بسم الله الرحمن الرحیم

انتشارات دانشگاه فردوسی مشهد، شمارهٔ ۳۱۲

تحلیل بقاء

نویسنده

ار. جی. میلر

مترجمان

ابوالقاسم بزرگ‌نیا - حجّت رضایی پژند

۱۳۸۰

میلر، روپرت Miller, Rupert G.
تحلیل بقاء / نویسنده ار. جی. میلر؛ مترجمان ابوالقاسم بزرگ‌نیا، حجت رضایی پژند. ـ مشهد: دانشگاه فردوسی مشهد، ۱۳۸۰.
۱۸۶ ص.: جدول، نمودار. ـ (انتشارات دانشگاه فردوسی مشهد؛ ۳۱۲).
۸۳۰۰ ریال (ISBN: 964-6335-83-7)
فهرست‌نویسی بر اساس اطلاعات فیپا.
عنوان به انگلیسی Survival analysis.
واژه‌نامه.
کتابنامه.

۱. تحلیل داده‌های زمان افست. ۲. تحلیل بقاء (زیست‌سنجی). الف. بزرگ‌نیا، ابوالقاسم، ۱۳۱۲ - ، مترجم. ب. رضایی پژند، حجت، مترجم. ج. دانشگاه فردوسی (مشهد). د. عنوان.

QA ۲۷۶ / م ۹ پ ۷ ۵۱۹/۲
۱۳۸۰
کتابخانه ملی ایران ۱۶۸۹۵-۸۰ م

تحلیل بقاء

نوشتهٔ

ار. جی. میلر

مترجمان

ابوالقاسم بزرگ‌نیا، حجت رضایی پژند

وزیری، ۱۸۸ صفحه، ۱۰۰۰ نسخه، چاپ اوّل، پاییز ۱۳۸۰
امور فنّی و چاپ : مؤسسهٔ چاپ و انتشارات دانشگاه فردوسی
بها : ۸۳۰۰ ریال

شابک ۹۶۴-۶۳۳۵-۸۳-۷ (ISBN: 964-6335-83-7)

پیش‌گفتار نویسنده:

در بهار سال ۱۹۸۰، آموزش روشهای مفید و کاربردی در تحلیل بقاء به دانشجویان تحصیلات تکمیلی رشتهٔ آمار کاربردی در دانشگاه استانفورد، به این جانب پیشنهاد شد. با توجّه به سفارشهای ارزشمند برادافرون، این کتاب تهیه و تنظیم شده است.

در سرفصلهای این نوشته، بیشتر از سمینارهای تحلیل بقاء در گروه غددشناسی کالیفرنیای شمالی -که به صورت زنجیره‌ای چاپ می‌شود- و همچنین از نوشته‌های آرت پترسن استفاده شده است.

بیل بران، ضمن تشویق ما، در تألیف این کتاب کمک کرده است. از نکات ارزشمند جری‌هالپرن و تری‌ترینو و نظریّات ویرایشی الاین‌یونگ، در این کتاب استفاده شده است.

کارولادکلو، با تایپ دقیق و سریع خود و همکاری در سرعت چاپ و ترسیمات بسیار عالی ماریاجد، به چاپ کتاب کمک نموده‌اند.

انتشار کتاب توسّط مؤسسهٔ تحقیقاتی علوم دارویی عمومی گرانت، پشتیبانی شده است.

رپرت - ج - میلر و جی‌ار
گیل گانگ
آلوارو مونوز

استانفورد، کالیفرنیا
جولای ۱۹۸۱

پیش گفتار مترجمان:

آمار حیاتی از شاخه‌های اصلی آمار به شمار می‌رود. بسیاری از روشهای آماری این شاخه به تحقیقات پزشکی وابسته است. رشد روزافزون تحقیقات پزشکی و نبود نوشتاری کاربردی-نظری نیاز به نگارش چنین کتابی را تأیید می کند. نوشتاری کوتاه و پرمحتوی به شیوهٔ کتابهای درسی، می‌تواند این نیاز را تا اندازه‌ای برطرف کند. یکی از کتابهای بسیار مفید در این زمینه "تحلیل بقاء"، نوشتهٔ میلر است. چاپ دوباره آن در سال ۱۹۹۸، بیانگر استقبال از شیوهٔ بیان و مطالب ارزشمند آن است. مثالهای خوب، مسائل حل شده در پایان کتاب، روشهای مناسب، ارائهٔ قضایای لازم و اثبات روابط به صورت کوتاه، این کتاب را کاملاً کاربردی نموده است. در حالی که بیان نظری آن تا حدّ امکان حفظ شده است. کمی حجم و غنی بودن مطالب آن بیانگر توان بالای علمی مؤلّف و انتقال مناسب آن به خوانندگان است. ذکر مراجع مختلف و مربوط به هر مبحث، یکی از فواید این اثر و راه ارزنده‌ای برای دنبال کردن مطالب جانبی، اثبات قضایا و پژوهش در آمار کاربردی است. با توجّه به موارد بالا به ترجمهٔ این کتاب ارزشمند اقدام شده است. امید است بدین وسیله در پیشبرد آمار کاربردی خدمت کوچکی کرده باشیم.

مطالب این کتاب در حدّ دو تا سه واحد درسی برای دانشجویان کارشناسی رشته‌های آمار، داروسازی و رشته‌های مهندسی است. علاوه برآن مرجعی مفید و کوتاه برای دانشجویان کارشناسی ارشد، دکترا، پژوهشگران مختلف علوم پزشکی، مهندسی و علوم است. محققین و دانشجویان رشته‌های داروسازی، بیمه گری و آمارحیاتی بیشترین استفاده را از این کتاب می‌برند.

در پایان لازم می‌دانیم از زحمات آقایان مهندس تبرّایی، کدکنی و فنائی (چاپخانه دانشگاه مشهد)، همچنین آقای مهندس امین تشکر و قدردانی نماییم. چاپ زیبا، سریع و دقیق آقای فردوسی مکان (آمارپژوه) کمک مهمّی در ارائه این اثر داشته است. از ایشان قدردانی و تشکر می‌شود. طبیعی است که این ترجمه خالی از اشتباه و کاستی نیست، تذکّر صاحب‌نظران ما را خوشحال و چاپ‌های بعدی را بهبود خواهد بخشید.

ابولقاسم بزرگ‌نیا

حجت رضایی پژند

فهرست مطالب

عنوان	صفحه

فصل اوّل

مقدّمه‌ای در مفاهیم بقاء

تحلیل بقاء از نظر علم آمار، عبارت است از: استفاده از فنون مختلف آماری در تحلیل متغیّرهای تصادفی نامنفی. نوعاً، مقدار این متغیّر تصادفی زمان شکست یک مؤلّفهٔ فیزیکی (مکانیکی یا الکتریکی) یا زمان مرگ یک واحد زنده (سلّول، بیمار، حیوان و غیره) است. ممکن است این متغیّر، زمان یادگیری یک مهارت باشد، یا حتی امکان دارد به زمان هیچ ارتباطی نداشته باشد. برای مثال، متغیّر می‌تواند مبلغ پرداختی یک شرکت بیمه در وضعیّت خاصّی باشد.

در برخی موارد، یک بیمار بهبود یافته و مبلغ کل پرداختی بیمه او معلوم است. در موارد دیگر بیماری هنوز ادامه دارد و تنها مبلغ پرداختی تا آن زمان معلوم است.

پیشینه و منشأ تحلیل بقاء کارهایی است که درگذشته در مورد جداول طول عمر انجام شده است. شکل جدید تحلیل بقاء از نیم قرن گذشته با کاربردهای مهندسی مورد توجّه و مطالعه قرار گرفته است.

در جنگ جهانی دوم، علاقه‌مندی بسیاری به قابلیّت اعتماد تجهیزات جنگی به وجود آمد. این علاقه‌مندی به قابلیّت اعتماد، موضوع را به کارهای نظامی و فرآورده‌های تجاری کشاند. قبلاً بیشتر تحقیقات آماری مورد استفاده در مهندسی بر الگوهای پارامتری متمرکز بود. در دو دههٔ اخیر تحقیقات آزمایشگاهی پزشکی افزایش یافته است. در این آزمایشها بیشتر به روشهای ناپارامتری توجّه شده است. این کتاب هر دو روش پارامتری و ناپارامتری را مورد بحث قرار می‌دهد. ولی بیشتر به روشهای جدید ناپارامتری و کاربرد آنها در تحقیقات پزشکی تأکید شده است.

REFERENCE

Leavitt and Olshen, unpublished report (1974), give the insurance example.

۱.۱ تابعهای بقاء و نرخهای شکست

فرض کنید متغیّر تصادفی $T>0$ دارای تابعهای چگالی $f(t)$ و توزیع $F(t)$ باشد. تابع بقاء $S(t)$، به صورت زیر تعریف می‌شود:

$$S(t)=1-F(t)=P(T>t)$$

نرخ شکست یا تابع شکست $\lambda(t)$، به شکل زیر تعریف می‌شود:

$$\lambda(t)=\frac{f(t)}{1-F(t)}$$

در علم امراض مسری، تابع $\lambda(t)$ را نرخ مرگ و میر نامند. نرخ شکست به صورت زیر قابل تفسیر است:

$$\lambda(t)dt \cong P(t<T<t+dt \mid T>t)$$

$$= P\left\{\text{رخ دهد } (t, t+dt) \text{ پیشامد در بازهٔ } \middle| \text{ بقاء بیشتر از } t\right\}$$

از $\lambda(t)$ انتگرال می‌گیریم، داریم:

$$\int_0^t \lambda(u)du = \int_0^t \frac{f(u)}{1-F(u)}du = -\log\Big[1-F(u)\Big]_0^t$$

$$= -\log\big[1-F(t)\big] = -\log S(t)$$

در نتیجه می‌توان رابطهٔ مهم زیر را نتیجه گرفت:

$$S(t)=e^{-\int_0^t \lambda(u)du}$$

توجّه نمایید که $F(\infty)=1$ (یعنی $S'(\infty)=0$) خواهد بود، اگر و فقط اگر $\int_0^\infty \lambda(u)du=\infty$ باشد. مفاهیم بالا را می‌توان برای حالاتی که T دارای تابع چگالی نیست نیز تعمیم داد. به عبارت دیگر هنگامی که تابع توزیع F گسسته باشد. در بحث‌هایمان F را پیوسته در نظر می‌گیریم و هر جا لازم باشد، مفاهیم و رابطه‌ها را برای حالت گسسته اصلاح می‌کنیم.

۲.۱ انواع برشها (Censoring)

مطالب ارائه شده در این کتاب در بیشتر منابع آماری موجوداند. آنچه تحلیل بقاء را از سایر مباحث آماری جدا می‌کند، عمل برش است. به‌طور کلّی، یک مشاهدهٔ بریده شده، فقط شامل قسمتی از اطّلاعات مربوط به متغیّرهای تصادفی مورد نظر است. در این کتاب سه نوع برش را مورد بحث و توجّه قرار می‌دهیم.

فرض کنید $T_۱$، $T_۲$، ...، T_n، متغیّرهای مستقل و هم‌توزیع (iid) با تابع توزیع F باشند. برشها را به شرح زیر بررسی می‌کنیم:

۱.۲.۱ برش نوع اوّل

فرض کنید t_c عددی ثابت باشد، که آن را زمان برش می‌نامیم. به جای مشاهدهٔ $T_۱$ تا T_n (متغیّرهای مورد توجّه)، فقط متغیّرهای $Y_۱$ تا Y_n را به شرح زیر مشاهده می‌کنیم:

$$Y_i = \begin{cases} T_i & T_i \le t_c \\ t_c & T_i > t_c \end{cases}$$

توجّه کنید که تابع توزیع Y در $y = t_c$، دارای جرم مثبت $P(T > t_c) > ۰$، است.

۲.۲.۱ برش نوع دوم

فرض کنید $r < n$ عددی ثابت و $T_{(۱)} < T_{(۲)} < \cdots < T_{(n)}$، آماره‌های مرتّب $T_۱$ تا T_n باشند. اگر r امین شکست رخ دهد، مشاهده در این حالت پایان می‌پذیرد. بنابراین، فقط می‌توانیم $T_{(۱)}$ تا $T_{(r)}$ را مشاهده کنیم. مشاهدات مرتّب شده کامل به شرح زیراند:

$$\begin{aligned} Y_{(۱)} &= T_{(۱)} \\ Y_{(۲)} &= T_{(۲)} \\ &\vdots \\ Y_{(r)} &= T_{(r)} \\ Y_{(r+۱)} &= T_{(r)} \\ &\vdots \\ Y_{(n)} &= T_{(r)} \end{aligned}$$

هر دو نوع، برش اوّل و دوم را در مهندسی به کار می‌برند. برای مثال، تعدادی

لامپ یا ترانزیستور موجود است. تمام آنها را در زمان $t=0$ مورد آزمایش قرار می‌دهیم و زمان شکست را یاداشت می‌کنیم. ممکن است طول عمر بعضی از ترانزیستورها طولانی باشد و نخواهیم به مدت طولانی منتظر بمانیم تا آزمایش به پایان برسد. بنابراین، امکان دارد، آزمایش را در زمان t_c، که از قبل تعیین شده است، پایان دهیم. در این صورت برش نوع اول انجام شده است. در حالت دیگر از قبل ندانیم چه زمانی برای برش مناسب است. لذا، تصمیم بگیریم، که در انتظار بمانیم تا نسبت معیّنی از ترانزیستورها، مانند: $\frac{r}{n}$ بسوزند. در این‌جا، برش نوع دوم رخ داده است.

۳.۲.۱ برش نوع سوم (برش تصادفی)

این نوع برش که به برش تصادفی نیز معروف است، به صورت زیر تعریف می‌شود: فرض کنید، C_1 تا C_n، متغیّرهای iid با تابع توزیع G باشند. C_i زمان برش مربوط به T_i است. در این حالت فقط می‌توانیم (Y_1, δ_1)، ...، (Y_n, δ_n) را مشاهده کنیم. که در آن Y_i و δ_i به شرح زیر تعریف می‌شوند:

$$Y_i = \min(T_i, C_i) = T_i \wedge C_i$$

$$\delta_i = I(T_i \le C_i) = \begin{cases} 1 & T_i \le C_i \quad (\text{یعنی: } T_i \text{ بریده نمی‌شود}) \\ 0 & T_i > t_c \quad (\text{یعنی: } T_i \text{ بریده می‌شود}) \end{cases}$$

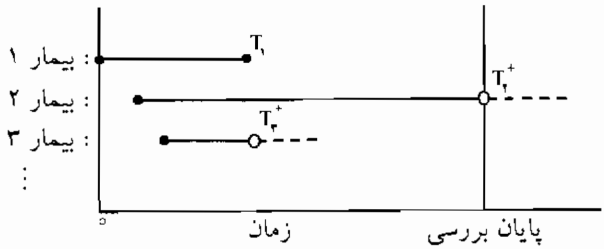
توجّه شود که Y_1 تا Y_n، متغیّرهای iid با توزیع H هستند. علاوه بر این، δ_1 تا δ_n، شامل اطّلاعات مربوط به برش هستند. تذکر این‌که، در برش نوع اوّل و دوم می‌توانستیم ببینیم که چه مشاهداتی بریده می‌شوند. به این علّت، تعریف δ_iها به طور آشکار ضروری نبود.

برش تصادفی در کارهای پزشکی و در مورد افراد یا آزمایشهای بالینی استفاده می‌شود. در یک آزمایش بالینی، ممکن است بیماران برای معالجه در زمانهای مختلف مراجعه کنند. سپس، هر بیمار به یکی از روشهای ممکن معالجه شود. حال اگر بخواهیم طول عمر آنها را مطالعه کنیم، عمل برش به یکی از صورتهای زیر انجام می‌شود:

۱. **عدم مراجعه:** ممکن است بیمار تصمیم بگیرد به پزشک دیگری مراجعه کند. در نتیجه، دیگر او را نخواهیم دید.

۲. **قطع معالجه:** ممکن است به علّت عوارض بد جانبی، درمان آن را متوقف کنیم. یا این‌که، ممکن است بیمار هنوز در تماس باشد ولی از ادامهٔ معالجه خودداری کند.

۳. **اتمام بررسی:** نمودار زیر یک بررسی ممکن را تشریح می‌کند:



در این‌جا، بیمار (۱) در زمان $t=0$ مورد بررسی قرار گرفته و در زمان T_1 فوت شده است. در نتیجه، یک مشاهدهٔ بریده نشده به دست آمده است. بیمار (۲) مورد بررسی قرار گرفته و در پایان بررسی هنوز زنده است. در نتیجه یک مشاهدهٔ بریده نشده T_2^+ به دست آمده است. بالاخره، بیمار (۳) مورد بررسی قرار گرفته و قبل از پایان بررسی، معالجه را قطع نموده است. در نتیجه، یک مشاهدهٔ بریده شدهٔ T_3^+ به دست آمده است.

دربارهٔ برش تصادفی، فرض اساسی زیر را در نظر می‌گیریم:

فرض: متغیّرهای تصادفی C_i و T_i مستقل‌اند.

بدون این فرض نتایج کمتری در دسترس است. معقول به نظر می‌رسد که فرض کنیم: مطالعه در زمانهای تصادفی شروع، به تصادف بررسی و قطع می‌شود. با این وجود، اگر دلیلی برای قطع بررسی در جریان معالجه وجود داشته باشد، در این صورت ممکن است بین T_i و C_i، رابطه‌ای وجود داشته باشد.

۴.۲.۱ انواع دیگر برش

انواع دیگر برش در منابع مختلف مورد بررسی قرار گرفته‌اند. انواع قبلی برش به راست و چپ تقسیم می‌شوند. اگر متغیّر مورد مطالعه بسیار بزرگ باشد و نخواهیم آن را به‌طور کامل مشاهده کنیم، آن را "راست برش" نامند. به‌طور مشابه "چپ برش" قابل تعریف است. برای مثال، در برش تصادفی چپ، تنها می‌توان (Y_1, ε_1)، ...، (Y_n, ε_n) را مشاهده نمود، که Y_i و ε_i به شرح زیرند:

$$Y_i = \max(T_i, C_i) = T_i \vee C_i$$

$$\varepsilon_i = I(C_i \le T_i)$$

مثال. کودکان آفریقایی: در این مثال هر دو برش راست و چپ انجام می‌شود. یک روان‌پزشک می‌خواهد سنّ گروه خاصی از کودکان آفریقایی را برای انجام یک عمل ویژه بداند. وقتی به روستای مورد نظر می‌رسد، تعدادی از بچّه‌ها از قبل می‌دانند آن عمل را چگونه انجام دهند. در نتیجه، این کودکان مشاهدات بریده شده چپ را ارائه می‌کنند. بعضی از کودکان یادگیری را شروع کرده‌اند و سنّ یادگیری آنها را می‌توان ثبت کرد. هنگام بازگشت پژوهشگر، هنوز برخی از کودکان روستایی این عمل را یاد نگرفته‌اند. آنها مشاهدات بریده شده راست را نتیجه می‌دهند.

REFERENCE
Leiderman et al., Nature (1974).
Turnbull, JASA (1974).

برشهای راست و چپ؛ هر دو حالتهای خاصّی از برش فاصله‌ای هستند، که در آنها فقط مشاهداتی مورد توجّه‌اند که در یک بازه قرار گیرند. اگر T_i برش تصادفی راست باشد، مشاهدات T_i در بازهٔ $[C_i, \infty)$ قرار می‌گیرند. همچنین، اگر T_i برش تصادفی چپ باشد، مشاهدات T_i در بازهٔ $(-\infty, C_i]$ قرار می‌گیرند.

در برابر برش (Censoring) فاصله‌ای، قطع مشاهدات (Truncation) قرار دارد. در این‌جا، اگر متغیّر مورد علاقه در خارج بازهٔ معیّنی قرار گیرد، از آن چشم‌پوشی می‌شود. برای مثال، فرض کنید بخواهیم توزیع و میانگین اندازهٔ یک عنصر داخل سلّول را تعیین کنیم. طبیعی است، به علّت محدودیّت وسایل اندازه‌گیری، اگر اندازهٔ عنصر، کوچکتر از اندازهٔ تعیین شده باشد، قابل تشخیص نیست.

نمادهای مورد استفاده:

متغیّر تصادفی T_i را برای زمان بقاء و C_i را برای زمان برش در نظر می‌گیریم. متغیّرهای تصادفی مشاهده شده، عبارت‌اند از: $Y_i = T_i \wedge C_i$ و $\delta_i = I(T_i \le C_i)$. نمادهای دیگر مورد استفاده به شرح زیراند:

۱) $X_i \sim F$: زمان بقاء و $Y_i \sim G$: زمان برش است، متغیّرهای مشاهده‌ای به صورت $Z_i \simeq X_i \wedge Y_i \sim H$ و $\delta_i = I(X_i \le Y_i)$ است.

این نمادها به نظر خوب می‌رسند. زیرا به راحتی می‌توان بین متغیّر تصادفی و تابعهای توزیع تفاوت قائل شد. ولی در کاربردهای رگرسیونی که بعداً به آن می‌پردازیم،

X را برای متغیّر مستقل به کار می‌بریم.

۲) $X_i^\circ \sim F^\circ$: زمان بقاء و $Y_i \sim G$: زمان برش است. متغیّرهای مشاهده‌ای به صورت $Z_i \simeq X_i^\circ \wedge Y_i$ و $\delta_i = I(X_i^\circ \leq Y_i)$ است.

در گزارش اعداد واقعی، مناسبتر است مشاهدهٔ بریده نشده را بـا (T_i) و مشـاهدهٔ بریده شده را با (T_i^+)، نشان دهیم. در این صورت ممکن است داده‌ها به شکل زیر باشند. ۵، $^+$۱۱، ۶٫۵، $^+$۱۴ و ... . اعداد ۵ و ۶٫۵ بریده نشده و ۱۱ و ۱۴ بریده شده‌اند.

فصل دوم

الگوهای پارامتری

۱ توزیعها

۱.۱ توزیع نمایی:

الگوی نمایی دارای نرخ شکست ثابت است. یعنی: $\lambda(t) = \lambda > 0$، در نتیجه موارد زیر را داریم:

$$\int_0^t \lambda(u)du = \lambda t$$

$$S(t) = e^{-\int_0^t \lambda(u)du} = e^{-\lambda t}$$

$$f(t) = -\frac{d}{dt}S(t) = \lambda e^{-\lambda t}$$

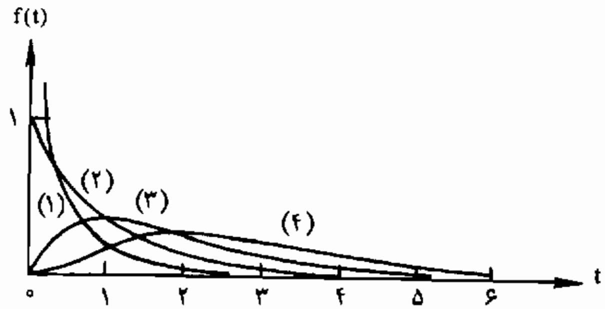
$$E(T) = \frac{1}{\lambda} \quad \text{و} \quad Var(T) = \frac{1}{\lambda^2}$$

۲.۱ توزیع گاما (دو پارامتری)

الگوی گاما تعمیمی از الگوی نمایی است:

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1}.e^{-\lambda t} \qquad \alpha\ ,\ \lambda > 0$$

امید ریاضی و واریانس آن $E(T) = \frac{\alpha}{\lambda}$ و $Var(T) = \frac{\alpha}{\lambda^2}$ هستند. منحنی تابع چگالی گاما با λی ثابت و به ازاء αهای مختلف در نمودار (۱)، ارائه شده است:



نمودار ۱. تابع چگالی گاما برای $\lambda = ۱$ و (۱): $\alpha = \frac{۱}{۲}$، (۲): $\alpha = ۱$، (۳): $\alpha = ۲$، (۴): $\alpha = ۳$.

متأسّفانه الگوی گاما، عبارت بسته‌ای برای $S(t)$ و $\lambda(t)$ ندارد.

$$S(t) = ۱ - \int_{۰}^{t} f(u)du = ۱ - \left(\frac{\text{تابع گاما ناقص}}{\text{تابع گاما کامل}}\right)$$

۳.۱ توزیع وایبل

الگوی وایبل تعمیم دیگری از توزیع نمایی است.

$$S(t) = e^{-(\lambda t)^{\alpha}} \qquad \alpha\ ,\ \lambda > ۰$$

در نتیجه در موارد زیر به دست می‌آید:

$$\int_{۰}^{t} \lambda(u)du = (\lambda t)^{\alpha}$$

$$\lambda(t) = \alpha\lambda(\lambda t)^{\alpha - ۱}$$

$$f(t) = \lambda(t).S(t) = \alpha\lambda(\lambda t)^{\alpha-۱}.e^{-(\lambda t)^{\alpha}}$$

در الگوی وایبل، $E(T)$ و $Var(T)$ دارای عبارت بسته‌ای نیستند. با این حال، شکل سادهٔ $\lambda(t)$ و $S(t)$، الگوی وایبل را در تحلیل بقاء مفید ساخته است. نمودار (۲)، منحنی الگوی وایبل را نشان می‌دهد.

۴.۱ توزیع رایلی

این توزیع دارای مشخصات زیر است:

$$\lambda(t)=\lambda_0+\lambda_1 t$$

$$\int_0^t \lambda(u)du=\lambda_0 t+\frac{1}{2}\lambda_1 t^2$$

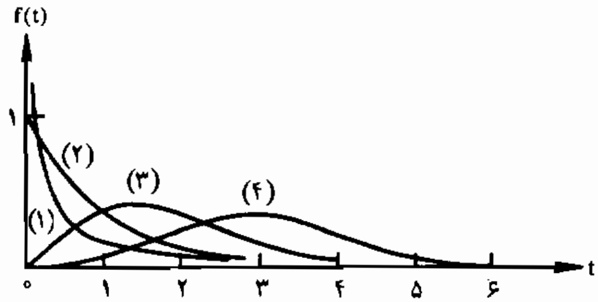
$$S(t)=\exp(-\lambda_0 t-\frac{1}{2}\lambda_1 t^2)$$

$$f(t)=(\lambda_0+\lambda_1 t)\exp(-\lambda_0 t-\frac{1}{2}\lambda_1 t^2)$$

گشتاورهای رایلی عبارت بسته‌ای ندارند.

نرخ شکست خطّی را می‌توان به صورت چندجمله‌ای عمومیت داد؛ داریم:

$$\lambda(t)=\sum_{i=0}^{p}\lambda_i t^i$$



نمودار ۲. چگالی وایبل در ازای (۱): $\lambda = 1$ و $\alpha = \frac{1}{2}$، (۲): $\lambda = 1$ و $\alpha = 1$، (۳): $\lambda = 0/5$ و $\alpha = 2$، (۴): $\lambda = 0/3$ و $\alpha = 3$.

۵.۱ توزیع لگ نرمال (دو پارامتری)

در این توزیع فرض می‌شود $\log T_i$ به صورت نرمال توزیع شده باشد.

$$\log T_i \simeq N(\mu,\sigma^2)$$

$\lambda(t)$ و $S(t)$ دارای صورت بسته‌ای نیستند.

$$S(t)=1-P(T<t)=1-P\{\log T<\log t\}$$

$$=1-P\left\{\frac{\log T-\mu}{\sigma}<\frac{\log t-\mu}{\sigma}\right\}=1-\Phi\left(\frac{\log t-\mu}{\sigma}\right)$$

توزیع لگ نرمال می‌تواند برای داده‌های بریده نشده مفید باشد. یک تبدیل لگاریتمی

داده‌ها را به الگوی خطّی استاندارد تبدیل می‌کند.

۶.۱ توزیع پارتو

این توزیع به شرح زیر است:

$$S(t)=\left(\frac{a}{t}\right)^{\alpha} I_{(a,\infty)}(t) \qquad \alpha\ ,\ a>0$$

در نتیجه:

$$f(t)=\frac{\alpha a^{\alpha}}{t^{\alpha+1}} I_{[a,\infty)}(t)$$

$$\lambda(t)=\frac{\alpha}{t} I_{[a,\infty)}(t)$$

گشتاورهای این توزیع به آسانی محاسبه می‌شوند، ولی ممکن است نامتناهی باشند.

۷.۱ IFR و IFRA

اگر F یا f دارای تابع نرخ شکست صعودی باشند، آنها را IFR نامند. در این صورت $\lambda(t)$ تابعی صعودی است. اگر F یا f دارای میانگین نرخ شکست صعودی باشند، آنها را IFRA نامند. در این صورت تابع زیر صعودی است:

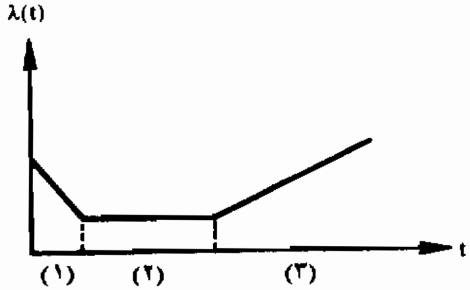
$$\frac{1}{t}\int_0^t \lambda(u)du$$

تعاریف مشابهی برای توابع DFR و DFRA وجود دارد.

FR ثابت	IFR	DFR
نمایی	وایبل $(\alpha>1)$	وایبل $(\alpha<1)$
	گاما $(\alpha>1)$	گاما $(\alpha<1)$
	رایلی $(\lambda_1>0)$	رایلی $(\lambda_1<0)$
		پارتو $(t>a)$

مفاهیم توزیعهای IFR و IFRA در کاربردهای مهندسی مفیدند. به ویژه در مطالعهٔ دستگاههای متشکّل از چندین مؤلّفه. این دو معمولاً در آمار حیاتی کاربرد زیادی ندارند.

برای مثال در بررسی‌های علم بیماریهای مسری، خطر بقاء دراز مدّت، معمولاً به شکل وام حمّام است، که در آن، زمان به سه دوره همانند شکل زیر تقسیم می‌شود:



(۱): دورهٔ خردسالی
(۲): دورهٔ بزرگسالی
(۳): دورهٔ کهولت

REFERENCE

Barlow and Proschan, Statistical Theory of Reliability and Life Testing (1975).

۲ برآورد کردن

۱.۲ حداکثر درست‌نمایی

الگوی برش تصادفی را در نظر می‌گیریم. (توجّه شود که در این‌جا برش نوع اوّل را با فرض $c_i \equiv t_c$، در نظر می‌گیریم. درست‌نمایی برای برش نوع دوم، مشابه نوع اوّل است. با این تفاوت که به علّت منظور کردن ترتیب یک ضریب ثابت وجود دارد).

تابع درست‌نمایی زوج مرتّب (y_i, δ_i) به شرح زیر است:

$$L(y_i, \delta_i) = \begin{cases} f(y_i) & \delta_i = ۱ \quad \text{(بدون برش)} \\ S(y_i) & \delta_i = ۰ \quad \text{(بریده شده)} \end{cases}$$

$$= f(y_i)^{\delta_i} . S(y_i)^{۱-\delta_i}$$

تابع درست‌نمایی نمونهٔ کامل نیز به شرح زیر است:

$$L = L(y_۱, \ldots, y_n, \delta_۱, \ldots, \delta_n) = \prod_{i=۱}^{n} L(y_i, \delta_i)$$

$$= \left(\prod_{u} f(y_i)\right)\left(\prod_{c} S(y_i)\right)$$

در واقع تابعهای درستنمایی برای برش تصادفی، به شرح زیر است:

$$L(y_i, \delta_i) = \begin{cases} f(y_i)[1-G(y_i)] & \delta_i = 1 \\ g(y_i)S(y_i) & \delta_i = 0 \end{cases}$$

$$L = \left(\prod_u f(y_i)\right)\left(\prod_c S(y_i)\right)\left(\prod_c g(y_i)\right)\left(\prod_u [1-G(y_i)]\right)$$

اگر فرض شود که زمان برش با زمان بقاء ارتباط ندارد، دو حاصل‌ضرب آخری $\prod_c g(y_i)$ و $\prod_u [1-G(y_i)]$، شامل پارامترهای مجهول نیستند. در نتیجه،این دو را می‌توان در بیشینه کردن L، ثابت فرض کرد.

فرض کنید، $\underline{\theta} = (\theta_1, \ldots, \theta_p)'$، بردار پارامتر باشد. محاسبهٔ $\max L(\underline{\theta})$ با محاسبهٔ جواب $\hat{\underline{\theta}}$ معادلات درستنمایی زیر یکسان است:

$$\frac{\partial}{\partial \theta_j} \log L(\underline{\theta}) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log L_{\underline{\theta}}(y_i, \delta_i)$$

$$= \sum_u \frac{\partial}{\partial \theta_j} \log f_{\underline{\theta}}(y_i) + \sum_c \frac{\partial}{\partial \theta_j} \log S_{\underline{\theta}}(y_i) = 0 \qquad j = 1, 2, \ldots, p$$

معمولاً، محاسبهٔ جواب به کمک رایانه و روشهای عددی امکان‌پذیر است.

روشهای نیوتن-رافسون و چوب‌خطّی

نمادهای زیر را تعریف می‌کنیم:

$$L_i(\underline{\theta}) = L_{\underline{\theta}}(y_i, \delta_i) \qquad i = 1, 2, \ldots, n$$

$$\frac{\partial}{\partial \underline{\theta}} \log L(\underline{\theta}) = \left(\frac{\partial}{\partial \theta_1} \log L(\underline{\theta}), \ldots, \frac{\partial}{\partial \theta_p} \log L(\underline{\theta})\right)'$$

$$\frac{\partial^2}{\partial \underline{\theta}^2} \log L(\underline{\theta}) = \begin{pmatrix} \frac{\partial^2}{\partial \theta_1 \partial \theta_1} \log L(\underline{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} \log L(\underline{\theta}) \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} \log L(\underline{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_p \partial \theta_p} \log L(\underline{\theta}) \end{pmatrix}$$

در نتیجه معادلات درست‌نمایی به صورت زیر خواهد بود:

$$\sum_i \frac{\partial}{\partial\theta_j} \log L_i(\underline{\theta}) = 0 \qquad j=1,\dots,p$$

یا

$$\frac{\partial}{\partial\underline{\theta}} \log L(\underline{\theta}) = 0$$

فرض کنید بردار $\underline{\hat{\theta}}^{\circ} = (\hat{\theta}_1^{\circ},\dots,\hat{\theta}_p^{\circ})'$، یک مقدار اولّیه برای جواب باشد. در صورت بسط حول $\underline{\hat{\theta}}^{\circ}$، داریم:

$$\sum_i \frac{\partial}{\partial\theta_j} \log L_i(\underline{\hat{\theta}}) = \sum_i \frac{\partial}{\partial\theta_j} \log L_i(\underline{\hat{\theta}}^{\circ}) +$$

$$+ \sum_k (\hat{\theta}_k - \hat{\theta}_k^{\circ}) \sum_i \frac{\partial^2}{\partial\theta_k \partial\theta_j} \log L_i(\underline{\hat{\theta}}^{\circ}) + \cdots = 0 \qquad j=1,\dots,p$$

یا

$$\frac{\partial}{\partial\underline{\theta}} \log L(\underline{\hat{\theta}}) = \frac{\partial}{\partial\underline{\theta}} \log L(\underline{\hat{\theta}}^{\circ}) + \frac{\partial^2}{\partial\underline{\theta}^2} \log L(\underline{\hat{\theta}}^{\circ})\,(\underline{\hat{\theta}} - \underline{\hat{\theta}}^{\circ}) + \cdots = \underline{0}$$

فرض کنید $\underline{\hat{\theta}}^1$ جوابی باشد، که در آن از جملات درجه دوم و بیشتر چشم‌پوشی شده است.

$$\underline{\hat{\theta}}^1 = \underline{\hat{\theta}}^{\circ} + \left(-\frac{\partial^2}{\partial\underline{\theta}^2} \log L(\underline{\hat{\theta}}^{\circ}) \right)^{-1} \frac{\partial}{\partial\underline{\theta}} \log L(\underline{\hat{\theta}}^{\circ}) \qquad (1)$$

بردار $\left(\frac{\partial}{\partial\underline{\theta}}\right) \log L(\underline{\hat{\theta}}^{\circ})$ را بردار چوب‌خط (یا امتیاز) در $\underline{\hat{\theta}}^{\circ}$ نامند. ماتریس زیر را ماتریس اطّلاع نمونه در $\underline{\hat{\theta}}^{\circ}$ نامند.

$$\underline{i}(\underline{\hat{\theta}}^{\circ}) = -\frac{\partial^2}{\partial\underline{\theta}^2} \log L(\underline{\hat{\theta}}^{\circ})$$

توجّه شود که رابطهٔ زیر صادق است. $\underline{I}(\underline{\theta})$، اطّلاع فیشر مشاهدهٔ iام است.

$$E(\underset{\sim}{i}(\hat{\underline{\theta}}^{\circ})) = \left(-E\frac{\partial^{2}}{\partial\theta_k\partial\theta_j}\log L(\underline{\theta})\right) = \underset{\sim}{I}(\underline{\theta})$$

متذکّر می‌شود که $\underset{\sim}{I}(\underline{\theta})$ اطّلاع فیشر برای کلّ نمونه است.

$$\underset{\sim}{I}(\underline{\theta}) = \sum_{i=1}^{n}\underset{\sim}{I}_i(\underline{\theta}) = n\underset{\sim}{I}_1(\underline{\theta})$$

در این‌جا $\underset{\sim}{I}_i(\underline{\theta})$ اطّلاع فیشر مربوط به مشاهدهٔ iام است.

روش تکراری که در (۱) به کار رفت، روش نیوتن-رافسون نامیده می‌شود. اگر اطّلاعات نمونه در (۱) را با اطّلاع فیشر جابه‌جا کنیم، داریم:

$$\hat{\underline{\theta}}^{1} = \hat{\underline{\theta}}^{\circ} + \underset{\sim}{I}^{-1}(\hat{\underline{\theta}}^{\circ})\frac{\partial}{\partial\underline{\theta}}\log L(\hat{\underline{\theta}}^{\circ}) \qquad (۲)$$

روش تکراری به کار رفته در (۲) را روش چوب‌خطّی (یا امتیازی) گویند. روش (۲) باید در بعضی از موارد سریعتر همگرا شود. ولی در عمل، هنگامی که عمل برش وجود دارد، نمی‌توان از $\underset{\sim}{I}(\underline{\theta})$ در (۲) استفاده کرد.

REFERENCES

Rao, Linear Statistical Inference (1965), Section 5g.

Gross and Clark, Survival Distributions (1975), Chapter 6.

Kalbfleisch and The Statistical Analysis of Failure Time Data (1980), Section 3.7.

بازه‌های اطمینان و آزمونها

برای دو برش تصادفی نوع اوّل با شرایط همواری، رابطهٔ زیر را داریم:

$$\hat{\underline{\theta}} \overset{a}{\sim} N(\underline{\theta}, \underset{\sim}{I}^{-1}(\underline{\theta}))$$

معمولاً این نتیجه برای برش نوع دوم نیز برقرار است. ولی، اثبات آن مشکل است. نماد " $\overset{a}{\sim}$ " به این معنی است، که رابطه به طور مجانبی برقرار است.

برای آزمون فرض: $H_0: \underline{\theta} = \underline{\theta}^{\circ}$ یا برای ساختن بازه‌های اطمینان، معمولاً از سه روش استفاده می‌شود:

۱ روش والد: تحت فرض H_0، داریم:

$$(\hat{\underline{\theta}} - \underline{\theta}^{\circ})'\underset{\sim}{I}(\underline{\theta}^{\circ})(\hat{\underline{\theta}} - \underline{\theta}^{\circ}) \overset{a}{\sim} \chi^2_p$$

که در این رابطه از $\underset{\sim}{I}(\hat{\underline{\theta}})$ به جای $\underset{\sim}{I}(\underline{\theta}^{\circ})$ می‌توان استفاده کرد.

۲ روش نیمن-پیرسن-نسبت درست‌نمایی ویلکس: تحت فرض H_0، داریم:

$$-2 \log \frac{L(\underset{\sim}{\theta}^{\circ})}{L(\hat{\underset{\sim}{\theta}})} \overset{a}{\sim} \chi_p^2$$

۳ روش رائو: تحت فرض H_0، داریم:

$$\frac{\partial}{\partial \theta} \log L(\underset{\sim}{\theta}^{\circ})' \underset{\sim}{I}^{-1}(\underset{\sim}{\theta}^{\circ}) \frac{\partial}{\partial \theta} \log L(\underset{\sim}{\theta}^{\circ}) \overset{a}{\sim} \chi_p^2$$

توجّه کنید که روش رائو از برآورد حداکثر درست‌نمایی (MLE) استفاده نمی‌کند. همچنین، محاسبات تکراری را لازم ندارد. علاوه بر آزمون، اغلب مایل به محاسبهٔ $\tilde{\theta}$ نیز هستیم. اگر $\hat{\underset{\sim}{\theta}}$ و $\underset{\sim}{I}(\underset{\sim}{\theta}^{\circ})$ را داشته باشیم، روش والد ساده‌تر خواهد بود.

در حالت برش لازم است به جای $\underset{\sim}{I}(\underset{\sim}{\theta})$ از $\underset{\sim}{i}(\underset{\sim}{\theta})$ استفاده شود. زیرا، معمولاً محاسبهٔ $\underset{\sim}{I}(\underset{\sim}{\theta})$ مشکل است. همچنین، افرن و هینکلی پیشنهاد می‌کنند، که استفاده از $\underset{\sim}{i}(\underset{\sim}{\theta})$ برای بازهٔ اطمینان از $\underset{\sim}{I}(\underset{\sim}{\theta})$، حتی زمانی که $\underset{\sim}{I}(\underset{\sim}{\theta})$ قابل محاسبه باشد، بهتر است. با این وجود در این‌گونه موارد نظرها متفاوت است.

REFERENCES

Rao, Linear Statistical Inference (1965), Section 6e.
Efron and Hinkley, Biometrika (1978).

مثال ۱ نمایی: تحت برش تصادفی، فرض کنید n_u، تعداد مشاهدات بریده نشده باشد، در این صورت داریم:

$$L = \lambda^{n_u} \exp\left(-\lambda \sum_u t_i - \lambda \sum_c c_i\right) = \lambda^{n_u} \exp\left(-\lambda \sum_{i=1}^{n} y_i\right)$$

$$\log L = n_u \log \lambda - \lambda \sum_{i=1}^{n} y_i$$

$$\frac{\partial}{\partial \lambda} \log L = \frac{n_u}{\lambda} - \sum_{i=1}^{n} y_i$$

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n} y_i}$$

$$\frac{\partial^{2}}{\partial \lambda^{2}} \log L = \frac{-n_u}{\lambda^{2}}$$

$$i(\underline{\lambda}) = \frac{n_u}{\lambda^{2}}$$

توجّه شود، که $\hat{\lambda} = n_u \Big/ \sum_{i=1}^{n} y_i$، برآورد MLE تحت برش نـــوع اوّل و دوم -بـه خوبـی برآورد تحت برش تصادفی- نیز هست.

برای ساختن بازه‌های اطمینان و اجرای آزمونها، توزیع $\hat{\lambda}$ مورد نیاز است.

(الف) اگر مشاهدات بریده نشوند، داریم:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i} = \frac{1}{\bar{T}}$$

در رابطهٔ بالا: T_1 تا T_n، متغیّرهای iid با توزیع نمایی و تابع چگالی زیراند:

$$f_{T_i}(t) = \lambda e^{-\lambda t}$$

در نتیجه، $S = \sum_{i=1}^{n} T_i$ دارای چگالی گاما به شرح زیر است.

$$f_S(t) = \frac{\lambda^{n}}{\Gamma(n)} t^{n-1} . e^{-\lambda t}$$

بنابراین: $2\lambda \sum_{i=1}^{n} T_i \sim \chi^{2}_{2n}$ یا به عبارت معادل:

$$\frac{2n\lambda}{\hat{\lambda}} \sim \chi^{2}_{2n}$$

یعنی: $\frac{2n\lambda}{\hat{\lambda}}$ یک آمارهٔ محوری است و آن را می‌توان برای آزمون و بازهٔ اطمینان مـــورد استفاده قرار داد (علامت" ~ " به معنی هم‌توزیع بودن است).

