hat{\phi}_N=\frac{1}{N}\sum_{i=1}^{N}\phi\left(\theta^{(i)}\right)$$

اگر با استفاده از زنجیره مارکف (نامنظم،کاهش ناپذیر و دارای توزیع پایا با تابع چگالی احتمال $\pi(\theta|x)$) نمونه‌ها تولید گردند، آنگاه با استفاده از قضیه تحویل ناپذیری،

$$\hat{\phi}_N\underset{N\to\infty}{\to}E_{\pi}\left[\phi(\theta)\right]$$

برآورد $\hat{\phi}_N$ میانگین تحویل ناپذیری نامیده می‌شود.

همچنین برای بسیاری از زنجیره‌ها، اگر واریانس $\phi(\theta)$ متناهی باشد، قضیه حد مرکزی ثابت می‌گردد و به صورت هندسی همگرایی اتفاق می‌افتد. این روش شبیه سازی در فصل اول مورد بررسی قرار خواهد گرفت.

فصل دوم

روش شبیه سازی زنجیره مارکف – مونت – کارلو

در این فصل به بررسی شبیه سازی به روش زنجیره مارکف– مونت – کارلو، که از آن به روش MCMC(Markov Chain Monte Carlo) یاد می‌شود، و چگونگی پیدایش این روش می‌پردازیم.

۲–۱ مارکف و زنجیره‌ها

استفاده از روش زنجیره مارکف– مونت – کارلو (MCMC) به منظور ارزیابی کمیت‌های انتگرالی گستره‌ای حداقل ۵۰ ساله دارد. آغاز آن توسط گیمان و گیمان (منبع[۷]) با انتشار «مقدمه‌ای از نمونه‌گیر گیبس»(Gibbs sampler) در سال ۱۹۸۴ به عنوان یک روش برای به دست آوردن کمیت‌های پسین مشکل در فرایندهای تجدید تصویر بود و پس از آن با انتشار مقاله «راهنمای انتگرال‌گیری» توسط گلفاند و اسمیت(منبع[۵]) در سال ۱۹۹۰، سرعت انتشار مقالات مونت – کارلو مارکفی به صورت نمایی رشد کرد. حال اینکه گذشته این توسعه‌ها، به پیدایش تولید داده‌ها توسط متروپلیس و همکاران در سال ۱۹۵۳ باز می‌گردد.

تفاوتی عمده‌ای که در بین روش‌های شبیه سازی مونت – کارلو و شبیه سازی زنجیره مارکف– مونت – کارلو ایجاد می‌گردد ساختار وابسته بین مقادیر شبیه سازی شده متوالی مشخص شده می‌باشد. روش‌های مونت – کارلو مجموعه‌ای از مقادیر شبیه سازی شده مستقل، مطابق با توزیع احتمال خواسته شده را تولید می‌کنند در حالی که روش‌های زنجیره مارکف– مونت – کارلو مقادیری را تولید می‌کنند که هر مقدار وابستگی اندکی به مقدار ماقبل خود را دارند. مفهوم پایه‌ای آن است که این زنجیره یک بار به اندازه و طول کافی شروع می‌شود و زنجیره راه را برای رسیدن به توزیع پسین مورد نظر خواسته شده پیدا می‌کند و می‌توان این توزیع را با در نظر گرفتن زنجیره‌ای سرگردان حول آن خلاصه نمود، بنابراین به طور خلاصه آماره‌هایی از مقادیر ثبت شده تولید می‌شود.

۲–۱–۱) زنجیره مارکف چیست؟

در ابتدا تعریف مفاهیم اولیه در زمینه زنجیره‌های مارکف نیاز می‌باشد.

- **فرآیند تصادفی**

یک فرایند تصادفی مجموعه‌ای متوالی از کمیت‌های تصادفی تعریف شده بر فضای معلوم و یکسان Θ می‌باشد و به صورت $\left\{\theta^{[t]}: t \in T\right\}$ نمایش داده می‌شود. معمولاً، اما نه ضرورتاً، T مجموعه‌ای از اعداد صحیح مثبت متوالی، زوج فاصله‌های زمانی به صورت $\left\{\theta^{[t=0]}, \theta^{[t=1]}, \theta^{[t=2]}, \ldots\right\}$ می‌باشد.

