MVE420:
Existentiell risk, Fermis paradox och det stora filtret

Vilhelm Verendel
2015-03-27

Definition av risk

Två användningar av begreppet *risk*

Definition av risk

Två användningar av begreppet *risk*

Definition av risk

Två användningar av begreppet *risk*

1. Risk, generellt: potentiellt negativa framtida konsekvenser

Definition av risk

Två användningar av begreppet *risk*

1. Risk, generellt: potentiellt negativa framtida konsekvenser
2. Distinktion inom beslutsteori:
    1. *Risk*: händelser med *kända* sannolikheter
    2. *Osäkerhet*: händelser med *okända* sannolikheter

Definition av risk

Två användningar av begreppet *risk*

1. Risk, generellt: potentiellt negativa framtida konsekvenser
2. Distinktion inom beslutsteori:
   1. *Risk*: händelser med *kända* sannolikheter
      Exempel: "Risk att förlora vid roulettebordet"
   2. *Osäkerhet*: händelser med *okända* sannolikheter
      Exempel: "Risk för kärnvapenkrig de närmsta 50 åren"

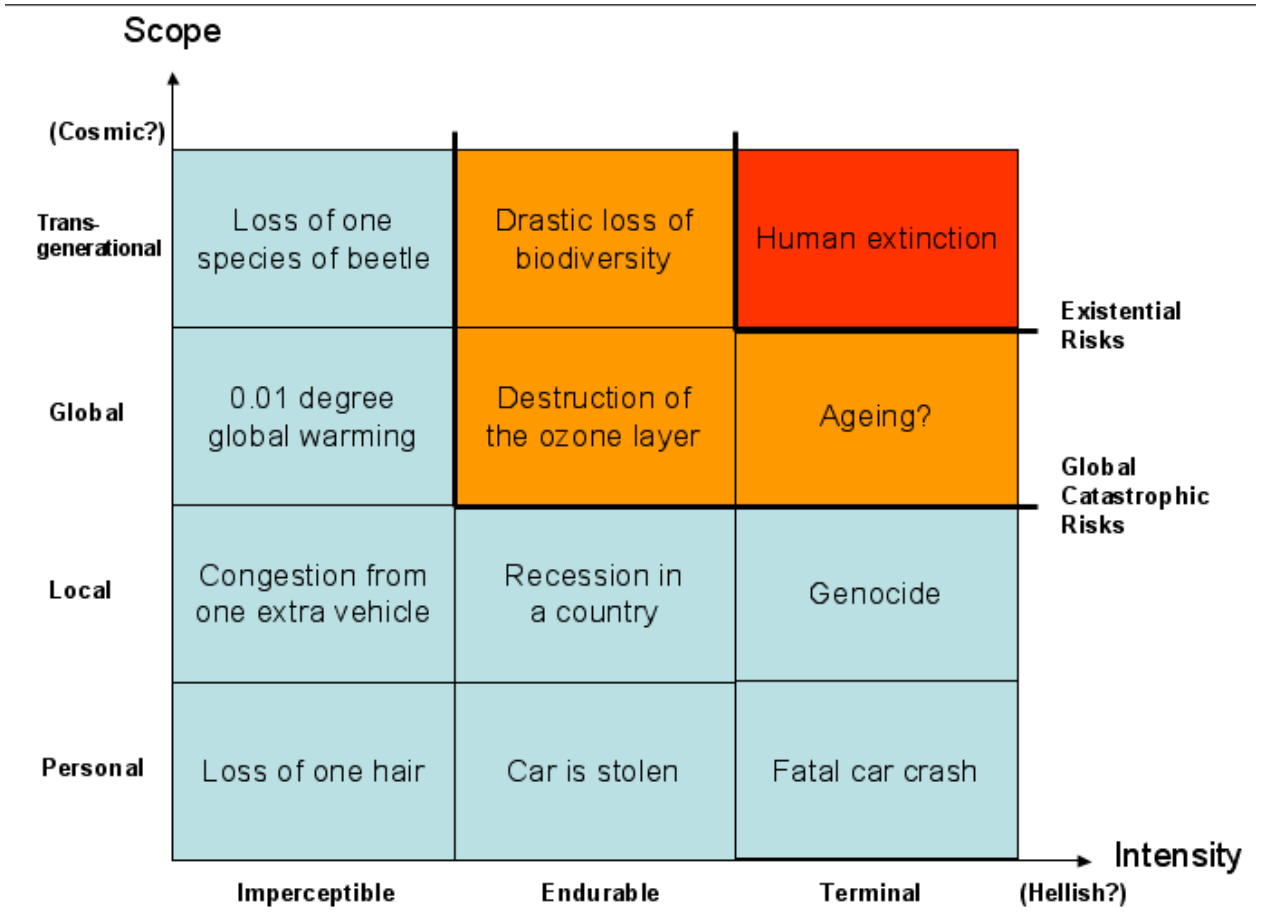
Definition av risk

Två användningar av begreppet *risk*

1. Risk, generellt: potentiellt negativa framtida konsekvenser
2. Distinktion inom beslutsteori:
   1. *Risk*: händelser med *kända* sannolikheter
      Exempel: "Risk att förlora vid roulettebordet"
   2. *Osäkerhet*: händelser med *okända* sannolikheter
      Exempel: "Risk för kärnvapenkrig de närmsta 50 åren"

