MVE420:
Beslutsteori och rationalitet under osäkerhet

Vilhelm Verendel

24 April 2015

Bakgrund

Vi vill resonera noga om framtida risker och hur vi bör hantera dem. De kunskapsteoretiska problem hopar sig: en blandning av prediktion, spekulation och osäkerhet och vi har ingen data...

Bakgrund

Vi vill resonera noga om framtida risker och hur vi bör hantera dem.
De kunskapsteoretiska problem hopar sig: en blandning av prediktion, spekulation och osäkerhet och vi har ingen data...

Problemen innefattar bland annat

- Situationer som ännu inte finns (orealiserade teknologier) och experiment inte möjliga eller inte önskvärda (destruktiva teknologier)
- Vi har ingen tidsserie (vilket inte utesluter nya sorters händelser)
- Händelserna kan vara så osannolika att vi inte kan förvänta oss att ha sett dem, men ha så stora konsekvenser att vi inte kan ignorera dem (existentiell risk)

Hur kan vi tänka? Hur bör vi tänka? Vilka modeller och ansatser finns?

Tre relevanta frågor för beslut och framtiden

Vi verkar inte alltid välkalibrerade för att bedöma framtida risker: tillgänglighetsheuristik, förankringseffekt, kvantitetsblindhet, ..., vilket lyfter frågor som:

1. Kan framtiden hanteras rationellt?
2. Vad är rationellt att tro?
3. Hur bör alternativ jämföras?

Ser ut att vara olika frågor, men är mycket relaterade.

Tre relevanta frågor för beslut och framtiden

Vi verkar inte alltid välkalibrerade för att bedöma framtida risker: tillgänglighetsheuristik, förankringseffekt, kvantitetsblindhet, ..., vilket lyfter frågor som:

1. Kan framtiden hanteras rationellt? (Teorier om rationalitet.)
2. Vad är rationellt att tro? (Bayesiansk analys ger ett delsvar.)
3. Hur bör alternativ jämföras? (Maximera förväntad nytta?)

Ser ut att vara olika frågor, men är mycket relaterade.

Vad är rationalitet?

Kort svar:
Det finns flera teorier, men ingen har sett perfekt rationalitet i praktiken :-)

Två olika ansatser:

1. Rationalitet är något vi definierar (t ex ett teoribygge)
2. Rationalitet är något vi upptäcker (vi förstår vad det är när vi ser det)

Beslutsteori

- Rationalitet som definition: implicerar typiskt maximering av *förväntad nytta* och i vissa fall Bayesiansk uppdatering när en rationell agent gör observationer
- Under vilka förutsättningar? Varför förväntad nytta?

Motiverande exempel

1. Du vill ha en frukt. Välj mellan apelsin, banan, päron.
2. Du ska singla en slant. Vad är sannolikheten för en krona?
3. Du parkerar bilen på gatan. Vad är sannolikheten att den blir stulen inatt?
4. Du ska genomgå en medicinsk operation. Vad är sannolikheten att du överlever?
5. Vilken är sannolikheten för kärnvapenkrig de kommande 50 åren?

Vad skiljer de här situationerna åt?

Motiverande exempel

1. Du vill ha en frukt. Välj mellan apelsin, banan, päron.
2. Du ska singla en slant. Vad är sannolikheten för en krona?
3. Du parkerar bilen på gatan. Vad är sannolikheten att den blir stulen inatt?
4. Du ska genomgå en medicinsk operation. Vad är sannolikheten att du överlever?
5. Vilken är sannolikheten för kärnvapenkrig de kommande 50 åren?

Vad skiljer de här situationerna åt?
Säkerhet, till risk, till osäkerhet.

Rationalitet i tre olika fall

1. Säkra utfall (optimering, sökproblem i stora mängder)
2. Risk: kända (objektiva) sannolikheter
3. Osäkerhet: okända (subjektiva) sannolikheter

Rationalitet i tre olika fall

1. Säkra utfall (optimering, sökproblem i stora mängder)
   T ex: välj mellan apelsin, banan, päron
2. Risk: kända (objektiva) sannolikheter
   T ex: spela roulette
3. Osäkerhet: okända (subjektiva) sannolikheter
   T ex: sannolikheten för kärnvapenkrig de kommande 50 åren

Framtiden är mer eller mindre osäker. Vi koncentrerar oss på fall 2 och 3.

Tidiga tankar om förväntad nytta: Pascals vad

Pascal (1600-talet): Borde jag tro på Gud?

	Gud existerar	Gud existerar inte
Tro	∞	$-c$
Tro inte	$-\infty$	0

Tidiga tankar om förväntad nytta: Pascals vad

Pascal (1600-talet): Borde jag tro på Gud?

	Gud existerar	Gud existerar inte
Tro	∞	$-c$
Tro inte	$-\infty$	0

Låt p vara sannolikheten att Gud existerar

Tidiga tankar om förväntad nytta: Pascals vad

Pascal (1600-talet): Borde jag tro på Gud?

	Gud existerar	Gud existerar inte
Tro	∞	$-c$
Tro inte	$-\infty$	0

Låt p vara sannolikheten att Gud existerar
Pascal: även för mycket små $p > 0$, borde vi tro på Gud

Tidiga tankar om förväntad nytta: Pascals vad

Pascal (1600-talet): Borde jag tro på Gud?

	Gud existerar	Gud existerar inte
Tro	∞	$-c$
Tro inte	$-\infty$	0

Låt p vara sannolikheten att Gud existerar
Pascal: även för mycket små $p > 0$, borde vi tro på Gud
Nämnde inte "nytta" explicit, men argumentet berör förväntad nytta

St. Petersburg-paradoxen

Bernoulli (1738):

Ett mynt skall singlas och antalet gånger det blir krona räknas till första klaven. Låt X vara antalet kronor innan första klaven. Vinsten som erbjuds är 2^Xkr.

St. Petersburg-paradoxen

Bernoulli (1738):

Ett mynt skall singlas och antalet gånger det blir krona räknas till första klaven. Låt X vara antalet kronor innan första klaven. Vinsten som erbjuds är 2^Xkr.

Fråga: Hur mycket skulle du betala för att delta i detta lotteri?

St. Petersburg-paradoxen

Bernoulli (1738):

Ett mynt skall singlas och antalet gånger det blir krona räknas till första klaven. Låt X vara antalet kronor innan första klaven. Vinsten som erbjuds är 2^Xkr.

Fråga: Hur mycket skulle du betala för att delta i detta lotteri?

Möjliga utfall: $1, 2, 4, ..., \infty$

St. Petersburg-paradoxen

Bernoulli (1738):

Ett mynt skall singlas och antalet gånger det blir krona räknas till första klaven. Låt X vara antalet kronor innan första klaven. Vinsten som erbjuds är 2^Xkr.

Fråga: Hur mycket skulle du betala för att delta i detta lotteri?

Möjliga utfall: $1, 2, 4, ..., \infty$

Väntevärdet för vinst:
$E(2^X) = \sum_i^\infty 2^i Pr(i \text{ kronor}) = \infty$

St. Petersburg-paradoxen

Bernoulli (1738):

Ett mynt skall singlas och antalet gånger det blir krona räknas till första klaven. Låt X vara antalet kronor innan första klaven. Vinsten som erbjuds är 2^Xkr.

Fråga: Hur mycket skulle du betala för att delta i detta lotteri?

Möjliga utfall: $1, 2, 4, ..., \infty$

Väntevärdet för vinst:
$E(2^X) = \sum_i^\infty 2^i Pr(i \text{ kronor}) = \infty$
Bernoullis observation: människor betalar inte hur mycket som helst

Bernoullis förklaring: introducera *nytta* (eng. *utility*)

Bernoullis idé: människor vill inte direkt ha pengar utan *nytta* **u(x)**
u icke-linjär i x:
$E(2^X) = \sum_{i=0}^{\infty} 2^i Pr(i \text{ kronor}) = \infty$
$EU(2^X) = \sum_{i=0}^{\infty} u(2^i) Pr(i \text{ kronor}) < \infty$

Bernoullis förklaring: introducera *nytta* (eng. *utility*)

Bernoullis idé: människor vill inte direkt ha pengar utan *nytta* **u(x)**
u icke-linjär i x:
$E(2^X) = \sum_{i=0}^{\infty} 2^i Pr(i \text{ kronor}) = \infty$
$EU(2^X) = \sum_{i=0}^{\infty} u(2^i) Pr(i \text{ kronor}) < \infty$

Med hjälp av en nyttofunktion på t ex formen $u(x) = c\ log(x)$ blir väntevärdet inte oändligt.

Möjligtvis en förklaring, men: varför skulle personer betrakta väntevärdet? Varför inte någon annan kombination av väntevärde, varians, ..., någon mer komplicerad funktion av stokastiska variabeln 2^X?

Moderna teorier om rationalitet motiverar maximering av väntevärde på ett noggrannt sätt.

Moderna teorier om rationalitet

Översiktlig skiss:
Grundläggande byggblock för teorin: *preferenser*

Moderna teorier om rationalitet

Översiktlig skiss:
Grundläggande byggblock för teorin: *preferenser*
En beslutsfattare ställd inför att välja från en beslutsmängd A

Moderna teorier om rationalitet

Översiktlig skiss:
Grundläggande byggblock för teorin: *preferenser*
En beslutsfattare ställd inför att välja från en beslutsmängd A
En *preferens* $\succcurlyeq$ rangordnar parvisa möjligheter från mängden A

Moderna teorier om rationalitet

Översiktlig skiss:
Grundläggande byggblock för teorin: *preferenser*
En beslutsfattare ställd inför att välja från en beslutsmängd A
En *preferens* $\succcurlyeq$ rangordnar parvisa möjligheter från mängden A
A i allmänhet stor och komplicerad, men lekfullt exempel:
$A = \{\text{apelsin, banan, päron}\}$
En preferens: $\text{apelsin} \succcurlyeq \text{banan}, \text{banan} \succcurlyeq \text{päron}, \text{apelsin} \succcurlyeq \text{päron}$

Moderna teorier om rationalitet

Översiktlig skiss:
Grundläggande byggblock för teorin: *preferenser*
En beslutsfattare ställd inför att välja från en beslutsmängd A
En *preferens* $\succcurlyeq$ rangordnar parvisa möjligheter från mängden A
A i allmänhet stor och komplicerad, men lekfullt exempel:
$A = \{\text{apelsin, banan, päron}\}$
En preferens: $\text{apelsin} \succcurlyeq \text{banan}, \text{banan} \succcurlyeq \text{päron}, \text{apelsin} \succcurlyeq \text{päron}$

Teorier om rationalitet: föreslår vilken sorts preferenser som är rationella.
Praktisk beslutsregel: *representationsteorem* för att kunna göra kvantitativ analys inom ekonomi, beslutsteori och spelteori.
Ger grundlig teoretisk motivering för när det är rimligt att använda *förväntad nytta*.

