



به نام خداوند بخشنده مهربان

روش‌های آماری و اقتصادسنجی

در تحلیل و مدل‌سازی داده‌های حمل و نقلی

محمد مهدی بشارتی

besharati@iut.ac.ir

احتمالات

فصل
سوم

بخش اول: مفاهیم احتمال

بخش دوم: متغیر تصادفی



✓ مهندسی حمل و نقل: برنامه ریزی، طراحی، ساخت، اجرا و نگهداری بخش‌های مختلف سیستم‌های حمل و نقلی

✓ این کار، مستلزم درک عمیقی از عرضه و تقاضاست؛

- عرضه سیستم (مثلا طول جاده‌ها، طول خطوط راه آهن، تنظیمات چراغ‌های راهنمایی و غیره) و
- تقاضای سیستم (مثلا تعداد وسایل نقلیه‌ای که می‌خواهند از یک بزرگراه استفاده کنند، تعداد قطارهای عبوری از یک مسیر راه آهن، تعداد کشتی‌ها در یک بندر و غیره)

✓ این امر یک فرآیند بسیار پیچیده و دربردارنده متغیرهایی است که ماهیت آنها تصادفی و احتمالاتی است.



✓ به عبارت دیگر، مهندسین حمل و نقل همچون سایر رشته‌های مهندسی، همواره با **مسائل غیرقطعی** روبرو هستند.

سوال: منظور از مسائل غیرقطعی چیست؟

منظور از مسائل غیرقطعی، پدیده‌هایی است که قبل از وقوع، نتیجه آنها قابل پیش‌بینی نبوده و در هر بار رخ دادن در شرایط یکسان، می‌توانند نتایج مختلفی را به دنبال داشته باشند. به این‌گونه پدیده‌ها، **آزمایش‌های تصادفی** گفته می‌شود.

آزمایش تصادفی عبارتست از عملی که در شرایط ثابت، در هر بار تکرار، منجر به یکی از اعضای فضای نمونه (S) می‌شود.



مثال‌هایی از آزمایش‌های تصادفی؟

- تعداد خودروهایی که در ساعت اوج ترافیک صبح، از یک مقطع بزرگراهی عبور می‌کنند.
- نسبت روزانه تعداد مسافرین مذکر به مسافرین مونث در هر مبدا-مقصد سامانه مترو
- تعداد سفرهای روزانه تولید شده در یک ناحیه از یک شهر
- تعداد تصادفات سالانه در یک تقاطع معین
- و



البته باید توجه داشت که:

مهندسين اغلب ترجیح می دهند این مسائل غیرقطعی و احتمالاتی را به مدل‌هایی ساده تبدیل کرده و مطالعه کنند. بسیاری از فرمول‌های مهندسی که در ظاهر، یک فرمول با فرم بسته به نظر می‌رسند؛ در واقع، شامل این مفاهیم احتمالاتی می‌شوند.

✓ برای مثال، **آیین‌نامه‌های طراحی**، به وسیله قضیه‌های احتمالاتی و نمونه‌های آماری توسعه داده می‌شوند. اما در نهایت برای سهولت در استفاده، این روابط را به صورت "جبری" و نه "احتمالاتی" برای استفاده مهندسين ارائه می‌کنند.

✓ به علاوه، **ضریب اطمینانی** که در بسیاری از روابط مهندسی استفاده می‌گردد، از همین مفهوم سرچشمه می‌گیرد.



قبل از اینکه هر آزمایشی صورت گیرد، می‌توان فهرستی از نتایج موردانتظار برای آن را معین کرد و تمامی نتایج آزمایش را در یک مجموعه نشان داد.

فضای نمونه (S)، مجموعه‌ای است متشکل از تمام نتایج ممکن که با توجه به شرایط آزمایش و برحسب موضوع موردنظر (برای یک مشاهده یا آزمایش تصادفی) قابل تصور است.

توجه شود که هر فضای نمونه، یک مجموعه است. این مجموعه ممکن است قابل شمارش (یا گسسته) و یا غیرقابل شمارش (یا پیوسته) باشد.

