بسم الله الرحمن الرحيم

دانشکده ریاضی و رایانه

بخش آمار

پایان نامه تحصیلی برای دریافت درجه کارشناسی ارشد

رشته آمار ریاضی

توزیع های چندمتغیره بریده شده $t_{(\nu)}$ و لاپلاس

مؤلف :

سمیه نیک زاده عباسی

استاد راهنما :

دکترعلیرضا عربپور

استاد مشاور :

دکتر احد جمالیزاده

بهمن ماه ۱۳۹۱

این پایان نامه

به عنوان یکی از شرایط احراز کارشناسی ارشد

به

بخش آمار – دانشکده ریاضی و رایانه

دانشگاه شهید باهنر کرمان

تسلیم شده است و هیچگونه مدرکی به عنوان فراغت از تحصیل دوره مزبور شناخته نمی شود.

دانشجو :سمیه نیک زاده عباسی

استاد راهنما : دکتر علیرضا عربپور

استاد مشاور: دکتر احد جمالیزاده

داور ۱ : دکتر محسن مددی

داور ۲ : دکتر وحید امیرزاده

نماینده تحصیلات تحمیلی دانشگاه: دکتر محسن رضا پور

تقدیم به پدر و مادر عزیزم

که عشق و زندگی را به من آموختند.

به پاس زحمات بی دریغشان.

و تقدیم به آنان که قلبی مهربان و بخشنده دارند.

تشکر و قدردانی

سپاس خدایی را که درگذر از تمام مراحل و مشکلات زندگی یار و یاور ماست و بهترین هدایت‌گر اوست و با تشکر از استادید دلسوز و پر تلاشم آقایان دکتر عرب پور و دکتر جمالی زاده که با رهنمودهای دلسوزانه و بی‌دریغشان، مرا در تدوین این پایان نامه یاری نمودند و راه درست تحقیق و پژوهش را به من نشان دادند. همچنین بر خود لازم می دانم از آقای دکتر مددی و آقای دکتر امیرزاده به جهت تقبل داوری این پایان نامه تشکر نمایم و در پایان از خانواده عزیزم، به خصوص پدر و مادر مهربانم به خاطر صبر و بردباریشان سپاسگزاری می‌نمایم.

چکیده :

توزیع‌های بریده شده کاربرد بسیار وسیعی در علوم مختلف دارند. طرح‌ها و پیشنهاد های زیادی برای گسترش چندین توزیع بریده شده خاص ارائه شده است، اما درمورد توزیع‌های چند متغیره t ولاپلاس بریده شده بحث اندکی مطرح شده است. گشتاورهای یک توزیع بریده شده دربسیاری از موقعیت‌های کاربردی مورد نیاز می‌باشند. در واقع دو گشتاوراول، مخصوصاً میانگین و واریانس از اهمیت ویژه ای برخوردار هستند.

در این پایان نامه، توزیع‌های نرمال، t و لاپلاس بریده شده را در حالت‌های یک متغیره و چند متغیره بررسی می کنیم و عبارات صریحی را برای گشتاورهای این توزیع ها ارائه می دهیم. همچنین با استفاده از روش نمونه گیری گیبز و روش نمونه گیری برشی به نمونه گیری ازاین توزیع‌ها، می‌پردازیم.

واژه‌های کلیدی: توزیع بریده شده، توزیع نرمال بریده شده، توزیع t بریده شده، توزیع لاپلاس بریده شده، نمونه گیری گیبز، نمونه گیری برشی.

فهرست مطالب

فصل اول

مقدمات

۱-۱ مقدمه

از آنجـایی کـه جمـع آوری و مـدل‌سـازی داده‌هـا نقـش اساسـی در تحقیقـات علمـی دارد، توزیع‌های چندمتغیره در تحلیل‌های آماری مورد توجه واقع شده‌اند. در واقع توزیع‌های چندمتغیره در مدل‌سازی‌های مربوط به نتایج آزمایش‌های تصادفی به‌کار می‌روند.

با وجود نقش مهم توزیع نرمال چند متغیره در آمار، علاقه زیادی به ساخت توزیع‌های غیر نرمـال[1] وجود دارد. در سال‌های اخیر پیشرفت‌های قابل ملاحظه‌ای در زمینه ساخت توزیع‌های چند متغیـره چوله شده است. این توزیع‌ها، کاربرد زیادی در حل مشکلات موجود در علوم مهندسی، اقتصاد و پزشکی دارند که گنتون[2] (۲۰۰۴) برای اولین بار نتایج و برنامه های کاربردی در این زمینه را مطرح کرد.

