
بسم الله الرحمن الرحیم

حمایت از حقوق پدیدآورندگان

پایان نامۀ حاضر، حاصل پژوهشهای نگارنده در دورۀ **کارشناسی ارشد** رشتۀ **آمار** گرایش **آمار ریاضی** است که در بهمن ۱۳۹۲ در **دانشکده علوم** دانشگاه **یاسوج** به راهنمایی **دکتر انوشیروان غفاری پور** و مشاورۀ **دکتر حیدر علی مردانی فرد** از آن دفاع شده است و کلیۀ حقوق مادی و معنوی آن متعلق به دانشگاه یاسوج است.

دانشکده علوم
گروه ریاضی

پایان نامه کارشناسی ارشد   رشته آمار گرایش آمار ریاضی

توزیع فرآیندهای وایبل دو متغیره مارشال اولکین

استاد راهنما

دکتر انوشیروان غفاری پور

پژوهشگر

داود بربست

بهمن ۱۳۹۲

توزیع فرآیندهای وایبل دو متغیره مارشال اولکین

به وسیلهٔ

داود بربست

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی از فعالیت‌های تحصیلی لازم برای اخذ درجهٔ کارشناسی ارشد

در رشتهٔ:

آمار

در تاریخ .... توسط هیأت داوران زیر بررسی و با درجهٔ .... به تصویب نهایی رسید.

۱- استاد راهنما:	دکتر انوشیروان غفاری پور	با مرتبهٔ علمی استادیار	امضاء
۲- استاد مشاور:	دکتر حیدر علی مردانی فرد	با مرتبهٔ علمی استادیار	امضاء
۳- استاد داور داخل گروه:	دکتر آرش اردلان	با مرتبهٔ علمی استادیار	امضاء
۴- استاد داور خارج گروه:	دکتر کاووس خورشیدیان	با مرتبهٔ علمی استادیار	امضاء
۵- نمایندهٔ تحصیلات تکمیلی دانشگاه:	دکتر زهرا رفیعی	با مرتبهٔ علمی استادیار	امضاء

تقدیم به:

پدر و مادر خوبم

و

همسرم یاور همیشگی ام

قدردانی

سپاس خداوندگار حکیم را که با لطف بی‌کران خود، آدمی را زیور عقل آراست.

در آغاز وظیفه خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر انوشیروان غفاری پور، صمیمانه تشکر و قدردانی کنم که قطعاً بدون راهنمایی‌های ارزنده ایشان، این مجموعه به انجام نمی‌رسید.

از جناب آقای دکتر حیدر علی مردانی فرد که زحمت مطالعه و مشاوره این رساله را تقبل فرمودند و در آماده سازی این رساله، به نحو احسن اینجانب را مورد راهنمایی قرار دادند، کمال امتنان را دارم.

در پایان، بوسه می‌زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می‌کنم وجود مقدسشان را و تشکر می‌کنم از همسر عزیزم به پاس عاطفه سرشار و وجود پر ثمرشان، که در این سردترین روزگاران، بهترین پشتیبان من بودند. سلامتی و توفیق روز افزون همهٔ بزرگوارانی را که در تهیه این کار پژوهشی، اینجانب را مورد لطف و مساعدت خود قرار دارند، از خداوند بزرگ مسئلت دارم.

داود بربست

بهمن ۱۳۹۲

چکیده

یکی از مدل‌های پر کاربرد برای تحلیل داده های دو متغیری توزیع نمایی دو متغیری مارشال-الکین است. توزیع های دو متغیری زیادی توسط مارشال اولکین معرفی شده اند که یکی از آن ها توزیع وایبل دو متغیره مارشال اولکین که دارای دو نوع است. نوع اول آن با تغییر متغیر در توزیع نمایی دو متغیری مارشال اولکین بدست می آید و نوع دوم آن با افزودن یک پارامتر به نوع اول آن بدست می آید.

