نظریه مفصل و مدل‌سازی وابستگی

دکتر حکیم بکری‌زاده

امروز کتابخوانی و علم‌آموزی نه تنها یک وظیفه‌ی ملی، که یک واجب دینی است[1].

مقام معظم رهبری

در عصر حاضر یکی از شاخصه‌های ارزیابی رشد، توسعه و پیشرفت فرهنگی هر کشوری میزان تولید کتاب، مطالعه و کتاب‌خوانی مردم آن مرز و بوم است. ایران اسلامی نیز از دیرباز تاکنون با داشتن تمدنی چندهزارساله و مراکز متعدد علمی، فرهنگی، کتابخانه‌های معتبر، علما و دانشمندان بزرگ با آثار ارزشمند تاریخی، سرآمد دولت‌ها و ملت‌های دیگر بوده و در عرصه فرهنگ و تمدن جهانی به‌سان خورشیدی تابناک همچنان می‌درخشد و با فرزندان نیک‌نهاد خویش هنرنمایی می‌کند. چه کسی است که در دنیا با دانشمندان فرزانه و نام‌آور ایرانی همچون ابوعلی سینا، ابوریحان بیرونی، فارابی، خوارزمی و ... همچنین شاعران برجسته‌ای نظیر فردوسی، سعدی، مولوی، حافظ و ... آشنا نباشد و در مقابل عظمت آنها سر تعظیم فرود نیاورد. تمامی این افتخارات ارزشمند، برگرفته از میزان عشق و علاقه فراوان ملت ما به فراگیری علم و دانش از طریق خواندن و مطالعه منابع و کتاب‌های گوناگون است. به شکرانه الهی، تاریخ و گذشته ما، همیشه درخشان و پربار است. ولی اکنون در این زمینه در چه جایگاهی قرار داریم؟ آمار و ارقام ارائه‌شده از سوی مجامع و سازمان‌های فرهنگی در مورد سرانه مطالعه هر ایرانی، برایمان چندان امیدوارکننده نمی‌باشد.

کتاب، دروازه‌ای به سوی گستره دانش و معرفت است و کتاب خوب، یکی از بهترین ابزارهای کمال بشری است. همه دستاوردهای بشر در سراسر عمر جهان، تا آنجا که قابل کتابت بوده است، در میان دست‌نوشته‌هایی است که انسان‌ها پدید آورده و می‌آورند. در این مجموعه بی‌نظیر، تعالیم الهی، درس‌های پیامبران به بشر، و همچنین علوم مختلفی است که سعادت بشر بدون آگاهی از آن‌ها امکان‌پذیر نیست. کسی که با دنیای زیبا و زندگی‌بخش کتاب ارتباط ندارد بی‌شک از مهم‌ترین دستاورد انسانی و نیز از بیشترین معارف الهی و بشری محروم است. با این دیدگاه، به‌روشنی می‌توان ارزش و مفهوم رمزی عمیق در این حقیقت تاریخی را دریافت که اولین خطاب خداوند متعال به پیامبر گرامی اسلام(ص) این است که »بخوان!« و در اولین سوره‌ای که بر آن فرستاده عظیم‌الشأن خداوند، فرود آمده، نام «قلم» به تجلیل یاد

1. https://farsi.khamenei.ir/message-content?id=2696

شده‌است: «اِقْرَأْ وَ رَبُّکَ الْاَکْرَمُ. اَلَّذی عَلَّمَ بِالْقَلَم» در اهمیت عنصر کتاب برای تکامل جامعه انسانی، همین بس که تمامی ادیان آسمانی و رجال بزرگ تاریخ بشری، از طریق کتاب جاودانه مانده‌اند.