یک روش برای ساختن توزیع‌های غیرنرمال، استفاده از مدل‌های گزینشی[3] می‌باشد. در ایـن روش مدلی را درنظر می گیریم که بردار تصادفی $\mathbf{X} \in \mathbb{R}^p$ دارای تابع چگالی احتمال متقـارن، $f(\boldsymbol{x};\boldsymbol{\theta})$ است و $\boldsymbol{\theta}$ یک بردار از پارامترهای نامعلوم باشد.

در آنالیز آماری، فرض می کنیم که یک نمونه تصادفی $\mathbf{X}_1,\dots,\mathbf{X}_\mathrm{n}$ از توزیع $f(\boldsymbol{x};\boldsymbol{\theta})$ مـی‌توانـد به منظور استنباط در مورد بردار $\boldsymbol{\theta}$ مشاهده شود. با این حال در بسیاری از شرایط ممکن است یک نمونه تصادفی در دسترس نباشد، به عنوان مثال هنگامی که گرفتن نمونه دشوار و یا هزینـه زیـادی داشته باشد. اگر تابع چگالی احتمال به وسیله تابع وزنی نامنفی $w(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\eta})$ تغییـر شـکل دهـد، بـه طوری که $\boldsymbol{\eta}$ نشان دهنده یک بردار از پارامترهای نامعلوم اضافی باشد، سـپس داده مشـاهده شـده یک نمونه تصادفی از توزیع وزنی با تابع چگالی احتمال زیر خواهد بود:

$$f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\eta})\,\frac{w(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\eta})}{\mathrm{E}\{w(\mathbf{X};\boldsymbol{\theta},\boldsymbol{\eta})\}}\,. \qquad (۱-۱)$$

[1] Nonnormal Distributions
[2] Genton
[3] Selection Model

در حالت خاص اگر داده مشاهده شده فقط از یک بخش از جامعه مورد نظر انتخاب شود، آنگاه (۱-۱) یک مدل گزینشی می‌باشد.

در این فصل توزیع‌های گزینشی را مورد توجه قرار می‌دهیم. به این ترتیب که دربخش ۱-۲ به بیان تعاریف پایه می‌پردازیم. در بخش ۱-۳ توزیع های گزینشی را تعریف می‌کنیم و در بخش ۱-۴ ویژگی‌های این توزیع‌ها را مورد بررسی قرار می‌دهیم. در بخش ۱- ۵ توزیع‌های بیضوی گزینشی و در نهایت در بخش ۱-۶ توزیع نرمال گزینشی را تعریف می‌کنیم.

۱-۲ تعاریف پایه

تعریف ۱-۲-۱: بردار تصادفی $\mathbf{X}$ را دارای توزیع نرمال p متغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس $\boldsymbol{\Sigma}$ گوییم وآن را با نماد $\mathbf{X} \sim \mathrm{N_p}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ نشان می‌دهیم، اگر دارای تابع چگالی احتمال زیر باشد:

$$f(\boldsymbol{x}) = 2\pi^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

که $\boldsymbol{\Sigma}$ یک ماتریس معین مثبت است و $\boldsymbol{\Sigma}^{-1}$ معکوس ماتریس $\boldsymbol{\Sigma}$ می‌باشد.

تعریف ۱-۲-۲: بردار p بعدی تصادفی $\mathbf{X}$ را دارای توزیع t با بردار میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس $\boldsymbol{\Sigma}$ و ν درجه آزادی گوییم، اگر دارای تابع چگالی احتمال زیر باشد وآن را با نماد $\mathbf{X} \sim \mathrm{t_p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ نشان می‌دهیم.

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{p}{2}}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\left(1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right),^{-\frac{(\nu+p)}{2}}$$

تعریف ۱-۲-۳: توزیع نرمال p متغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس $\boldsymbol{\Sigma} > 0$ ($\boldsymbol{\Sigma}$ ماتریس معین مثبت است) برروی بیضی‌هایی با مرکز $\boldsymbol{\mu}$ ثابت است اگر

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = k.$$

کلاس کلیه توزیع‌ها با این ویژگی را کلاس توزیع‌های بیضوی[1] می‌نامیم که تابع چگالی آن به صورت زیر می‌باشد:

$$f_{EC_p}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},h) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, h\{(\boldsymbol{x}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\},$$

به طوری‌که h تابع مولد توزیع بیضوی نامیده می‌شود.