(ب) برای برش نوع دوم می‌توان نوشت:

$$\sum_{i=1}^{n} Y_i = T_{(1)} + T_{(2)} + \cdots + (n-r) T_{(r)}$$

$$= nT_{(1)} + (n-1)[T_{(2)} - T_{(1)}] + \cdots + (n-r+1)[T_{(r)} - T_{(r-1)}]$$

با استفاده از نتایج فرایند پواسن و زمانهای انتظار نمایی، داریم:

$$T_{(1)} = \{\text{کمینه } (n-1) \text{ متغیّر iid نمایی } T_i \text{ با پارامتر } \lambda\} \sim n\lambda e^{-n\lambda t}$$

$$nT_{(1)} \sim \lambda e^{-\lambda t}$$

$$T_{(2)} - T_{(1)} = \{\text{کمینه } (n-1) \text{ متغیّر iid نمایی } T_i \text{ با پارامتر } \lambda\} \sim (n-1)\lambda e^{-(n-1)\lambda t}$$

$$(n-1)[T_{(2)} - T_{(1)}] \sim \lambda e^{-\lambda t}$$

تا آخر. چون متغیّرهـــای $nT_{(1)}$، $(n-1)[T_{(2)} - T_{(1)}]$ و ... و $(n-r+1)[T_{(r)} - T_{(r-1)}]$ مستقل‌اند، پس داریم:

$$2\lambda \sum_{i=1}^{n} Y_i \sim \chi^2_{2r}$$

بنابراین، برای ساختن بازه‌های اطمینان و آزمونها، می‌تـــوان از $2r\lambda/\hat{\lambda}$ و ارتبـــاط آن بـــا توزیع χ^2، استفاده کرد. در این حالت، درجهٔ آزادی دو برابر آماره‌های مرتّب بریده نشده است.

(پ) اگر برش از نوع تصادفی یا نوع اوّل باشد، چاره‌ای جز به‌کارگیری نظریهٔ مجـــانبی نداریم. با توجّه به محاسبات قبل، داریم:

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n} y_i} \qquad \text{و} \qquad \frac{\partial^2}{\partial \lambda^2} \log L = \frac{-n_u}{\lambda^2}$$

در نتیجه: $\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda^2/n_u}} \overset{a}{\sim} N(0,1)$. در این حالت می‌تـــوان بـــه‌جای n_u از $E(n_u)$ اســـتفاده نمود، به شرط آن‌که این میانگین معلوم باشد.

می‌توان تقریب نرمال را با تبدیل برآورد بهبود بخشید. به کمک روش دلتا (بعـــداً تعریف خواهد شد) و این‌که $\hat{\lambda} \overset{a}{\sim} N(\lambda, \frac{\lambda^2}{n_u})$، داریم:

$$\log\hat{\lambda} \overset{a}{\sim} N(\log\lambda, \frac{۱}{n_u})$$

توجّه شود که واریانس $\log\hat{\lambda}$، یعنی: $\frac{۱}{n_u}$، به پارامتر مجهول بستگی ندارد. این یک حقیقت تجربی است که، تبدیل کردن یک برآورد -برای حذف وابستگی واریانس به پارامتر مجهول- معمولاً، تمایل دارد که با کاهش چولگی، همگرایی به نرمال را بهبود بخشد.

REFERENCE

Epstein and Sobel, JASA (1953), is a classic paper.

روش دلتا: فرض کنید، متغیّر تصادفی Y دارای واریانس $\sigma^۲$ و میانگین μ است (که به اختصار به صورت $Y \sim (\mu, \sigma^۲)$ می‌نویسیم)، همچنین، فرض کنید توزیع تابع $g(Y)$ را می‌خواهیم. ابتدا، $g(Y)$ را پیرامون μ بسط می‌دهیم، داریم:

$$g(Y) = g(\mu) + (Y-\mu)g'(\mu) + \cdots$$

اگر از جملات درجات بالاتر چشم‌پوشی کنیم، تقریب: $g(Y) \approx (g(\mu), \sigma^۲(g'(\mu))^۲)$، به دست می‌آید. در این عبارت " $\approx$ " به معنی توزیع تقریبی است.

علاوه براین، اگر $Y \overset{a}{\sim} (\mu, \sigma^۲)$، آن‌گاه داریم:

$$g(Y) \overset{a}{\sim} (g(\mu), \sigma^۲(g'(\mu))^۲)$$

روش چندمتغیّره نیز به کار می‌رود. فرض کنید رابطهٔ زیر را داریم:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^۲ & \sigma_{xy} \\ & \sigma_y^۲ \end{pmatrix} \right)$$

همچنین فرض می کنیم، مایل به یافتن توزیع $g(X, Y)$ هستیم. بسط آن به شرح زیر است:

$$g(X,Y) = g(\mu_x, \mu_y) + (X-\mu_x)\frac{\partial}{\partial x}g(\mu_x, \mu_y) + (Y-\mu_y)\frac{\partial}{\partial y}g(\mu_x, \mu_y) + \cdots$$

در نتیجه، داریم:

$$g(X,Y) = \left(g(\mu_x, \mu_y), \sigma_x^۲(\frac{\partial}{\partial x}g)^۲ + ۲\sigma_{xy}\frac{\partial}{\partial x}g\frac{\partial}{\partial y}g + \sigma_y^۲(\frac{\partial}{\partial y}g)^۲ \right)$$

علاوه براین، اگر (X, Y) دارای توزیع مجانبی نرمال باشند، آن‌گاه $g(X, Y)$ نیز دارای توزیع مجانبی نرمال خواهد بود.

روش دلتا بسیار مفید است. مثلاً، می‌توان آن را برای محاسبهٔ تقریبی $\mathrm{Var}(\frac{\overline{X}}{\overline{Y}})$ یا $\mathrm{Var}(\overline{X}\,\overline{Y})$ به کار برد.

مثال۲ وایبل: اگر توزیع وایبل را با پارامتر $\gamma = \lambda^{\alpha}$ بنویسیم، می‌توان مشتقات آن را ساده‌تر محاسبه کرد.

$$S(t) = e^{-(\lambda t)^{\alpha}} = e^{-\gamma t^{\alpha}}$$

$$f(t) = \gamma \alpha t^{\alpha-1} . e^{-\gamma t^{\alpha}}$$

در این صورت، داریم:

$$L = (\gamma\alpha)^{n_u}\left(\prod_u t_i^{\alpha-1}\right)\exp\left(-\gamma\sum_u t_i^{\alpha}\right)\exp\left(-\gamma\sum_c c_i^{\alpha}\right)$$

$$= (\gamma\alpha)^{n_u}\left(\prod_u t_i^{\alpha-1}\right)\exp\left(-\gamma\sum_{i=1}^{n} y_i^{\alpha}\right)$$

$$\log L = n_u \log\gamma + n_u \log\alpha + (\alpha-1)\sum_u \log t_i - \gamma\sum_{i=1}^{n} y_i^{\alpha}$$

$$\frac{\partial}{\partial\gamma}\log L = \frac{n_u}{\gamma} - \sum_{i=1}^{n} y_i^{\alpha}$$

$$\frac{\partial}{\partial\alpha}\log L = \frac{n_u}{\alpha} + \sum_u \log t_i - \gamma\sum_{i=1}^{n} y_i^{\alpha}\log y_i$$

بنابراین، برآورد حداکثر درست‌نمایی $(\hat{\alpha}, \hat{\gamma})$ به شرح زیر است:

$$\hat{\gamma} = \frac{n_u}{\sum_{i=1}^{n} y_i^{\hat{\alpha}}}$$

$$\frac{n_u}{\hat{\alpha}} + \sum_u \log t_i - \hat{\gamma}\sum_{i=1}^{n} y_i^{\hat{\alpha}}\log y_i = 0$$

این معادلات را باید به روش عددی حل کرد. روش نیوتن-رافسون بــه مــاتریس اطّــلاع فیشر نمونهٔ زیر -که در مسأله ۳ محاسبه شد- نیاز دارد.

$$-\frac{\partial^{2}}{\partial \underline{\theta}^{2}}\log L=-\begin{pmatrix}\frac{\partial^{2}}{\partial\gamma^{2}}\log L & \frac{\partial^{2}}{\partial\gamma\,\partial\alpha}\log L\\ & \frac{\partial^{2}}{\partial\alpha^{2}}\log L\end{pmatrix}$$

روش نیوتن-رافسون به مقادیر اولّیهٔ $\hat{\gamma}_{\circ}$ و $\hat{\alpha}_{\circ}$ نیز نیاز دارد. توجّه نمایید، که برای محاسبهٔ مقادیر اولّیهٔ معقول، روابط زیر قابل استفاده است:

$$S(t)=e^{-\gamma t^{\alpha}}$$

$$\log S(t)=-\gamma t^{\alpha}$$

$$\log(-\log S(t))=\log\gamma+\alpha\log t$$

بنابراین، اگر برآورد $\hat{S}(t_i)$ را داشته باشیم، به کمک رگرســیون $\log(-\log\hat{S}(t_i))$ روی $\log t_i$ و محاسبهٔ ضریب رگرسیون $\hat{\alpha}$ و مقدار ثابت $\log\hat{\gamma}_{\circ}$، نتیجه مطلوب حاصل می‌شود. مقدار انتخابی ممکن $\hat{S}(t_i)$، همان برآورد کاپلان-میر است، که بعداً بحث خواهــد شــد. همچنین می‌توان از یک تابع توزیع تجربی -که از برش چشم‌پوشـــی می‌کنــد- اســتفاده کرد.

REFERENCE

Cohen, Technometrics (1965), treats the MLE and gives additional references.

برآورد $S(t)$. برآورد تابع بقاء، یکی از اهداف اصلی تحلیل بقاء است.

$$S(t)=\exp\left(-\int_{\circ}^{t}\lambda(u)\,du\right)$$

برای مثال، یکی از مقیاسهای معالجهٔ سرطان این است که احتمال زنده بودن بیمــار را بــه مدّت حداقل پنج سال حساب کنیم. در مطالعات مهندسی تابع بقاء را تابع اعتماد نـــامند و معمولاً آن را با $R(t)$ نشان می‌دهند. برای مثال، دانستن قابلیّت اعتماد یک مؤلّفه در یک دستگاه بعد از هزار ساعت کار آن دستگاه.

با داشتن MLE، برآورد تابع بقاء در دو حالت نمایی یا وایبل بسیار ساده است.

حالت نمایی $\hat{S}(t) = e^{-\hat{\lambda} t}$

حالت وایبل $\hat{S}(t) = e^{-(\hat{\lambda} t)^{\hat{\alpha}}} = e^{-\hat{\gamma} t^{\hat{\alpha}}}$

همچنین، برای هر مقدار ثابت t، تابع بقاء $\hat{S}(t)$ خود تابعی از $\hat{\lambda}$ یا $(\hat{\gamma}, \hat{\alpha})$ است. بنابراین به دست آوردن تقریبی از توزیع $\hat{S}(t)$ به روش دلتا امکان‌پذیر است. علاوه براین، می‌توان با تبدیل log log، همگرایی به نرمال را بهبود بخشید.

در حالت نمایی، داریم:

$$S(t) = e^{-\lambda t}$$

$$\log[-\log S(t)] = \log \lambda + \log t$$

$$\log[-\log \hat{S}(t)] = \log \hat{\lambda} + \log t$$

$$\widehat{\mathrm{Var}}\{\log[-\log \hat{S}(t)]\} \cong \frac{1}{n_u}$$

در حالت وایبل، داریم:

$$S(t) = e^{-\gamma t^{\alpha}}$$

$$\log[-\log S(t)] = \log \gamma + \alpha \log t$$

$$\log[-\log \hat{S}(t)] = \log \hat{\gamma} + \hat{\alpha} \log t$$

$$\widehat{\mathrm{Var}}\{\log[-\log \hat{S}(t)]\} \cong \frac{\mathrm{Var}(\hat{\gamma})}{\hat{\gamma}^2} + 2\mathrm{Cov}(\hat{\gamma}, \hat{\alpha}) \frac{\log t}{\hat{\gamma}} + \mathrm{Var}(\hat{\alpha})(\log t)^2$$

۲.۲ ترکیب خطّی آماره‌های مرتّب

در این بخش فقط توزیع وایبل بررسی می‌شود. روش کار قابل تعمیم است.

با تغییر پارامتر و تبدیل مناسب می‌توان مسألهٔ برآورد λ و α را در توزیع وایبل به برآورد پارامتر مبدأ و مقیاس تبدیل کرد. با دوباره نویسی داریم:

$$P(Y > t) = e^{-(\lambda t)^{\alpha}} = \exp\{-\exp[\alpha(\log \lambda + \log t)]\} = \exp\left\{-\exp\left(\frac{\log t - \mu}{\sigma}\right)\right\}$$

در روابط بالا، $\mu = -\log\lambda$ و $\sigma = \frac{1}{\alpha}$ است. در این صورت داریم:

$$P(\log Y > t) = P(Y > e^{t}) = \exp\left\{-\exp\left(\frac{t-\mu}{\sigma}\right)\right\}$$

در معادلهٔ قبل دیده می‌شود که μ و σ همان پارامترهای موقعیّت و مقیاس متغیّر تصادفی $\log Y$ هستند. این نتیجه بسیار مهّم است، زیرا در آمار قضایای زیادی برای برآورد دو پارامتر موقعیّت و مقیاس وجود دارد.

فرض کنید، می‌خواهیم احتمال بقاء را برای زمان ثابت $t_{\circ}$ حساب کنیم، داریم:

$$S(t_{\circ}) = P(Y > t) = \exp\left\{-\exp\left(\frac{\log t_{\circ} - \mu}{\sigma}\right)\right\}$$

با تعریف: $Y^{\circ} = \frac{Y}{t_{\circ}}$ و $\mu_{\circ} = \mu - \log t_{\circ}$، می‌توان عبارت را کوتاهتر نوشت. داریم:

$$S(t_{\circ}) = P(\log Y^{\circ} > \circ) = \exp\left\{-\exp\left(\frac{\mu_{\circ}}{\sigma}\right)\right\}$$

در این رابطه $\mu_{\circ}$ و σ، پارامترهای موقعیّت و مقیاس متغیّر تصادفی $\log Y^{\circ}$ هستند. اگر بتوانیم یک بازهٔ اطمینان برای نسبت $\frac{\mu_{\circ}}{\sigma}$ پیدا کنیم، با دوبار استفاده از تبدیل نمایی، یک بازهٔ اطمینان برای $S(t_{\circ})$ به دست می‌آید.

با استفاده از ترکیب خطّی آماره‌های مرتّب داریم:

$$\hat{\mu} = \sum_{i=1}^{n} a_i \log Y^{\circ}_{(i)} \qquad \text{و} \qquad \hat{\sigma} = \sum_{i=1}^{n} b_i \log Y^{\circ}_{(i)}$$

در این دو رابطه $\sum_{i=1}^{n} a_i = 1$ و $\sum_{i=1}^{n} b_i = \circ$ و a_1 تا a_n و همچنین b_1 تا b_n به گونه‌ای انتخاب می‌شوند که در یک شرط مجانبی بهینه صدق کنند. این روش به خصوص برای برش نوع دوم بسیار مناسب است، زیرا داریم:

$$a_{r+1} = \cdots = a_n = \circ$$

$$b_{r+1} = \cdots = b_n = \circ$$

بنابراین، برآوردها بر مبنای مشاهدات بریده نشده انجام می‌پذیرد.

REFERENCE
Johns and Lieberman, Technometrics (1966).

توزیع‌های فرین

تابع زیر یکی از سه توزیع حدّی فرین است.

$$G_1 = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty$$

یک توزیع حدّی فرین، توزیعی است مانند G، به گونه‌ای که اگر X_1، ... و X_n، متغیّرهای iid با توزیع F باشند، آنگاه $\max\{X_1,\ldots,X_n\}$ -با فرض این‌که به طور مناسب استاندارد شده باشند- در توزیع به G همگراست. یکی دیگر از توزیعهای حدّی به شرح زیر است:

$$G_2(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0 \\ 1 & x > 0 \end{cases}$$

اگر دنبالهٔ فوقانی توزیع وایبل به طور مناسب مقیاس‌بندی شود، برابر با دنبالهٔ تحتانی G_2 است. همچنین، دنبالهٔ فوقانی توزیع لگاریتم متغیّر وایبل (۳) برابر با دنبالهٔ تحتانی G_1 است. تابع توزیع G_2 از استاندارد کردن حدّ متغیّر زیر به دست می‌آید:

$$\max\{X_1,\ldots,X_n\} - x_0$$

در این رابطه x_0 نقطهٔ قطع فوقانی است (یعنی: $F(x_0)=1$ و $F(x_0-)<1$). در نتیجه داریم:

$$-\max\{X_1 - x_0,\ldots,X_n - x_0\} = \min\{x_0 - X_1,\ldots,x_0 - X_n\}$$

یک متغیّر تصادفی وایبل را می‌توان به صورت کمینه (یعنی اوّلین شکست) تعداد زیادی زمانهای شکست بالقوه، تفسیر کرد. دستگاه با شکست اوّلین مؤلّفه از کار می‌افتد.

۳.۲ برآوردگرهای دیگر

در این بخش فرض می‌شود که برآوردگرها دارای الگوی نمایی و بدون برش هستند.

برآوردگرهای اریب اصلاح شده. در این مبحث، روش کار دارای اهمیّت بیشتری از

نتایج به دست آمده است. فرض کنید احتمال بقاء را به صورت زیر برآورد می کنیم، که در آن $\bar{T}=\frac{1}{n}\sum_{i=1}^{n}T_i$ است.

$$\hat{S}(t)=e^{-\hat{\lambda}t}=e^{-t/\bar{T}}$$

$$E[\hat{S}(t)]\neq e^{-\lambda t}$$

در نتیجه $\hat{S}(t)$ یک برآورد اریب است. می‌توان به روش دلتا اریبی را کاهش داد. با فرض: $\theta=E(T)$، داریم:

$$e^{-t/\bar{T}}=e^{-t/\theta}+(\bar{T}-\theta)\frac{t}{\theta^2}e^{-t/\theta}+\frac{1}{2}(\bar{T}-\theta)^2\left[\left(\frac{t}{\theta^2}\right)^2-\frac{2t}{\theta^3}\right]e^{-t/\theta}+\cdots$$

$$E(e^{-t/\bar{T}})=e^{-t/\theta}+0+\frac{1}{2}\frac{\theta^2}{n}\left[\left(\frac{t}{\theta^2}\right)^2-\frac{2t}{\theta^3}\right]e^{-t/\theta}+\cdots$$

$$=\left[1+\frac{1}{2n}\left(\frac{t^2}{\theta^2}-\frac{2t}{\theta}\right)\right]e^{-t/\theta}+\cdots$$

در نتیجه:

$$\tilde{S}(t)=\frac{e^{-\hat{\lambda}t}}{1+\frac{1}{2n}(t^2\hat{\lambda}^2-2t\hat{\lambda})}$$

$\tilde{S}(t)$ محاسبه شده در بالا باید اریبی کمتری از برآوردگر $\hat{S}(t)$ داشته باشد. علاوه براین، ثابت می‌شود که، میانگین مربع خطای $\tilde{S}(t)$ کمتر از $\hat{S}(t)$ است.
برآورد جک‌نایف، که دربارهٔ آن بعداً بحث می‌شود، نیز تصحیح اریبی بالا را به همراه دارد.

برآوردگرهای نااریب با حداقل واریانس (UMVUE). در محاسبهٔ UMVUE برای $S(t)$ از برآورد نااریب $U=I(T_1>t)$، استفاده می‌شود. در این صورت با در نظر گرفتن آمارهٔ بسنده $S=\sum_{i=1}^{n}T_i$ و قضیهٔ رائو-بلکول، UMVUE معادل $E(U|S)$ است. در

نتیجه داریم:

$$\tilde{S}(t) = E(U|S=s) = \left(1-\frac{t}{s}\right)^{n-1} . I(t<s)$$

برآوردهای بیزی. فقط یادآوری می‌کنیم که برآوردهای بیز را می‌توان بـا اسـتفاده از توزیع پیشین گاما به دست آورد.

REFERENCES

Basu, Technometrics (1964), derives UMVUEs.

Zacks and Even, JASA (1966), compares mean square errors.

Gaver and Hoel, Technometrics (1970), look at estimators in the framework of sampling from a Poisson process.

۳ الگوهای رگرسیونی

در کاربردهای پزشکی امکان دارد که مدّت بقاء به مقدار دارو یا تشعشع بسـتگی داشته باشد. همچنین، در کاربردهای مهندسی طول عمر یک لامپ ممکن است به دما یـا عوامل دیگر وابسته باشد. فرض کنید Y متغیّر وابسته و X متغیّر مستقل باشد. دو الگـوی زیر برای توزیع نمایی پیشنهاد می‌شود:

(الف) الگوی خطّی

$$E(T) = \alpha + \beta X$$

در محاسبهٔ برآوردها، از روش حداکثر درست‌نمایی استفاده می‌کنیم. عیب این الگو ایـن است که اگر $\hat{\beta}$ منفی باشد، امکان دارد برآورد $E(T)$ منفی باشد.

(ب) الگوی خطّی لگاریتمی

$$E(T) = \alpha e^{\beta X}$$

$$\log E(T) = \log\alpha + \beta X$$

در این‌جا نیز از برآورد حداکثر درست‌نمایی استفاده می‌شود.

REFERENCES

Feigl and Zelen, Biometrics (1965), discuss the uncensored case for both the linear and log – linear models.

Zippin and Armitage, Biometrics (1966), discusses the censored case

for the linear model.
Glasser, JASA (1967), discusses the censored case for the log – linear model.
Zippin and Lamborn, Stanford Univ. Tech. Report No. 20 (1969), discuss the censored case for the log – linear model goodness of fit tests.
Mantel and Myers, JASA (1971), discuss the censored case for the multiple linear model.

۴ الگوهایی با کسرهای بقاء

۱.۴ نمونه‌ای به حجم واحد.

فرض کنید:

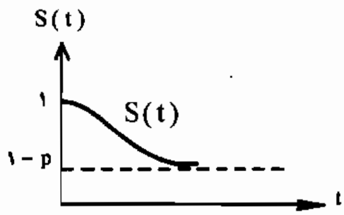
$$p = P(\text{مرگ}) \qquad \text{و} \qquad ۱ - p = P(\text{بقاء})$$

نسبت $۱ - p$ را کسر بقاء نامند. با فرض: $P(T \le t | \text{مرگ}) = ۱ - e^{-\lambda t}$، تابع درستنمایی به صورت زیر است:

$$L(y, \delta) = \begin{cases} p\lambda e^{-\lambda y} & \delta = ۱ \quad (\text{بدون برش}) \\ (۱-p) + pe^{-\lambda y} & \delta = ۰ \quad (\text{با برش}) \end{cases}$$

در محاسبهٔ برآوردها از حداکثر درستنمایی استفاده می‌کنیم.

الگوهای با کسر بقاء را گاهی برای آزمایشهای کوتاه به کار می‌برند. در این حالت فرض نمی‌شود که تابع بقاء $S(t)$ لزوماً به صفر نزدیک می‌شود. در عوض امکان دارد $S(t)$ مطابق شکل زیر باشد:

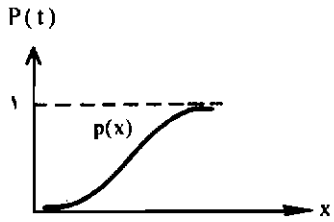


۲.۴ رگرسیون

$p(x)$ را به صورت زیر فرض می‌کنیم

$$p(x) = P(\text{مرگ} | x) = \frac{e^{\alpha+\beta x}}{۱ + e^{\alpha+\beta x}}$$

این تابع را لوجستیک گویند و به صورت نمودار زیر است:



همچنین فرض کنید: $P(T \leq t | \text{مرگ}) = ۱ - e^{-\lambda t}$. در این صورت تابع درستنمایی به صورت زیر خواهد بود:

$$L(y, \delta, x) = \begin{cases} p(x)\lambda e^{-\lambda y} & \delta = ۱ \quad (\text{بدون برش}) \\ ۱ - p(x) + p(x)e^{-\lambda y} & \delta = ۰ \quad (\text{با برش}) \end{cases}$$

برای به دست آوردن برآوردها از حداکثر درستنمایی استفاده می‌شود.

REFERENCE

Farewell, Biometrika (1977).

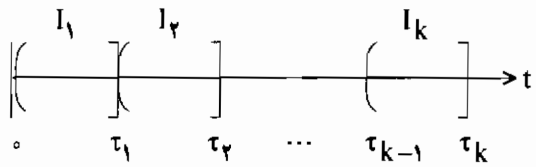
فصل سوم

روشهای ناپارامتری (یک نمونه)

۱ جدولهای طول عمر

روش رسمی برآورد $S(t)$ در علوم بیماریهای مسری و بیمه‌گری، همان روش بیمه‌گری است که به جدول طول عمر بستگی دارد و در زیر آن را شرح می‌دهیم.

فرض کنید زمان به دنباله‌های ثابتی از بازه‌های I_1، ...، I_k، افراز شده باشد. این بازه‌ها تقریباً همیشه (و نه لزوماً) مساوی در نظر گرفته می‌شوند. برای جمعیت انسانها این فاصله‌ها معمولاً یک سال است.



بنا به تعریف، در یک جدول طول عمر موارد زیر فرض می‌شوند:

n_i = تعداد افراد زنده در شروع بازهٔ I_i.

d_i = تعداد افراد مرده در طول بازهٔ I_i.

ℓ_i = تعداد گم شده‌ها در طول بازهٔ I_i.

w_i = تعداد خارج شده‌ها در طول بازهٔ I_i.

p_i = احتمال زنده بودن در طول بازهٔ I_i، در صورتی که در ابتدای بازه زنده بوده است.

$q_i = 1 - p_i$

جدول ۱، مثالی از یک جدول طول عمر است. I_1 تا I_5، هر کدام به مدّت یکسال‌اند. ستون (۲) شامل n_i، ستون (۳) شامل d_i، ستون (۴) شامل ℓ_i و ستون (۵) شامل w_i است. می‌خواهیم $S(5)$ را برآورد کنیم.

جدول ۱. محاسبهٔ نرخ بقاء پنج ساله

سالهای بعد از تشخیص بیماری (۱)	تعداد زنده‌ها در ابتدای بازه (۲)	مرده در طول بازه (۳)	گم شده‌ها (۴)	خارج شده‌ها (۵)	تعداد مؤثر در معرض خطر مرگ $(۲)-\frac{۱}{۲}[(۴)+(۵)]$ (۶)	نسبت مرگ و میر $(۳)/(۶)$ (۷)	نسبت زنده‌ها $۱-(۷)$ (۸)	نسبت تجمعی زنده‌ها تا پایان بازه $\prod_{i=۱}^{k}(۸)_i$ (۹)
۰-۱	۱۲۶	۴۷	۴	۱۵	۱۱۶٫۵	۰٫۴	۰٫۶۰	۰٫۶۰
۱-۲	۶۰	۵	۶	۱۱	۵۱٫۵	۰٫۱	۰٫۹۰	۰٫۵۴
۲-۳	۳۸	۲	—	۱۵	۳۰٫۵	۰٫۰۷	۰٫۹۳	۰٫۵۰
۳-۴	۲۱	۲	۲	۷	۱۶٫۵	۰٫۱۲	۰٫۸۸	۰٫۴۴
۴-۵	۱۰	—	—	۶	۷	۰٫۰۰	۱٫۰۰	۰٫۴۴

مرجع: Cutler and Ederer, J. Chronic Dis. (1958).

۱.۱ روش کاهش نمونه

برای برآورد $S(\tau_k)$، تنها افرادی را مورد توجّه قرار می‌دهیم که در بازهٔ $[0,\tau_k)$ در معرض خطر قرار دارند (یعنی: در بازهٔ مورد توجّه). موارد زیر را داریم:

$$n = n_1 - \sum_{i=1}^{k} \ell_i - \sum_{i=1}^{k} w_i$$

$$d = \sum_{i=1}^{k} d_i$$

$$\hat{S}(\tau_k) = 1 - \frac{d}{n}$$

در مثال جدول (۱)، داریم: $n = 126 - 12 - 56 = 60$ و $d = 56$ و $\hat{S}(5) = 1 - \frac{56}{60} = 0/067$

عیب روش کاهش نمونه این است که از اطّلاعات موجود در ℓ_i و w_i چشم‌پوشی می‌کند. این یک برآورد اریب (نقصانی) از $S(t)$ است.

۲.۱ روش بیمه‌گری

می‌توان احتمال بقاء $S(\tau_k)$ را به صورت حاصل‌ضرب چند احتمال تجزیه کرد. داریم:

$$\begin{aligned} S(\tau_k) &= P(T > \tau_k) \\ &= P(T > \tau_1).P(T > \tau_2 \mid T > \tau_1) \cdots P(T > \tau_k \mid \tau_{k-1}) \\ &= p_1 \cdot p_2 \cdots p_k \end{aligned}$$

در این رابطه: $p_i = P(T > \tau_i \mid T > \tau_{i-1})$.

روش بیمه‌گری، یک برآوردی جدا برای هر p_i، ارائه می‌دهد. سپس، با ضرب این برآوردها، برآورد $S(\tau_k)$ نتیجه می‌شود.

در برآورد p_i، می‌توان از $1 - \frac{d_i}{n_i}$، استفاده کرد، به شرطی که مشاهده‌ای در I_i گم یا خارج نشده باشد. با این وجود، اگر ℓ_i و w_i صفر نباشند، فرض می‌کنیم که به‌طور متوسّط، افرادی که در بازهٔ I_i حذف شده‌اند، در معرض خطر در نصف بازه بوده‌اند. بنابراین، حجم مؤثّر نمونه را به صورت زیر تعریف می‌کنیم:

$$n'_i = n_i - \frac{۱}{۲}(\ell_i + w_i) \quad و \quad \hat{q}_i = \frac{d_i}{n'_i} \quad و \quad \hat{p}_i = ۱ - \hat{q}_i$$

در نتیجه، برآورد روش بیمه‌گری معادل: $\hat{S}(\tau_k) = \prod_{i=۱}^{k} p_i$ است. در جدول (۱)، ستون (۶) شامل n'، ستون (۷) شامل $\hat{q}_i$، ستون (۸) شامل $\hat{p}_i$ و در ستون (۹) دیده می‌شود که $S(۵) = ۰/۴۴$ است.

برای بهبودبخشی در پیدا کردن جانشینی برای حجم نمونهٔ مؤثّر، تلاشهای زیادی به انجام رسیده است. ولی اگر یک برآورد دقیقتری برای $S(t)$ لازم شود، حدّ حاصل‌ضرب برآوردگر کاپلان-میر، روش مناسبی خواهد بود.

۳.۱ واریانس $\hat{S}(\tau_k)$

برای برآورد واریانس $\hat{S}(\tau_k)$، از رابطهٔ زیر استفاده می‌شود:

$$\log \hat{S}(\tau_k) = \sum_{i=۱}^{k} \log \hat{p}_i$$

با فرض $n'_i \hat{p}_i \sim \text{Binomial}(n'_i, p_i)$ و استفاده از روش دلتا، داریم:

$$\text{Var}(\log \hat{p}_i) \cong \text{Var}(\hat{p}_i)\left[\frac{d}{dp_i}(\log p_i)\right]^۲$$

$$\cong \frac{p_i q_i}{n'_i} \cdot \frac{۱}{p_i^۲} = \frac{q_i}{n'_i p_i}$$

فرض می‌شود $\log \hat{p}_۱$ و $\cdots$ و $\log \hat{p}_k$ مستقل‌اند.

$$\text{Var}[\log \hat{S}(\tau_k)] \cong \sum_{i=۱}^{k} \frac{q_i}{n'_i p_i}$$

$$\widehat{\text{Var}}[\log \hat{S}(\tau_k)] = \sum_{i=۱}^{k} \frac{\hat{q}_i}{n'_i \hat{p}_i} = \sum_{i=۱}^{k} \frac{d_i}{n'_i (n'_i - d_i)}$$

حال با استفادهٔ دوباره از روش دلتا، داریم:

$$\widehat{\text{Var}}[\hat{S}(\tau_k)] \cong \hat{S}^۲(\tau_k) \sum_{i=۱}^{k} \frac{d_i}{n'_i (n'_i - d_i)}$$

تساوی فوق به "رابطهٔ گرینوود" معروف است.

۴.۱ انواع جدولهای طول عمر

جدول (۱) و (۲)، مثالی از یک جدول طول عمر گروهی است. یک گروه یا دسته، مجموعه‌ای از افراد است که در جریان بررسی، مورد مطالعه قرار دارند. افراد در معرض خطر در ابتدای بازهٔ I_i، همان افرادی هستند که در بازهٔ قبلی (I_{i-1}) زنده مانده‌اند (نمرده یا گم یا خارج نشده‌اند).

نوعی دیگر از جدول طول عمر، جدول طول عمر جاری است. در این جدول گروهی از افراد با سنّ τ_{i-1} در نظر گرفته می‌شوند، که در ابتدای بازهٔ $I_i = (\tau_{i-1}, \tau_i]$ در معرض خطر قرار دارند. این گروه از افراد، کاملاً متفاوت از افرادی هستند که در بازهٔ قبلی I_{i-1} در معرض خطر بوده‌اند. نوعاً، گروههای سنّی مختلف در جمعیت در یک زمان بررسی می‌شوند.

REFERENCES

Berkson and Gage, Proc. Staff Meet. Mayo Clin. (1950).

Cutler and Ederer, J. Chronic Dis. (1958).

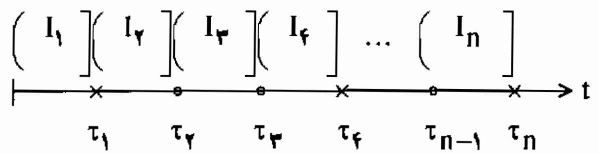
Elveback, JASA (1958).

Chiang, Stochastic Processes in Biostatistics (1968), Chapter 9.

Breslow and Crowley, Ann. Stat. (1974).

۲ برآوردگر حدّی حاصل‌ضرب (کاپلان - میر)

برآوردگر حدّی حاصل‌ضرب (PI)، همانند برآوردگر بیمه‌گری است با این تفاوت که طول بازه‌های I_i متغیّراند. در واقع τ_i، انتهای راست بازه I_i مشاهدهٔ بریده شده یا بریده نشده مرتبه iام است.



نماد " × ": نشانگر بریده نشده و **نماد** " o ": نشانگر بریده شده هستند.

به خاطر داشته باشید، که مشاهدات تکراری وجود ندارند و زوجهای (Y_1, δ_1) تا (Y_n, δ_n) را مشاهده می‌کنیم. فرض کنید $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$، آماره‌های مرتّب

Y_1 تا Y_n باشند. همچنین، فرض کنید: $\delta_{(i)}$ مقدار δ متناظر با $Y_{(i)}$ باشد. یعنی: اگر $Y_{(i)} = Y_j$، آن‌گاه $\delta_{(i)} = \delta_j$، باشد. توجّه شود که $\delta_{(1)}$ تا $\delta_{(n)}$ مرتّب نیستند، فرض کنید $R(t)$ تابع خطر در زمان t باشد. یعنی: مجموعهٔ افرادی که تا زمان t^- زنده هستند. همچنین n_i تعداد اعضای $R(Y_i)$، که در زمان $Y_{(i)}^-$ زنده هستند، d_i تعداد افرادی که در زمان $Y_{(i)}$ فوت شده و p_i، احتمال زنده بودن در طول I_i باشد (به شرطی که در ابتدای بازه I_i زنده باشند). به عبارت معادل:

$$p_i = P(T > \tau_i \mid T > \tau_{i-1}) \qquad \text{و} \qquad q_i = 1 - p_i$$

برآوردهای $\hat{p}_i$ و $\hat{q}_i$، به شرح زیراند:

$$\hat{p}_i = 1 - \hat{q}_i = \begin{cases} 1 - \dfrac{1}{n_i} & \delta_{(i)} = 1 \quad (\text{بریده نشده}) \\ 1 & \delta_{(i)} = 0 \quad (\text{بریده شده}) \end{cases}$$

حال، برآورد PL، در صورت نبود تکرار، به شرح زیر است:

$$\hat{S}(t) = \prod_{y_{(i)} \le t} \hat{p}_i = \prod_{u: y_{(i)} \le t} \left(1 - \frac{1}{n_i}\right) = \prod_{y_{(i)} \le t} \left(1 - \frac{1}{n_i}\right)^{\delta_{(i)}}$$

$$= \prod_{y_{(i)} \le t} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{(i)}} = \prod_{y_{(i)} \le t} \left(\frac{n - i}{n - i + 1}\right)^{\delta_{(i)}}$$

REFERENCE
Kaplan and Meier, JASA (1958).

یادآوریهای لازم:

(الف) برای مشاهدات مکرّر بریده نشده، فرض کنید قبل از زمان τ، دقیقاً n فرد زنده هستند، و در زمان τ، تعداد d تن فوت می‌شوند. فاصلهٔ d فوت را به فاصله‌های بی‌نهایت کوچک تقسیم می‌کنیم به گونه‌ای که عامل مربوط به d فوت در برآوردهای حدّی حاصل‌ضرب به شکل زیر در آید:

$$\left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{m-1}\right) \cdots \left(1 - \frac{1}{m-d-1}\right) = \frac{m-d}{m} = 1 - \frac{d}{m}$$

(ب) اگر مشاهدات بریده شده و بریده نشده تکراری باشند، مشاهدات بریده نشده را قبل از مشاهدات بریده شده در نظر بگیرید.

(پ) اگر آخرین مشاهده (مرتّب شده) $y_{(n)}$ بریده شده باشد، آن‌گاه برای $\hat{S}(t)$، به‌طوری که در بالا تعریف شد، داریم:

$$\lim_{t\to\infty} \hat{S}(t) > 0$$

گاهی بهتر است که برای $t \ge y_{(n)}$ تعریف کنیم $\hat{S}(t)=0$، یا تصوّر کنیم که اگر $\delta_{(n)}=0$ باشد، $t > y_{(n)}$ تعریف نشده است.

با توجه به یادآوریهای (الف) و (ب)، $Y'_{(1)} < Y'_{(2)} < \cdots < Y'_{(r)}$، زمانهای جداگانهٔ حیات را نشان می‌دهند. همچنین:

$$\delta'_i = \begin{cases} 1 & \text{اگر مشاهده در زمان } y'_{(j)} \text{ بریده نشده باشد} \\ 0 & \text{اگر مشاهده در زمان } y'_{(j)} \text{ بریده شده باشد} \end{cases}$$

$n_j =$ تعداد $R(y'_{(j)})$ ها

$d_j =$ تعداد فوت شده‌ها در زمان $y'_{(j)}$

با توجّه به موارد بالا، برآورد PL برای حالت تکراری، به شرح زیر است:

$$\hat{S}(t) = \prod_{u:y'_{(j)} \le t} \left(1-\frac{d_j}{n_j}\right) = \prod_{y'_{(j)} \le t} \left(1-\frac{d_j}{n_j}\right)^{\delta'_{(j)}}$$

مثال. مطالعهٔ درمان AML: یک آزمایش بالینی به منظور ارزشیابی تأثیر یک روش شیمیایی روی بیماری (AML) صورت گرفته است. بعد از رسیدن به مرحلهٔ خاصّی، بیماران به تصادف به دو گروه تقسیم شده‌اند. گروه اوّل تحت درمان شیمیایی قرار گرفته‌اند و گروه دوم یا شاهد، بدون درمان مورد مطالعه قرار گرفته‌اند. هدف این است که آیا درمان شیمیایی زمان شفایافتن را به تأخیر می‌اندازد. یعنی: آیا سبب افزایش زمان بهبودی می‌شود.

برای یک تحلیل اوّلیّه در طول دورهٔ آزمایش، داده‌ها (برحسب ۱۰/۷۴) به شرح صفحه بعد بوده است. طول بهبودی کامل برحسب هفته است.

گروه تیمار ۹ , ۱۳ , ۱۳⁺ , ۱۸ , ۲۳ , ۲۸⁺ , ۳۱ , ۳۴ , ۴۵⁺ , ۴۸ , ۱۶۱⁺

گروه شاهد ۵ , ۵ , ۸ , ۸ , ۱۲ , ۱۶⁺ , ۲۳ , ۲۷ , ۳۰ , ۳۳ , ۴۳ , ۴۵

برآوردگر PL (کاپلان-میر) برای گروه تیمار، به صورت زیر محاسبه می‌شود:

$$\hat{S}(۰)=۱$$

$$\hat{S}(۹)=\hat{S}(۰)\times\frac{۱۰}{۱۱}=۰/۹۱$$

$$\hat{S}(۱۳)=\hat{S}(۹)\times\frac{۹}{۱۰}=۰/۸۲$$

$$\hat{S}(۱۸)=\hat{S}(۱۳)\times\frac{۷}{۸}=۰/۷۲$$

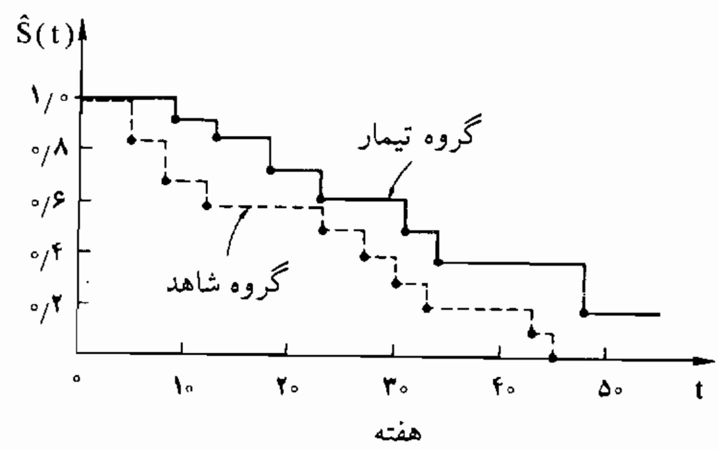
$$\hat{S}(۲۳)=\hat{S}(۱۸)\times\frac{۶}{۷}=۰/۶۱$$

$$\hat{S}(۳۱)=\hat{S}(۲۳)\times\frac{۴}{۵}=۰/۴۹$$

$$\hat{S}(۳۴)=\hat{S}(۳۱)\times\frac{۳}{۴}=۰/۳۷$$

$$\hat{S}(۴۸)=\hat{S}(۳۴)\times\frac{۱}{۲}=۰/۱۸$$

در نمودار (۳)، برآوردگرهای PL برای گروه تیمار و گروه شاهد نشان داده شده است.



نمودار ۳ برآورد تابع بقاء برای مطالعهٔ درمان AML

REFERENCE

Embury et al. , West. J. Med. (1977).

واریـانس $\hat{S}(t)$: با استفاده از روش محاسبهٔ واریانس بیمه‌گری و در حالت بدون تکرار، داریم:

$$\hat{Var}[\hat{S}(t)] = \hat{S}^2(t) \sum_{y_{(i)} \le t} \frac{\hat{q}_i}{n_i \hat{p}_i}$$

$$= \hat{S}^2(t) \sum_{y_{(i)} \le t} \frac{\delta_{(i)}}{(n-i)(n-i+1)}$$

در حالت وجود تکرار، داریم:

$$\hat{Var}[\hat{S}(t)] = \hat{S}^2(t) \sum_{y'_{(i)} \le t} \frac{\delta'_{(j)} d_j}{n_j(n_j - d_j)}$$

این تساویها به روابط گرین‌وود معروف است.

بررسی این روابط به روشنی جدول طول عمر نیست، زیرا تعـداد جمـلات حـاصل‌ضرب تصادفی است. همچنین، بیشتر جملات آن وابسته‌اند. با این وجود، بعداً به صورت تقریب مجانبی واریانس $\hat{S}(t)$، تحقیق خواهد شد.

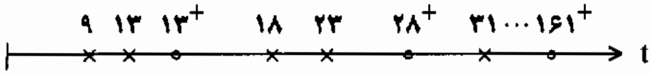
توماس و میر، سه روش مختلف ساختن بازهٔ اطمینان را مطالعه کرده‌اند. در یکی از روشها از تقریب واریانس $\hat{Var}[\hat{S}(t)]$، استفاده شده است. همچنین به یادآوریهای پایـان فصل ششم، بخش (۴.۱)، مراجعه فرمایید.

REFERENCE

Thomas and Grunkemeier, JASA (1975).

۱.۲ الگوریتم تجدید نظر در توزیع به راست

افرون، روش دیگری را برای محاسبهٔ برآوردگر PL معرّفی کرد. آن را بـا مثـال (AML) تشریح می‌کنیم. نمودار (n = ۱۱) زمان بقاء را رسم می‌کند.



