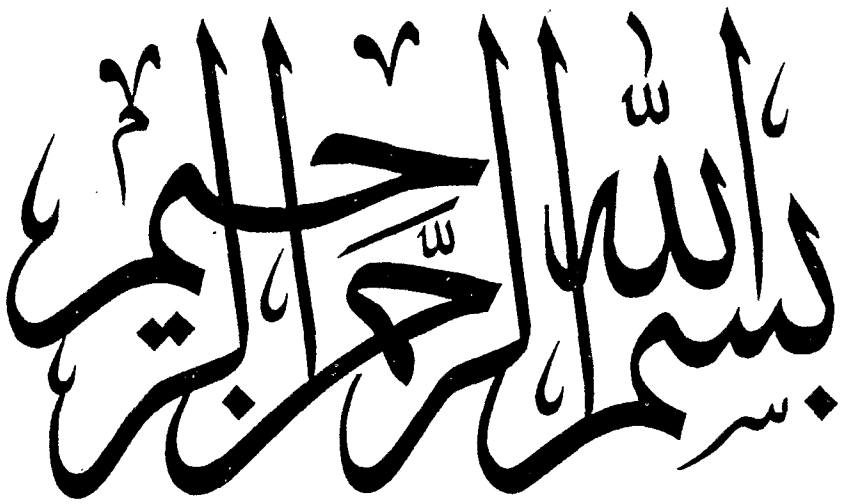
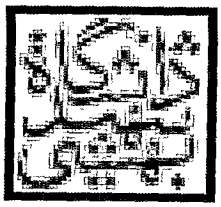


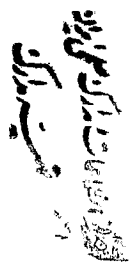


دانشگاه شهید بهشتی

دانشکده علوم ریاضی

گروه آمار

رساله دوره دکتری در رشته آمار



استنباط آماری کلاسیک و بیزی در توزیعهای آمیخته متناهی

توسط:

محمد بهرامی

استاد راهنما :

دکتر محمد رضا مشکانی

استاد مشاور :

دکتراحمد خدادادی

بهمن ماه ۱۳۸۶

چکیده

یک توزیع آمیخته ترکیبی از دو یا چند توزیع آماری است. وقتی نمونه گیری از یک جامعه غیر همگن متشکل از دو یا چند زیرجامعه هر یک با توزیعی متفاوت انجام می گیرد، با توزیعی آمیخته سروکار پیدا می کنیم. برای مثال، توزیع زمان خرابی آمیخته ای از مؤلفه های سالم و معیوب، توزیع وزن حیوانات با گروههای سنی متفاوت یا مدت زمانی که بیماران قلبی پس از یک عمل جراحی در گروههای سنی متفاوت زندگی خواهند کرد. دراین پایان نامه ابتدا توزیعهای آمیخته را درحالت کلی معرفی نموده و روشهای برآورد کردن پارامترها را دراین گونه مدلها بررسی می کنیم. سپس درحالتهای خاص برای توزیعهای آمیخته بقا مانند آمیخته دو توزیع نمایی و دو توزیع نرمال با به کارگیری روش نیوتون-رافسون و الگوریتم EM- برآوردهای ماکسیمم درست نمایی را که در واقع برآوردهای کلاسیک پارامترها می باشند محاسبه می کنیم. درادامه برآوردهای بیزی و بیزسلسله مراتبی پارامترهای موجود درتوزیع های آمیخته را با به کارگیری الگوریتم نمونه گیری گیبز به دست آورده و در حالت کلیتر با مجهول در نظرگرفتن تعداد مؤلفه ها در یک توزیع آمیخته و با استفاده از فرایند زاد و مرگ و الگوریتم مربوط به جهش برگشت پذیر برآورد های بیزی و بیز سلسله مراتبی پارامترها و تعداد مؤلفه های نامعلوم در مدل آمیخته را مشخص می کنیم. در پایان عوامل موثر بر زمان بقای بیماران مبتلا به لوسمی حاد را که از بیمارستان امید اصفهان جمع آوری شده است بررسی کرده و از آنجا که داده های مذکوریک توزیع آمیخته را نشان می داد، با برازش یک چگالی آمیخته شامل دو مؤلفه نرمال، میانگین زمان بقای این بیماران را به دست آورده ایم.