۱-۳ توزیع‌های گزینشی[2]

تعریف ۱-۳-۱: فرض کنید $\mathbf{U} \in \mathbb{R}^{\mathbf{q}}$ و $\mathbf{V} \in \mathbb{R}^{\boldsymbol{p}}$ دو بردار تصادفی باشند و $\mathbf{C}$ یک زیر مجموعه از $\mathbb{R}^{\mathrm{q}}$ باشد. آرلانو – واله و دیگران[3] (۲۰۰۶) توزیع‌های گزینشی را به صورت توزیع شرطی $\mathbf{V}$ به شرط $\mathbf{U} \in \mathbf{C}$ تعریف کردند. یک بردار تصادفی p بعدی $\mathbf{X}$ دارای توزیع گزینشی چند متغیره از $\mathbf{C}$ و $\mathbf{V}$ و $\mathbf{U}$ است اگر

$$\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in \mathbf{C}).$$

این توزیع را با نماد $\mathbf{X} \sim \mathrm{SLCT}_{\mathrm{p,q}}$ نشان می‌دهیم.

اگر $\mathbf{C} = \mathbb{R}^{\mathbf{q}}$ باشد در این حالت هیچ گزینشی وجود ندارد بنابراین فرض می‌کنیم که $\mathrm{P}(\mathbf{U} \in \mathbf{C}) > 0$ باشد.

اگر $\mathbf{U} = \mathbf{V}$ باشد آنگاه بردار $\mathbf{X}$ دارای توزیع بریده شده می باشد.

---

[1] Elliptically Contoured Density (EC)

[2] Selection Distribution (SLCT)

[3] Arellano – Valle et al

۱-۴ ویژگی‌های اصلی توزیع گزینشی

به‌راحتی می‌توان ویژگی‌های اصلی بردار تصادفی $\mathbf{X}$ که دارای توزیع گزینشی می‌باشد را با بررسی خصوصیات بردارهای تصادفی $\mathbf{V}$ و $\mathbf{U}$ مورد مطالعه قرار داد. در واقع اگر $\mathbf{V}$ دارای تابع چگالی احتمال f_{V} باشد آنگاه $\mathbf{X}$ دارای تابع چگالی زیر است:

$$f(\boldsymbol{x}) = f_{\mathrm{V}}(\boldsymbol{x})\ \frac{\mathrm{P}(\,\mathbf{U} \in \mathbf{C}\ |\ \mathbf{V}{=}\boldsymbol{x})}{\mathrm{P}(\mathbf{U} \in \mathbf{C})}. \qquad (۱-۲)$$

بیشتر ویژگی‌های بردار تصادفی $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in \mathbf{C})$ را می‌توان مستقیماً از روی تعریف نیز بررسی کرد، به عنوان نمونه سه حالت زیر را در نظرمی گیریم.

حالت اول: وقتی‌که $(\mathbf{U}\,,\mathbf{V})$ دارای تابع چگالی توأم $f_{\mathbf{U},\mathbf{V}}$ باشد آنگاه $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in \mathbf{C})$ دارای تابع چگالی زیر است:

$$f(\boldsymbol{x}) = \frac{\int_{\mathbf{C}} f_{\mathbf{U},\mathbf{V}}(\boldsymbol{u},\boldsymbol{x})d\boldsymbol{u}}{\int_{\mathbf{C}} f_{\mathbf{U}}(\boldsymbol{u})d\boldsymbol{u}}, \qquad (۱-۳)$$

به طوری‌که $f_{\mathbf{U}}$ توزیع حاشیه‌ای متغیر $\mathbf{U}$ می‌باشد.

حالت دوم: برای‌هر تابع برل g، داریم $\mathrm{g}(\mathbf{X}) \overset{d}{=} (\mathrm{g}(\mathbf{V})|\ \mathbf{U}\ \in \mathbf{C})$. علاوه براین $(\mathrm{g}(\mathbf{V})|\ \mathbf{U}\ \in \mathbf{C})$ به وسیله تبدیل $(\mathbf{U},\mathbf{V})$ به $(\mathbf{U}\,,\mathrm{g}(\mathbf{V}))$ تعیین می گردد.

آنگاه تابع چگالی $\mathrm{g}(\mathbf{X})$ را می‌توان با قرار دادن $\mathrm{g}(\mathbf{X})$ به جای $\mathbf{X}$ و $\mathrm{g}(\mathbf{V})$ به جای $\mathbf{V}$ در (۱-۲) و (۱-۳) بدست آورد.