برآورد معمولی $S(t)$، با فرض بریده نشدن به هر یک از زمانهای مشاهده شده، جـرم $\frac{1}{11}$ را نسبت می‌دهد. اوّلین زمان بریده شده را در نظر می‌گیریم (13^+)، چون فوتـی در 13^+

صورت نگرفته، لذا، در جایی و در سمت راست آن رخ می‌دهد. معقول به نظر می‌رسد که جرم $\frac{۱}{۱۱}$ نقطهٔ ۱۳^{+} را به‌طور مساوی بین تمام مشاهدات راست ۱۳^{+}، توزیع کنیم. بنابراین، جرم $\frac{۱}{۸}(\frac{۱}{۱۱})$ را به ۱۸، ۲۳، ۲۸^{+}، ... اضافه می‌کنیم. حال دومین مشاهدهٔ بریده شدهٔ (۲۸^{+}) را در نظر می‌گیریم. جرم $\frac{۱}{۱۱}+\frac{۱}{۸}(\frac{۱}{۱۱})$ نقطهٔ ۲۸^{+} را بین تمام نقاط سمت راست ۲۸^{+} توزیع می‌کنیم. در مورد بقیهٔ مشاهدات بریده شده نیز به‌طور مشابه عمل می‌کنیم. نتایج مربوط به PL در جدول صفحه بعد آمده است:

REFERENCE
Efron, Proc. Fifth Berkeley Symp. IV (1967), pp. 831-853.

۲.۲ خودسازگاری

برای سادگی، فرض می‌کنیم تکرار وجود ندارد. برآوردگر $\hat{SC}(t)$ را خودسازگار می‌نامند، اگر به شرح زیر باشد:

$$\hat{SC}(t)=\frac{1}{n}\left[\sum_{i=1}^{n}(1).I(y_{(i)}>t)+\sum_{i=1}^{n}(0).I(y_{(i)}\le t,\delta_{(i)}=1)+\sum_{i=1}^{n}\frac{\hat{SC}(t)}{\hat{SC}(y_{(i)})}I(y_{(i)}\le t,\delta_{(i)}=0)\right] \qquad (۴)$$

که در آن $\hat{SC}(t)\Big/\hat{SC}(y_{(i)})$، برآورد احتمال شرطی بقاء بعد از t است، با این شرط که در زمان $y_{(i)}$ زنده بوده است. توجّه کنید که (۴) معادل رابطهٔ زیر است:

$$\hat{SC}(t)=\frac{1}{n}\left[N_y(t)+\sum_{y_{(i)}\le t}(1-\delta_{(i)})\frac{\hat{SC}(t)}{\hat{SC}(y_{(i)})}\right] \qquad (۵)$$

که در آن $N_y(t)=\#(y_i>t)$.

(نماد" ( )#"، یعنی: تعدادی که در این شرط صدق می‌کنند.)

برآوردگر PL تنها برآوردگر خودسازگار برای $t<y_{(n)}$ است. اثبات به قرار زیر است:

$y_{(i)}$	جرم اوّلیه	جرم بعد از اوّلین تجدید	جرم بعد از دومین تجدید	جرم بعد از سومین تجدید	$\hat{S}(y_{(i)})$
۹	$\frac{۱}{۱۱}$ = ۰/۰۹	۰/۰۹	۰/۰۹	۰/۰۹	۰/۹۱
۱۳	۰/۰۹	۰/۰۹	۰/۰۹	۰/۰۹	۰/۸۲
۱۳^{+}	۰/۰۹	۰	۰	۰	
۱۸	⋮	۰/۰۹ + ($\frac{۱}{۸}$)(۰/۰۹) = ۰/۱۰	۰/۱۰	۰/۱۰	۰/۷۲
۲۳		۰/۱۰	۰/۱۰	۰/۱۰	۰/۶۱
۲۸^{+}		۰/۱۰	۰	۰	
۳۱		⋮	۰/۱۰ + ($\frac{۱}{۵}$)(۰/۱۰) = ۰/۱۲	۰/۱۲	۰/۴۹
۳۴			۰/۱۲	۰/۱۲	۰/۳۷
۴۵^{+}			۰/۱۲	۰	
۴۸			⋮	۰/۱۲ + ($\frac{۱}{۲}$)(۰/۱۲) = ۰/۱۸	۰/۱۸
۱۶۱^{+}				۰/۱۸	

با توجه به رابطهٔ (۵)، برآوردگر خودسازگار در شرایط زیر صدق می‌کند:

$$\hat{SC}(t)=\frac{N_y(t)}{n-\sum_{y_{(i)}\leq t}\left(\frac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)}$$

$$=\begin{cases}1 & t<y_{(1)} \\ \dfrac{N_y(t)}{n-\sum_{i=1}^{k}\left(\dfrac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)} & y_{(k)}\leq t<y_{(k+1)} \quad k=1,2,\dots,n-1\end{cases} \tag{۶}$$

باید ثابت کرد که اگر $\hat{SC}(t)$ در رابطهٔ (۶) صدق کند، آن‌گاه $\hat{SC}(t)$ با برآوردگر PL، $\hat{S}(t)$ برابر است. ابتدا، توجّه کنید اگر $t<y_{(1)}$، آن‌گاه

$$\hat{S}(t)=1-\hat{SC}(t)$$

همچنین، $\hat{S}(t)$ و $\hat{SC}(t)$ در بازهٔ $[y_{(k)}, y_{(k+1)})$ برای $k=1,2,\dots,n-1$، ثابت است. بنابراین، تنها کافی است نشان دهیم که جهش تابع $\hat{SC}(t)$ در نقطهٔ $y_{(k)}$ برابر جهش تابع $\hat{S}(t)$ است.

(الف) اگر $\delta_{(k)}=0$ باشد، از (۶) نتیجهٔ زیر به دست می‌آید:

$$N_y(y_{(k)}^-)-1=N_y(y_{(k)})=\hat{SC}(y_{(k)})\left[n-\sum_{i=1}^{k}\left(\frac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)\right]$$

$$=\hat{SC}(y_{(k)})\left[n-\sum_{i=1}^{k-1}\left(\frac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)\right]-1$$

$$=\hat{SC}(y_{(k)})\left[\frac{N_y(y_{(k)}-1)}{\hat{SC}(y_{(k)}-1)}\right]-1$$

در نتیجه:

$$\hat{SC}(y_{(k)})=\hat{SC}(y_{(k)}^{-})$$

پس اگر $\delta_{(k)}=0$ باشد، $\hat{SC}(t)$ در نقطهٔ $y_{(k)}$ دارای جهش نیست. در نتیجه در $t=y_{(k)}$ با $\hat{S}(t)$ برابر است.

(ب) اگر $\delta_{(k)}=1$ باشد، از (۶) نتیجهٔ زیر به دست می‌آید:

$$\hat{SC}(y_{(k)})=\frac{N_y(y_{(k)})}{n-\sum_{i=1}^{k}\left(\frac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)}=\frac{N_y(y_{(k)})}{N_y(y_{(k)}^{-})}\times\frac{N_y(y_{(k)}^{-})}{n-\sum_{i=1}^{k-1}\left(\frac{1-\delta_{(i)}}{\hat{SC}(y_{(i)})}\right)}$$

$$=\frac{n-k}{n-k+1}\hat{SC}(y_{(k)}^{-})$$

در نتیجه، اگر $\delta_{(k)}=1$ باشد، $\hat{SC}(t)$ در نقطهٔ $y_{(k)}$ دارای جهش است لذا داریم:

$$\frac{\hat{SC}(y_{(k)})}{\hat{SC}(y_{(k)}^{-})}=\frac{n-k}{n-k+1}$$

که در نقطهٔ $t=y_{(k)}$ با $\hat{S}(t)$ برابر است.

الگوریتم خودسازگاری

برآوردگر ساده زیر را در نظر می‌گیریم:

$$\hat{S}^{0}(t)=\frac{N_y(t)}{n}$$

این برآوردگر را با استفاده از رابطهٔ بازگشتی زیر می‌توان بهبود بخشید:

$$\hat{S}^{(j+1)}(t)=\frac{1}{n}\left[N_y(t)+\sum_{y_{(i)}\le t}(1-\delta_{(i)})\frac{\hat{S}^{(j)}(t)}{\hat{S}^{(j)}(y_{(i)})}\right]$$

در واقع، $\hat{S}^{(j)}(t)$ به‌طور یکنواخت در چند مرحلهٔ متناهی به برآوردگر PL، همگرا

می‌شود. این الگوریتم محاسباتی می‌تواند در مسائل کلّی برش مفید باشد.

REFERENCES

Efron, Proc. Fifth Berkeley Symp. IV (1967).

Turnbull, JASA (1974).

______, JRSS B (1976).

۳.۲ برآوردگر حدّاکثر درست‌نمایی تعمیم‌یافته

در شرایط معمولی، فرض می‌کنیم که مشاهدهٔ $\underline{X}$ دارای توزیع احتمال P_θ، صادق در رابطهٔ: $dP_\theta(\underline{x}) = f_\theta(\underline{x})d_\mu(\underline{x})$ است. $\mu(\underline{x})$، اندازهٔ کراندار برای ردهٔ $\{P_\theta\}$ است. به دست آوردن برآوردگر حدّاکثر درست‌نمایی معادل حدّاکثر کردن $L(\theta) = f_\theta(\underline{x})$، است.

در حالت مورد بحث، فرض کنید اندازهٔ احتمال مشاهده برابر P_F باشد، که به توزیع مجهول F بستگی دارد. ردهٔ $\{P_F\}$ دارای اندازهٔ کراندار نیست، بنابراین به یک تعریف کلّی‌تر برای حدّاکثر درست‌نمایی نیاز داریم.

کیفر و لفویتز، تعریف زیر را پیشنهاد نموده‌اند. فرض کنید $\mathcal{P} = \{P\}$ ردهٔ اندازه‌های احتمال باشد. برای دو عضو P_1 و P_2 در $\mathcal{P}$ تابع زیر، که مشتق رادون-نیکودین P_1 نسبت به $P_1 + P_2$ است، را تعریف می‌کنیم.

$$f(\underline{x}; P_1, P_2) = \frac{dP_1(\underline{x})}{d(P_1 + P_2)}$$

اندازهٔ احتمال $\hat{P}$ را برآوردگر حدّاکثر درست‌نمایی تعمیم یافته (GMLE) گوئیم هرگاه رابطهٔ زیر برای هر $P \in \mathcal{P}$ برقرار باشد:

$$f(\underline{x}; \hat{P}, P) \geq f(\underline{x}; P, \hat{P}) \qquad (۷)$$

این تعمیم، شامل تعریف معمولی MLE نیز هست.

برآوردگر کاپلان-میر PL، مقدار GMLE تابع F را ارائه می‌کند. اثبات به شرح زیر است:

اگر یک اندازهٔ احتمال $\hat{P}$ به $\underline{x}$ احتمال مثبت نسب دهد، آن‌گاه $f(\underline{x}; P, \hat{P}) = ۰$ است، مگر این‌که P نیز به $\underline{x}$ احتمال مثبتی نسبت دهد. پس، برای امتحان (۷) به ازای $P \in \mathcal{P}$، کافی است، آن را برای Pهایی امتحان کنیم که $P\{\underline{x}\} > ۰$ باشد. در نتیجه (۷) به صورت زیر در می‌آید:

$$\hat{P}(\underline{x}) \geq P(\underline{x}) \qquad (۸)$$

چون $\hat{S}$ به نقطهٔ $\underline{x}=[(y_۱,\delta_۱),\ldots,(y_n,\delta_n)]$، جرم مثبتی را نسبت می‌دهد. تنها کافی است آن اندازه‌های احتمال P را در نظر بگیریم که جرم مثبت به این نقطه نسبت می‌دهند. سپس، نشان دهیم که $\hat{S}$ مقدار احتمال $P\{(y_۱,\delta_۱),\ldots,(y_n,\delta_n)\}$ را حداکثر می‌کند. برای چنین P ی، رابطهٔ زیر را داریم:

$$L = P\{(y_۱,\delta_۱),\ldots,(y_n,\delta_n)\}$$

$$= \prod_{i=۱}^{n} P\{T=y_{(i)}\}^{\delta_{(i)}} P\{T>y_{(i)}\}^{۱-\delta_{(i)}}$$

فرض کنید P احتمال p_i را به فاصلهٔ نیم‌باز $[y_{(i)}, y_{(i+۱)})$ با شرط $y_{(n+۱)}=+\infty$، نسبت می‌دهد. برای مقادیر ثابت $p_۱$ تا p_n، تابع درست‌نمایی در ازاء $P\{T=y_{(i)}\}=p_i$ با شرط $\delta_{(i)}=۱$، حداکثر می‌شود. همچنین اگر $\delta_{(i)}=۰$ باشد، L به ازاء $P\{y_{(i)}<T<y_{(i+۱)}\}=p_i$ حداکثر می‌شود. پس برای مقادیر ثابت $p_۱$ تا p_n، مقدار حداکثر L به صورت زیر است:

$$\prod_{i=۱}^{n} p_i^{\delta_{(i)}} \left(\sum_{j=۱}^{n} p_j\right)^{۱-\delta_{(i)}}$$

با توجّه به مسألهٔ (۸)، دیده می‌شود که رابطهٔ (۹) به ازاء مقدار زیر حدّاکثر می‌شود:

$$\hat{p}_i = \prod_{j=۱}^{i-۱} \left(۱-\frac{\delta_{(i)}}{n-j+۱}\right)\frac{\delta_{(i)}}{n-i+۱}$$

که این متناظر با $\hat{S}$ است. اثبات در حالت تکراری به‌طور مشابه انجام می‌شود.

REFERENCES
Kiefer and Wolfowitz, Ann. Math. Stat. (1956).
Kaplan and Meier, JASA (1958).
Johansen, Scand. J. Stat. (1978).

۴.۲ سازگاری

می‌دانیم $S(t)$ به شرح زیر است:

$$S(t)=S_T(t)=P(T>t)=1-F(t)$$

تابع S^* را به صورت زیر تعریف می‌کنیم:

$$S^*(t)=S_Y(t)=P(Y>t)=1-H(t)$$
$$=[1-F(t)][1-G(t)]$$

توابع بقاء جزئی را به صورت زیر تعریف می‌کنیم:

$$S_u^*(t)=P\{Y>t,\delta=1\}=\int_t^{\infty}[1-G(u)]dF(u)$$

$$S_c^*(t)=P\{Y>t,\delta=0\}=\int_t^{\infty}[1-F(u)]dG(u)$$

در این صورت داریم:

$$S^*(t)=S_u^*(t)+S_c^*(t)$$

نشان خواهیم داد که $S(t)$ را می‌توان به صورت تابعی از $S_u^*(t)$ و $S_c^*(t)$ تعریف کرد.

(الف) فرض کنید $S_u^*(t)$ پیوسته است.

$$\int_0^t \frac{dS_u^*(u)}{S_u^*(u)+S_c^*(u)}=\int_0^t \frac{-[1-G(u)]dF(u)}{[1-F(u)][1-G(u)]}$$
$$=\int_0^t \frac{-dF(u)}{1-F(u)}=\log[1-F(u)]\Big|_0^t=\log S(t)$$

بنابراین داریم:

$$S(t)=\exp\left[\int_0^t \frac{dS_u^*(u)}{S_u^*(u)+S_c^*(u)}\right]$$

(ب) فرض کنید: S_u^* دارای جهش در t باشد. ولی S_c^* در t پیوسته باشد، داریم:

$$\log\frac{S_u^*(t^+)+S_c^*(t^+)}{S_u^*(t^-)+S_c^*(t^-)}=\log\frac{[1-F(t^+)][1-G(t^+)]}{[1-F(t^-)][1-G(t^-)]}$$

$$=\log\frac{[1-F(t^+)]}{[1-F(t^-)]}=\log\frac{S(t^+)}{S(t^-)}$$

(تساوی دوم از پیوستگی S_c^* در t به دست می‌آید، که در آن $G(t^+)=G(t^-)$)
بنابراین:

$$S(t^+)=S(t^-)=\exp\left\{\log\left[\frac{S_u^*(t^+)+S_c^*(t^+)}{S_u^*(t^-)+S_c^*(t^-)}\right]\right\}$$

اگر توزیعهای F و G جهش مشترک نداشته باشند، آن‌گاه از (الف) و (ب) رابطهٔ زیر نتیجه می‌شود:

$$S(t)=\exp\left\{c\int_0^t \frac{dS_u^*(u)}{S_u^*(u)+S_c^*(u)}+d\sum_{u\le t}\log\left[\frac{S_u^*(u^+)+S_c^*(u^+)}{S_u^*(u^-)+S_c^*(u^-)}\right]\right\} \qquad (۱۰)$$

در این رابطه $c\int$، یعنی انتگرال روی بازه‌های پیوستهٔ S_u^* و $d\sum$، یعنی مجموع روی نقاط جهش S_u^* است، عبارت (۱۰) را رابطهٔ پترسن نامند و نشان می‌دهد که $S(t)$ را می‌توان به صورت تابعی از S_u^* و S_c^* و t نشان داد؛ یعنی:

$$S(t)=\psi(S_u^*, S_c^*; t)$$

رابطهٔ پترسن ثابت می‌کند، که برآوردگر PL مربوط به $\hat{S}(t)$، سازگار است. اثبات آن به شرح زیر است. دو تابع توزیع تجربی را به صورت زیر تعریف می‌کنیم:

$$\hat{S}_u^*(t)=\frac{۱}{n}\sum_{i=۱}^{n} I(Y_i>t, \delta_i=۱)$$

$$\hat{S}_c^*(t)=\frac{۱}{n}\sum_{i=۱}^{n} I(Y_i>t, \delta_i=۰)$$

می‌توان دید که برآوردگر PL برابر زیر است:

$$\hat{S}(t)=\psi(\hat{S}_u^*, \hat{S}_c^*; t)$$

به شرط آن‌که هر تکرار بین مشاهدات بریده شده و نشده به عنوان یک مشاهدهٔ بریده نشده بعد از برش در نظر گرفته شود. توجّه شود که چون $\hat{S}_u^*$ گسسته است، $\psi(\hat{S}_u^*, \hat{S}_c^*; t)$، تنها شامل مجموع روی جهشهای $\hat{S}_u^*$ است.

بنابر قضیهٔ گلیونکو-کانتلی، داریم:

$$\hat{S}_u^*(t) \xrightarrow{a.s} S_u^*(t)$$

$$\hat{S}_c^*(t) \xrightarrow{a.s} S_c^*(t)$$ به‌طور یکنواخت در t

(نماد " $\xrightarrow{a.s}$ " همگرایی تقریباً همه جا را نشان می‌دهد). همچنین، ψ یک تابع پیوسته از S_u^* و S_c^* در اندازهٔ کوچکترین کران بالاست. یعنی اگر داشته باشیم:

$$\left\| S_u^* - S_u^{**} \right\| = \sup_t \left| S_u^*(t) - S_u^{**}(t) \right| \to ۰$$

و

$$\left\| S_c^* - S_c^{**} \right\| \to ۰$$

آن‌گاه داریم:

$$\psi(S_u^*, S_c^*; t) \to \psi(S_u^{**}, S_c^{**}; t)$$

بنابراین، داریم:

$$\hat{S}(t) = \psi(\hat{S}_u^*, \hat{S}_c^*; t) \xrightarrow{a.s} \psi(S_u^*, S_c^*; t) = S(t)$$

REFERENCE
Peterson, JASA (1977).

۵.۲ نرمال مجانبی

نشان می‌دهیم که اگر F و G بر $[۰, T]$ پیوسته و $F(t) < ۱$ باشد، با شرط $n \to \infty$، آن‌گاه:

$$Z_n(t) = \sqrt{n}\,[\hat{S}(t) - S(t)] \xrightarrow{w} Z(t)$$

که در آن $Z(t)$ یک فرایند گاوسی با گشتاورهای زیر است:

$$E[Z(t)] = ۰$$

$$\operatorname{Cov}[Z(t_۱), Z(t_۲)] = S(t_۱).S(t_۲) \times \int_۰^{t_۱ \wedge t_۲} \frac{dF_u(u)}{[۱ - H(u)]^۲}$$

$$= S(t_۱).S(t_۲) \times \int_۰^{t_۱ \wedge t_۲} \frac{dF_u(u)}{[۱ - F(u)][۱ - H(u)]}$$

که در آن:

$$F_u(t) = P(Y \le t, \delta = 1) = \int_0^t [1 - G(u)]dF(u)$$

$$1 - H(u) = [1 - F(u)][1 - G(u)]$$

اثبات شامل تابع نرخ شکست است، که در بخش بعدی به آن می‌پردازیم.

توجّه شود که $Z_n(t)$، به‌طور ضعیف ($\overset{w}{\rightarrow}$) به فرایند گاوسی $Z(t)$، میـــل می کنــد. یعنی: به ازاء هر t_1 تا t_k، $Z_n(t_1)$ تا $Z_n(t_k)$ دارای توزیـــع مجـانبی نرمـال چندمتغـیّره هستند و دنبالهٔ اندازه‌های احتمال Z_n، ملایم است، به‌گونه‌ای که $f(Z_n)$ در توزیع به $f(Z)$ -برای هر تابع پیوسته در اندازهٔ کوچکترین کران بالا- همگراست.

یک حالت خاص نتیجهٔ بالا، به شرح زیر است:

$$\hat{S}(t) \underset{\sim}{\overset{a}{}} N\left(S(t), \frac{S^2(t_2)}{n}\int_0^t \frac{dF_u(u)}{[1-H(u)]^2}\right)$$

بـرای واریـانس مجـانبی $\hat{S}(t)$ می‌تـوان یـک تقریـب بـه دسـت آورد. زیـرا $F_u(t) = P(Y \le t, \delta = 1)$ و $H(t) = P(Y \le t)$. موارد زیر را در نظـر می‌گـیریم (بـدون تکرار)

$$d\hat{F}_u(y_{(i)}) = \frac{\delta_{(i)}}{n}$$

$$1 - \hat{H}(y_{(i)}) = 1 - \frac{i}{n} = \frac{n-i}{n}$$

$$1 - \hat{H}(y_{(i)}^-) = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$$

اگر در واریانس مجانبی به‌جای $[1-H(u)]^2$ از $[1-H(u)][1-H(u^-)]$ استفاده کنیم و برآوردهای بالا را در آن قرار دهیم؛ داریم:

$$A\hat{V}ar[\hat{S}(t)] = \frac{\hat{S}^2(t)}{n}\sum_{y_{(i)} \le t}\frac{\delta_{(i)}/n}{[(n-i)/n][(n-i+1)/n]}$$

$$= \hat{S}^2(t)\sum_{y_{(i)} \le t}\frac{\delta_{(i)}}{(n-i)(n-i+1)}$$

که دقیقاً رابطهٔ گرین‌وود است. ($A\hat{V}ar$ به معنی واریانس مجانبی است).

REFERENCES
Billingsley, Convergence of Probability Measures (1968), for weak convergence.
Breslow and Crowley, Ann. Stat. (1974).

۳ برآوردگرهای تابع نرخ شکست

می‌دانیم تابع نرخ شکست به صورت: $\lambda(t)=\dfrac{f(t)}{1-F(t)}$، تعریف می‌شود. همچنین، مشکلات برآورد $\lambda(t)$ معادل برآورد تابع چگالی است. حالت ساده‌تر، برآورد تابع نرخ شکست تجمّعی: $\Lambda(t)=\int_0^t \lambda(u)du$، است.

توابع Λ و S با رابطهٔ $S(t)=e^{-\Lambda(t)}$ با هم ارتباط دارند. برای سادگی فرض می‌کنیم تکرار وجود ندارد. نلسون، $\Lambda(t)$ را به صورت زیر برآورد می‌کند:

$$\hat{\Lambda}(t)=\hat{\Lambda}_2(t)=\sum_{y_{(i)}\le t}\frac{\delta_{(i)}}{n-i+1}$$

برآورد پترسن Λ به شرح زیر است:

$$\hat{\Lambda}_1(t)=\sum_{y_{(i)}\le t}-\log\left(1-\frac{\delta_{(i)}}{n-i+1}\right)$$

دو برآوردگر بالا بسیار نزدیک هم‌اند، زیرا، برای مقادیر کوچک x، $\log(1-x)\cong -x$ است. برآوردگر پترسن متناظر برآوردگر PL تابع بقاء است. داریم:

$$\hat{S}_1(t)=e^{-\hat{\Lambda}_1(t)}=\prod_{y_{(i)}\le t}\left(1-\frac{\delta_{(i)}}{n-i+1}\right)=\hat{S}(t)$$

در حالی که، برآورد نلسون متناظر یک برآوردگر متفاوت دیگر تابع بقاست.

$$\hat{S}_2(t)=e^{-\hat{\Lambda}_2(t)}$$

فلمینگ و هارینگتن، $\hat{S}_2(t)$ را به عنوان یک برآوردگر دیگر تابع بقاء پیشنهاد می‌کنند و نشان می‌دهند که در بعضی از مواقع دارای میانگین مربع خطای کوچکتری است.

REFERENCES

Nelson, J. Qual. Tech. (1969).

______, Technometrics (1972).

Peterson, JASA (1977).

Fleming and Harrington, unpublished manuscript (1979).

نرمال مجانبی

از نتایج استاندارد توابع توزیع، داریم:

$$\sqrt{n}\,[\hat{F}_u(t)-F_u(t)] \xrightarrow{w} Z_{F_u}(t)$$

$$\sqrt{n}\,[\hat{H}(t)-H(t)] \xrightarrow{w} Z_H(t)$$

که: Z_{F_u} و Z_H، فرایندهای گاوسی‌اند. با بسط $\hat{\Lambda}(t)$ داریم:

$$\hat{\Lambda}(t)=\int_0^t \frac{d\hat{F}_u(u)}{1-\hat{H}(u^-)}$$

$$=\int_0^t \left[\frac{1}{1-H}+\frac{\hat{H}-H}{(1-H)^2}+\cdots\right]\left[dF_u+d(\hat{F}_u-F_u)\right]$$

$$=\int_0^t \frac{dF_u}{1-H}+\int_0^t \frac{\hat{H}-H}{(1-H)^2}dF_u+\int_0^t \frac{d(\hat{F}_u-F_u)}{1-H}+\cdots$$

$$=\Lambda(t)+\int_0^t \frac{\hat{H}-H}{(1-H)^2}dF_u+\frac{(\hat{F}_u-F_u)(t)}{1-H(t)}-\int_0^t \frac{\hat{F}_u-F_u}{(1-H)^2}dH+\cdots$$

تساوی آخر از انتگرال جزء به جزء به دست می‌آید. با تبدیل و ضرب در $\sqrt{n}$ داریم:

$$\sqrt{n}\,[\hat{\Lambda}(t)-\Lambda(t)]=\int_0^t \frac{\sqrt{n}\,(\hat{H}-H)}{(1-H)^2}dF_u$$

$$+\frac{\sqrt{n}\,(\hat{F}_u-F_u)(t)}{1-H(t)}-\int_0^t \frac{\sqrt{n}\,(\hat{F}_u-F_u)}{(1-H)^2}dH+\cdots$$

$$\xrightarrow{w} \int_0^t \frac{Z_H}{(1-H)^2}dF_u+\frac{Z_{F_u}(t)}{1-H(t)}-\int_0^t \frac{Z_{F_u}}{(1-H)^2}dH=Z_\Lambda(t)$$

حدّ $Z_{\Lambda}(t)$ برابر میانگین موزون فرایندهای گاوسی است و خود نیز یک فرایند گاوسی است. داریم:

$$E[Z_{\Lambda}(t)] = ۰$$

$$Cov[Z_{\Lambda}(t_۱), Z_{\Lambda}(t_۲)] = \int_۰^{t_۱ \wedge t_۲} \frac{dF_u}{(۱-H)^۲}$$

با استفاده از رابطه و تقریب $\hat{S}(t) = e^{-\hat{\Lambda}(t)}$، توزیع مجانبی $\hat{S}(t)$ به دست می‌آید.

$$e^{-\hat{\Lambda}(t)} = e^{-\Lambda(t)} - [\hat{\Lambda}(t) - \Lambda(t)]e^{-\Lambda(t)} + \cdots$$

$$\hat{S}(t) \cong S(t) - [\hat{\Lambda}(t) - \Lambda(t)]S(t) + \cdots$$

$$\sqrt{n}\,[\hat{S}(t) - S(t)] \cong -\sqrt{n}\,[\hat{\Lambda}(t) - \Lambda(t)]S(t) + \cdots$$

$$\overset{w}{\rightarrow} Z(t)$$

به‌گونه‌ای که $Z(t)$ یک فرایند گاوسی با $E[Z(t)] = ۰$ و کوواریانس زیر است:

$$Cov[Z(t_۱), Z(t_۲)] = S(t_۱).S(t_۲) \times \int_۰^{t_۱ \wedge t_۲} \frac{dF_u}{(۱-H)^۲}$$

REFERENCES
Breslow and Crowley, Ann. Stat. (1974).
Aalen, Scand. J. Stat. (1976).
______, Ann. Stat. (1978).

۴ برآوردگرهای تنومند

در مسائل برآورد، اغلب می‌توان پارامتر مورد علاقه را به صورت تابعی مانند: $\theta = T(F)$ نشان داد، که در آن F، تابع توزیع مربوط است.

اگر برش نباشد برآوردگر معمول به صورت: $\hat{\theta} = T(F_n)$ است، که F_n، تابع توزیع تجربی است. ولی اگر برش وجود داشته باشد، یک برآوردگر معقول به صورت: $\hat{\theta} = T(\hat{F})$ است، که در آن $\hat{F} = ۱ - \hat{S}$ و $\hat{S}$ برآوردگر PL است.

۱.۴ میانگین

$$\theta = T(F) = \int_{0}^{\infty} x\, dF(x) = \int_{0}^{\infty} [1 - F(x)]dx = \int_{0}^{\infty} S(t)dt$$

در صورت نبود برش، داریم:

$$\hat{\theta} = T(F_n) = \int_{0}^{\infty} x\, dF_n(x) = \bar{x} = \int_{0}^{\infty} [1 - F_n(x)]dx$$

در صورت وجود برش، داریم:

$$\hat{\theta} = T(\hat{F}) = \int_{0}^{\infty} x\, d\hat{F}(x) = \int_{0}^{\infty} \hat{S}(t)dt$$

$$AVar(\hat{\theta}) = \frac{1}{n}\int_{0}^{\infty} \frac{1}{[1 - H(s)]^{2}} \left(\int_{s}^{\infty} S(u)du \right)^{2} dF_u(s)$$

در حالت بدون تکرار، داریم:

$$A\hat{V}ar(\hat{\theta}) = \sum_{i=1}^{n} \left(\int_{y_{(i)}}^{\infty} \hat{S}(u)du \right)^{2} \frac{\delta_{(i)}}{(n-i)(n-i+1)}$$

اگر $y_{(n)}$ بریده شود، آن‌گاه با شرط $t \to \infty$، برآورد $\hat{S}(t)$ به صفر میل نمی‌کند. در نتیجه حاصل انتگرال بی‌نهایت خواهد شد. سه راه‌حل را بررسی می‌کنیم:

۱ تعریف دوبارهٔ آخرین مشاهده. با تعویض $\delta_{(n)} = 0$ به $\delta_{(n)} = 1$؛ از داده‌ها AML برای تشریح استفاده می‌کنیم:

$$\hat{\theta} = 9 \times 0{,}091 + 13 \times 0{,}091 + 18 \times 0{,}102 + 23 \times 0{,}102 + 31$$
$$\times 0{,}123 + 34 \times 0{,}123 + 48 \times 0{,}184 + (161 \times 0{,}184)$$
$$= 23{,}011 + (29{,}624) = 52{,}635$$

دنباله و بخصوص آخرین مشاهده، وزن زیادی دارند. علّت این است که برآوردگر PL وزنهای صعودی را به آخرین مشاهدات و به چولگی توزیع می‌دهد.

۲ میانگین محدود شده (مایر و سندر). به ازای مقدار ثابت S_0، میانگینی را در بازهٔ $[0, S_0)$ تعریف نموده و آن‌را به شرح زیر برآورد می‌کنیم:

$$\hat{\theta}=\int_{0}^{S_0}\hat{S}(t)dt$$

۳ حدّ متغیّر (سوسارلا و وان‌رایزیـن). $\theta=\int_{0}^{\infty}S(t)dt$ را با $\hat{\theta}=\int_{0}^{S_n}\hat{S}(t)dt$ برآورد می‌کنیم. در این برآورد، $\{S_n\}$ دنباله‌ای از اعدادی است که به‌طور یکنواخت به ∞ میـــل می‌کند.

متأسفانه، انتخاب مناسب S_n به F و G بستگی دارد و در این مورد روش کاربردی دقیق وجود ندارد.

REFERENCES
Kaplan and Meier, JASA (1958).
Meier, Perspwctives in Prob. and Stat. (1975).
Sander, Stanford Univ. Tech. Report No. 8. (1975).
Susarla and Van Ryzin, Ann. Stat. (1980).

۲.۴ برآوردگرهای - L

یک فرض اساسی برای استفاده از برآوردگرهای - L، این است که توزیع اوّلیّهٔ F نسبت به θ متقارن باشد. نوعاً، زمانهای بقاء توزیع متقارن ندارنـــد. زیـــرا مثبت‌انـــد. ولـــی می‌توان قبل از برآورد داده‌ها را با استفاده از یک تبدیل متقـــارن کـــرد. بـــرای مثـــال بـــا لگاریتم‌گیری. یک برآوردگر - L به صورت زیر است.

$$\hat{\theta}=\int_{-\infty}^{+\infty}xJ(\hat{F}(x))d\hat{F}(x)$$

در این رابطه J بر $[0,1]$ تعریف شده، نسبت به $\frac{1}{2}$ متقارن و در رابطهٔ زیر صدق می‌کند:

$$\int_{0}^{1}J(u)du$$

یک برآوردگر مهم L، با پیرایش کردن میانگین به شرح زیر به دست می‌آید:

$$J(u)=\frac{1}{1-2\alpha}I_{[\alpha,1-\alpha]}(u)$$

با داده‌های بریده شده، واریانس مجانبی یک برآوردگر - L به صورت زیر است:

$$\mathrm{AVar}(\hat{\theta})=\frac{۱}{n}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} S(t)\,J(S(t))\,S(u)$$

$$.J(S(u))\left\{\int_{-\infty}^{t\wedge u}\frac{dF_u(s)}{[۱-H(s)]^{۲}}\right\}dt\,du$$

REFERENCES
Sander, Stanford Univ. Tech. Report No. 8. (1975).
Reid Ann. Stat. (1981).

۳.۴ برآوردگر - M

در این‌جا نیز یک فرض اساسی متقارن بودن F است. در نتیجه ابتدا باید داده‌ها را تبدیل کرد. برآوردگر - M مربوط به $\hat{\theta}$ جواب معادلهٔ زیر است:

$$\int_{-\infty}^{+\infty}\psi(x-\hat{\theta})\,d\hat{F}(x)=۰$$

تابع $\psi(x-\theta)$، تعمیم $f'(x-\theta)/f(x-\theta)$ است. در نتیجه، برآوردگرهای - L، تعمیم برآوردگرهای حداّکثر درست‌نمایی‌اند. برآوردگر دو وزنی توکی متناظر با تابع زیر است:

$$\psi(x)=\begin{cases} x(۱-x^{۲})^{۲} & |x|\le ۱\\ ۰ & |x|>۱\end{cases}$$

در کاربردهای واقعی باید داده‌ها با یک مقیاس برآورد شده، مقیاس‌بندی شوند.
واریانس مجانبی یک برآوردگر - M در حالت برش، به شرح زیر است:

$$\mathrm{AVar}(\hat{\theta})=\frac{۱}{n}\int_{-\infty}^{+\infty}\frac{۱}{[۱-H(s)]^{۲}}.\left(\int_{s}^{+\infty}\frac{۱}{E\psi'}S(t)\,\psi'(t-\theta)dt\right)^{۲}dF_u(s)$$

در این‌جا: $.E\psi'=\int_{-\infty}^{+\infty}\psi'(t-\theta)\,dF(t)$

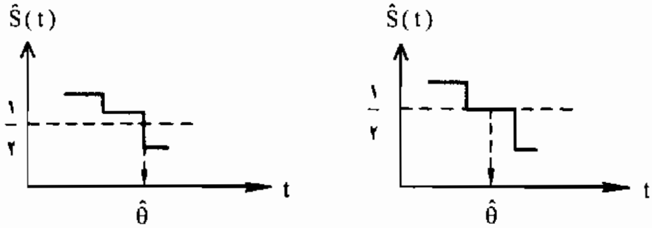
REFERENCE
Reid Ann. Stat. (1981).

تا زمان حال برآوردگرهای - L و M با داده‌های بریده شده به‌طور آزمایشگاهی بررسی

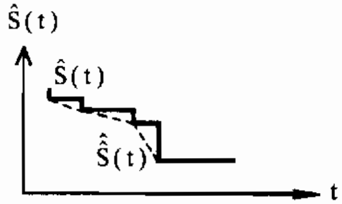
شده‌اند. عیوب و محاسن آنها بررسی نشده است. در حالی که برآوردگر میانه کاربرد عملی زیادی دارد.

۴.۴ میانه

با توجّه به $\theta = S^{-1}(\frac{1}{2})$، یک برآوردگر معقول برای θ، عبارت است از: $\hat{\theta} = \hat{S}^{-1}(\frac{1}{2})$ اگر $\hat{S}^{-1}(\frac{1}{2})$ دارای جواب منحصر به فرد نباشد، آن‌گاه $\hat{\theta}$ را نقطهٔ میانی بازهٔ شامل جوابها در نظر می‌گیریم:



شواهد تجربی نشان دهندهٔ این است، که این برآوردگر سرراست، خیلی بزرگ به نظر می‌رسد. برآوردگر PL با افزایش t، جهش بزرگتری می‌دهد و به خاطر مشاهدات بریدهٔ کنار گذاشته شده، شکاف بین مشاهدات بریده نشده با t افزایش پیدا نمی‌کند. بنابراین $\hat{\theta}$ بزرگتر می‌شود. برای رفع این مشکل $\hat{S}(t)$ را به صورت خطی هموار از $\hat{S}(t)$ تعریف می‌کنیم و $\hat{\hat{\theta}} = \hat{\hat{S}}^{-1}(\frac{1}{2})$



برای مثال، در داده‌ها AML، داریم:

$$\hat{S}(۲۳) = ۰/۶۱۴$$

$$\hat{S}(۳۱) = ۰/۴۹۱$$

$$\hat{\hat{\theta}} = ۳۱ - \frac{۸(۰/۰۰۹)}{(۰/۱۲۳)} = ۳۰/۴۱۵$$

باید واریانس $\hat{\theta}$ را محاسبه کرد. واریانس مجـانبی بـه شـرح: $AVar(\hat{\theta}) = \frac{AVar(\hat{S}(\theta))}{f^{۲}(\theta)}$ است. $AVar(\hat{S}(\theta))$ را می‌توان به کمک رابطهٔ گرین وود برآورد کرد. ولی، f یک تابع چگالی مجهول بوده و برآورد آن مشکل است.

REFERENCES

Sander, Stanford Univ. Tech. Report No. 5 (1975), discusses the asymptotic variance.

Földes, Rejto, and Winter, unpublished manuscript (1978), discuss density estimation using censored data.

Reid, Ann. Stat. (1981), discusses the asymptotic variance..

______ and Iyengar, unpublished notes (1978), consider estimates of the variance.

Efron, Stanford Univ. Tech. Report No. 53 (1980), uses the bootstrap to measure the var iability of $\hat{\theta}$.

۵ برآوردگرهای بیزی

فرض می‌کنیم تکرار وجود ندارد. با در نظر گرفتن $N_y(t) = \#(y_i > t)$، داریم:

$$\hat{S}(t) = \prod_{y_{(i)} \le t} \left[\frac{n-i}{n-i+۱}\right]^{\delta_{(i)}}$$

$$= \prod_{y_{(i)} \le t} \left[\frac{n-i+۱}{n-i}\right]^{-\delta_{(i)}} \cdot \frac{۱}{n}\left\{\frac{n}{n-۱} \cdot \frac{n-۱}{n-۲} \cdots \frac{N_y(t)+۱}{N_y(t)}\right\}\frac{N_y(t)}{۱}$$

$$= \frac{N_y(t)}{n} \prod_{y_{(i)} \le t} \left[\frac{n-i+۱}{n-i}\right]^{۱-\delta_{(i)}}$$

سوسالار و وان‌رایزین، نشان داده‌اند که برآوردگر بیز $S(t)$ دارای صورت زیر نیز هست:

$$\hat{S}_{\alpha}(t) = \frac{\alpha(t,\infty) + N_y(t)}{\alpha(۰,\infty) + n} \times \prod_{y_{(i)} \le t} \left[\frac{\alpha[y_{(i)},\infty) + (n-i+۱)}{\alpha[y_{(i)},\infty) + (n-i)}\right]^{۱-\delta_{(i)}}$$

برآوردگر $\hat{S}_{\alpha}(t)$، برآوردگر بیز با تابع زیان زیر است:

$$L(\hat{\delta}, S) = \int_{0}^{\infty} [\hat{\delta}(t) - S(t)]^{2} \, d\, w(t)$$

که در این رابطه، w هر تابع نامنفی صعودی و یا فرایند پیشین دیرکله $\mathcal{P}_{\alpha}$ با پـــارامتر α بر خانواده $\{P\}$ در تمام توزیعهای ممکن است. پارامتر α یــک انـدازهٔ متنــاهی نـامنفی بر $(0, \infty)$ است.

اندازهٔ احتمال تصادفی P را دارای فرایند پیشین دیرکله با پارامتر α گوییم، هرگاه برای هر افراز اندازه‌پذیر β_1 تا β_k از $(0, \infty)$ داشته باشیم:

$$(P(\beta_1), \ldots, P(\beta_k)) \sim \text{Dirichlet}(\alpha(\beta_1), \ldots, \alpha(\beta_k))$$

توجّه شود که توزیع دیرکله $(\alpha_1, \ldots, \alpha_k)$ دارای چگالی زیر است:

$$f(x_1, \ldots, x_k) \propto x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_k^{\alpha_k - 1} . I(x_i \geq 0, x_1 + \cdots + x_k = 1)$$

توجّه نمایید که در ازاء $k = 2$، توزیع دریکله دقیقاً همان توزیع بتاست.

فرض می‌کنیم مشاهدهٔ X دارای توزیع $\mathcal{P}_{\theta}$ است، که در آن θ طبــق یــک توزیــع پیشین انتخاب می‌شود. در حالت ناپارامتری، متغیّر T با توزیع P -که P طبق توزیــع $\mathcal{P}_{\alpha}$ انتخاب می‌شود- مشاهده می‌شود. به عبارت دیگر، زمان بقاء T توسّط $\mathcal{P}_{\alpha}$ به دست می‌آید و سپس، P را تولید نموده و P نیز متغیّر T را تولید می‌کند. می‌توان تساوی زیر را اثبات کرد:

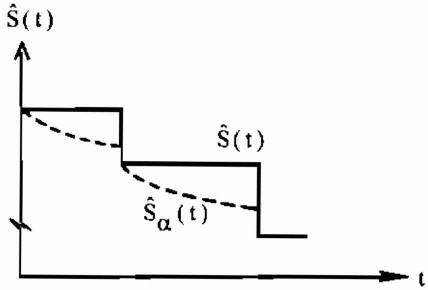
$$P\{T \in A\} = \frac{\alpha(A)}{\alpha(0, \infty)} \qquad (11-1)$$

معادلهٔ (۱-۱۱)، تفسیری از پارامتر α را ارائه می‌کند. نسبت $\frac{\alpha(A)}{\alpha(0, \infty)}$، حدس اوّلیّــه بــر احتمال مجموعهٔ A است. برای مثال اگر فرض کنیم T دارای توزیع نمایی با میانگین $\frac{1}{\lambda_0}$ باشد، آن‌گاه:

$$\frac{\alpha(t, \infty)}{\alpha(0, \infty)} = e^{-\lambda_0 t}$$

همچنیــن، جــرم کــل $\alpha(0, \infty)$، شــدّت بـــاور پیشـــین را نشـــان می‌دهـــد. بـــرای مثال، $\alpha(0, \infty) = 10$، بیان می‌کند که باور پیشین ما ارزش ده مشاهده را دارد.