کلید واژه ها: الگوریتم- EM، برآورد بیزی، برآورد بیز سلسله مراتبی، برآورد ماکسیمم درست نمایی، بیماری لوسمی، تابع گاما، توزیع آمیخته، جهش برگشت پذیر، داده های کامل، فرایند زاد و مرگ، متغیرهای پنهان، مدل بقا، نمونه گیری گیبز.

تشکر و قدر دانی

سپاس پروردگار یکتا را که توفیق بندگی، تحصیل علم و انجام این پایان نامه را به من عطا فرمود.

اکنون که در پرتو عنایات او موفق به انجام مراحل این رساله شده ام، برخود لازم می دانم از تمامی عزیزانی که در این امر مرا یاری داده اند سپاسگزاری نمایم.

در ابتدا از استاد ارجمند جناب آقای دکتر مشکانی که مسئولیت راهنمایی اینجانب را برعهده داشته و در مراحل مختلف این پژوهش همواره از راهنماییهای ارزنده ایشان برخوردار بوده ام تشکر و قدر دانی می نمایم.

از استاد گرامی جناب آقای دکتر احمد خدادادی که در مقام استاد مشاور در این تحقیق مرا از راهنماییهای خود برخوردار ساختند، سپاسگزارم.

همچنین از آقایان دکتر محسن محمد زاده، دکتر حمید علوی مجد، دکتر عبدالرحیم شهلایی و دکترمحمد رضا فرید روحانی که زحمت داوری این رساله را تقبل کردند و با انتقادات و پیشنهادات خود موجب تقویت این رساله گردیدند، تشکر و قدردانی می نمایم.

از استاد گرامی جناب آقای دکتر محمد قاسم وحیدی اصل مدیر محترم وقت گروه آمار به خاطر بذل توجه و پیگیری های مداوم در انجام مراحل اداری کار سپاسگزارم.

تقدیم

به تنها دخترم، نازنین

فهرست مطالب

پیشگفتار

در مراکز درمانی یکی از پرسش هایی که غالباً بیماران مطرح می کنند این است که " چه آینده ای در انتظار من است؟ " به این سؤال نمی توان با اطمینان پاسخ داد، زیرا همیشه پیشگویی های لازم برای پاسخ به اینگونه سؤالات یک جزء نامعلوم دارد. معمولاً بهترین راهنمای پیشگویی ها تجربهٔ سایر بیماران است. حتی وقتی پیامد نهایی را با قدری اطمینان بتوان پیش گویی کرد، ترتیب واقعی وقایع می‌تواند بین بیماران بسیار متفاوت باشد. مثلاً وضع بیماری را در نظر بگیرید که اخیراً برای او تشخیص ایدز داده شده است. در این مورد شانس بهبودی واقعاً صفر است، در نتیجه به دوران قابل پیش بینی بقا توجه می شود. برای پاسخ به این سئوال پزشک به پژوهشهای انجام شده دربارهٔ پیشرفت بیماری ایدز مراجعه می کند. معمولاً اینگونه اطلاعات برای گروههای زیادی از بیماران جمع آوری می شود. با جمع آوری و بررسی زمان وقایع بحرانی هر بیمار (مثلاً، تاریخ تشخیص، بروز تظاهرات بیشتر بیماری و مرگ) روند پیشرفت بیماری را می توان به مراحلی تقسیم کرد.یک روش تعیین سیر طبیعی یک بیماری، برآورد متوسط دوره زمانی موسوم به زمان بقا است، که از زمان تشخیص تا مرگ بیمار ادامه دارد. به طور کلی داده های بقا، داده هایی هستند که برای اندازه گیری زمان وقوع تصادفی از یک پیشامد آغازین (مانند تولد یک انسان) تا یک پیشامد نهایی (مانند مرگ همان انسان) در نظرگرفته می شود. برای مقایسه میانگین زمانهای بقا در داده های کامل در صورت نرمال بودن داده ها یا تبدیل آنها به نرمال از روشهای متداول آماری (مانند آزمون t- یا تحلیل واریانس) استفاده می شود. اما وجه تمایز داده های بقا با سایر داده ها وجود مشاهدات سانسوریده[1] است. مثلاً ممکن است پیشامد نهایی برای بعضی از موارد تحت بررسی مشاهده نشود، که این حالت را سانسور راست می نامند. زمان لازم تا وقوع یک پیشامد یک متغیر تصادفی پیوسته مثبت با یک تابع چگالی است. البته تعریف یک نقطه شروع نیز ضروری است که

1-Censored Data

معمولاً آن را زمان صفر در نظر می گیریم. مثلاً در مورد پیشامد مرگ، زمان صفر همان لحظه تولد است.