Preferenser: definition

Beslutsmängd A
Definition: en *preferensrelation* $\succcurlyeq \subseteq A \times A$

Preferenser: definition

Beslutsmängd A
Definition: en *preferensrelation* $\succcurlyeq \subseteq A \times A$
Tolkning: "a föredras svagt över b" som $(a, b) \in \succcurlyeq$ På kortform: $a \succcurlyeq b$

Preferenser: definition

Beslutsmängd A
Definition: en *preferensrelation* $\succcurlyeq \subseteq A \times A$
Tolkning: "a föredras svagt över b" som $(a, b) \in \succcurlyeq$ På kortform: $a \succcurlyeq b$

Exempel:
$A = \{\text{apelsin, banan, päron}\}$
En preferens: $\succcurlyeq = \{(\text{apelsin, banan}), (\text{banan, päron}), (\text{apelsin, päron})\}$

Preferenser: definition

Beslutsmängd A
Definition: en *preferensrelation* $\succcurlyeq \subseteq A \times A$
Tolkning: "a föredras svagt över b" som $(a, b) \in \succcurlyeq$ På kortform: $a \succcurlyeq b$

Exempel:
$A = \{\text{apelsin, banan, päron}\}$
En preferens: $\succcurlyeq = \{(\text{apelsin, banan}), (\text{banan, päron}), (\text{apelsin, päron})\}$
På kortform: $\text{apelsin} \succcurlyeq \text{banan}, \text{banan} \succcurlyeq \text{päron}, \text{apelsin} \succcurlyeq \text{päron}$

Preferenser: definition

Beslutsmängd A
Definition: en *preferensrelation* $\succcurlyeq \subseteq A \times A$
Tolkning: "a föredras svagt över b" som $(a, b) \in \succcurlyeq$ På kortform: $a \succcurlyeq b$

Exempel:
$A = \{\text{apelsin, banan, päron}\}$
En preferens: $\succcurlyeq = \{(\text{apelsin, banan}), (\text{banan, päron}), (\text{apelsin, päron})\}$
På kortform: $\text{apelsin} \succcurlyeq \text{banan}, \text{banan} \succcurlyeq \text{päron}, \text{apelsin} \succcurlyeq \text{päron}$

Rationalitet handlar om vilka par tagna ur $A \times A$ som kan och måste finnas i $\succcurlyeq$

Representationteorem, generell förklaring

Notera att $\succcurlyeq$ är en ickekvantitativ parvis rangordning.
När kan vi representera $\succcurlyeq$ som maximering av en realvärd **nytto**funktion?

Representationteorem, generell förklaring

Notera att $\succcurlyeq$ är en ickekvantitativ parvis rangordning.
När kan vi representera $\succcurlyeq$ som maximering av en realvärd **nytto**funktion?
Representationsteorem visar under vilka förutsättningar *kvantitativ representation* av $\succcurlyeq$ existerar som maximering av förväntad nytta

Representationteorem, generell förklaring

Notera att $\succcurlyeq$ är en ickekvantitativ parvis rangordning.
När kan vi representera $\succcurlyeq$ som maximering av en realvärd **nytto**funktion?
Representationsteorem visar under vilka förutsättningar *kvantitativ representation* av $\succcurlyeq$ existerar som maximering av förväntad nytta

Med **rationella** preferenser $\succcurlyeq$ ($\succcurlyeq$ uppfyller **axiom** $R_1..R_n$)

1. Existerar en nyttofunktion **u** med egenskapen att
2. Förväntad nytta *EU* rangordnar preferenserna i $\succcurlyeq$ kvantitativt

$a \succcurlyeq b$ om och endast om $EU(a) \geq EU(b)$ för alla $a, b \in A$

Representationteorem, generell förklaring

Notera att $\succcurlyeq$ är en ickekvantitativ parvis rangordning.
När kan vi representera $\succcurlyeq$ som maximering av en realvärd **nytto**funktion?
Representationsteorem visar under vilka förutsättningar *kvantitativ representation* av $\succcurlyeq$ existerar som maximering av förväntad nytta

Med **rationella** preferenser $\succcurlyeq$ ($\succcurlyeq$ uppfyller **axiom** $R_1..R_n$)

1. Existerar en nyttofunktion **u** med egenskapen att
2. Förväntad nytta EU rangordnar preferenserna i $\succcurlyeq$ kvantitativt

$a \succcurlyeq b$ om och endast om $EU(a) \geq EU(b)$ för alla $a, b \in A$

Förväntad nytta EU är typiskt $EU(a) = \sum_{x \in C(a)} u(x)p(x)$

där $C(a)$: mängden av potentiella konsekvenser vid val av a

Två huvudsakliga teorier

Vilka är då $R_1..R_n$ som axiom? Skiljer sig åt mellan teorier:

1. Neumann-Morgenstern (1944): $\succcurlyeq$ på "lotterier" (eng. *lotteries*)
   - Risk, kända sannolikheter
2. Savage (1954): $\succcurlyeq$ på "handlingar" (eng. *acts*)
   - Osäkerhet, okända sannolikheter

Neumann-Morgenstern (1944) om "lotterier"

Låt X vara en mängd alternativ, och låt $\succcurlyeq$ vara en preferensrelation på ΔX. Då existerar en (NM)-nyttofunktion $u : X \to R$ som ger en förväntad nytta-representation av $\succcurlyeq$ om och endast om $\succcurlyeq$ uppfyller (NM1) (svagt ordnad), (NM2) (kontinuitet), och (NM3) (oberoende).

Beslutsmängd $A = \{L_1, L_2, ..., L_n\}$ beskriver "lotterier" för beslutsfattaren

Neumann-Morgenstern (1944) om "lotterier"

Låt X vara en mängd alternativ, och låt $\succcurlyeq$ vara en preferensrelation på ΔX. Då existerar en (NM)-nyttofunktion $u : X \to R$ som ger en förväntad nytta-representation av $\succcurlyeq$ om och endast om $\succcurlyeq$ uppfyller (NM1) (svagt ordnad), (NM2) (kontinuitet), och (NM3) (oberoende).

Beslutsmängd $A = \{L_1, L_2, ..., L_n\}$ beskriver "lotterier" för beslutsfattaren
Exempel: köpa 100kr-lott eller inte? Potentiell vinst $10000kr$.
Alternativen $X = \{-100kr, 0kr, 10000kr\}$ beskriver olika konsekvenser

Neumann-Morgenstern (1944) om "lotterier"

Låt X vara en mängd alternativ, och låt $\succcurlyeq$ vara en preferensrelation på ΔX. Då existerar en (NM)-nyttofunktion $u : X \to R$ som ger en förväntad nytta-representation av $\succcurlyeq$ om och endast om $\succcurlyeq$ uppfyller (NM1) (svagt ordnad), (NM2) (kontinuitet), och (NM3) (oberoende).

Beslutsmängd $A = \{L_1, L_2, ..., L_n\}$ beskriver "lotterier" för beslutsfattaren
Exempel: köpa 100kr-lott eller inte? Potentiell vinst $10000kr$.
Alternativen $X = \{-100kr, 0kr, 10000kr\}$ beskriver olika konsekvenser
Inte spela: $L_1 = (0kr)$; Spela: $L_2 = (0.99 : -100kr, 0.01 : +10000kr)$

Neumann-Morgenstern (1944) om "lotterier"

Låt X vara en mängd alternativ, och låt $\succcurlyeq$ vara en preferensrelation på ΔX. Då existerar en (NM)-nyttofunktion $u : X \to R$ som ger en förväntad nytta-representation av $\succcurlyeq$ om och endast om $\succcurlyeq$ uppfyller (NM1) (svagt ordnad), (NM2) (kontinuitet), och (NM3) (oberoende).

Beslutsmängd $A = \{L_1, L_2, ..., L_n\}$ beskriver "lotterier" för beslutsfattaren
Exempel: köpa 100kr-lott eller inte? Potentiell vinst $10000kr$.
Alternativen $X = \{-100kr, 0kr, 10000kr\}$ beskriver olika konsekvenser
Inte spela: $L_1 = (0kr)$; Spela: $L_2 = (0.99 : -100kr, 0.01 : +10000kr)$

Mängden av alla lotterier $\Delta X = \{(p_1 : -100kr, p_2 : 0kr, p_3 : 10000kr)\}$
Beslutsmängd: faktiska lotterier $A \subset \Delta X$. Observerbara val $\succcurlyeq \in A \times A$

Neumann-Morgenstern (1944) om "lotterier"

Låt X vara en mängd alternativ, och låt $\succcurlyeq$ vara en preferensrelation på ΔX. Då existerar en (NM)-nyttofunktion $u : X \to R$ som ger en förväntad nytta-representation av $\succcurlyeq$ om och endast om $\succcurlyeq$ uppfyller (NM1) (svagt ordnad), (NM2) (kontinuitet), och (NM3) (oberoende).

Beslutsmängd $A = \{L_1, L_2, ..., L_n\}$ beskriver "lotterier" för beslutsfattaren
Exempel: köpa 100kr-lott eller inte? Potentiell vinst $10000kr$.
Alternativen $X = \{-100kr, 0kr, 10000kr\}$ beskriver olika konsekvenser
Inte spela: $L_1 = (0kr)$; Spela: $L_2 = (0.99 : -100kr, 0.01 : +10000kr)$

Mängden av alla lotterier $\Delta X = \{(p_1 : -100kr, p_2 : 0kr, p_3 : 10000kr)\}$
Beslutsmängd: faktiska lotterier $A \subset \Delta X$. Observerbara val $\succcurlyeq \in A \times A$

Notera att: NM1-3 är axiom över **alla lotterier** ΔX, sätter därför struktur på beslutsfattarens observerbara val mellan faktiska lotterier i A

Neumann-Morgenstern (1944): 3 axiom

Teorin antar kända sannolikheter p för lotterierna.
Här är de tre axiomen för $\succcurlyeq$:

NM1: svag ordning
NM2: kontinuitet
NM3: oberoende

NM1: svag ordning

För alla $P, Q, R \in \Delta X$:

1. **Komplett**: $P \succcurlyeq Q$ eller $Q \succcurlyeq P$ för alla P, Q
   Problem: Kan vara ganska krävande (exempel längre fram)

2. **Transitiv**: om $P \succcurlyeq Q$ och $Q \succcurlyeq R$ så $P \succcurlyeq R$

NM1: svag ordning

För alla $P, Q, R \in \Delta X$:

1. **Komplett**: $P \succcurlyeq Q$ eller $Q \succcurlyeq P$ för alla P, Q
   Problem: Kan vara ganska krävande (exempel längre fram)

2. **Transitiv**: om $P \succcurlyeq Q$ och $Q \succcurlyeq R$ så $P \succcurlyeq R$

(NM1) utesluter cykliska preferenser, t ex
$\succcurlyeq = \{$(apelsin, banan), (banan, päron), (päron, apelsin)$\}$

En aktör med cykliska preferenser på en marknad skulle kunna få monetära problem... (eng. *money pump*)

NM2: kontinuitet

Definition av strikt preferens $\succ$ för P över Q: $P \succcurlyeq Q$ och inte $Q \succcurlyeq P$

NM2: kontinuitet

Definition av strikt preferens $\succ$ för P över Q: $P \succcurlyeq Q$ och inte $Q \succcurlyeq P$

Kontinuitet: Om $P \succ Q \succ R$, så existerar alltid $\alpha, \beta \in (0, 1)$ så att,

$$\alpha P + (1 - \alpha)R \succ Q \succ \beta P + (1 - \beta)R$$

NM2: kontinuitet

Definition av strikt preferens $\succ$ för P över Q: $P \succcurlyeq Q$ och inte $Q \succcurlyeq P$

Kontinuitet: Om $P \succ Q \succ R$, så existerar alltid $\alpha, \beta \in (0, 1)$ så att,

$$\alpha P + (1 - \alpha)R \succ Q \succ \beta P + (1 - \beta)R$$

Tankeexempel:
P=+10kr, Q=0 kr, R=döden
Antag preferenserna $P \succ Q \succ R$
Om kontinuitet håller, för höga $\alpha < 1$: $(\alpha : P, (1 - \alpha) : R) \succ Q$
Första axiomet (NM1) är att kunna göra jämförelsen.
Kontinuitet (NM2): att den sortens prioriteringar existerar.

NM2: tankeexperiment om kontinuitet

Antag att du är i färd att köpa en tidning för 10 kr.

Men, du noterar att den ges bort gratis på andra sidan av den trafikerade gatan. Skulle du korsa gatan och få den gratis?

NM2: tankeexperiment om kontinuitet

Antag att du är i färd att köpa en tidning för 10 kr.

Men, du noterar att den ges bort gratis på andra sidan av den trafikerade gatan. Skulle du korsa gatan och få den gratis?