دانشگاه پیام‌نور با گستره جغرافیایی ایران‌شمول خود با هدف آموزش برای همه، همه‌جا و همه‌وقت، به‌عنوان دانشگاهی کتاب‌محور در نظام آموزش عالی کشورمان، افتخار دارد جایگاه اندیشه‌سازی و خردورزی بخش عظیمی از جوانان جویای علم این مرز و بوم باشد. تلاش فراوانی در ایام طولانی فعالیت این دانشگاه انجام پذیرفته تا با بهره‌گیری از تجربه‌های گرانقدر استادان و صاحب‌نظران برجسته کشورمان، کتاب‌ها و منابع آموزشی درسی شاخص و خودآموز تولید شود. در آینده هم، این مهم با هدف ارتقای سطح علمی، روزآمدی و توجه بیشتر به نیازهای مخاطبان دانشگاه پیام‌نور با جدیت ادامه خواهد داشت. به‌طور قطع استفاده از نظرات استادان، صاحب‌نظران و دانشجویان محترم، ما را در انجام این وظیفه مهم و خطیر یاری‌رسان خواهد بود. پیشاپیش از تمامی عزیزانی که با نقد، تصحیح و پیشنهادهای خود ما را در انجام این وظیفه خطیر یاری می‌رسانند، سپاسگزاری می‌نماییم. لازم است از تمامی اندیشمندانی که تاکنون دانشگاه پیام‌نور را منزلگه اندیشه‌سازی خود دانسته و ما را در تولید کتاب و محتوای آموزشی درسی یاری نموده‌اند، صمیمانه قدردانی گردد. موفقیت و بهروزی تمامی دانشجویان و دانش‌پژوهان عزیز آرزوی همیشگی ما است.

دانشگاه پیام‌نور

فهرست مطالب

پیشگفتار

یکی از مباحث مهم در نظریه احتمال و آمار، ساخت توزیع‌های چندمتغیره به‌منظور بیان و تحلیل وابستگی میان متغیرها است؛ موضوعی که از دیرباز مورد توجه آماردانان بسیاری قرار گرفته است. در سال‌های اخیر، یکی از روش‌های مؤثر و نوین در این زمینه، استفاده از توابع مفصل بوده است. مفهوم مفصل نخستین‌بار به‌صورت تحلیلی توسط اسکلار (۱۹۵۹) در قالب قضیه‌ای مطرح شد که ارتباط میان توابع توزیع تک‌بعدی و توابع توزیع چندمتغیره را تبیین می‌کرد. براساس این قضیه، هر توزیع توأم را می‌توان به‌صورت یکتایی با استفاده از یک مفصل و توزیع‌های حاشیه‌ای پیوستهٔ متناظر بیان کرد. توابع مفصل این امکان را فراهم می‌سازند که ساختار وابستگی میان متغیرها به‌صورت یک مدل مستقل از توزیع‌های حاشیه‌ای توصیف شود؛ به عبارت دیگر، مفصل، وابستگی متغیرها را از توزیع‌های حاشیه‌ای جدا می‌کند و بدین‌ترتیب، ابزار قدرتمندی برای مدل‌سازی نوع و شدت ارتباط بین متغیرها در اختیار پژوهشگران قرار می‌دهد.

در دهه‌های اخیر، پیشرفت‌های چشمگیری در نظریه و کاربرد توابع مفصل حاصل شده است. این توابع نه‌تنها به دلیل ویژگی ناپارامتری بودن، بلکه به سبب توانایی آن‌ها در بیان دقیق ساختار وابستگی میان متغیرهای تصادفی، جایگاه ویژه‌ای در آمار مدرن پیدا کرده‌اند. از سوی دیگر، علوم جدید و فناوری‌های روز نشان داده‌اند که بسیاری از پدیده‌های طبیعی را نمی‌توان با فرض استقلال میان عوامل توضیح داد. در چنین

شرایطی، برای تحلیل پدیده‌هایی که میان متغیرهای آن‌ها وابستگی وجود دارد، لازم است ابزارهایی برای اندازه‌گیری و مدل‌سازی وابستگی مورد استفاده قرار گیرد. در این میان، توابع مفصل نقش اساسی دارند؛ زیرا مفصل‌ها می‌توانند ساختار وابستگی را به‌صورت یک مدل دقیق نمایش دهند و بر پایهٔ آن، روابط میان متغیرها را توصیف کنند. همچنین، در مسائل مالی و سنجش ریسک نیز کاربرد گسترده‌ای دارند، چرا که در این حوزه‌ها، شناخت ساختار وابستگی میان متغیرها در برآورد و تحلیل ریسک نقش تعیین‌کننده‌ای ایفا می‌کند.