Om inget nämns använder vi *risk* generellt (negativa konsekvenser)

En mängd framtida risker (Bostrom)



Vad vet vi om existentiell risk?

Myndigheten för Samhällsskydd och Beredskap (2014)

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.
- En knapp majoritet (54%) tror att mänskligheten kommer att dö ut.

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.
- En knapp majoritet (54%) tror att mänskligheten kommer att dö ut.
- En av sex tror att mänskligheten kommer att dö ut inom 500 år. Ytterst få (1%) tror att mänskligheten dör ut inom deras livstid.

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.
- En knapp majoritet (54%) tror att mänskligheten kommer att dö ut.
- En av sex tror att mänskligheten kommer att dö ut inom 500 år. Ytterst få (1%) tror att mänskligheten dör ut inom deras livstid.
- Tre av tio (28%) av dem som tror att mänskligheten någon gång kommer att gå under tror att klimatförändringar kommer att orsaka detta

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.
- En knapp majoritet (54%) tror att mänskligheten kommer att dö ut.
- En av sex tror att mänskligheten kommer att dö ut inom 500 år. Ytterst få (1%) tror att mänskligheten dör ut inom deras livstid.
- Tre av tio (28%) av dem som tror att mänskligheten någon gång kommer att gå under tror att klimatförändringar kommer att orsaka detta
- Drygt en av tio tror att undergången kommer att orsakas av resursbrist (13%), världsomfattande krig (12%), att solen slocknar (11%) eller en pandemi (10%).

Myndigheten för Samhällsskydd och Beredskap (2014)

- 6683 intervjuer med 18-åringar i Sverige.
- En knapp majoritet (54%) tror att mänskligheten kommer att dö ut.
- En av sex tror att mänskligheten kommer att dö ut inom 500 år. Ytterst få (1%) tror att mänskligheten dör ut inom deras livstid.
- Tre av tio (28%) av dem som tror att mänskligheten någon gång kommer att gå under tror att klimatförändringar kommer att orsaka detta
- Drygt en av tio tror att undergången kommer att orsakas av resursbrist (13%), världsomfattande krig (12%), att solen slocknar (11%) eller en pandemi (10%).

Expertundersökning på Global Catastrophic Risks-konferensen 2008

Uppskattad sannolikhet för undergång innan 2100: medianen 19%

Varför tänka mer på existentiell risk?

Låt oss undersöka vad som menas med begreppet

Varför tänka mer på existentiell risk?

Låt oss undersöka vad som menas med begreppet

Definition av Bostrom (2013)

An *existential risk* is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development.

Varför tänka mer på existentiell risk?

Låt oss undersöka vad som menas med begreppet

Definition av Bostrom (2013)

An *existential risk* is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development.

1. Definitionen berör "Earth-originating intelligent life" - borde det inte vara människor?

Varför tänka mer på existentiell risk?

Låt oss undersöka vad som menas med begreppet

Definition av Bostrom (2013)

An *existential risk* is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development.

1. Definitionen berör "Earth-originating intelligent life" - borde det inte vara människor?
2. Definitionen fångar totala undergångshändelser, men verkar inte begränsad till dessa. Kan även andra händelser vara existentiella?

Varför tänka mer på existentiell risk?

Låt oss undersöka vad som menas med begreppet

Definition av Bostrom (2013)

An *existential risk* is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development.

1. Definitionen berör "Earth-originating intelligent life" - borde det inte vara människor?
2. Definitionen fångar totala undergångshändelser, men verkar inte begränsad till dessa. Kan även andra händelser vara existentiella?
3. Definitionen innehåller "drastic destruction of its potential": vilken potential är det som avses?

“Earth-originating intelligent life”

Den stora bilden

Universum: $\approx -13.8 \cdot 10^9$y

“Earth-originating intelligent life”

Den stora bilden

Universum: $\approx -13.8 \cdot 10^9$y
Jorden: $-4.5 \cdot 10^9$y

“Earth-originating intelligent life”

Den stora bilden

Universum: $\approx -13.8 \cdot 10^9$y
Jorden: $-4.5 \cdot 10^9$y
Prokaryotes: $-3.5 \cdot 10^9$y

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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

“Earth-originating intelligent life”

Den stora bilden

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
Angry Birds: -5y

“Earth-originating intelligent life”

Den stora bilden

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
Angry Birds: -5y

Är det rimligt att mänskligheten 2015 är “slutprodukten”?

