دانشگاه علامه طباطبائی
دانشکده‌ی اقتصاد
گروه آمار، ریاضی و کامپیوتر
پایان‌نامه برای دریافت درجه‌ی کارشناسی ارشد **آمار ریاضی**

عنوان

رهیافت درستنمایی تجربی و کاربرد آن در تحلیل بقا

پژوهشگر

محدثه صفاکیش

استاد راهنما

دکتر حمیدرضا نواب‌پور

استاد مشاور

دکتر فرزاد اسکندری

بهمن ۱۳۸۸

بسم الله الرحمن الرحیم

تأیید پایان‌نامه‌ی کارشناسی ارشد توسط دانشجو

عنوان پایان‌نامه: رهیافت درستنمایی تجربی و کاربرد آن در تحلیل بقا

نام دانشجو: محدثه صفاکیش

شماره‌ی دانشجویی: ۸۶۱۱۲۸۲۰۷

استاد راهنما: دکتر حمیدرضا نواب‌پور

این‌جانب **محدثه صفاکیش** دانشجوی کارشناسی‌ارشد رشته‌ی **آمار ریاضی** دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی گواهی می‌نمایم پژوهش‌های ارائه شده در پایان‌نامه با عنوان مذکور توسط شخص این‌جانب انجام شده است و درستی مطالب نگارش یافته مورد تأیید می‌باشد. همچنین گواهی می‌نمایم مطالب مندرج در پایان‌نامه تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط این‌جانب یا فرد دیگری در هیچ کجا ارائه نشده است و در نگارش متن پایان‌نامه شیوه‌ی نگارش مصوب دانشکده‌ی اقتصاد را به‌طور کامل رعایت نموده‌ام. چنان‌چه در هر زمان خلاف آن‌چه گواهی نموده‌ام مشاهده گردد خود را از آثار حقیقی و حقوقی ناشی از دریافت مدرک کارشناسی ارشد محروم می‌دانم و هیچ‌گونه ادعایی نخواهم داشت.

تاریخ: **امضا دانشجو:**

اهدائیه

تقدیم به همه ی آن هایی که

حقشان بر گردنم است

سپاس‌گزاری

سپاس خدای را که هر توفیقی در گرو عنایت اوست. اکنون که با یاری او توانسته‌ام تلاشی هر چند ناچیز را در راه کسب دانش به انجام رسانم، بر خود لازم می‌دانم از استاد راهنمای بزرگوارم، جناب آقای دکتر حمیدرضا نواب‌پور، که به پایان رساندن این تحقیق جز با راهنمایی‌های پدرانه و هدایت‌های بی‌دریغ ایشان میسر نبود، قدردانی نمایم.

از استاد مشاورم جناب آقای دکتر فرزاد اسکندری که تذکراتشان باعث غنای پایان‌نامه شد، تشکر می‌نمایم. همچنین از جناب آقای دکتر نادر نعمت‌الهی و جناب آقای دکتر محمد بامنی‌مقدم که زحمت داوری این اثر را به عهده داشتند سپاس‌گزارم.

در پایان، از خانواده‌ام، به‌ویژه پدر و مادرم که با حمایت‌های خویش، همواره مرا پشتیبانی کرده‌اند نهایت سپاس و قدرشناسی را دارم.

امیدوارم بتوانم از عهده ادای حق این عزیزان برآیم.

اسفند ۸۸

لف!

فهرست مطالب

فهرست جدول‌ها

فهرست نمادها و علایم اختصاری

تعریف	نماد
$\hat{\beta}$ را b تعریف کنید	$\hat{\beta} := b$
بردار a	$\mathbf{a}$
نرم اقلیدسی بردار a	$\lVert\mathbf{a}\rVert$
تابع توزیع متغیر و بردار تصادفی $\mathbf{x}$	$F(.)$
تابع چگالی بردار $\mathbf{X}_i = \mathbf{x}_i$	$F\{\mathbf{x}_i\}$
تابع درستنمایی تجربی	$L(F)$
تابع نسبت درستنمایی تجربی	$R(F)$
تابع لگاریتم درستنمایی تجربی موزون	$\ell_w(F)$
تابع درستنمایی شبه تجربی	$\hat{\ell}(p)$
آماره‌ی نسبت درستنمایی تجربی نیم‌رخ	$\ell_s(\beta^{(1)})$
همگرایی در توزیع	$\xrightarrow{\mathscr{L}}$