با توجّه به رابطهٔ $\frac{\alpha(t,\infty)}{\alpha(\circ,\infty)}=e^{-\lambda_\circ t}$، مقادیر $\hat{S}_\alpha(t)$ و $\hat{S}(t)$، مطابق شکل زیر قابل مقایسه‌اند:



سوسارلا و وان‌رایزین نشان دادند که در بسیاری از حالات، $\hat{S}_\alpha$ دارای میانگین مربع خطای کوچکتری از $\hat{S}$ -حتی اگر حدس پیشین درست نباشد- است. برآورد بیزی در حالت تکراز مطابق زیر است:

$$\hat{S}_\alpha(t)=\frac{\alpha(t,\infty)+N_y(t)}{\alpha(\circ,\infty)+n}\times\prod_{y'_{(j)}\le t}\left[\frac{\alpha[y'_{(j)},\infty)+N_y(y'_{(j)}-)}{\alpha[y'_{(j)},\infty)+N_y(y'_{(j)})}\right]^{1-\delta'_{(j)}}$$

REFERENCES

Ferguson ,Ann. Stat. (1973), discusses the Dirichlet process prior.
Susarla and Van Ryzin, JASA (1976), derive the Bayes estimate in the censored case.
______ and ______, Ann. Stat. (1978b), study the asymptotic behavior of Bayes estimates.
Ferguson and Phadia, Ann. Stat. (1979), examine more general prior distributions.
Rai, Susarla, and Van Ryzin, Comm. Stat. B (1980), look at mean square errors.

برآوردگرهای تجربی بیزی. به‌جای کاربرد یک حدس پیشین برای α، می‌توان نمونه را برای برآورد α به کاربرد.

REFERENCES

Susarla and Van Ryzin, Ann. Stat. (1978a).
Phadia, Ann. Stat. (1980).

فصل چهارم

روشهای ناپارامتری (دو نمونه)

برای نمونهٔ اوّل: فرض کنید T_1 تا T_m، متغیّرهای iid با توزیع F_1 و C_1 تا C_m، متغیّرهای iid با توزیع G_1 باشند. C_i زمان برش مربوط به T_i است. می‌توان داده‌های (X_1, δ_1) تا (X_m, δ_m) را به شرح زیر مشاهده نمود:

$$X_i = T_i \wedge C_i \qquad \text{و} \qquad \delta_i = I(T_i \leq C_i)$$

برای نمونهٔ دوم: فرض کنید U_1 تا U_n، متغیّرهای iid با توزیع F_2 و D_1 تا D_n، متغیّرهای iid با توزیع G_2 باشند. D_j زمان برش مربوط به U_j است. می‌توان مشاهدات (Y_1, ε_1) تا (Y_n, ε_n) را به شرح زیر مشاهده نمود:

$$Y_j = U_j \wedge D_j \qquad \text{و} \qquad \varepsilon_j = I(U_j \leq D_j)$$

معمولاً، در مسألهٔ دو نمونه‌ای آزمون فرض $H_0: F_1 = F_2$، مورد نظر است.

مثال. آزمایش بالینی فرضی: این آزمایش توسّط بایرون وی‌ام و بران جی‌آر، مطابق نمودارهای (۴. الف) و (ب)، بنا شده است. فرض نمایید مشاهدات X مربوط به تیمار A و مشاهدات Y مربوط به تیمار B باشند.

Rx A: ۳, ۵, ۷, ۹⁺, ۱۸

Rx B: ۱۲, ۱۹, ۲۰, ۲۰⁺, ۳۳⁺

۱ آزمون گهان

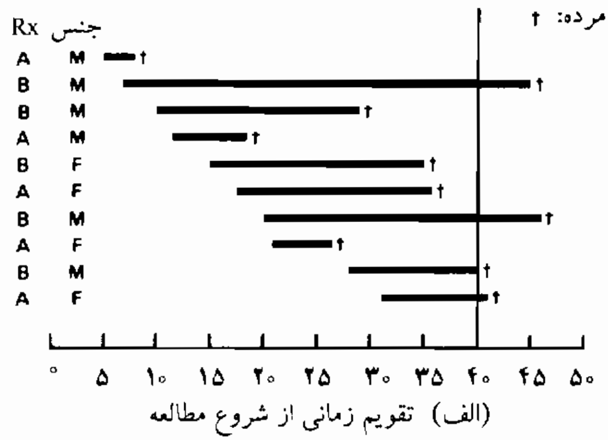
این آزمون تعمیمی از آزمون ویلکاکسن است. **فرض کنید مشاهدات دو نمونه به**

صورت زیر باشد:

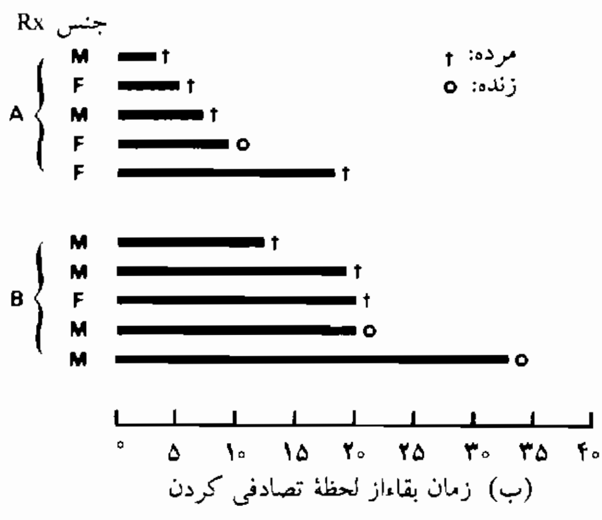
$$X_1, \ldots, X_m \;;\; Y_1, \ldots, Y_n$$

این ترکیب را مرتّب می‌کنیم، داریم:

$$Z_{(1)}, Z_{(2)}, \ldots, Z_{(m+n)}$$



نمودار ۴. (الف) بررسی زمان بقاء ده بیمار سرطانی که به تصادف تحت تیمار (A) و (B) قرار گرفته‌اند.



نمودار ۴. (ب) زمان بقاء از لحظهٔ تصادفی کردن ده بیمار سرطانی جدا از زمان ختم مطالعه در ۴۰ = t.

R_{1i} را رتبهٔ X_i و $R_1 = \sum_{i=1}^{m} R_{1i}$ در نظر می‌گیریم. فرض $H_\circ$ رد می‌شود، هرگاه R_1 خیلی کوچک یا خیلی بزرگ باشد. به کمک جدولهای مربوط، برای نمونه‌های کوچک و تقریب نرمال برای نمونه‌های بزرگ و به شرح زیر، آزمون را انجام می‌دهیم. در رابطهٔ زیر $E_\circ(R_1)$ و $Var_\circ(R_1)$، گشتاورهای تحت فرض صفراند.

$$\frac{R_1 - E_\circ(R_1)}{\sqrt{Var_\circ(R_1)}} = \frac{R_1 - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \overset{a}{\sim} N(0,1)$$

صورت من-ویتنی آزمون ویلکاکسن مفید خواهد بود. با توجّه به تعریف زیر می‌توان R_1 را مطابق بسط زیر نوشت:

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1 & X_i > Y_j \\ 0 & X_i = Y_j \\ -1 & X_i < Y_j \end{cases}$$

می‌توان نشان داد:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij}$$

در نتیجه:

$$R_1 = \frac{m(m+n+1)}{2} + \frac{1}{2}U$$

برای اثبات توجّه شود، که اگر مشاهدات به صورت:

$$X_{(1)} < \cdots < X_{(m)} < Y_{(1)} < \cdots < Y_{(n)}$$

درآیند، آن‌گاه $R_1 = \frac{m(m+1)}{2}$ می‌شود. برای هر تعویض زوج X و Y، R_1 به اندازهٔ یک واحد صعود می‌کند و تعداد این تعویض‌ها برابر $\sum_i \sum_j \frac{1}{2}(U_{ij}+1)$ است؛ بنابراین داریم:

$$R_1 = \frac{m(m+1)}{2} + \sum_i \sum_j \frac{1}{2}(U_{ij}+1) = \frac{m(m+1)}{2} + \frac{mn}{2} + \frac{1}{2}U$$

$$= \frac{m(m+n+1)}{2} + \frac{1}{2}U$$

آزمون من-ویتنی فرض H_0 را در صورتی رد می‌کند، که U یا $|U|$ خیلی بزرگ باشد. در نمونه‌های کوچک از جدولها و در نمونه‌های بزرگ از تقریب نرمال استفاده می‌کنیم.

$$\frac{U-E_0(U)}{\sqrt{Var_0(U)}}=\frac{U}{\sqrt{\frac{mn(m+n+1)}{3}}} \overset{a}{\sim} N(0,1)$$

برای داده‌های بریده شده، U_{ij} را به صورت زیر تعریف می‌کنیم:

$$U_{ij}=\begin{cases}+1 & t_i>u_j \text{ یا } (X_i>Y_j,\varepsilon_j=1) \text{ یا } (X_i=Y_j,\delta_i=0,\varepsilon_j=1)\\ 0 & \text{در سایر جاها}\\ -1 & t_i<u_j \text{ یا } (X_i<Y_j,\delta_i=1) \text{ یا } (X_i=Y_j,\delta_i=1,\varepsilon_j=0)\end{cases}$$

$$U=\sum_{i=1}^{m}\sum_{j=1}^{n}U_{ij}$$

فرض H_0 در صورت بزرگ بودن U یا $|U|$ رد می‌شود. آمارهٔ U به طور مجانبی دارای توزیع نرمال است. ولی، برای محاسبهٔ معیارهای آزمون تنها گشتاورهای U کفایت می‌کند.

۱.۱ میانگین و واریانس U

در حالت بریده نشده، میانگین و واریانس را می‌توان با استفاده از قضیهٔ جایگشت محاسبه کرد. تحت H_0، نمونه‌گیری از m مهره بدون جایگذاری را از جعبه‌ای شامل $m+n$ مهره به شماره‌های Z_1، ...، Z_{m+n} در نظر بگیرید، که زیرنویسهای m مهره نمونه‌گیری شده به عنوان مقادیر X_1 تا X_m و زیرنویسهای n مهره باقی‌مانده به عنوان Y_1 تا Y_n در نظر گرفته می‌شوند. فرض کنید $E_{0,p}(U)$ و $Var_{0,p}(U)$، گشتاورهای تحت الگوی جایگشت باشند. در نتیجه:

$$E_{0,p}(U)=0=E_0(U)$$

$$Var_{0,p}(U)=\frac{mn(m+n+1)}{3}=Var_0(U)$$

در حالت برش، گهان قضیهٔ جایگشت را تحت فرض قویتر زیر به کار می‌برد:

$$H_{\circ}^{*}: F_1 = F_2 \quad , \quad G_1 = G_2$$

فرض کنید نمونهٔ مرکب به صورت: (Z_1, ξ_1) تا (Z_{n+m}, ξ_{n+m}) باشد. نمونه‌ای شامل m مهره بدون جایگذاری را از جعبه‌ای شامل n + m مهره به صورت (Z_1, ξ_1) تا (Z_{n+m}, ξ_{n+m}) در نظر می‌گیریم. زیرنویسهای نمونهٔ m تایی را به صورت (X_1, δ_1) تا (X_m, δ_m) و زیرنویسهای n مهرهٔ باقی‌مانده را با (Y_1, ε_1) تا (Y_n, ε_n) در نظر می‌گیریم. در این صورت:

$$E_{\circ,P}^{*}(U) = \circ$$

$$Var_{\circ,P}^{*}(U) = \text{(به مقاله گهان بخش ۳.۴ صفحهٔ ۲۵۶ مراجعه شود)}$$

مقدار واریانس به صورت عبارتی پیچیده در می‌آید، که از آن صرفنظر می‌شود. در عوض از روش مانتل مقدار $Var_{\circ,P}^{*}(U)$ به صورت ساه‌تری به دست می‌آید.

۲.۱ روش محاسباتی مانتل برای $Var_{\circ,P}^{*}(U)$

بنابه تعریف:

$$U_{k\ell} = U((Z_k, \xi_k), (Z_\ell, \xi_\ell))$$

$$U_{ij} = \begin{cases} +1 & (Z_k > Z_\ell, \xi_\ell = 1) \text{ یا } (Z_k = Z_\ell, \xi_k = \circ, \xi_\ell = 1) \\ \circ & \text{در سایر جاها} \\ -1 & (Z_k < Z_\ell, \xi_k = 1) \text{ یا } (Z_k = Z_\ell, \xi_k = 1, \xi_\ell = \circ) \end{cases}$$

$$U_k^{*} = \sum_{\substack{\ell=1 \\ \neq k}}^{m+n} U_{k\ell} \quad , \quad U = \sum_{k=1}^{m+n} U_k^{*} I(k \in I_1)$$

که در آن I_1، مجموعهٔ اعداد صحیح در نمونه یک است. توجه شود که U معادل آماره گهان است، زیرا $U_{k_1 k_2} = -U_{k_2 k_1}$. در نتیجه اگر k_1 و k_2 عضو I_1 باشند، یکدیگر را حذف می‌کنند.

برای محاسبهٔ توزیع جایگشت U، فرض کنید U_1^{*} تا U_{n+m}^{*}، تحت $H_{\circ}$ معلوم باشند. یک نمونهٔ m تایی بدون جایگذاری از این U_k^{*}ها انتخاب می‌کنیم. سپس U، مجموع این m مقدار را تشکیل می‌دهیم. با استفاده از نتایج نمونه‌گیری از جمعیتهای متناهی، داریم:

$$\mathrm{Var}^{*}_{\circ,P}(U) = m\left(\frac{۱}{m+n-۱}\sum_{i=۱}^{m+n}(U_i^*)^{۲}\right)\left(۱-\frac{m}{m+n}\right)$$

$$= \frac{mn}{(m+n)(m+n-۱)}\sum_{i=۱}^{m+n}(U_i^*)^{۲}$$

۳.۱ مثال. به کمک داده‌های مثال قبل (۴. الف و ب)، داریم:

Z	Rx	# < Z	# > Z	U^*
۳	A	۰	۹	−۹
۵	A	۱	۸	−۷
۷	A	۲	۷	−۵
۹^+	A	۳	۰	۳
۱۲	B	۳	۵	−۲
۱۸	A	۴	۴	۰
۱۹	B	۵	۳	+۲
۲۰	B	۶	۲	+۴
۲۰^+	B	۷	۰	+۷
۳۳^+	B	۷	۰	+۷

$$U = -۹-۷-۵+۳+۰ = -۱۸$$

$$E^{*}_{\circ,P}(U) = ۰ \qquad \text{و} \qquad \mathrm{Var}^{*}_{\circ,P}(U) = \frac{(۵)(۵)(۲۸۶)}{(۱۰)(۹)} = ۷۹/۴۴$$

تحت فرض $H_\circ$، داریم:

$$\frac{U}{\sqrt{\mathrm{Var}^{*}_{\circ,P}(U)}} = \frac{-۱۸}{۸/۹۱} = -۲/۰۲ \overset{a}{\sim} N(۰,۱)$$

بنابراین $P = 0{,}022$، مقدار P برای آزمون یک طرفه است.

REFERENCES
Gehan, Biometrika (1965).
Mantal, Biometrics (1967).

۴.۱ واریانس تحت H_0

نتایج مباحث قبل دربارهٔ واریانس با فرض: $H_0^*: F_1 = F_2\ ,\ G_1 = G_2$، به دست آمده‌اند. حال واریانس جایگشت تحت فرض $H_0: F_1 = F_2$ با طرح برش ثابت چیست؟ فرض کنید V_1، ...، V_{m+n}، نمونهٔ مرکّب از T_1، ...، T_m و U_1، ... U_n، باشد. تحت فرض H_0، از آن V_k را بدون جایگذاری نمونه‌گیری کرده و در عبارت زیر قرار می‌دهیم:

$$(\_, C_1)\ \text{و} \cdots \text{و}\ (-, C_m)\ ;\ \ (-, D_1)\ \text{و} \cdots \text{و}\ (-, D_n)$$

به کمک این مشاهدات، صورت زیر مورد توجّه است:

$$(X_1, \delta_1)\ \text{و} \cdots \text{و}\ (X_m, \delta_m)\ ;\ \ (Y_1, \varepsilon_1)\ \text{و} \cdots \text{و}\ (Y_n, \varepsilon_n)$$

$$(X_1, \delta_1), \dots, (X_m, \delta_m);\ \ (Y_1, \varepsilon_1), \dots, (Y_n, \varepsilon_n)$$

با توجّه به این‌که تمام T_i، C_i، U_j و D_jها قابل مشاهده نیستند، متأسفانه، نمی‌توان تمام (X_i, δ_i) و (Y_j, ε_j)ها را ساخت.

هاید $E_0(\mathrm{Var}^*_{0,P}(U))$ را با $\mathrm{Var}_0(U)$، مقایسه می‌کند. داریم:

$$\mathrm{Var}_0(U) = E_0(U^2) = E\left\{\left(\sum_{i=1}^{m}\sum_{j=1}^{n} U_{ij}\right)^2\right\}$$

$$= mn\,E_0(U_{ij}^2) + mn(n-1)\,E_0(U_{ij}U_{ij'}) + m(m-1)n\,E_0(U_{ij}U_{i'j})$$

$$+\, m(m-1)n(n-1)E_0(U_{ij}U_{i'j'})$$

$$E_0(\mathrm{Var}^*_{0,P}(U)) = \text{مجموع جملات مشابه}$$

اگر با شرط $\frac{m}{m+n} \to \lambda$ و $0 < \lambda < 1$، n و m به بی‌نهایت میل کنند، آن‌گاه داریم:

$$R^{۲} = \underset{m,n\to\infty}{p-\lim} \frac{Var^{*}_{۰,p}(U)}{Var_{۰}(U)} = \lim_{m,n\to\infty} \frac{E_{۰}(Var^{*}_{۰,p}(U))}{Var_{۰}(U)}$$

$$= ۳\lambda(۱-\lambda) + \Big\{\lambda^{۳} P\{C_۱ \wedge C_۲ \wedge C_۳ > T_۱ \wedge T_۲ \wedge T_۳\} +$$

$$+(۱-\lambda)^{۳} P\{D_۱ \wedge D_۲ \wedge D_۳ > T_۱ \wedge T_۲ \wedge T_۳\}\Big\}$$

$$\times \Big\{\lambda \; P\{C_۱ \wedge C_۲ \wedge D_۱ > T_۱ \wedge T_۲ \wedge T_۳\} +$$

$$+(۱-\lambda) \; P\{C_۱ \wedge D_۱ \wedge D_۲ > T_۱ \wedge T_۲ \wedge T_۳\}\Big\}^{-۱} \qquad (۱۱)$$

با توجّه به رابطهٔ (۱۱) داریم: $R^{۲} > ۳\lambda(۱-\lambda)$. اگر $\lambda = \frac{۱}{۲}$، آن‌گاه $R^{۲} > \frac{۳}{۴}$ یا $R > ۰/۸۷$. پس، اگر حجم نمونه‌ها برابر باشند، $SD^{*}_{۰,p}(U)$، نمی‌تواند خیلی کوچکتر از $SD_{۰}(U)$ باشد. در این‌جا نوع برش مهم نیست.

فرض کنید توزیعهای بریده شده به صورتهای مختلف لهمن باشند. یعنی:

$$(۱-G_۱)^{r_۱} = ۱-F \qquad و \qquad (۱-G_۲)^{r_۲} = ۱-F$$

که در آن $r_۱$ و $r_۲$ به مقادیر زیر و به صورت $p_۱ = \frac{۱}{r_۱+۱}$ و $p_۲ = \frac{۱}{r_۲+۱}$ وابسته‌اند.

$p_۱ = P(C_۱ < T_۱) = P($مشاهده در جمعیت ۱ بریده شده است$)$

$p_۲ = P(D_۱ < U_۱) = P($مشاهده در جمعیت ۲ بریده شده است$)$

هاید در جدول ۲، مقادیر R را برای $\lambda = ۰/۵$، تحت توزیعهای لهمن با تغییر سطوح بریدهٔ $p_۱$ و $p_۲$ محاسبه کرده است. جدول به خاطر تعیین حالتهای $|R-۱| < ۰/۰۵$، افراز شده است. جدول ۳، با فرض $\lambda = ۰/۲$ مساوی جدول ۲ است. از جدول ۲ دیده می‌شود که آزمون گهان (طرحهای برش را مساوی فرض می‌کند) از یک انحراف معیار تقریباً صحیح -حتی وقتی که احتمالهای برش تفاوت زیادی دارند- استفاده می‌کند. علاوه بر این، هنگامی که یک نمونه چهار برابر دیگری باشد (جدول ۳)، آزمون گهان از یک انحراف معیار تقریباً صحیح برای دامنهٔ وسیعی از احتمالهای بریده شده، استفاده می‌کند.

REFERENCES
Gilbert, Univ. Chicago thesis (1962), was the first to calculate $Var_0(U)$.
Hyde Stanford Univ. Tech. Report No. 30 (1977).

۲ آزمون مانتل-هانزل

۱.۲ جدول ۲×۲. فرض کنید دو جمعیت داریم، به گونه‌ای که یک فرد می‌تواند در هر دو جمعیت یکی از دو مشخصه‌ای را دارا باشد. برای مثال، جمعیت ۱، شامل بیماران سرطانی تحت تیمار معیّن و جمعیت ۲، شامل بیماران تحت تیمار دیگر است. ممکن است، بیماران هر یک از دو جمعیت در عرض یک سال بمیرند یا زنده بمانند. می‌توان داده‌ها را در یک جدول ۲×۲ و به شرح زیر خلاصه نمود. همچنین p_1 و p_2 را به شرح زیر تعریف می‌کنیم:

	مرده	زنده	
جمعیت ۱	a	b	n_1
جمعیت ۲	c	d	n_2
	m_1	m_2	n

$$p_1 = P(\text{مردن} \mid \text{در جمعیت ۱})$$

$$p_2 = P(\text{مردن} \mid \text{در جمعیت ۲})$$

برای آزمون $H_0: p_1 = p_2$، از آمارهٔ زیر استفاده می‌شود، که در آن $\hat{p}_1 = \frac{a}{n_1}$، $\hat{p}_2 = \frac{c}{n_2}$ و $\hat{p} = \frac{m_1}{n}$ است:

$$\chi^2 = \left[\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}\right]^2 = \frac{n(ad-bc)^2}{n_1 n_2 m_1 m_2}$$

یا با توجّه به تصحیح پیوستگی داریم:

$$\chi_c^2 = \frac{n(|ad-bc| - n/2)^2}{n_1 n_2 m_1 m_2}$$

جدول ۲. مقادیر R برای $\lambda = ۰٫۵$ و توزیعهای بریده شده به صورتهای مختلف لهمن هستند.

	p_1									
p_2	۰٫۰۰	۰٫۱۰	۰٫۲۰	۰٫۳۰	۰٫۴۰	۰٫۵۰	۰٫۶۰	۰٫۷۰	۰٫۸۰	۰٫۹۰
۰٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۱	۱٫۰۲	۱٫۰۴	۱٫۰۹	۱٫۲۰
۰٫۱۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۲	۱٫۰۴	۱٫۰۷	۱٫۱۸
۰٫۲۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۱	۱٫۰۳	۱٫۰۶	۱٫۱۶
۰٫۳۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۲	۱٫۰۵	۱٫۱۳
۰٫۴۰	۱٫۰۱	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۴	۱٫۱۱
۰٫۵۰	۱٫۰۱	۱٫۰۱	۱٫۰۱	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۲	۱٫۰۹
۰٫۶۰	۱٫۰۲	۱٫۰۲	۱٫۰۱	۱٫۰۱	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۶
۰٫۷۰	۱٫۰۴	۱٫۰۴	۱٫۰۳	۱٫۰۲	۱٫۰۱	۱٫۰۱	۱٫۰۰	۱٫۰۰	۱٫۰۰	۱٫۰۴
۰٫۸۰	۱٫۰۹	۱٫۰۷	۱٫۰۶	۱٫۰۵	۱٫۰۴	۱٫۰۲	۱٫۰۱	۱٫۰۰	۱٫۰۰	۱٫۰۱
۰٫۹۰	۱٫۲۰	۱٫۱۸	۱٫۱۶	۱٫۱۳	۱٫۱۱	۱٫۰۹	۱٫۰۶	۱٫۰۴	۱٫۰۱	۱٫۰۰

جدول ۳. مقادیر R برای $\lambda = ۰٫۲$ و توزیعهای بریده شده به صورتهای مختلف لهمن هستند.

p_2	p_1: ۰٫۰۰	۰٫۱۰	۰٫۲۰	۰٫۳۰	۰٫۴۰	۰٫۵۰	۰٫۶۰	۰٫۷۰	۰٫۸۰	۰٫۹۰
۰٫۰۰	۱٫۰۰	۱٫۰۱	۱٫۰۲	۱٫۰۴	۱٫۰۶	۱٫۰۹	۱٫۱۴	۱٫۲۱	۱٫۳۳	۱٫۶۵
۰٫۱۰	۰٫۹۹	۱٫۰۰	۱٫۰۱	۱٫۰۳	۱٫۰۵	۱٫۰۸	۱٫۱۲	۱٫۱۸	۱٫۲۹	۱٫۵۹
۰٫۲۰	۰٫۹۸	۰٫۹۹	۱٫۰۰	۱٫۰۱	۱٫۰۳	۱٫۰۶	۱٫۰۹	۱٫۱۵	۱٫۲۶	۱٫۵۳
۰٫۳۰	۰٫۹۷	۰٫۹۸	۰٫۹۹	۱٫۰۰	۱٫۰۲	۱٫۰۴	۱٫۰۷	۱٫۱۲	۱٫۲۲	۱٫۴۷
۰٫۴۰	۰٫۹۶	۰٫۹۷	۰٫۹۷	۰٫۹۹	۱٫۰۰	۱٫۰۲	۱٫۰۵	۱٫۰۹	۱٫۱۸	۱٫۴۰
۰٫۵۰	۰٫۹۵	۰٫۹۵	۰٫۹۶	۰٫۹۷	۰٫۹۸	۱٫۰۰	۱٫۰۲	۱٫۰۶	۱٫۱۴	۱٫۳۳
۰٫۶۰	۰٫۹۴	۰٫۹۴	۰٫۹۵	۰٫۹۶	۰٫۹۷	۰٫۹۸	۱٫۰۰	۱٫۰۳	۱٫۰۹	۱٫۲۶
۰٫۷۰	۰٫۹۳	۰٫۹۳	۰٫۹۳	۰٫۹۴	۰٫۹۵	۰٫۹۶	۰٫۹۷	۱٫۰۰	۱٫۰۵	۱٫۱۸
۰٫۸۰	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۳	۰٫۹۳	۰٫۹۴	۰٫۹۵	۰٫۹۷	۱٫۰۰	۱٫۰۹
۰٫۹۰	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۲	۰٫۹۳	۰٫۹۵	۱٫۰۰

متغیّر χ^{2}، تقریباً دارای توزیع χ_{1}^{2} است. این یک تقریب تحت فرض H_{0} برای توزیع شرطی گسستهٔ دقیق است. با معلوم بودن n_{1}، n_{2}، m_{1} و m_{2}، متغیّر تصادفی A، عدد مربوط به خانهٔ (۱,۱) جدول ۲ × ۲، دارای توزیع فوق هندسی به شرح زیر است:

$$P(A=a)=\frac{\binom{n_{1}}{a}\binom{n_{2}}{m_{1}-a}}{\binom{n}{m_{1}}}$$

دو گشتاور اوّلیهٔ توزیع فوق هندسی، به شرح زیر است:

$$E_{0}(A)=\frac{n_{1}m_{1}}{n} \qquad \text{و} \qquad Var_{0}(A)=\frac{n_{1}n_{2}m_{1}m_{2}}{n^{2}(n-1)}$$

در نتیجه داریم:

$$ad-bc=n(a-E_{0}(A))$$

$$n_{1}n_{2}m_{1}m_{2}=n^{2}(n-1)Var_{0}(A)$$

$$\chi^{2}=\frac{n(ad-bc)^{2}}{n_{1}n_{2}m_{1}m_{2}}=\frac{n}{n-1}\left[\frac{a-E_{0}(A)}{\sqrt{Var_{0}(A)}}\right]^{2}$$

۲.۲ دنباله‌ای ازجدولهای ۲ × ۲

فرض کنید، دنباله‌ای از جدولهای ۲ × ۲ در اختیار داریم. برای مثال، در k بیمارستان، بیماران تیمار ۱ یا تیمار ۲ را دریافت می‌کنند و رفتار آنها ثبت می‌شود. امکان دارد بیمارستانها دارای تفاوت باشند، لذا، نمی‌توانیم k جدول را در یک جدول ۲ × ۲ خلاصه کنیم. در نتیجه فرض زیر را آزمون می‌کنیم:

$$H_{0}: p_{11}=p_{12}\ , \ldots ,\ p_{k1}=p_{k2}$$

که p_{i1} و p_{i2} به شرح زیراند:

$p_{i1}=P($ مردن | بیمارستان i، تیمار ۱ $)$

$p_{i2}=P($ مردن | بیمارستان i، تیمار ۲ $)$

جداول به شرح زیراند:

	مرده	زنده	
تیمار ۱	a_k		n_{k1}
تیمار ۲			n_{k2}
	m_{k1}	m_{k2}	n_k

بیمارستان kام

...

	مرده	زنده	
تیمار ۱	a_1		n_{11}
تیمار ۲			n_{12}
	m_{11}	m_{12}	n_1

بیمارستان یکم

با استفاده از آمارهٔ مانتل-هانزل، داریم:

$$MH = \frac{\sum_{i=1}^{k} (a_i - E_0(A_i))}{\sqrt{\sum_{i=1}^{k} (Var_0(A_i))}}$$

در صورت استفاده از تصحیح پیوستگی، داریم:

$$MH_c = \frac{\left|\sum_{i=1}^{k} (a_i - E_0(A_i))\right| - \frac{1}{2}}{\sqrt{\sum_{i=1}^{k} (Var_0(A_i))}}$$

اگر جدولها مستقل باشند یا این‌که k مقدار ثابتی بوده و $n_i \to \infty$، یا هنگامی که $k \to \infty$ میل کند، آن‌گاه: $MH \overset{a}{\sim} N(0,1)$ است. همچنین، جدولها هم‌توزیع‌اند.

در تحلیل بقاء آمارهٔ MH به صورت زیر به کار برده می‌شود. فرض کنید: $(Z_{(1)}, \xi_{(1)})$ تا $(Z_{(m+n)}, \xi_{(m+n)})$، نمونهٔ مرتّب شده ترکیبی باشد. برای هر زمان بریده شده یک جدول 2×2 بسازید. آمارهٔ MH را برای این دنباله از جدولها برای آزمون $H_0: F_1 = F_2$، بسازید.

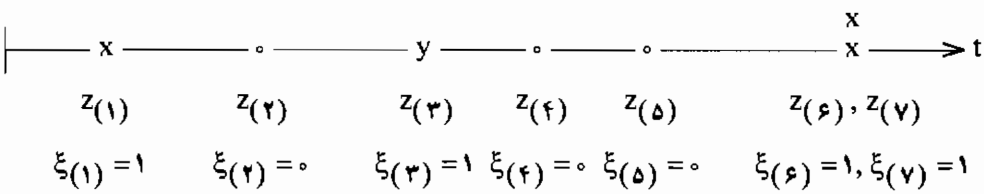
این جدولها مستقل نیستند. زیرا برای مثال، $\Re(z_{(1)})$ و $\Re(z_{(3)})$ تقریباً برابرند. ولی هنوز نرمال بودن مجانبی برقرار است. تغییر طرحهای برشی بر آمارهٔ MH بی‌تأثیر است. این آزمون را آزمون لگاریتم رتبه نیز گویند (به فصل ششم، بخش (۱.۲)، مراجعه کنید).

۳.۲ مثال

محاسبهٔ آمارهٔ MH (مربوط به شکلهای ۴. الف و ۴. ب)، در جدول شماره۴ داده شده است. ستون z شامل مشاهدات مرتّب بریده نشده است. چهار ستون بعدی n، m_1، n_1 و a، جدولهای (۲×۲) را می‌سازند. ستون بعدی مقدار $E_\circ(A)=\frac{n_1 m_1}{n}$ را محاسبه می‌کند. حاصل‌ضرب دو ستون آخر $\frac{m_1(n-m_1)}{(n-1)}$ و $(1-\frac{n_1}{n})\frac{n_1}{n}$، مقدار $Var_\circ(A)$ را می‌دهند. بهتر است، مقدار $Var_\circ(A)$ را به این طریق محاسبه کنیم. زیرا، مقدار $\frac{m_1(n-m_1)}{(n-1)}$ معمولاً برابر یک و $(1-\frac{n_1}{n})\frac{n_1}{n}$ برابر حاصل‌ضرب نسبت‌های دو نمونه است.

$$MH=\frac{\text{مجموع ستون } E_\circ(A) - \text{مجموع ستون } a}{\sqrt{\text{مجموع}\left(\text{ستون } (1-\frac{n_1}{n})\frac{n_1}{n} \times \text{ستون } \frac{n_1(n-m)}{n-1}\right)}}=\frac{۲٫۳۱}{۱٫۰۲}=۲٫۲۶$$

برای آزمون یک طرفهٔ $P=۰٫۰۱۲$ و $MH_C=\frac{۲٫۳۱-۰٫۵}{۱٫۰۲}=۱٫۷۷$ و $P=۰٫۰۳۸$ است. دنبالهٔ جدولهای ۲ × ۲ مانتل-هانزل به شرح زیر است:



	D	A
X	۱	
Y	۰	

n_1

	D	A
X	۰	
Y	۱	

n_2

	D	A
X	۲	
Y	۰	

n_3

۴.۲ نرمال مجانبی

برای اثبات نرمال مجانبی، فرض می‌کنیم تکرار وجود ندارد و موارد زیر را در نظر می‌گیریم:

جدول ۴. محاسبات مربوط به آمارهٔ مانتل-هانزل در آزمایش فرضی بالینی بران.

z	n	m_1	n_1	a	$E_0(A)$	$a-E_0(A)$	$\frac{m_1(n-m_1)}{n-1}$	$\frac{n_1}{n}(1-\frac{n_1}{n})$
۳	۱۰	۱	۵	۱	۰/۵۰	۰/۵۰	۱	۰/۲۵۰۰
۵	۹	۱	۴	۱	۰/۴۴	۰/۵۶	۱	۰/۲۴۶۹
۷	۸	۱	۳	۱	۰/۳۸	۰/۶۲	۱	۰/۲۳۴۴
۱۲	۶	۱	۱	۰	۰/۱۷	−۰/۱۷	۱	۰/۱۳۸۹
۱۸	۵	۱	۱	۱	۰/۲۰	۰/۸۰	۱	۰/۱۶۰۰
۱۹	۴	۱	۰	۰	۰	۰	۱	۰
۲۰	۳	۱	۰	۰	۰	۰	۱	۰
مجموع				۴	۱/۶۹	۲/۳۱		

$$N = n + m$$

$$\hat{H}(t) = \frac{1}{N}\sum_{i=1}^{N} I(Z_i \le t)$$

$$\hat{H}_1(t) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \le t)$$

$$\hat{H}_u(t) = \frac{1}{N}\sum_{i=1}^{N} I(Z_i \le t, \xi_i = 1)$$

$$\hat{H}_{1u}(t) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \le t, \delta_i = 1)$$

حال می‌توان صورت MH را به شرح زیر نوشت:

$$\sum_{i=1}^{k}(a_i - E_\circ(A_i)) = m\left\{\int_\circ^\infty d\hat{H}_{1u}(s) - \int_\circ^\infty \frac{1-\hat{H}_1(s^-)}{1-\hat{H}(s^-)}\, d\hat{H}_u(s)\right\}$$

زیرا، $E(A_i) = \frac{m_{i1}n_{i1}}{n_i}$. که در آن a_i، m_{i1}، n_{i1} و n_i از جدول ۲×۲، متناظر با مشاهدهٔ iام بریده نشده به دست می‌آید:

	D	A	
X	a_i		n_{i1}
Y			
	m_{i1}		n_i

چون فرض کرده‌ایم تکرار وجود ندارد، پس $m_{i1} = 1$. اگر s_i زمان iامین مشاهده بریده نشده باشد، داریم:

$$n_{i1} = \#(\text{های باقی‌مانده در زمان } X \; s_i^-) = m(1-\hat{H}_1(s_i^-))$$

$$n_i = \#(\text{های باقی‌مانده در زمان } Z \; s_i^-) = N(1-\hat{H}_1(s_i^-))$$

حال می‌توان صورت MH را برحسب توابع توزیع تجربی نوشت و می‌توان از روش

مشابه نرمال مجانبی برآوردگر PL، استفاده کرد.

REFERENCES
Mantel and Haenszel, J. Natl. Cancer Inst. (1962).
Crowley, JASA (1977).
Lininger et al., Biometrika (1979).

۳ ردهٔ آزمونهای تارون- وایر

بعد از ساختن جدول ۲×۲ برای هر مشاهده بریده شده، تارون و وایر، وزنی را برای هر جدول پیشنهاد می‌کنند، به گونه‌ای که:

$$\sum_{i=1}^{k} w_i[a_i - E_0(A_i)] = \sum_{i=1}^{k} w_i\left[a_i - \frac{m_{i1} n_{i1}}{n_i}\right] \qquad (12)$$

و برای واریانس

$$\sum_{i=1}^{k} w_i^2 \operatorname{Var}_0(A_i) = \sum_{i=1}^{k} w_i^2\left[\frac{m_{i1}(n_i - m_{i1})}{n_i - 1}\right]\times\left[\left(\frac{n_{i1}}{n_i}\right)\left(1-\frac{n_{i1}}{n_i}\right)\right] \qquad (13)$$

سه حالت ویژه مهم وجود دارد:

(الف) $w_i = 1$، در این صورت آمارهٔ MH به دست می‌آید.
(ب) $w_i = n_i$، در این صورت آمارهٔ گهان به دست می‌آید.
(پ) $w_i = \sqrt{n_i}$، در این صورت آمارهٔ تارون-وایر به دست می‌آید.

یادآوریها:

(الف) از کدام آزمون استفاده کنیم؟ آمارهٔ گهان به مشاهدات مقدماتی وزن بیشتری می‌دهد، در حالی که آمارهٔ MH به تمام مشاهدات وزن یک می‌دهد. پیشنهاد تارون-وایر بین دو روش قبلی قرار دارد. بنابه اظهار آنان وزنهای $w_i = \sqrt{n_i}$، تأثیر فراوانی در دامنهٔ تغییرات دارد.

(ب) هر چند (۱۲) با آمارهٔ گهان U برابر است، ولی $\widehat{\operatorname{Var}}_{TW}(U)$ که از رابطهٔ (۱۳) به دست می‌آید، برابر $\operatorname{Var}^*_{0,P}(U)$ نیست. به طور مجانبی $\widehat{\operatorname{Var}}_{TW}(U)$ معادل واریانس U، تحت فرض H_0 است، در حالی که $\operatorname{Var}^*_{0,P}(U)$، واریانس تحت H_0^* است.

مثال. با مراجعه به جدول ۴، که در آن آمارهٔ MH را محاسبه کرده‌ایم، داریم:

$$\sum_{i=۱}^{k} n_i(a_i - E_\circ(A_i)) = (۱۰)(۰/۵۰) + (۹)(۰/۵۶) + (۸)(۰/۶۲)$$

$$+ (۶)(-۰/۱۷) + (۵)(۰/۸۰) = ۱۷/۹۸$$

که برابر آمارهٔ U در آمارهٔ گهان است، البتّه بدون علامت و مقدار گرد شده، همچنین داریم:

$$\hat{\mathrm{Var}}_{TW}(U) = \sum n_i^۲ \left[\frac{m_{i۱}(n_i - m_{i۱})}{n_i - ۱}\right]\left[\left(\frac{n_{i۱}}{n_i}\right)\left(۱ - \frac{n_{i۱}}{n_i}\right)\right]$$

$$= (۱۰^۲)(۰/۲۵) + (۹^۲)(۰/۲۴۶۹) + (۸^۲)(۰/۲۳۴۴)$$

$$+ (۶^۲)(-۰/۱۳۸۹) + (۵^۲)(۰/۱۶) = ۶۹$$

$$\mathrm{Var}^*_{\circ,P}(U) = ۷۹/۴۴$$

در نتیجه داریم:

$$\sqrt{\hat{\mathrm{Var}}_{TW}(U)} = ۸/۳۱ \qquad \text{و} \qquad \sqrt{\mathrm{Var}^*_{\circ,P}(U)} = ۸/۹۱$$

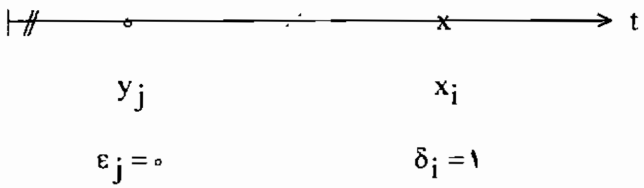
REFERENCE
Tarone and Ware, Biometrika (1977).

۴ آزمون افرون

به خاطر بیاورید که در ساختن آمارهٔ گهان، تابع امتیاز را به صورت زیر تعریف کردیم:

$$U_{ij} = \begin{cases} +۱ & t_i > u_j \\ \circ & \text{در سایر جاها} \\ -۱ & t_i < u_j \end{cases}$$

فرض کنید، حالت زیر را داریم:



آزمون گهان، امتیاز $U_{ij} = ۰$ را به این زوج -با صرفنظر از بزرگی X_i نسبت به Y_i- می‌دهد. افرون پیشنهاد می‌کند که امتیاز زیر را به آنها نسبت دهیم:

$$U_{ij} = \hat{P}\{T_i > U_j \mid (x_i, \delta_i), (y_j, \varepsilon_j)$$

برای حالت مورد نظر داریم:

$$U_{ij} = \hat{P}\{U_j < x_i \mid U_j > y_j\} = \frac{\hat{F}_۲(x_i) - \hat{F}_۲(y_j)}{۱ - \hat{F}_۲(y_j)}$$

در این رابطه $\hat{F}_۲$ برآوردگر PL کاپلان-مایر برای جمعیت ۲ است.

استفاده از این امتیازها و کاربرد ۱ و ۰ به جای ۱ و ۱-، آمارهٔ زیر را نتیجه می‌دهد:

$$\int_۰^\infty [۱ - \hat{F}_۱(u)]\, d\hat{F}_۲(u) = \hat{P}\{T_i > U_j\} \qquad (۱۴)$$

برآوردگر $\hat{P}(T_i > U_j)$ همان GMLE احتمال $P(T_i > U_j)$ است، که پارامتر آمارهٔ ویلکاکسن در حالت بریده نشده است، یعنی:

$$\frac{۱}{mn} U \xrightarrow{a.s.} P(X > Y)$$

در حالت بریده نشده: $U_{ij} = ۱$ یا ۰.

برآوردگر (۱۴)، در دنباله پایدار نیست و مانع کاربرد وسیع آن می‌شود.

REFERENCE

Efron, Proc. Fifth Berkeley Symp. IV (1967).

فصل پنجم

روشهای ناپارامتری K نمونه

برای نمونهٔ iام $(i=1,2,\ldots,K)$، فرض کنید T_{i1} تا T_{in_i}، متغیّرهای iid بوده، که توزیع هر یک F_i باشد. همچنین C_{i1} تا C_{in_i}، متغیّرهای iid با تابع توزیع G_i باشند. C_{ij} زمان برش مربوط به T_{ij} است. مشاهدات عبارت‌اند از: (X_{i1}, δ_{i1}) تا $(X_{in_i}, \delta_{in_i})$، که در آن:

$$X_{ij} = T_{ij} \wedge C_{ij} \qquad \text{و} \qquad \delta_{ij} = I(T_{ij} \le C_{ij})$$

فرض زیر مورد نظر ماست:

$$H_0: F_1 = \cdots = F_K$$

۱ آزمون گهان تعمیم‌یافته (برسلو)

با استفاده از تابع امتیاز مسألهٔ دو نمونه‌ای، فرضهای زیر را در نظر می‌گیریم:

$$W_i = \sum_{j=1}^{n_i} \sum_{\substack{i'=1 \\ \neq i}}^{K} \sum_{j'=1}^{n_{i'}} U((X_{ij}, \delta_{ij}), (X_{i'j'}, \delta_{i'j'}))$$

$$\underline{W} = (W_1, \ldots, W_K)'$$

برسلو ماتریس کوواریانس مجانبی $\underline{W}$، تحت فرض محدود کنندهٔ زیر را به دست آورد:

$$H_0^*: F_1 = \cdots = F_K \quad ; \quad G_1 = \cdots = G_K$$

با تعریف $N = \sum^{K} n_i$، فرض می‌کنیم با شرط $N \to \infty$ و $i=1,2,\ldots,K$، $\frac{n_i}{N} \to \lambda_i$

$$\underline{W} \overset{a}{\sim} N(\underline{\mu}^*_\circ, N^3 \underline{\Sigma}^*_\circ)$$

که در این رابطه $\underline{\mu}^*_\circ = \circ$ و

$$\underline{\Sigma}^*_\circ = \left(\int_\circ^\infty [1 - H(u)]^2 d H_u(u) \right) \times \begin{pmatrix} \lambda_1(1-\lambda_1) & & -\lambda_i \lambda_j \\ -\lambda_i \lambda_j & \ddots & \\ & & \lambda_K(1-\lambda_K) \end{pmatrix}$$

و

$$H_i(t) = P(X_{i1} \le t)$$

$$H_{iu}(t) = P(X_{i1} \le t, \delta_{i1} = 1)$$

$$H(t) = \lambda_1 H_1(t) + \cdots + \lambda_K H_K(t)$$

$$H_u(t) = \lambda_1 H_{1u}(t) + \cdots + \lambda_K H_{Ku}(t)$$

چون ماتریس کوواریانس مجانبی به پارامترهای مجهول بستگی دارد، جانشین می‌کنیم.

$$\hat{\lambda}_i = \frac{n_i}{N}$$

$$\hat{H}(t) = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} I(X_{ij} \le t)$$

$$\hat{H}_u(t) = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} I(X_{ij} \le t, \delta_{ij} = 1)$$

REFERENCE

Breslow, Biometrika (1970).