یکی از مدل های سری زمانی فرآیند های کمینه سازی (مینیمم سازی) است که ما در اینجا فرآیند های مینیمم سازی با دو نوع توزیع وایبل دو متغیره مارشال اولکین را معرفی کرده و شرایط مانایی این فرآیند ها را مورد بررسی قرار داده ایم. در نهایت برآورد شش پارامتر فرآیند مینیمم سازی با نوع اول توزیع دو متغیری مارشال-الکین را بدست آورده و به شبیه سازی و تحلیل با داده های واقعی پرداخته‌ایم.

در فصل اول سعی شده برخی مفاهیم مورد نیاز شرح داده شود از جمله این موارد می توان به مباحث اصلی در فرآیند تصادفی، تابع بقا، تابع نرخ مخاطره و قضیه ارگودیک را نام برد.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فهرست علائم اختصاری

امید ریاضی فرآیند تصادفی N_t	$E(N_t)$
کواریانس بین N_s و N_t	$cov(N_s, N_t)$
همبستگی بین N_S و N_t	$corr(N_s, N_t)$
big O از مرتبه h	$O(h)$
تابع بقای متغیر تصادفی T	$\bar{F}(t)$
تابع نرخ مخاطره	$\lambda(t)$
فضای احتمال	$(\Omega, \mathbb{F}, P)$
تابع توزیع	cdf
توزیع نمایی دو متغیره مارشال الکین	$MOBE$
توزیع وایبل دو متغیره مارشال الکین	$MOBVW$
اتورگرسیو از مرتبه p	$AR(p)$

فصل ۱

مفاهیم و تعاریف اولیه

فرآیندهای تصادفی مجموعه مهمی از مباحث آماری است که جای خود را نه تنها در علوم پایه و پژوهش های محض علمی باز کرده است بلکه کاربرد های فراوانی در زمینه های مختلف علمی و صنعتی برای آنها وجود دارد. به طور کلی فرآیند های تصادفی سیستم هایی را مورد بحث و بررسی قرار می دهد که با گذشت زمان تحت نوسانات تصادفی قرار می‌گیرند. اکثر پدیده های طبیعی تحت فعل و انفعالات احتمالی قرار دارند و این زمینه را برای استفاده از الگو های فرایندهای تصادفی در مطالعه آن ها فراهم می سازد. آزمایش تصادفی نقطه شروع مباحث مرتبط با حساب احتمالات است. متغیر تصادفی تابعی است از فضای نمونه به زیر مجموعه‌ای از اعداد حقیقی و فرآیندهای تصادفی به دنباله ای از متغیرهای تصادفی اشاره دارد.[۲]

در فصل اول به معرفی مفاهیمی می پردازیم که در فصل های پیش رو برای شناخت موضوع اصلی به آنها نیازمندیم. این مفاهیم عبارت اند از فرآیند تصادفی و برخی مباحث اصلی آن، فرآیند شمارشی، فرآیند پواسن، فرآیند مارکوف، تابع بقا، توزیع وایبل و آزمون نیکویی برازش توزیع.

۱-۱ فرآیند های تصادفی

فرض کنید T مجموعه ای دلخواه و به ازای هر $t \in T$، X_t متغیری تصادفی باشد. فرض کنید $E \subset R$ مجموعه ی ثابتی باشد و مقادیر متغیر تصادفی X_t در داخل این مجموعه باشد. در این صورت مجموعه‌ی $\{X_t : t \in T\}$ را فرآیند تصادفی[1] با مجموعهٔ اندیس گذار T و فضای حالت E می گوییم. اگر $A \subset E$ (یا $x \in E$) و $X_t \in A$ (یا $X_t = x$)، می گوییم فرآیند در زمان (مرحله) t در مجموعه A (یا در حالت x) قرار دارد. اگر T مجموعه ای شمارا باشد، فرآیند را زمان- گسسته و اگر مجموعه‌هایی به صورت $(\circ, \infty)$ یا $(-\infty, \infty)$ باشد آن را زمان پیوسته می گوییم. مثلاً اگر X_n نشان دهنده شدت n امین زمین لرزه در ناحیه ای باشد این فرآیند زمان گسسته، و اگر y_t حجم آب منبع دانشکده در زمان t باشد فرآیند y_t از نوع زمان پیوسته است.[۱]

مثال ۱-۱-۱. فرض کنید سکه ای را متوالیا پرتاب می کنیم متغیر تصادفی X_t، $t = ۱, ۲, ۳, ...$ را به صورت زیر تعریف می کنیم.