چکیده

در بیش‌تر مطالعه‌هایی که صورت می‌گیرد، علاقه‌مند به استنباط درباره‌ی توزیع جامعه و پارامترهای آن هستیم. از جمله روش‌های معمول برای براورد پارامترهای جامعه -زمانی‌که توزیع معلوم است- روش ماکسیمم درستنمایی است. براوردهای حاصل از این روش در حالت حدی ویژگی‌های مطلوب بسیاری دارند. نااریبی، مینیمم واریانس و توزیع حدی نرمال به همراه روش دلتا استنباط پیرامون پارامترها و توابع آن‌ها را میسر می‌سازد.

در صورت ناشناخته بودن توزیع جامعه، روش‌های ناپارامتری بسیاری وجود دارند. برخی از این روش‌ها مانند روش باز نمونه‌گیری خودگردان، بر پایه‌ی تکرار نمونه‌گیری از یک نمونه‌ی اولیه پایه‌ریزی شده‌اند.

در این پایان‌نامه رهیافت ناپارامتری درستنمایی تجربی به منظور استفاده‌ی بهتر از اطلاعات کمکی برای استنباط درباره‌ی پارامترهای جامعه معرفی می‌شود. توزیع حدی آماره‌ی نسبت درستنمایی تجربی، به دست آمده و چگونگی ساختن ناحیه‌ی اطمینان با استفاده از آن بیان می‌شود. همچنین استفاده از این روش در براورد مدل رگرسیون میانه‌ی زمان شکست در حضور داده‌های سانسور از راست نشان داده می‌شود و ناحیه‌ی اطمینان بردار پارامترهای مدل به دست می‌آید. علاوه بر این آماره‌ی نسبت درستنمایی تجربی نیم‌رخ با هدف محاسبه‌ی ناحیه‌ی اطمینان برای زیر بردار دلخواه از پارامترها تعریف شده و توزیع حدی آن به دست می‌آید. سپس توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی نیم‌رخ با به‌کارگیری روش خودگردان ساخته شده و از آن برای محاسبه‌ی مرز ناحیه‌ی اطمینان استفاده می‌شود.

سرانجام به عنوان کاربردی از روش‌های بیان شده در پایان‌نامه، مدل رگرسیون میانه‌ی طول عمر برای مجموعه‌ی داده‌های مربوط به بیماران مبتلا به سرطان مغز استخوان با در نظر گرفتن مجموعه‌ای از متغیرهای کمکی، براورد می‌شود. همچنین ناحیه‌ی اطمینان برای بردار پارامترها و بازه‌های اطمینان برای تک تک ضریب‌های رگرسیونی به دست می‌آید.

واژه‌های کلیدی: *درستنمایی تجربی؛ معادله‌ی براورد؛ ناحیه‌ی اطمینان؛ براوردگر کاپلان- مه‌یر؛ رگرسیون میانه؛ سانسور از راست؛ درستنمایی تجربی نیم‌رخ؛ روش خودگردان درستنمایی تجربی.*

فصل ۱
کلیات

۱.۱ مقدمه

در بیش‌تر بررسی‌هایی که صورت می‌گیرد، جامعه تحت مطالعه حتی در صورت مشخص بودن توزیع جامعه (در حالت پارامتری) شامل کمیت‌های نامعلومی نظیر میانگین، واریانس و ... است که ما علاقه‌مند به براورد آن‌ها به‌عنوان پارامترهای جامعه هستیم. به این منظور روش‌های مختلفی مثل روش گشتاوری، ماکسیمم درستنمایی، کمترین توان‌های دوم در تحلیل رگرسیونی و ... وجود دارند.

روش ماکسیمم درستنمایی با وجودی که کاملاً وابسته به توزیع جامعه است براوردگرهایی ارائه می‌دهد که در حالت حدی دارای ویژگی‌های زیراند:

۱- نااریب هستند،

۲- دارای واریانس مینیمم می‌باشند، و

۳- توزیع نرمال دارند.

