TMA4240 Statistikk H2004:
Oppsummering kontinuerlige sannsynlighetsfordelinger

Kontinuerlig uniform fordeling

$$f(x) = \frac{1}{B-A}, \qquad A \leq x \leq B.$$

Fenomén

En kontinuerlig størrelse (vekt, lengde, tid), som aldri kan bli mindre enn en gitt verdi A og aldri større enn en annen gitt verdi B. Eventuelt, en størrelse som i praksis ligger langs en sirkel og nullpunktet er valgt fritt. (F.eks. vinkel/orientering.) Se eksemplene.

Sannsynligheten

er den samme for alle like store intervaller innenfor A og B.

Forventningsverdi og varians:

$$\mu = \mathrm{E}(X) = \frac{A+B}{2},$$
$$\sigma^2 = \mathrm{Var}(X) = \frac{(B-A)^2}{12}.$$

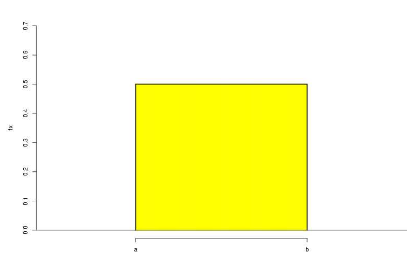
Eksempel

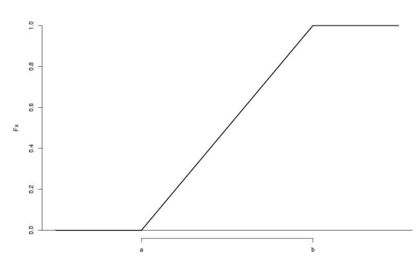
- De vanligste tilfeldig-tall generatorene på en datamaskin gir denne fordelingen med $A = 0$ og $B = 1$.
- Ta et gummistrikk med lengde L og merk av et nullpunkt og en skala. Stram strikken i alle retninger til den ryker ett sted. Punktet den ryker er kontinuerlig uniformt fordelt mellom $A = 0$ og $B = L$.
- Fra "the Cartoon Guide"; en hund som snurrer rundt en akse og stopper i en vilkårlig retning.

Kommentar

Fordelingen kan sees på som et grensetilfelle av diskret uniform fordeling hvis punktene $x_1, x_2, \ldots, x_k$ er jevnt fordelt og $k \to \infty$.

Plott av fordeling og kumulativ fordeling:





Normalfordelingen *

$$f(x) = n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty.$$

Fenomén

I teorien en kontinuerlig størrelse som kan ta alle mulige reelle verdier. I praksis brukes ofte fordelingen selv om det antakelig finnes en øvre og/eller nedre grense. Målte størrelser som er avhengig av et stort antall små faktorer blir gjerne tilnærmet normalfordelt.

Sannsynlighetstettheten

er symmetrisk om forventningsverdien, og har makspunkt (mode/ekstremalpunkt), median og forventningsverdi i samme punkt. Det er ingen "brå" endringer i sannsynlighet noe sted, to vendepunkter i hhv. $\mu - \sigma$ og $\mu + \sigma$.

Forventningsverdi og varians

$$\mu = \mathrm{E}(X) = \mu,$$
$$\sigma^2 = \mathrm{Var}(X) = \sigma^2.$$

(Valget av bokstaver på parameterne i normalfordelingen er gjort i samsvar med hva som er forventningsverdi og varians.)

Eksempel

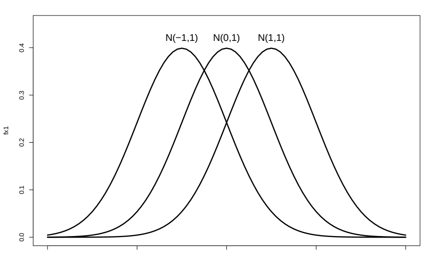
Om størrelser i naturen faktisk **er** normalfordelte, er vel en filosofisk debatt. Poenget er at normalfordelingen har en del gode teoretiske egenskaper som gjør den godt egnet til å tilnærme alle typer målinger hvor feilen like ofte gir for lavt som for høyt resultat. Eksempler på slike målesituasjoner vil du se mange av videre i kurset og seinere.