“Earth-originating intelligent life”

Den stora bilden

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
Angry Birds: -5y

Är det rimligt att mänskligheten 2015 är “slutprodukten”?

Uppskattningar: jorden kan vara beboelig ytterligare 1 miljard år.
Därav intelligent liv som härrör från oss.

Vilken “drastic destruction” av potential avses?

Exempel av Parfit (1984)

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

Vilken “drastic destruction” av potential avses?

Exempel av Parfit (1984)

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

1. Peace.
2. A nuclear war that kills 99% of the world's existing population.
3. A nuclear war that kills 100%.

Vilken "drastic destruction" av potential avses?

Exempel av Parfit (1984)

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

1. Peace.
2. A nuclear war that kills 99% of the world's existing population.
3. A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences?

Vilken "drastic destruction" av potential avses?

Exempel av Parfit (1984)

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

1. Peace.
2. A nuclear war that kills 99% of the world's existing population.
3. A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2).
Which is the greater of these two differences?
Most people believe that the greater difference is between (1) and (2). I believe that the difference between (2) and (3) is very much greater.

Parfit: undergång vore mycket värre

Varför?

1. 99% av mänskligheten: $6.93 \cdot 10^9$ människor
2. 100%: $7 \cdot 10^9$ människor
3. Att förlora det senare förhindrar något mer:

Parfit: undergång vore mycket värre

Varför?

1. 99% av mänskligheten: $6.93 \cdot 10^9$ människor
2. 100%: $7 \cdot 10^9$ människor
3. Att förlora det senare förhindrar något mer: *alla framtida generationer*

Framtida potential på jorden

Ett exempel (Bostrom)

1. Antag hållbart liv på jorden för 1 miljard människor
2. Antag genomsnittlig livslängd på 100 år
3. Antag 1 miljard år framför oss på jorden

Framtida potential på jorden

Ett exempel (Bostrom)

1. Antag hållbart liv på jorden för 1 miljard människor
2. Antag genomsnittlig livslängd på 100 år
3. Antag 1 miljard år framför oss på jorden

Under förenklande antaganden: $\approx 10^{16}$ framtida personer på jorden
I absoluta tal, jämför detta med $7 \cdot 10^9$

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv

Försiktiga uppskattningar av forskare

Vintergatan uppskattas till mellan 100000-120000 ljusår i diameter Hair och Hedman, 2012:

1. Antag teknologi för att färdas med 0.25% ljusets hastighet
2. Efter 50m år, uniform spridning över 130000 ljusår
3. Kort period på den kosmiska tidsskalan

Försiktiga uppskattningar av forskare

Vintergatan uppskattas till mellan 100000-120000 ljusår i diameter Hair och Hedman, 2012:

1. Antag teknologi för att färdas med 0.25% ljusets hastighet
2. Efter 50m år, uniform spridning över 130000 ljusår
3. Kort period på den kosmiska tidsskalan

Antag efter tusentals år av teknikutveckling, att vi kan färdas med 10% ljushastighet. Då når vi många stjärnor inom hundratalet år.

Försiktiga uppskattningar av forskare

Vintergatan uppskattas till mellan 100000-120000 ljusår i diameter
Hair och Hedman, 2012:

1. Antag teknologi för att färdas med 0.25% ljusets hastighet
2. Efter 50m år, uniform spridning över 130000 ljusår
3. Kort period på den kosmiska tidsskalan

Antag efter tusentals år av teknikutveckling, att vi kan färdas med 10% ljushastighet. Då når vi många stjärnor inom hundratalet år.

Andra galaxer: miljoner ljusår bort, t ex, Andromeda på 2.5m ljusår
Poängen: kort på universums tidsskala

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden
3. Argumentet oberoende av rätt storleksordning: poängen är enorm framtida potential för civilisation med teknologi för rymdkolonisering

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden
3. Argumentet oberoende av rätt storleksordning: poängen är enorm framtida potential för civilisation med teknologi för rymdkolonisering

Vi kan förstå "drastic destruction" i absoluta termer som existentiell risk: att inte realisera de scenarier som är inom ramen för våra möjligheter

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden
3. Argumentet oberoende av rätt storleksordning: poängen är enorm framtida potential för civilisation med teknologi för rymdkolonisering

Vi kan förstå "drastic destruction" i absoluta termer som existentiell risk: att inte realisera de scenarier som är inom ramen för våra möjligheter

