دانشگاه تربیت مدرس

دانشکده **علوم ریاضی**

رساله دوره دکتری آمار

اثرات تصادفی چوله گاوسی در مدل‌های خطی تعمیم‌یافته فضایی

توسط

فاطمه حسینی

استاد راهنما

دکتر محسن محمدزاده

اسفند ۱۳۸۹

قدردانی

سپاس و ستایش معبود یگانه را که پرتو الطاف بی‌شمارش بر لحظه لحظهٔ زندگی‌ام ساطع و آشکار است. حمد و ثنا می‌گزارم او را که فکرت و اندیشه را در بستر روحم روان ساخت و بهره‌گیری از خوان گستردهٔ دانش اساتیدم را نصیب و روزی‌ام گردانید.

برخود لازم می‌دانم از استاد گرامیم جناب آقای دکتر محسن محمدزاده که در طی سالها، همواره راهنمای اینجانب بوده‌اند و از ابتدا تا انتهای این پایان‌نامه مرا صبورانه و صمیمانه یاری نمودند و همواره با راهنمایی‌ها و فکرهای ارزشمند خود راهگشای من در مشکلات بوده‌اند، قدردانی و سپاسگزاری نمایم. از اساتید محترم و جناب آقای دکتر گل علیزاده و جناب آقای دکتر خالدی که با راهنمایی‌های ارزنده‌شان نقش موثری در بهبود این رساله داشته‌اند، کمال تشکر را دارم. همینطور از اساتید محترم جناب آقای دکتر شفیعی و جناب آقای دکتر نعمت‌اللهی به سبب حضور در جمع داوران سپاسگزاری می‌نمایم و امیدوارم همواره در پناه الطاف الهی محفوظ و موفق باشند.

از تمام دوستانی که در این مرحله از زندگی مرا به هر شکل ممکن یاری نموده‌اند، مخصوصاً همسر گرامیم آقای امید کریمی و پدر و مادر عزیزم صمیمانه تشکر می‌کنم.

فاطمه حسینی

تهران – اسفند ۱۳۸۹

تقدیم به

گوهران والای تمامی سالهای زندگی‌ام

پدر و مادرم

صفا بخش تمامی لحظه‌های خوشبختی‌ام

امید

قشنگ‌ترین واژه زندگی‌ام

آرمان

چکیده

مدل‌های آمیخته خطی تعمیم‌یافته فضایی برای مدل‌بندی پاسخ‌های فضایی گسسته به کار می‌روند، که در آن‌ها ساختار همبستگی فضایی داده‌ها از طریق متغیرهای پنهان با توزیع نرمال در نظر گرفته می‌شود. هر چند فرض نرمال بودن توزیع متغیرهای پنهان موجب سهولت محاسبات می‌شود، اما در عمل به دلیل غیرقابل مشاهده بودن متغیرهای پنهان، بررسی نرمال بودن این متغیرها مقدور نیست و پذیرش ناصحیح این فرض می‌تواند روی دقت برآورد پارامترها و پیشگوها تأثیر سوء داشته باشد. در این رساله استفاده از دو خانواده توزیع‌های چوله نرمال و چوله نرمال بسته که شامل خانوادهٔ نرمال نیز هستند، برای توزیع متغیرهای پنهان پیشنهاد شده و پیشگویی متغیرهای پنهان و برآورد ماکسیمم درستنمایی پارامترهای مدل مورد مطالعه قرار گرفته است. سپس پیشگویی فضایی بیزی متغیرهای پنهان چوله ارائه و در دو مطالعه شبیه‌سازی مورد ارزیابی قرار گرفته است. نحوه کاربست مدل آمیخته خطی تعمیم یافته با متغیرهای پنهان چوله، روی داده‌های دمای هوا به نمایش گذاشته و براساس معیار اعتبارسنجی متقابل برتری مدل پیشنهادی بر مدل نرمال نشان داده شده است. طولانی بودن زمان محاسبه برآوردها و پیشگویی‌های بیزی انگیزه‌ای برای معرفی روش بیز تقریبی گردید که از سرعت بالاتری برخوردار است. این روش در یک مطالعه شبیه‌سازی و دو مجموعه داده واقعی به کار گرفته شده و دقت پیشگویی مدل با دو فرض نرمال و چوله نرمال بودن متغیرهای پنهان بررسی و عملکرد دو روش بیز تقریبی و معمولی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی : توزیع چوله نرمال بسته، مدل آمیخته خطی تعمیم‌یافته، متغیر پنهان، استنباط بیز تقریبی.