بر همین اساس، کتاب حاضر مقدمه‌ای است بر نظریهٔ مفصل و مدل‌سازی وابستگی می‌باشد که در آن مفاهیم بنیادین، نظریه‌های اصلی و اندازه‌های وابستگی با بیانی ساده و روشن، همراه با مثال‌ها و تمرین‌های متنوع تشریح شده‌اند. در این اثر، تلاش شده است ضمن معرفی یافته‌های جدید پژوهشگران، بستری مناسب برای درک و کاربرد توابع مفصل در زمینه‌های مختلف علمی فراهم شود.

امید است که این کتاب گامی کوچک اما مؤثر در جهت آشنایی دانشجویان کلیه مقاطع تحصیلی و پژوهشگران با نظریه و کاربرد توابع مفصل باشد و زمینه‌ساز تحقیقات گسترده‌تر در این حوزه گردد.

حکیم بکری‌زاده

فهرست نمادها

تابع چگالی	f
تابع توزیع	F
تابع توزیع بقاء	$\overline{F}$
کلاس تمام توابع مفصل	L
مفصل	C
تابع چگالی مفصل	c
مفصل بقاء	$\overline{C}$
تابع مولد مفصل ارشمیدسی	$\varphi(t)$
تابع محدب مفصل مقادیر فرین	$A(t)$
ضریب همبستگی پیرسن	ρ
ضریب همبستگی کندال	τ_k
ضریب همبستگی اسپیرمن	ρ_S
وابستگی دمی قوی بالا	λ_U
وابستگی دمی قوی پایین	λ_L
وابستگی دمی ضعیف بالا	χ_U
وابستگی دمی ضعیف پایین	χ_L
مفصل مستقل	Π
تابع وزنی	$W(x)$
متغیر تصادفی وزنی	X^w
تابع چگالی وزنی	f^w

فصل اول

مروری بر متغیر تصادفی (پیوسته)

هدف کلی

هدف کلی این فصل، مروری بر مفاهیم ویژگی‌های متغیر تصادفی پیوسته و توزیع‌های احتمال، به‌منظور درک مفاهیم بنیادی احتمال در حالت پیوسته و یادگیری نحوه مدل‌سازی و تحلیل پدیده‌های تصادفی است.

هدف‌های یادگیری

انتظار می‌رود فراگیر پس از مطالعه این فصل بتواند به سؤالات زیر پاسخ دهد.

۱. متغیر تصادفی چیست و چگونه تعریف می‌شود.
۲. ویژگی‌های یک متغیر تصادفی کدام‌اند.
۳. تابع توزیع احتمال چیست و چه خصوصیات مهمی دارد.
۴. تابع توزیع احتمال توأم چیست و چه ویژگی‌هایی دارد.
۵. در چه شرایطی دو متغیر تصادفی از یکدیگر مستقل هستند.
۶. نحوه محاسبه میانگین، واریانس و سایر ویژگی‌های آماری در توزیع‌های پیوسته چگونه است.

مقدمه

در دنیای آمار و احتمال، **متغیر تصادفی** یکی از اساسی‌ترین و پرکاربردترین مفاهیم است که پایه بسیاری از تحلیل‌های آماری را تشکیل می‌دهد. یک متغیر تصادفی در واقع تابعی است که هر نتیجه ممکن از یک آزمایش تصادفی را به یک مقدار عددی نسبت می‌دهد. به

بیان ساده‌تر، متغیر تصادفی راهی برای کمی کردن نتایج غیرقطعی و تصادفی است که در پدیده‌های مختلف مشاهده می‌کنیم. اهمیت این مفهوم در آمار به حدی است که بدون درک صحیح آن، نمی‌توان بسیاری از روش‌های پیشرفته تحلیل داده را فهمید یا به‌کار برد. هر متغیر تصادفی با توزیع احتمال آن به‌طور کامل مشخص می‌شود. توزیع احتمال در واقع توصیفی ریاضی از رفتار احتمالی متغیر است که نشان می‌دهد چه مقادیری با چه احتمالاتی رخ می‌دهند. در این راستا، در این فصل، به بررسی اجمالی از مفاهیم و یژگی‌های متغیر تصادفی پیوسته و توزیع‌های احتمال، به منظور درک مفاهیم بنیادی احتمال در حالت پیوسته و یادگیری نحوه مدل‌سازی و تحلیل پدیده‌های تصادفی است.