$$X_t = \begin{cases} ۱, & \text{اگر دفعه t ام شیر بیاید} \\ \circ, & \text{اگر دفعه t ام خط بیاید} \end{cases}$$

در این صورت $\{X_t : t \geqslant ۱\}$ فرآیند تصادفی زمان گسسته و فضای حالت آن $E = \{\circ, ۱\}$ است.اگر نتیجه یک آزمایش به صورت $\omega = HHTTTH...$ باشد، آنگاه

$X_۱(\omega) = X_۲(\omega) = X_۶(\omega) = ۱$ و $X_۳(\omega) = X_۴(\omega) = X_۵(\omega) = \circ$

با استفاده از اطلاعات فرآیند در مراحل قبل می توان مراحل بعد را پیش بینی[۲] کرد.

۱-۱-۱ مباحث اصلی در فرآیند تصادفی

نموّ: در فرآیند $\{X_t : t \geqslant \circ\}$ تفاضل $X_t - X_s$، $s < t$ رانمو فرآیند در فاصله $(s, t]$ گویند. مثلاً اگر فرآیند در زمان های ۵ و ۱۰ به ترتیب در حالت های ۶ و ۱۵ باشد، آن‌گاه نموّ آن

[۱] Stochastic Processes

[۲] prediction

در فاصله $[5,10)$ برابر ۹=۶-۱۵ است. می گوییم یک فرآیند دارای نموّهای مستقل است اگر به ازای هر n و هر $0 \leqslant t_1 < t_2 < t_3 < ... < t_n$ نموّهای زیر که متغیر تصادفی هستند مستقل باشند.

$$X_0, X_{t_1} - X_0, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$$

نموّ مانا: فرآیند تصادفی $\{X_t : t \geqslant 0\}$ یک فرآیند ماناست اگر و فقط اگر به ازای $s, t \in T$ نمو X از s به t یعنی $X_t - X_s$ با نمو X در زمان دیگر با طول ثابت مانند $X_{t+h} - X_{s+h}$ هم توزیع باشد. در این نوع فرآیند ساختار نموّ فرآیند در فاصله های زمانی یکسان با هم برابر است و نقطه شروع در این فرآیند مهم نیست و همچنین توزیع $X_t - X_s$ فقط به $t - s$ بستگی دارد.
به عبارت دیگر فرآیند دارای نموّ های مستقل است اگر تغیرات مقادیر آن روی فاصله های زمانی نامتداخل مستقل باشند و دارای نموّ های مانا است اگر $X_{t+s} - X_t$ برای تمام t ها دارای یک توزیع باشند یعنی توزیع تغیرات مقادیر در بین هر دو نقطه فقط به فاصله آن دو نقطه بستگی داشته باشد. □
مانای اکید: فرآیند $\{X_t : t \geqslant 0\}$ را مانای اکید می گوییم هرگاه به ازای $n \geqslant 1$ و هر $t_1, t_2, t_3, ..., t_n$ و هر $h \geqslant 0$ توزیع توأم $X_{t_1}, X_{t_2}, ..., X_{t_n}$ همان توزیع توأم متغیر های تصادفی $X_{t_1+h}, X_{t_2+h}, ..., X_{t_n+h}$ باشد.
با فرض $n = 1$ نتیجه می شود که در هر فرآیند مانای اکید به ازای هر s, t، X_t و X_s هم توزیع اند $(h = t - s)$ و لذا توزیع X_t به t بستگی ندارد. یعنی تمام متغیر تصادفی تشکیل دهنده فرآیند، هم توزیع اند. از اینجا نتیجه می شود که $E(X_t)$ (تابع میانگین) به t بستگی ندارد و مقداری ثابت است.

