دانشکده علوم ریاضی و آمار

گروه آمار

**پایان‌نامه کارشناسی ارشد**
**آمار ، گرایش آمار ریاضی**

**عنوان**

# استنباط آماری تحت سانسور فزاینده بازه ای نوع $I$

استاد راهنما

**دکتر فاطمه یوسف زاده**

استاد مشاور

**دکتر مجید رضایی**

نگارنده

**افسانه قلی پور**

شهریور ۱۳۹۱

# چکیده

در آزمون های طول عمر صنعتی یا پزشکی موارد زیادی وجود دارد که زمان دقیق شکست مولفه ای مشخص نباشد و یا طول عمر آن مولفه در یک بازه معلوم باشد. در این حالت با سانسور بازه ای نوع ۱ مواجه هستیم که فقط تعداد شکست ها در بازه زمانی از پیش تعیین شده ثبت می شود. اگر در این بازه های زمانی تعدادی از واحدهای آزمایشی به طور تصادفی از آزمایش حذف گردند، در این صورت سانسور فزاینده بازه ای نوع ۱ رخ داده است.
برآورد پارامترها به وسیله روش های مختلف از جمله روش های درستنمایی ماکسیمم، تقریب نقطه-میانی، الگوریتم EM، گشتاوری، نمودار احتمال و درستنمایی ماکسیمم تقریبی (AML) در توزیع های مختلف طول عمر از قبیل توزیع نمایی تعمیم یافته (GE) ، توزیع گامپرتز و توزیع بئر نوع $XII$ تحت سانسور فزاینده بازه ای نوع $I$ مورد بررسی قرار می گیرد. با استفاده از مطالعات شبیه سازی روش های مختلف برآورد را بر حسب میانگین، میانه، اریبی و میانگین مربعات خطا (MSE) مقایسه می کنیم. سرانجام، به منظور نشان دادن کاربرد، روش های برآورد بر روی داده های حقیقی در توزیع نمایی تعمیم یافته و بر روی داده های شبیه سازی شده در توزیع های گامپرتز و بئر نوع $XII$ مورد بررسی و تحقیق قرار می گیرند.

واژگان کلیدی: اریبی، الگوریتم EM ، توزیع بئر نوع $XII$، توزیع گامپرتز، توزیع نمایی تعمیم یافته، روش درستنمایی ماکسیمم تقریبی، روش تقریب نقطه-میانی، روش گشتاوری، روش نمودار احتمال، سانسور فزاینده بازه ای نوع $I$، میانگین مربعات خطا.
تعداد صفحات پایان نامه: ۱۲۱

# خدایا...[1]

به من زیستنی عطا کن که در لحظه مرگ، بر بی‌ثمری لحظه‌ای که برای زیستن گذشته است، حسرت نخورم و مُردنی عطا کن که بر بیهودگیش، سوگوار نباشم. بگذار تا آن را، خود انتخاب کنم، اما آنچنان که تو دوست می‌داری.

تو می‌دانی و همه می‌دانند که شکنجه دیدن بخاطر تو، زندانی کشیدن بخاطر تو و رنج بردن به پای تو تنها لذت بزرگ زندگی من است، از شادی توست که من در دل می‌خندم، از امید رهایی توست که برق امید در چشمان خسته‌ام می‌درخشد و از خوشبختی توست که هوای پاک سعادت را در ریه‌هایم احساس می‌کنم. نمی‌توانم خوب حرف بزنم. نیروی شگفتی را که در زیر کلمات ساده و جمله‌های ضعیف و افتاده، پنهان کرده‌ام دریاب، دریاب.

تو می‌دانی و همه می‌دانند که زندگی از تحمیل لبخندی بر لبان من، از آوردن برق امیدی در نگاه من، از برانگیختن موج شعفی در دل من، عاجز است.

تو، چگونه زیستن را به من بیاموز، چگونه مردن را خود خواهم آموخت.

به من توفیق تلاش در شکست، صبر در نومیدی، رفتن بی‌همراه، جهاد بی‌سلاح، کار بی‌پاداش، فداکاری در سکوت، دین بی‌دنیا، مذهب بی‌عوام، عظمت بی‌نام، خدمت بی‌نان، ایمان بی‌ریا، خوبی بی‌نمود، گستاخی بی‌خامی، قناعت بی‌غرور، عشق بی‌هوس، تنهایی در انبوه جمعیت، و دوست داشتن بی‌آنکه دوست بداند، روزی کن.

اگر تنهاترین تنهاشوم، باز خدا هست

او جانشین همه نداشتن‌هاست...

---

[1] مناجاتی از دکتر علی شریعتی.

# سپاس گزاری. . . می کنم

سپاسگزاری جداگانه خود را در این مکان وارد نمایید.

افسانه قلی پور
شهریور ۱۳۹۱

# فهرست مطالب


**۱ مقدمات و مفاهیم اولیه** ۴

۱.۱ سانسور ۵

۱.۱.۱ سانسور از راست ۶

۲.۱.۱ سانسور نوع $I$ ۷

۳.۱.۱ سانسور نوع $II$ ۷

۴.۱.۱ سانسور فزاینده ۸

۵.۱.۱ سانسور فزاینده نوع $I$ ۸

۶.۱.۱ سانسور فزاینده نوع $II$ ۹

۷.۱.۱ سانسور فزاینده بازه ای نوع $I$ ۱۰

۲.۱ روش های برآورد پارامترها ۱۱

۱.۲.۱ الگوریتم EM ۱۱

۲.۲.۱ روش گشتاوری ۱۳

۳.۲.۱ روش درستنمایی نیمرخی ۱۴

۴.۲.۱ روش نمودار احتمال ۱۵

**۲ توزیع نمایی تعمیم یافته** $(GE)$ ۱۷

۱.۲ معرفی توزیع نمایی تعمیم یافته ۱۸

۲.۲ برآورد پارامترهای توزیع نمایی تعمیم یافته تحت سانسور فزاینده بازه ای نوع $I$ ۲۲