توزیع متغیر تصادفی زمان بقا را در حالت پارامتری، معمولاً به صورت یک توزیع نمایی، گاما یا وایبول در نظر می گیرند. اما حالاتی نیز وجود دارند که در مطالعه بقا با داده‌هایی سروکار داریم که به دلیل ناهمگن بودن جمعیت آماری نمی توان با توزیع های متداول برای آنها تابع بقای مناسبی را به دست آورد. از اینرو به توزیع هایی روی می آوریم که به صورت ترکیبی خطی از دو یا چند توزیع آماری باشند. اینگونه توزیع ها همان توزیع های آمیخته اند. در این رساله استنباط آماری در باره زمانهای بقایی را که به صورت آمیخته ای از دو یا چند توزیع نمایی پارامتری و دو یا چند توزیع نرمال پارامتری باشند بررسی می کنیم. ابتدا برآورد ماکسیمم درست نمایی پارامتر های چنین مدل های آمیخته را با نوشتن برنامه هایی در نرم افزار *MATLAB* با بکارگیری الگوریتم *EM*- به دست می آوریم.

در مرحله بعد مسئله را از دیدگاه استنباط بیزی بررسی کرده و شیوه های استنباطی لازم را بسط می دهیم. به عنوان کاربردی ازروش پیشنهادی و با استفاده از نرم افزار پیشرفته *WINBUGS* برآوردهای بیزی و برآوردهای بیز سلسله مراتبی این پارامترها را در مورد داده های مربوط به زمان بقا بیماران لوسمی حادکه طی سال های۱۳۷۳ تا ۱۳۸۲ به بیمارستان امید اصفهان مراجعه کرده اند، را محاسبه می کنیم.

فصل اول

تعاریف و مقدمات

تحلیل بقا

۱-۱-مقدمه

به منظور روشن بودن تعریفها و مفهوم هایی که در فصلهای آینده بـه کـارخواهیم بـرد، در این فصل به کوتاهی به بیان تعریفها و مفهوم هایی که در تحلیل بقا به کار می روند می پردازیم. در این راستا نمادگذاریها و فرمولهای را که برای توزیعها به کار می روند بیان می کنیم و خلاصه ای از روشهای برآورد را نیز توضیح می دهیم.

۱-۲-تعریف ها و مقدمات تحلیل بقا

بقا[۱]: مدت زمان لازم از یک نقطه شروع (مانند تشخیص بیماری) تا زمـان وقـوع پیـشامدی خاص(مانند وقوع مرگ) را بقا گویند.

تابع بقا: برای محاسبهٔ احتمال بقا تا زمانی معین از تابعی استفاده می کنیم که به آن تابع بقا گفته می شود. فرض کنید T یک متغیر تـصادفی بـا مقـادیر مثبـت باشـد. اگـر F_T تـابع توزیـع متغیـر تصادفی T باشد، آنگاه تابع بقا که آن را با S_T نشان می دهیم به صورت زیر تعریف می شود.

$$S_T(t)=P(T>t)=1-F_T(t) \qquad (۱,۲,۱)$$

تابع خطر[۲]: فرض کنید f_T تابع چگالی متغیر تصادفی مثبت T باشد. تابع خطر یا نـرخ شکـست لحظه ای به شرط زنده بودن فرد تا زمان t که آن را با $h(t)$ نشان می دهیم عبارت است از:

$$h(t)=\frac{f(t)}{S(t)} \qquad (۲,۲,۱)$$

صورت دقیق تر تابع خطر به شکل زیر است:

$$h(t)=\lim_{\Delta t\to\infty}\frac{P(t<T\le t+\Delta t\,|\,t\le T)}{\Delta t} \qquad (۳,۲,۱)$$