با فرض $n = 2$ نتیجه می شود که در هر فرآیند مانای اکید به ازای هر s, t، توزیع توام X_t و X_s و توزیع توأم X_0, X_{t-s} هم توزیع هستند $(h = -s)$ و لذا توزیع توأم X_t, X_s فقط بستگی به تفاضل s و t دارد. بنابراین هرگونه محاسبه مبتنی بر توزیع توأم X_t, X_s فقط بستگی به تفاضل s و t دارد. از جمله این محاسبات می تواند تابع کواریانس یا همبستگی باشد که در زیر آمده است.

$$cov(X_s, X_t) = E(X_t X_s) - E(X_s)E(X_t)$$

چون در این گونه فرآیند ها $E(X_t)$ و $E(X_s)$ مقداری ثابت اند با توجه به تابع کواریانس در فرآیند مانای اکید، نتیجه می‌گیریم که $E(X_tX_s)$ به $t-s$ بستگی دارد. □

مانای ضعیف: فرآیند $\{X_t,\ t\in T\}$ را مانای ضعیف می گوییم هرگاه اولاً دارای تابع میانگین ثابت باشد ثانیاً تابع کواریانس آن فقط بستگی به تفاضل $t-s$ داشته باشد (یعنی تابعی از $t-s$ باشد). با توجه به این تعریف هر فرآیند مانای اکید حتماً مانای ضعیف هست ولی عکس آن برقرار نیست چرا که شاید توزیع X_t با وجود ثابت بودن تابع میانگین متغیر باشد. □

مثال ۱-۱-۲. فرض کنید A و B دو متغیر تصادفی ناهمبسته باشند. به طوری که داشته باشیم $E(A)=E(B)=\circ$ و $\sigma_A^{\text{۲}}=\sigma_B^{\text{۲}}=\sigma^{\text{۲}}$. همچنین فرض کنید ω نقطه‌ی دلخواهی در فاصله $(\circ,\pi)$ باشد. به ازای هر $t\geqslant\circ$ متغیر تصادفی زیر را در نظر بگیرید.

$$X_t=A\cos(\omega t)+B\sin(\omega t)$$

نشان می دهیم که فرآیند $\{X_t:\ t\geqslant\circ\}$ فرآیندی مانای ضعیف است ولی مانای اکید نیست. با محاسبه تابع میانگین و کواریانس خواهیم داشت

$$E(X_t)=E(A)\cos(\omega t)+E(B)\sin(\omega t)=\circ$$

$$cov(X_t,X_s)=E(X_tX_s)=E[(A\cos(\omega t)+B\sin(\omega t))(A\cos(\omega s)+B\sin(\omega s))]$$

$$=E[A^{\text{۲}}\cos(\omega t)\cos(\omega s)+AB\cos(\omega t)\sin(\omega s)+BA\sin(\omega t)\cos(\omega s)$$

$$+B^{\text{۲}}\sin(\omega t)\sin(\omega s)]=\cos(\omega t)\cos(\omega s)E(A^{\text{۲}})+\cos(\omega t)\sin(\omega s)E(AB)$$

$$+\sin(\omega t)\cos(\omega s)E(BA)+\sin(\omega t)\sin(\omega s)E(B^{\text{۲}})$$

بنا بر فرض $E(AB)=E(BA)=\circ$ و $E(B^{\text{۲}})=E(A^{\text{۲}})=\sigma^{\text{۲}}$ (زیرا امید A و B برابر صفر و هر دو ناهمبسته اند) پس

$$\begin{aligned} cov(X_t,X_s) &= \sigma^{\text{۲}}\big(\cos(\omega t)\cos(\omega s)+\sin(\omega t)\sin(\omega s)\big)\\ &= \sigma^{\text{۲}}\cos(\omega(t-s))\end{aligned}$$

ملاحظه می شود که در این فرآیند تابع میانگین ثابت و تابع همبستگی آن فقط به تفاضل t و s بستگی دارد. پس این فرآیند مانای ضعیف است.

در این فرآیند اگر $t=\frac{\pi}{۲\omega}$، آن گاه $X_t=B$ و اگر $s=\frac{\pi}{۴\omega}$ آن گاه $X_s=\frac{\sqrt{۲}}{۲}(A+B)$ و ملاحظه می شود که با تغییر s و t مقادیری که فرآیند اختیار می کند تغییر می کند و در نتیجه توزیع X_t با تغییر t ثابت نمی ماند یعنی فرآیند مانای اکید نیست.