فهرست مندرجات

لیست اشکال

فصل ۱

مفاهیم مقدماتی

۱.۱ مقدمه

نلدر و ودربرن (۱۹۷۲) اولین کسانی بودند که مدل‌های خطی تعمیم‌یافته[۱] (GLM) را معرفی نمودند و مک‌کلاح و نلدر (۱۹۸۹) مدل‌های خطی تعمیم‌یافته را برای مدل‌بندی متغیرهای پاسخ گسسته پیشنهاد دادند. در این مدل‌ها با فرض استقلال مشاهدات، با استفاده از یک تابع پیوند بین میانگین مشاهدات و متغیرهای کمکی ارتباط خطی برقرار می‌شود. در حالتی که بین مشاهدات همبستگی وجود دارد، تعمیمی از مدل‌های مذکور به نام مدل‌های آمیخته خطی تعمیم‌یافته[۲] (GLMM) استفاده می‌شود. در این مدل‌ها فرض استقلال مشاهدات به استقلال شرطی تعدیل و همبستگی بین آن‌ها با اضافه کردن اثرات تصادفی[۳] از طریق متغیرهای پنهان[۴] به مدل در نظر گرفته

[۱] Generalized Linear Models

[۲] Generalized Linear Mixed Models

[۳] Random Effects

[۴] Latent Variables

می‌شود.

در بسیاری از علوم محیطی، پزشکی، ژنتیک و ... با داده‌هایی مواجه می‌شویم که برحسب موقعیت قرار گرفتن آن‌ها به یکدیگر وابسته‌اند، به این‌گونه مشاهدات داده‌های فضایی[۵] گویند. آمار فضایی شاخه‌ای از علم آمار است که به بررسی متغیرهایی می‌پردازد که ساختار فضایی از خود نشان می‌دهند. این شاخه از آمار می‌کوشد بین مقادیر مختلف یک متغیر و فاصله و جهت قرارگیری آن‌ها نسبت به هم ارتباطی برقرار کند، این ارتباط فضایی ساختار فضایی نام دارد. در حالتی که پاسخ‌های فضایی پیوسته باشند، با در نظر گرفتن یک میدان تصادفی مانا[۶]، تحلیل فضایی و پیشگویی مقادیر نامعلوم در موقعیت‌های معلوم با روش کریگیدن[۷] امکان‌پذیر است. اما در مواردی که پاسخ‌های فضایی گسسته هستند، معمولاً مدل آمیخته خطی تعمیم‌یافته فضایی[۸] (SGLMM) بکار گرفته می‌شود. یک مسئله مهم در مدل‌های SGLM پیشگویی متغیرهای پنهان در موقعیت‌های فاقد مشاهده می‌باشد، که مستلزم برآورد پارامترهای مدل و متغیرهای پنهان در موقعیت‌های دارای مشاهده پاسخ می‌باشد.

چون در مدل‌های آمیخته خطی تعمیم‌یافته، تابع درستنمایی برخلاف مدل‌های خطی به دلیل ناگاوسی بودن متغیر پاسخ و وجود متغیرهای پنهان فرم بسته‌ای ندارد، برآورد پارامترها به راحتی امکان‌پذیر نمی‌باشد. لذا در اکثر مقالات با پذیرش فرض نرمال بودن متغیرهای پنهان به ارائه راه حلی برای برآورد پارامترهای مدل و متغیرهای پنهان با ماکسیمم کردن توابع درستنمایی،

[۵] Spatial Data

[۶] Stationary Random Field

[۷] Kriging

[۸] Spatial Generalized Linear Mixed Model

شبه درستنمایی تاوانیده[۹] یا درستنمایی سلسله مراتبی[۱۰] به روش‌های عددی پرداخته شده است. از جمله مک‌کلاچ (۱۹۹۷) با بکار بردن روش‌های عددی مثل ماکسیمم‌سازی امید ریاضی مونت کارلویی[۱۱] (MCEM)، نیوتن رافسون مونت کارلویی[۱۲] (MCNR) و ماکسیمم درستنمایی شبیه‌سازی شده[۱۳] (SML)، برآورد ماکسیم درستنمایی[۱۴] (ML) پارامترها را در مدل‌های GLM با اثرات تصادفی غیرفضایی به دست آورد. همچنین در مدل‌های غیرفضایی GLM، پن و تامپسون (۲۰۰۷) برآورد شبه مونت کارلو[۱۵] و لین (۲۰۰۷) برآورد پارامترهای مدل GLM پواسنی را با ترکیب درستنمایی نما[۱۶] و شبه درستنمایی تاوانیده مورد مطالعه قرار دادند. ناتارجان و کَس (۲۰۰۰)، با بسط روش‌های بیزی در تحلیل مدل‌های GLM، نحوه انتخاب پیشین در این مدل‌ها را مطالعه کردند.