۱.۲.۲ برآورد درستنمایی ماکسیمم ۲۲

۲.۲.۲ روش تقریب نقطه-میانی $(Mid-point)$ ۲۳

۳.۲.۲ الگوریتم $EM$ ۲۴

۴.۲.۲ روش گشتاوری ۲۷

۵.۲.۲ روش درستنمایی ماکسیمم تقریبی $(AMLE)$ . . . . . ۲۸
۶.۲.۲ روش نمودار احتمال . . . . . ۳۵
۳.۲ مطالعه شبیه سازی . . . . . ۳۶
۱.۳.۲ الگوریتم شبیه سازی . . . . . ۳۷
۲.۳.۲ طرح های شبیه سازی . . . . . ۳۸
۳.۳.۲ برازش مدل . . . . . ۳۹
۴.۳.۲ نتایج شبیه سازی . . . . . ۴۰

**۳ توزیع گامپرتز** **۴۳**
۱.۳ معرفی توزیع گامپرتز . . . . . ۴۴
۲.۳ برآورد پارامتر های توزیع گامپرتز تحت سانسور فزاینده بازه ای نوع $I$ . . . . ۴۸
۱.۲.۳ برآورد درستنمایی ماکسیمم . . . . . ۴۸
۲.۲.۳ روش تقریب نقطه-میانی . . . . . ۴۹
۳.۲.۳ روش الگوریتم $EM$ . . . . . ۵۱
۴.۲.۳ روش نمودار احتمال . . . . . ۵۳
۳.۳ مطالعه شبیه سازی . . . . . ۵۴
۱.۳.۳ طرح های شبیه سازی . . . . . ۵۴
۲.۳.۳ داده های شبیه سازی شده . . . . . ۵۵
۳.۳.۳ نتایج شبیه سازی . . . . . ۵۶

**۴ توزیع بئر** **۵۸**
۱.۴ معرفی توزیع بئر نوع $XII$ . . . . . ۵۹
۲.۴ برآورد پارامترهای توزیع بئر نوع $XII$ تحت سانسور فزاینده بازه های نوع $I$ . ۶۲
۱.۲.۴ روش درستنمایی ماکسیمم . . . . . ۶۳
۲.۲.۴ روش تقریب نقطه-میانی . . . . . ۶۴
۳.۲.۴ روش الگوریتم $EM$ . . . . . ۶۵
۴.۲.۴ روش درستنمایی ماکسیمم تقریبی $(AMLE)$ . . . . . ۶۷
۵.۲.۴ روش نمودار احتمال . . . . . ۷۳
۳.۴ مطالعه شبیه سازی . . . . . ۷۴
۱.۳.۴ طرح های شبیه سازی . . . . . ۷۵

۲.۳.۴ داده های شبیه سازی شده . . . . . ۷۵
۳.۳.۴ نتایج شبیه سازی . . . . . ۷۶
۴.۴ فاصله اطمینان تقریبی . . . . . ۷۸

**آ برنامه های شبیه سازی در $R$** **۸۲**
آ.۱ برنامه های برآورد پارامترهای توزیع نمایی تعمیم یافته $(GE)$ . . . . . ۸۳
آ.۱.۱ برنامه شبیه سازی یک نمونه سانسور فزاینده بازه ای نوع $I$ در توزیع $GE$ ۸۳
آ.۱.۲ برنامه برآورد پارامترهای توزیع $GE$ به روش $MLE$ . . . . . ۸۴
آ.۱.۳ برنامه برآورد پارامترهای توزیع $GE$ به روش تقریب نقطه-میانی . . . ۸۶
آ.۱.۴ برنامه برآورد پارامترهای توزیع $GE$ به روش الگوریتم $EM$ . . . . . ۸۷
آ.۱.۵ برنامه برآورد پارامترهای توزیع $GE$ به روش گشتاوری . . . . . ۸۹
آ.۱.۶ برنامه برآورد پارامترهای توزیع $GE$ به روش $AMLE$ . . . . . ۹۰
آ.۱.۷ برنامه برآورد پارامترهای توزیع $GE$ به روش نمودار احتمال . . . . . ۹۳
آ.۲ برنامه های برآورد پارامترهای توزیع گامپرتز . . . . . ۹۴
آ.۲.۱ برنامه شبیه سازی یک نمونه سانسور فزاینده بازه ای نوع $I$ در توزیع گامپرتز . . . . . ۹۴
آ.۲.۲ برنامه برآورد پارامترهای توزیع گامپرتز به روش $MLE$ . . . . . ۹۵
آ.۲.۳ برنامه برآورد پارامترهای توزیع گامپرتز به روش تقریب نقطه-میانی . ۹۷
آ.۲.۴ برنامه برآورد پارامترهای توزیع گامپرتز به روش الگوریتم $EM$ . . . ۹۸
آ.۲.۵ برنامه برآورد پارامترهای توزیع گامپرتز به روش نمودار احتمال . . . ۱۰۰
آ.۳ برنامه های برآورد پارامترهای توزیع بئر نوع $XII$ . . . . . ۱۰۱
آ.۳.۱ برنامه شبیه سازی یک نمونه سانسور فزاینده بازه ای نوع $I$ در توزیع بئر نوع $XII$ . . . . . ۱۰۱
آ.۳.۲ برنامه برآورد پارامترهای توزیع بئر نوع $XII$ به روش $MLE$ . . . . ۱۰۲
آ.۳.۳ برنامه برآورد پارامترهای توزیع بئر نوع $XII$ به روش تقریب نقطه-میانی . . . . . ۱۰۴
آ.۳.۴ برنامه برآورد پارامترهای توزیع بئر نوع $XII$ به روش الگوریتم $EM$ ۱۰۵
آ.۳.۵ برنامه برآورد پارامترهای توزیع بئر نوع $XII$ به روش $AMLE$ . . . ۱۰۷
آ.۳.۶ برنامه برآورد پارامترهای توزیع بئر نوع $XII$ به روش نمودار احتمال ۱۰۹