- Civilisationskollaps eller inlåsning på låg teknologisk nivå

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden
3. Argumentet oberoende av rätt storleksordning: poängen är enorm framtida potential för civilisation med teknologi för rymdkolonisering

Vi kan förstå "drastic destruction" i absoluta termer som existentiell risk: att inte realisera de scenarier som är inom ramen för våra möjligheter

- Civilisationskollaps eller inlåsning på låg teknologisk nivå
- Att vi aldrig når en potentiell teknologisk nivå för rymdkolonisering

Mer optimistiska scenarier: framtida potential i universum

Mer än 10^{16} framtida liv?
Två scenarier från Bostrom:

1. Rymdkolonisering, nuvarande biologi: 10^{34} liv
2. Uploading: 10^{54} teoretiskt möjligt, väldigt spekulativa antaganden
3. Argumentet oberoende av rätt storleksordning: poängen är enorm framtida potential för civilisation med teknologi för rymdkolonisering

Vi kan förstå "drastic destruction" i absoluta termer som existentiell risk: att inte realisera de scenarier som är inom ramen för våra möjligheter

- Civilisationskollaps eller inlåsning på låg teknologisk nivå
- Att vi aldrig når en potentiell teknologisk nivå för rymdkolonisering
- Notera: *givet* det som är möjligt för mänskligheten att realisera

Tre kategorier av existentiell risk

1. Risker från naturen
2. Risker från oönskade konsekvenser av mänskliga handlingar
3. Risker från avsiktliga konsekvenser av mänskliga handlingar

Framtida existentiella risker: naturen

- Asteroider
- Sjukdomar
- Storskaliga vulkanutbrott
- Döende stjärnor och kosmologiska fenomen
- ...

Framtida existentiella risker: människor

- Krig (kärnvapen, nya kraftfulla vapen)
- Miljöpåverkan (t ex riskabel geoengineering för klimathotet)
- Bioteknik
- Nanoteknik
- Artificiell intelligens
- ...
- "Unknown unknowns": framtida uppfinningar och innovationer

Svårt att dra en tydlig gräns mellan oönskade/avsiktliga konsekvenser

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
Angry Birds: -5y

- Naturliga risker: vi har överlevt i ett par hundra tusen år.
- Kraftfulla teknologier: begränsad eller ingen tidsserie.

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier
- Experter bedömer gränsen för global katastrof till 1 km i diameter

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier
- Experter bedömer gränsen för global katastrof till 1 km i diameter
- Data: detta har skett ungefär 1 gång på 1000000 år

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier
- Experter bedömer gränsen för global katastrof till 1 km i diameter
- Data: detta har skett ungefär 1 gång på 1000000 år
- Uppskattning: sannolikheten för detta det nästa århundrade ≈ 0.0001

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier
- Experter bedömer gränsen för global katastrof till 1 km i diameter
- Data: detta har skett ungefär 1 gång på 1000000 år
- Uppskattning: sannolikheten för detta det nästa århundrade ≈ 0.0001
- Viktigt antagande: vi har skäl att tro att statistiken är *stationär* under kommande århundraden/årtusenden

Exempel: asteroidnedslag (Häggström 2015)

- Kollision av stora asteroider mot jorden är existentiella risker
- Sibirien, 2013: en meteroit $17 - 20m$ i diameter skadar 1500 människor
- Shoemaker-Levy, en komet $2 - 5$ km i diameter, träffar Jupiter (1994) vilket skulle varit en global katastrof
- Chicxulub, en asteroid med minst 10 km i diameter, träffade jorden 66 miljoner år sedan och ledde till massutrotning av dinosaurier
- Experter bedömer gränsen för global katastrof till 1 km i diameter
- Data: detta har skett ungefär 1 gång på 1000000 år
- Uppskattning: sannolikheten för detta det nästa århundrade ≈ 0.0001
- Viktigt antagande: vi har skäl att tro att statistiken är *stationär* under kommande århundraden/årtusenden
- Stationär, kvalitativt: vår närvaro, våra handlingar samt andra omständigheter verkar inte påverka dessa risker nämnvärt

Andra naturliga risker

- Naturliga pandemier
- Supervulkaner
- Supernovor
- Solen
- ...

Liknande statistiskt läge för naturliga risker: små risker per århundrade
För framtida risker från nya teknologier har vi ingen välgrundad statistik

Konceptuella problem med existentiella risker

Tre sorters problem: psykologiska, statistiska, konceptuella

1. Ett antal psykologiska effekter och vanliga sorters felslut kan störa ut vår förmåga till riskbedömning
2. Statistiska/kunskapsteoretiska: bedöma *händelser som aldrig inträffat*
3. Vissa sorters existentiell risk går *per definition* inte att mäta mer än en gång...