۱-۱ تعاریف مقدماتی

فرض کنید نماد $\Re$ نشان‌دهنده خط حقیقی معمولی $(-\infty,\infty)$، $\overline{\Re}$ نشان‌دهنده خط حقیقی توسعه‌یافته $[-\infty,\infty]$، $\overline{\Re}^2$ نشان‌دهنده صفحه حقیقی $\overline{\Re}\times\overline{\Re}$ و B یک حاصل‌ضرب دکارتی دو فاصله بسته به‌صورت $B=[x_1,x_2]\times[y_1,y_2]$ یک مستطیل در $\overline{\Re}^2$ باشد که رأس‌های مستطیل نقاط (x_1,y_1)، (x_2,y_1)، (x_1,y_2) و (x_2,y_2) می‌باشند. همچنین فرض کنید، مربع واحد I^2 حاصل‌ضرب $I\times I$ باشد ($I=[0,1]$).

تابع حقیقی دومتغیره F، تابعی است که دامنه آن را با $DomF$، که یک زیرمجموعه از $\overline{\Re}^2$ است و برد آن را با $RanF$ نشان می‌دهیم که یک زیرمجموعه از $\Re$ است.

تعریف ۱-۱. فرض کنید S_1 و S_2 زیرمجموعه‌های ناتهی $\overline{\Re}$ و F تابعی است که دامنه آن $DomF=S_1\times S_2\subseteq\overline{\Re}^2$ و برد آن $RanF\subseteq\Re$ می‌باشد.

اگر $B=[x_1,x_2]\times[y_1,y_2]$ یک مستطیل باشد که رئوس آن متعلق به $DomF$ است، آنگاه F حجم مجموعه B به‌صورت زیر تعریف می‌شود.

$$V_F(B)=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\ .$$

تعریف ۱-۲. تابع F در تعریف (۱-۱) را دو-صعودی گویند اگر برای هر مستطیل B که رئوس آن‌ها در دامنه F قرار دارد. داشته باشیم؛ $V_F(B)\geq 0$.

توجه شود که دو–صعودی بودن F، معادل با نانزولی بودن آن نیست. مثال زیر می‌تواند گویای این مطلب باشد.

مثال ۱–۱. فرض کنید F روی $I^2=[0,1]\times[0,1]$ با ضابطه $F(x,y)=\max(x,y)$، تعریف شده باشد. در این صورت F تابعی نانزولی از x و y است، برای؛

$$V_F(I^2)=\max(1,1)-\max(1,0)-\max(0,1)+\max(0,0)=-1<0\ .$$

بنابراین F تابعی دو–صعودی نیست.

تعریف ۱–۳. فرض کنید a_1 کمترین مقدار مجموعه S_1 و b_1 کمترین مقدار مجموعه S_2 باشد. گوییم تابع F از $S_1\times S_2$ به $\Re$ جهت‌دار است، اگر برای هر (x,y) در $S_1\times S_2$ داشته باشیم:

$$F(a_1,y)=0=F(x,b_1)$$

مثال ۱–۲. فرض کنید F با ضابطه $F(x_1,x_2)=\dfrac{(x_1+1)(e^{x_2}-1)}{x_1+2e^{x_2}-1}$ یک تابع تعریف شده روی دامنه $[-1,1]\times[0,\infty]$ باشد. آنگاه $F(x,0)=0$ و $F(-1,y)=0$ است؛ بنابراین F جهت‌دار است.

تعریف ۱–۴. فرض کنید a_2 بیشترین مقدار مجموعه S_1 و b_2 بیشترین مقدار مجموعه S_2 باشد. گوییم تابع F از $S_1\times S_2$ به $\Re$ صادق است، اگر توابع F_1 با دامنه S_1 و F_2 با دامنه S_2 وجود داشته باشند به‌طوری‌که؛

$$F_1(x)=F(x,b_2),\qquad \forall x\in S_1$$
$$F_2(y)=F(a_2,y),\qquad \forall y\in S_2$$

در این صورت توابع F_1 و F_2 را توابع حاشیه‌ای تابع F می‌نامند.