حالت سوم: با توجه به توزیع توأم $(\mathbf{X}_1\ ,\mathbf{X}_2) \overset{d}{=} [(\mathbf{V_1},\mathbf{V_2})|\mathbf{U} \in \mathbf{C}]$، می‌توان توزیع گزینشی شرطی $(\,\mathbf{X}_2|\,\mathbf{X}_1 = \boldsymbol{x}_1\,)$ را با استفاده از $(\mathbf{V_2}|\,\mathbf{V_1} = \boldsymbol{x_1}\,,\ \mathbf{U} \in \mathbf{C}\,)$ به دست آورد. با استفاده از رابطه (۱-۲) تابع چگالی به صورت زیر خواهد بود:

$$f_{\mathbf{X_2}|\,\mathbf{X_1}=x_1}(\boldsymbol{x_2}) = f_{\mathbf{V_2}|\mathbf{V_1}=x_1}(\boldsymbol{x_2})\ \frac{\mathrm{P}(\,\mathbf{U} \in \mathbf{C}\,|\,\mathbf{V_1}{=}\boldsymbol{x_1}\,,\mathbf{V_2}{=}\boldsymbol{x_2}\,)}{\mathrm{P}(\,\mathbf{U} \in \mathbf{C}\,|\,\mathbf{V_1}{=}\boldsymbol{x_1}\,)}.$$

همچنین افراز $\mathbf{U} = (\mathbf{U_1}, \mathbf{U_2})$ را در نظر می‌گیریم به طوری‌که $(\mathbf{U_1}, \mathbf{V_1})$ و $(\mathbf{U_2}, \mathbf{V_2})$ مستقل هستند و فرض می‌کنیم که $\mathbf{C} = \mathbf{C_1} \times \mathbf{C_2}$ به طوری‌که $\mathbf{U} \in \mathbf{C}$ هم ارز با $\mathbf{U_1} \in \mathbf{C_1}$ و $\mathbf{U_2} \in \mathbf{C_2}$ می‌باشد. بنابراین می‌توان نتیجه گرفت که $\mathbf{X_1} \overset{d}{=} (\mathbf{V_1} \mid \mathbf{U} \in \mathbf{C}) \overset{d}{=} (\mathbf{V_1} \mid \mathbf{U_1} \in \mathbf{C_1})$ و $\mathbf{X_2} \overset{d}{=} (\mathbf{V_2} \mid \mathbf{U} \in \mathbf{C}) \overset{d}{=} (\mathbf{V_2} \mid \mathbf{U_2} \in \mathbf{C_2})$ هم ارز و از هم مستقل می‌باشند. آرلانو – واله و گنتون[1] (۲۰۰۵) و آرلانو – واله و آزالینی[2] (۲۰۰۶) جزییات بیشتری را هنگامی که $\mathbf{C}$ به صورت یک مجموعه خاص تعریف شده باشد، بیان کرده اند.

۱-۵ توزیع‌های بیضوی گزینشی

یک خانواده معروف از توزیع‌های گزینشی درحالتی ساخته می‌شود که $(\mathbf{U}, \mathbf{V})$ دارای توزیع توأم بیضوی چند متغیره به صورت زیر باشند:

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim \mathrm{EC_{q+p}}\left(\xi = \begin{pmatrix} \boldsymbol{\xi}_\mathbf{U} \\ \boldsymbol{\xi}_\mathbf{V} \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_\mathbf{U} & \boldsymbol{\Delta}^\mathrm{T} \\ \boldsymbol{\Delta} & \boldsymbol{\Omega}_\mathbf{V} \end{pmatrix}, h^{(q+p)}\right).$$

به طوری‌که $\boldsymbol{\xi}_\mathbf{U} \in \mathbb{R}^\mathrm{q}$ و $\boldsymbol{\xi}_\mathbf{V} \in \mathbb{R}^p$ بردارهای مکان هستند، $\boldsymbol{\Omega}_\mathbf{U} \in \mathbb{R}^{q\times q}$ و $\boldsymbol{\Omega}_\mathbf{V} \in \mathbb{R}^{p\times p}$ و $\Delta \in \mathbb{R}^{p\times q}$ ماتریس‌های کوواریانس می‌باشند علاوه براین $h^{(q+p)}$ نیز تابع مولد چگالی می‌باشد. توزیع بیضوی گزینشی را با نماد $\mathrm{SLCT-EC_{p,q}}\ (\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, \mathbf{C})$ نشان می دهیم.