واژه‌نامه فارسی به انگلیسی ۱۱۱

مراجع ۱۱۴

# پیش گفتار

سانسور یک پدیده متداول در آزمون‌های طول عمر و مطالعات قابلیت اعتماد می‌باشد. داده‌های سانسور شده وقتی رخ می‌دهند که در آن آزمایشگر نتواند اطلاعات کامل از تمام واحدهای آزمایشی مورد مطالعه به دست آورد. انواع سانسور بر اساس اینکه آزمایش چه موقع به پایان می‌رسد و چگونه داده‌ها از آزمایش‌های طول عمر جمع آوری شوند، پدید می‌آید. کریچ [۲] (۱۹۴۹) برای اولین بار عنوان سانسور شده را برای این نمونه‌ها پیشنهاد کرد. سانسور تحت بررسی در این پایان نامه، سانسور فزاینده بازه‌ای نوع I می باشد. سانسور فزاینده بازه‌ای نوع $I$ توسط آگاروالا [۳] (۲۰۰۱) معرفی شد و آگاروالا استنباط آماری را برای توزیع نمایی برپایه داده‌های سانسور شده فزاینده بازه‌ای نوع $I$ انجام داد. پس از آن نویسندگانی چون نگ [۴] و ونگ [۵] (۲۰۰۹) و چن [۶] و لیو [۷] (۲۰۱۰) به ترتیب به بررسی توزیع‌های وایبل و نمایی تعمیم یافته تحت این نوع سانسور پرداختند. برآورد پارامترهای چندین توزیع طول عمر از قبیل توزیع‌های نمایی تعمیم یافته، گامپرتز و بئر نوع XII به وسیله روش‌های مختلف از جمله روش گشتاوری، نمودار احتمال، الگوریتم $EM$ و درستنمایی ماکسیمم و درستنمایی ماکسیمم تقریبی تحت این نوع سانسور مورد بررسی قرار گرفته است. این تحقیق در شاخه‌های مختلف علوم پزشکی، مهندسی و صنعت و.... کاربرد زیادی دارد. چون معمولاً در این نوع جوامع، با نمونه‌های سانسور شده بسیاری برمی‌خوریم. هم چنین در شاخه‌های علمی نظیر ریاضی، آمار، فیزیک، اقتصاد، حسابداری ، علوم کامپیوتر و .... استفاده می‌شود.

توزیع نمایی تعمیم یافته ($GE$) توسط مودهولکر [۸] و اسریواستاوا [۹] (۱۹۹۳) معرفی شد. پس از آن نویسندگان زیادی این توزیع را تحت مطالعه و بررسی قرار دادند و بیان کردند که این توزیع می‌تواند به طور موثر در تحلیل داده‌های طول عمر به کار رود از قبیل رکب [۱۰] و احسان الله [۱۱] (۲۰۰۱) ، رکب (۲۰۰۲) ، زهنگ [۱۲] (۲۰۰۲) و سرهان [۱۳] (۲۰۰۷). گوپتا [۱۴] و کوندو [۱۵] (۱۹۹۹) توزیع

---

[۲] Kerrich
[۳] Aggarwala
[۴] NG
[۵] Wang
[۶] Chen
[۷] Lio
[۸] Mudholkar
[۹] Srivastava
[۱۰] Raqab
[۱۱] Ahsanollah
[۱۲] Zheng
[۱۳] Sarhan
[۱۴] Gupta
[۱۵] Kundu

$GE$ را به عنوان یک جایگزین برای توزیع‌های وایبل و گاما در موقعیت‌های مختلف به کار بردند و ویژگی‌های مختلف این توزیع را مطالعه و بررسی کردند. گوپتا و کوندو (۲۰۰۱) روش های مختلف برآورد پارامترهای توزیع نمایی تعمیم یافته را ارائه دادند. جاهین [۱۶] (۲۰۰۴) و رکب و مدی [۱۷] (۲۰۰۵) استنباط بیزی برای پارامترهای مجهول توزیع $GE$ مورد بررسی و تحقیق قرار دادند. اخیرا، گوپتا و کوندو (۲۰۰۷) تحقیق گسترده‌ای برای توزیع $GE$ بر اساس یک نمونه تصافی کامل انجام دادند.

توزیع گامپرتز به طور گسترده در توصیف مرگ و میر انسان ها به کار می‌رود. وو [۱۸] و همکاران (۲۰۰۳) و وو و یو [۱۹] (۲۰۰۵) به ترتیب برآوردهای درستنمایی ماکسیمم، فاصله اطمینان دقیق و نواحی اطمینان دقیق برای پارامترهای توزیع گامپرتز و توزیع بئر نوع $XII$ بر اساس نمونه سانسور اولین شکست به دست آوردند. این توزیع برای اولین بار توسط گامپرتز [۲۰] (۱۸۲۵) معرفی شد و پس از آن توسط نویسندگان زیادی مورد مطالعه و تحقیق قرار گرفت. به عنوان مثال رید [۲۱] (۱۹۸۳)، ماکانی [۲۲] (۱۹۹۱)، رائو [۲۳] و داماراجو [۲۴] (۱۹۹۲)، فرنسس [۲۵] (۱۹۹۴)، وو و لی [۲۶] (۱۹۹۹)، گریج [۲۷] و همکاران (۲۰۰۴) خواص توزیع گامپرتز را مورد مطالعه قرار دادند و برآوردهای درستنمایی ماکسیمم را برای پارامترهای این توزیع به دست آوردند. چن (۱۹۹۷) فاصله اطمینان دقیق و ناحیه اطمینان دقیق برای پارامترهای توزیع گامپرتز تحت نمونه سانسور نوع $II$ به دست آورد.

