18.440: Lecture 25

Covariance and some conditional expectation exercises

Scott Sheffield

MIT

Outline

Covariance and correlation

Paradoxes: getting ready to think about conditional expectation

Outline

Covariance and correlation

Paradoxes: getting ready to think about conditional expectation

A property of independence

- If X and Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

A property of independence

- If X and Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- Just write $E[g(X)h(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy$.

A property of independence

- If X and Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- Just write $E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy$.
- Since $f(x,y) = f_X(x)f_Y(y)$ this factors as $\int_{-\infty}^{\infty} h(y)f_Y(y)dy \int_{-\infty}^{\infty} g(x)f_X(x)dx = E[h(Y)]E[g(X)]$.

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])$.

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])$.
- Note: by definition $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])$.
- Note: by definition $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.
- Covariance (like variance) can also written a different way. Write $\mu_x = E[X]$ and $\mu_Y = E[Y]$. If laws of X and Y are known, then μ_X and μ_Y are just constants.

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])$.
- Note: by definition $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.
- Covariance (like variance) can also written a different way. Write $\mu_x = E[X]$ and $\mu_Y = E[Y]$. If laws of X and Y are known, then μ_X and μ_Y are just constants.
- Then

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y] =$$

$$E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y = E[XY] - E[X]E[Y].$$

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])$.
- Note: by definition $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.
- Covariance (like variance) can also written a different way. Write $\mu_x = E[X]$ and $\mu_Y = E[Y]$. If laws of X and Y are known, then μ_X and μ_Y are just constants.
- Then

$$\mathrm{Cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y] =$$

$$E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y = E[XY] - E[X]E[Y].$$

- Covariance formula $E[XY] - E[X]E[Y]$, or "expectation of product minus product of expectations" is frequently useful.

Defining covariance and correlation

- Now define covariance of X and Y by $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])$.
- Note: by definition $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.
- Covariance (like variance) can also written a different way. Write $\mu_x = E[X]$ and $\mu_Y = E[Y]$. If laws of X and Y are known, then μ_X and μ_Y are just constants.
- Then

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] =$$

$$E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y = E[XY] - E[X]E[Y].$$

- Covariance formula $E[XY] - E[X]E[Y]$, or "expectation of product minus product of expectations" is frequently useful.
- Note: if X and Y are independent then $\mathrm{Cov}(X, Y) = 0$.

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.

Basic covariance facts

- Using $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.
- **General statement of bilinearity of covariance:**

$$\mathrm{Cov}(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \mathrm{Cov}(X_i, Y_j).$$

Basic covariance facts

- Using $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ as a definition, certain facts are immediate.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.
- **General statement of bilinearity of covariance:**

$$\mathrm{Cov}(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j).$$

- Special case:

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{(i,j):i<j} \mathrm{Cov}(X_i, X_j).$$

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by

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

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by

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

- Correlation doesn't care what units you use for X and Y. If $a > 0$ and $c > 0$ then $\rho(aX + b, cY + d) = \rho(X, Y)$.

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by

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

- Correlation doesn't care what units you use for X and Y. If $a > 0$ and $c > 0$ then $\rho(aX + b, cY + d) = \rho(X, Y)$.
- Satisfies $-1 \le \rho(X, Y) \le 1$.

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by

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

- Correlation doesn't care what units you use for X and Y. If $a > 0$ and $c > 0$ then $\rho(aX + b, cY + d) = \rho(X, Y)$.
- Satisfies $-1 \leq \rho(X, Y) \leq 1$.
- Why is that? Something to do with $E[(X + Y)^2] \geq 0$ and $E[(X - Y)^2] \geq 0$?

Defining correlation

- Again, by definition $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by
$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$
- Correlation doesn't care what units you use for X and Y. If $a > 0$ and $c > 0$ then $\rho(aX + b, cY + d) = \rho(X, Y)$.
- Satisfies $-1 \leq \rho(X, Y) \leq 1$.
- Why is that? Something to do with $E[(X + Y)^2] \geq 0$ and $E[(X - Y)^2] \geq 0$?
- If a and b are positive constants and $a > 0$ then $\rho(aX + b, X) = 1$.