در حالت کلی می‌توان توزیع انتخابی را به وسیله تعیین توزیع کناری برای $\mathbf{V}$ و توزیع شرطی $(\mathbf{U} \mid \mathbf{V} = \boldsymbol{x})$ بدست آورد. برای مثال تابع چگالی توزیع $\mathrm{SLCT} - \mathrm{E}C_{\mathrm{p,q}}\ (\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, \mathbf{C})$ به همین صورت به دست می آید.

$$\mathbf{V} \sim \mathrm{EC}_p\left(\boldsymbol{\xi}_\mathbf{V}, \boldsymbol{\Omega}_\mathbf{V}, h^{(p)}\right),$$

$$(\mathbf{U} \mid \mathbf{V} = \boldsymbol{x}) \sim \mathrm{EC}_q\left(\boldsymbol{\xi}_\mathbf{U} + \Delta^\mathbf{T}\boldsymbol{\Omega}_\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\xi}_\mathbf{V}), \boldsymbol{\Omega}_\mathbf{U} - \Delta^T\boldsymbol{\Omega}_\mathbf{V}^{-1}\Delta, h_{v(\mathrm{x})}^{(q)}\right). \quad (۱-۵)$$

که در آن $h_{v(\mathbf{x})}^{(q)}(\boldsymbol{u}) = h^{(q+p)}(\boldsymbol{u} + v(\mathbf{x}))/h^{(p)}(v(\mathbf{x}))$ تابع مولد شرطی می باشد.

---
[1] Arellano – Valle and Genton
[2] Arellano – Valle and Azzalini

همچنین $v(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\xi}_{\mathbf{V}})^{T}\boldsymbol{\Omega}_{\mathbf{V}}^{-1}(\mathbf{x}-\boldsymbol{\xi}_{\mathbf{V}})$ است.

بردار تصادفی r بعدی $\mathbf{Y}$ دارای توزیع بیضوی با تابع چگالی زیر است که آن را با نماد $\mathbf{Y} \sim \mathrm{EC}_r\left(\boldsymbol{\xi}_{\mathbf{Y}}, \boldsymbol{\Omega}_{\mathbf{Y}}, h^{(r)}\right)$ نشان می دهیم:

$$f_r\left(\boldsymbol{y}; \boldsymbol{\xi}_Y, \boldsymbol{\Omega}_{\mathbf{Y}}, h^{(r)}\right) = |\boldsymbol{\Omega}_{\mathbf{Y}}|^{-\frac{1}{2}} h^{(r)}\left((\mathbf{y}-\boldsymbol{\xi}_{\mathbf{Y}})^{T}\boldsymbol{\Omega}_{\mathbf{y}}^{-1}(\mathbf{y}-\boldsymbol{\xi}_{\mathbf{Y}})\right).$$

و $\mathrm{P}(\mathbf{Y} \in \mathbf{C})$ را به صورت $\bar{F}_r\left(\mathbf{C};\ \boldsymbol{\xi}_{\mathbf{Y}}, \boldsymbol{\Omega}_{\mathbf{Y}}, h^{(r)}\right)$ نشان می‌دهیم. به طوری که $\bar{F}_r\left(\mathbf{C};\ \boldsymbol{\xi}_{\mathbf{Y}}, \boldsymbol{\Omega}_{\mathbf{Y}}, h^{(r)}\right) = \int_{\mathbf{C}} f_r\left(\boldsymbol{y}; \boldsymbol{\xi}_Y, \boldsymbol{\Omega}_{\mathbf{Y}}, h^{(r)}\right) d\boldsymbol{y}$ می باشد. آنگاه تابع چگالی $\mathbf{X} \overset{d}{=} (\mathbf{V} \mid \mathbf{U} \in \mathbf{C})$ به صورت زیر است:

$$f(\boldsymbol{x}) = f_p\left(\boldsymbol{x};\ \boldsymbol{\xi}_{\mathbf{V}}, \boldsymbol{\Omega}_{\mathbf{V}},\ h^{(p)}\right)\ \frac{\overline{\mathrm{F}}_{\mathrm{q}}\left(\mathbf{C}\,;\, \boldsymbol{\xi}_{\mathbf{U}}+\Delta^{\mathbf{T}}\boldsymbol{\Omega}_{\mathbf{V}}^{-1}(\mathbf{x}-\boldsymbol{\xi}_{\mathbf{V}}),\, \boldsymbol{\Omega}_{\mathbf{U}}-\Delta^{T}\boldsymbol{\Omega}_{\mathbf{V}}^{-1}\Delta,\, h_{v(\mathbf{x})}^{(q)}\right)}{\overline{\mathrm{F}}_{\mathrm{q}}\left(\mathbf{C}\,;\, \boldsymbol{\xi}_{\mathbf{U}}, \boldsymbol{\Omega}_{\mathbf{U}},\, h^{(q)}\right)}. \quad (۱-۶)$$