انواع آزمونها

۱ آزمون χ^2 امنیبوس: از آمارهٔ زیر استفاده می‌شود:

$$\frac{1}{N^3 \int_\circ^\infty (1-\hat{H})^2 d\hat{H}_u} \sum_{i=1}^{K} \frac{W_i^2}{\hat{\lambda}_i} \overset{a}{\sim} \chi^2_{K-1}$$

این آماره معادل $\underline{W}'(\underline{\Sigma}_{\circ}^{*})^{-}\underline{W}$ است، که $(\underline{\Sigma}_{\circ}^{*})^{-}$ معکوس تعمیم یافته $\underline{\Sigma}_{\circ}^{*}$ است.

اگر برش نباشد، این آماره به طور مجانبی معادل آمارهٔ کروسکال-والیس است. یادآوری می‌شود که اگر R_{ij} مرتبهٔ X_{ij} در میان N مشاهده باشد، داریم:

$$R_i = \sum_{j=1}^{n_i} R_{ij} \qquad \text{و} \qquad \overline{R}_i = \frac{1}{n_i} R_i \qquad \text{و} \qquad \overline{R}_{\bullet} = \frac{1}{N}\sum_{i=1}^{K} R_i$$

حال، آماره کروسکال-والیس، به شرح زیر است:

$$\frac{12}{N(N+1)}\sum_{i=1}^{K} n_i(\overline{R}_i - \overline{R}_{\bullet})^2 = \left(\frac{12}{N(N+1)}\sum_{i=1}^{K}\frac{n_i^2}{n_i}\right) - 3N(N+1)$$

این آماره دارای توزیع χ^2_{K-1}، تحت فرض $H_{\circ}$ است.

۲ **آزمون روند:** فرض کنید، اگر جمعیت‌ها همگی برابر نباشد، آن‌گاه قابل مرتّب شدن هستند. برای مثال می‌توان جمعیت‌ها را برحسب اندازهٔ مقدار دارو به صورت: $d_1 < \cdots < d_K$، مرتّب کرد. d_i، مقدار دارویی است که جمعیت iام دریافت می‌کند. در حالتهای دیگر، ممکن است از قبل بدانیم هرگاه جمعیت‌ها با هم متفاوت باشند، باید به طور یکنواخت تغییر کنند و این تغییر همبستگی عددی نیست.

$$\underline{\ell} = (d_1, \ldots, d_K)'$$

در هنگامی که متغیّرهای کمّی در دسترس نباشد، از تعریف زیر استفاده می‌کنیم:

$$\underline{\ell} = (-(K-1), \ldots, -3, -1, +1, +3, \ldots, +(K-1))' \qquad \text{K زوج}$$

$$\underline{\ell} = \left(-\frac{(K-1)}{2}, \ldots, -1, 0, +1, \ldots, +\frac{(K-1)}{2}\right)' \qquad \text{K فرد}$$

آبسلون و توکی پیشنهاد می‌کنند که مقابله‌های خطّی مرتبهٔ ۲ یا مرتبهٔ ۲ تا ۴، که برای مقادیر زوج K به شرح زیر تشریح می‌شود، مورد استفاده قرار گیرد.

$$\underline{\ell} = (-2(K-1), -(K-3), -(K-5), \ldots, +(K-5), +(K-3), +2(K-1))'$$

$$\underline{\ell} = (-4(K-1), -2(K-3), -(K-5), \ldots, +(K-5), +2(K-3), +4(K-1))'$$

برای نرمال کردن $\underline{W}$، می‌نویسیم: $\overline{W}_i = \frac{W_i}{n_i(N-n_i)}$ و $\underline{\overline{W}} = (\overline{W}_1, \ldots, \overline{W}_K)'$.

حال $\underline{c}$ را چنان در نظر بگیرید که، $\underline{\ell}'\,\overline{\underline{W}} = \underline{c}'\,\underline{W}$ باشد، در نتیجه داریم:

$$\frac{\underline{c}'\,\underline{W}}{\sqrt{N^{3}\,\underline{c}'\,\underline{\Sigma}^{*}_{0}\,\underline{c}}} \overset{a}{\sim} N(0,1)$$

آمارهٔ بالا را می‌توان برای آزمون فرض $H_0: F_1 = \cdots = F_K$ در مقابل $H_1: F_1 < \cdots < F_K$ به کار برد.

اگر کمیّتهای اندازه‌پذیر در دست باشند، روشهای رگرسیونی قابل استفاده‌اند. آنها را در فصل ششم بررسی می‌کنیم.

۱.۱ ماتریس کوواریانس جایگشت

متغیّر W^{*} را به صورت زیر تعریف می‌کنیم:

$$W^{*}_{ij} = \sum_{\substack{i'=1 \\ (i',j')\neq(i,j)}}^{K} \sum_{j'=1}^{n_{i'}} U((X_{ij},\delta_{ij}),(X_{i'j'},\delta_{i'j'}))$$

متغیّرهای $W^{*}_{Kn_K}$، ...، W^{*}_{K1} و ... و $W^{*}_{1n_1}$، ...، W^{*}_{11} را بــه صــورت $W^{*}_{1},\ldots,W^{*}_{N}$ در نظر می‌گیریم. برای محاسبهٔ توزیع جایگشت $\underline{W}$، فرض کنید W^{*}_{1} تا W^{*}_{N} معلوم‌اند. تحــت فرض H_0 و بدون جایگذاری W^{*}_{i}ها را نمونه‌گیری می‌کنیــم. فــرض کنیــد W_1 مجمــوع اوّلین n_1 مشاهدهٔ نمونه، W_2 مجموع دومین n_2 نمونه و تا آخر باشد.

ماتریس کوواریانس $\underline{W} = (W_1,\ldots,W_K)'$، تحت این طرح نمونه‌گیری به شــکل زیر است:

$$\underline{\Sigma}^{*}_{0,p} = \frac{1}{N}\left(\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(W^{*}_{ij})^{2}}{N-1}\right)\times\begin{pmatrix} n_1(N-n_1) & & -n_i n_j \\ & \ddots & \\ -n_i n_j & & n_K(N-n_K)\end{pmatrix}$$

ماتریس $\underline{\Sigma}^{*}_{0,p}$ را می‌توان به جای $N^{3}\hat{\underline{\Sigma}}^{*}_{0}$ به کار برد. زیرا، به طور مجانبی معادل‌اند.

REFERENCE

Marcuson and Nordbrock, Biom. Zeit. / Biom. J. (1981).

۲.۱ توزیع تحت $H_\circ$

تحت فرض: $H_\circ: F_1 = \cdots = F_K$، ثابت می‌شود که

$$\underline{W} \overset{a}{\sim} N(\underline{\mu}_\circ , N^{3}\underline{\Sigma}_\circ)$$

که درایه‌های $\underline{\Sigma}_\circ$، به شرح زیراند:

$$\sigma_{ij}^\circ = -\lambda_i \lambda_j \int_\circ^\infty (1-H_i)(1-H_j)\,dH_u \qquad i \neq j$$

$$\sigma_{ii}^\circ = \lambda_i \int_\circ^\infty [(1-H)(1-H_i) - \lambda_i(1-H_i)^2]\,dH_u$$

برای برآورد $\underline{\Sigma}_\circ$، به جای جایگذاری مقدار در H، H_i و H_u، ساده‌تر است که از آمارهٔ ML، استفاده کنیم.

REFERENCE
Breslow, Biometrika (1970).

۲ آزمون مانتل-هانزل تعمیم‌یافته (تارون و وایر)

فرض کنید، نمونهٔ ترکیبی مرتّب شده به صورت: $(Z_{(1)}, \xi_{(1)}), \ldots, (Z_{(N)}, \xi_{(N)})$ و همچنین به فرض $\Re_{(i)} = \Re(Z_{(i)}^-)$ باشد.
برای هر نقطهٔ زمانی بریده نشده، جدول 2×2 را به شرح زیر می‌سازیم. این جدول به ازای $k = 2$، ترانهادهٔ جدولهای 2×2 فصل چهارم است.

	۱	۲	$\cdots$	K	
مرده	$a_{i1} = \circ$	$a_{i2} = 1$	$\cdots$	$a_{iK} = \circ$	m_{i2}
زنده			$\cdots$		m_{i1}
	n_{i1}	n_{i2}	$\cdots$	n_{iK}	$N_i = \#\Re_{u(i)}$

تحت فرض $H_\circ: F_1 = \cdots = F_K$، داریم:

$$E_\circ(\underline{A}_i) = (E_\circ(A_{i1}), \ldots, E_\circ(A_{iK}))' = \left(\frac{m_{i1} n_{i1}}{N_i}, \ldots, \frac{m_{i1} n_{iK}}{N_i}\right)'$$

$$\underline{\Sigma}_\circ(\underline{A}_i) = \left(\frac{m_{i1} n_{i2}}{N_i - 1}\right) \times \begin{pmatrix} \frac{n_{i1}}{N_i}\left(1 - \frac{n_{i1}}{N_i}\right) & & -\frac{n_{ik}}{N_i}\frac{n_{i\ell}}{N_i} \\ -\frac{n_{ik}}{N_i}\frac{n_{i\ell}}{N_i} & \ddots & \\ & & \frac{n_{iK}}{N_i}\left(1 - \frac{n_{iK}}{N_i}\right) \end{pmatrix}$$

بنا به تعریف داریم:

$$\underline{a} - E_\circ(\underline{A}) = \sum_i w_i(\underline{a}_i - E_\circ(\underline{A}_i))$$

$$\underline{\Sigma}_\circ = \sum_i w_i^2 \underline{\Sigma}_\circ(\underline{A}_i)$$

در این روابط، w_i وزنهای نسبت داده شده‌اند. سه حالت خاص به شرح زیر وجود دارد:

(الف) $w_i = 1$، در این صورت آزمون تعمیم یافته MH به دست می‌آید.

(ب) $w_i = N_i$، در این صورت آزمون تعمیم یافته گهان به دست می‌آید.

(پ) $w_i = \sqrt{N_i}$، کارایی بالایی روی دامنهٔ تغییرات می‌دهد.

در بخش اوّل، یک ماتریس کوواریانس مجانبی از آمارهٔ تعمیم یافته گهان را تحت فرض محدود کنندهٔ $H_\circ^*$ به دست آوردیم. این ماتریس تحت $H_\circ$ برابر $\underline{\Sigma}_\circ$ بوده و از نظر محاسباتی از قبلی ساده‌تر است.

REFERENCE

Tarone and Ware, Biometrika (1977).

انواع آزمونها

۱ آزمون χ^2. چون $\underline{\Sigma}_\circ$ منفرد است، یکی از جمعیتها، مثلاً اوّلی را حذف می‌کنیم. بنابه تعریف $\underline{a}_{-1} - E_\circ(\underline{A}_{-1})$ و $\underline{\Sigma}_{\circ,-1}$ به ترتیب برابر $\underline{a}_{-1} - E_\circ(\underline{A}_{-1})$ و $\underline{\Sigma}_{\circ,-1}$ باشند، که در

آنها جمعیت اوّل حذف شده است. حال تحت فرض $H_۰$، داریم:

$$W = (\underline{a}_{-۱} - E_۰(\underline{A}_{-۱}))' \underline{\Sigma}_{۰,-۱}^{-۱} (\underline{a}_{-۱} - E_۰(\underline{A}_{-۱})) \overset{a}{\sim} \chi^۲_{K-۱}$$

اگر هر کدام از جمعیت‌ها را حذف کنیم، W تغییر نمی‌کند.

برای آزمون تعمیم یافتهٔ مانتل-هانزل $(w_i = ۱)$ آزمون تقریبی زیر را داریم:

$$\sum \frac{(O-E)^۲}{E} = \sum_{k=۱}^{K} \frac{(a_k - E_۰(A_k))^۲}{E_۰(A_k)} \approx \chi^۲_{K-۱}$$

که در آن $a_k = \sum_i a_{ik}$ و $E_۰(A_k) = \sum_i E_۰(A_{ik}) = \sum_i \frac{m_{i۱} n_{ik}}{N_i}$.

هرچند این آزمون به علّت نامساوی $\sum \frac{(O-E)^۲}{E} \le W$ تا حدودی محتاطانه است. ولی، در کاربرد ساده‌تر است. زیرا در آن به محاسبهٔ معکوس ماتریس نیازی نیست.

REFERENCES

Peto and Pike, Biometrics (1973).

Peto et al. , British J. Cancer (1976, 1977).

۲ آزمون روند. فرض $H_۱$ را به صورت $H_۱: F_۱ < \cdots < F_K$، در نظر می‌گیریم. برای انتخاب $\underline{\ell}$ مطابق بخش اوّل، از آمارهٔ $\underline{\ell}'(\underline{a} - E_۰(\underline{A}))$، که به طور مجانبی نرمال است، استفاده می‌کنیم.

REFERENCE

Tarone, Biometrika (1975).

فصل ششم

روشهای ناپارامتری: رگرسیون

۱ الگوهای نرخ شکست متناسب کاکس

فرض کنید T_1 تا T_n و C_1 تا C_n، متغیّرهای تصادفی مستقل باشند. متغیّر C_i زمان برش مربوط به زمان بقاء T_i است. مشاهدات به صورت زیراند، که در آن $Y_i = T_i \wedge C_i$ و $\delta_i = I(T_i \leq C_i)$ است.

$$(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$$

همچنین، مقادیر معلوم به صورت: x_1، ...، $\underline{x}_n$ بوده که در آن $\underline{x}_i = (x_{i1}, \ldots, x_{ip})'$، بردار متغیّرهای وابسته به متغیّر تابع T_i است. به خاطر بیاورید که تابع نرخ شکست را به صورت زیر تعریف کردیم، که در آن وابستگی T از طریق $\underline{x}$، منظور شده است.

$$\lambda(t; \underline{x}) = \frac{f(t; \underline{x})}{1 - F(t; \underline{x})}$$

در این الگو فرض می‌شود که: $\lambda(t; \underline{x}) = e^{\underline{\beta}' \underline{x}} \lambda_0(t)$ باشد، که بردار: $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$، ضرایب رگرسیون است. نرخ شکست برابر حاصل‌ضرب یک عدد در تابع $\lambda_0(t)$ است، که این ضریب عددی به ضرایب رگرسیون و کوواریانسها وابسته است. قضیه می‌تواند در این حالت نیز برقرار باشد که، به جای $e^{\underline{\beta}' \underline{x}}$، هر مقدار محسوس $h(\underline{\beta}' \underline{x})$ را، که در آن h مثبت باشد، جانشین کنیم. ضرایب رگرسیون $\underline{\beta}$ و تابع نرخ شکست $\lambda_0(t)$، هیچ کدام معلوم نیستند.

خانواده‌ای از توزیعها را "خانوادهٔ توزیعهای لهمن" گوییم، هرگاه تابع توزیعی مانند F_0 وجود داشته باشد، به‌گونه‌ای که در ازای هر F در این خانواده

رابطهٔ $1-F=(1-F_\circ)^\gamma$ برقرار باشد. در این رابطه γ یک عدد مثبت است. می‌توان رابطه را برحسب تابع بقاء و به صورت $S=S_\circ^\gamma$، نیز نوشت.

از الگوی نرخهای متناسب نتیجه می‌شود که توابع توزیع آنها تشکیل یک خانوادهٔ توزیعهای لهمن می‌دهند. اثبات آن به شرح زیر است:

$$S(t;\underline{x})=\exp\left\{-\int_\circ^t \lambda(u;\underline{x})du\right\}=\exp\left\{-e^{\underline{\beta}'\underline{x}}\int_\circ^t \lambda_\circ(u)du\right\}$$

$$=\exp\left\{-\int_\circ^t \lambda_\circ(u)du\right\}^{e^{\underline{\beta}'\underline{x}}}=S_\circ(t)^{e^{\underline{\beta}'\underline{x}}}$$

که در رابطهٔ بالا $S_\circ(t)=\exp\left\{-\int_\circ^t \lambda_\circ(u)du\right\}$ است. حالت خاص $p=1$ را در نظر می‌گیریم:

$$x_i=\begin{cases}1 & \text{اگر مشاهده } i \text{ام از جمعیت ۱ باشد}\\ \circ & \text{اگر مشاهده } i \text{ام از جمعیت ۲ باشد}\end{cases}$$

$$e^{\beta x_i}=\begin{cases}e^{\beta}=\gamma & \text{اگر مشاهده } i \text{ام از جمعیت ۱ باشد}\\ 1 & \text{اگر مشاهده } i \text{ام از جمعیت ۲ باشد}\end{cases}$$

در نتیجه توابع بقاء برای جمعیت ۱ و ۲، به صورت زیر ارتباط دارند:

$$S_1(t)=S_2^\gamma(t)$$

۱.۱ تحلیل درست‌نمایی شرطی

در نوشته‌های کاکس داریم: فرض کنید $\lambda_\circ(t)$ اختیاری باشد. هیچ اطّلاعی دربارهٔ β از فاصله‌های زمانی، که در آنها هیچ شکستی رخ نداده، به دست نمی‌آید. زیرا، امکان دارد، $\lambda_\circ(t)$ به طور قابل درکی با صفر متّحد باشد. بنابراین، روی آن لحظه‌هایی که در آنها شکست رخ می‌دهد، شرطی می‌کنیم. در زمان گسسته نیز روی مشاهدات تکراری شرطی می‌کنیم. به هر حال، نیاز به روشی داریم که تمام $\lambda_\circ(t)$ را تحلیل نماید.

در نظر گرفتن این توزیع شرطی اجباری به نظر می‌رسد.

فرض کنید تکرار وجود نداشته باشد. حالتی که تکرار وجود دارد را بعداً بررسی می‌کنیم. زمانهای مشاهده‌ای را به صورت $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$، مرتب می کنیم. فرض کنید، $\delta_{(i)}$ تابع نشانگر و $\underline{x}_{(i)}$، متغیّر وابسته به $y_{(i)}$ باشد، می‌نویسیم: $\Re_{(i)} = \Re(y_{(i)}-)$. برای هر زمان بریده نشدهٔ $y_{(i)}$، داریم:

$$P\{[y_{(i)}, y_{(i)} + \Delta y] \text{ یک مرگ در بازهٔ } \mid \Re_{(i)}\} \cong \sum_{j \in \Re_{(i)}} e^{\underline{\beta}'\underline{x}} \lambda_\circ(y_{(i)}) \Delta y$$

$$P\{y_{(i)} \text{ در زمان } (i) \text{ مرگ } \mid y_{(i)} \text{ در زمان } \Re_{(i)} \text{ یک مرگ در }\} = \frac{e^{\underline{\beta}'\underline{x}_{(i)}}}{\sum_{j \in \Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}}$$

اگر حاصل‌ضرب سه احتمال شرطی را حساب کنیم، درست‌نمایی شرطی به دست می‌آید.

$$L_c(\underline{\beta}) = \prod_u \frac{e^{\underline{\beta}'\underline{x}_{(i)}}}{\sum_{j \in \Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}}$$

کاکس پیشنهاد می کند، که از درست‌نمایی شرطی به جای درست‌نمایی معمولی استفاده کنیم. به ویژه پیدا کردن برآورد حداکثر درست‌نمایی از بردار امتیاز و ماتریس اطّلاعات نمونه، به شرح زیر استفاده شود:

$$\frac{\partial}{\partial \underline{\beta}} \log L_c(\underline{\beta}) = \left(\frac{\partial}{\partial \beta_1} \log L_c(\underline{\beta}), \ldots, \frac{\partial}{\partial \beta_p} \log L_c(\underline{\beta}) \right)'$$

$$\underline{i}(\underline{\beta}) = -\frac{\partial^2}{\partial \underline{\beta}^2} \log L_c(\underline{\beta}) = -\begin{pmatrix} \frac{\partial^2}{\partial \beta_1 \partial \beta_1} \log L_c(\underline{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_1 \partial \beta_p} \log L_c(\underline{\beta}) \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial \beta_p \partial \beta_1} \log L_c(\underline{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_p \partial \beta_p} \log L_c(\underline{\beta}) \end{pmatrix}$$

می‌خواهیم معادلات زیر را، که معمولاً به روش تکرار منجر می‌شود، حل کنیم:

$$\frac{\partial}{\partial \underline{\beta}} \log L_c(\underline{\beta}) = \underline{0}$$

فرض کنید، $\hat{\underline{\beta}}^{0}$ یک حدس اولیّه باشد، داریم:

$$\hat{\underline{\beta}}^{1} = \hat{\underline{\beta}}^{0} + \underline{i}^{-1}(\hat{\underline{\beta}}^{0}) \frac{\partial}{\partial \underline{\beta}} \log L_c(\hat{\underline{\beta}}^{0})$$

اگر $\hat{\underline{\beta}}$ جواب معادله باشد، داریم:

$$\hat{\underline{\beta}} \overset{a}{\underset{\sim}{}} N(\underline{\beta}, \underline{i}^{-1}(\underline{\beta}))$$

با مشتق‌گیری از عبارت زیر، رابطهٔ بردار امتیاز و ماتریس اطّلاعات نمونه به دست می‌آید، داریم:

$$\log L_c(\underline{\beta}) = \sum_{u} \left[\underline{\beta}' \underline{x}_{(i)} - \log\left(\sum_{j \in \mathfrak{R}_{(i)}} e^{\underline{\beta}' \underline{x}_j} \right) \right]$$

$$\frac{\partial}{\partial \beta_k} \log L_c(\underline{\beta}) = \sum_{u} \left(x_{(i)k} - \frac{\sum_{j \in \mathfrak{R}_{(i)}} x_{jk} e^{\underline{\beta}' \underline{x}_j}}{\sum_{j \in \mathfrak{R}_{(i)}} e^{\underline{\beta}' \underline{x}_j}} \right)$$

$$i_{k\ell}(\underline{\beta}) = -\frac{\partial^2}{\partial \beta_k \partial \beta_\ell} \log L_c(\underline{\beta})$$

$$= \sum_{u} \left(\frac{\sum_{j \in \mathfrak{R}_{(i)}} x_{jk} x_{j\ell} e^{\underline{\beta}' \underline{x}_j}}{\sum_{j \in \mathfrak{R}_{(i)}} e^{\underline{\beta}' \underline{x}_j}} - \frac{\sum_{j \in \mathfrak{R}_{(i)}} x_{jk} e^{\underline{\beta}' \underline{x}_j}}{\sum_{j \in \mathfrak{R}_{(i)}} e^{\underline{\beta}' \underline{x}_j}} \times \frac{\sum_{j \in \mathfrak{R}_{(i)}} x_{j\ell} e^{\underline{\beta}' \underline{x}_j}}{\sum_{j \in \mathfrak{R}_{(i)}} e^{\underline{\beta}' \underline{x}_j}} \right)$$

برای آزمون $H_0: \underline{\beta} = \underline{0}$، کاکس از آمارهٔ نوع راتو به شرح زیر، که به طور مجانبی دارای توزیع χ^2_p تحت فرض H_0 است، استفاده می‌کند.

$$\left(\frac{\partial}{\partial \underline{\beta}} \log L_c(\underline{\circ})\right)' \underline{i}^{-1}(\underline{\circ}) \left(\frac{\partial}{\partial \underline{\beta}} \log L_c(\underline{\circ})\right)$$

بردار امتیاز و ماتریس اطّلاعات نمونه در $\underline{\beta} = \underline{\circ}$، دارای صورت سادهٔ زیر است:

$$\frac{\partial}{\partial \beta_k} \log L_c(\underline{\circ}) = \sum_u (x_{(i)k} - \bar{x}_{(i)k})$$

$$i_{k\ell}(\underline{\circ}) = L_c(\underline{\beta}) = \sum_u \left(\frac{1}{n_i} \sum_{j \in \Re_{(i)}} x_{jk} x_{j\ell} - \bar{x}_{jk} \bar{x}_{j\ell} \right)$$

$$= \sum_u \left(\frac{1}{n_i} \sum_{j \in \Re_{(i)}} (x_{jk} - \bar{x}_{(i)k}) \times (x_{j\ell} - \bar{x}_{(i)\ell}) \right)$$

که در این روابط، $n_i = \#\Re_{(i)}$ و $\underline{\bar{x}}_{(i)} = \dfrac{\sum_{j \in \Re_{(i)}} \underline{x}_j}{n_i}$ است.

ماتریس اطّلاع نمونه عبارت است از: مجموع ماتریسهای کوواریانس برای مجموعه‌های نرخ شکست مشاهدات بریده نشده.

حالت خاص $p = 1$ و x_i را به شرح زیر در نظر می‌گیریم:

$$x_i = \begin{cases} 1 & \text{اگر } i \text{ در نمونه ۱ باشد} \\ 0 & \text{اگر } i \text{ در نمونه ۲ باشد} \end{cases}$$

در این صورت با توجّه به نمادهای آزمون MH، داریم:

$$\frac{\partial}{\partial \beta} \log L_c(\underline{\circ}) = \sum_u (x_{(i)} - \bar{x}_{(i)}) = \sum_u \left(a_i - \frac{n_{i1}}{n_i} \right)$$

$$i(\underline{\circ}) = \sum_u \left(\frac{1}{n_i} \sum_{j \in \Re_{(i)}} x_j^2 - \bar{x}_{(i)}^2 \right)$$

$$= \sum_u \frac{n_{i1}}{n_i} \left(1 - \frac{n_{i1}}{n_i} \right)$$

در نتیجه اگر تکرار نباشد، آزمون کاکس و MH یکسان هستند.

REFERENCES

Cox, JRSS B (1972).

Prentice and Kalbfleisch, Biometrics (1979), has a nice survey of the Cox procedure.

Kalbfleisch and Prentice, The Statistical Analysis of Failure Time Data (1980), is an excellent new text on the Cox approach.

۲.۱ بررسی درست‌نمایی شرطی

درست‌نمایی حاشیه‌ای برای رتبه‌ها. مجدداً فرض کنید تکرار وجـــود نـدارد. بـا وجود تکرار استدلال زیر صحیح نیست.

فرض نمایید داده‌ها بریده نشـــده‌اند و Y_1 تـا Y_n مسـتقل و Y_i دارای توزیـع F_i و چگالی f_i باشد. دو نماد $\underline{Y}=(Y_1,\ldots,Y_n)$ و $\underline{R}=(R_1,\ldots,R_n)$، کـــه در آن R_i، رتبـه مشاهدهٔ Y_i است را در نظر می‌گیریم. در این صورت احتمال بردار رتبهٔ $\underline{r}$ به صورت زیـر است:

$$p(\underline{r}) = \int\cdots\int_{u_1<\cdots<u_n} \prod_{i=1}^{n} f_{(i)}(u_i)\, du_1\cdots du_n$$

در این رابطه $f_{(i)}$ چگالی متناظر $y_{(i)}$ است. برای مثال با $n=3$ و $\underline{r}=(3,1,2)$، داریم:

$$p(\underline{r}) = P\{R_1=3, R_2=1, R_3=2\} = \iiint_{u_1<u_2<u_3} f_2(u_1) f_3(u_2) f_1(u_3)\, du_1\, du_2\, du_3$$

کلب‌فلیچ و پرنتیس رابطهٔ زیر را اثبات نموده‌اند:

$$F_i(t) = 1-\exp\left(-e^{\underline{\beta}'\underline{x}_i}\int_0^t \lambda_0(u)du\right)$$

در نتیجه داریم:

$$p(\underline{r})\} = \prod_{i=1}^{n} \frac{e^{\underline{\beta}'\underline{x}_{(i)}}}{\sum_{j\in\Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}}$$

در حالت برش با استفاده از نماد $(Y_۱, \delta_۱), \ldots, (Y_n, \delta_n)$ و فرض این که Y_i دارای توزیع F_i و چگالی f_i است، بردار رتبه را به صورت زیر تعریف می‌کنیم:

$$\underline{R}^{u/c} = (R_۱^{u/c}, \ldots, R_n^{u/c})$$

که در آن:

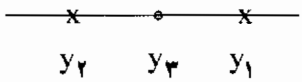
$$R_i^{u/c} = \begin{cases} \text{رتبهٔ } Y_i \text{ در میان مشاهدات بریده نشده با شرط} & \delta_i = ۱ \\ \text{رتبهٔ مشاهدهٔ بریده نشدهٔ قبلی با شرط} & \delta_i = ۰ \end{cases}$$

و بردار نشانگر $\underline{\delta}$ به صورت $\underline{\delta} = (\delta_۱, \ldots, \delta_n)$، است.
حال احتمال $(\underline{r}^{u/c}, \underline{\delta})$ به صورت زیر است:

$$p(\underline{r}^{u/c}, \underline{\delta}) = \int \cdots \int_{u_۱ < \cdots < u_{n_u}} \prod_{i=۱}^{n_u} \left\{ f_{u(i)}(u_i) \times \prod_{j \in C_{i,\, i+۱}} [۱ - F_j(u_i)] \right\} du_۱ \cdots du_{n_u}$$

در این رابطه، $f_{u(i)}$، چگالی متناظر iامین مشاهده بریده نشدهٔ مرتب، $C_{i,\, i+۱}$، مجموع اندیسهای متناظر مشاهدات بریده شده واقع در بین مشاهدات بریده نشدهٔ iام و $(i+۱)$ام و n_u تعداد کلّ مشاهدات بریده نشده است.

برای مثال، $\underline{r} = (۲, ۱, ۱)$ و $\underline{\delta} = (۱, ۱, ۰)$، متناظر با شکل زیر است:



احتمال رتبه به شرح زیر است:

$$p((۲, ۱, ۱), (۱, ۱, ۰)) = \iint_{u_۱ < u_۲} f_۲(u_۱)[۱ - F_۳(u_۱)] \times f_۱(u_۲)\, du_۱\, du_۲$$

کلب‌فلیش و پرنتیس نشان دادند که اگر $F_i(t)$ به صورت زیر باشد، امتحان قابل محاسبه است.

$$F_i(t) = ۱ - \exp\left(-e^{\underline{\beta}' \underline{x}_i} \int_۰^t \lambda_۰(u)\, du \right)$$

$$p(\underline{r}^{u/c}, \underline{\delta}) = \prod_{u} \frac{e^{\underline{\beta}' \underline{x}_{(i)}}}{\sum_{j \in \Re_{(i)}} e^{\underline{\beta}' \underline{x}_j}} = L_c$$

REFERENCE
Kalbfleisch and Prentice, Biometrika (1973).

درست‌نمایی جزئی. دنبالهٔ کمیّتهای تصادفی زیر را در نظر بگیرید:

$$(X_1, S_1;\ X_2, S_2;\ \ldots;\ X_m, S_m)$$

در رگرسیون با برش، فرض کنید: $y_{u(i)}$، مشاهد iام بریده نشده باشد. همچنین، فرض کنید: متغیّر X_i دارای تمام اطّلاعات بریده شده در $[y_{u(i-1)}, y_{u(i)})$ -با این اطّلاع که یک شکست در زمان $y_{u(i)}$ رخ داده- باشد. S_i را متغیّری در نظر می‌گیریم که شامل مشاهدهٔ خاصی باشد، که با همبستگی $\underline{x}_{u(i)}$ در زمان $y_{u(i)}$ شکست می‌خورد.
درست‌نمایی حاشیه‌ای S_1 تا S_m به شرح زیر است:

$$p(S_1, \ldots, S_m | \underline{\beta}) = \prod_{i=1}^{m} p(S_i | S_1, \ldots, S_{i-1}; \underline{\beta})$$

و درست‌نمایی شرطی S_1 تا S_m با فرض X_1 تا X_m نیز مطابق زیر است:

$$p(S_1, \ldots, S_m | X_1, \ldots, X_m; \underline{\beta}) = \prod_{i=1}^{m} p(S_i | S_1, \ldots, S_{i-1}; X_1, \ldots, X_m; \underline{\beta})$$

و بالاخره درست‌نمایی کامل به شرح زیر است:

$$p(X_1, \ldots, X_m; S_1, \ldots, S_m | \underline{\beta})$$

$$= \prod_{i=1}^{m} p(X_i, S_i | X_1, \ldots, X_{i-1}; S_1, \ldots, S_{i-1}; \underline{\beta})$$

$$= \prod_{i=1}^{m} p(X_i | X_1, \ldots, X_{i-1}; S_1, \ldots, S_{i-1}; \underline{\beta}) \times$$

$$\times \prod_{i=1}^{m} p(S_i | X_1, \ldots, X_{i-1}; S_1, \ldots, S_{i-1}; \underline{\beta})$$

کاکس، عبارت دوم حاصل‌ضرب (عبارت زیر) را درست‌نمایی جزئی نامیده است.

$$\prod_{i=1}^{m} p(S_i \mid X_1, \ldots, X_{i-1}, X_i; S_1, \ldots, S_{i-1}; \underline{\beta})$$

در رگرسیون با داده‌های بریده شده، درست‌نمایی جزئی با L_c برابر است، که ما آن را درست‌نمایی شرطی گوییم. از مقایسه تعاریف درست‌نمایی‌های شرطی و جزئی، معلوم می‌شود که، درست‌نمایی جزئی، یک درست‌نمایی شرطی یا حاشیه‌ای واقعی نیست. کاکس مدّعی است که: درست‌نمایی جزئی شامل بیشترین اطّلاعات دربارهٔ $\underline{\beta}$ برای رگرسیون با داده‌های بریده شده است. لذا، می‌توان از جملهٔ اوّل حاصل‌ضرب، یعنی عبارت زیر چشم پوشی کرد.

$$\prod_{i=1}^{m} p(X_i \mid X_1, \ldots, X_{i-1}; S_1, \ldots, S_{i-1}; \underline{\beta})$$

بدون آن‌که چیز زیادی از دست برود، افرون، اطّلاع فیشر را در درست‌نمایی جزئی با اطّلاع فیشر در درست‌نمایی کامل برای تعدادی از الگوها مقایسه کرده است. معمولاً اطّلاع در L_c بسیار بالاست. با کارایی بیش از ۹۰٪. در حالات نادری L_c، حدّاقل به اندازهٔ درست‌نمایی کامل شامل اطّلاعات است.

REFERENCES

Cox, Biometrika (1975).

Efron, JASA (1977).

Oakes, Biometrika (1977).

۳.۱ بررسی نرمال مجانبی

در مقالهٔ ۱۹۷۲، کاکس بیان شده، که $\hat{\underline{\beta}}$، جواب معادلهٔ $\frac{\partial}{\partial \underline{\beta}} \log L_c(\underline{\beta}) = \underline{0}$، به طور مجانبی دارای توزیع نرمال است. باکس در مقالهٔ ۱۹۷۵ خود یک شناسهٔ کاشف که مشابه شناسهٔ درست‌نمایی حداکثر استاندارد است را ارائه می‌دهد.

تسیاتیس، اثباتی از نرمال مجانبی $\hat{\underline{\beta}}$ با کاربرد انتگرال و فرایندهای تصادفی ارائه داده است. این اثبات مشابه اثباتی است که توسّط برسلو و کرولی برای نرمال مجانبی برآوردگر PL و اثبات ارائه شده توسّط کرولی برای آمارهٔ MH است.

بیلی با کاربرد تصویرهای هاجک یک شناسه ارائه داده است.

REFERENCES
Cox, Biometrika (1975).
Bailey, Univ. of Chicago thesis (1979).
Tsiatis, Ann. Stat. (1981).

۴.۱ برآورد $S(t;\underline{x})$

تحت الگوی نرخ شکست کاکس، داریم:

$$S(t;\underline{x}) = \exp\left(-e^{\underline{\beta}'\underline{x}} \int_{\circ}^{t} \lambda_{\circ}(u)du\right) = \exp\left(-e^{\underline{\beta}'\underline{x}} \Lambda_{\circ}(t)\right) = S_{\circ}(t)^{e^{\underline{\beta}'\underline{x}}}$$

که در آن $S_{\circ}(t) = e^{-\Lambda_{\circ}(t)}$، است.

برای برآورد $S(t;\underline{x})$، به جای $\underline{\beta}$ از $\hat{\underline{\beta}}$ استفاده می‌کنیم. ولی $\Lambda_{\circ}(t)$ یا $S_{\circ}(t)$ را چگونه برآورد کنیم.

برسلو فرض می‌کند که $\lambda_{\circ}(t)$ بین مشاهدات بریده نشده ثابت است، داریم:

$$\hat{\lambda}_{\circ,B}(t) = \frac{۱}{(y_{u(i)} - y_{u(i-۱)}) \sum_{j\in\Re_{u(i)}} e^{\underline{\beta}'\underline{x}_j}} \quad , \; y_{u(i-۱)} < t < y_{u(i)}$$

و $S_{\circ}(t)$ را به صورت زیر برآورد می‌کنیم:

$$\hat{S}_{\circ,B}(t) = \prod_{y_{(i)}\le t} \left(۱ - \frac{\delta_{(i)}}{\sum_{j\in\Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}} \right)$$

توجّه شود که $\hat{\Lambda}_{\circ,B}(t) = \int_{\circ}^{t} \hat{\lambda}_{\circ,B}(u)du$ و $\hat{S}_{\circ,B}(t)$، حتی به ازای $t = y_{(i)}$، سازگار نیستند، یعنی:

$$\hat{S}_{\circ,B}(t) \neq e^{-\hat{\Lambda}_{\circ}(t)}$$

همچنین $\hat{S}_{\circ,B}(t)$ می‌تواند مقادیر منفی اختیار کند. تسیاتیس از مقدار زیر استفاده

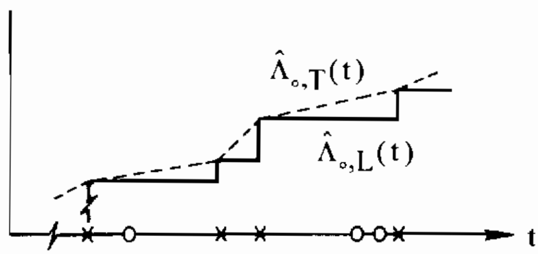
می‌کند:

$$\hat{S}_{\circ,T}(t)=e^{-\hat{\Lambda}_{\circ,T}(t)}$$

که در آن $\hat{\Lambda}_{\circ,B}(t)=\sum\limits_{y_{(i)}\le t}\dfrac{\delta_{(i)}}{\sum\limits_{j\in\Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}}$ است. ولی هنگامی که $\underline{\hat{\beta}}=\underline{\circ}$ باشد، $\hat{S}_{\circ,T}$ به برآوردگر PL، ساده نمی‌شود. توجّه نمایید که $\hat{\Lambda}_{\circ}(t)$ یک تابع پلّه‌ای است.

اینک یک صورت خطّی $\hat{\Lambda}_{\circ,B}(t)=\sum\limits_{y_{(i)}\le t}\dfrac{\delta_{(i)}}{\sum\limits_{j\in\Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}}$ را با تعریف زیر به کار می‌برد، که معادل انتگرال برسلوی $\lambda_{\circ}(t)$ است.

$$\hat{S}_{\circ,L}(t)=e^{-\hat{\Lambda}_{\circ,L}(t)}$$



تسیاتیس و لینک هر دو $\widehat{\mathrm{Var}}(\hat{S}_{\circ}(t))$ را محاسبه کرده‌اند. اوّلی با استفاده از الگوی درست‌نمایی و دومی به روش دلتا. لینک با استفاده از روشهای مونت‌کارلو، فاصله‌های اطمینان مربوط به $\hat{S}_{\circ,L}(t)$، $\log\hat{S}_{\circ,L}(t)$ و $\operatorname{logit}\hat{S}_{\circ,L}(t)=\log\dfrac{S_{\circ,L}(t)}{1-\log\hat{S}_{\circ,L}(t)}$ را بررسی و اثبات می‌کند که احتمال پوشاندن $\hat{S}_{\circ,L}(t)$ خیلی کم، از $\operatorname{logit}\hat{S}_{\circ,L}(t)$ خیلی زیاد و از $\log\hat{S}_{\circ,L}(t)$ تقریباً درست است. این نتایج مربوط به پوشش فاصله‌های اطمینان، برای برآوردگر PL نیز برقرار است.

برآوردگرهای دیگر $S(t;\underline{x})$، که از نظر محاسباتی پیچیده‌تراند توسّط کاکس و افرون پیشنهاد شده است.

REFERENCES

Breslow, JRSS B (1972), in Discussion on Cox's paper.

______, Biometrics (1974).

Tsiatis, Univ. Wisconsin Tech. Report No. 524 (1978).

______, Ann. Stat. (1981).

Link, Stanford Univ. Tech. Report No. 45 (1979).

۵.۱ داده‌های گسسته یا طبقه‌ای

زمانهای متمایز و مرتّب شدهٔ بقاء را به صـــورت: $y'_{(۱)} < \cdots < y'_{(r)}$، می‌نویســیم و فرض زیر را در نظر می گیریم:

$$\Re_{(i)} = \text{نرخ شکست در زمان } y'_{(i)}-$$

$$\wp_{(i)} = \text{(مجموعه مرگهای جدا در زمان } y'_{(i)}\text{) مرگ در زمان } y'_{(i)}$$

$$d_i = \#(\wp_{(i)})$$

کاکس پیشنهاد می کند که از کمیّت زیر استفاده شود:

$$L_c = \prod_{i=۱}^{r} P\{\wp_{(i)} \mid \Re_{(i)}, d_i\}$$

با:

$$P\{\wp_{(i)} \mid \Re_{(i)}, d_i\} = \frac{\exp\left(\sum_{j\in\wp_{(i)}} \underline{\beta}'\underline{x}_j\right)}{\sum_{\wp^*_{(i)}} \exp\left(\sum_{j\in\wp^*_{(i)}} \underline{\beta}'\underline{x}_j\right)}$$

در رابطهٔ قبل، مجموع مخرج روی تمام زیر مجموعه‌های $\wp^*_{(i)} \subset \Re_{(i)}$ در نظر گرفتــه می‌شود، به گونه‌ای که $\#\wp^*_{(i)} = d_i$ باشد. برای $i = ۱, ۲, \ldots, r$، تعـــداد $\binom{n_i}{d_i}$ زیرمجموعــه وجود دارد، که حتی برای مجموعهٔ داده‌های متوسّط هم از نظر محاسباتی مناسب به نظـــر نمی‌رسد.

یک تابع درست‌نمایی دیگر به شرح زیر توسّط برسلو و دیگران پیشنهاد شده است. اگر تعداد تکرارها زیاد نباشد، در عمل معقول به نظر می‌رسد:

$$L_c = \prod_{i=1}^{r} \frac{\exp\left(\sum_{j\in\wp_{(i)}} \underline{\beta}'\underline{x}_j\right)}{\left(\sum_{j\in\Re_{(i)}} e^{\underline{\beta}'\underline{x}_j}\right)^{d_i}}$$

پرنتیس محور زمان را به صورت: $0 = a_0 < a_1 < \cdots < a_{r-1} < a_r = \infty$، $A_j = [a_{j-1}, a_j)$، افراز می‌کند. اگر یک زمان بقاء در بازهٔ A_j قرار گیرد، آن‌گاه زمان j را ثبت می‌کنیم.