و عبارت است از احتمال اینکه نابودی در بازه $[t,t+\Delta t)$ رخ دهد به شرط آنکه فرد یـا شـئ مـورد نظر تا زمان t زنده مانده باشد. به عبارت دیگر تابع خطر مشخص کننده احتمـال شکـست آنـی در لحظهٔ $T=t$ می باشد. برای زمان بقا پیوسته T می توان نوشت:

---

[۱] -Survival
[۲] -Hazard Function

$$h(t)=\frac{f(t)}{S(t)}=-\frac{S'(t)}{S(t)} \qquad (۴,۲,۱)$$

و با توجه به اینکه $S(۰)=۱$ داریم:

$$S(t)=\exp\left(-\int_{۰}^{t} h(u)du\right) \qquad (۵,۲,۱)$$

۱-۳ سانسور شدگی

یکی از مسائل مهم در تحلیل بقا وجود سانسورشدگی[۳] یا گمشدگی در داده هاست. بدین معنا که بعضی از افراد تحت مطالعه در طول دورهٔ مورد مطالعه ممکن است به نوعی از مطالعه خارج شوند. مثلاً ممکن است دیگر تمایلی به شرکت در آزمایش پزشکی نداشته باشند یا اینکه به دلیلی قادر به ثبت وضعیت آنها بر حسب وقوع پیشامد نباشیم. برای مثال به دلیلی غیر از پیشامد تعریف شده بیماز یا فرد مورد مطالعه فوت کند و یا اینکه بیمار یا فرد مورد نظر پس از اتمام دوره پیگیری هنوز زنده باشد و پیشامد مورد نظر برایش رخ نداده باشد. بنابراین مشخص نیست که پیامد مورد نظر برای این دسته ازافراد چه زمانی رخ می دهدویا اینکه اصلاًرخ خواهد داد.

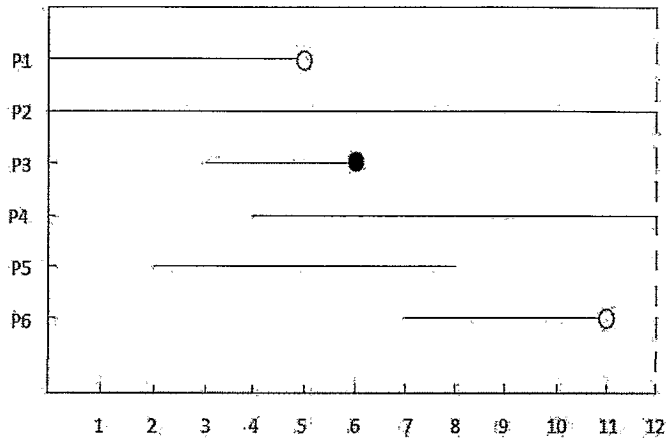
تنها چیزی که می دانیم این است که تا پایان مطالعه پیشامد هنوز اتفاق نیفتاده است. این موارد را سانسور شدگی می نامیم. برای نشان دادن سانسورشدگی برای فرد i –ام در یک نمونهٔ n تایی تحت مطالعه می توان از تابع نشانگر زیر استفاده کرد.

$$v_i=\begin{cases}1 & T_i \le c \\ 0 & T_i > c\end{cases}, \qquad i=1,2,...,n$$

اگر پیشامد در طول دوره مورد مطالعه رخ دهد، مقدار v_i برابر یک و درغیر این صورت صفر است. متغیر T_iزمان بقای مربوط به فرد i – ام مورد مطالعه و c زمان بقای سانسوریده آن می باشد. نمودار ۱-۱ به عنوان مثال وضعیت شش بیمار مبتلا به ایدز را در یک دورهٔ ۱۲ ماهه مورد مطالعه نشان می دهد. فرض کنید علامت o به منزله مرگ بیمار بر اثر بیماری ایدزو نشان دهنده مرگ بیمار براثر بیماری وعلامت • به معنی مرگ بیمار براثر سایر عوامل باشد.

---

[۳] -Censoring



نمودار ۱-۱ سا نسورشدگی در داده های بقاء

بیمار P_1: پس از ۵ ماه مطالعه بر اثر بیماری ایدز فوت شده است.

بیمار P_2: سانسور شدگی پس از ۱۲ ماه رخ داده است.

بیمار P_3: از ماه سوم وارد مطالعه شده و براثر حادثه ای غیراز بیماری ایدز در ماه ششم فوت شده که باز هم نوعی سانسورشدگی است.