Svar ja verkar innebära att acceptera en risk, om än väldigt liten, av att förlora sitt liv för en tidning.

$\alpha < 1$: $(\alpha : P, (1 - \alpha) : R) \succ Q$

NM3: Oberoende

Oberoende av "irrelevanta alternativ" för varje $P, Q, R \in \Delta X$:

NM3: Oberoende

Oberoende av "irrelevanta alternativ" för varje $P, Q, R \in \Delta X$:

För $\alpha \in (0, 1)$,

$$P \succcurlyeq Q$$
om och endast om
$$(\alpha : P, (1-\alpha) : R) \succcurlyeq (\alpha : Q, (1-\alpha) : R)$$

NM3: Oberoende

Oberoende av "irrelevanta alternativ" för varje $P, Q, R \in \Delta X$:

För $\alpha \in (0, 1)$,

$$P \succcurlyeq Q$$
om och endast om
$$(\alpha : P, (1 - \alpha) : R) \succcurlyeq (\alpha : Q, (1 - \alpha) : R)$$

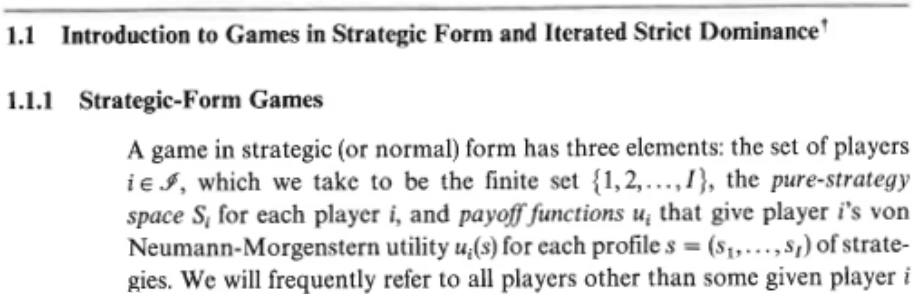
Exempel: lotterier rangordnade efter utfall där de skiljer
$(100, 1) \succcurlyeq (90, 1) \rightarrow (100, 0.1, -200, 0.9) \succcurlyeq (90, 0.1, -200, 0.9)$
Problem: bl a effekten av förankring i mer komplicerade situationer.

Reflektion över Neumann-Morgensterns teori

1. *Nytta* i modern beslutsteori beskriver preferenser kvantitativt
2. Om dessa preferenser är "bra" eller "dåligt" för beslutsfattaren i mer allmän mening är osagt, termen nytta är inlånad
3. Teorin beskriver en rationell beslutsfattare enbart om vi anser att axiomen NM1-3 beskriver en rationell beslutsfattare
4. Om vi anser att (NM1)-(NM3) är rationella, så kan vi beskriva rationellt beslutsfattande med maximering av förväntat nyttovärde
5. Konceptuellt: från parvis rangordning till kvantitativ förväntad nytta
6. Ett påstående om att vi *bör* vara rationella i denna mening innehåller även det en sorts värdering: om att vi *bör* följa teorins antaganden.

Antas ofta i ekonomisk analys och spelteori

T ex i grundbok spelteori av Fudenberg and Tirole:



1.1 Introduction to Games in Strategic Form and Iterated Strict Dominance[†]

1.1.1 Strategic-Form Games

A game in strategic (or normal) form has three elements: the set of players $i \in \mathcal{I}$, which we take to be the finite set $\{1, 2, \ldots, I\}$, the *pure-strategy space* S_i for each player i, and *payoff functions* u_i that give player i's von Neumann-Morgenstern utility $u_i(s)$ for each profile $s = (s_1, \ldots, s_I)$ of strategies. We will frequently refer to all players other than some given player i

Vanligtvis betyder maximering av förväntad nytta implicit rationellt beslutsfattande.

Savage (1954): Bayesiansk Rationalitet

Neumann-Morgensterns teori antog objektiva (kända) sannolikheter och beskriver bara rationalitet över specialfallet sådana situationer. Tänk om vi inte alls har några kända sannolikheter?

Savage (1954): Bayesiansk Rationalitet

Neumann-Morgensterns teori antog objektiva (kända) sannolikheter och beskriver bara rationalitet över specialfallet sådana situationer. Tänk om vi inte alls har några kända sannolikheter?

Om vi startar enbart med preferenser över en beslutsmängd, *utan* nyttofunktion, *utan* sannolikheter. Inga tal som beskriver beslutsfattarens situation. Vad kan göras?

Savage (1954): Bayesiansk Rationalitet

Neumann-Morgensterns teori antog objektiva (kända) sannolikheter och beskriver bara rationalitet över specialfallet sådana situationer. Tänk om vi inte alls har några kända sannolikheter?

Om vi startar enbart med preferenser över en beslutsmängd, *utan* nyttofunktion, *utan* sannolikheter. Inga tal som beskriver beslutsfattarens situation. Vad kan göras?

Savages teori: 7 axiom $P1 - P7$

1. Bevis under $P1 - P7$ att en *nyttofunktion* $\mathbf{u}$ och en *sannolikhet* $\mathbf{p}$ existerar, representerar preferenser med förväntad nytta.
2. Att vara konsistent rationell över tid: hur spelare radionellt bör lära sig ("uppdatera tro") av observationer genom att uppdatera $\mathbf{p}$ Bayesianskt (motiverar stor verktygslåda av Bayesiansk statistik)

Savages teorigenombrott: Kontrast mot tidigare syn

Keynes:

"By uncertain knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty ... The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence ... About these matters *there is no scientific basis on which to form any calculable probability whatever*. We simply do not know." - Keynes, 1920-talet

Savages situation

Startar helt kvalitativt, med abstrakta mängder:

- Världen är i ett av många olika möjliga *tillstånd* S (eng. states)
- Det finns en beslutsmängd A av *handlingar* (eng. acts)
- Det finns en mängd *konsekvenser* C (eng. consequences)
- Som i Neumann-Morgensterns teori, finns en preferensrelation $\succcurlyeq$ över parvisa val från A

En handling: en funktion från tillstånd till konsekvenser
Varje handling $f \in A$ är på formen, $f : S \to C$
Beslutsfattaren väljer mellan handlingar

Ett exempel (Savages eget)

You are making an omelette, and have cracked 5 good eggs. What about the 6th egg? It can be good, or it can be rotten. How should you go about it?

Ett exempel (Savages eget)

You are making an omelette, and have cracked 5 good eggs. What about the 6th egg? It can be good, or it can be rotten. How should you go about it?

Act	State	
	Good	Rotten
Break into bowl	Six-egg omelet	No omelet and five good eggs destroyed
Break into saucer	Six-egg omelet and a saucer to wash	Five-egg omelet and a saucer to wash
Throw away	Five-egg omelet and one good egg destroyed	Five-egg omelet

Här har $|S| = 2$ tillstånd, $|A| = 3$ handlingar, $|C| = 6$ konsekvenser

Savage P1-P7: sketch

Preferensrelationen $\succcurlyeq$ är definierad över par av handlingar från beslutsmängden A.
Liknande Neumann-Morgenstern, låt mängden F vara mängden av *alla* möjliga handlingar (funktioner) från S to C
Savage: när $\succcurlyeq$ över A är på en viss form relativt F finns en väntevärdesmaximering som representerar $\succcurlyeq$.

Teori: När $\succcurlyeq$ följer $(P1) - (P7)$ med avseende på F, för varje $f, g \in F$:

Savage P1-P7: sketch

Preferensrelationen $\succcurlyeq$ är definierad över par av handlingar från beslutsmängden A.
Liknande Neumann-Morgenstern, låt mängden F vara mängden av *alla* möjliga handlingar (funktioner) från S to C
Savage: när $\succcurlyeq$ över A är på en viss form relativt F finns en väntevärdesmaximering som representerar $\succcurlyeq$.

Teori: När $\succcurlyeq$ följer $(P1) - (P7)$ med avseende på F, för varje $f, g \in F$:

P1-P7: vissa likheter med Neumann/Morgensterns axiom, och några fler.
Använder Neumann/Morgensterns resultat som en del i ett mer allmänt resultat.

Savages bevis

Om $\succcurlyeq$ följer $P1 - P7$ så:

- Existerar det en sannolikhet p över S och en nyttofunktion $u : C \to \mathbb{R}$

För varje $f, g \in A$,

$$f \succcurlyeq g$$

om och endast om

$$\sum_{s \in S} u(f(s)p(s) \geq \sum_{s \in S} u(g(s))p(s)$$

- Existens: p unik (följer sannolikhetsaxiom och är en sannolikhet)
- Unikhet: u unik upp till positiv linjärtransform

Reflektion över Savages teori

En Savage-Bayes-rationell beslutsfattare har preferenser sådana att

- En sannolikhet p och en nyttofunktion u representerar preferenser med maximering av förväntad nytta
- Objektiva sannolikheter behövs inte för detta!
- Under vissa väldigt allmänna förutsättningar finns skäl för att
  1. En beslutsfattare börjar med en prior och maximerar väntevärde på nytta
  2. Vid ny information göra Bayesiansk uppdatering av en sannolikhet, men vilka skäl för detta? Andra möjligheter?

Effekten av observation om världen

- Savages ramverk (P1)-(P7) berör beslutsfattande i ett statiskt ramverk. Inte en beslutsfattare som får ny information. Sannolikheten p fås över de olika tillstånden S.
- Att få information om att världen befinner sig i tillstånd $S' \subseteq S$ skulle ju kunna få en beslutsfattare att ändra sina preferenser $\succcurlyeq$ till $\succcurlyeq_{S'}$
- Ett antagande om tidskonsistenta preferenser ger Bayesiansk uppdatering: men kräver att beslutsfattaren aldrig blir förvånad (se t ex, Häggström 2015)
- Rationalitet över tid och under osäkerhet blir mer öppet om vi inte kan anta detta

Bayesianism

Efter observation D, välj f om och endast om

$$\sum_{s \in S} u(f(s))p(s|D) \geq \sum_{s \in S} u(g(s))p(s|D)$$

för alla g.

- Detta teoretiska tankeexempel har blivit ett ideal för Bayesiansk uppdatering.
- Nackdel med krav på tidskonsistens: starkt antagande om en Bayesiansk agent som aldrig kan bli förvånad

Sammanfattat hittills, rationalitet

- I ekonomisk analys och beslutsteori motsvaras rationalitet av en beslutsregel som maximerar förväntad nytta
- Bernoullis idé har gradvis förfinats till omfattande matematiska och filosofiska teorier om rationalitet och representationsteorem
- Distinktionen *risk* kontra *osäkerhet*
- Teorin om rationalitet under osäkerhet talar för Bayesianska agenter

En tillämpning av Bayesiansk analys

En Bayesiansk analys av existentiell risk

Antag att

Vi upptäcker liv på Mars, eller någon annan planet, på en teknologisk nivå liknande mänskligheten...

Några möjliga reaktioner

1. Detta är bra!
   - Vi ser liv överallt runtom oss på jorden, generellt är liv något bra
   - Skönt att inte vara ensamma i universum

Några möjliga reaktioner

1. Detta är bra!
   - Vi ser liv överallt runtom oss på jorden, generellt är liv något bra
   - Skönt att inte vara ensamma i universum
2. Vissa menar: Tvärt om, det vore mycket oroande!
   - Vi måste tänka noga kring vad det här betyder för vår egen situation
   - Bostrom: detta vore oroväckande för hela mänsklighetens framtid

Några möjliga reaktioner

1. Detta är bra!
   - Vi ser liv överallt runtom oss på jorden, generellt är liv något bra
   - Skönt att inte vara ensamma i universum
2. Vissa menar: Tvärt om, det vore mycket oroande!
   - Vi måste tänka noga kring vad det här betyder för vår egen situation
   - Bostrom: detta vore oroväckande för hela mänsklighetens framtid
3. Vetenskaplig analys:
   - Låt oss göra en enkel statistisk modell av situationen
   - Se på vilken betydelse observationen (liv på Mars) får

Hur skulle det kunna vara dåligt? Nick Bostrom, 2008:

What could be more fascinating than discovering life that had evolved entirely independently of life here on Earth? Many people would also find it heartening to learn that we are not entirely alone in this vast cold cosmos.
...