با توجّه به عبارت: $\alpha_j = \exp\left(-\int_{a_{j-1}}^{a_j} \lambda_0(t)dt\right)$، که احتمال شرطی یک فرد با متغیّر وابستهٔ $\underline{x} = \underline{x}$ واقع در بازهٔ A_j -با فرض این‌که در بازهٔ A_{j-1} زنده بوده- است، می‌توان احتمال مشاهدهٔ iام را که در ابتدای A_j زنده است، به صورت زیر محاسبه کرد:

$$\prod_{k=1}^{j-1} \alpha_k^{e^{\underline{\beta}'\underline{x}_i}}$$

$$P\{Y_i = j, \delta_i\} = \left(\prod_{k=1}^{j-1} \alpha_k^{e^{\underline{\beta}'\underline{x}_i}}\right)\left(1-\alpha_j^{e^{\underline{\beta}'\underline{x}_i}}\right)^{\delta_i}$$

درست‌نمایی کامل به صورت: $L = \prod_{i=1}^{n} P\{Y_i = j, \delta_i\}$ بوده که تابعی از پارامترهای مجهول $\underline{\beta}$ و α_1 تا α_r است.

برای برآورد این پارامترها از حداکثر درست‌نمایی استفاده می‌کنیم. توجّه شود که α_iها به صورت زیر محدود می‌شوند:

$$0 < \alpha_j < 1 \quad \text{و} \quad j = 1, \ldots, r \quad \text{و} \quad \sum_{j=1}^{r} \alpha_j = 1$$

اگر α_r را حذف نموده و فرض کنیم: $\gamma_j = \log(-\log\alpha_j)$ با $j = 1, \ldots, r-1$، به طوری که:

$$-\infty < \gamma_j < +\infty \quad \text{و} \quad j = 1, \dots, r-1$$

آن‌گاه حداکثر کردن، نسبت به γ_1 تا γ_{r-1}، ساده‌تر است از این‌که نسبت به α_1 تا α_r حداکثر کنیم. زیرا، در این صورت نگران مقادیر مرزی نیستیم. همچنین روش نیوتن-رافسون سریعتر همگرا خواد بود.

REFERENCES
Cox, JRSS B (1972).
Kalbfleisch and Prentice, JRSS B (1972), and
Peto, JRSS B (1972), in Discussion on Cox's paper.
Breslow, Biometrics (1974).
Prentice and Gloeckler, Biometrics (1978).

۶.۱ متغیّرهای وابسته به زمان

به حالتی که متغیّرها با زمان تغییر می‌کنند، تعمیم می‌دهیم. در نتیجه همراه با $Y_i = T_i \wedge C_i$ و $\delta_i = I(T_i \le C_i)$، مقادیر $\underline{x}_i(t)$ را مشاهده می‌کنیم. در الگوی نرخ شکست متناسب، فرض می‌شود، که تابع نرخ شکست iامین شاهد به شرح زیر باشد:

$$\lambda_i(t) = e^{\underline{\beta}' \underline{x}_i(t)} \lambda_0(t)$$

در نتیجه داریم:

$$P\{ y_{(i)} \text{ در زمان } (i) \text{ مرگ} \mid y_{(i)} \text{ در زمان } \Re_{(i)} \text{ یک مرگ در} \} = \frac{e^{\underline{\beta}' \underline{x}_{(i)}(y_{(i)})}}{\sum_{j \in \Re_{(i)}} e^{\underline{\beta}' \underline{x}_j(y_{(i)})}}$$

درست‌نمایی شرطی به صورت زیر در می‌آید:

$$L_c = \prod_u \frac{e^{\underline{\beta}' \underline{x}_{(i)}(y_{(i)})}}{\sum_{j \in \Re_{(i)}} e^{\underline{\beta}' \underline{x}_j(y_{(i)})}}$$

در این حالت که به زمان بستگی دارد، هیچ‌گونه اثباتی برای نرمال مجانبی وجود ندارد. همچنین، برای حجم نمونهٔ متوسّط یا بزرگ محاسبات کُند و پرهزینه خواهد بود.

۷.۱ مثال ۱: داده‌های پیوند قلب استانفورد. آیا بیماران قلبی که قلب خود را عوض می‌کنند از بیمارانی که چنین نمی‌کنند، عمر بیشتری دارند؟ نوعاً، یک بیمار، قلب خود را هنگامی عوض می‌کند که قلبی وجود داشته باشد. در این موقعیّت می‌گوییم که بیمار از جمعیت بدون تعویض به جمعیت تعویض شده‌ها منتقل شده است. متغیّر وابسته که انتقال را نشان می‌دهد از ۰ به ۱ تغییر می‌کند. متغیّرهای وابستهٔ دیگر شامل سن، زمان انتظار برای انتقال، زمان شروع معالجه و ... است.

REFERENCES
Turnbull, Brown, and Hu, JASA (1974).
Crowley and Hu, JASA (1977).

۸.۱ مثال ۲: فرزند خواندگی و آبستنی

آیا زوجهایی با دارا بودن مشکل نازایی و عقیمی، که بچّه‌ای را به فرزندی قبول کرده‌اند، بیشتر از زوجی که این عمل را انجام نداده‌اند، مایل به آبستنی هستند یا خیر؟ در این‌جا، امکان دارد زوجها از جامعهٔ فرزند خوانده‌دار منتقل شوند. برش هنگامی رخ می‌دهد که زوج کوشش خود را در آبستن شدن متوقّف کنند.

REFERENCES
Lamb and Leurgans, Amer. J. Obstet. Gyn. (1979).
Leurgans, Stanford Univ. Tech. Report No. 57 (1980).

۲ الگوی خطّی

الگوی خطّی استاندارد به صورت $T_i = \alpha + \beta x_i + e_i$ و یا به شکل زیر است:

$$T_i = \alpha + \underline{\beta}' \underline{x} + e_i \quad , \; i = ۱, \ldots, n$$

که در آن $e_۱$ تا e_n دارای خاصیّت iid با توزیع مشترک F هستند. فرض کنید $C_۱$ تا C_n مستقل باشند. C_i زمان برش مربوط به T_i است. مشاهدات به شرح زیرند:

$$Y_i = T_i \wedge C_i \qquad \text{و} \qquad \delta_i = I(T_i \le C_i)$$

الگوهای زمانی شتاب داده شده: الگوهای خطّی به وسیلهٔ الگوهای زمانی شتاب داده شده، به الگوهای خطّی شکست مربوط می‌شوند. فرض کنید $Z_۰$ **زمان بقاء با تابع نرخ شکست زیر باشد:**

$$\lambda_\circ = \frac{f_\circ(z)}{1 - F_\circ(z)}$$

همچنین، فرض می‌کنیم که زمان بقاء یک فرد با متغیّر وابستهٔ $\underline{x}$ با متغیّر زیر هم‌توزیع باشد.

$$Z_{\underline{x}} = e^{\underline{\beta}'\underline{x}} Z_\circ$$

توجّه شود که اگر $\underline{\beta}'\underline{x} > \circ$ باشد، آن‌گاه $Z_{\underline{x}}$ از $Z_\circ$ کوتاهتر بوده و می‌گوییم کـه متغیّر وابسته، زمان شکست را شتاب داده است. نرخ شکست $Z_{\underline{x}}$، به شرح زیر است:

$$\lambda_{\underline{x}}(z) = \frac{f_{\underline{x}}(z)}{1 - F_{\underline{x}}(z)} = \frac{f_\circ(e^{\underline{\beta}'\underline{x}} z) e^{\underline{\beta}'\underline{x}}}{1 - F_\circ(e^{\underline{\beta}'\underline{x}} z)} = \lambda_\circ(e^{\underline{\beta}'\underline{x}} z) e^{\underline{\beta}'\underline{x}}$$

با تعریف $T_{\underline{x}} = \log Z_{\underline{x}}$، داریم:

$$E(T_{\underline{x}}) = \underline{\beta}'\underline{x} + E(\log Z_\circ) = \underline{\beta}'\underline{x} + \alpha$$

در نتیجه الگوی زمانی شتاب داده شده با یک الگوی لگاریتمی خطّی، به شرح زیر برابر است:

$$T_{\underline{x}} = \alpha + \underline{\beta}'\underline{x} + e$$

به‌گونه‌ای که $e = \log Z_\circ - E(\log Z_\circ)$ است. در کاربرد روشهای الگوی خطّی برای داده‌های بقاء، اغلب لازم است که داده‌ها را توسّط لگاریتم‌گیری تبدیل کنیم تا متقارن شوند. در نتیجه از این دیدگاه، الگوی زمان شتاب داده شده مناسب است.

REFERENCES

Prentice and Kalbfleisch, Biometrics (1979), and

Kalbfleisch and Prentice, The Statistical Analysis of Failure Time Data (1980), both discuss the accelerated time model.

۱.۲ آزمونهای رتبهٔ خطّی

اگر برش نباشد، تواناترین آمارهٔ رتبه برای آزمون $H_\circ: \beta = \circ$ در مقابل $H_\circ: \beta \neq \circ$، در حالت $p = 1$ به صورت: $\frac{d}{d\beta} \log p(\underline{r})\Big|_{\beta=\circ}$ است، که $p(\underline{r})$ احتمال بردار رتبهٔ $\underline{r}$ است.

همچنین، در صورت وجود برش می‌توان از آمارهٔ زیر استفاده کرد:

$$\frac{d}{d\beta}\log p(\underline{r}^{u/c},\underline{\delta})\Big|_{\beta=0}$$

که با توجّه به بخش اوّل، داریم:

$$p(\underline{r}^{u/c},\underline{\delta})=\int\cdots\int_{u_1<\cdots<u_{n_u}}\prod_{i=1}^{n_u}\left\{f_{u(i)}(u_i)\times\prod_{j\in C_{i,i+1}}[1-F_j(u_i)]\right\}du_1\cdots du_{n_u}$$

$$f_i(u)=f(u-\beta x_i)$$

می‌توان نشان داد که:

$$\frac{d}{d\beta}\log p(\underline{r}^{u/c},\underline{\delta})\Big|_{\beta=0}=\sum_{i=1}^{n_u}\left\{x_{u(i)}c_i+\left(\sum_{j\in C_{i,i+1}}x_j\right)C_i\right\}$$

که در این رابطه، داریم:

$$c_i=\left(\prod_{j=1}^{n_u}n_{u(j)}\right)\int\cdots\int_{u_1<\cdots<u_{n_u}}\left\{-\frac{d}{du_i}\log f(u_i)\right\}$$

$$\times\prod_{j=1}^{n_u}\left\{f(u_j)[1-F(u_j)]^{m_{u(j)}}\right\}du_1\cdots du_{n_u}$$

$$C_i=\left(\prod_{j=1}^{n_u}n_{u(j)}\right)\int\cdots\int_{u_1<\cdots<u_{n_u}}\left\{-\frac{d}{du_i}\log[1-F(u_i)]\right\}$$

$$\times\prod_{j=1}^{n_u}\left\{f(u_j)[1-F(u_j)]^{m_{u(j)}}\right\}du_1\cdots du_{n_u}$$

$$m_{u(j)}=\#\text{ in }C_{j,j+1}$$

فرض کنید توزیع خطا، توزیع مقادیر فرین باشد. که تابع چگالی و بقاء آن به شرح زیر است:

$$f(t)=e^{t-e^{t}} \qquad \text{و} \qquad ۱-F(t)=e^{-e^{t}}$$

حال مقدار C_i، به صورت زیر خواهد بود:

$$c_i=\sum_{j=۱}^{i}\frac{۱}{n_{u(j)}}-۱$$

$$C_i=\sum_{j=۱}^{i}\frac{۱}{n_{u(j)}}$$

در نتیجه تواناترین آمارهٔ رتبه در این حالت به صورت: $-\sum_{u}(x_{(i)}-\bar{x}_{(i)})$، است. این عبارت برابر مخرج آمارهٔ کاکس در آزمون $H_۰:\beta=۰$ است (بخش اوّل) و آن را آزمون رتبهٔ لگاریتمی می‌نامند.

REFERENCES

Peto and Peto, JRSS A (1972), introduce linear rank tests and coin the term "log rank test".

Latta, Biometrika (1977), establishes a connection between linear rank tests and Efron's test.

Morton, Biometrika (1978), discusses permutation theory for linear rank tests.

Prentice, Biometrika (1978), give the preceding derivation of the linear rank test statistic and calculates its variance.

Kalbfleisch and Prentice, The Statistical Analysis of Failure Time Data (1980), Chapter 6.

۲.۲ برآوردگرهای کمترین مربعات

برای سادگی فرض کنید $p=۱$ باشد، یعنی: $E(T_i)=\alpha+\beta x_i$. این برآوردگر را می‌توان به بیش از یک متغیّر وابسته تعمیم داد.

برآوردگرهای میلر. برای مشاهدات بریده نشده، برآوردهای $\hat{\alpha}$ و $\hat{\beta}$، عبارت زیر را کمینه می‌سازند:

$$\sum_{i=۱}^{n}(y_i-\alpha-\beta x_i)^۲=n\int_{-\infty}^{+\infty}z^۲\,dF_n(z)$$

به طوری که F_n توزیع تجربی $z_۱$ تا z_n است و $z_i=y_i-\alpha+\beta x_i$.

در صورت وجود برش، میلر پیشنهاد می‌کند که عبارت زیر را کمینه سازیم:

$$n\int_{-\infty}^{+\infty} z^{2}\, d\hat{F}(z)=\sum_{i=1}^{n}\hat{w}_i(\beta)(y_i-\alpha-\beta x_i)^{2}$$

که در آن $\hat{F}$ برآوردگر PL برآوردگر مبنای (z_1,δ_1)، ...، (z_n,δ_n) است و وزنها $\hat{w}_1(\beta)$ تا $\hat{w}_n(\beta)$، جهشهای برآوردگر PL هستند. در نگاه اول می‌توان نتیجه گرفت که مجموع موزون مربعها به مشاهدات بریده شده بستگی ندارند. با این وجود که در اصل، برآوردگر PL و در نتیجه هر کوواریانس از وزنها به مشاهدات بریده شده وابسته‌اند.

اگر آخرین مشاهده بریده شده باشد و $\delta_{(n)}=0$، آن را بریده نشده تغییر می‌دهیم. در این صورت $\sum_1^n \hat{w}_i(\beta)=1$ است.

وزنها را فقط به صورت توابعی از β نوشته‌ایم. زیرا، اضافه کردن ثابتی مانند α به هر T_i، نتیجه‌اش فقط یک انتقال برآوردگر PL و جهشهای آن است و در نتیجه وزنها به α بستگی ندارند.

برای محاسبهٔ $\hat{\alpha}$ و $\hat{\beta}$، نسبت به α مشتق می‌گیریم، داریم:

$$\hat{\alpha}=\sum_{i=1}^{n}\hat{w}_i(\beta)y_i-\beta\sum_{i=1}^{n}\hat{w}_i(\beta)x_i$$

اگر این مقدار را در مجموع موزون مربعات قرار دهیم، فقط تابعی از β به دست می‌آید.

$$f(\beta)=\sum \hat{w}_i(\beta)(y_i-\hat{\alpha}-\beta x_i)^{2}$$

که با تجسس می‌توان آن را کمینه کرد.

چون تابع $f(\beta)$ پیوسته نیست، ممکن است روش تجسس خسته کننده باشد، به‌ویژه در ابعاد بالا. میلر روش بهتر زیر را به عنوان راه‌حل دیگری پیشنهاد نموده است. در این روش، برآورد اوّلیّه را به صورت زیر تعریف می‌کنیم، که $\hat{\beta}^{\circ}$ همان شیب خط کمترین مربعات مشاهدات بریده نشده است.

$$\hat{\beta}^{\circ}=\frac{\sum_u y_i(x_i-\bar{x}_u)}{\sum_u (x_i-\bar{x}_u)^{2}}$$

با این حدس اوّلیّه $\hat{\beta}^{\circ}$، داریم: $\hat{z}_i^{\circ}=y_i-\hat{\beta}^{\circ}x_i$، $i=1,\dots,n$. فرض کنید $\hat{F}^{\circ}$، برآوردگر PL

بر مبنای $(\hat{z}_1^\circ, \delta_1)$ تا $(\hat{z}_n^\circ, \delta_n)$ و $\hat{w}_1(\hat{\beta}^\circ)$ تا $\hat{w}_n(\hat{\beta}^\circ)$، جهشهای $\hat{F}^\circ$ باشند. برآورد جدید به صورت زیر تعریف می‌شود:

$$\hat{\beta}^1 = \frac{\sum_u \hat{w}_i^*(\hat{\beta}^\circ) y_i (x_i - \bar{x}_u^*)}{\sum_u \hat{w}_i^*(\hat{\beta}^\circ)(x_i - \bar{x}_u^*)^2}$$

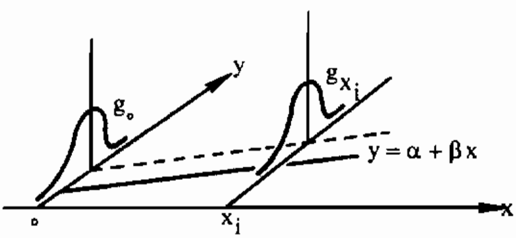
که در این رابطه $\hat{w}_i^*(\hat{\beta}^\circ) = \frac{\hat{w}_i(\hat{\beta}^\circ)}{\sum_u \hat{w}_i(\hat{\beta}^\circ)}$ و $\bar{x}_u^* = \sum_u \hat{w}_i^*(\hat{\beta}^\circ) x_i$.

با دوباره نرمال کردن وزنهای $\hat{w}_i^*(\hat{\beta}^\circ)$، به ما اجازه می‌دهد که از آخرین جملهٔ مرتّب شده $\hat{z}_i^\circ$ -در هر دو حالت بریده شده و نشده- چشم‌پوشی کنیم. تنها مشاهدات برید نشده در مجموع ظاهر می‌شوند. روش معمول تعریف دوبارهٔ آخرین جملهٔ مرتّب شده $\hat{z}_i^\circ$ -به‌گونه‌ای که بریده شود، اگر بریده شده است- نتایج کمتر پایداری را در برآورد تکراری β ارائه می‌کند. ولی هنوز می‌توان آن را برای برآورد α به کار برد.

روش بالا را تکرار می‌کنیم با این امید که همگرا شود. با این وجود، امکان دارد دنبالهٔ برآوردگرهای β در یک دور که بین دو مقدار نوسان می‌کند، قرار بگیرد. در این حالت متوسّط دو مقدار را انتخاب می‌کنیم.

فرض نمایید که تغییرپذیری مربوط به وزنهای $\hat{w}_i^*(\hat{\beta})$ قابل حذف باشد. داریم:

$$\hat{\mathrm{Var}}(\hat{\beta}) = \frac{\sum_u \hat{w}_i^*(\hat{\beta}^\circ)(y_i - \hat{\alpha} - \hat{\beta} x_i)}{\sum_u \hat{w}_i^*(\hat{\beta}^\circ)(x_i - \bar{x}_u^*)^2}$$



برای محاسبهٔ برآورد واریانس بالا، مانند اثبات سازگاری برآوردهای $\hat{\alpha}$ و $\hat{\beta}$، به این فرض بستگی دارد که توزیع بریده شدهٔ مشاهدهٔ iام به صورت: $G_{x_i}(c) = G_\circ(c - \beta x_i)$،

باشد. برای تابع توزیعی مانند $G_\circ$، اگر چگالی آن ($g_\circ$) مانند شکل زیر باشد، آن‌گاه G_{x_i} دارای چگالی g_{x_i} خواهد بود. در واقع g_{x_i}، انتقال یافتهٔ $g_\circ$ به اندازهٔ βx_i است.

REFERENCE
Miller, Biometrika (1976).

برآوردگر باکلی-جیمز. در این الگو فرض می‌شود: $E(T_i)=\alpha+\beta x_i$. متأسفانه، نمی‌توان T_i را مشاهده کنیم. ولی، فقط Y_i قابل مشاهده است، همچنین $E(Y_i)\neq\alpha+\beta x_i$. در نتیجه باکلی و جیمز متغیّرهای کاذب زیر را تعریف می‌کنند:

$$Y_i^* = Y_i\delta_i + E(T_i \mid T_i > Y_i)(1-\delta_i)$$

محاسبهٔ $E(Y_i^*)$ به شرح زیر است:

$$\begin{aligned}
E(Y_i^*) &= \int_\circ^\infty u(1-G_i(u))\,dF_i(u) + \int_\circ^\infty \left[\int_u^\infty \frac{s\,dF_i(s)}{1-F_i(u)}\right](1-F_i(u))\,dG_i(u)\\
&= \int_\circ^\infty u(1-G_i(u))\,dF_i(u) + \int_\circ^\infty \left[\int_\circ^s dG_i(u)\right] s\,dF_i(s)\\
&= \int_\circ^\infty u(1-G_i(u))\,dF_i(u) + \int_\circ^\infty G_i(s)s\,dF_i(s)\\
&= \int_\circ^\infty u\,dF_i(u)\\
&= \alpha+\beta x_i
\end{aligned}$$

بنابراین، اگر بتوانیم y_1^* تا y_n^* را مشاهده کنیم، معقول به نظر می‌رسد که از برآوردگرهای زیر استفاده کنیم:

$$\hat{\alpha} = \bar{y}^* - \hat{\beta}\bar{x} \qquad \text{و} \qquad \hat{\beta} = \frac{\sum_{i=1}^{n} y_i^*(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{۱۵}$$

چون نمی‌توانیم تمام y_1^* تا y_n^* را مشاهده کنیم، لذا، از برآوردگرهای مشاهدات گم شده

استفاده می‌کنیم. اگر $\delta_i = ۰$ باشد، تعریف زیر را در نظر می‌گیریم:

$$\hat{E}(T_i \mid T_i > y_i) = \hat{\beta} x_i + \frac{\sum_{\hat{z}_k > \hat{z}_i} \hat{w}_k(\hat{\beta}) \hat{z}_k}{۱ - \hat{F}(\hat{z}_i)} \tag{۱۶}$$

که در این رابطه $\hat{z}_i = y_i - \hat{\beta} x_i$ و $\hat{F}$ برآوردگر PL بر مبنای $(\hat{z}_۱, \delta_۱)$ تا $(\hat{z}_n, \delta_n)$ است. $\hat{w}_۱(\hat{\beta})$ تا $\hat{w}_n(\hat{\beta})$، جهشهای $\hat{F}$ است. در این صورت تعریف زیر را داریم:

$$\hat{y}_i^* = y_i \delta_i + \left[\hat{\beta} x_i + \frac{\sum_{\hat{z}_k > \hat{z}_i} \hat{w}_k(\hat{\beta}) \hat{z}_k}{۱ - \hat{F}(\hat{z}_i)} \right] (۱ - \delta_i) \tag{۱۷}$$

با توجّه به این که معادلات (۱۵)، $\hat{\beta}$ را به صورت تابعی از y_i^* و معادله (۱۷)، y_i^* را به صورت تابعی از $\hat{\beta}$ می‌دهد، لذا، باید روش تکراری را برای محاسبهٔ آنها به کار ببریم. مانند برآورد میلر، امکان دارد دنبالهٔ برآوردهای β بین دو مقدار نوسان کند، در این حالت نیز مقدار متوسّط این دو را انتخاب می‌کنیم.

باکلی و جیمز مدّعی هستند که اگر برآوردگرهای β نوسان کنند، آن‌گاه تفاضل بین این دو مقدار، کوچکتر از برآورد میلر است. علاوه بر این اعتبار روش آنها به فرضهای توزیع بریده شدهٔ G_i بستگی ندارد.

باکلی و جیمز، برآورد واریانس را به صورت زیر ارائه می‌کنند:

$$\hat{Var}(\hat{\beta}) = \frac{\hat{\sigma}_u^۲}{\sum_u (x_i - \bar{x}_u)^۲}$$

که در آن $\hat{\sigma}_u^۲$ به صورت زیر است:

$$\hat{\sigma}_u^۲ = \frac{۱}{n_u - ۲} \sum_u (y_i - \bar{y}_u - \hat{\beta}(x_i - \bar{x}_u))^۲$$

در این جا اثبات آن ارائه نمی‌شود.

REFERENCE

Buckley and James, Biometrika (1979).

تذکّرها:

(الف) روش باکلی و جیمز یک روش ناپارامتری و مشابه روش نرمال اشمی و هان است. با توجّه به تعریف $W_i = \frac{T_i - \alpha - \beta x_i}{\sigma}$، اگر F نرمال باشد. آن‌گاه داریم:

$$E(T_i \mid T_i > y_i) = E\left(\sigma W_i + \alpha + \beta x_i \mid W_i > \frac{y_i - \alpha - \beta x_i}{\sigma}\right)$$

$$= \alpha + \beta x_i + \frac{\sigma \int_{(y_i-\alpha-\beta x_i)/\sigma}^{\infty} w\,\phi(w)\,dw}{1-\Phi\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)} = \alpha + \beta x_i + \frac{\sigma\,\phi\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)}{1-\Phi\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)}$$

در این رابطه ϕ و Φ به ترتیب تابعهای چگالی و توزیع نرمال استاندارد است. اشمی و هان به جای (۱۶)، برآورد زیر را به کار برده‌اند.

$$\hat{E}(T_i \mid T_i > y_i) = \hat{\alpha} + \hat{\beta} x_i + \frac{\hat{\sigma}\,\phi\left(\frac{y_i - \hat{\alpha} - \hat{\beta} x_i}{\sigma}\right)}{1-\Phi\left(\frac{y_i - \hat{\alpha} - \hat{\beta} x_i}{\hat{\sigma}}\right)}$$

REFERENCE

Schmee and Hahn, Technometrics (1979).

(ب) هر دو روش پارامتری و ناپارامتری، شبیه الگوریتم EM در نظریّهٔ حداکثر درست‌نمایی‌اند.

REFERENCE

Dempster, Laird, and Rubin, JRSS B (1977).

برآوردگر کول-سوزارلا-وان رایزین

تعریف $Y_i^* = \frac{\delta_i Y_i}{1 - G(Y_i)}$ را در نظر می‌گیریم، داریم:

$$E(Y_i^*) = \int_0^{\infty} \frac{u}{1-G(u)}(1-G(u))\,dF_i(u) = \int_0^{\infty} u\,dF_i(u) = \alpha + \beta x_i$$

بنابراین، اگر بتوانیم y_1^* تا y_n^* را مشاهده کنیم، می‌توانیم بــه طـور معمــول α و β را بــه کمک (۱۵) برآورد کنیم. متأسّفانه، نمی‌توان همـهٔ y_1^* تـا y_n^* را مشـاهده کنیـم، ولـی، می‌توانیم برآوردها را جانشین نماییم. برای این کار، به جای G از برآوردگر PL استفاده شود، که در آن نقش متغیّرهای بقاء و بریده شـده عـوض می‌شـود. ایـن پژوهشـگران پیشنهاد می‌کنند که از برآوردگر بیز G استفاده شود.

امتیاز بزرگ این روش بی‌نیازی از تکرار است. همچنین بیشـتر بـر مبنـای یـک توزیع بریده شده عمل می‌کند تا با توزیعهای برشی انتقال یافته‌ای، مانند برآوردگر میلر. با این وجود، y_i^*ها مقادیر خاصّی هستند. اینها یا صفرانــد یـا مقـادیر y_iهـا را افزایـش می‌دهند. رفتار این برآوردگرها در این جا ارزشیابی نمی‌شود.

REFERENCE

Koul, Susarla, and Van Ryzin, unpublished manuscript (1979).

مثال. داده‌های پیوند قلب استانفورد: روشهای کاکس-میلر و باکلی-جیمـز را بـه کمک داده‌های جـدول (۵)، مقایســه می‌کنیـم. در رگرسـیون اوّلـی (شـکل ۵) متغیّـر وابسته: $\log_{10}$(زمان بقاء) است، که در آن زمان بقاء تا زمان مرگ بــه علّـت نپذیرفتـن قلب اهدایی و متغیّر کمکی امتیاز مربوط بــه طـرز عمـل اسـت. در رگرسـیون دومـی (شکل ۶)، متغیّر وابسته: $\log_{10}$(زمان بقاء) است، که در آن زمان بقاء تـا لحظـهٔ مـرگ -صرفنظر از این‌که به علّت نپذیرفتن قلب اهدایی یا علل دیگر باشد- است. متغیّر کمکی سن است. اگر زمان بقاء صفر باشد آن را به ۱ تغییر می‌دهیم تا لگاریتم آن قابل محاسـبه باشد. مقایسه‌های هر سه روش در جدولهای شماره ۶ و ۷، ارائه شده‌اند.

هر سه روش، نتایج پیچیده‌ای را در رگرسـیون روی سـن نشـان می‌دهنـد: روش کاکس، نشانگر این است که اثر سن بسیار معنی‌دار است. روش میلر مدّعی است که سن اثری ندارد و روش باکلی و جمیز، ادّعای وسطی را بیان می‌کند. در این حالت می‌توان به کمک الگوی برش، برآوردگرهای میلر را کنار گذاشت.

همچنین دربارهٔ درجهٔ اعتبار اثر ناهمخوانی امتیاز عــدم توافـق وجـود دارد. بـاید تلاشهای بیشتری را در مورد این‌که کدام الگو (زمانی شتاب داده شده یا نــرخ شکسـت متناسب) برای این داده‌ها مناسبتراند، انجام داد.

REFERENCES

Millre, Biometrika (1976).

Buckley and James, Biometrika (1979).

جدول ۵. داده‌های پیوند قلب استانفورد

روزهای تحمل	زمان بقاء	۱=مرده ۰=زنده	۱=نپذیرفتن ۰=پذیرفتن	مقادیر نابرابر T5	سن در زمان Tx
۳	۱۵	۱	۰	۱٫۱۱	۵۴٫۳
۴	۳	۱	۰	۱٫۶۶	۴۰٫۴
۷	۶۲۴	۱	۱	۱٫۳۲	۵۱٫۰
۱۰	۴۶	۱	۱	۰٫۶۱	۴۲٫۵
۱۱	۱۲۷	۱	۰	۰٫۳۶	۴۸٫۰
۱۳	۶۴	۱	۱	۱٫۸۹	۵۴٫۶
۱۴	۱۳۵۰	۱	۱	۰٫۸۷	۵۴٫۱
۱۶	۲۸۰	۱	۱	۱٫۱۲	۴۹٫۵
۱۸	۲۳	۱	۰	۲٫۰۵	۵۶٫۹
۲۰	۱۰	۱	۱	۲٫۷۶	۵۵٫۳
۲۱	۱۰۲۴	۱	۱	۱٫۱۳	۴۳٫۴
۲۲	۳۹	۱	۱	۱٫۳۸	۴۲٫۸
۲۳	۷۳۰	۱	۱	۰٫۹۶	۵۸٫۴
۲۴	۱۳۶	۱	۱	۱٫۶۲	۵۲٫۰
۲۵	۱۷۷۵	۰	۰	۱٫۰۶	۳۳٫۳
۲۸	۱	۱	۰	۰٫۴۷	۵۴٫۲
۳۰	۸۳۶	۱	۱	۱٫۵۸	۴۵٫۰
۳۲	۶۰	۱	۱	۰٫۶۹	۶۴٫۵
۳۳	۱۵۳۶	۰	۰	۰٫۹۱	۴۹٫۰
۳۴	۱۵۴۹	۰	۰	۰٫۳۸	۴۰٫۶
۳۶	۵۴	۱	۱	۲٫۰۹	۴۹٫۰
۳۷	۴۷	۱	۱	۰٫۸۷	۶۱٫۵
۳۸	۰	۱	۰	۰٫۸۷	۴۱٫۵
۳۹	۵۱	۱			۵۰٫۵
۴۰	۱۳۶۷	۰	۰	۰٫۷۵	۴۸٫۶

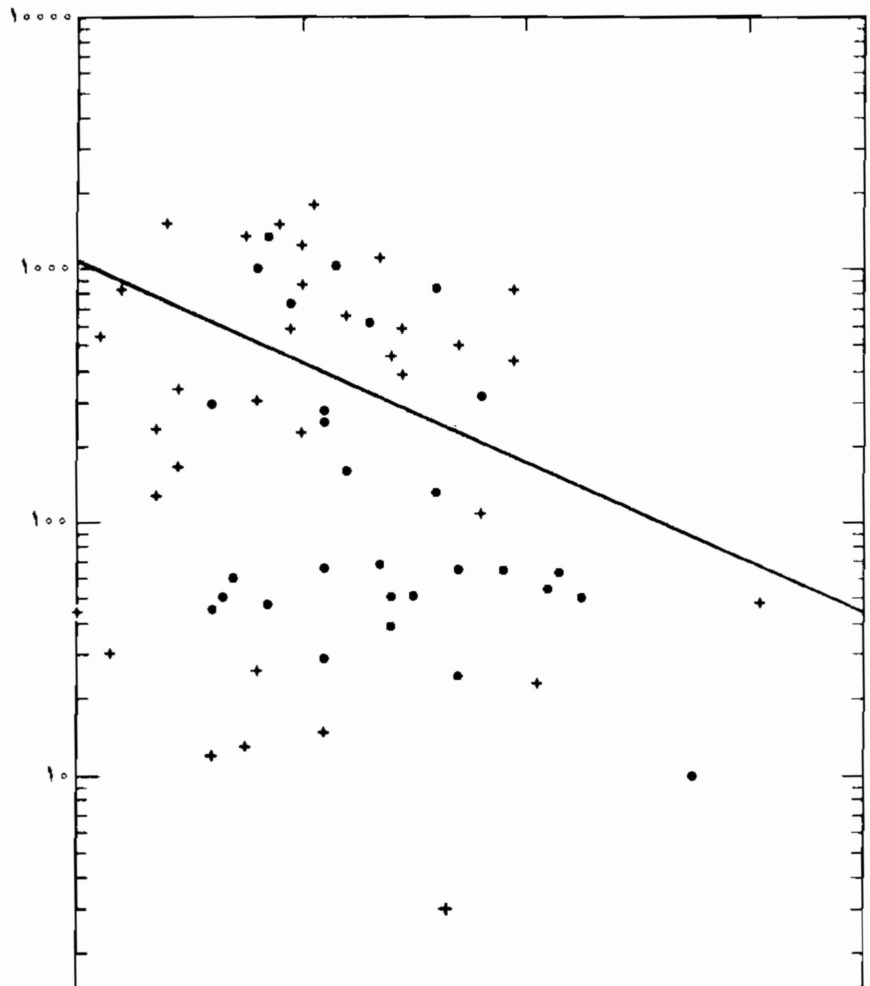
ادامهٔ جدول ۵.

سن در زمان Tx	مقادیر نابرابر T۵	۱=نپذیرفتن ۰=پذیرفتن	۱=مرده ۰=زنده	زمان بقاء	روزهای تحمل
۴۵/۵	۰/۹۸	۰	۰	۱۲۶۴	۴۱
۳۶/۲	۰/۰	۰	۱	۴۴	۴۵
۴۸/۶	۰/۸۱	۱	۱	۹۹۴	۴۶
۴۷/۲	۱/۳۸	۱	۱	۵۱	۴۷
۳۶/۸	۱/۳۵	۰	۰	۱۱۰۶	۴۹
۴۶/۱			۱	۸۹۷	۵۰
۴۸/۸	۱/۰۸	۱	۱	۲۵۳	۵۱
۴۷/۵			۱	۱۴۷	۵۳
۵۲/۵	۱/۵۱	۱	۱	۵۱	۵۵
۳۸/۹	۰/۹۸	۰	۰	۸۷۵	۵۶
۴۸/۱	۱/۸۲	۱	۱	۳۲۲	۵۸
۴۱/۶	۰/۱۹	۰	۰	۸۳۸	۵۹
۴۹/۱	۰/۶۶	۱	۱	۶۵	۶۰
۳۲/۷	۱/۹۳	۰	۰	۸۱۵	۶۳
۴۸/۹	۰/۱۲	۰	۱	۵۵۱	۶۴
۵۱/۳	۱/۱۲	۱	۱	۶۶	۶۵
۱۹/۷	۱/۰۲	۰	۱	۲۲۸	۶۷
۴۵/۲	۱/۶۸	۱	۱	۶۵	۶۸
۴۸/۰	۱/۲۰	۰	۰	۶۶۰	۶۹
۵۳/۰	۱/۶۸	۱	۱	۲۵	۷۰
۴۷/۵	۰/۹۷	۰	۰	۵۸۹	۷۱
۲۶/۷	۱/۴۶	۰	۰	۵۹۲	۷۲
۵۶/۴	۲/۱۶	۱	۱	۶۳	۷۳
۲۹/۲	۰/۶۱	۰	۱	۱۲	۷۴
۵۲/۲	۱/۷۰	۰	۰	۴۹۹	۷۶

ادامهٔ جدول ۵.

سن در زمان Tx	مقادیر نابرابر T5	۱=نپذیرفتن ۰=پذیرفتن	۱=مرده ۰=زنده	زمان بقاء	روزهای تحمل
۴۹٫۳	۰٫۸۱	۰	۰	۳۰۵	۷۸
۵۴٫۰	۱٫۰۸	۱	۱	۲۹	۷۹
۴۶٫۵	۱٫۴۱	۰	۰	۴۵۶	۸۰
۵۲٫۹	۱٫۹۴	۰	۰	۴۳۹	۸۱
۵۳٫۴	۳٫۰۵	۰	۱	۴۸	۸۳
۴۲٫۸	۰٫۶۰	۱	۱	۲۹۷	۸۴
۴۸٫۹	۱٫۴۴	۰	۰	۳۸۹	۸۶
۴۶٫۴	۲٫۲۵	۱	۱	۵۰	۸۷
۵۴٫۴	۰٫۶۸	۰	۰	۳۳۹	۸۸
۵۱٫۴	۱٫۳۳	۱	۱	۶۸	۸۹
۵۲٫۵	۰٫۸۲	۰	۱	۲۶	۹۰
۴۵٫۸	۰٫۱۶	۰	۰	۳۰	۹۲
۴۷٫۸	۰٫۳۳	۰	۰	۲۳۷	۹۳
۴۳٫۸	۱٫۲۰	۱	۱	۱۶۱	۹۴
۴۰٫۳			۱	۱۴	۹۵
۲۶٫۷	۰٫۴۶	۰	۰	۱۶۷	۹۶
۲۳٫۷	۱٫۷۸	۰	۰	۱۱۰	۹۷
۲۸٫۹	۰٫۷۷	۰	۰	۱۳	۹۸
۳۵٫۲	۰٫۶۷	۰	۰	۱	۱۰۰

نمودار ۵. بقاء در مقابل امتیازهای ناهم‌خوان T_5.
"+" = زنده یا مرگ رد شده
"•" = مرگ رد شده
"−" = خط کمترین مربعات کاپلان-مایر

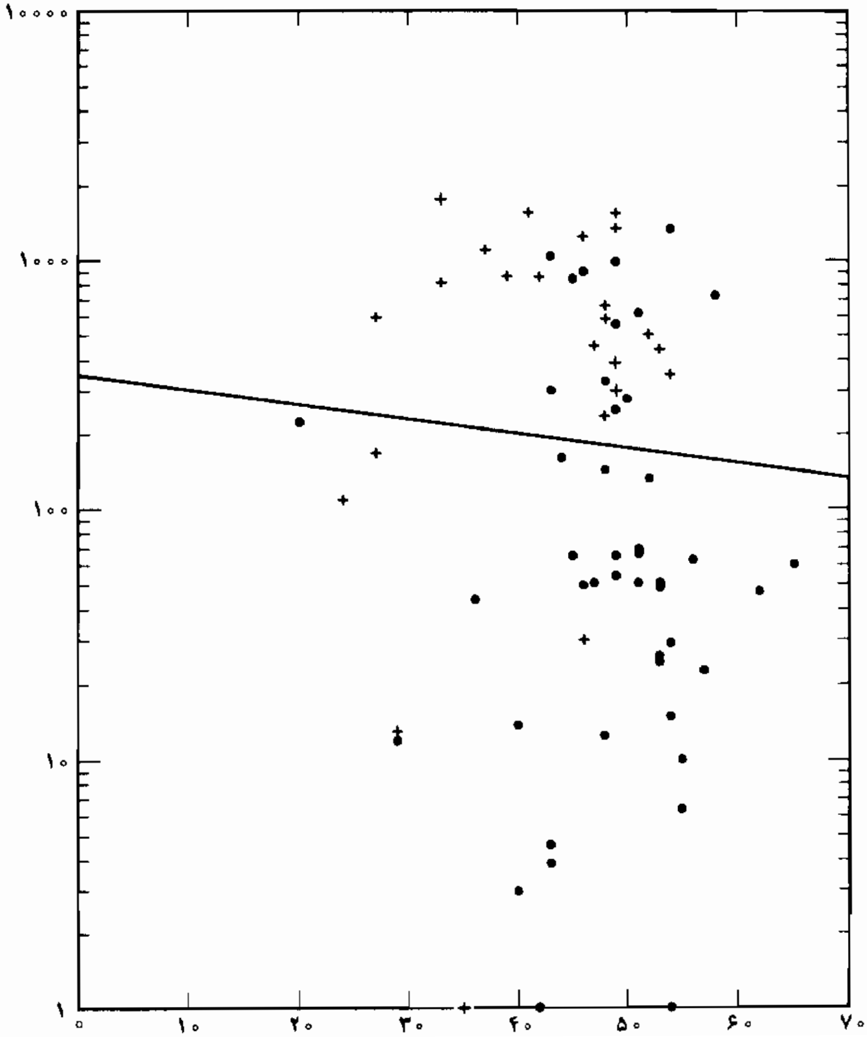


نمودار ۶. بقاء در مقابل سن.

"+" = زنده

"•" = مرده

"–" = خط کمترین مربعات کاپلان-مایر



جدول ۶. رگرسیون لگاریتم زمان بقاء بر امتیاز ناهم‌خوان

روش	$\hat{\alpha}$	$\hat{\beta}$	$\hat{SD}(\hat{\beta})$
کاکس	—	۱٫۰۷۶	۰٫۳۶۸
میلر	۳٫۰۳۶	−۰٫۳۹۴	—
تعمیم میلر	۳٫۱۲۰	−۰٫۴۵۲	۰٫۲۳۶
	۳٫۱۴۵	−۰٫۴۷۱	۰٫۲۳۴
باکلی-جیمز	—	—	—

جدول ۷. رگرسیون لگاریتم زمان بقاء بر سن

روش	$\hat{\alpha}$	$\hat{\beta}$	$\hat{SD}(\hat{\beta})$
کاکس	—	۰٫۰۵۷۵	۰٫۰۲۳۳
میلر	۲٫۵۳۷	−۰٫۰۰۵۸	—
تعمیم میلر	۲٫۱۱۱	۰٫۰۰۳۶	۰٫۰۱۶۶
	۲٫۱۷۱	۰٫۰۰۲۴	۰٫۰۱۶۳
باکلی-جیمز	۳٫۵۸۲	−۰٫۰۲۷۸	۰٫۰۱۴۹

فصل هفتم

نیکویی برازش

۱ روشهای ترسیمی

به راحتی می‌توان با نگاه کردن، خط را از منحنی تمیز داد. بنابراین برای روش رسم نمودار از اصل زیر استفاده می‌کنیم:

اصل اساسی. مقیاس محورهای مختصّات را به‌گونه‌ای اختیار می‌کنیم، که اگر الگو برقرار باشد، نمودار به شکل خط مستقیم و در غیر این صورت به شکل منحنی در آید.

معمولاً، دو نوع نمودار بقاء و نرخ شکست مورد استفاده قرار می‌گیرند، این دو مورد بسیار نزدیک هم‌اند. و در هر حالت مناسبترین انتخاب می‌شود.

(الف) نمودارهای بقاء

در این نمودارها یا $\hat{S}(y_{u(i)})$ را در مقابل $y_{u(i)}$ و یا $\hat{S}(t)$ را در مقابل t رسم می‌کنند. این نوع، حالت خاصی از نمودار $Q-Q$ یا نمودارهای احتمالی است.

REFERENCE

Wilk and Gnanadesikan, Biometrika (1968).

(ب) نمودارهای نرخ شکست

در این نمودارها یا $\hat{\Lambda}(y_{u(i)})$ در مقابل $y_{u(i)}$ و یا $\hat{\Lambda}(t)$ در مقابل t رسم می‌شود. برای این کار از رابطهٔ نلسون (فصل سوم، بخش سوم را ببینید) به صورت:

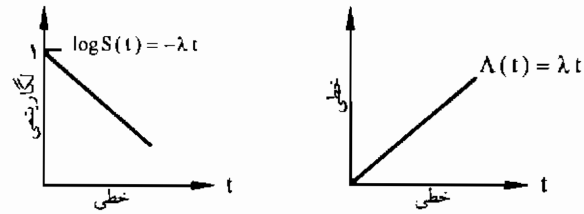
$\hat{\Lambda}_2(t) = \sum_{y_{(i)} \le t} \frac{\delta_{(i)}}{n-i+1}$ و یا رابطهٔ: $\hat{\Lambda}_1(t) = -\log\hat{S}(t)$، استفاده می‌شود.