بیمار P_4: از ابتدای ماه چهارم وارد مطالعه شده و تا پایان دوره، پیشامدی برای اواتفاق نیفتاده است.

بیمار P_5: از ابتدای ماه دوم وارد مطالعه شده و در پایان ماه هشتم از مطالعه خارج شده است.

بیمار P_6: از ماه هفتم وارد مطالعه شده و در پایان ماه یازدهم بر اثر بیماری ایدز فوت کرده است.

۱-۳-۱ سانسور راست و چپ

اگر از دست دادن فرد نمونه مدت زمانی پس از ورود او به مطالعه اتفاق افتد، این وضعیت را سانسورراست می گوییم، یعنی $T_i \geq t_0$ و اگر زمان بقا واقعی فرد در نمونه قبل از زمان شروع مطالعه باشد، یعنی پیشامد قبل از ورود فرد به مطالعه اتفاق افتاده باشد و او را از دست داده باشیم، به آن سانسورچپ می گوییم یعنی: $T_i < t_0$. در نمودار ۱-۱ مشاهدات مربوط به بیماران P_2، P_3، P_4 و P_5 همگی سانسور راست هستند.

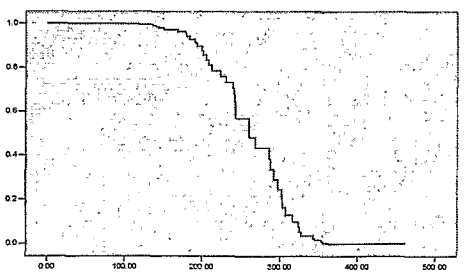
۱-۳-۲ سانسور بازه ای

این سانسورشدگی تلفیقی از سانسورشدگی راست و چپ است و در حقیقت حالتی است که پیشامد مورد نظر در یک بازهٔ زمانی مثلاً از t_a تا t_b رخ می دهد. بنابراین $t_a < T_i < t_b$ که در آن T_i زمان رخ دادن پیشامد برای فرد $i-$ام است.

۱-۴ تابع بقای تجربی

تابع بقاء $S_T(t)$ نسبت افرادی در نمونه را نشان می دهد که از مبدأ زمانی خود که همان لحظهٔ ورودشان به مطالعه است تا t واحد زمانی که پس از آن زنده مانده اند. مقدار آن در لحظهٔ t برابر است با نسبت تعداد افرادی که در زمان t زنده مانده اند به تعداد کل افراد شرکت کننده در مطالعه. در شروع مطالعه هنوز هیچ اتفاقی رخ نداده است اما با گذشت زمان پیشامد ها یکی پس از دیگری رخ می دهند. اگر مدت زمان مطالعه نامتناهی باشد، بالاخره هیچ فردی زنده نخواهد ماند، یعنی خم بقا نهایتاً به صفر می گراید. امّا چون در عمل مدت زمان مطالعه متناهی است، احتمالاً در پایان دوره هنوز تعدادی از افراد وجود دارند که پیشامد مورد نظر برای آنها اتفاق نیفتاده است و سانسور نشده اند.

از سوی دیگر چون تابع بقا بین دو شکست متوالی مقداری ثابت است، پس از مشاهده شکست در زمان t، بلافاصله کاهش می یابد و تابع پله ای به دست می آید. این تابع پله ای در واقع همان تابع بقا تجربی[۴] است که آن را با $\hat{S}(t)$ نشان می دهیم و نمودار آن در شکل ۱-۴ رسم شده است.



شکل۱-۴- نمودار تابع بقا تجربی در دورهٔ مورد مطالعه

[۴] -Empirical Survival Function

۱-۵ مدل رگرسیون خطرات متناسب کاکس

مانند تمام مباحث در استنباط آماری، در تحلیل بقا نیز فرض بر این است که زمان بقا برای تمام افراد یا اشیاء مورد آزمایش متغیرهای تصادفی مستقل و هم توزیع هستند. اما علاوه بر زمان بقا هر واحد، تعدادی متغیر دیگر نیز اندازه گیری می شوند که این متغیرها می توانند بر زمان بقا هر واحد تأثیر داشته باشند. این متغیرهارا متغیرهای تبیینی[۵] می نامیم. درمسئلهٔ تحلیل بقا مهمترین موضوعی که به آن توجه می شود تابع خطراست که آن را با $h(t)$ نشان دادیم. واضح است که اگر زمان بقا از توزیعی خاص پیروی کند که به بردار پارامتری $\underset{\sim}{\theta}$ بستگی داشته باشد، آنگاه تابع خطر نیز تابعی از پارامتر $\underset{\sim}{\theta}$ به صورت $h(t,\underset{\sim}{\theta})$ است.

