

Fra lotto til optimale strategier

10. oktober 2025

Eulers Venner, Aarhus Universitet 2025

Eulers Venner, 8 oktober 2025

Monty Hall Problemet

I det amerikanske TV-show
"Let's Make a Deal" kunne
deltagerne vinde en bil!



- Deltageren skulle åbne én af tre døre med numrene 1,2 eller 3.
- Bag den ene dør var der en bil.
- Bag de to andre døre var der en ged.

Eulers Venner, 8 oktober 2025

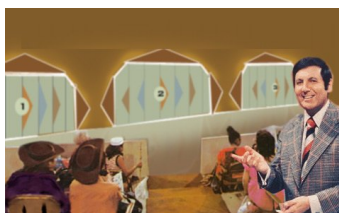


- Deltageren skulle først vælge en af dørene 1,2 eller 3.
- Derefter oplyste tv-værten nummeret på en af de ikke-udvalgte døre, som gemte på en ged.
- Deltageren fik derpå muligheden for at fastholde sit valg eller at vælge om.
- Hvilken mulighed giver størst sandsynlighed for at vinde bilen?

Eulers Venner, 8 oktober 2025

Der er 3 muligheder for placeringen af bilen:

Dør-nummer: $\begin{matrix} 1 & 2 & 3 \\ B & G & G \\ G & B & G \\ G & G & B \end{matrix}$



- Lad os f.eks. sige, at deltageren i første omgang vælger dør 1.
- Hvis hun fastholder dør 1, vil hun kun vinde i første tilfælde.
- Hvis hun vælger om, vil hun vinde i andet og tredje tilfælde.
- Sandsynligheden for at vinde er altså hhv. $\frac{1}{3}$ og $\frac{2}{3}$.

Eulers Venner, 8 oktober 2025

Ssh for 7 rigtige to gange i Lotto, hvis man spiller i 28 år

- Manden har spillet 10 rækker i Lotto hver uge siden 1989.
- Han fik 7 rigtige første gang i 2001 og anden gang i 2017.
- Der er 8.347.680 rækker i Lotto, hvoraf kun én har 7 rigtige. Sandsynligheden for at få 7 rigtige ved at spille på 10 rækker en enkelt uge er således
$$p_1 = \frac{10}{8.347.680} = 1.197938 \cdot 10^{-06}.$$
- Hvis man spiller 10 rækker i Lotto, hver uge i 28 år (dvs. i 1456 uger), bliver antallet af gange man får 7 rigtige binomialfordelt med parametre $(1456, p_1)$.
- Hvis $V \sim Binomial(1456, p_1)$ gælder
$$P(V \geq 2) = 1 - P(V = 0) - P(V = 1)$$
$$= 1 - \sum_{j=0}^1 \binom{1456}{j} p_1^j (1 - p_1)^{1456-j} = 1.518303 \cdot 10^{-06}.$$

Eulers Venner, 8 oktober 2025

Ssh for 7 rigtige to gange på landsplan

- Antag nu, at 1 million danskere spiller 10 rækker hver uge i 28 år.
- Hver af dem har så ssh $p_2 = 1.518303 \cdot 10^{-06}$ for at få 7 rigtige mindst 2 gange.
- Antallet A af spillere, der får 7 rigtige mindst 2 gange bliver således binomialfordelt med parametre $(1000000, p_2)$.
- $P(A \geq 1) = 1 - (1 - p_2)^{1000000} = 0.7809169.$
- $P(A \geq 2) = 1 - P(A = 0) - P(A = 1) = 0.4482819.$
- Faktisk skete det for to danskere over en periode på 34 år! Her bliver sandsynligheden $P(A \geq 2) = 0.6547017.$

Eulers Venner, 8 oktober 2025

Definition af konveks funktion

Lad I være et interval i $\mathbb{R}$, og betragt en funktion $\varphi: I \rightarrow \mathbb{R}$.

Vi siger da, at φ er konveks, hvis der for alle a, b i I og alle λ i $(0, 1)$ gælder uligheden:

$$\varphi(\lambda a + (1 - \lambda)b) \leq \lambda \varphi(a) + (1 - \lambda)\varphi(b).$$

Jensens Ulighed (klassiske version)

Lad $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ være en konveks funktion.