REFERENCES
Nelson, J. Qual. Tech. (1969).
______, Technometrics (1972).

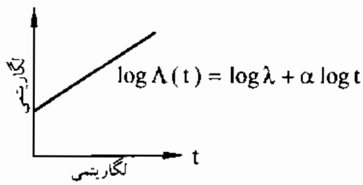
۱.۱ یک نمونه

نمودارهای چند توزیع در زیر رسم شده‌اند:

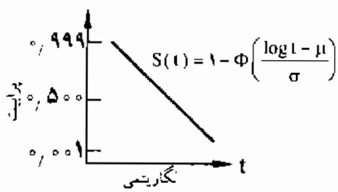
(الف) نمایی



(ب) وایبل



(ج) لوگ نرمال



(د) گاما و سایر توزیعها

بدون استفاده از کاغذهای نموداری، کمّیتها بر مبنای فرضهای پارامتری در مقابل کمّیتهای مبتنی بر برآوردگر PL، رسم می‌شود.

REFERENCES
Wilk, Gnanadesikan, and Huyett, Technometrics (1962), for the gamma distribution without censoring.

۲.۱ دو نمونه تا K نمونه

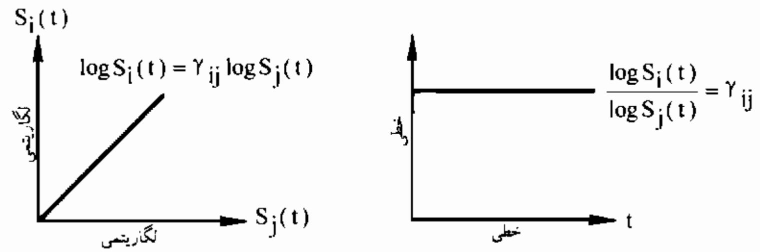
برای الگوهای پارامتری، موارد (الف) تا (د) را روی هر نمونه تکرار می‌کنیم. فرض کنید می‌خواهیم اعتبار الگوی نرخ شکست متناسب کاکس را بررسی کنیم. تحت الگوی $S_i(t)=S_j(t)^{\gamma_{ij}}$ برای چند مقدار از γ_{ij}، داریم:

$$\log S_i(t)=\gamma_{ij}\log S_j(t)$$

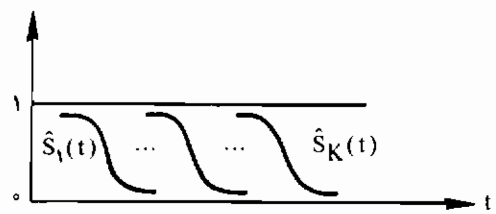
یا

$$\frac{\log S_i(t)}{\log S_j(t)}=\gamma_{ij}$$

برآوردهای جدای PL هر یک از $\hat{S}_1(t)$ تا $\hat{S}_K(t)$ را محاسبه و یکی از نمودارهای زیر را می‌سازیم:



برای وارسی الگوی خطّی، برآورد PL را به طور جداگانه برای هر یک از K نمونه محاسبه و آن را رسم می‌کنیم. تغییر مکانها را توسّط انتقال وارسی می‌کنیم.



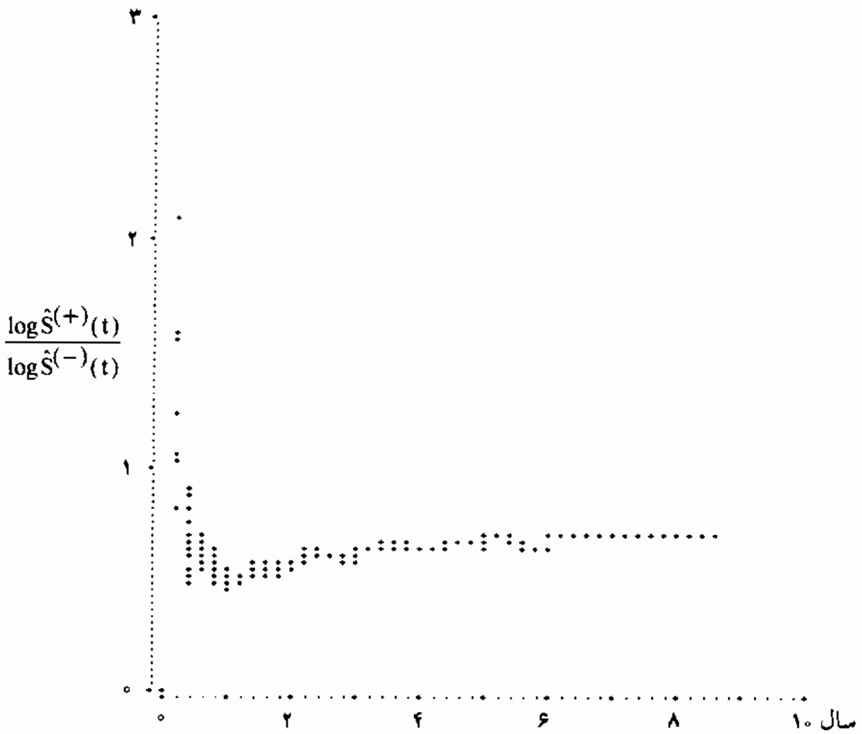
مثال: مطالعه DNCB. بیماران مرض هادکین، مورد حساسیّت قرار گرفته و سپس به طور مستمر در معرض شیمی درمانی دینی‌تروکلروبنزن (DNCB) قرار گرفته‌اند. جامعهٔ (+) شامل آن بیمارانی است که واکنش مثبت به DNCB نشان داده و جامعهٔ (−) شامل آن بیمارانی است که واکنش نشان نداده‌اند. بیماران می‌توانند در میان جمعیتها نقل

مکان کنند، زمان بقاء را تا زمان جایگذاری در نظر می گیریم.

آیا بیماران جامعهٔ (+) بیش از بیماران جامعهٔ (−) عمر می کنند. الگوی نرخ شکست متناسب کاکس به کار می‌رود. رسم $\log\hat{S}^{(+)}/\log\hat{S}^{(-)}$ در نمودار شماره ۷، نشان می‌دهد که برای زمان t نزدیک به صفر، نسبت لگاریتم به طور مستدلّی ثابت است، که نشان دهندهٔ برقراری الگوست.

REFERENCE

Gong, Stanford Univ. Tech. Report No. 57 (1980).



نمودار ۷. نمایش $\log\hat{S}^{(+)}(t)/\log\hat{S}^{(-)}(t)$ توسّط رایانه

۳.۱ رگرسیون

فرض کنید می‌خواهیم الگوی نرخ شکست متناسب را بررسی نماییم. در حالت یک بعدی، می‌توان محور x ها را به K بازه افراز کرد. برآوردگر PL به طور جداگانه برای هر بازه محاسبه نموده و روش K نمونه را به کار برد. اگر $\underline{x}$ چند بعدی باشد،

می‌توان فضای $\underline{x}$ را به K ناحیه افراز کرد. با این وجود، لازمهٔ طبقه‌بندی داده‌ها، زیاد بودن آنهاست و تعداد لازم به سرعت با بعد $\underline{x}$ افزایش می‌یابد.

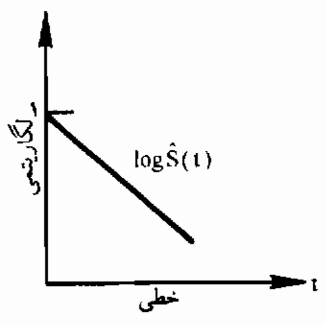
یک راه دیگر برای طبقه‌بندی به شرح زیر است. ابتدا تعریف زیر ارائه می‌شود:

$$\Lambda_{\underline{x}_i}(T_i) = e^{\underline{\beta}'\underline{x}_i} \int_0^{T_i} \lambda_0(u)du$$

در این صورت، تحت الگوی نرخ شکست متناسب، رابطهٔ زیر نشان می‌دهد که $\Lambda_{\underline{x}_i}(T_i)$ یک متغیّر نمایی واحد است.

$$P\{\Lambda_{\underline{x}_i}(T_i) > t\} = P\{T_i > \Lambda_{\underline{x}_i}^{-1}(t)\} = \exp\{-\Lambda_{\underline{x}_i}(\Lambda_{\underline{x}_i}^{-1}(t))\} = e^{-t}$$

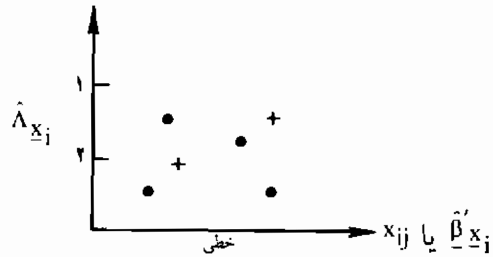
بنابراین، $(\Lambda_{\underline{x}_1}(Y_1), \delta_1)$ تا $(\Lambda_{\underline{x}_n}(Y_n), \delta_n)$ یک نمونه از توزیع نمایی واحد با برش است. چون $\Lambda_{\underline{x}_i}(Y_i)$ به پارامتر مجهول $\underline{\beta}$ و $\lambda_0(t)$ بستگی دارد، برآوردها را جانشین می‌کنیم. با تعریف $\hat{\Lambda}_i = \hat{\Lambda}_{\underline{x}_i}(Y_i) = e^{\hat{\underline{\beta}}'\underline{x}_i} \int_0^{Y_i} \hat{\lambda}_0(u)du$، فرض کنید $\hat{S}$ برآوردگر PL بر مبنای $(\hat{\Lambda}_1, \delta_1)$ تا $(\hat{\Lambda}_n, \delta_n)$، تحت الگوی نرخ شکست متناسب باشد، باید $\log\hat{S}(t)$ تقریباً تابع خطی از t شود.



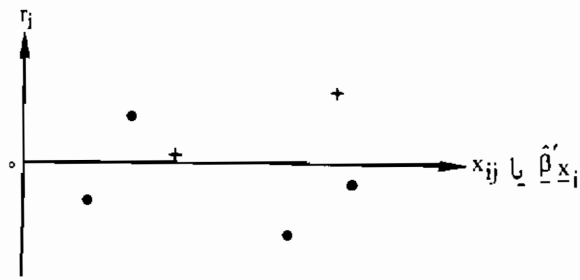
اگر نمودار $\log\hat{S}(t)$ در مقابل t خطّی نباشد، امکان دارد تصمیم‌گیری در مورد الگوی مناسب دیگری مشکل باشد.

تحت الگوی نرخ شکست متناسب، $(\hat{\Lambda}_1, \delta_1)$ تا $(\hat{\Lambda}_n, \delta_n)$، مقادیر بریده شده (بریده نشده)، تقریباً متغیّرهای تصادفی iid هستند. بنابراین، نباید رسم $\hat{\Lambda}_i(t)$ در مقابل یک متغیّر کمکی خاص همانند: x_{ij} یا در مقابل $\hat{\underline{\beta}}'\underline{x}_i$ هیچ‌گونه طرح مجانبی را آشکار

سازد. برآوردهای $\hat{\Lambda}_1$ تا $\hat{\Lambda}_n$ را مانده‌های تعمیم یافته می‌نامند.



در وارسی الگوی خطّی، در صورت زیاد بودن تعداد مشاهدات، افراز ناحیهٔ $\underline{x}$ بــه K زیــر ناحیــه و اســتفاده از روش K نمونــه پیشــنهاد می‌شــود. از طــرف دیگــر، می‌تــوان مانده‌های: $r_i = y_i - \hat{\underline{\beta}}'\underline{x}_i$ را در مقابل متغیّر کمکی x_{ij} یا $\hat{\underline{\beta}}'\underline{x}_i$ رسم کرد.



برای هر دو مورد نرخ شکست متناسب و الگوی خطّی، حساسیّت نمودار مانده‌ها را رســم تا اثر یک متغیّر وابستهٔ x_{ij} را کشف کنیم. امکان دارد این عمل با محاســبهٔ باقی‌مــانده و حذف این متغیّر وابسته در الگو انجام شود.

REFERENCES

Cox and Snell, JRSS B (1968), discuss generalized residuals.

Crowley and Hu, JASA (1977), plot generalized residuals for the Stanford heart transplant data.

Kay, Appl. Stat. (JRSS C) (1977), discusses plotting generalized residuals.

۲ آزمونها

۱.۲ یک نمونه

می‌خواهیم فرض: $H_0: F = F_0$ را با معلوم بودن F_0، آزمون کنیم.

(الف) تعمیم آزمون کلموگروف-اسمیرنف: فرض $H_\circ$ را می‌پذیریم، اگر

$$\sqrt{n}\left|\hat{F}(t)-F_\circ(t)\right|\le \hat{C}_n(t) \qquad t\ge \circ$$

که در آن، $\hat{F}(t)$ برآوردگر PL و $\hat{C}_n(t)$ را از جدول محاسبه کنیم. این آزمون را می‌توان برای ساختن فاصلهٔ اطمینان برای $F_\circ(t)$ نیز به کار برد، داریم:

$$P\left\{\hat{F}(t)-\frac{\hat{C}_n(t)}{\sqrt{n}}\le F_\circ(t)\le \hat{F}(t)+\frac{\hat{C}_n(t)}{\sqrt{n}} \quad ,\forall t\ge \circ\right\}=۱-\alpha$$

REFERENCES

Barr and Davidson, Technometrics (1973), and
Koziol and Byar, Technometrics (1975), and
Dufour and Maag, Technometrics (1978), consider Type I and Type II censoring.
Gillespie and Fisher, Ann. Stat. (1979), and
Hall and Wellner, Biometrika (1980), consider the PL estimator and random censoring.

(ب) تعمیم آزمون کرامر-وون مایسز: بعد از انجام یک تبدیل احتمالی انتگرال به‌گونه‌ای که $F_\circ(t)=t$ باشد، تابع توزیع، یکنواخت شود. آزمون کرامر-وون مایسز از آمارهٔ زیر استفاده می‌کند.

$$n\int_\circ^۱ (\hat{F}(t)-t)^۲\,dt$$

که در آن $\hat{F}$ برآوردگر PL است.

REFERENCES

Koziol and Green, Biometrika (1976), consider the PL estimator and random censoring.
Pettit and Stephens, Biometrika (1976), consider Type I and Type II censoring. Pettit specializes to the normal and exponential distributions in
Pettit, Biometrika (1976), and
_____, Biometrika (1977), respectively.

(ج) آزمون نوع مانتل-هانزل

REFERENCE

Hyde, Biometrika (1977).

(د) حدّ آزمون افرون

REFERENCE

Hollander and Proschan, Biometrics (1979).

(ه) خانواده‌های پارامتری

فرض کنید می‌خواهیم فرض: $\underline{\theta} \in \underline{\Theta}$, $H_\circ: F = F_{\underline{\theta}}$ را آزمون کنیم. به روش معمول یک برآورد $\hat{\underline{\theta}}$ را به دست می‌آوریم. سپس، وارسی می‌کنیم که آیا $\hat{F}$ به اندازهٔ کافی به $F_{\hat{\underline{\theta}}}$ نزدیک است یا خیر.

REFERENCE

Mihalko and Moore, Ann. Stat. (1980), consider χ^2 - tests for Type II censoring with estimates that are asymptotically equivalent to linear combinations of order statistics.

اگر $\underline{\Theta}_\circ \subset \underline{\Theta}$ و بخواهیم $H_\circ: \underline{\theta} \in \underline{\Theta}_\circ$ را آزمون کنیم، باید آزمون نسبت درست‌نمایی را به کار بریم.

REFERENCE

Turnbull and Weiss, Biometrics (1978), consider likelihood ratio tests for discrete or grouped data.

۲.۲ رگرسیون

(الف) خانواده‌های پارامتری

الگو را در یک الگوی بزرگتر می‌نشانیم (مثلاً، الگویی که دارای اثرات درجه ۲ یا ۳ اثرات متقابل است). آزمون می‌کنیم که آیا الگوی کوچکتر برقرار است؟ در واقع فرض زیر را آزمون می‌کنیم

$$H_\circ: \underline{\theta} \in \underline{\Theta}_\circ \subset \underline{\Theta}$$

(ب) آزمونهای $\chi^۲$

REFERENCES

Schoenfeld, Biometrika (1980), considers proportional hazards models with regions in the time × covariate space.

Lamborn, Stanford Univ. Tech. Report No. 21 (1969), looks at χ^2 – tests for exponential regression.

فصل هشتم

مباحث مختلف

۱ برآوردگر دو متغیّری کاپلان-مایر

فرض کنید، $\underline{T}_i = (T_{i1}, T_{i2})$، یک زوج از زمانهای شکست باشد. برای مثال، ممکن است، زمانهای از کار افتادن کلیّه‌های راست و چپ یا زمانهای تشخیص سرطان در شش راست و چپ باشد. امکان دارد، هر کدام یا هر دو زمان شکست به علّت متغیّر تصادفی بریده شدهٔ بعدی C_i، قابل مشاهده و بردار نشانگر به شرح زیر است:

$$\underline{Y}_i = (Y_{i1}, Y_{i2}) = (T_{i1} \wedge C_i, T_{i2} \wedge C_i)$$

$$\underline{\delta} = (\delta_{i1}, \delta_{i2}) = (I(T_{i1} \leq C_i), I(T_{i2} \leq C_i))$$

میونوز نشان داده است که چگونه می‌توان برآوردگر تعمیم یافته کاپلان-مایر دوبعدی را به کمک الگوریتم خودسازگاری و تجدید نظر در توزیع راست، محاسبه کرد. همچنین، او ثابت نمود که این برآوردگر، تعمیم برآورد حداکثر درست‌نمایی و یک برآوردگر سازگار توزیع دو متغیّری $F(t_1, t_2) = P\{T_{i1} \leq t_1, T_{i2} \leq t_2\}$ است.

کامپل الگو را با زمانهای برش دو متغیّری و رفتار شرایط داده‌های طبقه‌بندی شده در نظر گرفته است، همچنین، کوروار داده‌های طبقه‌بندی شده دو متغیّره با هر دو نوع برش راست و چپ را بررسی کرده است.

REFERENCES

Campbell, Purdue univ. Mimeoseries #79 – 25 (1979),

and

Korwar, unpublished manuscript (1980), treat bivariate grouped

data with censoring.
Muñoz, Stanford Univ. Tech. Report No. 60 (1980), defines the two – dimensinoal KM estimator through algorithms and proves it is the GMLE.
______, Stanford Univ. Tech. Report No. 61 (1980), proves consistency of the two – dimensional estimator.

۲ نرخ شکست رقیب

فرض کنید $\underline{T}_i = (T_{i1}, \ldots, T_{ip})$، یک بردار p بعدی از زمانهای شکست باشد. هر مختص، زمان شکست حاصل از علّت خاصّی مانند: از کار افتادن قلب، سرطان، خرابی کبد و غیره است. موضوع فقط هنگامی قابل مشاهده است که اوّلین شکست رخ دهد: زمانهای شکست تمام عوامل دیگر، توسّط از کار افتادگی دستگاه در اوّلین زمان شکست، بریده می‌شوند. کمّیتهای قابل مشاهده به صورت زیر است:

$$T_i = \min\{T_{i1}, \ldots, T_{ip}\}$$

و

$$\underline{\delta}_i = (\delta_{i1}, \ldots, \delta_{ip}) = (I(T_{i1} \le T_i), \ldots, I(T_{ip} \le T_i))$$

بردار نشانگر $\underline{\delta}$، علّت خاص شکست را نشان می‌دهد.

احتمال $P\{T_{ij} \le t, \delta_{ij} = 1\}$ را احتمال خام مردن به علّت j در زمان t می‌نامند. این احتمال، مستقیماً توسط نسبت مشاهده‌ای، برآورد می‌شود.

$$\frac{1}{n}\sum_{i=1}^{n} I(T_i \le t, \delta_{ij} = 1)$$

علّت خام، عبارت از $P(T_{ij} \le t)$ و اگر علّتها مستقل باشند، آن را می‌توان به طور سازگار به وسیلهٔ روش PL برآورد کرد، که در آن تمام زمانهای شکست را به علّت j، به علّت یکی از زیر مجموعه‌های ممکن علّتها در نظر می‌گیرد.

یک نتیجهٔ اصلی در نظریهٔ نرخهای شکست رقیب این است که بر مبنای نمونهٔ T_i و $\underline{\delta}_i$، $i = 1, \ldots, n$، نمی‌توان مدّعی شد که این رابطه

$$P\{T_{i1} \le t_1, \ldots, T_{ip} \le t_p\} = \prod_{j=1}^{p} P\{T_{ij} \le t_j\}$$

برقرار است یا زمانهای شکست T_{i1}، ...، T_{ip} وابسته‌اند. اثباتهای مختلف ایـــن نتیجـه بـا شرایط متفاوت، سالهای مورد بحث بوده است. به مقالات برمـــن، آلت‌شــولر، تســیاتیس، پترسن، لانگبرگ - پروچان - کوینزی نگاه کنید.

REFERENCES

Chiang, Introduction to Stochastic Processes in Biostatistics (1968), discusses the relationships between crude, net, and partial crude probabilities in Chapter 11.

Moeschberger and David, Biometrics (1971), consider parametric likelihood methods.

Gail, Biometrics (1975), is a review article.

Prentice et a1., Biometrics (1978), review competing risks from the hazard rate point of view.

Berman, Ann. Math. Stat. (1963),

Altshuler, Mathematical Biosciences (1970),

Tsiatis, Proc. Nat1. Acad. Sci. (1975),

Peterson, Stanford Univ. Tech. Report No. 13 (1975),

______, Proc. Natl. Acad. Sci. (1976), and

Langberg Proschan, and Quinzi, Ann. Stat. (1981), examine the identifiability question.

۳ برش وابسته

برای حالتی که زمانهای شکست و زمانهای برش وابسته‌اند، کار زیـــادی صــورت نگرفته است. بعضی از کارهای انجام شده در نرخهای شکســـت رقیــب وابســته در ایــن خصوص انجام شده است. در مقالات لاگاکوس و ویلیامز بحث‌های کلّـی و نتـایج ارائـه شده است.

REFERENCES

Williams and Lagakos, Biometrika (1977).

Lagakos and Williams, Biometrika (1978).

Lagakos, Biometrics (1979).

۴ روش جک‌نایف و بوت‌استراپ

فرض کنید پارامتر θ یک تابعی $T(F)$ از تابع توزیع F باشد. در بسیاری از موارد θ توسّط تابع توزیع نمونه F_n، به جای F در T، به صورت $\hat{\theta}=T(F_n)$، برآورد می‌شود. برآورد جک‌نایف θ برای یک نمونهٔ Y_1 تا Y_n، با توزیع F به صورت زیر تعریف می‌شود:

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i} \quad , \; i=1,\dots,n$$

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i = n\hat{\theta} - \frac{(n-1)}{n}\sum_{i=1}^{n}\tilde{\theta}_{-i}$$

که در آن $\tilde{\theta}_{-i} = T(F_{n-1,-i})$، برآورد θ با این فرض که نمونهٔ iام یعنی Y_i در نمونه حذف شده، می‌باشد.

در صورت نبودن برش، رابطهٔ زیر برای T به اندازهٔ کافی هموار قابل اثبات است.

$$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{1}{n(n-1)}\sum_1^n(\tilde{\theta}_i-\tilde{\theta})^2}} \overset{a}{\sim} N(0,1) \qquad (18)$$

برای بررسی (۱۸) در شرایط مختلف به مقالهٔ تجدید نظر شدهٔ میلر (۱۹۷۴) مراجعه شود. میلر نشان داده که (۱۸) برای داده‌هایی که به طور تصادفی بریده شده -با این فرض که $\hat{\theta}=T(\hat{F})$ و $\hat{F}$ برآوردگر PL است- برقرار است.

هموار بودن T که در (۱۸) استفاده می‌شود، با هموار بودن تابع تأثیر زیر در ارتباط است.

$$IC(y;F) = \lim_{\varepsilon\to 0}\frac{T((1-\varepsilon)F+\varepsilon\delta_y)-T(F)}{\varepsilon}$$

که در آن δ_y، تابع توزیعی است که به نقطهٔ y جرم واحد را نسبت می‌دهد. برای داده‌های بریده نشده، تابع جک‌نایف و تابع تأثیر به شکل زیر با هم در ارتباط‌اند.

$$(n-1)(\hat{\theta}-\hat{\theta}_{-i}) = \left.\frac{T((1-\varepsilon)F+\varepsilon\delta_y)-T(F)}{\varepsilon}\right|_{\varepsilon=-1/(n-1),\, F=F_n,\, y=y_i}$$

راید برای داده‌های بریده شده، توابع تأثیر (یعنی مشتقّات جزئی نسبت به F_u و F_c) را برای تابعی از برآوردگر PL، پیدا نموده است.

بوت‌استراپ افرون، به صورت زیر انجام می‌شود. فرض کنیـــد Y_1^* تـا Y_n^*، یـک نمونه باجایگذاری از y_1 تا y_n، در حالت بدون برش باشد. در صورت وجود برش، فرض کنید (Y_1^*, δ_1^*) تا (Y_n^*, δ_n^*)، نمونه‌ای باجایگذاری از (y_1, δ_1) تا (y_n, δ_1) باشد. در این صورت F_n^* و $\hat{F}^*$، به ترتیب: توزیع نمونهٔ بوت‌استراپ و برآوردگر PL هستند. همچنیـــن، $\hat{\theta}^* = T(F_n^*)$ یا $T(\hat{F}^*)$، بستگی به این دارد که داده‌ها بریده شده یا بریده نشـــده باشـند. این روش نمونه‌گیری N بار تکرار می‌شود تا $\hat{\theta}_1^*$، ...، $\hat{\theta}_N^*$، حاصل شود. توزیع تجربی $\hat{\theta}_1^*$ تا $\hat{\theta}_N^*$، برای محاسبهٔ توزیع تقریبی $\hat{\theta}$ به کار می‌رود. به ویژه کمّیتهـای محـوری توزیـع تجربی $\hat{\theta}^* - \hat{\theta}$ برای تقریب توزیع $\hat{\theta} - \theta$، مورد استفاده قرار می‌گیرد.

REFERENCES

Miller, Biometrika (1974), reviews the jackknife for uncensored data problems.

______, Stanford Univ. Tech. Report No. ۱۴ (۱۹۷۵), establishes the validity of jackknifing the PL estimator.

Reid, Ann. Stat. (۱۹۸۱), derives the influence functions for the PL estimator.

Efron, Ann. Stat. (۱۹۷۹), introduces bootstrapping for uncensored data problems.

______, Stanford Univ. Tech. Report No. ۵۳ (۱۹۸۰), studies bootstrapping

فصل نهم

مسائل

مسألهٔ ۱.

ثابت کنیـد توزیـع گامـا در ازای $\alpha > 1$ دارای IFR و در ازای $\alpha < 1$ دارای DFR است.

حل:

$$\frac{1}{\lambda(t)} = \frac{\int_t^\infty x^{\alpha-1} e^{-\lambda x} dx}{t^{\alpha-1} e^{-\lambda t}} = \int_t^\infty \left(\frac{x}{t}\right)^{\alpha-1} e^{-\lambda(x-t)} dx$$

با تغییر متغیّر $u = x - t$

$$= \int_0^\infty \left(1+\frac{u}{t}\right)^{\alpha-1} e^{-\lambda u} du$$

اگر $\alpha > 1$ باشد، عبارت: $\left(1+\frac{u}{t}\right)^{\alpha-1}$ برحسب t نزولی و $\lambda(t)$ صعودی خواهد بود.
اگر $\alpha < 1$ باشد، عبارت بالا برحسب t صعودی و $\lambda(t)$ نزولی خواهد بود.

مسألهٔ ۲.

تابع اطّلاع فیشر را برای یک مشاهده در توزیع نمایی با برش نوع اوّل بـه دسـت آورید.

حل:

فرض کنید t_c زمان برش، ثابت باشد. لگاریتم درست‌نمایی به شرح زیر است:

$$\delta \log\lambda - \delta\lambda y - (1-\delta)\lambda t_c$$

اگر نسبت به λ دوبار مشتق بگیریم، $-\frac{\delta}{\lambda^{۲}}$ به دست می‌آید. در نتیجه اطّلاع فیشر به شرح زیر خواهد بود:

$$I(\lambda)=\frac{۱}{\lambda^{۲}}E(\delta)=\frac{۱}{\lambda^{۲}}P\{T\le t_c\}=\frac{۱}{\lambda^{۲}}(۱-e^{-\lambda t_c})$$

مسألهٔ ۳.

ماتریس اطّلاع نمونه را برای توزیع وایبل تحت برش تصادفی به دست آورید.

حل:

از معادلهٔ شماره ۲ فصل دوم، داریم:

$$\frac{\partial}{\partial\gamma}\log L=\frac{n_u}{\gamma}-\sum_{i=۱}^{n}y_i^{\alpha}$$

$$\frac{\partial}{\partial\alpha}\log L=\frac{n_u}{\alpha}+\sum_{u}\log t_i-\gamma\sum_{i=۱}^{n}y_i^{\alpha}\log y_i$$

ماتریس اطّلاع نمونه در (γ,α)، به صورت زیر است:

$$-\begin{pmatrix}\frac{\partial^{۲}}{\partial\gamma^{۲}}\log L & \frac{\partial^{۲}}{\partial\gamma\,\partial\alpha}\log L\\ & \frac{\partial^{۲}}{\partial\alpha^{۲}}\log L\end{pmatrix}$$

مسألهٔ ۴.

از فوریه ۱۹۷۲ تا فوریه ۱۹۷۵، تعداد ۲۹ بیمار در دو بیمارستان تحت معالجه قرار گرفته و به تصادف به دو گروه تیمار و شاهد تقسیم شده‌اند. زمانهای بقاء (برحسب هفته) چهارده بیمار در گروه تیمار، به شرح زیر است:

۱۶^+ و ۱۶^+ و ۱۶^+ و ۱۲^+ و ۱۰^+ و ۱۰ و ۸ و ۷ و ۵ و ۴^+ و ۱^+ و ۱ و ۱ و ۱

توزیع نمایی $S(t)=\exp(-\lambda t)$ را فرض کنید و به سؤالات زیر پاسخ دهید.

(الف) مقدار λ را به روش حدّاکثر درست‌نمایی برآورد کنید و یک فاصلهٔ اطمینان ۹۵٪ را برای آن پیدا کنید.

(ب) مقدار $S(۱۶)$ را برآورد کنید و فاصلهٔ اطمینان ۹۵٪ را برای آن پیدا کنید.

(پ) میانهٔ زمان بقاء را برآورد کنید. فاصلهٔ اطمینان ۹۵٪ را برای آن پیدا کنید.

حل:

(الف) از مثال شماره ۱ فصل دوم بخش (۱.۲)، داریم:

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n} y_i} = \frac{7}{108} = 0/065 \qquad \text{و} \qquad \log \hat{\lambda} \overset{a}{\sim} N(\log \lambda, \frac{1}{n_u})$$

در نتیجه فاصلهٔ اطمینان ۹۵٪ برای λ، به شرح زیر است:

$$\left(\hat{\lambda} \exp\left(\frac{-Z_{0/025}}{\sqrt{n_u}}\right), \hat{\lambda} \exp\left(\frac{Z_{0/025}}{\sqrt{n_u}}\right)\right) = (0/031, 0/136)$$

(ب) داریم: $\hat{S}(16) = \exp(-\hat{\lambda} \times 16) = 0/355$. یک فاصلهٔ اطمینان ۹۵٪ برای $S(16)$ به شرح زیر است:

$$\left(e^{-0/136 \times 16}, e^{-0/031 \times 16}\right) = (0/113, 0/609)$$

(پ) داریم: $\hat{t}_{med} = \log(2)/\hat{\lambda} = 10/69$. یک فاصلهٔ اطمینان ۹۵٪ برای میانه به شرح زیر است:

$$\left(\frac{\log 2}{0/136}, \frac{\log 2}{0/031}\right) = (5/097, 22/36)$$

مسألهٔ ۵.

برای داده‌های بیمارستانی مسألهٔ شماره ۴، برآورد حدّ حاصل‌ضرب کاپلان-مایر، تابع بقاء را به دست آورید. نمودار آن و نمودار تابع بقـاء را تحـت فـرض نمـایی روی همان محور مختصّات خطّی-لگاریتمی رسم کنید. آیا فکر می‌کنید فرض نمایی در مدّت شانزده هفته صادق است.

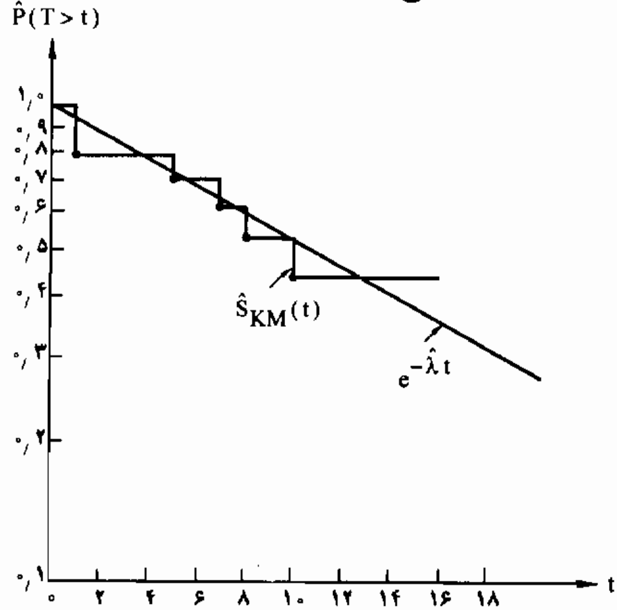
حل:

داریم: (شکل زیر را نیز ببینید)

$$\hat{S}(t) = \begin{cases} 1 & 0 \le t < 1 \\ 11/14 = 0/786 & 1 \le t < 5 \\ 11 \times 8/14 \times 9 = 0/698 & 5 \le t < 7 \end{cases}$$

$$\hat{S}(t)=\begin{cases}۱۱\times ۷/۱۴\times ۹=۰/۶۱۱ & ۱۰\leq t<۱۶\\ ۱۱\times ۶/۱۴\times ۹=۰/۵۲۴ & ۸\leq t<۱۰\\ ۱۱\times ۵/۱۴\times ۹=۰/۴۳۷ & ۷\leq t<۸\end{cases}$$

نتایج عادلانه نیکویی برازش عبارت است از: پانزده نمونه از بیست و یک نمونهٔ رسم شده توزیع نمایی را تأیید می‌کنند و پنج نمونه تأیید نمی‌کنند و یک مورد جواب نمی‌دهد.



مسألهٔ ۶.

از جدول طول عمر (جدول شماره ۱) خطای استاندارد $S(۵)$ را حساب کنید.

حل: t

با استفاده از رابطهٔ گرین‌وود، داریم:

$$\hat{Var}(\hat{S}(۵))\cong(۰/۴۴)^۲\left[\frac{۴۷}{۱۱۶/۵(۱۱۶/۵-۴۷)}+\frac{۵}{۵۱/۵(۵۱/۵-۵)}+\frac{۲}{۳۰/۵(۳۰/۵-۵)}+\frac{۲}{۱۶/۵(۱۶/۵-۲)}+\frac{۰}{۷(۷-۰)}\right]=۰/۰۰۳۶۰۸$$

بنابراین $\hat{SE}(\hat{S}(۵))\cong ۰/۰۶$.

مسألهٔ ۷.

به کمک داده‌های AML (بخش دوم از فصل سوم) خطای استاندارد $\hat{S}(۲۴)$ را در گروه تیمار حساب کنید.

حل:

با استفاده از رابطهٔ گرین‌وود، داریم:

$$\hat{Var}(\hat{S}(۲۴))=\left(\frac{۶\times۹}{۱۱\times۸}\right)^{۲}\left(\frac{۱}{۱۰\times۱۱}+\frac{۱}{۹\times۱۰}+\frac{۱}{۷\times۸}+\frac{۱}{۶\times۷}\right)=۰/۰۲۳۲۹$$

در نتیجه: $\hat{SE}(\hat{S}(۲۴))=۰/۱۵۲۶$.

مسألهٔ ۸.

در اثبات GMLE برآوردگر PL، نشان دهید حداکثر عبارت:

$$\prod_{i=۱}^{n} p_i^{\delta_{(i)}}\left(\sum_{j=i}^{n} p_j\right)^{۱-\delta_{(i)}}$$

به ازای مقدار زیر به دست می‌آید:

$$p_i=\frac{\delta_{(i)}}{n-i+۱}\prod_{j=۱}^{i-۱}\left(۱-\frac{\delta_{(j)}}{n-j+۱}\right)$$

حل:

فرض کنید: $\lambda_i=\frac{p_i}{\sum_{j=i}^{n} p_j}$، $i=۱,\dots,n$ در این صورت چون، $\sum_{j=۱}^{n} p_j=۱$ و $۱-\lambda_i=\sum_{j=i+۱}^{n} p_j\Big/\sum_{j=i}^{n} p_j$، داریم: $\sum_{j=i}^{n} p_j=\prod_{j=۱}^{i-۱}(۱-\lambda_j)$ و چون $\lambda_n=۱$ است، در نتیجه داریم:

$$\prod_{i=۱}^{n} p_i^{\delta_{(i)}}\left(\sum_{j=i}^{n} p_j\right)^{۱-\delta_{(i)}}=\prod_{i=۱}^{n}\lambda_i^{\delta_{(i)}}\prod_{j=۱}^{i-۱}(۱-\lambda_j)=\prod_{i=۱}^{n}\lambda_i^{\delta_{(i)}}(۱-\lambda_i)^{n-i}$$

با توجّه به نظریهٔ نمونه‌گیری دوجمله‌ای، هر حاصل‌ضرب به

ازای: $\hat{\lambda}_i = \frac{\delta_{(i)}}{n-i+\delta_{(i)}} = \frac{\delta_{(i)}}{n-i+1}$، حداکثر می‌شود. بنابراین، داریم:

$$\hat{p}_i = \hat{\lambda}_i\left(\sum_{j=i}^{n}\hat{p}_j\right) = \hat{\lambda}_i\prod_{j=1}^{i-1}(1-\hat{\lambda}_j) = \frac{\delta_{(i)}}{n-i+1}\prod_{j=1}^{i-1}\left(1-\frac{\delta_{(j)}}{n-j+1}\right)$$

مسألهٔ ۹.

ثابت کنید که الگوریتم تجدید نظر در توزیع به راست برآوردگر حدّ حاصل‌ضرب کاپلان-مایر را بدون تکرار می‌دهد.

حل:

دو راه اصلی برای اثبات این نتیجه وجود دارد:

(الف) با توجّه به الگوریتم بالا، کلّیهٔ نقاط $y_{(i)}$ در هر دو حالت بریده شده و بریده نشده، در ابتدا دارای جرم $\frac{1}{n}$ هستند. این الگوریتم از چپ به راست در میان آماره‌های مرتّب حرکت می‌کند. هنگامی که به $y_{(i)}-$ می‌رسد، تمام نقاط باقی‌ماندهٔ، $y_{(i)}$، $y_{(i+1)}$ و ... و $y_{(n)}$، دارای جرم مساوی، با توجّه به روش انجام الگوریتم، هستند. فرض کنید جرم کلّ باقی‌مانده $\tilde{S}(y_{(i)}-)$ باشد. با توجّه به مساوی بودن جرمها، $y_{(i)}$ دارای جرم $\tilde{S}(y_{(i)}-)/(n-i+1)$ است. اگر بریده نشده باشد این جرم حفظ می‌شود. اگر بریده شده باشد، این جرم در سمت راست توزیع، توزیع می‌شود.

چون برآوردگر PL، $\hat{S}$، مانند $\tilde{S}$ از یک شروع می‌شود و اندازهٔ جهشهای آن در مشاهدات بریده نشده به صورت: $\hat{S}(y_{(i)}-)/(n-i+1)$ و در حالت بریده شده صفر است، لذا، دو برآوردگر برابرند.

(ب) برای برآوردگر کاپلان-مایر، داریم:

$$\hat{\Delta}_{(i)} = \hat{S}(y_{(i)}-) - \hat{S}(y_{(i)}) = \prod_{j=1}^{i-1}\left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}} - \prod_{j=1}^{i}\left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}$$

$$= \prod_{j=1}^{i-1}\left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}\frac{\delta_{(i)}}{n-i+1} = \prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)^{-\delta_{(j)}} \times \frac{1}{n}\times\frac{n}{n-1}\times\cdots\times$$

$$\times\frac{n-i+2}{1}\times\frac{\delta_{(i)}}{n-i+1} = \frac{\delta_{(i)}}{n}\prod_{j=1}^{i-1}\left(\frac{n-j+1}{n-j}\right)^{1-\delta_{(j)}}$$

فرض کنید، $j_i > \cdots > j_1$، زیرنویسهای مشاهدات بریده شدهٔ قبل از $y_{(i)}$ باشند. برای الگوریتم تجدید نظر در توزیع به راست، جرم نسبت داده شده به $y_{(i)}$ با شرط $\delta_{(i)} = ۱$ برابر زیر است:

$$\tilde{\Delta}_{(i)} = \frac{۱}{n}\left(۱+\frac{۱}{n-j_۱}\right)\left(۱+\frac{۱}{n-j_۲}\right)\cdots\left(۱+\frac{۱}{n-j_i}\right) = \frac{۱}{n}\prod_{j=۱}^{i-۱}\left(\frac{n-j+۱}{n-j}\right)^{۱-\delta_{(j)}}$$

و اگر $\delta_{(i)} = ۰$ باشد، $\tilde{\Delta}_{(i)} = ۰$ می‌شود. این معادل $\hat{\Delta}_{(i)}$ است. بنابراین، الگوریتم تجدید نظر در توزیع به راست، برآوردگر PL را می‌دهد.

مسألهٔ ۱۰.

اگر برای برآوردگر PL، $\hat{S}(t)$ رابطهٔ:

$$ACov(\hat{S}(t_۱), \hat{S}(t_۲)) = \frac{S(t_۱)\,S(t_۲)}{n} \times \int_۰^{t_۱\wedge t_۲} \frac{d\,F_u(s)}{(۱-H(s))^۲}$$

را داشته باشیم، نشان دهید که برای $\hat{\mu} = \int_۰^{\infty} \hat{S}(t)dt$، رابطهٔ زیر برقرار است:

$$AVar(\hat{\mu}) = \frac{۱}{n}\int_۰^{\infty} \frac{۱}{(۱-H(s))^۲}\left(\int_s^{\infty} S(t)dt\right)^۲ d\,F_u(s)$$

که در آن AVar و ACov، به معنی واریانس و کوواریانس مجانبی است.

حل:

$$Var(\hat{\mu}) = E(\hat{\mu}^۲) - (E(\hat{\mu}))^۲$$

$$= E\left(\int_۰^{\infty}\int_۰^{\infty} \hat{S}(t_۱).\hat{S}(t_۲)\,dt_۱\,dt_۲\right) - \left(E\left(\int_۰^{\infty}\hat{S}(t)\,dt\right)\right)^۲$$

$$= \int_۰^{\infty}\int_۰^{\infty} Cov(\hat{S}(t_۱).\hat{S}(t_۲))\,dt_۱\,dt_۲$$

بنابراین، داریم:

$$AVar(\hat{\mu}) = \frac{۱}{n}\int_۰^{\infty}\int_۰^{\infty} S(t_۱).S(t_۲) \times \int_۰^{t_۱\wedge t_۲}\frac{d\,F_u(s)}{(۱-H(s))^۲}\,dt_۱\,dt_۲$$

با تعویض انتگرال (قضیه فوبینی)، داریم:

$$AVar(\hat{\mu})=\frac{1}{n}\int_{0}^{\infty}\frac{1}{(1-H(s))^{2}}\int_{s}^{\infty}S(t_1)dt_1\times\int_{s}^{\infty}S(t_2)dt_2\,dF_u(s)$$

$$=\frac{1}{n}\int_{0}^{\infty}\frac{1}{(1-H(s))^{2}}\left(\int_{s}^{\infty}S(t)dt\right)^{2}dF_u(s)$$

مسألهٔ ۱۱.

برای داده‌های AML در گروه شاهد، که به شکل زیر هستند، **(الف)** $\hat{\mu}$ و **(ب)** $\hat{Var}(\hat{\mu})$ را محاسبه کنید.