قضیه ۱–۱. فرض کنید S_1 و S_2 زیرمجموعه‌های ناتهی $\overline{\Re}$ باشند. اگر تابع F از $S_1\times S_2$ به $\Re$ تابعی دو–صعودی باشد. آنگاه برای هر $(x_1,y_1),(x_2,y_2)$ در $S_1\times S_2$ داریم:

$$\left|F(x_2,y_2)-F(x_1,y_1)\right|\le\left|F_1(x_2)-F_1(x_1)\right|+\left|F_2(y_2)-F_2(y_1)\right|.$$

برهان. با توجه به نامساوی مثلثی داریم؛

$$\begin{aligned}\left|F(x_2,y_2)-F(x_1,y_1)\right|&=\left|F(x_2,y_2)-F(x_1,y_2)+F(x_1,y_2)-F(x_1,y_1)\right|\\&\le\left|F(x_2,y_2)-F(x_1,y_2)\right|+\left|F(x_1,y_2)-F(x_1,y_1)\right|\end{aligned}$$

از آنجایی که تابع F، دو–صعودی است و در تعریف (۱–۳) صدق می‌کند، پس برای $x_1\le x_2$ داریم:

$$F(x_2,y_2)-F(x_1,y_2)\ge 0$$

و چون F در تعریف (۱–۴) نیز صادق است پس برای $y_2\le b_2$، داریم؛

$$0\le F(x_2,y_2)-F(x_1,y_2)\le F_1(x_2)-F_1(x_1).$$

همچنین، برای $x_1\ge x_2$، داریم:

$$F_1(x_2)-F_1(x_1)\le F(x_2,y_2)-F(x_1,y_2)\le 0.$$

در نتیجه خواهیم داشت:

$$\left|F(x_2,y_2)-F(x_1,y_2)\right|\le\left|F_1(x_2)-F_1(x_1)\right|,\quad x_1,x_2\in S_1\ .$$

همچنین، با استدلالی مشابه نتیجه گرفته می‌شود.

$$\left|F(x_2,y_2)-F(x_1,y_1)\right|\le\left|F_2(y_2)-F_2(y_1)\right|,\quad y_1,y_2\in S_2\ .$$

در نتیجه، با جایگزینی عبارات بالا در سمت راست نابرابری، قضیه اثبات می‌شود. توجه شود که هر تابع توزیع توأم در شرط‌های قضیه (۱–۱) صدق می‌کند.

۱–۲ متغیر تصادفی

متغیر تصادفی متغیری است که مقدار آن از اندازه‌گیری یک فرایند تصادفی به‌دست می‌آید. به‌طور رسمی، برای یک مدل احتمال با فضای نمونه S، تابع حقیقی X را که دامنه آن Sو بردش زیرمجموعه‌ای از اعداد حقیقی است یک متغیر تصادفی روی مدل

احتمال می‌نامند (یعنی؛ $X:S\to\Re$). در نظریه احتمال و آمار تابع توزیع احتمال بیانگر احتمال هر یک از مقادیر متغیر تصادفی (در مورد متغیر گسسته) و یا احتمال قرار گرفتن متغیر در یک بازه مشخص (در مورد متغیر تصادفی پیوسته) می‌باشد.

۱-۳ تابع توزیع احتمال

فرض کنید X یک متغیر تصادفی روی مدل احتمال باشد، تابع $F(x)=\Pr[X\le x]$ را که در آن x یک مقدار حقیقی است تابع توزیع احتمال متغیر تصادفی X می‌نامند. برای تابع توزیع احتمال F شرایط زیر همواره برقرار است:

الف) $\lim_{x\to-\infty} F(x)=0$ و $\lim_{x\to\infty} F(x)=1$،

ب) تابع توزیع احتمال F غیرنزولی است؛ یعنی؛ برای مقادیر $x_1\le x_2$ داشته باشیم؛ $F(x_1)\le F(x_2)$.

ج) تابع توزیع احتمال F همواره از راست پیوسته است؛ یعنی؛ $\lim_{x\to a^+} F(x)=F(a)$.