مثال:

- (فضای نمونه گسسته): فضای نمونه مدرک تحصیلی افراد متقاضی گواهینامه رانندگی
- (فضای نمونه پیوسته): فضای نمونه سرعت خودروها در یک خیابان، مثلاً

$$S = \{X | 0 \leq x \leq 250\}$$



برآمد (رخداد): به هر عضو از فضای نمونه، یک برآمد گویند.

پیشامد (E): به هر زیرمجموعه قابل اندازه گیری از فضای نمونه، یک پیشامد (E) گفته می شود.

سوال: برای دو فضای نمونه زیر، چند پیشامد را مثال بزنید؟

- (فضای نمونه گسسته): فضای نمونه مدرک تحصیلی افراد متقاضی گواهینامه رانندگی
- (فضای نمونه پیوسته): فضای نمونه سرعت خودروها در یک خیابان
- $S = \{X | 0 \leq x \leq 150\}$

مثلا اگر سرعت مجاز تردد در یک خیابان، ۶۰ کیلومتر بر ساعت باشد

- پیشامد اینکه سرعت خودرویی بیشتر از سرعت غیرمجاز باشد $A = \{X | x > 60\}$



پیشامد (E): به هر زیرمجموعه قابل اندازه گیری از فضای نمونه، یک پیشامد (E) گفته می شود.

احتمال پیشامد A : یک اندازه عددی که امکان وقوع پیشامد A را نشان می دهد.

$$P(A) = \frac{n(A)}{n(S)} = \frac{\text{تعداد اعضای پیشامد } A}{\text{تعداد اعضای فضای نمونه}}$$

احتمال پیشامد A ، عددی بین ۰ و ۱ است



برخی از مهم‌ترین روابط حاکم بر پیشامدها عبارتند از:

۱- برای هر پیشامد مانند A ، رابطه روبرو برقرار است: $0 \leq P(A) \leq 1$

۲- $P(\emptyset) = 0$ ، $P(S) = 1$ ، که در آن، \emptyset پیشامد تهی است.

۳- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



برخی از مهم ترین روابط حاکم بر پیشامدها عبارتند از:

۴- اگر A و B دو پیشامد دوبه دو **ناسازگار** باشند (یعنی: $(A \cap B) = \emptyset$)؛ آنگاه:

$$P(A \cup B) = P(A) + P(B)$$

۵- اگر A' متمم پیشامد A باشد: $P(A') = 1 - P(A)$

۶- در صورتیکه A و B دو پیشامد **مستقل** از هم باشند، آنگاه:

$$P(A \cap B) = P(A) \times P(B)$$

سوال: تفاوت دو ویژگی "ناسازگاری" و "استقلال" چیست؟



سوال: تفاوت دو ویژگی "ناسازگاری" و "استقلال" چیست؟

- دو پیشامد A و B را ناسازگار گوئیم هرگاه هیچ اشتراکی با هم نداشته باشند، به عبارت دیگر دو پیشامد نتوانند همزمان رخ دهند.
- ویژگی استقلال دو پیشامد، یک رابطه عددی است و نه مجموعه‌ای. دو پیشامد A و B مستقل هستند، هرگاه وقوع یکی از آنها، احتمال وقوع دیگری را تغییر ندهد.



احتمال شرطی

- گاهی اوقات، آگاهی از وقوع یک پیشامد، می تواند در محاسبه احتمال وقوع پیشامدهای دیگر موثر باشد.
- احتمال وقوع پیشامد B هنگامی که (بدانیم) پیشامد A اتفاق افتاده باشد را احتمال شرطی می نامند.
- احتمال وقوع B به شرط وقوع A از رابطه زیر بدست می آید.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