اگر $\hat{\theta}$ براورد ML برای پارامتر θ باشد، سه ویژگی بالا به‌صورت زیر خلاصه می‌شوند:

$$\hat{\theta} \longrightarrow N\Big(\theta, Var(\hat{\theta})\Big)$$

که در آن $Var(\hat{\theta}) = \frac{1}{I(\hat{\theta})}$ و $I(\theta)$ اطلاع فیشر نهفته در داده‌ها درباره‌ی θ است.

علاوه بر این هر تابع $g(\theta)$، از براوردگر ماکسیمم درستنمایی که دارای مشتق مرتبه‌ی اول پیوسته در θ، و $g'(\theta) \neq 0$ باشد، براساس روش دلتا دارای توزیع نرمال با میانگین $g(\theta)$ و واریانس $g'(\theta)^2 Var(\hat{\theta})^2$ است که در آن $Var(\hat{\theta})$ واریانس براوردگر ماکسیمم درستنمایی است؛

$$g(\hat{\theta}) \longrightarrow N\Big(g(\theta), g'(\theta)^2 Var(\hat{\theta})\Big)$$

همچنین ضریبی از تابع لگاریتم نسبت درستنمایی، دارای توزیع خی‌دو با v درجه‌ی آزادی است که v تعداد پارامترهایی است که باید براورد شوند. بنابر این با داشتن خصوصیت‌های بالا استنباط پیرامون پارامتر جامعه و هر تابع g از آن که دارای ویژگی‌های بالا باشد، به‌راحتی و با درنظر گرفتن توزیع حدی نرمال برای براوردگر ماکسیمم درستنمایی، امکان‌پذیر است.

در این روش تابع درستنمایی از روی تابع توزیع جامعه در صورت معلوم بودن به دست می‌آید.

در حالتی‌که توزیع جامعه مشخص نیست و امکان نرمال بودن آن نیز توسط آزمون‌های نیکویی برازش تأیید نمی‌شود و حتی تبدیل‌های نرمال کننده‌ی داده‌ها هم مؤثر واقع نشوند، استفاده از تقریب نرمال برای ساختن بازه‌ی اطمینان و آزمون فرض، درستی استنباط‌ها را مورد تردید قرار می‌دهد. در چنین شرایطی بهتر است از روش‌های مشابه ناپارامتری برای استنباط پیرامون پارامترهای جامعه استفاده شود. بعد از یافتن براورد نقطه‌ای پارامتر مورد نظر نیازمند داشتن براوردی از واریانس آن هستیم. بیش‌تر روش‌های باز نمونه‌گیری که وجود دارند به دنبال یافتن براوردی برای واریانس و میزان اریبی براوردگر به دست آمده برای پارامتر هستند. در این روش‌ها با تکرار نمونه‌گیری تصادفی ساده باجایگذاری از نمونه‌ی اصلی و محاسبه‌ی براوردگر موردنظر در تکرارهای زیاد نمونه‌گیری، توزیع نمونه‌گیری براوردگر موردنظر تقریب زده می‌شود. حال با استفاده از این توزیع و رابطه‌ی معمولی که برای محاسبه واریانس وجود دارد، براوردی از واریانس به دست می‌آید. از دیگر روش‌های موجود برای براورد خطای استاندارد و واریانس براوردگرها روش جک نایف است. این روش مشابه خودگردان براساس مفهوم تکرار پایه‌گذاری شده است. در این روش تعداد تکرارها مشخص و برابر با n (تعداد کل نمونه‌ی کامل) است. روش کار به این صورت است که هر بار یکی از مشاهده‌ها $(x_{(j)})$ حذف شده و براوردگر موردنظر $\hat{\theta}$، براساس n−۱ مشاهده‌ی باقی‌مانده محاسبه می‌شود $(\hat{\theta}_{(j)})$. سپس با توجه به میزان پراکندگی میان براوردهای حاصل، واریانس براوردگر مورد مطالعه، براورد می‌شود.