همچنین تعریفی دیگر از فرایند تصادفی را با استفاده از تعریف بیلینگزلی(منبع[۱۳] ص ۴۸۲) در مورد فرایند تصادفی به صورت زیر می‌توان ارائه کرد :

«یک فرایند تصادفی یک مجموعه $\left[X_t : t \in T\right]$ از متغیرهای تصادفی بر فضای احتمال $(\Omega, \Im, P)$ می‌باشد.».

با استفاده از روش زنجیره مارکف– مونت – کارلو فقط علاقه‌مند به این نوع محدودیت از فرایندهای تصادفی هستیم. یک فرایند تصادفی همچنین باید نسبت به تکیه‌گاه Θ تعریف شود، که برد مقادیر ممکن θ را معین کند. این تکیه‌گاه با توجه به اینکه متغیر مورد علاقه چگونه اندازه‌گیری شده باشد می‌تواند گسسته یا پیوسته باشد.

یک زنجیره مارکف، یک فرایند تصادفی با این ویژگی که هر مکان خاص در دنباله، $\theta^{[t]}$، فقط به مقدار ماقبل زنجیره، $\theta^{[t-1]}$، وابسته است، می‌باشد و بنابراین به طور شرطی از تمام مقادیر ماقبل‌تر، $\theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}$ مستقل است که می‌توان آن را به صورت زیر بیان کرد :

$$P\left(\theta^{[t]} \in A \mid \theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}, \theta^{[t-1]}\right) = P\left(\theta^{[t]} \in A \mid \theta^{[t-1]}\right) \quad (۱–۲)$$

که در آن A هر مجموعه مشخص (یک پیشامد یا محدوده‌ای از پیشامدها) بر فضای کل حالات می‌باشد؛ بنابراین یک زنجیره مارکف، حول تکیه‌گاه سرگردان باقی می‌ماند. این ویژگی به نتایج مطلوبی برای تولید نمونه‌های بزرگ از توزیع‌های حدی خواسته شده از زنجیره می‌رسد زیرا در نهایت زنجیره ناحیه‌ای از

تکیه‌گاه که بالاترین چگالی با آن را داشته باشد پیدا می‌کند و می‌تواند نمونه‌ای از این توزیع که استقلال کمی را داشته باشد تولید کند. این مقادیر، مقادیر نمونه‌ای هستند که می‌توان از آن‌ها برای توزیع پسین مورد علاقه خواسته شده استفاده می‌شود.

یک ارتباط بنیادی بین فرایندهای انتقال و تعریف احتمالات حرکت به سایر نقاط در تکیه‌گاه در جریان حرکت زنجیره وجود دارد. ساده‌ترین راه برای تفکر در مورد این ساختار در تعریف انتقال هسته مرکزی، K، به عنوان مکانیسم کلی برای شرح دادن احتمال حرکت به دیگر مکان‌های خاص بر اساس جریان زنجیره می‌باشد.(منبع[۱۵] ص ۱۴۱) فایده این نکته در این است که استنباط کردن در هر دو فضای گسسته و پیوسته یکسان است. انتقال هسته مرکزی نیاز دارد که $K\left(\theta,A\right)$ را به عنوان اندازه احتمال برای همهٔ نقاط θ در تکیه‌گاه به مجموعه $A\in\Theta$ تعریف کرد؛ بنابراین $K\left(\theta,A\right)$ نگاشت‌های بالقوه انتقال پیشامدها به احتمال رخداد آن‌ها است.