قضیه بیز

باتوجه به شکل روبرو؛

فرض کنید $\{A_i, i=1,2,\dots, n\}$ مجموعه‌ای از

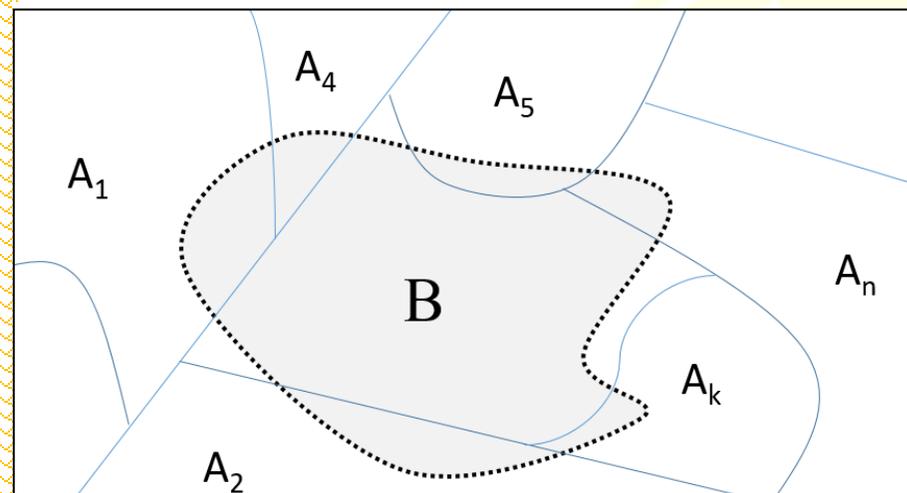
پیشامدهای **دو به دو ناسازگاری** است که فضای نمونه را

به گونه‌ای تقسیم می‌کنند که به ازای هر i ، $P(A_i)$

غیر صفر باشد.

اگر B پیشامدی از فضای نمونه باشد که $P(B) \neq 0$ ،

آنگاه به ازای $k=1,2,\dots, n$ خواهیم داشت؛



$$P(A_k|B) = \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)} = \frac{P(B|A_k) \cdot P(A_k)}{P(B)}$$



قضیه بیز

دو مورد از مهم‌ترین کاربردهای قضیه بیز، عبارتند از:

- (۱) محاسبه احتمالات شرطی
- (۲) ارائه برآوردهای دقیق‌تر برای پارامترهای جامعه با ترکیب منابع مختلف اطلاعاتی
(**رویکرد مدل‌سازی بیزین**).



مثال:

پلیس راهور در قطعه‌ای از یک راه، راننده‌ای را به جرم تخطی از سرعت مجاز متوقف کرده و از او تست الکل می‌گیرد. اگر نتیجه تست مثبت بوده باشد، با توجه به مفروضات زیر می‌خواهیم بدانیم احتمال اینکه او واقعا مست بوده، چقدر است؟ فرض کنید احتمال اینکه دستگاه به اشتباه نتیجه آزمایش را مثبت نشان دهد ۵٪ و احتمال اینکه به اشتباه نتیجه آزمایش را منفی نشان دهد ۱۰٪ می‌باشد. همچنین، فرض کنید از قبل می‌دانیم که به‌طور متوسط ۸٪ از رانندگانی که از این قطعه عبور می‌کنند مست هستند.

پیشامدهای زیر را در نظر می‌گیریم:

A_1 : رانندگی در حال مستی

A_2 : رانندگی در حال سلامت

B : مثبت بودن نتیجه آزمایش

$$P(A_1) = 0/08$$

$$P(A_2) = 1 - 0/08 = 0/92$$

$$P(B|A_1) = 0/90$$

$$P(B|A_2) = 0/05$$



مثال:

بدین ترتیب، براساس قضیه بیز خواهیم داشت؛

$$P(A_1|B) = \frac{p(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)} = \frac{P(B|A_1) \cdot P(A_1)}{P(B)}$$
$$= \frac{0/90 * 0/08}{0/90 * 0/08 + 0/05 * 0/92} \simeq 0/61$$

بنابراین، احتمال اینکه راننده‌ای که نتیجه تست الکل او مثبت بوده، واقعا در حال مستی باشد، 0/61 می‌باشد.

توجه شود که مقدار احتمال شناسایی صحیح راننده مست، به درصد رانندگان مست در میان جامعه موردنظر بستگی دارد. این موضوع یکی از ایده‌های اصلی در استفاده از این قضیه برای برآورد بیزین می‌باشد.