در مورد این روش‌ها به‌خصوص روش خودگردان، آن‌چه باعث ایجاد محدودیت می‌شود این است که در ابتدای امر نیازمند داشتن رابطه‌ای برای براورد پارامتر موردنظر هستیم؛ این امر زمانی‌که یکی از هدف‌های استنباط ارائه‌ی براورد نقطه‌ای برای پارامتر است چندان مناسب به‌نظر نمی‌رسد. همچنین بازه‌ای که این روش برای براوردگر به دست می‌دهد مطابق با همان رابطه‌ای است که برای تقریب نرمال وجود دارد و تنها براورد

واریانس از طریق ساختن توزیع نمونه‌گیری براوردگر به دست می‌آید.

با توجه به آن‌چه به اختصار در مورد روش‌های بازنمونه‌گیری در بالا گفته شد، دیده می‌شود که به‌کارگیری روشی که ویژگی‌های مطلوب‌تری داشته و فرضیه‌های اولیه‌ی کمتری نیاز داشته باشد، ضروری است. در مقابل همه‌ی روش‌های باز نمونه‌گیری، درستنمایی تجربی به‌عنوان یك روش ناپارامتری زمانی که توزیع جامعه نامشخص است و داده‌ها مستقل از هم تولید شده‌اند برای ساختن ناحیه‌های اطمینان و آزمون فرض روی پارامترهای مورد نظر جامعه ابداع شده است.

در حالت پارامتری ماکسیمم کننده‌ی تابع درستنمایی به‌عنوان براورد پارامتر موردنظر استفاده می‌شود و برای درستنمایی تجربی هم از روش مشابهی براوردگرهای ماکسیمم درستنمایی تجربی را ارائه می‌دهند. در حالت پارامتری ناحیه‌ی اطمینان و آزمون فرض بر مبنای تابع نسبت درستنمایی و استفاده از قضیه‌های حدی به دست می‌آیند، به طور مشابه تابع نسبت درستنمایی تجربی نیز تعریف شده و از آن برای ساختن ناحیه‌ی اطمینان و انجام آزمون فرض استفاده می‌شود. با توجه به این‌که لگاریتم نسبت درستنمایی تجربی به‌صورت حدی دارای توزیع خی‌دو است، بنابر این بدون نیاز به کمیت محوری ناحیه‌ی اطمینان به دست می‌آید. این خصوصیت در مسائلی که یافتن براورد برای واریانس براوردگرها کار مشکلی است، مزیت بزرگی به‌نظر می‌رسد.

از نظر هال و اسکالا (۱۹۹۰) وجه تمایز روش درستنمایی تجربی نسبت به سایر روش‌هایی که در موقعیت مشابه به کار می‌روند دارا بودن خصوصیت‌های زیر است.

۱- ناحیه‌های درستنمایی تجربی نیاز به براورد شاخص‌های خاص نظیر چولگی و ... ندارند، در حالی‌که در روش خودگردان تصحیح اریبی شتابیده نیاز به براوردی از چولگی است و روش خودگردان صدك-t بدون داشتن براورد کارا از خطای استاندارد براوردگر پارامتر موردنظر به نتیجه نمی‌رسد.

۲- ناحیه‌ی اطمینان در این روش بدون نیاز به داشتن کمیت محوری ساخته می‌شود. این خصوصیت در مسائلی که دست‌یابی به واریانس براوردگر مشکل است، -مثل حالتی‌که می‌خواهیم برای ضریب همبستگی بازه‌ی اطمینان بسازیم- مزیت بزرگی به‌نظر می‌رسد.

۳- ناحیه‌های اطمینان درستنمایی تجربی حافظ دامنه‌ی تغییرات پارامتر بوده و قابل تبدیل هستند. حافظ دامنه بودن به این معنا است که ناحیه‌ی به دست آمده از روش درستنمایی تجربی زیر مجموعه‌ی ناسره از دامنه‌ی تغییرات پارامتر است. برای مثال ناحیه‌ای که برای ضریب همبستگی به دست می‌آید حتماً در فاصله بین ۱- تا ۱+ قرار می‌گیرد. قابل تبدیل بودن یعنی برای ساختن ناحیه‌ی اطمینان برای پارامتر $g(\theta)$ کافی است تابع g روی ناحیه‌ی به دست آمده برای پارامتر θ اثر داده شود. دارا بودن این دو ویژگی برای ناحیه‌های درستنمایی تجربی در حالی است که ناحیه‌های خودگردان صدك-t هیچ یك از

این ویژگی‌ها را ندارند، و روش تصحیح اریبی شتابیده نیز تنها تحت تبدیل‌های پایا از پارامتر بدون تغییر است.