Mycket begränsad förmåga att direkt och kontrollerat mäta existentiell risk

Analys av existentiell risk

En blandning av spekulation, prediktion och osäkerhet. Vad kan vi göra?

Resonera så systematiskt och kritiskt det bara går om framtiden:

Analys av existentiell risk

En blandning av spekulation, prediktion och osäkerhet. Vad kan vi göra?

Resonera så systematiskt och kritiskt det bara går om framtiden:

- Statistik för naturliga risker: undre gräns för existentiell risk
  Minst ≈ 1 promille per århundrade
- Sannolikhet för framtida risker (utan stationära tidsserier)
  1. Experter
  2. Reasonera systematiskt under osäkerhet (Bayesiansk statistik)
  3. Generella argument om existentiell risk (utan data om specifika risker)

När kan (1) – (3) användas?

Möjlighet #1: Fråga experter

Expertundersökning på Global Catastrophic Risks-konferensen 2008
Medianer för uppskattade sannolikheter för existentiell risk detta århundrade:

- Vapen baserade på molekylär nanoteknik: 5%
- Superintelligent AI: 5%
- Krig (inklusive kärnvapen): 4%
- Avsiktligt skapade pandemier: 2%
- Kärnvapenkrig: 1%

- Total existentiell risk innan 2100: 19%
  Jämför med storlek på naturliga risker

Problem med expertforskning

- Experter har inte alltid rätt: modeller är ibland bättre än experter
- Många psykologiska faktorer kan påverka experters bedömningar
- Vetenskaplig standard: subjektiva omdömen längre från verkligheten
  Frågan vem som räknas som expert verkar ha en del godtycklighet

Möjlighet #2: Resonera under osäkerhet

Bayesiansk statistik

- Utgångsläge: det finns många hypoteser om storlek på existentiell risk
- Vi vill egentligen svara på frågan "vad är sannolikheten p för X?" men saknar möjligheten att göra direkta experiment

Bayesiansk statistik: kan kvantitativt beskriva osäkerheten vi har över p

Möjlighet #2: Resonera under osäkerhet

Bayesiansk statistik

- Utgångsläge: det finns många hypoteser om storlek på existentiell risk
- Vi vill egentligen svara på frågan "vad är sannolikheten p för X?" men saknar möjligheten att göra direkta experiment

Bayesiansk statistik: kan kvantitativt beskriva osäkerheten vi har över p

- Osäkerhet för p: beskrivs som en *prior*-fördelning

Möjlighet #2: Resonera under osäkerhet

Bayesiansk statistik

- Utgångsläge: det finns många hypoteser om storlek på existentiell risk
- Vi vill egentligen svara på frågan "vad är sannolikheten p för X?" men saknar möjligheten att göra direkta experiment

Bayesiansk statistik: kan kvantitativt beskriva osäkerheten vi har över p

- Osäkerhet för p: beskrivs som en *prior*-fördelning
- Uppskatta *prior*-fördelningar: t ex genom systematiska frågor till experter. Priors kan vägas ihop med observationer.

Möjlighet #2: Resonera under osäkerhet

Bayesiansk statistik

- Utgångsläge: det finns många hypoteser om storlek på existentiell risk
- Vi vill egentligen svara på frågan "vad är sannolikheten p för X?" men saknar möjligheten att göra direkta experiment

Bayesiansk statistik: kan kvantitativt beskriva osäkerheten vi har över p

- Osäkerhet för p: beskrivs som en *prior*-fördelning
- Uppskatta *prior*-fördelningar: t ex genom systematiska frågor till experter. Priors kan vägas ihop med observationer.
- Anpassning: givet modell och observationer D

Möjlighet #2: Resonera under osäkerhet

Bayesiansk statistik

- Utgångsläge: det finns många hypoteser om storlek på existentiell risk
- Vi vill egentligen svara på frågan "vad är sannolikheten p för X?" men saknar möjligheten att göra direkta experiment

Bayesiansk statistik: kan kvantitativt beskriva osäkerheten vi har över p

- Osäkerhet för p: beskrivs som en *prior*-fördelning
- Uppskatta *prior*-fördelningar: t ex genom systematiska frågor till experter. Priors kan vägas ihop med observationer.
- Anpassning: givet modell och observationer D kan *posterior*-fördelningen tas fram med Bayes sats:
  $P(p = x|D) = \frac{P(D|p=x)P(p=x)}{P(D)}$
- Problem: att använda en välgrundad prior (separata argument)

Möjlighet #3: generella argument om existentiell risk

Var finns data? T ex här.