تعریف ۱-۱-۳. تابع f را از مرتبه $o(g)$ گوییم و می نویسیم $f(h)=o(g(h))$، هر گاه برای $g(h)\neq ۰$ داشته باشیم $lim_{h\to ۰}\frac{f(h)}{g(h)}=۰$ یا $lim_{h\to\infty}\frac{f(h)}{g(h)}=۰$
برای مثال تابع $sinx^۲$ از مرتبه x می باشد زیرا

$$lim_{x\to ۰}\frac{sinx^۲}{x}=lim_{x\to ۰}x\frac{sinx^۲}{x^۲}=۰\times ۱=۰$$

تعریف ۱-۱-۴. فرآیند تصادفی $\{N_t:\ t\geqslant ۰\}$ را هر گاه N_t تعداد کل پیشامد هایی باشد که تا زمان t رخ داده است و در شرایط زیر صدق کند یک فرآیند شمارشی گوییم [۲].

الف) $N_t\geqslant ۰$؛

ب) N_t یک عدد صحیح باشد؛

ج) به ازای $t>s$، $N_t\geq N_s$ باشد؛

د) به ازای $t>s$، N_t-N_s برابر تعداد پیشامدهای اتفاق افتاده در فاصله $(s,t]$ می باشد.

□

تعریف ۱-۱-۵. فرآیند شمارشی $\{N_t:\ t\geqslant ۰\}$ را فرآیند پواسن با پارامتر $\lambda>۰$ گوییم اگر[۲]:

الف) $N_۰=۰$؛

ب) $\{N_t:\ t\geqslant ۰\}$ دارای نموّ های مستقل و مانا باشد؛

ج) به ازای $۰<s<t$ ، نموّ N_t-N_s دارای توزیع پواسن با پارامتر $\lambda(t-s)$ باشد به عبارت دیگر؛

$$P[N_t - N_s = k] = e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!} \qquad k = ۰, ۱, ۲, ...$$

□

از تعریف بالا ویژگی های زیر برای فرآیند پواسن حاصل می شود.

الف) به ازای هر $t > ۰$، N_t دارای توزیع پواسن با پارامتر λt می باشد، زیرا با توجه به شرط (د) بالا $N_t = N_t - N_۰$؛

ب) تابع میانگین در فرآیند پواسن باتوجه به ویژگی (الف) و مقدار امید ریاضی در توزیع پواسن، به صورت زیر است

$$E(N_t) = \lambda t$$

ج) تابع همبستگی فرآیند پواسن به شکل زیر محاسبه می شود. فرض کنید $۰ < s < t$، در این حالت

$$\begin{aligned}
cov(N_s, N_t) &= E(N_t N_s) - E(N_t)E(N_s) \\
&= E((N_t - N_s + N_s)(N_s - N_۰)) - \lambda^۲ ts \\
&= E((N_t - N_s)(N_s - N_۰)) + E(N_s^۲) - \lambda^۲ ts \\
&= E(N_t - N_s)E(N_s) + E(N_s^۲) - \lambda^۲ ts
\end{aligned}$$

با استفاده از استقلال نموّها و موارد زیر

$$E(N_t - N_s) = \lambda(t-s), \quad E(N_s) = \lambda s, \quad E(N_s^۲) = \lambda s + \lambda^۲ s^۲$$

خواهیم داشت

$$cov(N_s, N_t) = \lambda(t-s)\lambda s + \lambda s + \lambda^۲ s^۲ - \lambda^۲ ts = \lambda s$$