۴- شکل ناحیه‌ی درستنمایی تجربی از داده‌ها به دست می‌آید. بنابر این نیازی به تعیین شکل ناحیه قبل از ساختن ناحیه‌ی اطمینان مثل روش‌های خودگردان و تقریب نرمال نیست.

۲.۱ تعریف مفاهیم و واژه‌های اساسی

در این قسمت برخی از مفاهیم پایه‌ای به کار رفته در پایان‌نامه معرفی می‌شوند.

۱.۲.۱ تابع‌های $O(.), o(.)$

بارتل (۱۳۷۸)، برای مقایسه‌ی مرتبه‌ی بزرگی اعضای دنباله‌ها این دو تابع را بیان کرده است: فرض کنید $X=(x_n), Y=(y_n), y_n \neq 0$ دنباله‌هایی در $\mathbb{R}$ باشند، برای جمله‌های این دو دنباله از اعداد داریم:

الف) اگر $\lim_{n\to\infty} \frac{x_n}{y_n} = 0$ باشد، می‌گوییم مرتبه‌ی X کمتر از مرتبه‌ی Y است [سرعت افزایش جمله‌های دنباله‌ی Y بیش‌تر است.] و می‌نویسیم:

$$X = o(Y) \quad or \quad x_n = o(y_n)$$

ب) در صورتی که دنباله‌ی $\frac{x_n}{y_n}$ کران‌دار باشد، می‌گوییم مرتبه‌ی X بزرگ‌تر از مرتبه‌ی Y است [سرعت افزایش جمله‌های دنباله‌ی X بیش‌تر است.] و می‌نویسیم:

$$X = O(Y) \quad or \quad x_n = O(y_n)$$

در ریاضیات و سایر علوم مرتبط با آن از نماد O برای مطالعه‌ی رفتار حدی تابع مورد نظر زمانی‌که آرگومان آن به سمت مقدار مشخصی و یا ∞ میل می‌کند، استفاده می‌شود. توصیف تابع با استفاده از این نماد معمولاً تنها یک کران بالا برای نرخ رشد تابع ارائه می‌دهد. در هنگام استفاده از بسط مجانبی و سری‌های تیلور، Oی بزرگ برای توصیف بخش خطا در تقریب تابع ریاضی به کار می‌رود. مثال زیر را درنظر بگیرید.

مثال ۱. بسط سری تیلور تابع e^x را وقتی $x \to 0$ درنظر بگیرید.

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots = 1 + x + \frac{x^2}{2} + r_n$$

بنابر این نمایش باقیمانده‌ی r_n با استفاده از نماد معرفی شده وقتی $x \longrightarrow 0$، به‌صورت $|r_n| \leq C|x^3| = O(x^3)$ است و C مقدار ثابتی است.

در ادامه عبارت‌هایی که در کاربردها بیش‌تر مورد استفاده قرار می‌گیرند؛ بیان می‌شوند. برای تابع‌های f و $g \neq 0$ دلخواه داریم:

۱- $f(n) = o(g(n))$ اگر و فقط اگر داشته باشیم: $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

۲- از آن‌جا که $\lim_{n\to\infty} \frac{\log(n)}{n} = 0$، بنابر این $\log(n) = o(n)$.

۳- اگر $\lim_{n\to\infty} \frac{f(n)}{g(n)} = L$، که $0 \leq L < \infty$، آن‌گاه داریم: $f(n) = O(g(n))$.