Photography: Thierry Cohen

Exempel: Det stora filtret (Hanson, 1998)

Ett generellt argument om existentiell risk
... som svar på Fermis paradox

Den stora bilden:

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
Angry Birds: -5y
... Storleken på observerbara universum (Nature, 2012):

Den stora bilden:

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
Angry Birds: -5y
... Storleken på observerbara universum (Nature, 2012):
Antal stjärnor i vår galax: minst $300 \cdot 10^9$

Den stora bilden:

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
Angry Birds: -5y
... Storleken på observerbara universum (Nature, 2012):
Antal stjärnor i vår galax: minst $300 \cdot 10^9$
Antal planeter i vår galax: minst $100 \cdot 10^9$

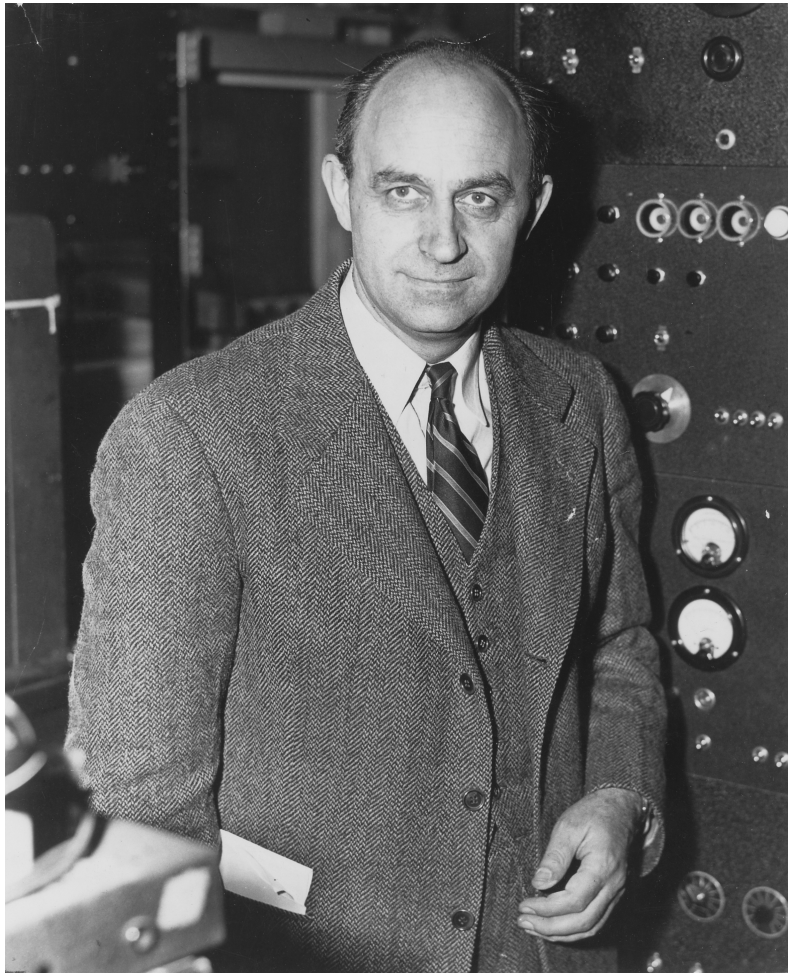
Den stora bilden:

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
Angry Birds: -5y
... Storleken på observerbara 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 $100 \cdot 10^9$

Den stora bilden:

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
Angry Birds: -5y
... Storleken på observerbara 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 det enorma antalet galaxer och planeter, 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.

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.

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).

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 några miljoner år.

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 några miljoner år.
5. Några miljoner år är långt för människan, men kort på den *kosmologiska* tidsskalan av miljarder år.

Fermis resonemang

Kosmologisk tids- och rumsskala:
13.8 miljarder år samt 10^{22} planeter, många äldre än jorden
Hypotetiskt scenario:

- Låt säga 10^{-4} för sannolikheten att intelligent liv skall uppstå
- Låt säga 10^{-4} för sannolikheten att intelligent liv når till rymdfärd
- Även med dessa mycket små chanser skulle det finnas mycket liv i vår galax, sedan lång tid tillbaka
- Det räcker med att en enda sådan civilisation skulle önska kolonisera galaxen för att ha gjort det för länge sedan

Fermis paradox: Vi saknar bekräftade tecken på utomjordiskt liv

Många möjliga förklaringar till Fermis tystnad

- De är ointresserade av oss
- De har varit här, men åkte igen
- De är osynliga för oss (“mörk materia/energi” 95% av universum)
- De vill stanna hemma
- ...
- Hanson (1998): Det stora filtret: det är väldigt svårt för liv att nå till teknologiska stadiet att det koloniserar galaxen

Hansons idé: det stora filtret

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.

