دانشکده‌ی علوم ریاضی

احتمال و کاربردآن — ۱۷ فروردین ۹۳

جلسه‌ی ۱۱: برخی مفاهیم احتمال و چند توزیع احتمال

مدرّس: دکتر شهرام خزائی — *نگارنده: مهدی یوسفی*

در این قسمت ابتدا با سه مفهوم اساسی در نظریه احتمال آشنا می‌شویم و سپس با معرفی چند توزیع احتمال این قسمت را به اتمام می‌رسانیم.

۱ تابع توزیع تجمعی

ایده‌ی تابع توزیع تجمعی از اینجا نشأت گرفته است که ما گاهی نیازمند محاسبه‌ی احتمال‌های زیر هستیم.

$$\Pr\{X=a\} \quad \Pr\{X\le a\} \quad \Pr\{X>a\}$$
$$\Pr\{X\ge a\} \quad \Pr\{a<X\le b\} \quad \Pr\{a\le X<b\}$$
$$\Pr\{a<X<b\}$$

تعریف ۱ *تابع توزیع تجمعی* [۱] *متغیر تصادفی X را با نماد $F_X(x)$ نشان می‌دهیم و به صورت زیر تعریف می‌کنیم.*

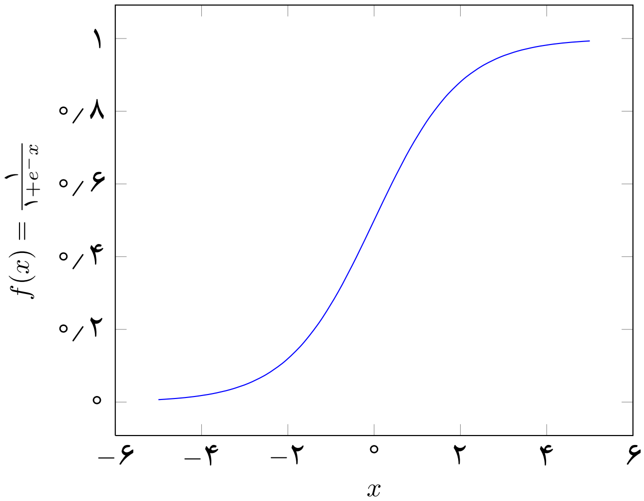
$$F_X(x)=\Pr\{X\le x\}$$

قضیه ۱ *تابعی مانند $F_X(x)$ می‌تواند تابع توزیع تجمعی یک متغیر تصادفی باشد به شرطی که اولاً غیرنزولی باشد ثانیاً در هر نقطه از راست پیوسته باشد. ثالثاً حد آن در بی‌نهایت در تساوی‌های زیر صدق کند.*

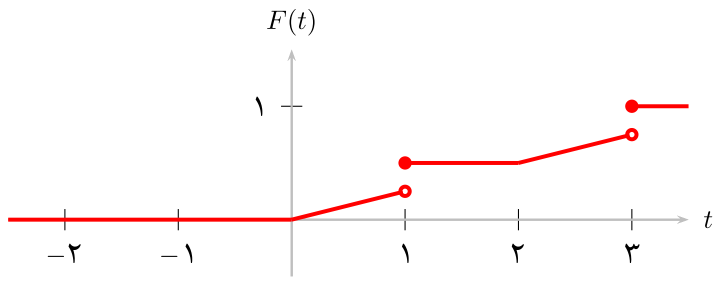
$$\lim_{x\to-\infty} F_X(x)=0 \qquad \lim_{x\to\infty} F_X(x)=1$$

مثال ۱ *آیا توابعی با نمودار های زیر می‌تواند معرف یک تابع توزیع تجمعی باشد.*
نمودار اول:

[۱] COMULATIVE DISTRIBUTION FUNCTION



بله زیرا هم صعودی است و در هر نقطه از راست پیوسته است و همچنین حد آن در مثبت بی‌نهایت برابر یک است. در منفی بی‌نهایت برابر صفر است.
نمودار دوم:



بله در هر نقطه از راست پیوسته است تابع صعودی است و شرط حدها هم پوشش داده می‌شود.