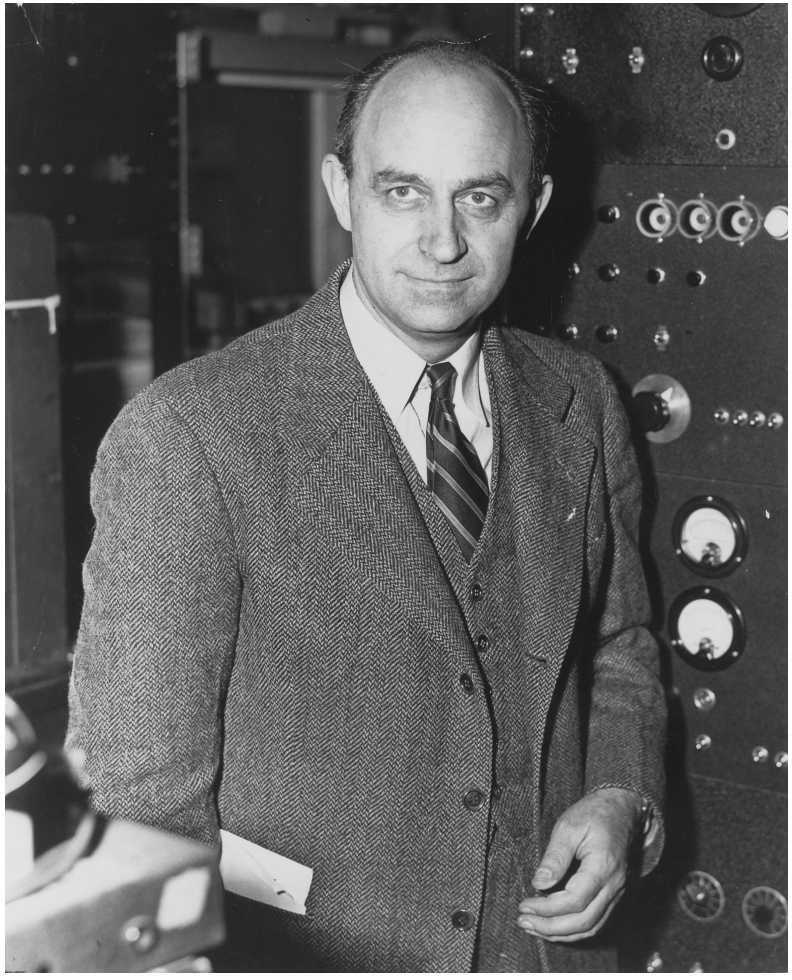
I hope that our Mars probes will discover nothing. It would be good news if we find Mars to be completely sterile. Dead rocks and lifeless sands would lift my spirit.
...

I'm hoping that our space probes will discover dead rocks and lifeless sands on Mars, on Jupiter's moon Europa, and everywhere else our astronomers look. It would keep alive the hope for a great future for humanity.

Det stora perspektivet

Universum: $\approx -13.8 \cdot 10^9$y
Jorden: $-4.5 \cdot 10^9$y
Prokaryotes: $-3.5 \cdot 10^9$y
Eukaryotes: $-1.7 \cdot 10^9$y
Homo sapiens: -200000y
Lämnar Afrika: -100000y
Jordbruk/komplexa samhällen: -12000y
Moderna nationer/stater: -400y
Elkraft: -120y
Kärnkraft: -60y
Internet: -30y
Mobilteknik: -15y
Storleken på synliga universum (Nature, 2012):
Antal stjärnor i vår galax: minst $300 \cdot 10^9$
Antal planeter i vår galax: minst $100 \cdot 10^9$
Antal galaxer i universum: minst $100 \cdot 10^9$
Antal planeter i universum: $\approx 10^{22}$

Enrico Fermi, 1950: Var är alla?



Trots enorma antalet planeter och andra galaxer, ser vi inga tecken på liv

Fermis resonemang

1. Solen är en typisk stjärna, relativt ung. Det finns miljarder av stjärnor i vår galax. Många är miljarder år äldre.
2. Vi kan förvänta oss att några av dessa stjärnor har planeter som liknar jorden. Några kan också utveckla intelligent liv.
3. Några av civilisationerna kan utveckla teknologin att färdas i rymden, något inom våra möjligheter redan nu (eller inom några tusen års teknisk utveckling).
4. Även med väldigt långsam rymdfärd, så kunde galaxen koloniseras på miljontals år. En enda sådan civilisation räcker.
5. Miljontals år är långt för människan, men kort på den *kosmologiska* tidsskalan av miljarder år.

Fermis paradox: vi ser inga tecken på liv, trots enorma antalet planeter

Många möjliga förklaringar

Varför ser vi inga tecken på liv, trots enorma antalet planeter?

- De är ointresserade av oss
- De har varit här, men åkte igen
- De är osynliga för oss (“mörk materia/energi” 94% av universum)
- De vill stanna hemma
- De har kommunikationsproblem
- ...
- Det stora filtret: det är väldigt svårt för liv att nå till teknologisk nivå att det koloniserar galaxen (Hanson, 1998)

Hypotes: Det stora filtret

Robin Hanson (1998):
There is a Great Filter between dead lifeless planets and advanced technological civilizations. All life in all civilizations eventually destroy themselves before acquiring the capacity to colonize space.

Hypotes: Det stora filtret

Robin Hanson (1998):
There is a Great Filter between dead lifeless planets and advanced technological civilizations. All life in all civilizations eventually destroy themselves before acquiring the capacity to colonize space.

Liv och död: skulle kunna förklara varför vi ej besökts/sett tecken på utomjordingar. Liv uppstår inte från första början, eller dör ut överallt.

Stora Filtret: potentiellt tidigare steg

Skulle det kunna ligga bakom oss?

1. Tur i universum? Rätt stjärna med rätt kemi och rätt avstånd från farliga föremål i rymden.
2. Naturliga hot? Asteroider, pandemier, döende stjärnor? Med oberoende händelser kommer civilisationer förr eller senare att lyckas ta sig förbi. Med den statistik vi ser för naturliga hot.
3. Uppkomst av självreplikerande molekyler? (RNA)
4. Prokaryotiskt liv? Uppstod först efter 1 miljard på jorden. Landmassor stelnar och hav bildas.
5. Från prokaryotiskt till eukarytiskt liv? Tog 1.8 miljarder år!
6. Steg kring ökande biologisk komplexitet? Kambriska explosionen, 500 miljoner år sedan.
7. Civilisationsutveckling? Social komplexitet? Stenåldern till industriella revolutionen?

Stora Filtret: potentiellt framtida steg

Kanske alla avancerade civilisationer upptäcker alldeles för kraftfull teknik?

- Kärnvapen (eller nya kraftfulla vapen i konflikt)
- Bioteknik
- Nanoteknik
- Artificiell intelligens
- "Unknown unknowns" ... saker vi har kvar att upptäcka

Annat: civilisationskollaps av t ex resursanvändning.

En enkel modell av liv och död i universum

Se varje planet som ett experiment, som beskrivs av två sannolikheter

p: liv uppstår på en planet och når teknologisk nivå mänsklig civilisation
q: detta når vidare teknologisk kapacitet med observerbar rymdkolonisering

I universum har vi runt $\mathbf{N} = 10^{22}$ experiment (planeter)

En enkel modell av liv och död

Förväntat antal planeter med liv och avancerad rymdteknologi:

$$Npq$$

1. Npq kan inte vara särskilt stort (Fermis tystnad)
2. Om N är stort ($\approx 10^{22}$), så måste pq vara litet
   Med $p = q = 10^{-9}$ skulle några tusen avancerade civilisationer finnas.
3. Liv på Mars skulle tala för ett lite större p än vi tidigare trott
4. Det talar för ett lite mindre q

Bostrom (2008)

If we discovered some very simple life forms on Mars in its soil or under the ice at the polar caps, it would show that the Great Filter must exist somewhere after that period in evolution. This would be disturbing, but we might still hope that the Great Filter was located in our past.
If we discovered a more advanced life-form, such as some kind of multi-cellular organism, that would eliminate a much larger stretch of potential locations where the Great Filter could be. The effect would be to shift the probability more strongly to the hypothesis that the Great Filter is ahead of us, not behind us.

Vad vet vi om p och q?

- De är sannolikheter i intervallet mellan 0 och 1 (gränserna 0%,100%)
- Vi vet att $p > 0$
- Vi tror också att $q > 0$

Kan vi säga något vidare om p och q?

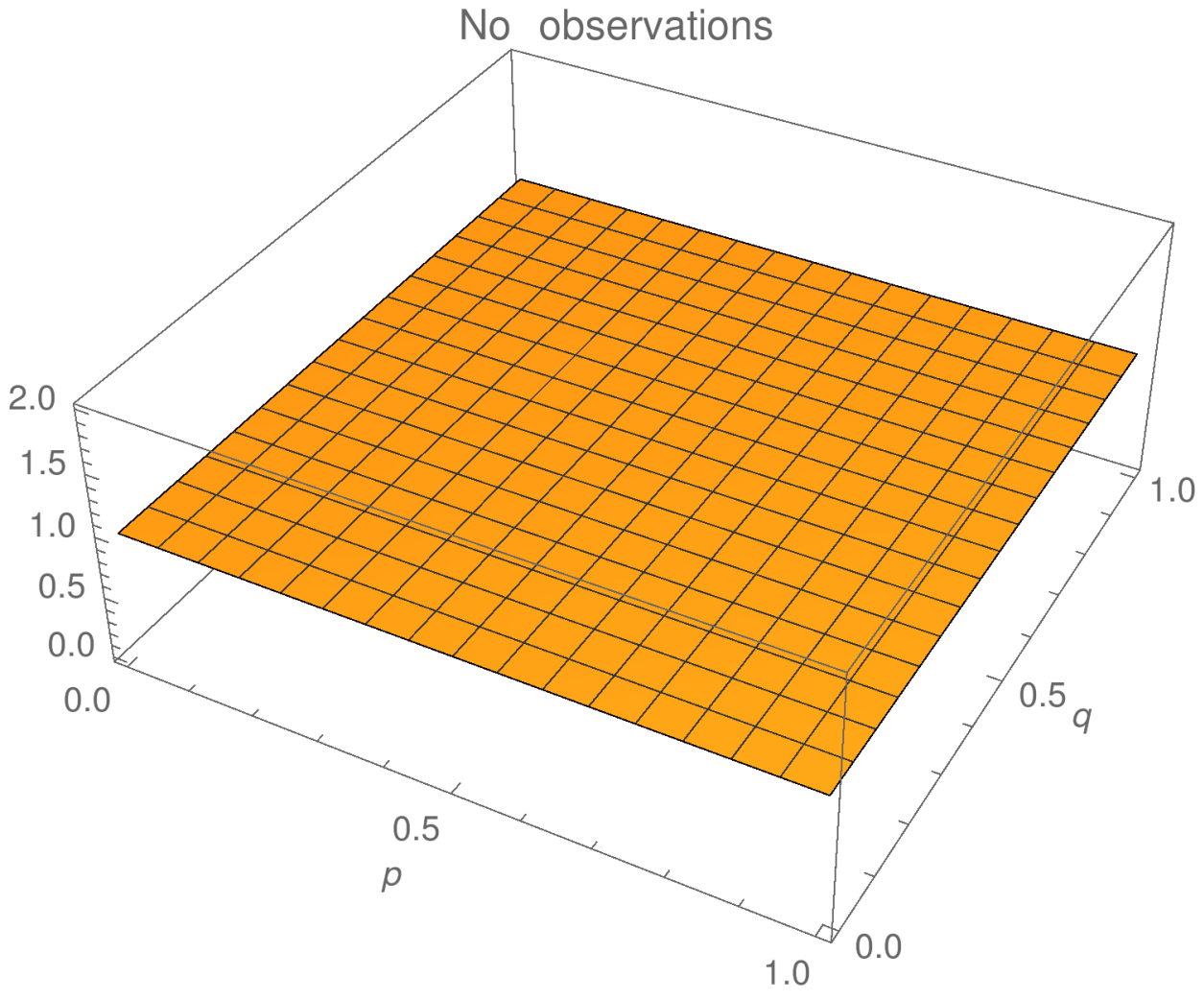
Bayesiansk statistik

Vi vet inte vad p och q är, men vi kan analysera osäkerheten

Bayesiansk analys:

1. Beskriv osäkerheten över värdena p och q
2. Bayes regel: hur observationer påverkar osäkerheten
3. Våra observationer: Fermis stora tystnad + Liv på Mars

Osäkerhet över p, q (prior-fördelning)



Modell

Varje planet som ett experiment, antag oberoende:

$X \sim Bin(N, pq)$ Antal supercivilisationer

$Y \sim Bin(1, p)$ En undersökt planet

Våra observationer:

$P(X = 0) = (1 - pq)^N$

$P(Y = 1) = p$

Bayesiansk statistik: ändra osäkerheten om p och q efter observationer.