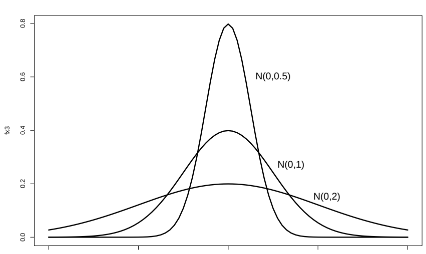
Kommentar

Kanskje den vanligste og viktigste kontinuerlige fordelingen. For å arbeide praktisk med normalfordelingen er det utviklet egne "normal-papir" for å plotte og kontrollere data mot den teoretiske fordelingen. (Kapittel 8.3) Fra sentralgrenseteoremet vet vi at et gjennomsnitt av mange uavhengige data blir tilnærmet normalfordelt (litt upresist formulert), spesielt er sum (og gjennomsnitt) av normalfordelte variabler alltid normalfordelt.

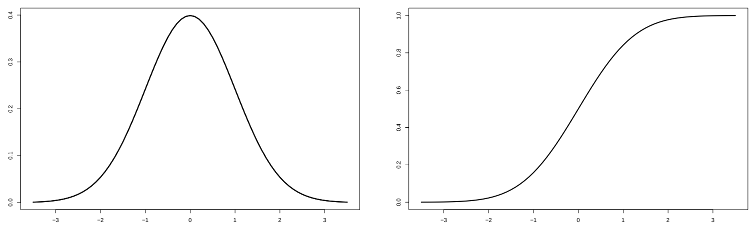
Husk transformasjonsformelen! $Z = \frac{X-\mu}{\sigma}$ gjør oss i stand til å bruke **én** tabell for **alle** normalfordelinger.

Sannsynlighetstetthet $f(x)$ for ulike valg av μ og σ





Standard normalfordelingen ($\mu = 0, \sigma = 1$), $f(x)$ og $F(x)$



Lognormalfordelingen

Denne er ikke pensum, men brukes i en del fagområder. Formel, forventningsverdi og varians finnes i tabellen og i kapittel 6.9 i læreboka. Poenget er at vi kan ha en størrelse i naturen med følgende egenskaper;

- Alltid positiv,
- Har en forventningsverdi som er "mest vanlig",
- Kan godt (i hvert fall teoretisk) bli ekstremt stor (flere størrelsesordener),
- Kan også bli veldig nær null (men aldri akkurat null - teoretisk),
- Ikke symmetrisk om gjennomsnitt/forventningsverdi,
- Gjerne påvirket av en rekke faktorer i naturen - som normalfordelte størrelser.

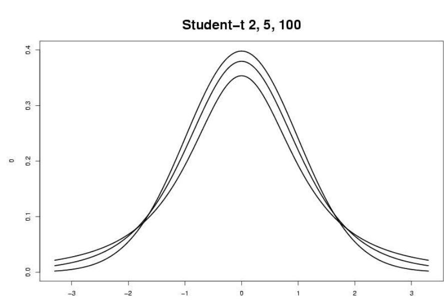
Et eksempel kan være bølgehøyde på åpent hav. I stedet for å bruke log-normalfordelingen kan vi ta logaritmen av alle målingene, og bruke normalfordelingen. Dvs. hvis vi antar bølgehøyde er log-normalfordelt, og du har målt to bølger på hhv. 3,2 meter og 7,9 meter, bruker du verdiene 0,51 og 0,90 og normalfordeling.

Student t-fordeling

$$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

Dette er en fordeling vi skal bruke, men bare med tabellverdier. Formelen for fordelingen trenger du ikke kjenne. Vi bruker t-fordelingen i statistikk når vi har målinger som (teoretisk) er normalfordelte, og vi ikke kjenner hverken variansen σ^2 eller forventningsverdien.

Fordelingen har forventningsverdi null, og er breiere (har tyngre "haler") enn normalfordelingen. Den er oppkonstruert og teoretisk, derfor har vi ikke eksempler på prosesser som faktisk har t-fordeling.



Eksponensialfordelingen

$$f(x) = \frac{1}{\beta} e^{-x/\beta} = \lambda e^{-\lambda x}, \qquad 0 \le x < \infty.$$

Fenomén

Nesten alltid tid mellom etterfølgende hendelser i en Poisson-prosess.

Sannsynligheten

for en hendelse i løpet av en viss tid framover i tid er den samme, uansett hvor lenge vi har ventet allerede. Sannsynlighetstettheten er avtagende overalt.

