سیر تاریخی روشهای تحلیل بقاء

اولین دیدگاههای تئوریک برای تحلیل بقاء که سریعا" توسعه یافت و ویژگی های کاربردی خویش را نشان داد مبتنی است بر روش های کلاسیک موسوم به " جدول طول عمر " که توسط *کاپلان*[1] و *مایر*[2] در سال ۱۹۵۸ ارائه شد. آنها در بررسی خویش با توجه به داده های ناتمام موجود در مطالعات بقاء، روشی ناپارامتری را برای برآورد احتمالات بقاء ارائه نمودند.

این روشها سریعا" مورد توجه سایر آمارشناسان قرار گرفت. ازجمله این افراد *اسوردراپ*[3] بود که در سال ۱۹۶۵، نه تنها از این روش استفاده نمود بلکه برای اولین بار ضعف های این روش را نیز مطرح نموده و استفاده از مدلهای مبتنی بر فرآیندهای تصادفی را پیشنهاد نمود. او اعلام کرد افراد در طول زندگی خویش تا مرگ از مراحل مختلفی از سطوح سلامتی عبور می نمایند، که زمان توقف هر فرد در هر سطح از سلامتی تصادفی بوده و طول عمر افراد از دنباله ای از فرآیند های تصادفی تشکیل شده است.

به دنبال *اسوردراپ*، *هوئم*[4] نیز در سال ۱۹۶۹ طی مقاله معروف خود ایده هایي را درمورد استفاده از فرآیندهای تصادفی در مدلهای تحلیل بقاء مطرح نمود. او برای اولین بار امکان استفاده از فرآیندهای " زنجیر مارکف[5] " را اعلام نمود.

اما *نلسون*[6] در سال ۱۹۶۹، با ارائه مباحث تئوریک خود در مورد "تابع مخاطره[7] " هم به توسعة مدل *کاپلان _ مایر* پرداخت و هم باعث شد که برای مدتی بحث استفاده از فرآیندهای تصادفی در مطالعات بقاء به حاشیه رانده شود. او با تئوریزه نمودن ارتباطات بین توابع بقاء و مخاطره، محاسبه احتمالات بقاء در زمانهای مختلفی از عمر را امکان پذیر نمود. اهمیت کار او آنجایي بود که نقش ناتمام بودن برخی از داده ها را در مدل نشان داد و اثر چنین داده هایي را در مدل لحاظ نمود.

آلتشالر[8] در مقاله ای برای اولین بار در مورد رقابت جویي مخاطرات[9] سخن گفت، و در این مسیر از روش *نلسون* استفاده نمود. البته او در این تحقیق که براساس داده های حاصل از یک مطالعه تجربی بر روی حیوانات بود، توجه چندانی به داده های ناتمام نداشت.

در سال ۱۹۷۲، *آلن*[10] در پایان نامه فوق لیسانس خود برآوردگری را برای تابع مخاطره ارائه نمود که بعدها به برآوردگر "*نلسون _ آلن* " معروف شد. بعدا" نشان داده شد که این برآوردگر برای زمانهای بزرگتر از بزرگترین زمان مشاهده شده چندان مناسب نیست، زیرا اگر در t_{max} تمام موارد با مرگ روبرو شوند مقدار بقاء صفر شده و اگر ناتمام در نظر گرفته شوند تابع بقاء قابل محاسبه نخواهد بود.

---

[1] E.L. Kaplan
[2] P. Meier
[3] E. Sverdrup
[4] J.M. Hoem
[5] Markov chain process
[6] W. Nelson
[7] Hazard function
[8] B. Altshuler
[9] Competing risk
[10] O.O. Aalen

در این سال ها اقدامات گسترده ای نیز در زمینه استفاده از روش های ناپارامتری صورت گرفت. از جملة آنها چگونگی استفاده از " آزمون ویلکاکسون تعمیم یافته " برای مقایسه داده های شامل داده های ناتمام است، که توسط *گهان*[11] ارائه شد. *منتل*[12] نیز کاربرد مدل های ناپارامتری را در تحلیل بقاء با کاربست آنها در تعیین طول عمر مبتلایان به بیماری سرطان نشان داد. همچنین *افرون*، نیز ضرورت استفاده از مدل های ناپارامتری را در تحلیل بقاء در مقالاتی شرح داد و راه هایي برای رفع مشکلات موجود در مطالعات بقاء ارائه نمود. بویژه، او برای رفع مشکلات موجود در کاربرد برآوردگر *نلسون _ آلن* پیشنهاداتی ارائه کرد، هرچند نتیجة کاربرد این پیشنهاد نیز بروز اریبي منفی خواهد بود. *برسلو*[13] نیز روشی را برای مقایسه طول عمر k نمونه مختلف که الگوی ناتمامی داده ها در آنها متفاوت است را براساس تعمیم آزمون "*کروسکال _ والیس* " ارائه نمود. *پتو*[14] پیشنهاداتی برای کاربرد مدلهای تغییر ناپذیر[15] ارائه نمود، و نشان داد که کاربرد چنین روشهایي بطور مجانبی کارا خواهند بود.