می‌توان $f(\boldsymbol{x})$ را با استفاده از رابطه (۱-۳) به صورت زیر نوشت:

$$f(\boldsymbol{x}) = \frac{\int_{\mathbf{c}} f_{q+p}(\boldsymbol{u}, \boldsymbol{x}\,; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})\, d\boldsymbol{u}}{\int_{\mathbf{c}}^{-} f_q(\boldsymbol{u}\ ; \boldsymbol{\xi}_{\mathbf{U}}, \boldsymbol{\Omega}_{\mathbf{U}},\ h^{(q)})\, d\boldsymbol{u}}.$$

همچنین می‌توان با توجه به انتخاب $h^{(q+p)}$ تابع چگالی توزیع‌های بیضوی گزینشی متفاوتی را بدست آورد. برای مثال وقتی $h^{(q+p)}(\boldsymbol{u}) = c(q+p, \nu)\{1+\boldsymbol{u}\}^{-(q+p+\nu)/2}$ باشد، تابع چگالی توزیع t گزینشی با ν درجه آزادی خواهیم داشت که به صورت زیر است:

$$f(\boldsymbol{x}) = t_p(\boldsymbol{x}\,; \boldsymbol{\xi}_{\mathbf{V}}, \Omega_{\mathbf{V}}, \nu)\ \frac{\overline{\mathrm{T}}_{\mathrm{q}}\left(\mathbf{C}\,;\Delta^{\mathbf{T}}\boldsymbol{\Omega}_{\mathbf{V}}^{-1}(\boldsymbol{x}-\boldsymbol{\xi}_{\mathbf{V}})+\boldsymbol{\xi}_{\mathbf{U}}, \frac{\nu+v(x)}{\nu+p}[\boldsymbol{\Omega}_{\mathbf{U}}-\Delta^{\mathbf{T}}\boldsymbol{\Omega}_{\mathbf{V}}^{-1}\Delta], \nu+p\right)}{\overline{\mathrm{T}}_{\mathrm{q}}(\mathbf{C}\,; \boldsymbol{\xi}_{\mathbf{U}}, \boldsymbol{\Omega}_{\mathbf{U}}, \nu)}.$$

(۱-۷)

به طوری که $c(r,\nu) = \Gamma[(r+\nu)/2]\nu^{\nu/2}/\Gamma[\nu/2]\,\pi^{r/2}$ و $\mathbf{Y} \sim t_r(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ که $t_r(.\,; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ تابع چگالی توزیع t، r متغیره می باشد و $\overline{\mathrm{T}}_r(\mathbf{C}\,; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \mathrm{P}(\mathbf{Y} \in \mathbf{C})$ است.

۱-۶ توزیع نرمال گزینشی

یکی از مهم‌ترین حالت‌ها در توزیع‌های بیضوی حالتی است که $\mathbf{U}$ و $\mathbf{V}$ دارای توزیع توأم نرمال چند متغیره به صورت زیر باشند:

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim \mathrm{N}_{\mathrm{q+p}}\left(\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{\mathbf{U}} \\ \boldsymbol{\xi}_{\mathbf{V}} \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\mathbf{U}} & \boldsymbol{\Delta}^{\mathbf{T}} \\ \boldsymbol{\Delta} & \boldsymbol{\Omega}_{\mathbf{V}} \end{pmatrix}\right).$$

آنگاه $\mathbf{X} \overset{d}{=} (\mathbf{V}\,|\mathbf{U} \in \mathbf{C})$ دارای توزیع نرمال گزینشی، با نماد $\mathrm{SLCT} - \mathrm{N}_{\mathrm{p,q}}(\boldsymbol{\xi}\,, \boldsymbol{\Omega}, \mathbf{C})$ می باشد که تابع چگالی و تابع مولد گشتاور آن به صورت زیر می باشند:

$$f(\boldsymbol{x}) = \phi_p(\boldsymbol{x}\,;\, \boldsymbol{\xi}_{\mathbf{V}}\,, \boldsymbol{\Omega}_{\mathbf{V}}) \quad \frac{\bar{\Phi}_{\mathrm{q}}(\mathbf{C}\,;\, \boldsymbol{\xi}_{\mathbf{U}} + \Delta^{\mathbf{T}} \boldsymbol{\Omega}_{\mathbf{V}}^{-\mathbf{1}}(\boldsymbol{x} - \xi_{\mathbf{V}})\,, \boldsymbol{\Omega}_{\mathbf{U}} - \Delta^{\mathbf{T}} \boldsymbol{\Omega}_{\mathbf{V}}^{-\mathbf{1}} \Delta)}{\bar{\Phi}_{\mathrm{q}}(\mathbf{C}\,;\, \xi_{\mathbf{U}}\,, \boldsymbol{\Omega}_{\mathbf{U}})} \qquad (۱-۸)$$