Lad videre n være et naturligt tal, lad $\lambda_1, \dots, \lambda_n$ være (strengt) positive tal, og lad $x_1, \dots, x_n$ være vilkårlige reelle tal.

Da gælder uligheden:

$$\varphi\left(\frac{\sum_{i=1}^n \lambda_i x_i}{\sum_{i=1}^n \lambda_i}\right) \leq \frac{\sum_{i=1}^n \lambda_i \varphi(x_i)}{\sum_{i=1}^n \lambda_i}.$$



Eulers Venner, 8 oktober 2025

Ækivalent formulering af klassiske Jensens Ulighed

Sætning (Jensens Ulighed)

Lad $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ være en konveks funktion.

Lad X være en diskret stokastisk variabel med $R_X = \{x_1, x_2, \dots, x_n\}$.

Så gælder der, at

$$\varphi(E[X]) \leq E[\varphi(X)]$$

Eksempel: $\varphi(x) = x^2$.

Bevis for Jensens Ulighed: Ved induktion efter n

$n = 1$: $R_X = \{x_1\}$, og $P_X(x_1) = 1$. Så gælder

$$\varphi(E[X]) = \varphi(x_1 P_X(x_1)) = \varphi(x_1) = \varphi(x_1) P_X(x_1) \overset{\text{LOTUS}}{\stackrel{!}{=}} E[\varphi(X)].$$

$n \geq 2$: Antag, at Jensens Ulighed gælder for $n - 1$. Vis da, at den også gælder for n .

Betragt altså en diskret stokastisk variabel X med $R_X = \{x_1, x_2, \dots, x_n\}$.

Definér en ny stokastisk variabel Y ved

$$R_Y = \{x_1, \dots, x_{n-1}\}, \quad \text{og} \quad P_Y(x_j) = \frac{P_X(x_j)}{1 - P_X(x_n)}, \quad j = 1, \dots, n - 1.$$

Bemærk:

$$\sum_{j=1}^{n-1} P_Y(x_j) = \frac{1}{1 - P_X(x_n)} \sum_{j=1}^{n-1} P_X(x_j) = \frac{1}{1 - P_X(x_n)} (1 - P_X(x_n)) = 1.$$

Husk: $R_Y = \{x_1, \dots, x_{n-1}\}$, og $P_Y(x_j) = \frac{P_X(x_j)}{1 - P_X(x_n)}$.

Vi finder nu, at

$$\begin{aligned} \varphi(E[X]) &= \varphi\left(\sum_{j=1}^n x_j P_X(x_j)\right) = \varphi\left((1 - P_X(x_n)) \sum_{j=1}^{n-1} x_j P_Y(x_j) + x_n P_X(x_n)\right) \\ &= \varphi\left((1 - P_X(x_n)) E[Y] + P_X(x_n) x_n\right) \\ &\overset{\varphi \text{ er konveks}}{\stackrel{!}{\leq}} (1 - P_X(x_n)) \varphi(E[Y]) + P_X(x_n) \varphi(x_n) \\ &\overset{\text{Jensen for } n-1}{\stackrel{!}{\leq}} (1 - P_X(x_n)) E[\varphi(Y)] + P_X(x_n) \varphi(x_n) \\ &\overset{\text{LOTUS}}{\stackrel{!}{=}} (1 - P_X(x_n)) \sum_{j=1}^{n-1} \varphi(x_j) P_Y(x_j) + P_X(x_n) \varphi(x_n) \\ &= \sum_{j=1}^{n-1} \varphi(x_j) P_X(x_j) + \varphi(x_n) P_X(x_n) \overset{\text{LOTUS}}{\stackrel{!}{=}} E[\varphi(X)]. \end{aligned}$$

Jensens Ulighed (klassiske version)

Lad $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ være en konveks funktion.

Lad videre n være et naturligt tal, lad $\lambda_1, \dots, \lambda_n$ være (strengt) positive tal, og lad $x_1, \dots, x_n$ være vilkårlige reelle tal.