تعریف ۱-۵. فرض کنید متغیر تصادفی X دارای تابع توزیع احتمال $F(x)$ باشد. تابع $F^{-1}(t)$ را نیمه معکوس $F(x)$ با دامنه I می‌نامیم اگر شرایط زیر برقرار باشد:

الف) اگر t در برد F باشد، آنگاه $F(F^{-1}(t))=t$،

ب) اگر t در برد F نباشد، آنگاه $F^{-1}(t)=\inf\{x\,|\,F(x)\ge t\}=\sup\{x\,|\,F(x)\le t\}$،

ج) اگر F اکیداً صعودی باشد، آنگاه $F^{(-1)}=F^{-1}$.

تعریف ۱-۶. فرض کنید F تابع توزیع احتمال توأم متغیرهای تصادفی X_1 و X_2 با توابع توزیع احتمال حاشیه‌ای $F_۱$ و $F_۲$ باشد. آنگاه F در نامساوی زیر صدق می‌کند:

$$\max\{F_1(x_1)+F_2(x_2)-1,0\}\le F(x_1,x_2)\le\min\{F_1(x_1),F_2(x_2)\}$$

کران‌های بالایی و پایینی رابطه بالا، خود تابع توزیع‌های توامی هستند که به کران‌های فرشه-هافدینگ معروف‌اند.

۱-۴ تابع توزیع احتمال توأم

تابع توزیع احتمال F از $\Re^p$ به فاصله $[0,1]$ را که برای هر $\tilde{x}=(x_1,x_2,...,x_p)\in\Re^p$ به‌صورت

$$F(x_1,x_2,...,x_p)=\Pr[X_1\le x_1,\, X_2\le x_2\,,\, ...\,, X_p\le x_p]$$

تعریف می‌شود تابع توزیع احتمال توأم$\left(X_1,X_2,...,X_p\right)$می‌نامند. در حالت دوبعدی ($p=2$) تابع بالا به‌صورت زیر تعریف می‌شود.

تعریف ۱-۷. تابع توزیع احتمال توأم متغیرهای تصادفی$\left(X,Y\right)$ را با $F(x,y)$نشان داده و به‌صورت زیر تعریف می‌شود:

$$F(x,y)=P[X\le x,Y\le y]$$

تابع $F(x,y)$ دارای ویژگی‌های زیر می‌باشد:

الف) $0\le F(x,y)\le 1$،

ب) $\lim\limits_{x\to-\infty}F(x,y)=0$ و $\lim\limits_{y\to-\infty}F(x,y)=0$،

ج) $\lim\limits_{y\to+\infty}F(x,y)=F_1(x)$ و $\lim\limits_{x\to+\infty}F(x,y)=F_2(y)$،

که در آن $F_1(x)$ و $F_2(y)$ به ترتیب توابع توزیع احتمال حاشیه‌ای X و Y می‌باشند،

د) برای هر $x_1,x_2,y_1,y_2\in\Re$ اگر $x_1\le x_2$ و $y_1\le y_2$ باشد. در این صورت داریم؛

$$F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\ge 0\ .$$

مثال ۱-۳. فرض کنید H یک تابع توزیع روی $\overline{\Re}^2$ به‌صورت زیر باشد؛

$$H(x,y)=\begin{cases}\dfrac{(x+1)(e^y-1)}{x+2e^y-1}, & (x,y)\in[-1,1]\times[0,\infty]\\ 1-e^{-y}\quad, & (x,y)\in(1,\infty]\times[0,\infty]\\ \quad 0\qquad, & elsewhere\end{cases}$$

در این صورت H یک تابع دو–صعودی است و $H(\infty,\infty)=0$. همچنین، H یک تابع توزیع توأم احتمال با توابع توزیع احتمال حاشیه‌ای به‌صورت زیر می‌باشد:

$$F(x)=U_{[-1,1]}(x) \text{ و } G(y)=\begin{cases} 0, & y\in[-\infty,0) \\ 1-e^{-y}, & y\in[0,\infty] \end{cases}$$

تعریف ۱–۸. فرض کنید (X,Y) یک زوج متغیر تصادفی (مطلقاً) پیوسته باشد، در این صورت داریم:

الف) $F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)dsdt$،

ب) $f(x,y)=\dfrac{\partial^2 F(x,y)}{\partial x\partial y}$،

ج) $f_1(x)=\int_{-\infty}^{+\infty} f(x,y)dy$ و $f_2(y)=\int_{-\infty}^{+\infty} f(x,y)dx$،

د) $f(x|y)=\dfrac{f(x,y)}{f_2(y)}$.