Defining correlation

- Again, by definition $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$.
- **Correlation** of X and Y defined by

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

- Correlation doesn't care what units you use for X and Y. If $a > 0$ and $c > 0$ then $\rho(aX + b, cY + d) = \rho(X,Y)$.
- Satisfies $-1 \leq \rho(X,Y) \leq 1$.
- Why is that? Something to do with $E[(X+Y)^2] \geq 0$ and $E[(X-Y)^2] \geq 0$?
- If a and b are positive constants and $a > 0$ then $\rho(aX + b, X) = 1$.
- If a and b are positive constants and $a < 0$ then $\rho(aX + b, X) = -1$.

Important point

- Say X and Y are uncorrelated when $\rho(X, Y) = 0$.

Important point

- Say X and Y are uncorrelated when $\rho(X, Y) = 0$.
- Are independent random variables X and Y always uncorrelated?

Important point

- Say X and Y are uncorrelated when $\rho(X, Y) = 0$.
- Are independent random variables X and Y always uncorrelated?
- Yes, assuming variances are finite (so that correlation is defined).

Important point

- Say X and Y are uncorrelated when $\rho(X, Y) = 0$.
- Are independent random variables X and Y always uncorrelated?
- Yes, assuming variances are finite (so that correlation is defined).
- Are uncorrelated random variables always independent?

Important point

- Say X and Y are uncorrelated when $\rho(X, Y) = 0$.
- Are independent random variables X and Y always uncorrelated?
- Yes, assuming variances are finite (so that correlation is defined).
- Are uncorrelated random variables always independent?
- No. Uncorrelated just means $E[(X - E[X])(Y - E[Y])] = 0$, i.e., the outcomes where $(X - E[X])(Y - E[Y])$ is positive (the upper right and lower left quadrants, if axes are drawn centered at $(E[X], E[Y])$) balance out the outcomes where this quantity is negative (upper left and lower right quadrants). This is a much weaker statement than independence.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Can we generalize this example?

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Can we generalize this example?
- What is variance of number of people who get their own hat in the hat problem?

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Can we generalize this example?
- What is variance of number of people who get their own hat in the hat problem?
- Define X_i to be 1 if ith person gets own hat, zero otherwise.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Can we generalize this example?
- What is variance of number of people who get their own hat in the hat problem?
- Define X_i to be 1 if ith person gets own hat, zero otherwise.
- Recall formula $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{(i,j):i<j} \mathrm{Cov}(X_i, X_j)$.

Examples

- Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with variance 1. For example, maybe each X_j takes values ± 1 according to a fair coin toss.
- Compute $\mathrm{Cov}(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Compute the correlation coefficient $\rho(X_1 + X_2 + X_3, X_2 + X_3 + X_4)$.
- Can we generalize this example?
- What is variance of number of people who get their own hat in the hat problem?
- Define X_i to be 1 if ith person gets own hat, zero otherwise.
- Recall formula $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{(i,j):i<j} \mathrm{Cov}(X_i, X_j)$.
- Reduces problem to computing $\mathrm{Cov}(X_i, X_j)$ (for $i \neq j$) and $\mathrm{Var}(X_i)$.

Outline

Covariance and correlation

Paradoxes: getting ready to think about conditional expectation

Outline

Covariance and correlation

Paradoxes: getting ready to think about conditional expectation

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.
- At the end of this period, a (biased) coin will be tossed. Banker will be assigned to hell forever with probability $1/n$ and heaven forever with probability $1 - 1/n$.

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.
- At the end of this period, a (biased) coin will be tossed. Banker will be assigned to hell forever with probability $1/n$ and heaven forever with probability $1 - 1/n$.
- After 10 days, banker reasons, "If I wait another day I reduce my odds of being here forever from $1/10$ to $1/11$. That's a reduction of $1/110$. A $1/110$ chance at infinity has infinite value. Worth waiting one more day."

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.
- At the end of this period, a (biased) coin will be tossed. Banker will be assigned to hell forever with probability $1/n$ and heaven forever with probability $1 - 1/n$.
- After 10 days, banker reasons, "If I wait another day I reduce my odds of being here forever from $1/10$ to $1/11$. That's a reduction of $1/110$. A $1/110$ chance at infinity has infinite value. Worth waiting one more day."
- Repeats this reasoning every day, stays in hell forever.