برحسب هفته ۴۵ و ۴۳ و ۳۳ و ۳۰ و ۲۷ و ۲۳ و $^+$۱۶ و ۱۲ و ۸ و ۸ و ۵ و ۵

حل:

(الف) برآوردگر کاپلان-مایر $\hat{S}(t)$ در جدول زیر داده شده است:

$t\in$	$[0,5)$	$[5,8)$	$[8,12)$	$[12,23)$	$[23,27)$	$[27,30)$	$[30,33)$	$[33,43)$	$[43,45)$	$[45,\infty)$
$\hat{S}(t)=$	1	$\frac{10}{12}$	$\frac{8}{12}$	$\frac{7}{12}$	$\frac{7}{12}\times\frac{5}{6}$	$\frac{7}{12}\times\frac{4}{6}$	$\frac{7}{12}\times\frac{3}{6}$	$\frac{7}{12}\times\frac{2}{6}$	$\frac{7}{12}\times\frac{1}{6}$	0

در این صورت، داریم:

$$\hat{\mu}=\int_{0}^{\infty}\hat{S}(t)dt=1\times5+\frac{10}{12}\times3+\frac{8}{12}\times4+\frac{7}{12}\times11+\frac{7}{12}\times\frac{5}{6}\times4+\frac{7}{12}\times\frac{4}{6}\times3+$$

$$+\frac{7}{12}\times\frac{3}{6}\times3+\frac{7}{12}\times\frac{2}{6}\times10+\frac{7}{12}\times\frac{1}{6}\times2=22/71$$

(ب)

$$\hat{Var}(\hat{\mu})=\sum_{u}\left(\int_{y_{(i)}}^{\infty}\hat{S}(t)dt\right)^{2}\frac{d_i}{n_i(n_i-d_i)}$$

$$=(17/71)^{2}\frac{2}{12\times10}+(15/21)^{2}\frac{2}{10\times8}+(12/54)^{2}\frac{1}{8\times7}+(6/125)^{2}\frac{1}{6\times5}+$$

$$+(4/18)^{2}\frac{1}{5\times4}+(3/01)^{2}\frac{1}{4\times3}+(2/14)^{2}\frac{1}{3\times2}+(0/19)^{2}\frac{1}{2\times1}=17/47$$

مسألهٔ ۱۲.

برای داده‌های هپاتیت (مسألهٔ ۴)، تیمار I و شاهد II، زمانهای بقاء گروهها برحسب هفته به شرح زیراند. با m=۱۴ و n=۱۵، مطلوب است محاسبهٔ (الف) آمارهٔ گهان، (ب) جایگشت آن و (پ) آمارهٔ استاندارد و مقدار P.

حل:

امتیازهای U^* در جدول زیر محاسبه شده‌اند:

Z	گروه	# < Z	# > Z	U^*
۱(۳)	I	۰	۲۶	−۲۶(۳)
۱^+	I	۳	۰	۳
۱^+	II	۳	۰	۳
۲^+	II	۳	۰	۳
۳(۲)	II	۳	۲۱	−۱۸(۲)
۳^+	II	۵	۰	۵
۴^+	I	۵	۰	۵
۵	I	۵	۱۸	−۱۳
۵+(۲)	II	۶	۰	۶(۲)
۷	I	۶	۱۵	−۹
۸	I	۷	۱۴	−۷
۱۰	I	۸	۱۳	−۵
۱۰^+	I	۹	۰	۹
۱۲^+	I	۹	۰	۹
۱۶+(۳)	I	۹	۰	۹(۳)
۱۶+(۸)	II	۹	۰	۹(۸)

(الف) آمارهٔ گهان برابر ۵۹ = $\sum_{II} U^*$ می‌شود.

(ب) واریانس جایگشت به شرح زیر است:

$$\frac{۱۴\times۱۵}{۲۹\times۲۸}\sum_{I,II}(U^*)^۲=۱۰۸۶/۷۲$$

(پ) آمارهٔ استاندارد برابر $۱/۷۹ = \frac{۵۹}{\sqrt{۱۰۸۶/۷۲}}$ به دست می‌آید، که متناظر با P یک‌طرفه ۰/۰۳۷۵، است.

REFERENCE

Gregory et al., New England Journal of Medicine (1976).

مسألهٔ ۱۳.

برای داده‌های بیمارستانی (مسألهٔ ۱۲)، موارد زیر را به دست آورید:

(الف) آمارهٔ MH و مقدار P متناظر آن.

(ب) صورت تارون-وایر آمارهٔ گهان و مقدار متناظر P.

حل:

محاسبات مانند جدول شماره ۴ از فصل چهارم، بخش دوم، انجام شده است.

z	n	m_1	n_1	a	$E_0(A)$	$n(a-E_0(A))$	$\frac{m_1(n-m_1)}{n-1}$	$\frac{n_1}{n}(1-\frac{n_1}{n})$	$n_1(n-n_1)$
۱	۲۹	۳	۱۴	۳	۱/۴۴۸	۴۵	۲/۷۸۶	۰/۲۴۹۷	۲۱۰
۳	۲۳	۲	۱۰	۰	۰/۸۶۹	−۲۰	۱/۹۰۹	۰/۲۴۵۷	۱۳۰
۵	۱۹	۱	۹	۱	۰/۴۷۴	۱۰	۱	۰/۲۴۹۳	۹۰
۷	۱۶	۱	۸	۱	۰/۵۰۰	۸	۱	۰/۲۵۰۰	۶۴
۸	۱۵	۱	۷	۱	۰/۴۶۷	۸	۱	۰/۲۴۸۹	۵۶
۱۰	۱۴	۱	۶	۱	۰/۴۲۸	۸	۱	۰/۲۴۴۹	۴۸

(الف)

$$MH=\frac{\text{مجموع ستون }\{a-E_{\circ}(A)\}}{\sqrt{\text{مجموع}\left(\text{ستون }\frac{n_1}{n}(1-\frac{n_1}{n})\times\text{ ستون }\frac{m_1(n-m_1)}{n-1}\right)}}=\frac{2/814}{\sqrt{2/1578}}=1/916$$

بنابراین، مقدار P یک طرفهٔ متناظر برابر ۰٫۰۲۷ است.

(ب) فرض کنید U_{TW} صورت تارون-وایر آمارهٔ گهان باشد، داریم:

$$U_{TW}=\frac{\text{مجموع ستون }\{n(a-E_{\circ}(A))\}}{\sqrt{\text{مجموع}\left\{\text{ستون }n_1(n-n_1)\times\text{ ستون }\frac{m_1(n-m_1)}{n-1}\right\}}}=\frac{59}{\sqrt{1091/23}}=1/786$$

در نتیجه مقدار P یک‌طرفه برابر ۰٫۰۳۷ است.

مسألهٔ ۱۴.

پنج نقطه داده‌های مربوط به تعویض قلب بیمارستان استانفورد را در نظر بگیرید (مثال فصل ششم، بخش چهارم)

(الف) فرض $H_{\circ}:\beta=1$ را در الگوی نرخ شکست متناسب با محاسبهٔ مقدار P به وسیلهٔ آمارهٔ کاکس را، که به صورت زیر است، آزمون کنید.

$$\left(\frac{\partial}{\partial\beta}\log L_c(1)\right)^2\Big/-\frac{\partial^2}{\partial\beta^2}\log L_c(1)$$

(ب) برآورد تسیاتیس-لینک از $S(t;x)$ را برای $x=1/5$ و $0\le t\le 297$، با استفاده از $\beta=1$، به دست آورید.

زمان بقاء (Y)	نمره‌های عمل (X)	شماره مریض
۵۴	۲٫۰۹	۳۶
127^+	۰٫۳۶	۱۱
۲۹۷	۰٫۶۰	۸۴
389^+	۱٫۴۴	۸۶
1536^+	۰٫۹۱	۳۳

حل:

(الف) فرض کنید:

i :	۱	۲	۳	۴	۵
x_i :	۲/۰۹	۰/۳۶	۰/۶۰	۱/۴۴	۰/۹۱

با توجّه به عبارت فصل ششم، بخش دوم، داریم:

$$\frac{\partial}{\partial\beta}\log L_c(1)=x_1+x_3-\frac{\sum_{j=1}^{5}x_j e^{x_j}}{\sum_{j=1}^{5}e^{x_j}}-\frac{\sum_{j=3}^{5}x_j e^{x_j}}{\sum_{j=3}^{5}e^{x_j}}=0/0965$$

$$-\frac{\partial^2}{\partial\beta^2}\log L_c(1)=\frac{\sum_{j=1}^{5}x_j^2 e^{x_j}}{\sum_{j=1}^{5}e^{x_j}}-\left(\frac{\sum_{j=1}^{5}x_j e^{x_j}}{\sum_{j=1}^{5}e^{x_j}}\right)^2+$$
$$+\frac{\sum_{j=3}^{5}x_j^2 e^{x_j}}{\sum_{j=3}^{5}e^{x_j}}-\left(\frac{\sum_{j=3}^{5}x_j e^{x_j}}{\sum_{j=3}^{5}e^{x_j}}\right)^2=0/5110$$

بنابراین، داریم:

$$\frac{\left(\frac{\partial}{\partial\beta}\log L_c(1)\right)^2}{-\frac{\partial^2}{\partial\beta^2}\log L_c(1)}=0/018$$

و مقدار P، که از χ^2 به دست می‌آید، تقریباً ۰/۹ می‌شود.

(ب) **برآورد تسیاتیس (بخش اوّل از فصل ششم) عبارت است از:**

$$\hat{\Lambda}_{\circ,T}(t)=\begin{cases} ۱ & \circ\le t<۵۴ \\ \dfrac{۱}{\sum_{j=۱}^{۵} e^{x_j}}=\circ/\circ۵۵۴ & ۵۴\le t<۲۹۷ \\ \dfrac{۱}{\sum_{j=۱}^{۵} e^{x_j}}+\dfrac{۱}{\sum_{j=۳}^{۵} e^{x_j}}=\circ/۱۷۲۷ & t=۲۹۷ \end{cases}$$

بنابراین، داریم:

$$\hat{S}_T(t;۱/۵)=e^{-\hat{\Lambda}_{\circ,T}(t)\,e^{۱/۵}}=\begin{cases} ۱ & \circ\le t<۵۴ \\ \circ/۷۸ & ۵۴\le t<۲۹۷ \\ \circ/۴۶ & t=۲۹۷ \end{cases}$$

برآوردگر مطلوب (بخش اوّل از فصل ششم) به شرح زیر است:

$$\hat{\Lambda}_{\circ,T}(t)=\begin{cases} \dfrac{\circ/\circ۵۵۴}{۵۴}\,t & \circ\le t<۵۴ \\ \dfrac{\circ/۱۷۲۷-\circ/\circ۵۵۴}{۲۹۷-۵۴}(t-۵۴)+\circ/\circ۵۵۴ & ۵۴\le t\le ۲۹۷ \end{cases}$$

بنابراین، داریم:

$$\hat{S}_L(t;۱/۵)=e^{-\hat{\Lambda}_{\circ,L}(t)\,e^{۱/۵}}=\begin{cases} e^{-\circ/\circ\circ۴۶\,t} & \circ\le t<۵۴ \\ \circ/۸۷۷\,e^{-\circ/\circ\circ۲۲t} & ۵۴\le t\le ۲۹۷ \end{cases}$$

مسألۀ ۱۵.

برای داده‌های AML (مثال بخش دوم از فصل سوم) دو گروه تیمار و شاهد به شرح زیرند:

گروه تیمار $۹, ۱۳, ۱۳^+, ۱۸, ۲۳, ۲۸^+, ۳۱, ۳۴, ۴۵^+, ۴۸, ۱۶۱^+$

گروه شاهد $۵, ۵, ۸, ۸, ۱۲, ۱۶^+, ۲۳, ۲۷, ۳\circ, ۳۳, ۴۳, ۴۵$

دو گروه را به شکل زیر مقایسه کنید:

(الف) با آمارهٔ گهان و واریانس جایگشت آن

(ب) با آمارهٔ ML

(پ) صورت تارون-وایر آمارهٔ گهان

در هر حالت آمارهٔ استاندارد و مقدار P متناظر را به دست آورید.

حل:

(الف) در محاسبهٔ نمرات مورد نیاز برای انجام آمارهٔ گهان، از جدول صفحه بعد استفاده می‌کنیم:

آمارهٔ گهان برابر $\sum_{NM} U^* = -۵۰$ و واریانس جایگشت ML، به شرح زیر است:

$$\frac{۱۱\times۱۲}{۲۳\times۲۲}\sum_{N,NM}(U^*)^۲ = ۹۱۲$$

بنابراین آمارهٔ استاندارد برابر با: $\frac{-۵۰}{\sqrt{۹۱۲}} = -۱/۶۵۶$ بوده، که متناظر مقدار P دوطرفهٔ ۰/۰۹، است.

Z	گروه	# < Z	# > Z	U^*
۵(۲)	NM	۰	۲۱	−۲۱(۲)
۸(۲)	NM	۲	۱۹	−۱۷(۲)
۹	M	۴	۱۸	−۱۴
۱۲	NM	۵	۱۷	−۱۲
۱۳	M	۶	۱۶	−۱۰
۱۳$^{+}$	M	۷	۰	۷
۱۶$^{+}$	NM	۷	۰	۷
۱۸	M	۷	۱۳	−۶
۲۳	M	۸	۱۱	−۳
۲۳	NM	۸	۱۱	−۳
۲۷	NM	۱۰	۱۰	۰
۲۸	M	۱۱		۱۱
۳	NM	۱۱	۸	۳
۳۱	M	۱۲	۷	۵
۳۳	NM	۱۳	۶	۷
۳۴	M	۱۴	۵	۹
۴۳	NM	۱۵	۴	۱۱
۴۵	NM	۱۶	۳	۱۳
۴۵$^{-}$	M	۱۷	۰	۱۷
۴۸$^{-}$	M	۱۷	۱	۱۶
۱۶۱$^{+}$	M	۱۸	۰	۱۸

(ب) و (ج) محاسبات در جدول زیر مشابه مسألهٔ ۱۳، انجام پذیرفته است. آمارهٔ استاندار MH (به جواب مسألهٔ ۱۳ توجّه کنید)، برابر با: $\frac{-۳٫۶۹}{\sqrt{۴٫۰۰۷۲}} = -۱٫۸۴$ است. بنابراین، مقدار P دوطرفه برابر ۰٫۰۶۶، است. صورت تارون آمارهٔ گهان برابر: $\frac{-۵۰}{\sqrt{۹۱۷٫۹۷}} = -۱٫۶۵$ است (جواب مسألهٔ ۱۳ را ببینید). بنابراین، مقدار P دوطرفه برابر ۰٫۰۹۹ به دست می‌آید.

z	n	$m_۱$	$n_۱$	a	$E_\circ(A)$	$n(a-E_\circ(A))$	$\frac{m_۱(n-m_۱)}{n-۱}$	$\frac{n_۱}{n}(۱-\frac{n_۱}{n})$	$n_۱(n-n_۱)$
۵	۲۳	۲	۱۱	۰	۰٫۹۵۶۵	−۲۲	۱٫۹۰۹	۰٫۲۴۹۵	۱۳۲
۸	۲۱	۲	۱۱	۰	۱٫۰۴۸	−۲۲	۱٫۹	۰٫۲۴۹۴	۱۱۰
۹	۱۹	۱	۱۱	۱	۰٫۵۷۹	۸	۱	۰٫۲۴۳۷	۸۸
۱۲	۱۸	۱	۱۰	۰	۰٫۵۵۵	−۱۰	۱	۰٫۲۴۶۹	۸۰
۱۳	۱۷	۱	۱۰	۱	۰٫۵۸۸	۷	۱	۰٫۲۴۲۲	۷۰
۱۸	۱۴	۱	۸	۱	۰٫۵۷۱	۶	۱	۰٫۲۴۴۹	۴۸
۲۳	۱۳	۲	۷	۱	۱٫۰۷۷	−۱	۱٫۸۳	۰٫۲۴۸۵	۴۲
۲۷	۱۱	۱	۶	۰	۰٫۵۴۵	−۶	۱	۰٫۲۴۷۹	۳۰
۳۰	۹	۱	۵	۰	۰٫۵۵۵	−۵	۱	۰٫۲۴۶۹	۲۰
۳۱	۸	۱	۵	۱	۰٫۶۲۵	۳	۱	۰٫۲۳۴۴	۱۵
۳۳	۷	۱	۴	۰	۰٫۵۷۱	−۴	۱	۰٫۲۴۴۹	۱۲
۳۴	۶	۱	۴	۱	۰٫۶۶۷	۲	۱	۰٫۲۲۲۲	۸
۴۳	۵	۱	۳	۰	۰٫۶	−۳	۱	۰٫۲۴۰۰	۶
۴۵	۴	۱	۳	۰	۰٫۷۵	−۳	۱	۰٫۱۸۷۵	۳
۴۸	۲	۱	۲	۱	۱٫۰	۰	۱	۰	۰

مسألهٔ ۱۶.

ثابت کنید که صورت نسخهٔ تارون-وایر آمارهٔ گهان (یعنی: $\sum n_i(a_i - E_0(A_i))$) با صورت آمارهٔ گهان (یعنی: $U = \sum\sum U_{ij}$) برابر است. بجز احتمالاً برای عامل ۱-، که به خاطر تکرارهاست.

حل:

از بخش اوّل فصل چهارم نتیجه می‌شود که:

$$U = \sum_{k=1}^{m+n} U_k^* I(k \in I_1)$$

به‌گونه‌ای که:

$$U_k^* = \sum_{\substack{\ell=1 \\ \ell \neq k}}^{m+n} U_{k\ell}$$

است. یعنی اگر مشاهده‌ای با در نمونه ۱ دارای زیرنویس k باشد، بریده می‌شود، آن‌گاه U_k^* برابر تعداد مشاهدات بریده نشدهٔ قبل از آن منهای تعداد مشاهدات بعد از آن است.

(بخش دوم از فصل چهارم را ببینید) $U_k^* = \sum_{j=1}^{k-1} m_{j1} - (n_k - m_{k1}) = \sum_{j=1}^{k} m_{j1} - n_k$

از طرف دیگر، اگر مشاهدهٔ kام بریده شده باشد، آن‌گاه U_k^* برابر تعداد مشاهدات بریدهٔ نشدهٔ قبل از آن است، یعنی: $U_k^* = \sum_{j=1}^{k} m_{j1}$. بنابراین، داریم:

$$U = \sum_{\substack{k=1 \\ c}}^{m+n} \sum_{j=1}^{k} m_{j1} I(k \in I_1) + \sum_{\substack{k=1 \\ u}}^{m+n} \left(\sum_{j=1}^{k} m_{j1} - n_k \right) I(k \in I_1)$$

که در این رابطه، c و u به ترتیب به معنی این است که، مجموع‌ها روی مشاهدات بریده شده و نشده انجام می‌شود. پس، داریم:

$$U=\sum_{k=1}^{m+n}\sum_{j=1}^{k} m_{j1} I(k\in I_1)-\sum_{k=1}^{m+n} n_k I(k\in I_1,\delta_k=1)$$

$$=\sum_{j=1}^{m+n} m_{j1}\sum_{k=j}^{m+n} I(k\in I_1)-\sum_{k=1}^{m+n} n_k a_k$$

$$=\sum_{j=1}^{m+n}(m_{j1}n_{j1}-n_j a_j)=\sum_{u}(m_{j1}n_{j1}-n_j a_j)$$

$$=\sum_{u} n_j(a_j-E_0(A_j))$$

دو تساوی آخر از این حقیقت ناشی می‌شود که m_j (در نتیجه a_j) -تعداد مشاهدات بریده نشده در z_j- برابر صفر است. به شرطی که z_j یک مشاهدهٔ بریده شده باشد. (این قرارداد را به خاطر بیاورید که تکرار بین مشاهدات بریده شده و بریده نشده توسّط این ملاحظه که مشاهدات بریده شده بزرگتراند، شکسته شده است).

مسألهٔ ۱۷.

نشان دهید که واریانس جایگشت مانتل برای آمارهٔ گهان بر $N^3=(m+n)^3$، تقسیم می‌شود. یعنی:

$$\frac{1}{N^3}\times\frac{mn}{(m+n)(m+n-1)}\sum_{i=1}^{m+n}(U_i^*)^2$$

به عبارت دیگر $\lambda(1-\lambda)\int_0^\infty (1-H(t))^2\, dH_u(t)$ همگراست.

همچنین اگر $N\to\infty$ میل کند، آن‌گاه $\frac{m}{N}\to\infty$ تحت فرض $H_0^*: F_1=F_2\ ;\ G_1=G_2$ میل خواهد کرد، به‌گونه‌ای که F و G پیوسته بوده و داشته باشیم:

$$H(t)=P\{Z\le t\}=\int_0^t (1-G(u))\, dF_u(u)+\int_0^t (1-F(u))\, dG_u(u)$$

$$H_u(t)=P\{Z\le t,\xi=1\}=\int_0^t (1-G(u))\, dF_u(u)$$

حل:

فرض کنید:

$$\hat{H}(t)=\frac{۱}{N}\sum_{i=۱}^{N} I(Z_i \le t)$$

$$\hat{H}_u(t)=\frac{۱}{N}\sum_{i=۱}^{N} I(Z_i \le t\,,\xi_i=۱)$$

باشد. آن گاه،

$$U_i^* = \begin{cases} (\text{تعداد بریده شده‌ها} < Z_i) - (\text{تعداد مشاهده} > Z_i) & \xi_i=۱ \\ (\text{تعداد بریده شده‌ها} < Z_i) - & \xi_i=۰ \end{cases}$$

$$= \begin{cases} N\hat{H}_u(Z_i-)-N(۱-\hat{H}(Z_i)) & \xi_i=۱ \\ N\hat{H}_u(Z_i-) & \xi_i=۰ \end{cases}$$

$$= N[\hat{H}_u(Z_i-)-\xi_i(۱-\hat{H}(Z_i))]$$

در نتیجه، داریم:

$$\frac{۱}{N^۳}\sum_{i=۱}^{N}(U_i^*)^۲=\frac{۱}{N^۳}\sum_{i=۱}^{N}N^۲[\hat{H}_u(Z_i-)-\xi_i(۱-\hat{H}(Z_i))]^۲$$

$$=\frac{۱}{N}\sum_{i=۱}^{N}[\hat{H}_u^۲(Z_i-)-\frac{۲}{N}\sum_{i=۱}^{N}\xi_i\,\hat{H}_u(Z_i-)((۱-\hat{H}(Z_i))+$$

$$+\frac{۱}{N}\sum_{i=۱}^{N}\xi_i(۱-\hat{H}(Z_i))^۲$$

$$=\int_۰^\infty \hat{H}_u^۲(t-)\,d\hat{H}(t)-۲\int_۰^\infty \hat{H}_u(t-)(۱-\hat{H}(t))\,d\hat{H}_u(t)+$$

$$+\int_۰^\infty (۱-\hat{H}(t))^۲\,d\hat{H}_u(t)$$

چون $\hat{H}(t)\xrightarrow{a.s.}H(t)$ و $\hat{H}_u(t)\xrightarrow{a.s.}H_u(t)$، به طور یکنواخت نسبت به t

وقتی‌که $N \to \infty$ میل کند، لذا، بنابر قضیهٔ گلیونکو-کانتلی، اگر $N \to \infty$، آن‌گاه داریم:

$$\frac{1}{N^3}\sum_{i=1}^{N}(U_i^*)^2 \xrightarrow{a.s.} \int_0^\infty H_u^2(t)\,dH(t) -$$

$$-2\int_0^\infty H_u(t)(1-H(t))\,dH_u(t) + \int_0^\infty (1-H(t))^2\,dH_u(t)$$

با انتگرال‌گیری جزء به جزء، داریم:

$$2\int_0^\infty H_u(t)(1-H(t))\,dH_u(t)$$

$$= H_u^2(t)(1-H(t))\Big|_0^\infty + \int_0^\infty H_u^2(t)\,dH(t) = \int_0^\infty H_u^2(t)\,dH(t)$$

در نتیجه، دو جملهٔ اوّل حد بالا حذف شده و همراه با:

$$\frac{mn}{(m+n)(m+n-1)} \to \lambda(1-\lambda)$$

نتیجه کامل می‌شود.

پایان

واژه‌نامه

A	
Accelerated time model	الگوی زمانی شتاب داده شده
Acturial method	روش بیمه‌گری
Asymptotic normality	نرمال مجانبی
B	
Bias-corrected estimator	برآوردگر تصحیح اریبی
Bootstrap	بوت‌استراپ
Bayesian estimate	برآوردگر بیزی
C	
Censoring	برش
Chi-square test	آزمون کی‌دو
Competing risk	نرخ شکست رقیب
Conditional likelihood	درستنمایی شرطی
Confidence interval	فاصلهٔ اطمینان
Consistency	سازگاری
Continuity correction	تصحیح پیوستگی
Cox model	الگوی کاکس
Cumulative hazard function	تابع نرخ شکست تجمّعی
Cohort life table	جدول طول عمر گروهی

Current life table	جدول طول عمر جاری
Crude probability	احتمال خام

D

Delta method	روش دلتا
Density function	تابع چگالی
Dirichlet process	فرایند دریکله
Discret data	داده‌های جدا - گسسته
Distribution function	تابع توزیع
Drop out	قطع ادامه کار
Draw back	باجایگذاری - عیب
Diagnosis	تشخیص بیماری

E

Exponential function	تابع نمایی
Exponential distribution	توزیع نمایی
Exponential density	چگالی نمایی
Extreme value function	تابع مقادیر فرین
Extreme value distribution	توزیع مقادیر فرین
Extreme value density	چگالی مقادیر فرین
Efron's test	آزمون افرون
Effective sample size	طول مؤثر نمونه

F

Force of mortality	نرخ مرگ‌ومیر
Full likelihood	درست‌نمایی کامل

G

Gamma distribution	توزیع گاما

Gamma density	چگالی گاما
Gehan test	آزمون گهان

H

Hypergeometric	فوق‌هندسی
Hazard function	تابع نرخ شکست
Hazard rate	نرخ شکست

I

IFR	تابع نرخ شکست صعودی
IFRA	تابع متوسّط نرخ شکست صعودی
Insurance	بیمه‌گری
Indentically distribution	هم‌توزیع
iid	مستقل و هم‌توزیع
Interval censoring	برش فاصله‌ای (نوع ۳) یا تصادفی
Influence function	تابع تأثیر
Itrative method	روش تکراری

J

Jack nifed method	روش جک‌نایف

K

Kaplan–Meier estimator	برآوردگر کاپلان-مایر

L

Least squares	کمترین مربعات
Linear Rank test	آزمون رتبهٔ خطّی
Log Rank test	آزمون رتبهٔ لگاریتمی
Lose to follow–up	عدم بازگشت
Left censoring	چپ برش (برش از چپ)

Logistic function	تابع لوجستیک
Log – linear model	الگوی خطّی لگاریتمی

M

Maximum likelihood (ML)	حداکثر درست‌نمایی
Maximum likelihood estimator (MLE)	برآوردگر درست‌نمایی
Mean	میانگین (حسابی)، معدّل
Median	میانه
Miller modified	تعمیم میلر
Mortality table	جدول طول عمر
Mariginal	حاشیه‌ای
Method of scoring	روش امتیازی (چوب خطّی)

N

Newton – Raphson method	روش نیوتن روفسون
Neyman – Pearson	نیمن - پیرسن
Naive estimator	برآوردگر ساده

O

Order statistices	آماره‌های مرتّب
Order populations	جامعه‌های مرتّب

P

Partial likelihood	درست‌نمایی نسبی
Permutation	جایگشت
Permutation theory	نظریهٔ جایگشت
Peterson's representation	نمایش پترسن
Plots	رسم به کمک نقطه‌یابی
Probability plots	رسم احتمالی
Probability paper	کاغذ احتمالی

Ploting position	احتمال تجربی
Poisson process	فرایند پواسن
Product – limitestimator	حدّ حاصل‌ضرب برآوردگر
Proportional hazards model	الگوهای نرخ شکست متناسب
Pivotal statistic	آمارهٔ محوری
Prior distribution	توزیع پیشین
Postorior	توزیع پسین
partial probability	احتمال جزئی

Q

Q–Q plots	رسم Q–Q
Quadratic	درجه دوم

R

Rank	رتبه
Redistribute – to – the – right	تجدید نظر در توزیع به راست
Reduced sample method	روش کاهش نمونه
Regression linear model	الگوی خطّی رگرسیون
Rao – Blackwell theorn	قضیهٔ رائو - بلکول
Restricted mean	میانگین محدود شده
Randome censoring	برش تصادفی
Right censoring	راست برش (برش از راست)
Risk	شکست - نرخ شکست
Reliability function	تابع اعتماد
Robust estimator	برآوردگر تنومند

S

Sample information matrix	ماتریس اطّلاع نمونه
Score vector	بردار امتیاز
Score method	روش امتیازی (چوب خطّی)

Self–consistency	خودسازگاری
Subdistribution function	تابع توزیع جزئی
Survival	بقاء
Survival function	تابع بقاء
Survival time	زمان بقاء
Surviving fraction	کسر بقاء
Single sample	نمونه‌ای به حجم واحد
Subsurviaval function	تابع بقاء جزئی

T

Test	آزمون
Ties	تکرار
Time dependent covarites	متغیّرهای وابسته به نرمال
Trend	روند
Trunction	قطع
Two by two tables	جدولهای ۲ × ۲
Trimed mean	پیرایش کردن میانگین
Tukey biweight estimator	برآوردگر دو وزنی توکی
Tarone–Ware class	ردهٔ تارون-وایر
Tarone–Ware generalized	تعمیم تارون-وایر

W

Weak convergence	همگرایی ضعیف
With draw	باجایگذاری
Wilks likelihood ratio	نسبت درست‌نمایی ویلکس
Weibul distribution	توزیع وایبل

U

Unbias	ناریب

منابع

The numbers in brackets after the references are the numbers of the pages on which the references are cited.

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. Scandinavian Journal of Statistics 3, 15-27. [69]

______ (1978). Nonparametric inference for a family of counting processes. Annals of Statistics 6, 701-726. [69]

Abelson, R. P. and Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. Annals of Mathematical Statistics 34, 1347-1369. [112]

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. Mathematical Biosciences 6, 1-11. [179]

Bailey, K. R. (1979). The general maximum likelihood approach to the Cox regression model. Ph.D. dissertation, University of Chicago, Chicago, Illinois. [133]

Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart, and Winston, New York. [15]

Barr, D. R. and Davidson, T. (1973). A Kolmogorov-Smirnov test for censored samples. Technometrics 15, 739-757. [173]

Basu, A. P. (1964). Estimates of reliability for some distributions useful in life testing. Technometrics 6, 215-219. [35]

Berkson, J. and Gage, R. P. (1950). Calculation of survival rates for cancer. Proceedings of the Staff Meetings of the Mayo Clinic 25, 270-286. [46]

Berman, S. M. (1963). Note on extreme values, competing risks and semi-Markov processes. Annals of Mathematical Statistics 34, 1104-1106. [179]

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York. [65]

Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57, 579-594. [109, 114]

______ (1972). Discussion on Professor Cox's paper. Journal of the Royal Statistical Society, Series B 34, 216-217. [136]

_____ (1974). Covariance analysis of censored survival data. Biometrics 30, 89-99. [136, 139]

_____ and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics 2, 437-453. [46, 65, 69]

Buckley, J. and James, I. (1979). Linear regression with censored data. Biometrika 66, 429-436. [153, 163]

Campbell, G. (1979). Nonparametric bivariate estimation with randomly censored data. Mimeoseries #79-25, Department of Statistics, Purdue University, West Lafayette, Indiana. [177]

Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics. Wiley, New York. [46, 179]

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. Technometrics 7, 579-588. [29]

Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B 34, 187-202. [127, 139]

_____ (1975). Partial likelihood. Biometrika 62, 269-276. [132]

_____ and Snell, E. J. (1968). A general definition of residuals. Journal of the Royal Statistical Society, Series B 30, 248-275. [172]

Crowley, J. (1974). Asymptotic normality of a new nonparametric statistic for use in organ transplant studies. Journal of the American Statistical Association 69, 1006-1011. [103]

______ and Hu, M. (1977). Covariance analysis of heart transplant survival data. Journal of the American Statistical Association 72, 27-36. [141, 172]

Cutler, S. J. and Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. Journal of Chronic Diseases 8, 699-712. [42, 46]

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1-22. [154]

Dufour, R. and Maag, U. R. (1978). Distribution results for modified Kolmogorov-Smirnov statistics for truncated or censored samples. Technometrics 20, 29-32. [173]

Efron, B. (1967). The two sample problem with censored data. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV. University of California Press, Berkeley, California. 831-853. [52, 57, 106]

______ (1977). The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 72, 557-565. [132]

______ (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics 7, 1-26. [182]

______ (1980). Censored data and the bootstrap. Technical Report No. 53 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [76, 182]

______ and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. Biometrika 65, 457-487. [21]

Elveback, L. (1958). Estimation of survivorship in chronic disease: The "actuarial" method. Journal of the American Statistical Association 53, 420-440. [46]

Embury, S. H., Elias, L., Heller, P. H., Hood, C. E., Greenberg, P. L. and Schrier, S. L. (1977). Remission maintenance therapy in acute myelogenous leukemia. Western Journal of Medicine 126, 267-272. [50]

Epstein, B. and Sobel, M. (1953). Life testing. Journal of the American Statistical Association 48, 486-502. [25]

Farewell, V. T. (1977). A model for a binary variable with time-censored observations. Biometrika 64, 43-46. [38]

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. Biometrics 21, 826-838. [36]

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Annals of Statistics 1, 209-230. [79]

______ and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. Annals of Statistics 7, 163-186. [79]

Fleming, T. R. and Harrington, D. P. (1979). Nonparametric estimation of the survival distribution in censored data. Unpublished manuscript. [67]

Földes, A., Rejtö, L. and Winter, B. B. (1978). Strong consistency properties of nonparametric estimators for randomly censored data. Part II: Estimation of density and failure rate. Unpublished manuscript. [76]

Gail, M. (1975). A review and critique of some models used in competing risk analysis. Biometrics 31, 209-222. [179]

Gaver, D. P., Jr. and Hoel, D. G. (1970). Comparison of certain small-sample Poisson probability estimates. Technometrics 12, 835-850. [35]

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52, 203-223. [89]

Gilbert, J. P. (1962). Random censorship. Ph.D. dissertation, University of Chicago, Chicago, Illinois. [94]

Gillespie, M. J. and Fisher, L. (1979). Confidence bands for the Kaplan-Meier survival curve estimate. Annals of Statistics 7, 920-924. [173]

Glasser, M. (1967). Exponential survival with covariance. Journal of the American Statistical Association 62, 561-568. [36]

Gong, G. (1980). Do Hodgkin's disease patients with DNCB sensitivity survive longer? Biostatistics Casebook, Vol. III, Technical Report No. 57 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [168]

Gregory, P. B., Knauer, C. M., Kempson, R. L. and Miller, R. (1976). Steroid therapy in severe viral hepatitis. New England Journal of Medicine 294, 681-686. [186, 196]

Gross, A. J. and Clark, V. A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. Wiley, New York. [20]

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. Biometrika 67, 133-143. [173]

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. Biometrics 35, 393-401. [174]

Hyde, J. (1977). Testing survival under right censoring and left truncation. Biometrika 64, 225-230. [174]

______ (1977). Life testing with incomplete observations. Technical Report No. 30 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [94]

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. Scandinavian Journal of Statistics 5, 195-199. [59]

Johns, M. V., Jr. and Lieberman, G. J. (1966). An exact asymptotically efficient confidence bound for reliability in the case of the Weibull distribution. Technometrics 8, 135-175. [32]

Kalbfleisch, J. and Prentice, R. L. (1972). Discussion on Professor Cox's paper. Journal of the Royal Statistical Society, Series B 34, 215-216. [139]

______ and ______ (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60, 267-278. [130]

——— and ——— (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York. [20, 127, 143, 146]

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481. [48, 59, 72]

Kay, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *Applied Statistics (Journal of the Royal Statistical Society, Series C)* **26**, 227-237. [172]

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics* **27**, 887-906. [59]

Korwar, R. M. (1980). Nonparametric estimation of a bivariate survivorship function with doubly censored data. Unpublished manuscript. [177]

Koul, H., Susarla, V. and Van Ryzin, J. (1979). Regression analysis with randomly right censored data. Unpublished manuscript. [155]

Koziol, J. A. and Byar, D. P. (1975). Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data. *Technometrics* **17**, 507-510. [173]

——— and Green, S. B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63**, 465-474. [173]

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics* **35**, 139-156. [179]

and Williams, J. S. (1978). Models for censored survival analysis: A cone class of variable-sum models. Biometrika 65, 181-189. [179]

Lamb, E. J. and Leurgans, S. (1979). Does adoption affect subsequent fertility? American Journal of Obstetrics and Gynecology 134, 138-144. [141]

Lamborn, K. (1969). On chi-squared goodness of fit tests for sampling from more than one population with possibly censored data. Technical Report No. 21 (T01 GM00025), Department of Statistics, Stanford University, Stanford, California. [175]

Langberg, N. A., Proschan, F. and Quinzi, A. J. (1981). Estimating dependent life lengths, with applications to the theory of competing risks. Annals of Statistics 9, 157-167. [179]

Latta, R. B. (1977). Generalized Wilcoxon statistics for the two-sample problem with censored data. Biometrika 64, 633-635. [146]

Leavitt, S. S. and Olshen, R. A. (1974). The insurance claims adjuster as patients' advocate: Quantitative impact. Report for Insurance Technology Company, Berkeley, California. [2]

Leiderman, P. H., Babu, D., Kagia, J., Kraemer, H. C. and Leiderman, G. F. (1973). African infant precocity and some social influences during the first year. Nature 242, 247-249. [7]

Leurgans, S. (1980). Does adoption affect fertility? A proportional hazards model. Biostatistics Casebook, Vol. III, Technical Report No. 57 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [141]

Lininger, L., Gail, M. H., Green, S. B. and Byar, D. P. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. Biometrika 66, 419-428. [103]

Link, C. L. (1979). Confidence intervals for the survival function using Cox's proportional hazard model with covariates. Technical Report No. 45 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [136]

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. Biometrics 23, 65-78. [89]

_____ and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22, 719-748. [103]

_____ and Myers, M. (1971). Problems of convergence of maximum likelihood iterative procedures in multiparameter situations. Journal of the American Statistical Association 66, 484-491. [36]

Marcuson, R. and Nordbrock, E. (1981). A K-sample generalization of the Gehan-Gilbert procedure for the analysis of arbitrarily censored survival data. Biometrische Zeitschrift/Biometrical Journal. [113]

Meier, P. (1975). Estimation of a distribution function from incomplete observations. Perspectives in Probability and Statistics. Papers in Honour of M. S. Bartlett (Ed. J. Gani). Academic Press, New York. 67-82. [72]

Mihalko, D. P. and Moore, D. S. (1980). Chi-square tests of fit for Type II censored data. Annals of Statistics 8, 625-644. [174]

Miller, R. G. (1974). The jackknife - a review. Biometrika 61, 1-15. [182]

______ (1975). Jackknifing censored data. Technical Report No. 14 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [182]

______ (1976). Least squares regression with censored data. Biometrika 63, 449-464. [150, 163]

Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-923. [179]

Morton, R. (1978). Regression analysis of life tables and related nonparametric tests. Biometrika 65, 329-333. [146]

Muñoz, A. (1980). Nonparametric estimation from censored bivariate observations. Technical Report No. 60 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [177]

______ (1980). Consistency of the self-consistent estimator of the distribution function from censored observations. Technical Report No. 61 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [177]

Nelson, W. (1969). Hazard plotting for incomplete failure data. Journal of Quality Technology 1, 27-52. [67, 165]

(1972). Theory and applications of hazard plotting for censored failure data. Technometrics 14, 945-966. [67, 165]

Oakes, D. (1977). The asymptotic information in censored survival data. Biometrika 64, 441-448. [132]

Peterson, A. V., Jr. (1975). Nonparametric estimation in the competing risks problem. Technical Report No. 13 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [179]

(1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proceedings of the National Academy of Sciences 73, 11-13. [179]

(1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. Journal of the American Statistical Association 72, 854-858. [63, 67]

Peto, R. (1972). Discussion on Professor Cox's paper. Journal of the Royal Statistical Society, Series B 34, 205-207. [139]

and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, Series A 135, 185-198. [146]

and Pike, M. C. (1973). Conservatism of the approximation $\Sigma(O-E)^2/E$ in the logrank test for survival data or tumor incidence data. Biometrics 29, 579-584. [117]

_____, Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. British Journal of Cancer 34, 585-612. [117]

_____, Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British Journal of Cancer 35, 1-39. [117]

Pettit, A. N. (1976). Cramér-von Mises statistics for testing normality with censored samples. Biometrika 63, 475-481. [173]

_____ (1977). Tests for the exponential distribution with censored data using Cramér-von Mises statistics. Biometrika 64, 629-632. [173]

_____ and Stephens, M. A. (1976). Modified Cramér-von Mises statistics for censored data. Biometrika 63, 291-298. [173]

Phadia, E. G. (1980). A note on empirical Bayes estimation of a distribution function based on censored data. Annals of Statistics 8, 226-229. [80]

Prentice, R. L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179. [146]

_____ and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34, 57-67. [139]

_____ and Kalbfleisch, J. D. (1979). Hazard rate models with covariates. Biometrics 35, 25-39. [127, 143]

_____, Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34, 541-554. [179]

Rai, K., Susarla, V. and Van Ryzin, J. (1980). Shrinkage estimation in nonparametric Bayesian survival analysis: A simulation study. Communications in Statistics, Simulation and Computation B9, 271-298. [79]

Rao, C. R. (1965). Linear Statistical Inference. Wiley, New York. [20, 21]

Reid, N. M. (1981). Influence functions for censored data. Annals of Statistics 9, 78-92. [73, 74, 76, 182]

_____ and Iyengar, S. (1979). Estimating the variance of the median. Unpublished notes. [76]

Sander, J. M. (1975). The weak convergence of quantiles of the product-limit estimator. Technical Report No. 5 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [76]

_____ (1975). Asymptotic normality of linear combinations of functions of order statistics with censored data. Technical Report No. 8 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California. [72, 73]

Schmee, J. and Hahn, G. J. (1979). A simple method for regression analysis with censored data. Technometrics 21, 417-432. [154]

Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 67, 145-153. [175]

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. Journal of the American Statistical Association 71, 897-902. [79]

_____ and _____ (1978a). Empirical Bayes estimation of a distribution (survival) function from right censored observations. Annals of Statistics 6, 740-754. [80]

_____ and _____ (1978b). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. Annals of Statistics 6, 755-768. [79]

_____ and _____ (1980). Large sample theory for an estimator of the mean survival time from censored samples. Annals of Statistics 8, 1001-1016. [72]

Tarone, R. E. (1975). Tests for trend in life table analysis. Biometrika 62, 679-682. [118]

_____ and Ware, J. (1977). On distribution-free tests for equality of survival distributions. Biometrika 64, 156-160. [105, 116]

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association 70, 865-871. [52]

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences 72, 20-22. [179]

_____ (1978). A heuristic estimate of the asymptotic variance of the survival probability in Cox's regression model. Technical Report No. 524, Department of Statistics, University of Wisconsin, Madison, Wisconsin. [136]

_____ (1981). A large sample study of Cox's regression model. Annals of Statistics 9, 93-108. [133, 136]

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. Journal of the American Statistical Association 69, 169-173. [7, 57]

_____ (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society, Series B 38, 290-295. [57]

_____, Brown, B. W., Jr. and Hu, M. (1974). Survivorship analysis of heart transplant data. Journal of the American Statistical Association 69, 74-80. [141]

_____ and Weiss, L. (1978). A likelihood ratio statistic for testing goodness of fit with randomly censored data. Biometrics 34, 367-375. [174]

Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika 55, 1-17. [165]

_____, Gnanadesikan, R. and Huyett, M. J. (1962). Probability plots for the gamma distribution. Technometrics 4, 1-20. [166]

Williams, J. S. and Lagakos, S. W. (1977). Models for censored survival analysis: Constant-sum and variable-sum models. Biometrika 64, 215-224. [179]

Zacks, S. and Even, M. (1966). The efficiencies in small samples of the maximum likelihood and best unbiased estimators of reliability functions. Journal of the American Statistical Association 61, 1033-1051. [35]

Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. Biometrics 22, 665-672. [36]

_____ and Lamborn, K. (1969). Concomitant variables and censored survival data in estimation of an exponential survival parameter, Part II. Technical Report No. 20 (T01 GM00025), Department of Statistics, Stanford University, Stanford, California. [36]

FERDOWSI UNIVERSITY OF MASHHAD

Publication No. 312

Survival Analysis

by

RUPERT G. MILLER, JR.

Translated by

ABOLGHASEM BOZORGNIA - HOJJAT REZAEE PAZHAND

FERDOWSI UNIVERSITY PRESS

2001