$$M_{\mathbf{X}}(\mathbf{s}) = \exp\left\{\mathbf{s}^{\mathrm{T}} \boldsymbol{\xi}_{\mathbf{V}} + \frac{1}{2} \mathbf{s}^{T} \boldsymbol{\Omega}_{\mathrm{V}} \boldsymbol{s}\right\} \quad \frac{\bar{\Phi}_{\mathrm{q}}\left(\mathbf{C}\,;\, \xi_{\mathbf{U}} + \Delta^{\mathbf{T}} \mathbf{s}\,, \boldsymbol{\Omega}_{\mathbf{U}}\right)}{\bar{\Phi}_{\mathrm{q}}(\mathbf{C}\,;\, \xi_{\mathbf{U}}\,, \boldsymbol{\Omega}_{\mathbf{U}})} \qquad (۱-۹)$$

فصل دوم

روش‌های نمونه گیری

۲- ۱ مقدمه

از روش‌های نمونه‌گیری گیبز[1] و الگوریتم متروپلیس[2] که جزو روش‌های‌نمونه‌گیری زنجیره مارکف می‌باشند، اغلب برای گرفتن نمونه از بسیاری توزیع‌ها مانند توزیع‌های پسین استفاده می‌شود. یکی از علل استفاده از روش‌های زنجیره مارکف به حداقل رساندن زمان برای گرفتن نمونه از این توزیع‌ها می‌باشد. (توضیحات بیشتر در پیوست د-۱ ذکر شده است)

روش نمونه‌گیری گیبز یکی از پرکاربردترین روش‌های نمونه‌گیری زنجیره مارکف است که توسط گلفاند و اسمیت[3] (۱۹۹۰) معرفی شده است. فرض کنید می‌خواهیم از یک توزیع n متغیره با تابع چگالی $\mathrm{p}(\boldsymbol{x})$، نمونه گیری کنیم. نمونه گیری گیبز حاصل‌گرفتن نمونه به صورت پی درپی از توزیع‌های شرطی $\mathrm{p}(x_i|x_j)$ می‌باشد، برای هر x_i به شرط اینکه x_j داده شده باشد و $j \neq i$. با تکرار این روند یک زنجیره مارکف تعریف می‌شود که به توزیع مورد نظر همگرا است. این روش نمونه‌گیری را هنگامی می‌توان استفاده کرد که بدانیم چه طور از توزیع‌های شرطی مورد نظر نمونه‌گیری کنیم.

یک روش دیگر برای ساختن نمونه بردار زنجیره مارکف، الگوریتم متروپلیس می‌باشد که با انتخاب تصادفی یک حالت پیشنهادی آغاز می‌شود و از یک توزیع پیشنهادی نمونه‌گیری می‌کند. حالت پیشنهادی، براساس نسبت چگالی‌های احتمال در حالت‌های پیشنهادی وکنونی به عنوان حلقه جدید زنجیره مارکف رد یا پذیرفته می‌شود. اگر حالت پیشنهادی رد شود، حالت بعدی همان حالت قبلی خواهد بود. (توضیحات بیشتر در پیوست د-۲ ذکر شده است)

روش‌های گیبز و متروپلیس کاربردهای زیادی دارند اما نیاز به روش‌های بهتری به منظور اجرای شبیه سازی در سطح گسترده می‌باشد. یکی از این روش‌ها، روش نمونه‌گیری برشی[4] است که توسط نیل[5] (۲۰۰۳) معرفی شده است. پیاده سازی و اجرای روش نمونه‌گیری برشی از روش گیبز ساده تر و کارایی آن نسبت به الگوریتم متروپلیس بیشتر است.

[1] Gibbs Sampling
[2] Metropolis
[3] Gelfand and Smith
[4] Slice Sampling
[5] Neal

از ویژگی‌های اصلی نمونه‌گیری برشی این است که ما می‌توانیم نقاط را به صورت یکنواخت از ناحیه تحت منحنی تابع چگالی نمونه‌برداری کنیم. یک زنجیره مارکف که به این توزیع یکنواخت همگرا خواهد بود از طریق نمونه‌گیری‌های متناوب از فاصله‌های افقی و عمودی اجرا خواهد شد.

اجرا کردن نمونه‌گیری برشی متکی بر نمونه‌گیری گیبز می‌باشد. مارکف با بروز رسانی این طرح به طورگسترده‌ای برای گرفتن نمونه تصادفی ازچگالی‌های حاشیه ای با توجه به توزیع توأم، استفاده کرده است.