اولین بار بریسلو و کلیتون (۱۹۹۳) از مدل SGLM، در مطالعات پزشکی استفاده کردند. دیگل و همکاران (۱۹۹۸) این مدل‌ها را برای تحلیل متغیرهای فضایی گسسته در یک ناحیه پیوسته به کار بردند و تحت رهیافت بیزی با مینیمم کردن میانگین توان دوم خطاها، پیشگوی بهینه را برای متغیرهای پنهان در نقاط جدید به دست آوردند. زانگ (۲۰۰۲) با ترکیب روش مونت کارلو و الگوریتم گرادیانت ماکسیمم‌سازی امید ریاضی (EMG)، الگوریتم جدید گرادیانت ماکسیمم‌سازی

[۹] Penalized Quasi Likelihood
[۱۰] Hierarchical Likelihood
[۱۱] Monte Carlo Expectation Maximization
[۱۲] Monte Carlo Newton Raphson
[۱۳] Simulated Maximum Likelihood
[۱۴] Maximum Likelihood
[۱۵] Quasi-Monte Carlo
[۱۶] Pseudo-Likelihood

امید ریاضی مونت کارلویی[۱۷] (MCEMG) را برای برآورد ML پارامترهای همبستگی و پیشگویی متغیرهای پنهان نرمال در مدل SGLM بیان نمود. کریستنسن و همکاران (۲۰۰۰) و کریستنسن و وگیترسون (۲۰۰۲) با استفاده از رهیافت بیزی و روش‌های MCMC به تحلیل مدل‌های SGLM پرداختند و نشان دادند که استفاده از الگوریتم لانجوین-هستینگس[۱۸] (LH)، برای شبیه‌سازی از توزیع پسین متغیرهای پنهان گاوسی مفید می‌باشد. کریستنسن (۲۰۰۴) با روش ماکسیمم درستنمایی و الگوریتم مونت کارلو، پارامترها و پیشگوی بهینه را در مدل‌های SGLM با فرض نرمال بودن متغیرهای پنهان به دست آورد. کریستنسن و همکاران (۲۰۰۶) با معرفی روش‌های MCMC استوار به تحلیل مدل‌های SGLM با متغیرهای پنهان نرمال پرداختند و اینسورث و دین (۲۰۰۶) برآوردهای بیزی را با برآوردهای شبه‌درستنمایی تاوانیده در مدل‌های SGLM مقایسه کردند. زو و پترسون (۲۰۰۷) الگوریتم ماکسیمم‌سازی امید ریاضی تقریبی تصادفی[۱۹] (SAEM)، را برای تحلیل تابع درستنمایی در مدل‌های SGLM معرفی کردند. لذا در تئوری‌های فضایی با پاسخ‌های گسسته نیز با فرض نرمال بودن متغیرهای پنهان اقدام به حل مسئله شده است. اما در عمل به دلیل غیر قابل مشاهده بودن متغیرهای پنهان در مدل‌های SGLM، بررسی نرمال بودن آن‌ها مقدور نیست و پذیرش ناصحیح این فرض می‌تواند برروی دقت برآورد پارمترها و پیشگوها تأثیر سوء داشته باشد. اخیراً کمارک و لسافری (۲۰۰۸) با در نظر گرفتن توزیع گاوسی آمیخته تاوانیده[۲۰] برای متغیرهای پنهان به مطالعه مدل‌های GLM پرداخته‌اند. در حالتی که پاسخ‌های فضایی پیوسته هستند، برای اولین بار کیم و مالیک (۲۰۰۴) تحلیل داده‌های فضایی را برای یک میدان تصادفی چوله

[۱۷] Monte Carlo Expectation Maximization Gradiant

[۱۸] Langevin-Hastings

[۱۹] Stochastic Approximation EM

[۲۰] Penalized Gaussian Mixture

گاوسی[۲۱](SGR) مورد بررسی قرار دادند و پیشگویی فضایی بیزی را در یک مثال کاربردی به کار گرفتند. کریمی و محمدزاده (۲۰۰۷، ۲۰۰۹، ۲۰۱۰) و کریمی و همکاران (۲۰۱۰) تحلیل داده‌های فضایی را برای یک میدان تصادفی چوله گاوسی بسته[۲۲] (CSG) در مسائل کاربردی مورد مطالعه قرار دادند.