Bayesiansk uppdatering

Generell Bayes regel:

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)}$$

Med prior $f_{P,Q}(p,q)$ och sannolikhet för D:

$$f_{P,Q}(p,q|D) = \frac{f_{P,Q}(D|p,q)f_{P,Q}(p,q)}{\int\int f_{P,Q}(D|p',q')f_{P,Q}(p',q')dp'dq'}$$

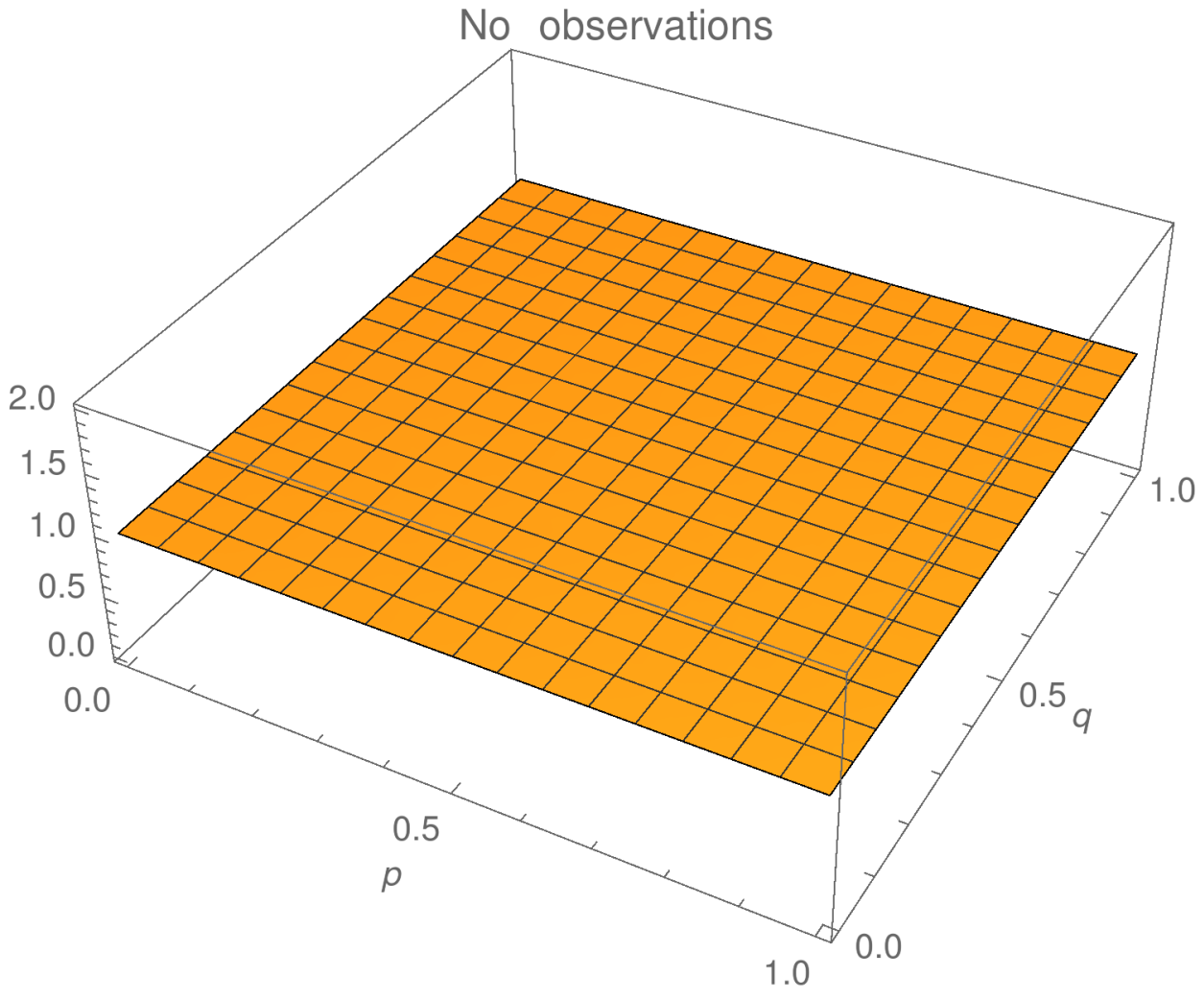
Posterior efter X=0:

$$f_{P,Q}(p,q|D) = \frac{(1-pq)^N}{\int_0^1\int_0^1(1-p'q')^N dp'dq'}$$

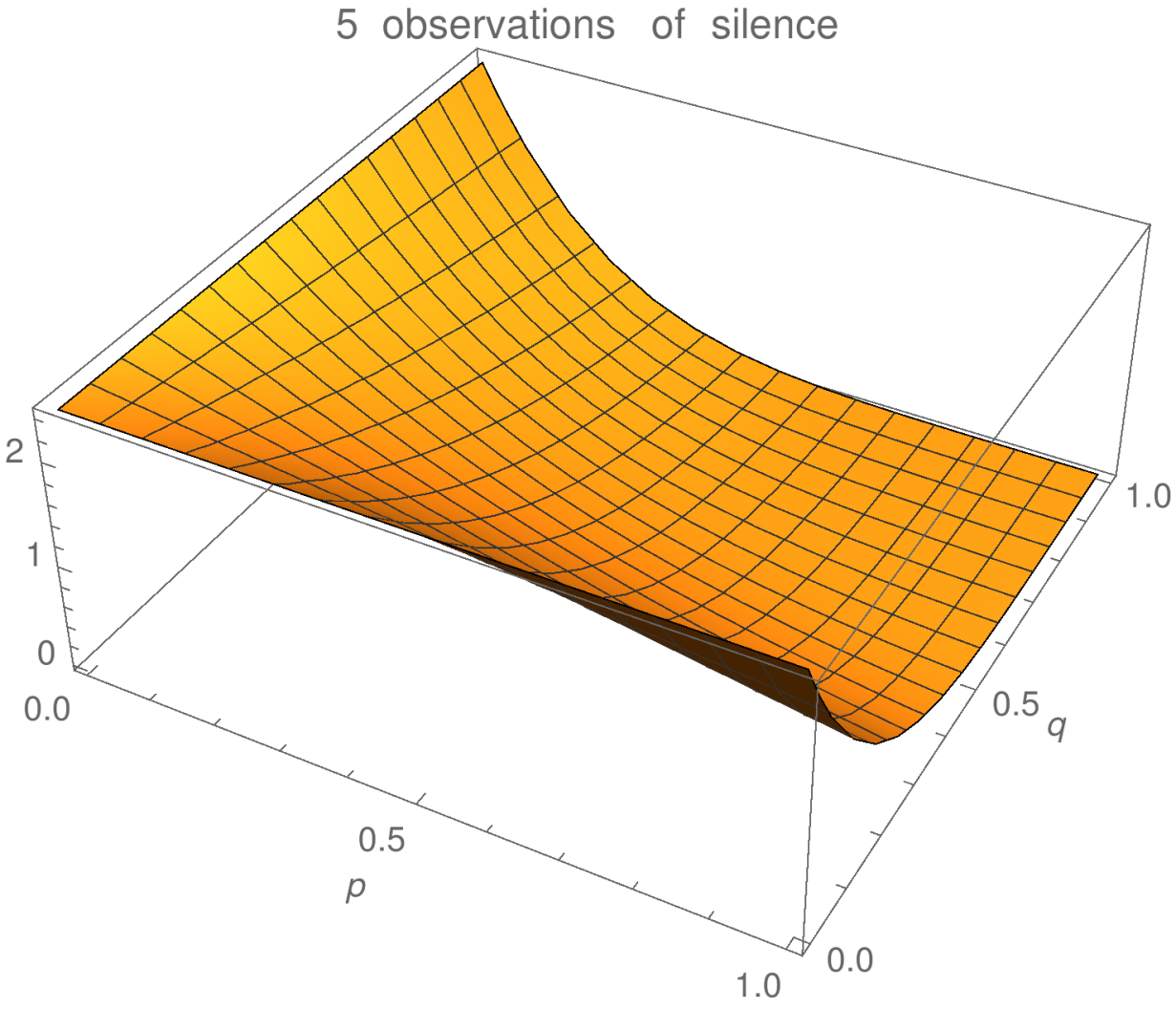
Posterior efter X=0,Y=1:

$$f_{P,Q}(p,q|D) = \frac{p(1-pq)^N}{\int_0^1\int_0^1 p'(1-p'q')^N dp'dq'}$$

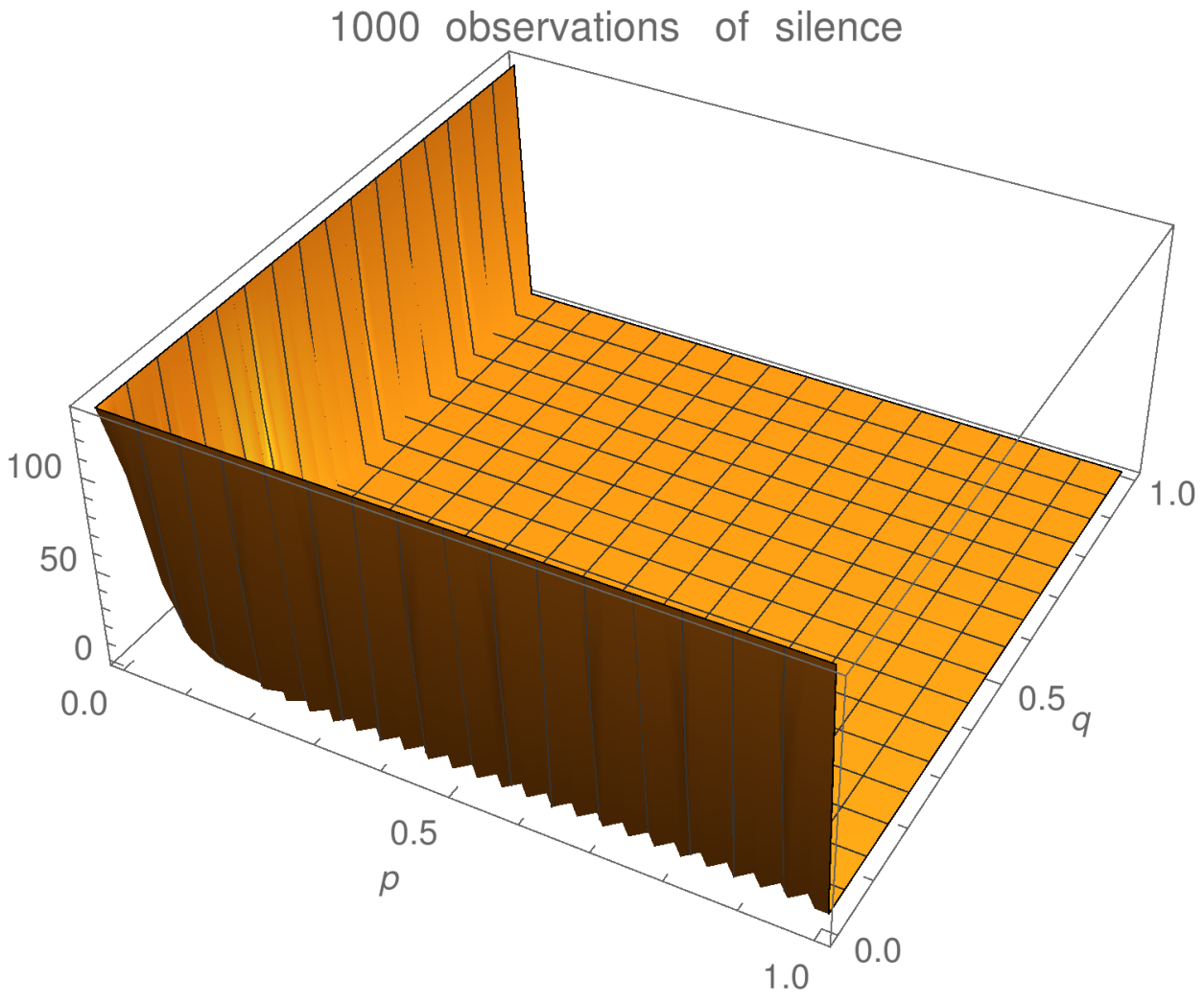
Inga observationer alls



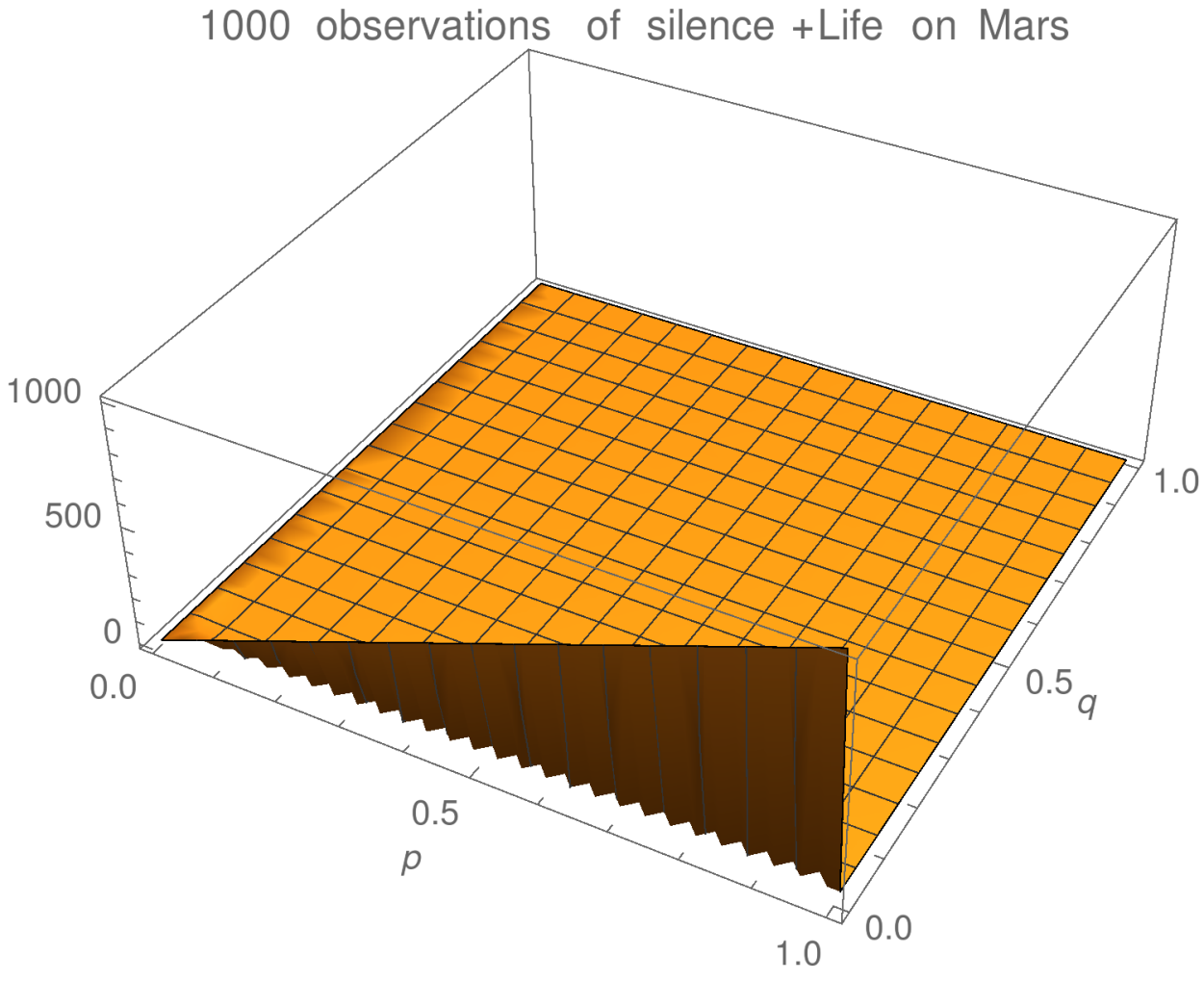
Bayesiansk uppdatering



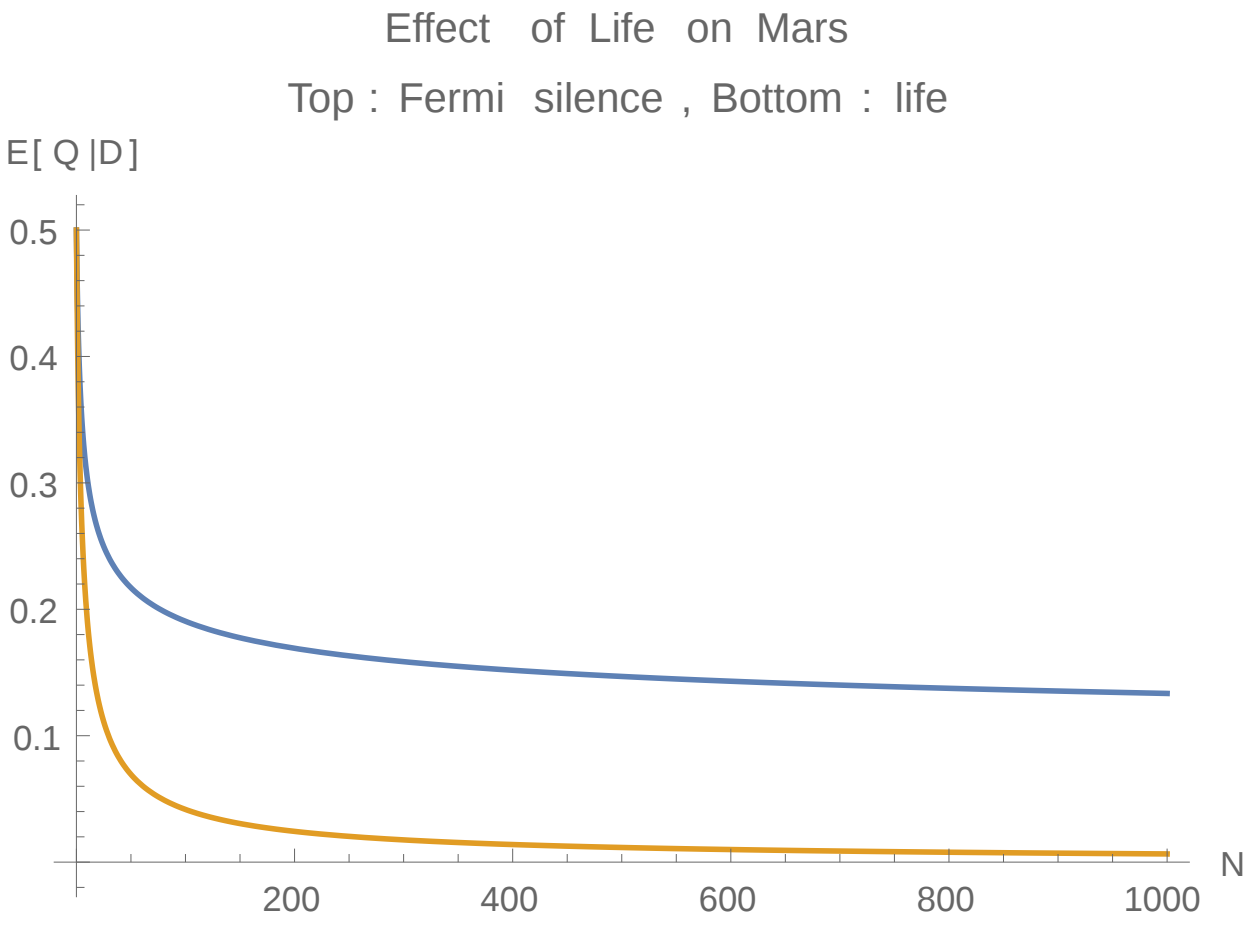
Bayesiansk uppdatering



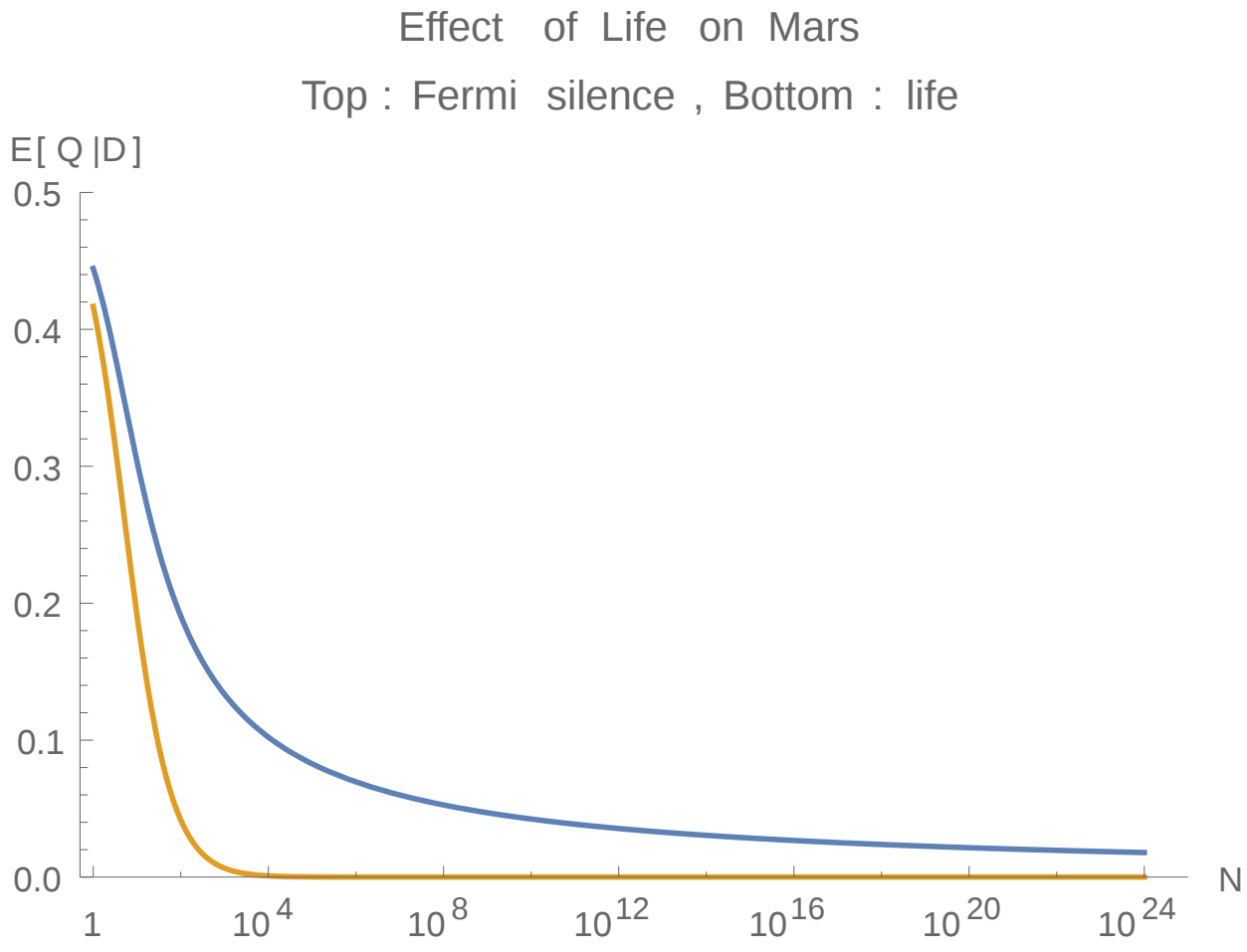
Bayesiansk uppdatering: efter liv på Mars