برای اولین بار توزیع بئر نوع $XII$ توسط بئر [۲۸] (۱۹۴۲) معرفی شد و در دو دهه اخیر به سبب اهمیت و کاربرد آن در موقعیت‌های عملی مورد توجه زیادی قرار گرفته است. برآورد درستنمایی ماکسیمم ($MLE$) پارامترهای توزیع بئر و روش‌های برازش این توزیع برای داده‌های طول عمر توسط وینگو [۲۹] (۱۹۸۳،۱۹۹۳) مورد مطالعه قرار گرفت. زیمر و همکاران (۱۹۹۸) با استفاده از این واقعیت که توزیع لگ-لجستیک حالت خاصی از توزیع بئر نوع $XII$ است، انعطاف پذیری

[۱۶] Jaheen
[۱۷] Madi
[۱۸] Wu
[۱۹] Yu
[۲۰] Gompertz
[۲۱] Read
[۲۲] Makany
[۲۳] Rao
[۲۴] Damaraju
[۲۵] Frances
[۲۶] Lee
[۲۷] Grage
[۲۸] Burr
[۲۹] Wingo

این توزیع را در مدل سازی زمان‌های شکست نشان داد. علی موسی [۳۰] و جاهین (۲۰۰۲) استنباط آماری برای توزیع بئر $XII$ را تحت داده های سانسور فزاینده نوع $II$ بررسی کردند. برآوردگرهای بیز و درستنمایی ماکسیمم توزیع بئر توسط سلیمان [۳۱] (۲۰۰۵) به دست آمده است.

این پایان نامه شامل ۴ فصل است. در فصل اول پایان نامه به معرفی مقدمات و مفاهیم اساسی اولیه که در طول پایان نامه از آنها استفاده شده است، می پردازیم. در فصل دوم، برآورد پارامترهای توزیع نمایی تعمیم یافته را به روش های مختلف ذکر شده به دست می آوریم و در پایان فصل مطالعات شبیه سازی به منظور مقایسه روش های مختلف برآورد بر اساس اریبی و میانگین مربعات خطا انجام شده است. برآورد پارامترهای این توزیع بر اساس داده‌های حقیقی بیماران دارای تومور بدخیم سلول‌های مغز استخوان به دست آمده است. در فصل سوم، پارامترهای توزیع گامپرتز با استفاده از روش‌های ماکسیمم درستنمایی، تقریب نقطه-میانی، الگوریتم $EM$ و نمودار احتمال برآورد شده‌اند. در این فصل داده‌های سانسور فزاینده بازه‌ای نوع $I$ از توزیع گامپرتز شبیه سازی شده‌اند و بر اساس این داده‌های شبیه سازی شده، برآورد پارامترهای توزیع گامپرتز را به دست آورده‌ایم. در فصل چهارم، برآورد پارامترهای یکی دیگر از توزیع‌های طول عمر به نام توزیع بئر نوع $XII$ را به روش‌های مختلف از جمله روش‌های درستنمایی ماکسیمم، تقریب نقطه میانی، الگوریتم $EM$، درستنمایی ماکسیمم تقریبی و نمودار احتمال مورد بررسی قرار گرفته است. در نهایت، روش‌های مختلف برآورد را با استفاده از شبیه سازی مقایسه می‌کنیم و پارامترهای توزیع بئر نوع $XII$ را بر اساس یک نمونه شبیه سازی شده از این توزیع، تحت سانسور فزاینده بازه‌ای نوع $I$ برآورد می‌کنیم. همچنین، فاصله اطمینان تقریبی بر اساس تقریب نرمال برای پارامترهای این توزیع تحت سانسور مورد نظر به دست می‌آیند.

[۳۰] Ali Mousa
[۳۱] Soliman

# فصل ۱
# مقدمات و مفاهیم اولیه

## ۱.۱ سانسور

در بسیاری از موارد، از جمله آزمون‌های طول عمر، آزمایش‌های کلینیکی، تحقیقات زیست شناسی، آنالیز بقا و دیگر زمینه‌های کاربردی علم آمار با نمونه‌هایی مواجه‌ایم که همه مشاهدات ثبت نشده‌اند و یا همه واحدهای نمونه به نتیجه نرسیده‌اند. این محدودیت‌ها ممکن است توسط آزمایشگر اعمال شود یا ماهیت آزمایش طوری باشد که خود به خود محدودیتی را در مشاهدات به وجود بیاورد. برای مثال بعضی از واحدهای آزمایشی ممکن است به طور ناخواسته و ناگهانی متوقف شوند و یا در آزمایشات بالینی ممکن است افراد به دلایلی از آزمایش انصراف دهند. کنار گذاشتن تعدادی از واحدهای آزمایشی در بعضی از مطالعات طول عمر که شامل واحدهای گران قیمت می‌باشند، باعث صرفه جویی زیادی در هزینه می‌شود و در بعضی از آزمایش‌های بالینی طول بقای بیمار ممکن است چند سال طول بکشد و یا ممکن است آزمایشگر قبل از مشاهده طول عمر بعضی از بیماران آزمایش را به پایان برساند. در چنین مواقعی با داده‌های سانسور شده مواجه‌ایم. در واقع، داده‌های سانسور شده وقتی رخ می‌دهند که در آن آزمایشگر نتواند اطلاعات کامل از تمام واحدهای آزمایشی مورد مطالعه به دست آورد. سانسور یک پدیده متداول در آزمون‌های طول عمر و مطالعات قابلیت اعتماد می‌باشد. انواع سانسور بر اساس اینکه آزمایش چه موقع به پایان می‌رسد و چگونه داده‌ها از آزمایش‌های طول عمر جمع آوری شوند، پدید می‌آید. کریچ [۱] (۱۹۴۹) برای اولین بار عنوان سانسور شده را برای این نمونه‌ها پیشنهاد کرد.