با توجه به این که خانواده توزیع‌های چوله نرمال[۲۳]، (آزالینی، ۱۹۸۵) و توزیع چوله نرمال بسته[۲۴](CSN)، (دامینگوس و همکاران، ۲۰۰۳) از خانواده توزیع‌های نرمال بزرگتر و از انعطاف‌پذیری بیشتری برخوردارند و شامل این خانواده می‌باشند، انتظار می‌رود پذیرش فرض چوله نرمال بودن توزیع متغیرهای پنهان در مدل‌های SGLM به واقعیت نزدیک‌تر از پذیرش فرض نرمال بودن آن‌ها باشد. لذا در مدل‌های آمیخته خطی تعمیم‌یافته، چون وجود توزیع متغیرهای پنهان ضروری است، پذیرش فرض چوله نرمال بودن توزیع متغیرهای پنهان بر فرض نرمال بودن ارجح به نظر می‌رسد. بنابراین در این رساله برای اجتناب از پذیرش فرض نرمال بودن، استفاده از خانواده توزیع‌های چوله نرمال و چوله نرمال بسته برای مدل‌بندی متغیرهای پنهان فضایی پیشنهاد می‌شود. سپس برآورد پارامترها و پیشگویی متغیرهای پنهان با رهیافت ماکسیمم درستنمایی و بیزی به دست آورده می‌شوند.

در این رساله، مفاهیم اولیه داده‌های فضایی، مدل‌های آمیخته خطی تعمیم‌یافته فضایی، توزیع چوله نرمال و چوله نرمال بسته در فصل ۱ ارائه شده‌اند. در فصل ۲، مدل‌های آمیخته خطی تعمیم یافته فضایی با متغیرهای پنهان چوله نرمال بسته معرفی و پیشگوی مینیمم میانگین توان دوم خطا[۲۵]

[۲۱] Skew Gaussian Random Field
[۲۲] Closed Skew Gaussian Random Field
[۲۳] Skew Normal
[۲۴] Closed Skew Normal
[۲۵] Mean-Squared Error

(MSE) برای متغیرهای پنهان ارائه می‌شوند. برای برآورد پارامترها روش ماکسیمم درستنمایی به کار گرفته می‌شود. در این مدل‌ها به دلیل وجود متغیرهای پنهان، تابع درستنمایی فرم پیچیده‌ای دارد و استفاده از روش ماکسیمم درستنمایی به راحتی میسر نیست و نیاز به حل عددی انتگرال‌های چند بعدی می‌باشد. لذا الگوریتم MCEMG که توسط زانگ (۲۰۰۲) برای به دست آوردن برآورد ماکسیمم درستنمایی پارامترها معرفی شده است، به مدل پیشنهادی تعمیم داده می‌شود و پارامترهای آن برآورد می‌گردد. در فصل ۳، پیشگویی و برآورد بیزی مدل‌های SGLM با متغیرهای پنهان چوله نرمال بسته ارائه و دو مطالعه شبیه‌سازی به منظور ارزیابی مدل‌ها براساس معیار میانگین توان دوم خطاها صورت می‌پذیرد. به علاوه به کارگیری مدل آمیخته خطی تعمیم یافته با متغیرهای پنهان چوله برای مدل‌بندی داده‌های دمای هوا در استان‌های غربی و شمال‌غربی ایران با استفاده از معیار اعتبارسنجی متقابل مورد ارزیابی قرار گرفته و کاهش میانگین توان دوم خطاهای اعتبار سنجی متقابل[۲۶] (CVMSE) برای مدل SGLM با متغیرهای پنهان چوله نرمال نسبت به نرمال نشان داده شده است. به دلیل طولانی بودن زمان محاسبه روش‌های بیزی برای این مدل‌ها، در فصل ۴ روش بیز تقریبی[۲۷] برای استنباط و پیشگویی فضایی معرفی می‌شود که نیازمند نمونه‌گیری‌های مونت کارلویی نیست و بسیار سریع است. تحلیل بیز تقریبی ارائه شده در یک مطالعه شبیه‌سازی و دو مجموعه داده واقعی مربوط به بارندگی و آلودگی آب دریاچه بکار گرفته شده و دقت پیشگویی مدل با دو فرض نرمال و چوله نرمال بودن متغیرهای پنهان بررسی و روش بیز تقریبی با روش‌های بیز معمولی مورد مقایسه و ارزیابی قرار گرفته است. در انتها به بحث و نتیجه‌گیری پرداخته شده است.