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.
- At the end of this period, a (biased) coin will be tossed. Banker will be assigned to hell forever with probability $1/n$ and heaven forever with probability $1 - 1/n$.
- After 10 days, banker reasons, "If I wait another day I reduce my odds of being here forever from $1/10$ to $1/11$. That's a reduction of $1/110$. A $1/110$ chance at infinity has infinite value. Worth waiting one more day."
- Repeats this reasoning every day, stays in hell forever.
- Standard punch line: this is actually what banker deserved.

Famous paradox

- Certain corrupt and amoral banker dies, instructed to spend some number n (of banker's choosing) days in hell.
- At the end of this period, a (biased) coin will be tossed. Banker will be assigned to hell forever with probability $1/n$ and heaven forever with probability $1 - 1/n$.
- After 10 days, banker reasons, "If I wait another day I reduce my odds of being here forever from $1/10$ to $1/11$. That's a reduction of $1/110$. A $1/110$ chance at infinity has infinite value. Worth waiting one more day."
- Repeats this reasoning every day, stays in hell forever.
- Standard punch line: this is actually what banker deserved.
- Fairly dark as math humor goes (and no offense intended to anyone...) but dilemma is interesting.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.
- **Even simpler variant:** infinitely many identical money sacks have labels $1, 2, 3, \ldots$ I have sack 1. You have all others.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.
- **Even simpler variant:** infinitely many identical money sacks have labels $1, 2, 3, \ldots$ I have sack 1. You have all others.
- You offer me a deal. I give you sack 1, you give me sacks 2 and 3. I give you sack 2 and you give me sacks 4 and 5. On the nth stage, I give you sack n and you give me sacks $2n$ and $2n+1$. Continue until I say stop.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.
- **Even simpler variant:** infinitely many identical money sacks have labels $1, 2, 3, \ldots$ I have sack 1. You have all others.
- You offer me a deal. I give you sack 1, you give me sacks 2 and 3. I give you sack 2 and you give me sacks 4 and 5. On the nth stage, I give you sack n and you give me sacks $2n$ and $2n+1$. Continue until I say stop.
- Lets me get arbitrarily rich. But if I go on forever, I return every sack given to me. If nth sack confers right to spend nth day in heaven, leads to hell-forever paradox.

- **Paradox:** decisions seem sound individually but together yield worst possible outcome. Why? Can we demystify this?
- **Variant without probability:** Instead of tossing $(1/n)$-coin, person deterministically spends $1/n$ fraction of future days (every nth day, say) in hell.
- **Even simpler variant:** infinitely many identical money sacks have labels $1, 2, 3, \ldots$ I have sack 1. You have all others.
- You offer me a deal. I give you sack 1, you give me sacks 2 and 3. I give you sack 2 and you give me sacks 4 and 5. On the nth stage, I give you sack n and you give me sacks $2n$ and $2n+1$. Continue until I say stop.
- Lets me get arbitrarily rich. But if I go on forever, I return every sack given to me. If nth sack confers right to spend nth day in heaven, leads to hell-forever paradox.
- I make infinitely many good trades and end up with less than I started with. "Paradox" is really just existence of 2-to-1 map from (smaller set) $\{2, 3, \ldots\}$ to (bigger set) $\{1, 2, \ldots\}$.

Money pile paradox

- You have an infinite collection of money piles with labeled $0, 1, 2, \ldots$ from left to right.

Money pile paradox

- You have an infinite collection of money piles with labeled $0, 1, 2, \ldots$ from left to right.
- Precise details not important, but let's say you have $1/4$ in the 0th pile and $\frac{3}{8}5^j$ in the jth pile for each $j > 0$. Important thing is that pile size is increasing exponentially in j.

Money pile paradox

- You have an infinite collection of money piles with labeled $0, 1, 2, \ldots$ from left to right.
- Precise details not important, but let's say you have $1/4$ in the 0th pile and $\frac{3}{8}5^j$ in the jth pile for each $j > 0$. Important thing is that pile size is increasing exponentially in j.
- Banker proposes to transfer a fraction (say $2/3$) of each pile to the pile on its left and remainder to the pile on its right. Do this simultaneously for all piles.