که در آن $f(x,y)$، $f_1(x)$، $f_2(y)$ و $f(x|y)$ به ترتیب، تابع چگالی احتمال توأم (X,Y)، تابع چگالی احتمال حاشیه‌ای X و Y و تابع چگالی احتمال شرطی X به‌شرط Y می‌باشند.

تعریف ۱–۹. متغیرهای تصادفی X و Y مستقل هستند هرگاه برای $x,y\in\Re$ داشته باشیم:

$$F(x,y)=F_1(x)F_2(y).$$

تعریف ۱–۱۰. فرض کنید (X,Y) یک زوج متغیر تصادفی باشند. گوییم X و Y تعویض‌پذیر هستند اگر (X,Y) و (Y,X) هم‌توزیع باشند.

تعریف ۱-۱۱ (تابع توزیع بقاء). تابع توزیع زنده ماندن یا بقای بیشتر از زمان t به‌صورت

$$\bar{F}(x) = \Pr(X \geq x) = 1 - F(x)$$

تعریف می‌شود. همچنین، برای زوج متغیرتصادفی (X,Y)با تابع توزیع توأم $F(x,y)$، تابع بقای توأم (X,Y) را با $\bar{F}(x,y)$ نمایش داده و به‌صورت زیر محاسبه می‌شود:

$$\bar{F}(x,y) = P(X > x, Y > y)\,.$$

به همین ترتیب، در حالت چندمتغیره تابع توزیع بقاء به‌صورت زیر تعریف می‌شود:

$$\bar{F}(x_1, x_2, ..., x_p) = \Pr[X_1 \geq x_1,\, X_2 \geq x_2\,, ...\,, X_p \geq x_p]\,.$$

حالت خاص:

۱. برای $p = 2$ تابع توزیع بقاء برابر است با:

$$\bar{F}(x_1, x_2) = 1 - F_1(x_1) - F_2(x_2) + F_1(x_1)F_2(x_2)\,.$$

۲. برای $p = 3$ تابع توزیع بقاء برابر است با:

$$\begin{aligned}\bar{F}(x_1, x_2, x_3) = 1 &- F_1(x_1) - F_2(x_2) - F_3(x_3) + F_{12}(x_1, x_2) \\ &+ F_{13}(x_1, x_3) + F_{23}(x_2, x_3) - F_{123}(x_1, x_2, x_3)\end{aligned}\,.$$

۱-۵ ویژگی‌هایی از متغیر تصادفی

برای متغیر تصادفی پیوسته X امید ریاضی X را با $E(X)$ نشان داده که به‌صورت زیر تعریف می‌شود:

$$E(X) = \int_{x \in \Re} x\, f(x)dx$$

و برای هر تابعی از متغیر تصادفی X مانند $\Phi(X)$ داریم؛

$$E[\Phi(X)] = \int_{x \in \Re} \Phi(x)\, f(x)dx\,.$$

همچنین، برای متغیر تصادفی پیوسته X واریانس X را با $Var(X)$ نشان داده که به‌صورت زیر تعریف می‌شود:

$$Var(X)=E(X^2)-E^2(X).$$

برای یک زوج متغیر تصادفی (مطلقاً) پیوسته (X,Y)، کوواریانس X و Y را با $Cov(X,Y)$ نشان داده و عبارت است از:

$$Cov(X,Y)=E(XY)-E(X)E(Y)،$$

که در آن؛ $E(XY)=\int_{y\in\Re}\int_{x\in\Re} xy\, f(x,y)dxdy$.

ضریب همبستگی X و Y را با $\rho(X,Y)$ نشان داده و برای یک زوج متغیر تصادفی (مطلقاً) پیوسته (X,Y)، به‌صورت زیر تعریف می‌شود:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

خودآزمایی فصل اول

۱. فرض کنید F روی I^2 با ضابطه $F(x,y)=(2x-1)(2y-1)$ تعریف شده باشد. بررسی کنید آیا تابع F روی I^2 دو–صعودی است؟

۲. به ازای هر $a\in\Re$ تابع توزیع F_a را به‌صورت زیر در نظر بگیرید:

$$F_a(x)=\begin{cases}1\ , \ x\in[a,\infty]\\ 0,\ x\in[-\infty,a)\end{cases}$$

در این صورت، تابع نیمه معکوس F_a را به‌دست آورید.