Bayesiansk uppdatering: effekt på q

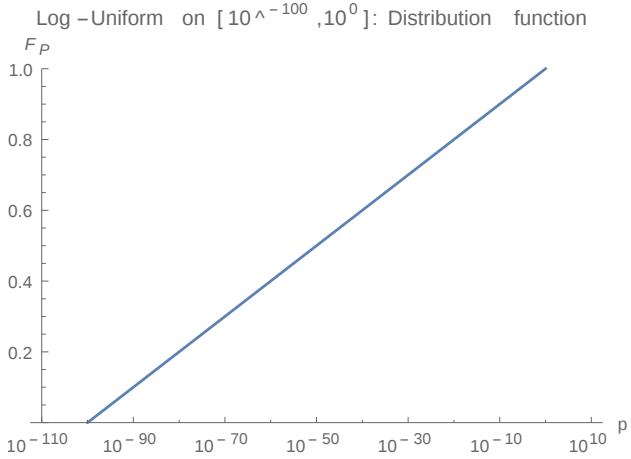


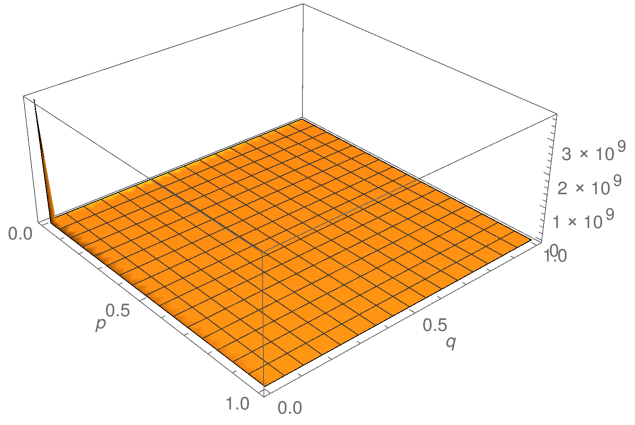
Bayesiansk uppdatering: effekt på q



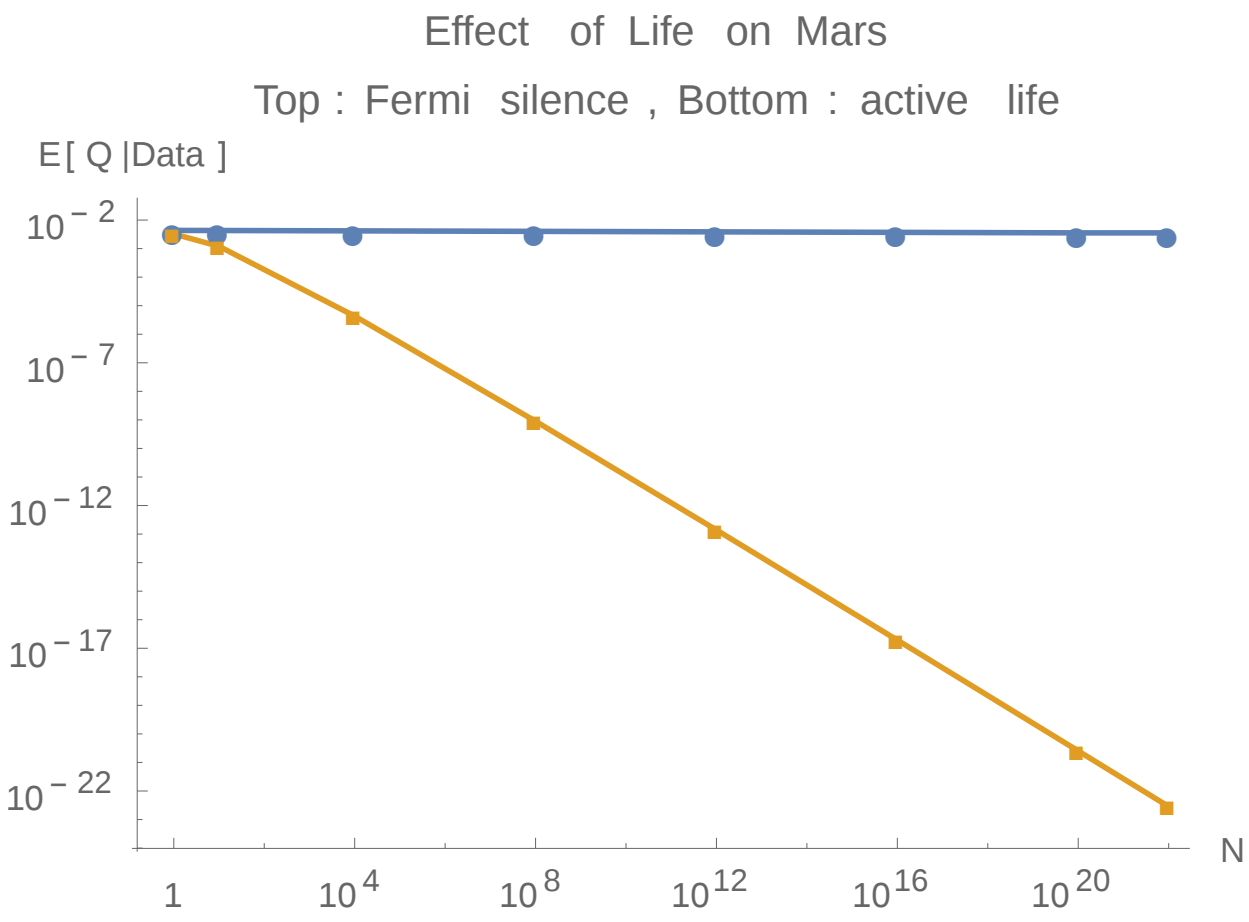
Mer rimlig första osäkerhet: log-uniform

Vi skulle kunna att tro att p och q skulle kunna vara riktigt små.
En mer rimlig prior är log-uniform.





log-uniform: effekt på q

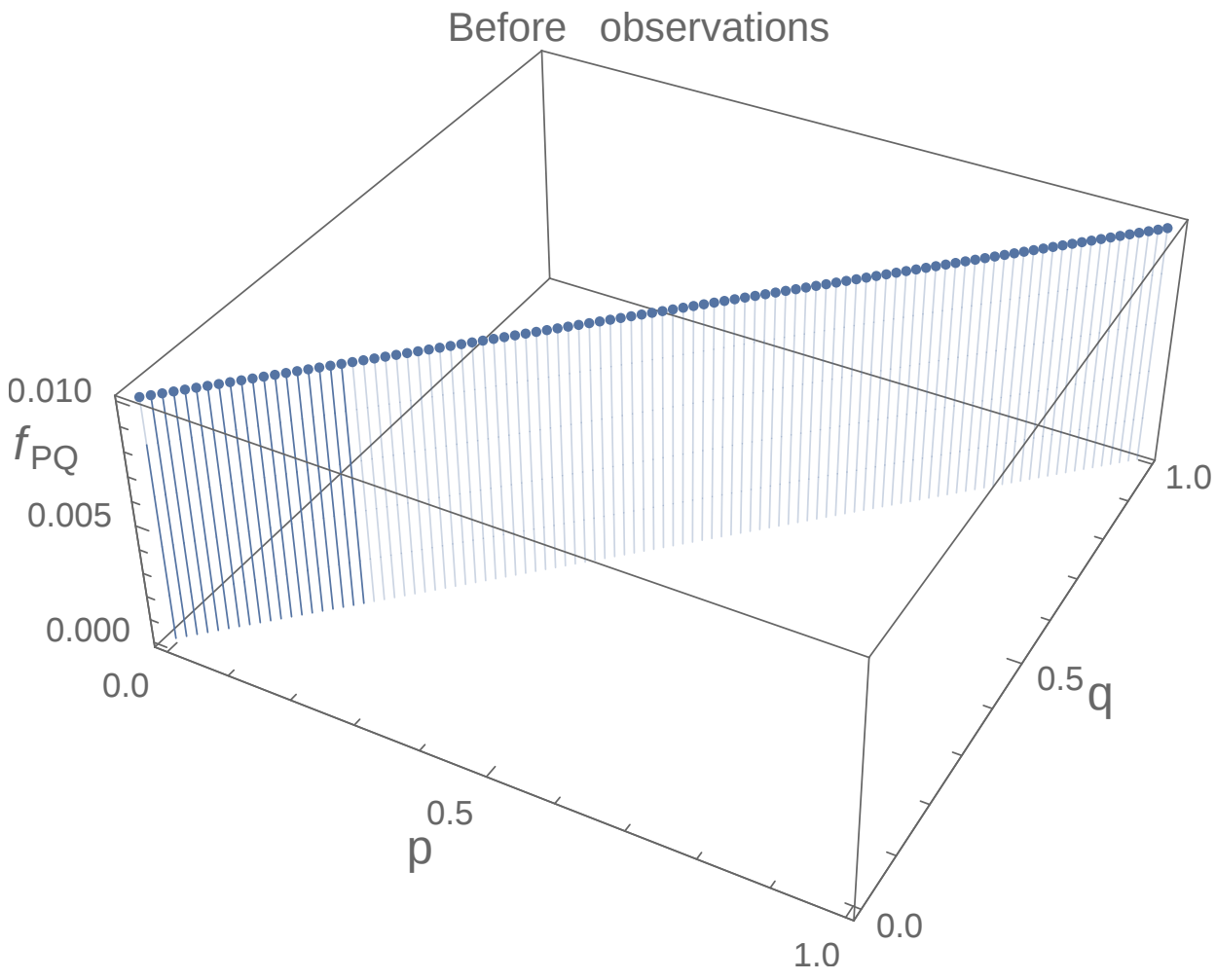


Stämmer Bostroms intuition alltid?