۱.۱ روابط پیرامون تابع توزیع تجمعی

حال مقادیر احتمال‌هایی را که در ابتدای این بخش آوردیم را محاسبه می‌کنیم.

$$\Pr\{X \leq a\} = F_X(a) \qquad \Pr\{X > a\} = ۱ - F_X(a)$$
$$\Pr\{X < a\} = \lim_{x \to a^-} F_X(x) = F_X(a^-) \qquad \Pr\{a < X \leq b\} = F_X(b) - F_X(a)$$
$$\Pr\{X = a\} = F_X(a) - F_X(a^-) \qquad \Pr\{X \geq a\} = ۱ - F_X(a^-)$$
$$\Pr\{a \leq X < b\} = F_X(b^-) - F_X(a^-) \qquad \Pr\{a < X < b\} = F_X(b^-) - F_X(a)$$

۲ امیدریاضی

امید ریاضی یکی از مفاهیم پایه‌ای در نظریه احتمال است که برای یک متغیر تصادفی مانند X با تابع جرم احتمال $p(x)$ به صورت زیر تعریف می‌شود.

تعریف ۲ *فرض کنید X یک متغیر تصادفی با برد $\{x_۱, x_۲, \cdots\}$ و تابع جرم احتمال $p(x)$ باشد. مقدار متوسط (امید ریاضی) متغیر تصادفی X با $\mathrm{E}[X]$ نشان داده می‌شوند و به صورت زیر تعریف می‌شود.*

$$\mathrm{E}[X] = \sum_i x_i p(x_i)$$

به بیان بهتر امید ریاضی، میانگین وزن‌دار مقادیری است که متغیر تصادفی X با وزنی برابر با مقدار احتمال خود می‌پذیرد. از این رو از امید ریاضی با واژه‌ی مقدار متوسط یا مقدار مورد انتظار هم یاد می‌کنند.
در آزمایش‌هایی که برآمد آن برد یا باخت است امید ریاضی اطلاعات سودمندی در اختیار ما قرار می‌دهد. هنگامی که این آزمایش‌ها بارها تکرار می‌شوند، عددی که امید ریاضی نشان می‌دهد، متوسط مقدار مورد انتظار ما برای تعداد بردها در هر بازی است.

نکته ۱ *هر عدد ثابت یک متغیر تصادفی است با احتمال یک و مقدار متوسط آن برابر مقدار همان عدد ثابت است.*

$$\mathrm{E}[\alpha] = \alpha$$

مثال ۲ *از نکته‌ی قبل نتیجه می‌شود چون $\mathrm{E}[X]$ عددی ثابت است، متوسط آن برابر خودش است؛ یعنی:*

$$\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]$$

۱.۲ مقدار متوسط تابعی از یک متغیر تصادفی

فرض کنید تابعی مثل $g : \mathbb{R} \to \mathbb{R}$ روی اعداد حقیقی داده شده‌باشد. هدف ما در این قسمت یافتن راهی برای محاسبه‌ی $\mathrm{E}[g(X)]$ است. نکته‌ی این قسمت را با ذکر یک مثال شروع می‌کنیم.

مثال ۳ *متغیر تصادفی X با مقادیر و احتمالات زیر مفروض است.*

$$\Pr\{X = -۱\} = ۰/۲ \quad \Pr\{X = ۰\} = ۰/۵ \quad \Pr\{X = ۱\} = ۰/۳$$

هدف ما محاسبه‌ی مقدار $\mathrm{E}[X^۲]$ است. برای این کار ابتدا متغیر تصادفی جدید $Y = X^۲$ را در نظر می‌گیریم. احتمال مقادیری که Y می‌پذیرد، در زیر آمده است:

$$\Pr\{Y = ۱\} = \Pr\{X = -۱\} + \Pr\{X = ۱\} = ۰/۵$$

$$\Pr\{Y = ۰\} = \Pr\{X = ۰\} = ۰/۵$$

بنابراین

$$\mathrm{E}[X^۲] = \mathrm{E}[Y] = ۱ \times ۰/۵ + ۰ \times ۰/۵ = ۰/۵$$

مشخص است که $\mathrm{E}[X^۲] \neq \mathrm{E}[X]^۲$.

اگر چه می‌توان راهبرد به کار برده شده در مثال را برای محاسبه‌ی $\mathrm{E}[g(X)]$ ها به کار برد ولی این راهبرد ممکن است برای برخی $g(X)$ خسته‌کننده باشد. لذا یک روش میان‌بر برای محاسبه آن به صورت قضیه‌ی زیر بیان می‌کنیم.