روش نمونه‌گیری برشی به راحتی برای توزیع‌های یک متغیره قابل اجرا می‌باشد. همچنین می‌توان با تعمیم دادن این روش از آن برای نمونه‌گیری از توزیع‌های چند متغیره استفاده کرد. گاهی برای گرفتن نمونه از توزیع‌های چند متغیره، روش نمونه‌گیری گیبز و الگوریتم متروپلیس استفاده می شوند با این حال برای اجرای نمونه‌گیری گیبز نیاز به ابداع روش‌های نمونه‌گیری ازتوزیع های غیراستاندارد یک متغیره می‌باشد و برای استفاده از الگوریتم متروپلیس نیز به منظور ایجاد یک نمونه کارا باید یک توزیع پیشنهادی مناسب پیدا کرد.

در این فصل ابتدا روش نمونه‌گیری گیبز را مورد بررسی قرار می‌دهیم سپس روش نمونه‌گیری برشی را در حالت های یک متغیره و چند متغیره بیان می‌کنیم که از آن برای نمونه‌گیری از توزیع‌های بریده شده در فصل چهارم استفاده خواهیم کرد.

۲-۲ نمونه‌گیری گیبز

در سال‌های اخیر مواردی همچون دسترسی مداوم و ارزان قیمت، سرعت بالای محاسبات، باعث تغییر بسیاری از روش‌های آماری شده است. پیشرفت‌های زیادی بر روی روش‌های الگوریتمی صورت گرفته که می‌توان به مواردی مانند الگوریتم EM، دمپستر وهمکاران[1] (۱۹۷۷) یا روش‌های باز نمونه‌گیری مانند بوت استرپ[2] (۱۹۸۲) اشاره کرد. در ادامه روش نمونه‌گیری گیبزکه نوع متفاوتی از روش‌های آماری می‌باشد، را مطرح می‌کنیم.

[1] Dempster
[2] Bootstrap

نمونه‌گیری گیبز با مقاله گلمن و گلمن [1] (۱۹۸۴) که درآن مدل‌های پردازش تصویر مطرح شـد، از محبوبیت زیادی برخوردارگشت. این نمونه‌گیری بر پایه روش‌هایی همچون متروپلیس، روسنبلاد، تـالر و تـالر [2] (۱۹۵۳) اسـتوار مـی‌باشـد. گلفانـد و اسـمیت [3] (۱۹۹۰) بـا بیـان کـردن کاربردهـای نمونه‌گیری گیبز در مسائل رایج آماری، محبوبیت این روش را افـزایش دادنـد. نمونـه‌گیـری گیبـز یک روش برای تولید متغیرهای تصادفی به طـور غیرمسـتقیم از یـک توزیـع، بـدون محاسـبه تـابع چگالی می‌باشد. در این روش از ویژگی‌های مقدماتی زنجیره های مـارکف اسـتفاده مـی‌شـود. بـا استفاده از نمونه‌گیری گیبز می‌توان از انجام محاسبات دشوار صرف نظر کرد و این محاسبات را با یک دنباله از محاسبات ساده تر جایگزین نمود. همچنین این روش تأثیر به سزایی برحل مشـکلات آماری دارد. اگرچه روش گیبز در مدل‌های بیزی کاربرد فراوانی دارد، اما بـرای انجـام محاسـبات کلاسیک نیز مفید خواهد بود.

۲-۳ روش نمونه‌گیری گیبز

فـرض‌کنیـد چگـالی تـوأم $f(x, y_1, \dots, y_p)$ را داشـته باشـیم و بخـواهیم ویژگـی‌هـای تـابع چگالی‌حاشیه‌ای $f(x)$ ماننـد میـانگین و واریـانس را بدسـت آوریـم. تـابع چگـالی حاشـیه ای بـه صورت زیر می باشد:

$$f(x) = \int \dots \int f(x, y_1, \dots, y_p) dy_1 \dots dy_p \,. \tag{۲-۱}$$

شاید استفاده از انتگرال یکی از معمولی‌ترین و ساده‌ترین راه‌ها برای بدست آوردن $f(x)$ باشد اما در بسیاری از موارد محاسبه انتگرال بالا دشوار است و باید از روش‌های عددی استفاده کرد. در چنین مواردی روش نمونه‌گیری‌گیبز یک راه مناسب برای بدست آوردن تابع چگالی $f(x)$ را ارائه می‌دهد. به جای محاسبه مستقیم انتگرال می‌توان با استفاده از روش گیبز و بدون استفاده از تابع چگالی $f(x)$ با شبیه سازی یک نمونه به اندازه کافی بزرگ، میانگین و واریانس یا هرمشخصه دیگر را محاسبه نمود.

---

[1] Gelman and Gelman

[2] Metropolis,Rosenbluth,Taller and Taller

[3] Gelfand and Smith