۲.۲.۱ تابع‌های $O_p(.), o_p(.)$

نماد مرتبه در احتمال، در نظریه‌ی آمار و احتمال به‌صورت موازی با نماد Oی بزرگ در ریاضیات به کار می‌رود. نماد Oی بزرگ همگرایی مجموعه‌ها و دنباله‌هایی از اعداد ترتیبی را بررسی می‌کند؛ در برابر مرتبه در احتمال همگرایی مجموعه‌هایی از متغیرهای تصادفی را بررسی می‌کند. همگرایی مورد استفاده در این نماد، همگرایی در احتمال است. برای یک مجموعه از متغیرهای تصادفی X_n و مجموعه‌ای از ثابت‌های a_n نماد $X_n = o_p(a_n)$ به این معنی است که مجموعه‌ی مقدارهای $\frac{X_n}{a_n}$ در احتمال به صفر همگراست. به عبارت بهتر $\lim_{n\to\infty} \frac{X_n}{a_n} = 0$. نماد $X_n = O_p(a_n)$ یعنی مجموعه‌ی مقدارهای $\frac{X_n}{a_n}$ در احتمال کران‌دار هستند.

۳.۲.۱ درستنمایی نیم‌رخ

درستنمایی نیم‌رخ در گروه روش‌هایی قرار دارد که در آن‌ها استنباط بر مبنای درستنمایی به دست آمده برای زیر مجموعه‌ای از پارامترها، انجام می‌شود.

فرض کنید در یك مسئله با بردار پارامتر $\theta = (\theta_1, \ldots, \theta_r)$ به استنباط پیرامون پارامتر تکی ψ علاقه‌مند هستیم، که ممکن است $\psi = \theta_i$ و یا $\psi = g(\theta)$ باشد. در حالت اول باقی بردار θ را پارامترهای مزاحم نامیده و آن‌ها را با τ نشان می‌دهیم. درستنمایی نیم‌رخ با معلوم درنظر گرفتن ψ به وسیله‌ی ماکسیمم کردن تابع درستنمایی نسبت به پارامترهای مزاحم به دست می‌آید. با توجه به توضیحات بالا درستنمایی نیم‌رخ برای ψ به‌صورت زیر تعریف می‌شود:

$$L(\psi) = \max_{\tau} L(\psi, \tau) = L(\psi, \hat{\tau})$$

۴.۲.۱ الگوریتم نیوتن-رافسون

فرض کنید در جستجوی یافتن ریشه‌ی تابع دلخواه $f(x)$ هستیم. همچنین می‌دانیم امکان به دست آوردن ریشه با استفاده از روش‌های تحلیلی وجود ندارد. در چنین حالتی از روش‌های عددی برای یافتن ریشه استفاده می‌شود.

یکی از روش‌های عددی معمول برای یافتن مقدار صفر تابع مشتق‌پذیر $f(.)$ استفاده از الگوریتم نیوتن-رافسون است. در ادامه چگونگی عملکرد این الگوریتم ارائه می‌شود.

گام ۱: x_0 را به‌عنوان مقدار اولیه‌ی ریشه درنظر می‌گیریم.

گام ۲: نقطه‌ی تلاقی خط مماس بر تابع $f(.)$ در نقطه‌ی x_0، با محور xها را یافته و x_1 می‌نامیم. x_1 به‌صورت زیر به دست می‌آید.

$$f(x_1)-f(x_0)=f'(x_0)(x_1-x_0) \quad , \quad f(x_1)=0$$

در نتیجه

$$x_1=x_0-\frac{f(x_0)}{f'(x_0)} \tag{۱.۲.۱}$$

گام ۳: در این مرحله الگوریتم x_1 را به‌عنوان مقدار اولیه‌ی جدید جایگزین x_0 کرده و گام ۲ را تکرار می‌کند. با توجه به رابطه‌ی (۱.۲.۱)، فرمول یافتن مقدار جدید (x_i) برحسب مقدار به دست آمده در مرحله‌ی قبل (x_{i-1}) به‌صورت زیر است.

$$x_i=x_{i-1}-\frac{f(x_{i-1})}{f'(x_{i-1})} \tag{۲.۲.۱}$$

تا زمانی‌که $|x_i-x_{i-1}|=\left|\frac{f(x_{i-1})}{f'(x_{i-1})}\right|$ بزرگ‌تر از مقدار تحمل باشد؛ مرحله‌های بالا تکرار می‌شوند. مقدار تحمل از قبل توسط پژوهشگر و معمولاً نزدیک به صفر اختیار می‌شود. هر چه مقدار تحمل به صفر نزدیک‌تر باشد، حاصل الگوریتم صحت بیش‌تری دارد.

