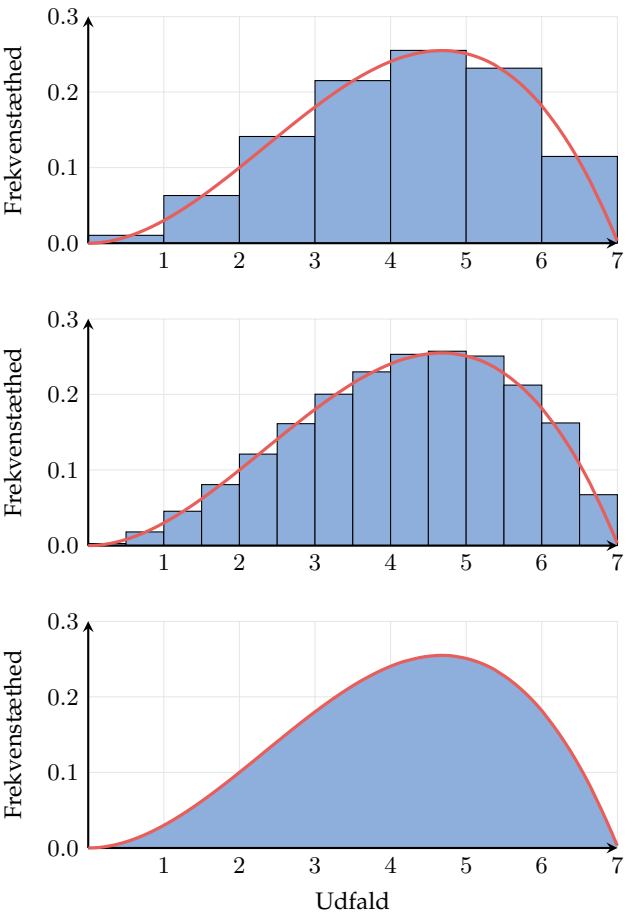


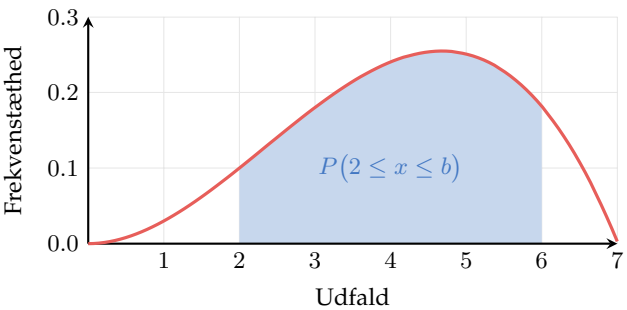
Sandsynlighedstæthed

Ligesom for statistiske fordelinger kan man ligeledes opstille sandsynlighedsfordelinger. Forskellen mellem de to er at en statistisk fordeling er baseret på et datasæt af observerede hændelser hvorimod en sandsynlighedsfordeling er en antagelse omkring den underliggende tilfældighed af udfaldene. Såfremt antagelsen om sandsynlighedsfordelingen er sand vil en statistisk fordeling over udfaldene konvergere mod sandsynlighedsfordelingen når antallet af observationer går imod uendeligt. Denne teori om grænseværdien af et statistisk datasæt kaldes *de store tals lov*, som trods et lidt misvisende navn understreger, at der i statistik er brug for mange tal for at kunne kvantisere sandsynligheden af det underliggende stokastiske forsøg. Et mere retvisende navn ville altså være *de mange tals lov*.

Betragtes således tilfældet hvor udfaldsmængden og derved observationsmængden er kontinuær, hvor der anvendes en grupperet statistisk fordeling, vil antallet af grupperinger blive større når antallet af observationer bliver større. Dette vil sige at værdiintervallet for grupperingen går mod nul når antallet af observationer går mod uendelig. På figuren nedenfor er situationen illustreret hvor det ses at jo når værdiintervallet bliver uendelig småt kan frekvenstætheden beskrives ved en funktion hvorved frekvensen udgør arealet under funktionskurven.



Funktionskurven f – som kan betragtes som en slags grænseværdi for den observerede frekvenstæthed – beskriver altså *sandsynlighedstætheden*. For at bestemme en sandsynlighed for hændelsen, at en stokastisk variabel x har en værdi i et givet interval $[a, b]$ skal den kumulerede sandsynlighedstæthed beregnes, svarende til det blå areal vist i figuren nedenfor.



Den kumulerede sandsynlighedstæthed beregnes ved det bestemte integrale af tæthedsfunktionen:

$$P(x \in [a, b]) = \int_a^b f(x) \cdot dx$$

Her skal bemærkes, at idet sandsynlighedstætheden essentielt repræsenterer frekvenstætheden skal der for integralet af denne gælde at:

$$\int_{-\infty}^{\infty} f(x) \cdot dx = 1$$

Sandsynlighederne for forskellige hændelser kan også beregnes hvis man kender stamfunktionen F til tæthedsfunktionen, også kaldet *fordelingsfunktionen*:

$$F(x) = \int f(x) \cdot dx$$

Ligesom sandsynlighedstætheden er grænseværdien af frekvenstætheden i statistik er fordelingsfunktionen grænseværdien til sumkurven og repræsenterer altså den kumulerede sandsynlighed. Dette gør, at sandsynlighed for forskellige hændelser, at en stokastisk variabel x har en værdi i et givet interval kan beregnes ud fra fordelingsfunktionen ved:

$$\begin{aligned} P(x \leq a) &= F(a) \\ P(a \leq x \leq b) &= F(b) - F(a) \\ P(x \geq a) &= 1 - F(a) \end{aligned}$$