و اگر $t = s$، آنگاه

$$cov(N_s, N_t) = var(N_t) = \lambda t$$

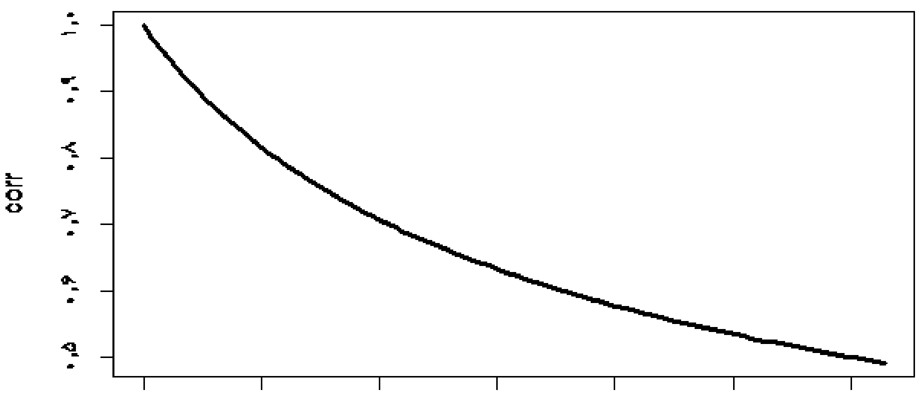
پس با توجه به محاسبات بالا، تابع همبستگی فرآیند پواسن به صورت زیر بدست می آید.

$$cov(N_s, N_t) = \lambda \min\{t, s\} \qquad (۱-۱)$$

با توجه به (۱-۱) می توان ضریب همبستگی N_t و N_s را بدست آورد.

$$\begin{aligned} corr(N_s, N_t) &= \frac{cov(N_s, N_t)}{\sqrt{var(N_t)}\sqrt{var(N_s)}} = \frac{\lambda \min\{t, s\}}{\sqrt{\lambda t}\sqrt{\lambda s}} = \frac{\min\{t, s\}}{\sqrt{st}} \\ &= \sqrt{\frac{s}{t}} \qquad (s < t) \\ &= \sqrt{\frac{s}{s+u}} \qquad (u > 0, t = s + u) \end{aligned}$$

از اینجا نتیجه می شود که هر چه فاصله بین s و t بیشتر باشد (u بزرگتر) ضریب همبستگی N_t و N_s کمتر می شود. نمودار تابع ضریب همبستگی(به عنوان تابعی از u و برای s ثابت) به صورت زیر است.



شکل ۱-۱: تابع ضریب همبستگی در فرآیند پواسن

از (الف) و (ب) نتیجه می شود که فرآیند پواسن مانای اکید نیست. از (ج) هم نتیجه می شود این فرآیند مانای ضعیف نیست.

د) با استفاده از تعریف فرآیند پواسن به آسانی می توان احتمالات زیر را بدست آورد:

$$P[N_{t+h} - N_t = 1] = \frac{e^{-\lambda h}(\lambda h)^1}{1!} = \lambda h e^{-\lambda h} = \lambda h[\textstyle\sum_{n=0}^{\infty}(-1)^n \frac{(\lambda h)^n}{n!}]$$
$$= \lambda h[1 - \lambda h + \frac{\lambda^2 h^2}{2!} + ...] = \lambda h - \lambda^2 h^2 + \frac{\lambda^3 h^3}{3!} + ... = \lambda h + o(h)$$

همچنین،

$$P[N_{t+h} - N_t = 0] = e^{-\lambda h} = 1 - \lambda h + o(h)$$

و داریم،

$$\begin{aligned} P[N_{t+h} - N_t \geqslant 2] &= 1 - P[N_{t+h} - N_t = 0] - P[N_{t+h} - N_t = 1] \\ &= 1 - (1 - \lambda h + o(h)) - \lambda h + o(h) \\ &= o(h) \end{aligned}$$

نکته ۱: فرض کنید فرآیند $\{N_t : t \geqslant 0\}$ یک فرآیند پواسن با پارامتر $\lambda > 0$ باشد و قرار دهید

X_1: زمان لازم برای اولین ورودی(رخ دادن اولین اتفاق)

X_2: زمان بین ورودی اول و دوم

$\vdots$

X_n: زمان بین ورودی $n-1$ام و nامی

X_nها متغیر های تصادفی مستقل و هم توزیع، با توزیع نمایی و میانگین $\frac{1}{\lambda}$ می باشد.

مثال ۱-۱-۶. فرض کنید فرآیند $\{N_t : t \geqslant 0\}$ یک فرآیند پواسن با پارامتر $\lambda > 0$ باشد و $\{X_n : n \geqslant 0\}$ نشان دهنده زمان بین ورودی ها باشند اگر بدانیم تعداد اتفاقات تا زمان t برابر یک باشد، توزیع X_1 را در فاصله $(0, t)$ بدست بیاورید[۲].