برای گسترش الگوریتم به حالت تابع برداری $f(\mathbf{x})$ کافی است در رابطه‌ی (۲.۲.۱) بردار گرادیان $\triangledown f(\mathbf{x}_i)$ جایگزین مشتق $f'(x_i)$ شود. در این حالت چنان‌چه $\left\|f(\mathbf{x}_{i-1})\left[\triangledown f(\mathbf{x}_{i-1})\right]^{-1}\right\|$ کوچک‌تر از مقدار تحمل باشد، متوقف می‌شود.

۵.۲.۱ قضیه‌ی بری-اسین

متغیرهای تصادفی مستقل و هم‌توزیع $X_1,\ldots,X_n$ را درنظر بگیرید. همچنین فرض کنید

$$E(X_i)=\mu, \quad Var(X_i)=\sigma^2, \quad \beta_3=E|X_i-\mu|^3<\infty, \quad i=1,2,\ldots,n$$

خطای ناشی از براورد توزیع ناشناخته‌ی جامعه توسط توزیع نرمال برای ساختن ناحیه‌ی اطمینان را با $e_n(x)$ نمایش می‌دهیم. بری (۱۹۴۱) و اسین (۱۹۴۵) یک کران بالا برای این خطا را به‌صورت زیر به دست آوردند.

$$\begin{aligned}\sup\{e_n(x)\} &= \sup_x \left| \Pr\left(\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{Var(\bar{X}_n)}} \le x\right) - \Phi(x) \right| \\ &\le \frac{M\beta_3}{\sigma^3\sqrt{n}} \\ &= O_p(\frac{1}{\sqrt{n}})\end{aligned}$$

مقدار ثابت M مستقل از n و توزیع Xها است. به عبارت بهتر خطای استفاده از تقریب توزیع نرمال از مرتبه‌ی $O_p(n^{-1/2})$ است.

۶.۲.۱ تحلیل بقا

مجموعه‌ای از روش‌ها و مدل‌ها برای تحلیل داده‌های بقا را، تحلیل بقا گویند. هر فرآیندی که منتظر رخداد حادثه‌ای خاص است، داده‌ی بقا تولید می‌کند. داده‌های معرف طول عمر، زمان انتظار، مقدار فشار وارد به قطعه‌های يك دستگاه، کیلومتر طی شده توسط اتومبیل و ... مثال‌هایی از انواع دادهای بقا می‌باشند. در این مبحث نکته‌ی قابل تذکر این است که از زمان، با معنای اصطلاحی مناسب با هر بررسی استفاده می‌شود. زمان گاهی به معنای زمان حقیقی و در برخی موارد هم منظور از زمان طول عمر، فشار و ... است.

لازم به ذکر است که بیش‌تر مفاهیم مربوط به تحلیل بقا از کالب‌فلیش و پرینتیس (۲۰۰۲) آمده است. برخی از این مفاهیم در زیر آورده شده‌اند.

۱.۶.۲.۱ شکست

منظور از شکست در تحلیل بقا، یعنی رخداد همان حادثه‌ای که منتظر وقوعش بودیم. برای مثال در يك بررسی بالینی درباره‌ی نحوه‌ی اثربخشی رژیم غذایی در درمان بیماری سرطان ریه، شکست یعنی مرگ به دلیل بیماری سرطان ریه.

۲.۶.۲.۱ زمان مبدأ

زمانی که بقا از آن به بعد اندازه‌گیری می‌شود، زمان مبدأ است. سن فرد هنگام ورود به مطالعه، فشار اولیه‌ای که به قطعه‌ی تولیدی وارد می‌شود، و در بسیاری از موارد زمان رخداد برخی حوادث نظیر ورود فرد به مطالعه یا زمان تشخیص يك بیماری خاص می‌توانند به‌عنوان زمان مبدأ درنظر گرفته شوند. البته زمان مبدأ برای همه‌ی افرادی که مورد بررسی قرار می‌گیرند لزوماً یکسان نیست و در اندازه‌گیری زمان بقا باید به این مسئله توجه شود.