در همین سالها *کاکس*، مدل رگرسیونی که بعدها بنام خود او معروف شد را ارائه نمود، که تمامی ایده های مطرح شده تا آن زمان را تحت الشعاع خویش قرار داد. با کاربرد این مدل رگرسیونی امکان بررسی اثر متغیرهای مختلف بر طول عمر بطور همزمان فراهم گردید. مدلهای مختلفی که قبلا" ارائه شده بودند تنها این امکان را فراهم می ساختند که طول عمر در دو یا بیش از دو گروه مقایسه شوند، اما مدل رگرسیونی *کاکس* به محققین اجازه داد که اثر چند متغیر را همزمان بر طول عمر بررسی کرده و اثر داده های ناتمام نیز در این مدل اِعمال می شد. مدل مذکور نیز بر مبنای پیش فرضهاي مختلفی از جمله این که "میزان مخاطرات شرطی هر فرد با بردار متغیرهای مستقل $\mathbf{Z}$ برابر است با حاصلضرب میزان مخاطرة پایه $h_0(x)$ و تابعی نامنفی از متغیرهای مستقل به نام $C(\cdot)$" استوار گردیده، و حتی *کاکس* پیشنهاد کرده برای راحتی تابع $C(\cdot)$ را نمایي درنظر بگیریم. این مدل چندی بعد توسط خود او توسعه یافت و چگونگی برآورد ضرائب مدل رگرسیونی با روشهای دقیقتر از جمله روش درستنمایي جزئی[16] تشریح گردید. در مدل توسعه یافته *کاکس* بررسی اثر متغیرهای کمی وکیفی و نیز متغیرهای وابسته به زمان بر احتمالات بقاء امکان پذیر است.

با ارائه مدل *کاکس*، مطالعات بسیاری از آمارشناسان بر این روش و چگونگی برطرف کردن ضعف های آن متمرکز شد. اکثر پیشنهاداتی که ارائه می شد با توجه به ناتمامی داده ها و نیز حضور متغیرهای کیفی در مدل به عنوان متغیرهای مستقل، براساس استفاده از آزمونها و روشهای ناپارامتری بود. *پرنتایس*[17] روشی ناپارامتری جهت استفاده در مدلهای رگرسیونی با داده های ناتمام از راست ارائه نمود. *کازیک*[18] در دو مقاله ای که منتشر کرد چگونگی استفاده از آزمون رتبه ها [19] را تشریح کرد. یکی از مقالات ارزشمندی که در این سالها ارائه شد، مقاله ای از *کلایتون*[20] و *کازیک* بود که در آن چگونگی استفاده از مدلهای تعمیم یافته چند متغیره در مدلهای مخاطرات متناسب تئوریزه شده است. تأثیر داده های ناتمام

---

[11] E.A. Gehan
[12] N. Mantel
[13] N.E. Breslow
[14] R. Peto
[15] Invariant
[16] Partial likelihood
[17] R.L. Prentice
[18] J. Cuzick
[19] Rank test
[20] D.G. Clyton

بر این مدل توسط *کالب فلیج*[21] و *پرتنایس* بررسی و مدل تکمیل گردید. آنها تمام ایده هایي که تا آن زمان در مورد روش تجزیه و تحلیل بقاء ارائه شده و از بنیانهای تئوریک مناسبی برخوردار بودند در کتاب خویش منتشر ساختند. این کتاب هنوز یکی از مراجع اصلی در مطالعات بقاء محسوب می شود.

استفاده از فرآیندهاي تصادفي در تحلیل بقاء ، از سال ۱۹۷۵ و توسط آلن مرسوم شد . او بااستفاده از مارتینگلهاي پیوسته_ زمان به بررسي جداول طول عمر پرداخت. تئوري مارتینگلهاي پیوسته _ زمان ، تازه چند سالي بود که توسط دانشمندان علم آمار در شاخة فرآیندهاي تصادفي ارائه شده بود ، که از جمله این افراد *ماير*[22] و *کونیتا* و *واتانابه*[23] بودند .