چون زمان بقا یعنی T یک متغیر تصادفی پیوسته با مقادیر غیر منفی است، باید مدلهایی را در نظر گرفت که در آنها متغیرغیر منفی باشد، مانند توزیع نمایی، گاما و وایبول. البته در بسیاری از موارد نمی توان یک مدل پارامتری مناسب برای داده های بقا تعیین کرد، زیرا چولگی داده های بقا گریز ناپذیراست. مدل خطرات متناسب در واقع یک مدل نیمه پارامتری در تحلیل داده های بقا می باشد. یعنی علاوه بر پارامترهای توزیع، فرض می کنیم تابع خطر به متغیرهای تبیینی مانند $\underset{\sim}{Z}=(Z_1,...,Z_K)'$ نیز بستگی داشته باشد، یعنی فرم تابع خطر به صورت $h\ (t\ ;\ \underset{\sim}{Z}\ ,\ \underset{\sim}{\theta}\ )$ باشد. یکی از مهمترین مدلهای نیمه پارامتری در تحلیل داده های بقا مدل کاکس(۱۹۷۲) است، که به آن مدل خطرات متناسب کاکس[۶] گفته می شود. در مدل کاکس متغیرهای تبیینی اثر ضربی بر روی تابع خطر دارد. یعنی :

$$h(t;\underset{\sim}{Z})=h_{\circ}(t)\psi(\underset{\sim}{Z},\underset{\sim}{\beta}) \qquad (۱,۵,۱)$$

که در آن $h_{\circ}(t)$را تابع خطر پایه یا مبنا می نامیم و به متغیرهای تبیینی بستگی ندارد. چنانچه داشته باشیم : $\underset{\sim}{Z}=\underset{\sim}{0}$، آن را شرط استاندارد می نامیم و به واحد آماری (فرد یا شئ) که برای آن $\underset{\sim}{Z}=\underset{\sim}{0}$ است، مولفه استاندارد گفته می شود. بردار $\underset{\sim}{\beta}$، بردار پارامترهای مجهول است که باید

---

[۵] - Explanatory Variables
[۶] - Cox Proportional Hazard Model

برآورد[۷] گردد و $\psi = (\underset{\sim}{Z}, \underset{\sim}{\beta})$ نیز تابعی از متغیرهای توضیحی است که به زمان بقا t بستگی ندارد. در ابتدا فرض بر این است که متغیرهای تبیینی $\underset{\sim}{Z}$ در طول زمان ثابت اند و تغییر نمی کنند. اما در مسئله کلی تر به این موضوع اشاره می شود که ممکن است برخی از متغیرهای تبیینی در طول زمان ثابت نمانند و بنابراین در تجزیه و تحلیل مؤثر باشند.

در بکارگیری مدل کاکس لازم است که نسبت خطر هر دو واحدی که در نمونه واقع اند در طول زمان ثابت باشد. فرض کنید $\underset{\sim}{Z}_1, \underset{\sim}{Z}_2$ بردار متغیرهای تبیینی برای دو واحد مورد نظر باشند. اگر فرض کنیم نسبت خطرات این دو واحد HR است آنگاه برآورد این نسبت در مدل خطرات متناسب عبارت است از:

$$\hat{HR} = \frac{\hat{h}(t;\underset{\sim}{Z}_1)}{\hat{h}(t;\underset{\sim}{Z}_2)} = \frac{\hat{h}_\circ(t).\psi(\underset{\sim}{Z}_1, \underset{\sim}{\beta})}{\hat{h}_\circ(t).\psi(\underset{\sim}{Z}_2, \underset{\sim}{\beta})} = \frac{\psi(\underset{\sim}{Z}_1, \underset{\sim}{\beta})}{\psi(\underset{\sim}{Z}_2, \underset{\sim}{\beta})} \qquad (۲,۵,۱)$$