قضیه ۲

$$\mathrm{E}[g(X)] = \sum_i g(x_i) p(x_i)$$

برهان. *برای اثبات همانند مثال قبل استفاده عمل می‌کنیم. همه‌ی $g(x_i)$ هایی که یک مقدار تولید می‌کنند در متغیری مثل y_j ذخیره می‌کنیم و سپس مراحل زیر را طی می‌کنیم.*

$$\begin{aligned}\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i)\\ &= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i)\\ &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)\\ &= \sum_j y_j \Pr\{g(X)=y_j\}\\ &= \mathrm{E}[g(X)]\end{aligned}$$

■

نکته ۲ *مقدار متوسط خطی است؛ یعنی:*

$$\mathrm{E}[aX+b] = a\mathrm{E}[X]+b$$

برهان. *از قضیه‌ی ۲ نتیجه می‌شود.*

$$\mathrm{E}[aX+b] = \sum_i (ax_i+b)p(x_i) = a\sum_i x_i p(x_i) + b\sum_i p(x_i) = a\mathrm{E}[X]+b$$

■

نکته ۳ *حالت کلی‌تر نکته‌ی قبل نیز برقرار است.*

$$\mathrm{E}[\alpha_۱ g_۱(X)+\alpha_۲ g_۲(X)+\ldots..+\alpha_n g_n(X)] = \alpha_۱\mathrm{E}[g_۱(X)]+\alpha_۲\mathrm{E}[g_۲(X)]+\cdots\cdots+\alpha_n\mathrm{E}[g_n(X)]$$

این ویژگی تحت خطی بودن مقدار متوسط شناخته می‌شود.

نکته ۴ *تعبیری دیگری برای مقدار متوسط این است که مقدار متوسط خطا را صفر می‌کند. یعنی:*

$$\mathrm{E}[X-\mathrm{E}[X]] = ۰$$

برهان.

$$\mathrm{E}[X-\mathrm{E}[X]] = \mathrm{E}[X]-\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]-\mathrm{E}[X] = ۰$$

■

مثال ۴ *برای متغیر تصادفی X با تابع جرم احتمال زیر، مطلوب است محاسبه‌ی $\mathrm{E}[X]$ و $\mathrm{E}[|X-\mathrm{E}[X]|]$.*

$$\Pr\{X=-۲\}=\frac{۱}{۴}\qquad \Pr\{X=۰\}=\frac{۱}{۲}\qquad \Pr\{X=۲\}=\frac{۱}{۴}$$

$$\mathrm{E}[X] = \frac{۱}{۴}\times(-۲)+۰\times\frac{۱}{۲}+\frac{۱}{۴}\times ۲ = ۰$$

$$\mathrm{E}[|X-\mathrm{E}[X]|] = |-۲-۰|\times\frac{۱}{۴}+|۰-۰|\times\frac{۱}{۲}+|۲-۰|\times\frac{۱}{۴} = ۱$$

۳ واریانس

در قسمت قبل دیدیم که $\mathrm{E}[X-\mathrm{E}[X]] = ۰$ است. در واقع میانگین وزندار فاصله مقادیر مختلف متغیر تصادفی X از میانگین وزنی خود صفر است. علت این امر هم واضح است و آن تقارن نسبی بین مقادیری از X است که از $\mathrm{E}[X]$ کمتر یا بیشتر هستند. گاهی ما نیازمند آگاهی از میزان پراکندگی مقادیر مختلف متغیر تصادفی حول میانگین خود هستیم. برای این منظور، مفهوم جدیدی را با عنوان واریانس به صورت زیر تعریف می‌کنیم.

تعریف ۳ *واریانس یک متغیر تصادفی X، که با $\sigma_X^۲$ نیز نشان داده می‌شود، به صورت زیر تعریف می‌شود:*

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^۲]$$

نکته ۵ *با توجه به ویژگی خطی بودن مقدار متوسط، معمولا از رابطه‌ی ساده‌تر زیر برای محاسبه‌ی واریانس استفاده می‌شود:*

$$\mathrm{Var}(X) = \mathrm{E}[X^۲] - \mathrm{E}[X]^۲$$

برهان.

$$\begin{aligned}\mathrm{E}[(X - \mathrm{E}[X])^۲] &= \mathrm{E}[X^۲ - ۲X\mathrm{E}[X] + \mathrm{E}[X]^۲] \\ &= \mathrm{E}[X^۲] - ۲\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}[X]^۲ \\ &= \mathrm{E}[X^۲] - \mathrm{E}[X]^۲\end{aligned}$$

■

فرض کنید مقدار t را به عنوان تخمینی از متغیر تصادفی X در نظر بگیریم. در این صورت می‌توان مقدار $\mathrm{E}[(X-t)^۲]$ را به عنوان معیاری برای سنجش خطا در نظر گرفت. حال می‌خواهیم بدانیم مقدار مینیم آن در کجاست. تعریف کنید:

$$f(t) = \mathrm{E}[(X-t)^۲] = \mathrm{E}[X^۲] - ۲t\mathrm{E}[X] + t^۲$$

داریم:

$$\frac{df}{dt} = -۲\mathrm{E}[X] + ۲t = ۰ \rightarrow t = \mathrm{E}[X]$$

این نشان دهنده‌ی این است مقدار متوسط به عنوان تخمینی از متغیر تصادفی X، مجذور متوسط خطای تخمین را مینیم می‌کند.
در اینجا با مفهوم دیگری آشنا می‌شویم که از تعریف واریانس به دست می‌آید.