احتمالات

فصل
سوم

بخش اول: مفاهیم احتمال

بخش دوم: متغیر تصادفی

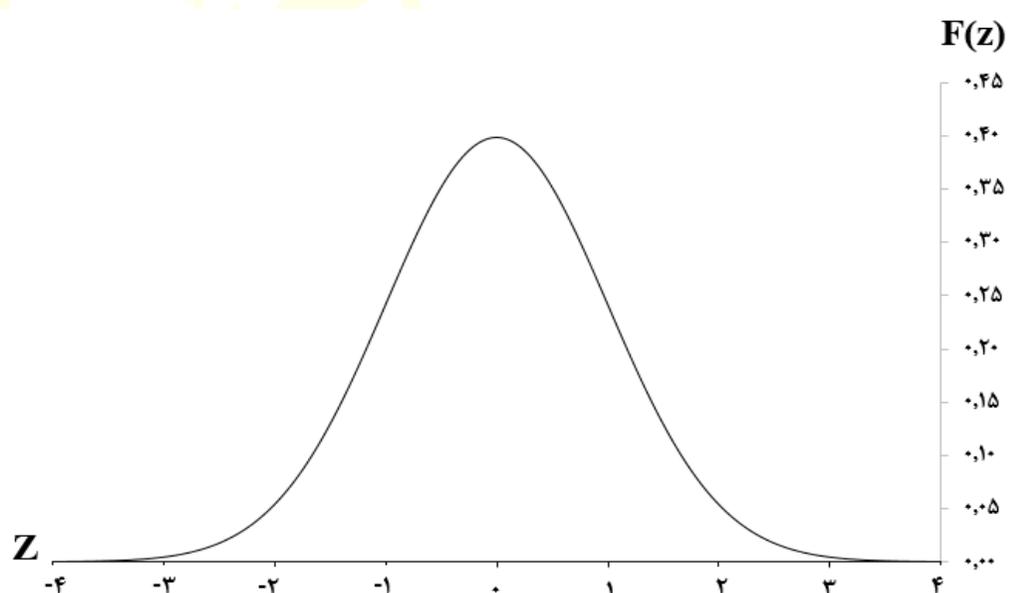
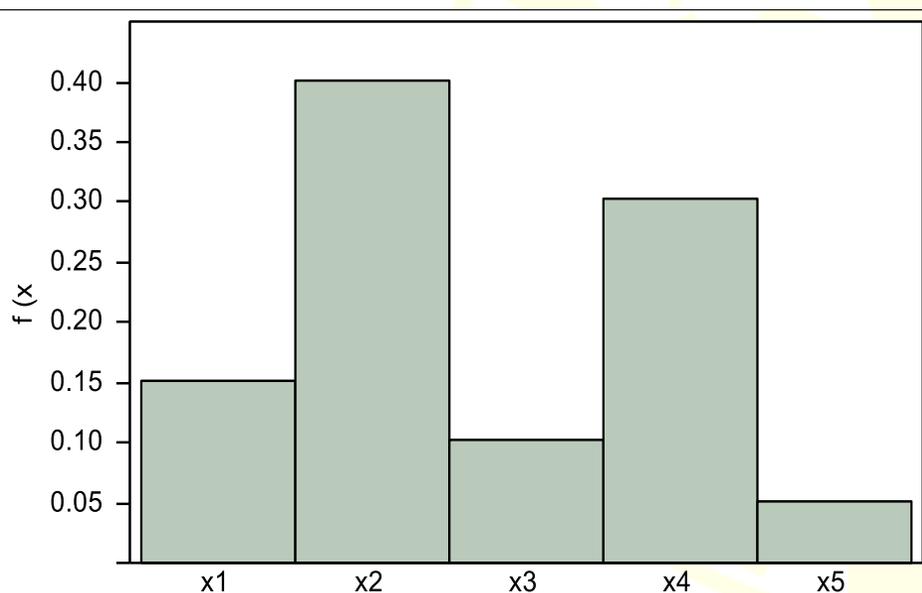