همان گونه که در(۲,۵,۱) ملاحظه می شود، این نسبت به t بستگی ندارد و نشان می دهدکه تابع خطرمربوط به واحد اول متناسب با تابع خطر واحد دوم به صورت: $\hat{h}(t;\underset{\sim}{z}_1) = \hat{k}.\hat{h}(t;\underset{\sim}{z}_2)$ می باشد. یک فرم بسیار ساده برای $\psi = (\underset{\sim}{z}, \underset{\sim}{\beta})$ عبارت است از:

$$\psi(\underset{\sim}{Z}, \underset{\sim}{\beta}) = \exp(Z_1\beta_1 + Z_2\beta_2 + ... + Z_k\beta_k) \quad ; k \geq 1 \qquad (۳,۵,۱)$$

اگر $\underset{\sim}{Z} = \underset{\sim}{0}$ باشد آنگاه $\psi(\underset{\sim}{0}, \underset{\sim}{\beta}) = 1$ و واحد در حالت استاندارد قرار دارد.

اگر $\underset{\sim}{\beta} = \underset{\sim}{0}$ آنگاه $\psi(\underset{\sim}{Z}, \underset{\sim}{0}) = 1$ و بدین مفهوم است که متغیرهای تبیینی تأثیری بر مقدار خطر ندارند.

۱-۵-۱ برآورد پارامترها در تحلیل بقا

یکی از روشهای متداول و نسبتاً کارا در برآورد پارامترهای مدلهای بقا روش ماکسیمم درست نمایی است. برآوردهای ماکسیمم درست نمایی دارای خاصیت ناوردایی[۸] بوده وتحت شرایط نظم عمومی بطور مجانبی کاراترین برآوردگر می باشند و توزیع مجانبی آنها نیز نرمال است (مود و

---

[۷] -Estimate
[۸] -Invariant Property

همکاران۱۹۷۴). با داشتن مقادیر p متغیر تبیینی و انتخاب یک نمونه n تایی می توان n ردیف مشاهده به صورت بردار $p+1$ –بعدی $(t_i, Z_{1i}, Z_{i2}, ..., Z_{ip})'$ را داشته باشیم که t_i زمان شکست i – امین مؤلفه می باشد. از اطلاعات فوق می توان برای برآورد پارامترهای $\underset{\sim}{\beta}$ استفاده کرد. فرض کنید زمانهای شکست $a_1 < a_2 < \cdots < a_n$ باشند. اگر برای واحدی در زمان a_j شکست رخ دهد آن را I_j می نامیم. بنابراین $I_j = i$ اگر و تنها اگر $t_i = a_j$. مجموع واحدهای در معرض خطر در زمان a_j را با $R(a_j)$ نشان می دهیم. در برخی موارد به آن مجموعه ریسک نیز گفته شده و داریم $R(a_j) = \{i \mid t_i \geq a_j\}$. تعداد واحد های موجود در $R(a_j)$ را با r_j نشان می دهیم. چون از $h_0(t)$ اطلاعاتی در دسترس نیست، بنابراین زمانهای a_j به تنهایی هیچ گونه اطلاعاتی از مقادیر $\underset{\sim}{\beta}$ را نمی دهند. پس اطلاعات مربوط به مقادیر $\underset{\sim}{\beta}$ را از I_j ها به دست می آوریم. مثلاً در زمان a_j زمانهای قبلی $a_1, a_2, \cdots, a_{j-1}$ معلوم بوده و می دانیم در هر یک از این زمانها برای کدام واحدپیشامد روی داده است. بنابراین مجموعه ریسک معلوم می‌شود. اکنون با شرط این که بدانیم برای یک واحد این مجموعه در زمان a_j پیشامدی رخ داده است، احتمال شرطی اینکه واحد مورد نظر، واحد i – ام باشد به صورت زیر به دست می آید.

$$\frac{h(a_j, \underset{\sim}{z}_i)}{\sum_{k \in R(a_j)} h(a_j, \underset{\sim}{z}_k)} = \frac{\psi(\underset{\sim}{z}_i, \underset{\sim}{\beta})}{\sum_{k \in R(a_j)} \psi(\underset{\sim}{z}_k, \underset{\sim}{\beta})} \qquad (۴,۵,۱)$$