تعریف ۴ *انحراف معیار متغیر تصادفی X به صورت زیر تعریف می‌شود:*

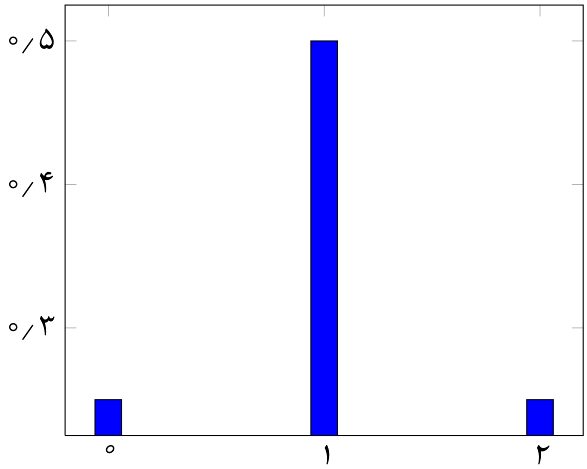
$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

نکته ۶ *فرض کنید متغیر تصادفی X را داشته باشیم و از آن متغیر تصادفی جدیدی بسازیم با نام X^* که رابطه‌ی بین این دو به صورت $X^* = \frac{X - \mathrm{E}[X]}{\sigma_X}$ باشد. آنگاه واریانس و مقدار متوسط آن به صورت زیر است.*

$$\mathrm{E}[X^*] = ۰ \qquad \mathrm{Var}(X^*) = ۱$$

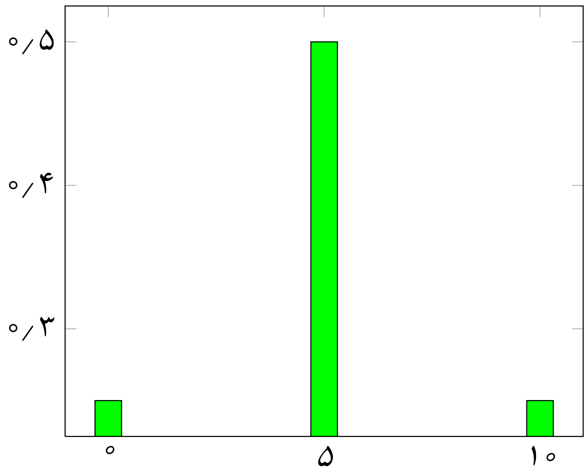
به این کار در اصطلاح نرمال کردن می‌گویند.

مثال ۵ *واریانس توزیع‌های احتمال زیر را به دست آورید.*
توزیع اول به صورت زیر است و $\mathrm{E}[X_۱] = ۱$



$$\mathrm{Var}(X_1) = (0-1)^2 \times (0.25) + (1-1)^2 \times (0.5) + (2-1) \times (0.25) = 0.5$$

توزیع دوم به صورت زیر است و $\mathrm{E}[X_2] = 5$



$$\mathrm{Var}(X_2) = (0-5)^2 \times (0.25) + (5-5)^2 \times (0.5) + (10-5) \times (0.25) = 12.5$$

۴ توزیع‌های احتمال

۱.۴ توزیع برنولی

اگر آزمایش تصادفی داشته باشیم که برآمدهای آن یا شکست است یا موفقیت، با توزیع برنولی مواجه هستیم. متغیر تصادفی آن را به صورت زیر تعریف می‌کنیم.

$$X = \begin{cases} 0 & failure \\ 1 & success \end{cases}$$

فرض می‌کنیم احتمال پیروزی برابر p باشد حال برای تابع جرم احتمال داریم.

$$p(0) = \Pr\{X = 0\} = 1 - p$$

$$p(1) = \Pr\{X = 1\} = p$$

متغیر تصادفی که تابع جرم احتمال آن از تابع جرم احتمال بالا پیروی کند، متغیر تصادفی برنولی با پارامتر p نامیده می‌شود.