- متغیر تصادفی، یکی از مفاهیم مهم در آمار و احتمالات است.
- متغیر تصادفی، تابعی است که به **اعضای فضای نمونه**، اعدادی را اختصاص می‌دهد.
- مرسوم است که: متغیرهای تصادفی را معمولاً با **حروف بزرگ** لاتین (مانند X) نمایش داده و هر مقدار معین از یک متغیر تصادفی را با **حروف کوچک** (x) نمایش می‌دهند.
- دو نوع متغیر تصادفی در متون مهندسی رایج هستند:
 - متغیرهایی که تنها مقادیر **گسسته** می‌گیرند. به عنوان مثال، فرض کنید یک مهندس تعداد تصادفات سالانه یک تقاطع را در نظر دارد. واضح است که مقادیر مورد نظر، اعداد صحیح نامنفی هستند.
 - متغیرهایی که می‌توانند تمام مقادیر حقیقی را در بازه مشخصی اختیار نمایند و به **متغیرهای تصادفی پیوسته** موسوم هستند. به عنوان مثال، سرعت وسایل نقلیه می‌تواند هر مقدار واقعی (بین صفر و حداکثر سرعت وسایل نقلیه) را اختیار نماید.



✓ یکی از مهم ترین چالش هایی که در برخورد با یک متغیر تصادفی با آن مواجه هستیم، برآورد احتمال وقوع (یا مشاهده) هریک از مقادیر ممکن (یا همان پیشامدهای) آن متغیر تصادفی است.

✓ برای این منظور از "**توزیع های احتمالاتی**" استفاده می شود (موضوع فصل بعد).





❖ مشخصه‌های اصلی توزیع‌های احتمالاتی

۱- امید ریاضی

امید ریاضی یا مقدار مورد انتظار یک متغیر تصادفی برابر است با مجموع حاصلضرب هر مقدار متغیر تصادفی در احتمال متناظر با آن مقدار.

به عبارت دیگر، امید ریاضی X میانگین وزنی متغیر X است که در آن، وزن هریک از مقادیر برابر است با احتمال مشاهده مقدار مورد نظر.

امید ریاضی، شاخصی از میزان تمرکز تابع احتمال می‌باشد.



❖ مشخصه‌های اصلی توزیع‌های احتمالاتی

۱- امید ریاضی

اگر X یک متغیر تصادفی **گسسته** و $f(x)$ مقدار احتمال متناظر با مشاهده x باشد، آنگاه مقدار مورد انتظار X برابر است با

$$E(X) = \sum_x x \cdot f(x)$$

در حالتی که X متغیر تصادفی **پیوسته** باشد؛ به جای عمل جمع، از انتگرال‌گیری استفاده می‌شود

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$



❖ مشخصه‌های اصلی توزیع‌های احتمالاتی

۱- امید ریاضی

اگر X_1, X_2, \dots, X_n و نیز Y متغیرهای تصادفی و a_i مقدار ثابت باشند، آنگاه، مهم‌ترین ویژگی‌های

$$E(a_0 + a_1X + a_2X^2 + \dots)$$

امیدریاضی عبارتند از:

$$E(a_0 + a_1X + a_2X^2 + \dots) = a_0 + a_1E(X) + a_2E(X^2) \quad (1)$$

$$E(X + Y) = E(X) + E(Y) \quad (2)$$

$$E(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + a_nE(X_n) \quad (3)$$

$$E(X.Y) = E(X).E(Y) \quad (4) \text{ اگر } X \text{ و نیز } Y \text{ متغیرهای تصادفی مستقل باشند:}$$

$$E(a_0) = a_0$$



❖ مشخصه‌های اصلی توزیع های احتمالاتی

۲- واریانس و کوواریانس

✓ واریانس و کوواریانس، شاخص هایی از میزان پراکندگی مقادیر متغیر تصادفی

اگر X یک متغیر تصادفی گسسته و $f(x)$ مقدار احتمال متناظر با مشاهده x باشد، آنگاه مقدار واریانس X برابر است با

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

در حالتی که X متغیر تصادفی پیوسته باشد؛ به جای عمل جمع، از انتگرال گیری استفاده می شود.