---

[۲۶] Cross Validation Mean Squared Errors

[۲۷] Approximate Bayesian Method

۲.۱ مدل‌های آمیخته خطی تعمیم یافته

مدل‌های خطی تعمیم‌یافته این امکان را فراهم می‌آورند که رگرسیون معمولی به مواردی که متغیر پاسخ غیرگاوسی باشد، توسیع داده شود. در مدل‌های خطی کلاسیک فرض بر این است که جمله خطا متغیرهای تصادفی مستقل با توزیع نرمال هستند، در حالی‌که در مدل‌های خطی تعمیم‌یافته با استفاده از یک تابع پیوند[۲۸] بین میانگین پاسخ‌ها و متغیرهای کمکی ارتباط خطی برقرار می‌باشد. واضح است که مدل‌های خطی تعمیم‌یافته کلاس متحدی از مدل‌های آماری فراهم می‌آورد که مدل‌های خطی کلاسیک را تعمیم داده است. برای توضیحات مقدماتی پیرامون مدل‌های خطی تعمیم‌یافته می‌توان به مک‌کلاح و نلدر (۱۹۸۹)، فارمیر و توتز (۱۹۹۱)، مک‌کلاچ و سیرل (۲۰۰۱) اشاره کرد. فرض کنید توزیع بردار مشاهدات پاسخ Y متعلق به خانواده نمایی به صورت

$$f(y|\eta,\phi) = \exp\{[y\eta - b(\eta)]/a(\phi) + c(y,\phi)\}$$

باشد، که در آن η پارامتر کانونی، ϕ پارامتر پراکندگی، $b(\cdot)$، $a(\cdot)$ و $c(\cdot)$ توابع معلوم هستند. همچنین $b(\cdot)$ یک تابع کومولانت[۲۹] است، بطوریکه $\mu = \mathrm{E}[Y] = \frac{\partial b(\eta)}{\partial \eta}$ و $Var[Y] = a(\phi)\frac{\partial^2 b(\eta)}{\partial \eta^2}$. اکنون با فرض استقلال مشاهدات، با استفاده از یک تابع پیوند معلوم g، داریم $g(\mu) = z'\beta$، که در آن z بردار متغیرهای کمکی و β بردار پارامترهای رگرسیونی است.

وقتی بردار پاسخ Y مستقل نباشند از مدل‌های آمیخته خطی تعمیم‌یافته، که توسیعی از مدل‌های خطی تعمیم‌یافته هستند، استفاده می‌شود که شامل متغیرهای پنهان برای بیان اثرات تصادفی می‌باشند. برای توضیحات مقدماتی پیرامون مدل‌های آمیخته خطی تعمیم‌یافته می‌توان به گلمن و

[۲۸] Link Function

[۲۹] Cumulant

همکاران (۲۰۰۴) اشاره کرد. فرض کنید Y بردار $n \times ۱$ مشاهدات متغیر پاسخ، Z و H به ترتیب ماتریس‌های $n \times p$ و $n \times q$ از متغیرهای کمکی مربوط به اثرات ثابت و متغیرهای پنهان باشند. با شرط روی بردار q بعدی متغیرهای پنهان $\boldsymbol{x}$، فرض می‌شود مشاهدات y_i به طور مشروط مستقل و از یک خانواده نمایی به صورت

$$f(y_i|\boldsymbol{x}) = \exp\{[y_i\eta_i - b(\eta_i)]/a(\phi) + c(\phi, y_i)\}, \quad i = ۱, \ldots, n,$$