نکته ۷ *مقدار متوسط و واریانس توزیع برنولی برابر است با:*

$$\mathrm{E}[X] = p \qquad \sigma_X^2 = p(1-p)$$

نکته ۸ *اگر X یک متغیر تصادفی برنولی با پارامتر p باشد؛ X^2 نیز یک متغیر تصادفی برنولی با پارامتر p است زیرا $X^2 = X$. بنابراین بدون محاسبه می‌توان گفت $\mathrm{E}[X^2] = p$.*

۲.۴ توزیع دو جمله‌ای

فرض کنید n آزمایش برنولی مستقل از هم داریم که احتمال پیروزی در هر کدام از آن‌ها برابر p است و متغیر تصادفی X را تعداد موفقیت‌ها (پیروزی‌ها) در n آزمایش تعریف می‌کنیم. متغیر تصادفی X، متغیر تصادفی با توزیع دوجمله‌ای با پارامترهای (n,p) خوانده می‌شود و تابع جرم احتمال آن به صورت زیر است.

$$p(k) = \Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k},\ \ k = 0, 1, \cdots, n$$

نکته ۹ *مقدار متوسط و واریانس متغیر تصادفی دوجمله‌ای با پارامترهای (n,p) به صورت زیر است:*

$$\mathrm{E}[X] = np \qquad \sigma_X^2 = np(1-p)$$

نکته ۱۰ *توزیع برنولی حالت خاصی از توزیع دوجمله‌ای است که در آن $n = 1$ است.*

نکته ۱۱ *نمودار توزیع دوجمله‌ای به شکل زنگوله است و مقدار ماکسیمم آن برابر است با:*

$$k = [(n+1)p]$$

۳.۴ توزیع پواسون

هر متغیر تصادفی که تابع جرم احتمال آن از فرمول زیر تبعیت کند، متغیر تصادفی پواسون با پارامتر λ نامیده می‌شود.

$$p(k) = \Pr\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

توزیع پواسون اگر چه در ظاهر توزیع پیچیده‌ای به نظر می‌رسد ولی بسیار توزیع ارزشمندی است و می‌توان تحت شرایط خاصی از آن برای تقریب توزیع دوجمله‌ای به کاربرد که در قالب نکته‌ای بیان خواهد شد.

نکته ۱۲

$$\mathrm{E}[X] = \lambda \qquad \mathrm{Var}(X) = \lambda$$

نکته ۱۳ *هنگامی که در یک آزمایش با متغیر تصادفی X با توزیع دوجمله‌ای با پارامترهای (n,p)، n مقداری بزرگ و p مقدار بسیار کمی را اختیار کند به طوری که $\lambda = np$ مقدار متوسط و مطلوبی بشود، در این صورت می‌توان از توزیع پواسون به عنوان تقریبی برای توزیع دوجمله‌ای استفاده کرد.*

برهان. *ابتدا تابع جرم احتمال توزیع دوجمله‌ای را در نظر می‌گیریم. در نظر داشته باشید که $\lambda = np$. داریم:*

$$\begin{aligned} p(k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} (\frac{\lambda}{n})^k (1-\frac{\lambda}{n})^{n-k} \\ &= \frac{n(n-1)\dots(n-k-1)}{n^k} \frac{\lambda^k}{k!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \end{aligned}$$

حال در این مرحله تحت شرایط فوق مقدار $p(k)$ به مقدار زیر میل می‌کند:

$$p(k) = \Pr\{X = k\} \to e^{-\lambda} \frac{\lambda^k}{k!}$$

که برابر با تابع جرم احتمال توزیع پواسون است. از این جهت می‌توان از توزیع پواسون برای تقریب مقادیر احتمال متغیر تصادفی X استفاده کرد. ■

۴.۴ توزیع هندسی

فرض کنید آزمایش‌های تصادفی پیاپی و مستقل از هم با برآمد پیروزی یا شکست، با احتمال پیروزی p انجام می‌شود تا یک پیروزی مشاهده شود. چنین متغیر تصادفی را متغیر تصادفی هندسی با پارامتر p می‌نامیم که دارای تابع جرم احتمال زیر است.

$$p(k) = (1-p)^{k-1} p, \ k = 1, 2, 3, \cdots$$

نکته ۱۴ *مقدار متوسط و واریانس متغیر تصادفی هندسی به صورت زیر است:*

$$\mathrm{E}[X] = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

سؤال ۱ *آیا می‌توانید مقدار متوسط توزیع‌های مطرح شده در این جلسه را به طور حسی توجیه کنید؟*