Var är det stora filtret?
Filtret kan ligga antingen bakom eller framför oss.

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. Jupiter-effekten.

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. Jupiter-effekten.
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.

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. Jupiter-effekten.
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)

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. Jupiter-effekten.
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.

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. Jupiter-effekten.
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!

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. Jupiter-effekten.
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.

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. Jupiter-effekten.
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

Stora Filtret: kvantifiera tidigare/framtida existentiell risk

Betrakta varje planet i universum som ett experiment, med parametrarna
$p \in [0, 1]$
$q \in [0, 1]$

Stora Filtret: kvantifiera tidigare/framtida existentiell risk

Betrakta varje planet i universum som ett experiment, med parametrarna
$p \in [0, 1]$
$q \in [0, 1]$

$p = P(\textbf{intelligent liv till teknologisk nivå mänskligheten})$
$q = P(\textbf{avancerad teknologisk nivå} | \textbf{intelligent liv})$

Stora Filtret: kvantifiera tidigare/framtida existentiell risk

Betrakta varje planet i universum som ett experiment, med parametrarna
$p \in [0, 1]$
$q \in [0, 1]$

$p = P(\textbf{intelligent liv till teknologisk nivå mänskligheten})$
$q = P(\textbf{avancerad teknologisk nivå} | \textbf{intelligent liv})$

I universum: cirka $N = 10^{22}$ experiment

Förväntat antal teknologiska civilisationer

Antag oberoende mellan planeter, och förväntade antalet avancerade civilisationer skulle bli

$$Npq$$

Även med $p = q = 10^{-9}$ och $N = 10^{22}$ skulle vi förvänta oss många avancerade civilisationer så här långt i universum.

Fermis paradox: det verkar som att pq måste vara väldigt litet, vilket kan vara svårt att tro

Förväntat antal teknologiska civilisationer

Antag oberoende mellan planeter, och förväntade antalet avancerade civilisationer skulle bli

$$Npq$$

Även med $p = q = 10^{-9}$ och $N = 10^{22}$ skulle vi förvänta oss många avancerade civilisationer så här långt i universum.

Fermis paradox: det verkar som att pq måste vara väldigt litet, vilket kan vara svårt att tro

En Bayesiansk analys av p och q görs vidare under kursens gång.

Vad bör vi göra åt existentiella risker?

Detta kan inte vetenskapen i sig själv ge svar på

- Vetenskapen försöker beskriva världen som den är
- Faktapåståenden: påståenden om vilka omständigheter som är sanna i världen omkring oss
- Värdepåståenden: påståenden om vilka faktapåståenden som vore önskvärda (om de kunde realiseras) alternativt icke-önskvärda

Vad bör vi göra åt existentiella risker?

Detta kan inte vetenskapen i sig själv ge svar på

- Vetenskapen försöker beskriva världen som den är
- Faktapåståenden: påståenden om vilka omständigheter som är sanna i världen omkring oss
- Värdepåståenden: påståenden om vilka faktapåståenden som vore önskvärda (om de kunde realiseras) alternativt icke-önskvärda
- Faktapåståenden är ofta viktiga för att nå våra mål, men räcker inte till för att beskriva vad vi bör göra
- Vilka mål som är önskvärda: kräver delvis värdepåståenden

Denna distinktion motiveras ofta med Humes lag

Humes lag

Humes lag: Vi kan inte härleda ett **bör** ur ett **är**

- Ett faktapåstående är sant eller falskt
- Ett värdepåstående kan inte enkelt avgöras som sant eller falskt i objektiv mening

Humes lag

Exempel från blogosfären

- Ett herrelöst lok skenar fram mot en grupp om fem personer som befinner sig på rälsen och inte kan ta sig därifrån
- Dessa fem personer kommer att dödas av loket om inte någon ingriper
- Du har möjlighet att rädda dessa fem genom att slå om en växel så att loket styrs in på ett stickspår

Slutsats: Du bör slå om växeln?

Humes lag

Exempel från blogosfären

- Fakta: Ett herrelöst lok skenar fram mot en grupp om fem personer som befinner sig på rälsen och inte kan ta sig därifrån
- Fakta: Dessa fem personer kommer att dödas av loket om inte någon ingriper
- Fakta: Du har möjlighet att rädda dessa fem genom att slå om en växel så att loket styrs in på ett stickspår
- (Ett möjligt) Värde: Vi bör alltid undvika att döda

Slutsats: Du bör slå om växeln
När det gäller existentiell risk behöver vi även precisera våra värderingar för att väga kostnader nu mot risker i framtiden