استخراج شده‌اند، به طوری که $\mu_i = \mathrm{E}[Y_i|\boldsymbol{x}] = \frac{\partial b(\eta_i)}{\partial \eta_i}$ و $Var[Y_i|\boldsymbol{x}] = a(\phi)\frac{\partial^۲ b(\eta_i)}{\partial \eta_i^۲}$. اکنون با فرض استقلال شرطی مشاهدات پاسخ، با استفاده از یک تابع پیوند g بین میانگین شرطی مشاهدات پاسخ و متغیرهای کمکی ارتباط خطی به صورت $g(\mu_i) = Z\boldsymbol{\beta} + H\boldsymbol{x}$ برقرار می‌شود. متغیرهای پنهان در این مدل‌ها می‌تواند بیانگر همبستگی فضایی، خوشه‌بندی یا فرم‌های دیگر از وابستگی بین پیشامدها باشد و اغلب فرض می‌شود دارای یک توزیع چند متغیره نرمال با میانگین صفر و کواریانس Σ_θ است، که در آن $\boldsymbol{\theta}$ بردار پارامترهای مربوط به همبستگی می‌باشد. یک روش معمول برای برآورد پارامترها در این مدل‌ها روش ماکسیمم درستنمایی است. تابع درستنمایی برای این مدل‌ها به صورت

$$L(\boldsymbol{\beta}, \phi, \boldsymbol{\theta}|y) = \int \prod_{i=۱}^{n} f(y_i|\boldsymbol{x}, \boldsymbol{\beta}, \phi) f(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x}, \qquad (۱.۲.۱)$$

می‌شود، که دارای فرم بسته‌ای نمی‌باشد و بعد انتگرال برابر با تعداد متغیرهای پنهان مدل می‌باشد. لذا برای به دست آوردن برآورد ماکسیمم درستنمایی این مدل‌ها روش‌های مختلفی ارائه شده است که می‌توان به الگوریتم ماکسیمم‌سازی امید ریاضی مونت کارلویی، الگوریتم نیوتن رافسون مونت کارلویی، ماکسیمم درستنمایی شبیه‌سازی شده و ... اشاره کرد.

۳.۱ مدل‌های آمیخته خطی تعمیم یافته فضایی

در آمار فضایی متغیر مورد اندازه‌گیری ممکن است گسسته یا پیوسته باشد. در حالتی که پاسخ‌های فضایی پیوسته باشند، با در نظر گرفتن یک میدان تصادفی مانا، تحلیل فضایی و پیشگویی مقادیر نامعلوم در موقعیت‌های معلوم با روش کریگیدن امکان‌پذیر است. اما در مواردی که پاسخ‌های فضایی گسسته هستند، تحلیل‌های فضایی با استفاده از مدل‌های آمیخته خطی تعمیم‌یافته امکان‌پذیر می‌شود. دیگل و همکاران (۱۹۹۸)، مدل‌های GLM را برای پاسخ‌های فضایی گسسته روی یک ناحیه پیوسته فضایی تعمیم و مدل‌های SGLM را تعریف کردند. در هر موقعیت فضایی s فرض کنید $Y(s)$ متغیر فضایی پاسخ غیر گاوسی (گسسته) و $Z_1(s),\dots,Z_p(s)$ متغیرهای کمکی قابل مشاهده و $\{X(s), s\in I\!R^2\}$ یک میدان تصادفی فضایی غیر قابل مشاهده باشد به طوری که $X(s)$ نشان‌دهنده اثر تصادفی در موقعیت s است. دیگل و همکاران (۱۹۹۸) مدل‌های SGLM را به صورت زیر تعریف نمودند.

- $\{X(s), s\in I\!R^2\}$ یک میدان تصادفی مانای گاوسی است، به طوری که برای هر s، $\mathrm{E}[X(s)]=0$ و برای هر $h\in I\!R$، $\mathrm{Cov}(X(s+h),X(s))=C(h;\boldsymbol{\theta})$، که در آن $\boldsymbol{\theta}\in I\!R^k$ بردار پارامتر تابع کواریانس می‌باشد.
- $\{Y(s), s\in I\!R^2\}$ به شرط $\{X(s), s\in I\!R^2\}$، مجموعه‌ای از متغیرهای تصادفی مستقل است و توزیع $Y(s)$ با میانگین شرطی $\mathrm{E}[Y(s)|X(s)]$ مشخص می‌شود.
- برای هر تابع پیوند معلوم g و پارامترهای رگرسیونی $\beta_1,\dots,\beta_p$، داریم

$$g\{\mathrm{E}[Y(s)|X(s)]\}=\sum_{j=1}^{p}Z_j(s)\beta_j+X(s).$$

در مدل‌های SGLM نیز توزیع شرطی $Y(s)|X(s)$ متعلق به خانواده نمایی در نظر گرفته می‌شود.