Money pile paradox

- You have an infinite collection of money piles with labeled $0, 1, 2, \ldots$ from left to right.
- Precise details not important, but let's say you have $1/4$ in the 0th pile and $\frac{3}{8}5^j$ in the jth pile for each $j > 0$. Important thing is that pile size is increasing exponentially in j.
- Banker proposes to transfer a fraction (say $2/3$) of each pile to the pile on its left and remainder to the pile on its right. Do this simultaneously for all piles.
- Every pile is bigger after transfer (and this can be true even if banker takes a portion of each pile as a fee).

Money pile paradox

- You have an infinite collection of money piles with labeled $0, 1, 2, \ldots$ from left to right.
- Precise details not important, but let's say you have $1/4$ in the 0th pile and $\frac{3}{8}5^j$ in the jth pile for each $j > 0$. Important thing is that pile size is increasing exponentially in j.
- Banker proposes to transfer a fraction (say $2/3$) of each pile to the pile on its left and remainder to the pile on its right. Do this simultaneously for all piles.
- Every pile is bigger after transfer (and this can be true even if banker takes a portion of each pile as a fee).
- Banker seemed to make you richer (every pile got bigger) but really just reshuffled your infinite wealth.

Two envelope paradox

- X is geometric with parameter $1/2$. One envelope has 10^X dollars, one has 10^{X-1} dollars. Envelopes shuffled.

Two envelope paradox

- X is geometric with parameter $1/2$. One envelope has 10^X dollars, one has 10^{X-1} dollars. Envelopes shuffled.
- You choose an envelope and, after seeing contents, are allowed to choose whether to keep it or switch. (Maybe you have to pay a dollar to switch.)

Two envelope paradox

- X is geometric with parameter $1/2$. One envelope has 10^X dollars, one has 10^{X-1} dollars. Envelopes shuffled.
- You choose an envelope and, after seeing contents, are allowed to choose whether to keep it or switch. (Maybe you have to pay a dollar to switch.)
- Maximizing conditional expectation, it seems it's always better to switch. But if you always switch, why not just choose second-choice envelope first and avoid switching fee?

Two envelope paradox

- X is geometric with parameter $1/2$. One envelope has 10^X dollars, one has 10^{X-1} dollars. Envelopes shuffled.
- You choose an envelope and, after seeing contents, are allowed to choose whether to keep it or switch. (Maybe you have to pay a dollar to switch.)
- Maximizing conditional expectation, it seems it's always better to switch. But if you always switch, why not just choose second-choice envelope first and avoid switching fee?
- Kind of a disguised version of money pile paradox. But more subtle. One has to replace "jth pile of money" with "restriction of expectation sum to scenario that first chosen envelop has 10^j". Switching indeed makes each pile bigger.

Two envelope paradox

- X is geometric with parameter $1/2$. One envelope has 10^X dollars, one has 10^{X-1} dollars. Envelopes shuffled.
- You choose an envelope and, after seeing contents, are allowed to choose whether to keep it or switch. (Maybe you have to pay a dollar to switch.)
- Maximizing conditional expectation, it seems it's always better to switch. But if you always switch, why not just choose second-choice envelope first and avoid switching fee?
- Kind of a disguised version of money pile paradox. But more subtle. One has to replace "jth pile of money" with "restriction of expectation sum to scenario that first chosen envelop has 10^j". Switching indeed makes each pile bigger.
- However, "Higher expectation given amount in first envelope" may not be right notion of "better." If S is payout with switching, T is payout without switching, then S has same law as $T - 1$. In that sense S is worse.

Moral

- Beware infinite expectations.

Moral

- Beware infinite expectations.
- Beware unbounded utility functions.

Moral

- Beware infinite expectations.
- Beware unbounded utility functions.
- They can lead to strange conclusions, sometimes related to “reshuffling infinite (actual or expected) wealth to create more” paradoxes.

Moral

- Beware infinite expectations.
- Beware unbounded utility functions.
- They can lead to strange conclusions, sometimes related to "reshuffling infinite (actual or expected) wealth to create more" paradoxes.
- Paradoxes can arise even when total transaction is finite with probability one (as in envelope problem).