[۱] Kerrich

دلایلی که باعث به وجود آمدن سانسور می‌شوند:۱-ممکن است یک شخص پیشامد موردنظر را قبل از پایان آزمایش تجربه نکند. ۲- عدم بازگشت شخص در طول دوران مطالعه ۳- کنار رفتن شخص از مطالعه به دلیل مرگ (اگر مرگ پیشامد موردنظر نباشد) یا دلایل دیگری مثل اثر نامطلوب روش درمانی مورد استفاده.

در اکثر آزمون‌های طول عمر زمان شکست همه واحدها معین نیست. لذا آزمایشگر با داده‌های سانسور شده مواجه است. فرض کنید $n$شی در یک آزمون طول عمر قرار گرفته‌اند و به دلایلی باید آزمایش قبل از شکست همه واحدها پایان یابد. در این صورت داده‌های حاصل تشکیل یک نمونه سانسور شده می‌دهند. این شرایط معمولا در آزمون‌های طول عمر، مطالعه سیستم‌ها و مباحث قابلیت اعتماد پدیدار می‌شود. از آنجا که آزمون‌های طول عمر زمان زیادی می‌برند، اغلب آنها را پس از شکست $r$ام متوقف می‌کنیم یا آنکه آزمایشگر تا زمان $t$ منتظر می‌ماند و سپس آزمایش را خاتمه می‌دهد. اکنون به معرفی انواع سانسور می‌پردازیم.

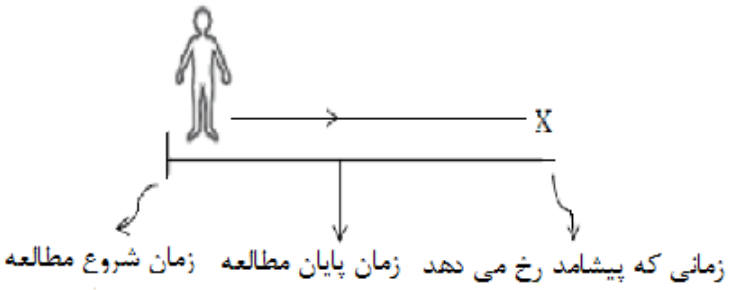
### ۱.۱.۱ سانسور از راست

این نمونه‌ها شامل $n$ مشاهده‌اند که $m$ تا از آنها کامل اندازه گیری شده در حالی که $r = n - m$ تای باقی مانده سانسور می‌شوند. تحت این سانسور، مطالعه تا زمان $t$ که از قبل آزمایشگر مشخص کرده است، ادامه داده می‌شود. در واقع، برای هر مشاهده سانسور شده فقط می‌دانیم که $X > t$ و برای مشاهدات اندازه گیری شده $X \leq t$ است. در نمونه‌های راست سانسور شده نوع$I$، $t$ یک نقطه ثابت معلوم و در نمونه‌های راست سانسور شده نوع $II$، $t = X_{m:n}$، $m$ امین آماره ترتیبی است. بنابراین، تحت طرح سانسور از راست مشاهدات عبارتند از:

$$Y_i = \begin{cases} X_i & X_i \leq t \\ T & X_i > t \end{cases}$$

نمونه‌های راست سانسور شده مکررا در آزمایش‌های طول عمر اتفاق می‌افتد.

**مثال ۱.۱.۱.** در یک آزمایش بالینی، برای ارزیابی یک نوع داروی جدید در حفظ بهبودی، ۲۱ بیمار مبتلا به نوعی از سرطان حاد خونی در طی یک سال مورد مطالعه قرار گرفته‌اند. اگر پس از یک سال (زمان پایان مطالعه) شخص بیمار هنوز تحت مراقبت باشد، یعنی پیشامد مورد نظر (بهبودی بیمار) در زمان پایان مطالعه هنوز رخ نداده باشد، آن گاه، زمان بقای شخص سانسور شده از راست در نظر گرفته می‌شود.




شکل ۱.۱: نمایش مشاهده راست سانسور شده

### ۲.۱.۱ سانسور نوع $I$

فرض کنید $n$ مولفه در یک آزمایش طول عمر قرار گرفته‌اند و آزمایش در زمان از پیش تعیین شده $t$ به پایان می‌رسد، به طوری که تنها زمان شکست مولفه هایی که قبل از این زمان از بین رفته‌اند، ثبت می‌شوند. داده‌هایی که به این ترتیب به دست می‌آیند یک نمونه سانسور شده نوع $I$ را می‌سازند. سانسور نوع $I$ ، سانسور زمانی نیز گفته می‌شود. تابع درستنمایی برای این طرح سانسور به صورت زیر می‌باشد :

$$L(\theta; r, x) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_i, \theta)[1 - F(t, \theta)]^{n-r}, x_1 < x_2 < ... < x_r < t \qquad (1.1)$$

**مثال ۱.۱.۲.** مطالعه ای را در نظر بگیرید که هدف آن بررسی اثر داروی خاص در خطر رد پیوند کلیه باشد. فرض کنید مطالعه به این صورت انجام شود که از ابتدای سال ۱۳۸۵ تا انتهای سال ۱۳۸۶ تمام بیمارانی که در بیمارستان سینا تحت پیوند کلیه قرار داشته‌اند، پیگیری می‌شوند و تمام موارد رد پیوند ثبت می‌شود. مطالعه فوق، یکی از آزمایشات تحلیل بقاست که داده‌های حاصل، نمونه‌ای از سانسور نوع $I$ هستند.