زمانی که تکیه‌گاه گسسته است K یک ماتریس نگاشت k×k، برای K عنصر گسسته در A است، که در آن هر ستون تعریف احتمال یک تغییر وضعیت از اولین دوره به تمام وضعیت‌های ممکن می‌باشد :

$$P_A\begin{pmatrix} p\left(\theta_1,\theta_1\right) & \cdots & p\left(\theta_1,\theta_k\right) \\ \vdots & & \vdots \\ p\left(\theta_k,\theta_1\right) & \cdots & p\left(\theta_k,\theta_k\right) \end{pmatrix} \quad (۲-۲)$$

که در آن سطرها نشان می‌دهند که زنجیره در این ردیف است و ستون‌ها نشانگر آن هستند که زنجیره به ردیف بعدی رفته است. مجموع سطرهای P_A برابر یکمی‌گردد و تابع چگالی شرطی را تعریف می‌کنند و تمام آن‌ها برای مقدار شروع یکسان مشخص شده می‌باشند و هر نتیجه ممکن در تکیه‌گاه را پوشش می‌دهند. به عنوان مثال برای سطر i – ام داریم :

$$\sum_{j=1}^{k} p\left(\theta_i,\theta_j\right)$$

هر عنصر ماتریس یک احتمال خوش رفتار $p\left(\theta_i,\theta_j\right)\geq 0;\forall i,j\in A$ می‌باشد.

زمانی که تکیه‌گاه پیوسته است آنگاه K یک تابع چگالی احتمال شرطی است : $f\left(\theta|\theta_i\right)$ و به درستی مفهوم احتمال را برای تمام $\theta\in A$ در مکان θ_i – ام معلوم، تعریف می‌کند.

یک ویژگی مهم از تغییر هسته مرکزی آن است که تغییر احتمالات بین دو وضعیت انتخاب شده برای تعداد دلخواه m گام را با ضرب نمودن می‌توان محاسبه کرد. برای مثال احتمال انتقال از وضعیت $\theta_i = x$ در زمان صفر به وضعیت $\theta_j = y$ در دقیقاً m مرحله به وسیله مجموع ضرب‌های زیر به دست می‌آید :

$$p^m\left(\theta_i^{[0]} = x \mid \theta_j^{[m]} = y\right) = \underbrace{\sum_{\theta_1}\sum_{\theta_2}\dots\sum_{\theta_{m-1}}}_{\text{تمام راههای ممکن}} \underbrace{\mathrm{p}\left(\theta_i,\theta_1\right)p\left(\theta_1,\theta_2\right)\dots p\left(\theta_{m-1},\theta_j\right)}_{\text{تغییرات حاصله}} \qquad (۲-۳)$$

بنابراین $p^m\left(\theta_i^{[0]} = x \mid \theta_j^{[m]} = y\right)$ یک ماتریس انتقال تصادفی نیز می‌باشد و این ویژگی عیناً برای تمام زنجیره‌های گسسته ثابت است و برای زنجیره‌های مارکف پیوسته فقط با اندکی تغییر سیگما به انتگرال به دست می‌آید.

ایده اصلی در رابطه (۲– ۳) آن است که احتمال کل انتقال یافتن از x به y حاصل کلیهٔ مراحل میانی مورد نیاز می‌باشد. به همین دلیل باید سیگما را بر تمام راه‌های ممکن رسیدن به y از x در نظر گرفت.

۲–۱–۲) معادلات چپمن – کلموگروف

رابطه (۲– ۳) منجر به این نکته می‌گردد که چگونه زنجیره احتمالات به یکدیگر متصل می‌شوند. معادلات خاص چپمن – کلموگروف چگونگی هم مرز بودن نتایج موفقیت‌ها با یکدیگر به صورت احتمالی را ارائه می‌دهند.

برای هر دو فضای وضعیت (تکیه‌گاه) گسسته و پیوسته سمت چپ رابطه (۲– ۳) را به صورت خلاصه ارائه می‌دهیم :

$$p^{m_1+m_2}\left(x,y\right) = \begin{cases} \displaystyle\sum_{\text{تمام } z\text{ها}} p^{m_1}\left(x,z\right)p\left(z,y\right) \\ \displaystyle\int_{z\ \text{بـرد}} p^{m_1}\left(x,z\right)p^{m_2}\left(z,y\right)dz \end{cases} \qquad (۲-۴)$$

معادلات چپمن – کلموگروف مخصوصاً برای حالت گسسته ظریف‌تر هستند زیرا رابطه (۱ – ۴) را می‌توان به صورت دنباله‌ای از حاصل ضرب‌های ماتریس انتقال ارائه داد :