$$\text{Var}(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot f(x) dx$$



❖ مشخصه‌های اصلی توزیع‌های احتمالاتی

۲- واریانس و کوواریانس

✓ واریانس و کوواریانس، شاخص‌هایی از میزان پراکندگی مقادیر متغیر تصادفی

کوواریانس متغیرهای تصادفی X و Y از رابطه زیر به دست می‌آید؛

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(X \cdot Y) - E(X) \cdot E(Y)$$



❖ مشخصه‌های اصلی توزیع‌های احتمالاتی

۲- واریانس و کوواریانس

برخی از مهم‌ترین ویژگی‌های واریانس و کوواریانس عبارتند از:

$$\text{Var}(a + bX) = b^2 \text{Var}(X) \quad (1)$$

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \quad (2)$$

$$\text{Cov}(aX, bY) = ab \text{Cov}(X, Y) \quad (3)$$

$$\text{Cov}(X, Y) = 0 \quad (4) \text{ اگر } X \text{ و } Y \text{ مستقل باشند،}$$

$$\text{Cov}(X+a, Y+b) = \text{Cov}(X, Y) \quad (5)$$



جمع متغیرهای تصادفی

رابطه‌هایی در مورد میانگین و واریانس ترکیبی خطی از n متغیر تصادفی (برای استفاده از مبحث توزیع نمونه‌گیری و استنباط آماری)

اگر X_1, X_2, \dots, X_n متغیرهای تصادفی مستقل باشند و $Y = \sum_{i=1}^n a_i X_i$

که در آن a_i ها مقادیر ثابت هستند؛ آنگاه امیدریاضی و واریانس Y از روابط زیر بدست می‌آید؛

$$E(Y) = \mu_Y = \sum_{i=1}^n a_i \cdot E(X_i)$$

$$var(Y) = \sigma_Y^2 = \sum_{i=1}^n a_i^2 \cdot var(X_i)$$



مثال

فرض کنید سفر روزانه یک شخص از خانه به محل کار از سه بخش تشکیل شده است.

انحراف استاندارد (دقیقه)	میانگین (دقیقه)	
۳	۱۰	۱. خودرو شخصی
۷	۴۰	۲. مترو
۲	۷	۳. پیاده روی

مشاهده می شود که مدت زمان سفر با هر یک از مدها، متغیر بوده و دارای میانگین و انحراف استاندارد معینی می باشد. مطلوبست محاسبه «میانگین مورد انتظار» و «واریانس» مدت زمان کل سفر؟



مثال

می‌توان گفت که زمان سفر کل (Y) ، یک متغیر تصادفی است؛ که از سه متغیر تصادفی دیگر (X_1, X_2, X_3) تشکیل شده است. بنابراین، طبق روابط ارائه شده، میانگین و واریانس زمان سفر کل بصورت زیر بدست می‌آید؛

$$\mu_Y = \sum_{i=1}^n a_i \cdot E(X_i) = 10 + 40 + 7 = 57$$

$$\text{var}(Y) = \sigma_Y^2 = \sum_{i=1}^n a_i^2 \cdot \text{var}(X_i) = (1^2 \times 3^2) + (1^2 \times 7^2) + (1^2 \times 2^2) = 62$$



در ریاضیات، عبارت فرم بسته (Closed-form expression) عبارتی است که بتوان آن را با استفاده از مجموعه‌ای متناهی از عملیات‌ها محاسبه کرد. عبارت‌های فرم بسته می‌توانند شامل ثابت‌ها، متغیرها، برخی عملیات‌های «آشنا» (مثل، $+$ ، $-$ ، \times ، \div)، و توابع ریاضی (مثل، ریشه عدد، توان، لگاریتم، توابع مثلثاتی، و تابع وارون هذلولوی) باشند، ولی معمولاً برای محاسبه آن‌ها نیازی به استفاده از حد نیست.

جواب‌های همه معادله‌های درجه دو که ضریب‌هایشان عدد مختلط باشد را می‌توان با جمع (ریاضی)، تفریق، ضرب (ریاضی)، تقسیم، و ریشه دوم محاسبه کرد. همه این موارد توابع مقدماتی‌اند. به عبارت دیگر، معادله درجه دو

$$ax^2 + bx + c = 0.$$

می‌توان با عبارت فرم بسته (Closed-form solution) زیر محاسبه کرد:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$