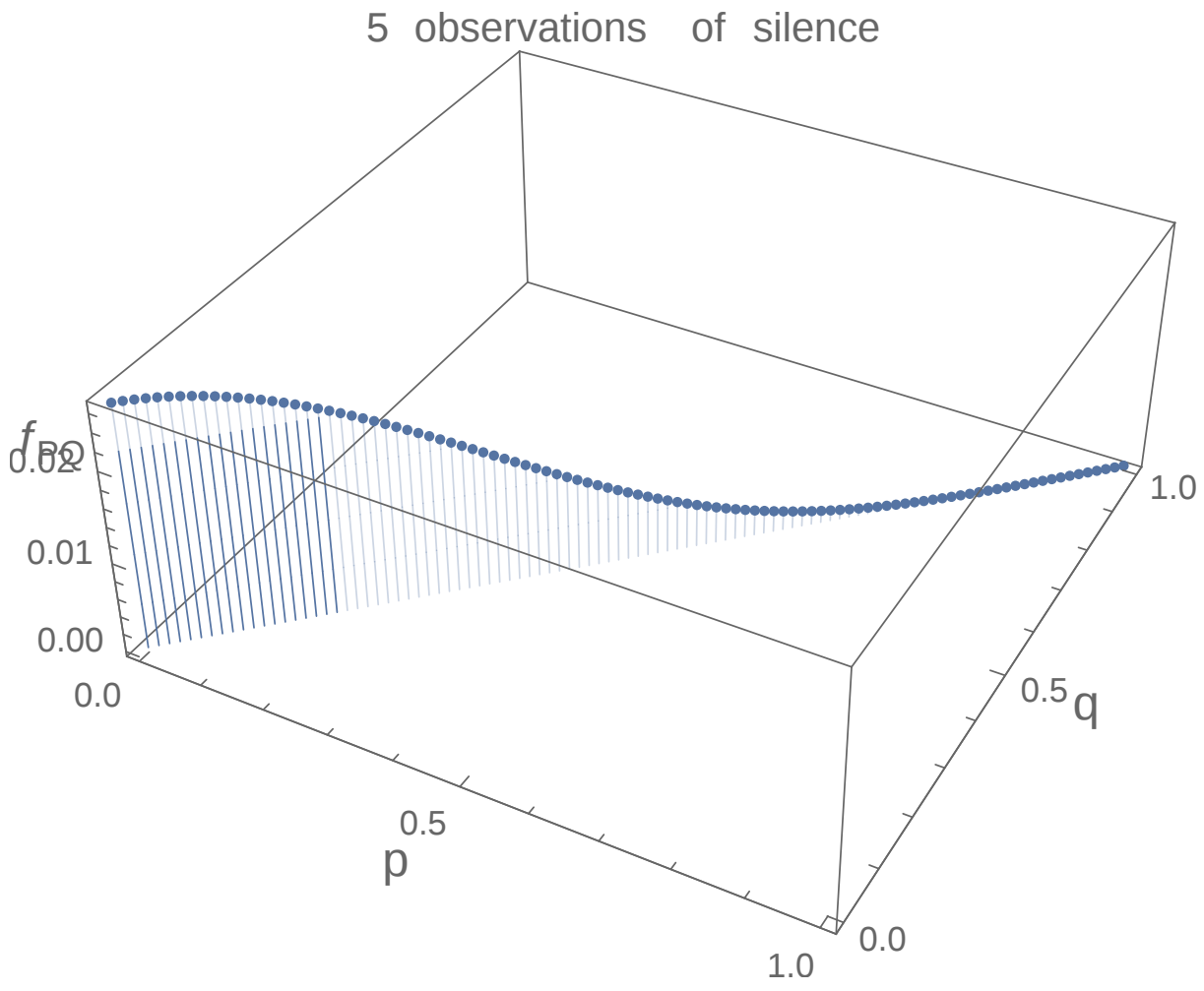
En frestande slutsats: liv på Mars sänker alltid vår tro för q.
Bostrom:

> "The effect would be to shift the probability more strongly to the hypothesis that the Great Filter is ahead of us, not behind us."

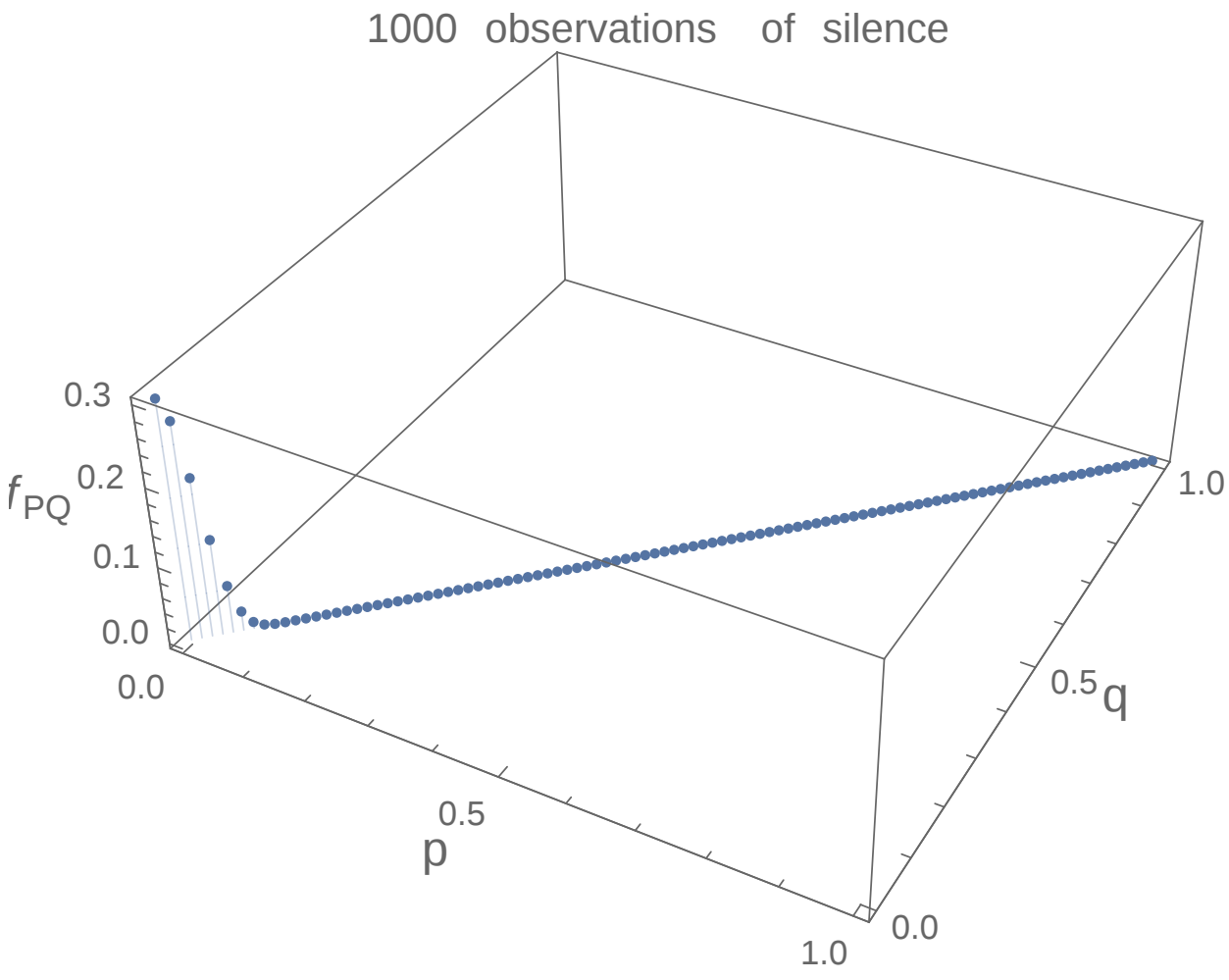
Ett motexempel mot Bostrom



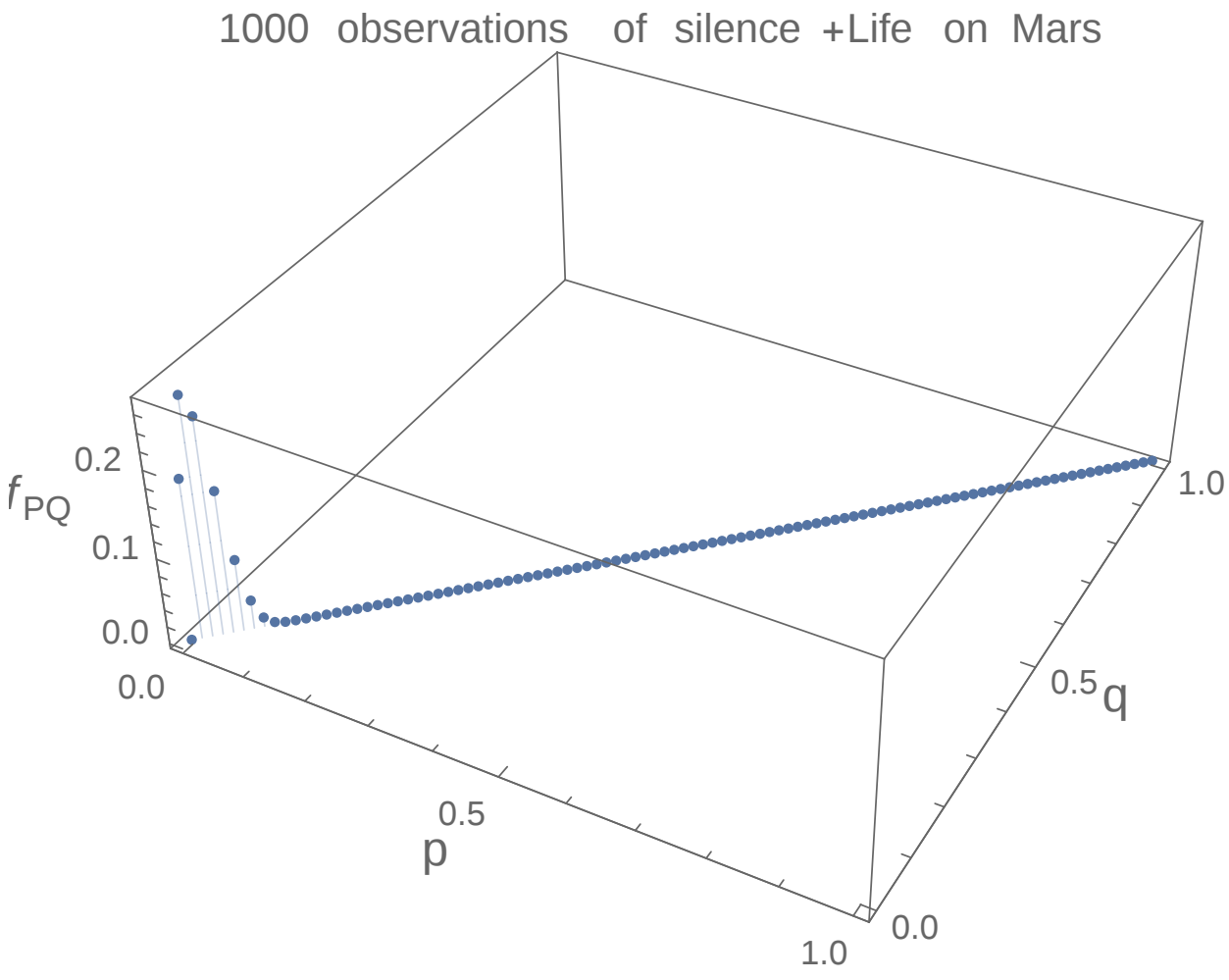
Ett motexempel mot Bostrom



Ett motexempel mot Bostrom



Ett motexempel mot Bostrom



Gäller även för andra priors

- Kan även gälla för priors inte enbart på diagonalen: täcker hela ytan
- Mycket av prior-massan på diagonalen: universum som befrämjar komplexitet på olika nivåer (p och q starkt korrelerade)

Slutsats

1. Fermis paradox: inga tecken på liv i ett stort gammalt universum
2. Det stora filtret har föreslagits för att förklara Fermis paradox
3. Liv på Mars kan tala för ett lägre q (liv på vår nivå svårt att överleva)
4. Matematik visar att det kan gälla, men motexempel finns!
5. Liknande argument för primitivt liv (lite mer komplicerad modell)
6. Ett exempel på Bayesiansk statistik och möjligtvis rationell uppskattning av existentiell risk.