آلن در سال ۱۹۷۸ درادامة رسالة دکتراي خویش این روشها را توسعه داد ، و با کمك فرآیند هـاي شمارشـي و برآوردگـر *نلسون –آلن* مدلي ساخت و از آن با کمـك انتگرالهـاي تـصادفي و تئـوري مارتینگلهـا در آزمونهـاي مقایـسه دو نمونـه اي استفاده نمود . او در مقالة بسیار معتبر خویش چگونگي برآورد نمودن احتمالات انتقال از هر دورة زمـاني بـه دورة بعـد را بـر اساس روشهاي ناپارامتري ارائه نمود.

او در ابتداي کار وقتي مي خواست از فرآیندهاي شمارشي استفاده نماید ، آنها را همگن[24] در نظر گرفته بود . براي حذف این پیش فرض غیر ضروري او با کمك *جانسن*[25] در سال ۱۹۷۸ بر اساس فرآیندهاي مارکفي مـدلي بـراي داده هـاي بقـاء (شامل موارد ناتمام) با استفاده از تبدیلات کانوني ارائه نمودند و چگونگي برآورد احتمالات بقاء را در یك ماتریس انتقال[26] براي زنجیرهاي مارکف غیر همگن تشریح کردند.

تقریبا" همزمان با *آلن* و *جانسن* ، اما جدا از آنها *فلمینگ*[27] روشي براي برآورد ناپارامتري مخاطرات همزمان بـا اسـتفاده از فرآیندهاي مارکفي غیر همگن ارائه نمود.

گیل[28] ، که در چارچوب مطالعات مربوط به رسالة دکتراي خویش بر تحلیل بقاء تمرکز یافته بود نشان داد کـه اسـتفاده از فرآیندهاي تصادفي در تحلیل بقاء ما را به نتایج دقیقتري رهنمون میشود.

آرژاس[29] و *هارا*[30] چگونگي بررسي اثر متغیرهاي مستقل بر طول عمر را در مدلهاي مبتنـي بـر فرآینـدهاي تـصادفي نـشان دادند.

آرژاس بعدا" اعلام و ثابت نمود که دقیقترین نتایج در تحلیل بقاء در مدلهاي مبتني بر فرآینـد هـاي تـصادفي اسـت . او بـه علت نوع عملکرد این مدلها آنها را پویا[31] نامید.

سرانجام در سـال ۱۹۹۱ ، دو تـیم یکـي شـامل *فلمینـگ و هـارینگتون*[32] و دیگـري شـامل *آندرسـن* ، *بورگـان*[33] ، *گیـل* و *کیدینگ*[34] در سال ۱۹۹۲ چگونگي استفاده از فرآیندهاي تصادفي در تحلیل بقاء را در دو کتاب تئوریزه نمودند . البته محـور

---

[21] J.D. Kalbflisch
[22] Meyer , 1966
[23] Kunita , Watanabe , 1967
[24] Homogeneous
[25] S. Johnsen
[26] Transition Matrix
[27] T.R. Feleming
[28] R.D. Gill
[29] E. Arjas
[30] P. Haara
[31] Dynamic
[32] Harington
[33] Borgan

کار گروه *فلمینگ – هارینگتون* بیشتر استفاده از فرآیندهاي شمارشي بوده و اساس کـار گـروه دیگـر اسـتفاده از مارتینگلهـا است.

در دهة اخیر مطالعات تئوریك در زمینة تحلیل بقاء در دو شاخة اصلي انجام میشود . *گروه اوّل* ، مطالعاتي اسـت کـه مـدل رگرسیوني کاکس را براي نقطة شروع پذیرفته و تلاش میکنند راههایي را براي افزایش کارآمدی نتایج حاصـل از ایـن مـدل بیابند . در چند سال گذشته استفاده از برآوردگرهاي بیزي ، و استفاده از الگوریتمهاي مختلف براي کاهش واریانس در این رابطه بسیار مرسوم شده است . دلیل تلاش پژوهشگران براي بهبود مدل *کاکس* ، سادگي کاربرد آن و استنتاج ساده از نتـایج حاصل از آن است . اما *گروه دوّم* ، که مطالعات خویش را بر مبناي استفاده از فرایندهاي تصادفي گذارده اند ، تلاش میکننـد با توجه به رفتار پدیده هاي مورد بررسي با استفاده از مدلهاي مختلف فرآیندي مدلهاي جدید تـري را ارائـه نمـوده و کـاربرد مدلهاي موجود را ساده تر نمایند .

[34] Keiding