۳. (توزیع لجستیک دو متغیره گامبل). فرض کنید متغیرهای تصادفی X و Y دارای تابع توزیع احتمال توأم به‌صورت زیر باشد:

$$H(x,y)=(1+e^{-x}+e^{-y})^{-1}$$

الف) ثابت کنید که H یک تابع توزیع دو–صعودی روی $\overline{\Re}$ می‌باشد.

ب) نشان دهید که متغیرهای تصادفی X و Y دارای توابع توزیع احتمال حاشیه‌های لجستیک به‌صورت زیر می‌باشند:

$$F(x)=(1+e^{-x})^{-1} \quad , \quad G(y)=(1+e^{-y})^{-1}$$

۴. فرض کنید متغیرهای تصادفی X و Y دارای تابع توزیع احتمال توأم به‌صورت زیر باشد:

$$H(x,y)=(1+e^{-x}+e^{-y}+(1-\theta)e^{-x-y})^{-1}, \ \ \theta\in[-1,+1]$$

الف) ثابت کنید که H یک تابع توزیع دو–صعودی روی $\overline{\Re}$ به ازای $\theta\in[-1,+1]$ می‌باشد.

ب) نشان دهید که متغیرهای تصادفی X و Y دارای توابع توزیع احتمال حاشیه‌های لجستیک استاندارد می‌باشند.

ج) ثابت کنید به ازای مقدار $\theta=0$ متغیرهای تصادفی X و Y مستقل می‌باشند.

۵. (توزیع نمایی دو متغیره گامبل). فرض کنید متغیرهای تصادفی X و Y دارای تابع توزیع احتمال توأم به‌صورت زیر باشد:

$$H(x,y)=1+e^{-x}+e^{-y}+e^{-(x+y+\theta xy)}, \ \ x\geq 0, y\geq 0, \theta\in[0,+1]$$

الف) ثابت کنید که H یک تابع توزیع دو–صعودی روی $\overline{\Re}$ به ازای $\theta\in[0,1]$ می‌باشد.

ب) توابع توزیع احتمال حاشیه‌ای متغیرهای تصادفی X و Y را به‌دست آورید.

ج) ثابت کنید به ازای مقدار $\theta=0$ متغیرهای تصادفی X و Y مستقل می‌باشند.

فصل دوم

مبانی و مفاهیم پایه مفصل

هدف کلی

هدف این فصل، آشنایی فراگیر با مبانی و مفاهیم پایه‌ای مفصل (Copula) است؛ مفاهیمی که زیربنای مباحث مربوط به توزیع‌های احتمال دوبعدی و چندبعدی را تشکیل می‌دهند. در این فصل، یادگیرنده با اصول اساسی مفصل آشنا می‌شود، روابط میان متغیرهای تصادفی را درک می‌کند و می‌آموزد چگونه از این مفاهیم برای تحلیل ساختار وابستگی میان متغیرها استفاده کند. این آشنایی، زمینه را برای فهم بهتر مباحث پیشرفته‌تر و کاربردهای عملی آن در علوم مختلف فراهم می‌سازد.

هدف‌های یادگیری

از فراگیر انتظار می‌رود پس از مطالعه این فصل، بتواند به سؤالات زیر پاسخ دهد.

۱. مفصل چیست و چه ویژگی‌هایی دارد.
۲. مهم‌ترین خصوصیات مفصل کدام‌اند.
۳. هدف‌ها و کاربردهای مفصل کدام‌اند.
۴. چند نمونه از کاربردهای مفصل در علوم مختلف را بیان کرده و به‌صورت خلاصه تشریح دهد.
۵. کدام قضیه، مبنای تحقیقات مربوط به مفصل و ساختار وابستگی بوده است.
۶. مفاهیم پایه یک مفصل کدام‌اند.
۷. پارامترهای یک مفصل چگونه برآورد می‌شوند.