### ۳.۱.۱ سانسور نوع $II$

اگر آزمایش را در زمان از کار افتادن $r$ امین واحد، $1 \leq r \leq n$ ، پایان دهیم نمونه سانسور شده نوع $II$ به دست می‌آید. فرض کنید $X_1 \leq X_2 \leq ... \leq X_r$ زمان‌های شکست مشاهده شده باشند. در این حالت $r$ ثابت و زمان اتمام آزمایش، یعنی$X_r$ ، تصادفی است. سانسور نوع $II$ سانسور شکست نیز نامیده می‌شود. تابع درستنمایی برای این طرح سانسور عبارت است از:

$$L(\theta; r, x) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_i, \theta)[1 - F(x_i, \theta)]^{n-r}, x_1 < x_2 < ... < x_r \qquad (2.1)$$

**مثال ۳.۱.۱.** در مثال فوق اگر آزمایش تا زمانی ادامه پیدا کند که تعداد مشخصی از بیماران مثلا $r$ تا دچار رد پیوند کلیه شوند، داده‌های حاصل، سانسور نوع $II$ هستند.

### ۴.۱.۱ سانسور فزاینده

یکی از ایرادات طرح‌های سانسور نوع $I$ و نوع $II$ اینست که امکان کنار گذاشتن واحدها در زمان‌هایی دیگر غیر از زمان پایان آزمایش را ندارند. در حالی که لازم است گاهی واحدهای آزمایشی را در مراحل مختلف آزمایش کنار بگذاریم. مخصوصا هنگامی که داده‌های مورد آزمایش پرهزینه هستند. در این حالت حذف واحدهای سالم آزمایشی می تواند باعث کاهش چشمگیری در هزینه و وقت آزمایش شود. به همین دلیل سانسورهای کلی تری به نام سانسورهای فزاینده مطرح می‌شوند. هرد[۲] (۱۹۶۵) برای اولین بار مسئله سانسور فزاینده را مطرح نمود. این نوع سانسور را معمولا سانسور چندگانه یا سانسور چندمرحله‌ای نیز می‌نامند. سانسور فزاینده در آزمایشات کلینیکی و صنعتی مفیدند. گرچه این نوع از سانسور در آزمون‌های طول عمر صنعتی مورد توجه بیشتری قرار گرفته است.

مواردی که سانسور فزاینده می‌تواند مناسب باشد: الف- در مواردی که برای مطالعات دیگر، نیاز به واحدهای آزمایشی داشته باشیم. ب- در مواردی که با توجه به ماهیت آزمایش انجام شده استفاده از سایر طرح‌های سانسور مقدور نباشد. ج- در مواردی که خرابی واحد آزمایشی به طور ضمنی نشان دهنده طول عمر تعداد دیگری از واحدهای آزمایشی باشد. د- در مواردی که محدودیت بودجه و امکانات وجود داشته باشد.

### ۵.۱.۱ سانسور فزاینده نوع $I$

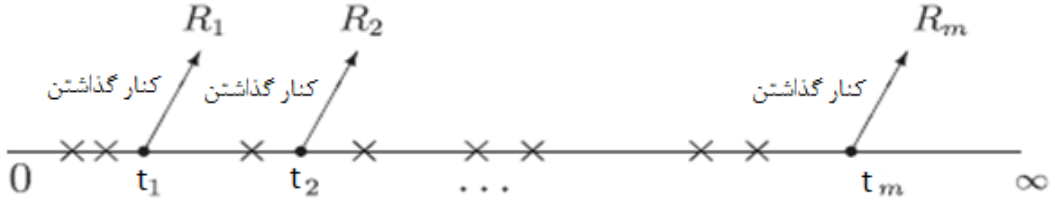
اگر آزمایشگر تصمیم بگیرد در زمان‌های از قبل تعیین شده $t_1, ..., t_m$ به ترتیب $R_1, ..., R_m$ واحد سالم به طور تصادفی از آزمایش حذف کند، در این صورت طرح سانسور فزاینده نوع $I$ رخ داده است. در زمان $t_i$ تعداد $R_i$ واحد سالم به تصادف از آزمایش حذف می‌شود. البته به شرط اینکه $R_i$ واحد سالم در زمان $t_i, i = 1, 2, ..., m$ ، وجود داشته باشد. یعنی تعداد واحدهای سانسور شده در زمان $t_i$ برابر است با ( تعداد واحدهای سالم در زمان $min(R_i, t_i$ . در نهایت آزمایش در زمان $t_m$ که هنوز $R_m$ واحد سالم باقی مانده، به پایان می یابد. اگر فرض کنیم $k$ واحد قبل از زمان $t_m$ شکست خورده اند، علاوه بر زمان های سانسور $t_i$ ، زمان های شکست $X_1, ..., X_k$ را نیز خواهیم داشت. بنابراین، در این طرح سانسور زمان شکست و زمان های سانسور یکسان نیستند.

[۲] Herd

تابع درستنمایی برای این نوع سانسور عبارتست از:

$$L(\theta) \propto \prod_{i=۱}^{k} f(x_i, \theta) \prod_{i=۱}^{m} [۱ - F(t_i, \theta)]^{R_i} . x_۱ < x_۲ < ... < x_k < t_m \qquad (۳.۱)$$

اگر در سانسور فزاینده نوع $I$ ، $R_i = ۰, i = ۱, ۲, ..., m-۱$ آن گاه نمونه سانسور فزاینده نوع $I$ به یک نمونه سانسور نوع $I$ تبدیل می‌شود. شکل ۲.۱، این طرح سانسور را نشان می‌دهد.