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

1. Vad vore reduktion av $P(B)$ med 10^{-8} till $P'(B) = P(B) - 10^{-8}$ (en miljondel av en procent) värt?

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

1. Vad vore reduktion av $P(B)$ med 10^{-8} till $P'(B) = P(B) - 10^{-8}$ (en miljondel av en procent) värt?
2. Förväntat antal räddade liv
   $U(10^{16}) \cdot P(B) - U(10^{16}) \cdot P'(B) = U(10^{8})$

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

1. Vad vore reduktion av $P(B)$ med 10^{-8} till $P'(B) = P(B) - 10^{-8}$ (en miljondel av en procent) värt?
2. Förväntat antal räddade liv
   $U(10^{16}) \cdot P(B) - U(10^{16}) \cdot P'(B) = U(10^{8})$
3. Så bör vi släppa allt och bara minimera existentiell risk?

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

1. Vad vore reduktion av $P(B)$ med 10^{-8} till $P'(B) = P(B) - 10^{-8}$ (en miljondel av en procent) värt?
2. Förväntat antal räddade liv $U(10^{16}) \cdot P(B) - U(10^{16}) \cdot P'(B) = U(10^{8})$
3. Så bör vi släppa allt och bara minimera existentiell risk?

Under vilka förutsättningar är det meningsfullt att föra sådana resonemang? Kan det vara rätt? Om det är fel, *varför* är det eventuellt fel?

Vad bör vi göra åt existentiell risk?

Låt B vara en existentiell risk: vilken är den acceptabla nivån på $P(B)$? Antag att vi vill minimera förväntad förlust i fallet där $N = 10^{16}$ och att varje liv i framtida generationer räknas lika. Det senare är ett *värdepåstående*.

1. Vad vore reduktion av $P(B)$ med 10^{-8} till $P'(B) = P(B) - 10^{-8}$ (en miljondel av en procent) värt?
2. Förväntat antal räddade liv
   $U(10^{16}) \cdot P(B) - U(10^{16}) \cdot P'(B) = U(10^{8})$
3. Så bör vi släppa allt och bara minimera existentiell risk?

Under vilka förutsättningar är det meningsfullt att föra sådana resonemang? Kan det vara rätt? Om det är fel, *varför* är det eventuellt fel?

Föreläsning 4 tar upp väntevärdesmaximering
Föreläsning 8 tar upp värderingar och framtida generationer

Välgrundade scenarier kräver något mer

Det föregående exemplet antyder att väldigt små risker skulle behöva omfattas av våra hänsyn till framtiden. Faran i att tro det motiverar godtyckligt handlande och kan lätt leda fel.

Välgrundade scenarier kräver något mer

Det föregående exemplet antyder att väldigt små risker skulle behöva omfattas av våra hänsyn till framtiden. Faran i att tro det motiverar godtyckligt handlande och kan lätt leda fel.

Platon: att veta något är ett specialfall av vad vi kan tro
För att veta något X krävs att

1. Tro: vi tror på X
2. Sann: det är så att X
3. Berättigad: vi har goda skäl (argument) för vår tro om X

Välgrundade scenarier kräver något mer

Det föregående exemplet antyder att väldigt små risker skulle behöva omfattas av våra hänsyn till framtiden. Faran i att tro det motiverar godtyckligt handlande och kan lätt leda fel.

Platon: att veta något är ett specialfall av vad vi kan tro
För att veta något X krävs att

1. Tro: vi tror på X
2. Sann: det är så att X
3. Berättigad: vi har goda skäl (argument) för vår tro om X

Ett klart problem för det Bayesianska ramverket, är om vi inte kräver berättigad tro. En subjektiv tro behöver också vara berättigad.
Bayesianismens akilleshäll: kritiken om en godtyckligt vald prior.

Sammanfattning

- Existentiella risker: både undergång men även permanent reduktion av potential
- De flesta sådana risker tros komma från mänskliga handlingar snarare än naturliga
- Hypotes i den här kursen: många risker kommer från avancerade teknologier
- Ett generellt argument för existentiell risk: det Stora Filtret
- Bayesiansk statistik kan användas för att resonera om osäkerhet, men avgörande är att ha en berättigad prior
- Mer om beslutsteori, värderingar, specifika teknologier i vidare föreläsningar

Vidareläsning

1. Bostrom (2013): Existential Risk Prevention as Global Priority `http://www.existential-risk.org/concept.pdf`
2. Hanson (1998): The Great Filter - Are We Almost Past It? `http://mason.gmu.edu/~rhanson/greatfilter.html`
3. Häggström (2015): Here Be Dragons: Science, Technology and the Future of Humanity
4. Stephen Webb (2002): Where is Everybody?
5. Parfit (1984): Reasons and Persons
6. MSB: `https://www.msb.se/Upload/Nyheter_press/MSB_18arsundersokning_2014.pdf`