همان گونه که در (۴,۵,۱) مشاهده می شود تابع خطر مبنا یعنی $h_0(a_j)$ حذف گردیده، بنابراین توزیع توأم تابع نشانگرهای I_j که در واقع همان تابع درست نمایی می باشد، به صورت زیر به دست خواهد آمد:

$$L(\underset{\sim}{\beta}) = \prod_{j=1}^{n} \frac{\psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta})}{\sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta})} \qquad (۵,۵,۱)$$

اکنون می توانیم با استفاده از (۵,۵,۱) برآوردهای ماکسیمم درست نمایی برای بردار پارامتری $\underset{\sim}{\beta}$ را به دست آوریم. ابتدا از (۵,۵,۱) لگاریتم می گیریم، داریم .

$$l(\underset{\sim}{\beta}) = \log L(\underset{\sim}{\beta}) = \sum_{j=1}^{n} \log \psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta}) - \sum_{j=1}^{n} \log\Big(\sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta}) \Big) \qquad (۶,۵,۱)$$

حال نسبت به β_P از $\ell(\underset{\sim}{\beta})$ مشتق می گیریم. بنابراین می توان نوشت:

$$\frac{\partial \ell(\underset{\sim}{\beta})}{\partial \beta_p} = \sum_{j=1}^{n} \frac{\frac{\partial}{\partial \beta_p} \psi(\underset{\sim}{z}_j, \underset{\sim}{\beta})}{\psi(\underset{\sim}{z}_j, \underset{\sim}{\beta})} - \sum_{j=1}^{n} \frac{\frac{\partial}{\partial \beta_p} \sum_{k \in R(a_j)} \psi(\underset{\sim}{z}_k, \underset{\sim}{\beta})}{\sum_{k \in R(a_j)} \psi(\underset{\sim}{z}_k, \underset{\sim}{\beta})} \qquad (۷,۵,۱)$$

همچنین داریم:

$$\frac{\partial^2 \ell(\underset{\sim}{\beta})}{\partial \beta_p \partial \beta_S} = \sum_{j=1}^{n} \frac{\left[\frac{\partial^2}{\partial \beta_p \partial \beta_S} \psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta})\right] \psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta}) - \frac{\partial^2}{\partial \beta_p} \psi(Z_j, \beta) . \frac{\partial^2}{\partial \beta_S} \psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta})}{\left(\psi(\underset{\sim}{Z}_j, \underset{\sim}{\beta})\right)}$$
$$- \sum_{j=1}^{n} \frac{\left[\frac{\partial^2}{\partial \beta_p \partial \beta_S} \sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta})\right] \left[\sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta})\right] - \frac{\partial^2}{\partial \beta_p} \sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \beta) . \frac{\partial^2}{\partial \beta_S} \sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta})}{\left(\sum_{k \in R(a_j)} \psi(\underset{\sim}{Z}_k, \underset{\sim}{\beta})\right)^2} \qquad (۸,۵,۱)$$

اکنون دسته معادلات (۷,۵,۱) را مساوی صفر قرار داده و برآوردهای مربوط به بردار پارامتری $\underset{\sim}{\beta}$ را به دست آورده و با بکارگیری معادله (۸,۵,۱) شرط ماکسیمم شدن درست نمایی را بررسی می کنیم. البته واضح است که این کار به وسیله روشهای عددی امکان پذیر است.

۱-۶- مدل های پارامتری

فرض کنید داده‌های بقا از یک توزیع خاص تبعیت کنندو T زمان بقا یا طول عمر فردی با توزیعی مانند $F_T(t|\underset{\sim}{\theta})$ و چگالی $f_T(t|\underset{\sim}{\theta})$ باشد، که در آنها $\underset{\sim}{\theta} = (\theta_1, \theta_2, ..., \theta_h)$ یک بردار پارامتری مجهول است و باید از روی داده‌های بقا برآورد گردد. در این بخش برآوردهای ماکسیمم درستنمایی و برآوردهای بیزی بردار پارامتری مجهول را برای این توزیعها بدست می‌آوریم.

تابع بقا را با $S(t|\underset{\sim}{\theta})$ نشان می‌دهیم و عبارت است از:

$$S(t|\underset{\sim}{\theta}) = P(T>t) = 1 - F(t|\underset{\sim}{\theta}) \qquad (۱,۶,۱)$$

در بسیاری از موارد T از یک توزیع خاص تبعیت می‌کند.