Da gælder uligheden:

$$\varphi\left(\frac{\sum_{i=1}^n \lambda_i x_i}{\sum_{i=1}^n \lambda_i}\right) \leq \frac{\sum_{i=1}^n \lambda_i \varphi(x_i)}{\sum_{i=1}^n \lambda_i}.$$



Sekretær Problemet

- Betragt n kugler, som er nummererede med forskellige (ukendte) hele tal.
- Antag, at kuglerne er skjult i et rør i en tilfældig rækkefølge.
- Spillet går ud på at trække kugler én efter én og på et tidspunkt beslutte sig for, at den kugle, *man netop har trukket*, er den med det højeste nummer.
- Man vinder, hvis man har ret; ellers taber man.

Strategien S_k

Lad k være et fast tal i $\{1, 2, \dots, n\}$, og betragt følgende strategi:

- 1 *Udtræk de første k kugler uden skelen til deres numre.*
- 2 *Udtræk derefter kugler, indtil man første gang får en kugle med højere nummer end samtlige de k først udtrukne.*

Tilfældet $k = \frac{n}{2}$.

- Del røret op i 2 dele I og II, som hver indeholder $\frac{1}{2}$ af kuglerne.
- Betragt placeringerne af de to kugler med de to højeste numre i forhold til de 2 dele af røret.
- Der er $2 \cdot 2 = 4$ mulige placeringer af de to kugler i forhold til de 2 dele af røret (som alle har sandsynlighed **omtrent** $1/4$).
- Blandt disse 4 muligheder er der 1 mulighed, hvor kuglen med det næsthøjeste nummer er placeret i del I, mens kuglen med det højeste nummer er placeret i del II.
- I dette tilfælde vil $S_{n/2}$ føre til, at man vinder spillet!
- Der gælder således at

$$P(S_{n/2} \text{ vinder}) \geq \frac{1}{4}.$$

Optimal strategi (for store n): $k = \frac{n}{e}$

For store værdier af n er strategien S_k optimal i tilfældet

“ $k = \frac{n}{e}$ ”, hvor $e \approx 2.718281828$.

I dette tilfælde gælder der (for store n), at

$$P(S_{n/e} \text{ vinder}) \approx \frac{1}{e} \approx 0.367879441.$$

Argument for den optimale strategi

Lad M betegne positionen af kuglen med det højeste nummer:
 $R_M = \{1, 2, \dots, n\}$, og $P(M = m) = \frac{1}{n}$ for alle m i R_M .

Vi ønsker først at bestemme $P(S_k \text{ vinder} \mid M = m)$ for m i R_M .

Det er klart, at

$$P(S_k \text{ vinder} \mid M = m) = 0, \quad \text{hvis } m \leq k.$$

Argument for den optimale strategi (fortsat)

Lad os nu se på tilfældet hvor $m \geq k + 1$:

- Betragt de første $m - 1$ kugler i røret.
- Om S_k vinder spillet afhænger alene af, om kuglen med det højeste nummer, blandt de $m - 1$ første, er placeret blandt de k første kugler.
- Det følger derfor, at

$$P(S_k \text{ vinder} \mid M = m) = \frac{k}{m - 1}.$$

Argument for den optimale strategi (fortsat)

Der gælder nu, at

$$\begin{aligned} P(S_k \text{ vinder}) &= \sum_{m=1}^n P(S_k \text{ vinder} \mid M = m) \cdot P(M = m) \\ &= \sum_{m=k+1}^n \frac{k}{m - 1} \cdot \frac{1}{n} = \frac{k}{n} \sum_{m=k+1}^n \frac{1}{m - 1} \stackrel{\ell=m-1}{=} \frac{k}{n} \sum_{\ell=k}^{n-1} \frac{1}{\ell} \\ &= \frac{k}{n} \sum_{\ell=k}^{n-1} \frac{1}{\ell} \frac{1}{n} \stackrel{f(x)=1/x}{=} \frac{k}{n} \sum_{\ell=k}^{n-1} f(\ell/n) \frac{1}{n} \approx \frac{k}{n} \int_{k/n}^1 \frac{1}{x} dx \\ &= \frac{k}{n} (\log(1) - \log(k/n)) = \frac{k}{n} \log(\frac{n}{k}). \end{aligned}$$

Her gælder der videre, at $x \log(1/x)$ er størst mulig, når $x = \frac{1}{e}$.

Dermed er $P(S_k \text{ vinder})$ størst, når $\frac{k}{n} \approx \frac{1}{e}$, dvs. når $k \approx \frac{n}{e}$.