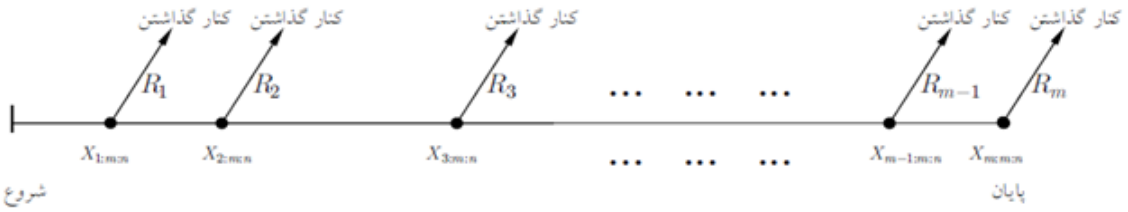

شکل ۲.۱: طرح سانسور فزاینده نوع $I$

### ۶.۱.۱ سانسور فزاینده نوع $II$

فرض کنید $n$ مولفه در زمان $t = ۰$ در یک آزمون طول عمر قرار دارند. با مشاهده اولین خرابی، $X_{۱:m:n}$، تعداد $R_۱$ مولفه سالم به طور تصادفی از آزمایش حذف می‌شوند. بلافاصله با مشاهده دومین خرابی، $X_{۲:m:n}$، تعداد $R_۲$ مولفه سالم از $n - R_۱ - ۲$ مولفه سالم باقیمانده به طور تصادفی از آزمایش حذف می‌شوند. این روند ادامه می‌یابد تا با مشاهده $m$ امین خرابی $X_{m:m:n}$ که $m$ از قبل مشخص است، همه $R_m$ مولفه سالم یعنی $R_m = n - m - \sum_{i=۱}^{m-۱} R_i$ ازآمایش کنار گذاشته می‌شوند. در سانسور فزاینده نوع $II$، $m$ آماره ترتیبی داریم که با نماد $X_{۱:m:n}, X_{۲:m:n}, ...X_{m:m:n}$ نشان می‌دهیم. در این حالت زمان‌های مربوط به حذف مولفه‌ها متغیر تصادفی است. در نتیجه اجرای این طرح سانسور، $m$ مقدار مرتب شده که آماره‌های ترتیبی راست سانسور شده نوع $II$ نامیده می‌شود، تولید می‌گردد. شکل (۳.۱)، این طرح سانسور را نشان می‌دهد. تابع درستنمایی برای این نوع سانسور عبارتست از:

$$L(\theta) = c \prod_{i=۱}^{m} f(x_i, \theta)[۱ - F(x_i, \theta)]^{R_i}, x_۱ < x_۲ < ... < x_m \qquad (۴.۱)$$

که در آن $c = n(n - R_۱ - ۱)(n - R_۱ - R_۲ - ۲)...(n - R_۱ - ... - R_{m-۱} - m + ۱)$ .
در صورتی که در سانسور فزاینده نوع $II$، $R_۱ = R_۲ = ... = R_{m-۱} = ۰$ و $R_m = n - m$ باشد،




شکل ۳.۱: طرح راست سانسور شده فزاینده نوع $II$

این طرح به طرح سانسور نوع $II$ از راست تبدیل می‌شود و اگر $R_1 = R_2 = ... = R_m = 0$ طرح مذکور به طرح کامل (فاقد سانسور) تبدیل می‌گردد که در آن تمام $n$ آماره مرتب معمول مشاهده می‌شود. برای جزئیات بیشتر طرح این طرح سانسور به آگاروالا [۳] و بالاکریشنان [۴] (۲۰۰۰) و چاندراسیکار [۵] و همکاران (۲۰۰۲) مراجعه کنید.

### ۷.۱.۱ سانسور فزاینده بازه ای نوع $I$

سانسور فزاینده بازه‌ای نوع $I$ ترکیبی از سانسور بازه ای نوع $I$ و سانسور فزاینده است. فرض کنید $n$ مولفه تحت آزمایش طول عمر در زمان $t_0$ قرار دارند و در $m$ زمان از پیش تعیین شده $t_1, t_2, ..., t_m$ که $t_m$ زمان پایان آزمایش است، تحت بررسی‌اند. در $i$ امین زمان بررسی، $t_i$، تعداد $X_i$ خرابی در بازه $(t_{i-1}, t_i]$ ثبت می‌شود و $R_i$ مولفه سالم به طور تصادفی انتخاب و از آزمایش حذف می‌شود. در نهایت، در زمان $t_m$، $X_m$ خرابی در بازه $(t_{m-1}, t_m]$ ثبت می‌شود و $R_m = n - \sum_{j=1}^{m} X_j - \sum_{j=1}^{m-1} R_j$ از آزمایش حذف می‌گردد. $R_i$ها می‌توانند به صورت درصدهای از پیش تعیین شده $p_1, ..., p_{m-1}$ و $p_m = 1$ از مولفه‌های سالم باقیمانده در زمان $t_i$، در نظر گرفته شوند $(i = 1, 2, ..., m)$ . اگر $y_i$ را تعداد مولفه‌های سالم در بازه $(t_{i-1}, t_i]$ در نظر بگیریم، می‌توان تعداد مولفه‌هایی که به طور تصادفی انتخاب و از آزمایش حذف می‌شود، را به صورت $R_i = \lfloor p_i y_i \rfloor, i = 1, 2, ..., m$ در نظر گرفت. بنابراین یک نمونه سانسور فزاینده بازه‌ای نوع $I$ را می‌توان به صورت $(X_i, R_i, t_i)$ نشان داد که حجم نمونه $n = \sum_{i=1}^{m}(X_i + R_i)$ است.

اگر $R_i = 0, i = 1, 2, ..., m-1$ آن گاه نمونه سانسور فزاینده بازه‌ای نوع $I$ به یک نمونه سانسور بازه‌ای نوع $I$ تبدیل می‌شود. تابع درستنمایی برای یک نمونه سانسور فزاینده بازه‌ای نوع $I$

[۳] Aggarwala
[۴] Balakrishnan
[۵] Chandrasekar

با تابع توزیع $F(t,\theta)$ عبارتست از:

$$L(\theta) \propto \prod_{i=۱}^{m}[F(t_i,\theta) - F(t_{i-۱},\theta)]^{X_i}[۱ - F(t_i,\theta)]^{R_i}. \qquad (۵.۱)$$

سانسور فزاینده بازه‌ای نوع $I$ توسط آگاروالا (۲۰۰۱) معرفی شد و او استنباط آماری را برای توزیع نمایی برپایه داده‌های سانسور شده فزاینده بازه ای نوع $I$ انجام داد. پس از آن نویسندگانی چون نگ [۶] و ونگ [۷] (۲۰۰۹) و چن [۸] و لیو [۹] (۲۰۱۰) به ترتیب به بررسی توزیع‌های وایبل و نمایی تعمیم یافته تحت این نوع سانسور پرداختند.