Forventningsverdi og varians

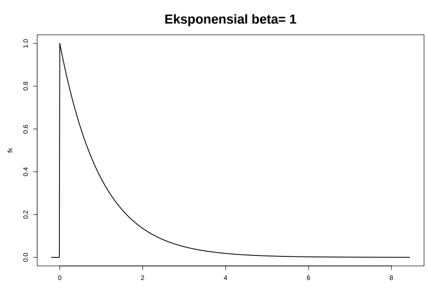
$$\mu = \mathrm{E}(X) = \beta = \frac{1}{\lambda},$$
$$\sigma^2 = \mathrm{Var}(X) = \beta^2 = \frac{1}{\lambda^2}.$$

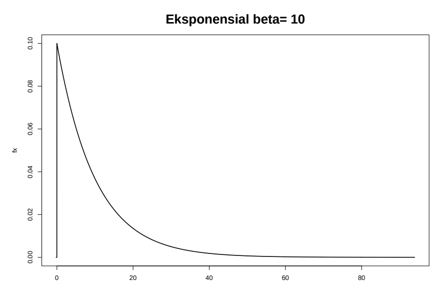
Eksempel

Levetider til elektriske komponenter, ventetider, avstand i tid mellom to jordskjelv eller mellom andre ulykker osv. modelleres gjerne med eksponensialfordelingen. Kravet er at hendelsen vi venter på kan inntreffe når som helst, dvs. uavhengig av aldring, slitasje o.l.

Kommentar

Det varierer om en bruker β eller λ som parameter, avhengig av situasjon og vane. Merk at β er forventet antall hendelser i løpet av en tidsenhet, mens λ er forventet rate eller hyppighet til hendelsene.





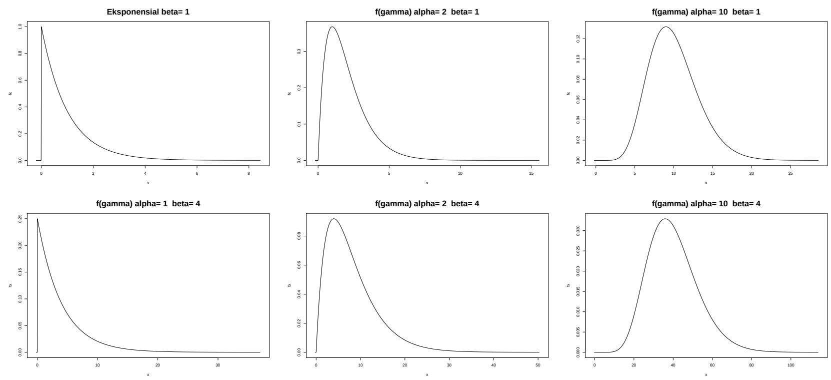
Gammafordelingen

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \qquad x \ge 0.$$

Hvis antakelsene til eksponensialfordelingen ikke er oppfylt, vil en gjerne prøve en fordeling som har omtrent samme egenskap. Her kommer gammafordelingen inn. I dette kurset vil du neppe bruke gammafordelingen spesielt ofte, så vi henviser til tabellen for formel, forventningsverdi og varians. Levetiden til mennesker eller andre/annet som utsettes for aldring, kan modelleres som en gammafordeling.

Summen av eksponensialfordelte variable er gammafordelt. Altså er tiden fram til hendelse nummer k gammafordelt. Eksponensialfordelingen er et spesialtilfelle av gammafordelingen med $\alpha = 1$ i formelen. (Se tabeller og formler.)

- $\alpha = 1$ gir eksponensialfordeling.
- $\alpha = \nu/2$ og $\beta = 2$ gir kji-kvadratfordeling.
- Hvis α er et heltall, kalles fordelingen Erlang-fordelingen (parameter $\alpha = k$).



χ^2*-fordelingen / kji-kvadratfordelingen*

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \qquad x \geq 0, \nu = 1, 2, 3, \ldots$$

Denne fordelingen er vel helst oppkonstruert for å kunne løse statistiske problemer, ikke fordi størrelser i naturen har denne fordelingen. Formelen behøver du ikke lære, men det blir viktig å huske at den er avhengig av en parameter ν (uttales [ny] - gresk n), som er et heltall. Vi sier at ν er antall *frihetsgrader*. Som sagt tidligere, er fordelingen et spesialtilfelle av gammafordelingen.

Sannsynlighetstettheten

er definert for alle $x > 0$ og har ett toppunkt.

Forventningsverdi og varians

$$\mu = \mathrm{E}(X) = \nu,$$
$$\sigma^2 = \mathrm{Var}(X) = 2\nu.$$