**مثال ۴.۱.۱.** در یک آزمایش بالینی ۱۱۲ بیمار که دارای تومور مغز استخوان می‌باشند، از زمان $t_۰ = ۰$ تحت مطالعه‌اند. آزمایش در ۹ زمان از پیش تعیین شده بر حسب ماه، ادامه پیدا می‌کند. در هر بازه زمانی داده شده تعداد بیمارانی که از آزمایش حذف می‌شوند، $R_i$ ، و تعداد شکست‌ها، $X_i$، داده شده است.

$X = (۱۸, ۱۶, ۱۸, ۱۰, ۱۱, ۸, ۱۳, ۴, ۱)$
$t = (۵/۵, ۱۰/۵, ۱۵/۵, ۲۰/۵, ۲۵/۵, ۳۰/۵, ۴۰/۵, ۵۰/۵, ۶۰/۵)$

جدول (۱.۱) مجموعه داده‌های حاصل از این بیماران را نشان می‌دهد.

## ۲.۱ روش های برآورد پارامترها

### ۱.۲.۱ الگوریتم EM

الگوریتم $EM$ یک روش تکراری برای برآورد پارامترها به روش ماکسیمم درستنمایی است، زمانی که تعدادی از متغیرهای تصادفی مدل مشاهده نشوند. متغیرهای مشاهده نشده می‌توانند شامل اطلاعات ناکامل و داده‌های گمشده باشند. فرض می‌شود که بردار تصادفی $X$ به صورت کامل مشاهده نمی‌شود، بنابراین می‌توان آن را به صورت $X = (Y, Z)$ تفکیک کرد که $Y$ داده‌های مشاهده شده و $Z$ بردار داده‌های مشاهده نشده است و دارای تابع چگالی $f(y, z|\theta)$ که $\theta \in \Theta$ باشد. می‌خواهیم $\theta$ را به روش ماکسیمم درستنمایی برآورد کنیم. اگر تابع درستنمایی داده‌های

[۶] Ng
[۷] Wang
[۸] Chen
[۹] Lio

| بازه های زمانی برحسب ماه | تعداد بیماران | تعداد بیمارانی که حذف می شوند |
|---|---|---|
| [۰,۵.۵) | ۱۱۲ | ۱ |
| [۵.۵,۱۰.۵) | ۹۳ | ۱ |
| [۱۰.۵,۱۵.۵) | ۷۶ | ۳ |
| [۱۵.۵,۲۰.۵) | ۵۵ | ۰ |
| [۲۰.۵,۲۵.۵) | ۴۵ | ۰ |
| [۲۵.۵,۳۰.۵) | ۳۴ | ۱ |
| [۳۰.۵,۴۰.۵) | ۲۵ | ۲ |
| [۴۰.۵,۵۰.۵) | ۱۰ | ۳ |
| [۵۰.۵,۶۰.۵) | ۳ | ۲ |
| [۶۰.۵,∞) | ۰ | ۰ |

جدول ۱.۱: مجموعه داده حاصل از بیماران

کامل را با $L(\theta|y,z)$ نشان دهیم و لگاریتم تابع درستنمایی داده‌های کامل را با $l(\theta|y,z)$ نشان دهیم، آن گاه

$$l(\theta|y,z) = lnL(\theta|y,z) = lnf(y,z|\theta).$$

$\widehat{\theta}$ برآورد $ML$ برای $\theta$ است هرگاه $l(\theta|y,z) \leq l(\widehat{\theta}|y,z), \theta \in \Theta$ . الگوریتم $EM$ زمانی مفید است که ماکسیمم کردن تابع درستنمایی داده‌های مشاهده شده $l(\theta|y)$ مشکل باشد ولی ماکسیمم کردن $l(\theta|y,z)$ راحت‌تر باشد. الگوریتم $EM$، $l(\theta|y,z)$ را با جایگزینی امید شرطی به شرط داده‌های مشاهده شده $y$ ماکسیمم می‌کند. اگر $\theta_{\circ}$ یک مقدار اولیه دلخواه از فضای پارامتری $\Theta$ باشد، آن گاه در اولین تکرار $Q(\theta|\theta_{\circ},y)$ به صورت زیر تعریف می‌شود:

$$Q(\theta|\theta_{\circ},y) = E_{\theta_{\circ}}(l(\theta|y,z)|Y)$$

حال $Q(\theta|\theta_{\circ},y)$ را بر حسب $\theta$ ماکسیمم می کنیم. اگر مقدار جدید $\theta$ که $Q(\theta|\theta_{\circ},y)$ را ماکسیمم می‌کند، $\widehat{\theta}_{(۱)}$ نامیده شود، در این صورت به ازای همه مقادیر $\theta \in \Theta$ ، $Q(\widehat{\theta}_{(۱)}|\theta_{\circ},y) \geq Q(\theta|\theta_{\circ},y)$ . الگوریتم شامل دو مرحله است:

۱-مرحله $E-$ : $Q(\theta|\widehat{\theta}_{(j-۱)},y)$ را به صورت زیر محاسبه می‌کنیم:

$$Q(\theta|\widehat{\theta}_{(j-۱)},y) = E_{\widehat{\theta}_{(j-۱)}}(l(\theta|y,z)|Y). \qquad (۶.۱)$$