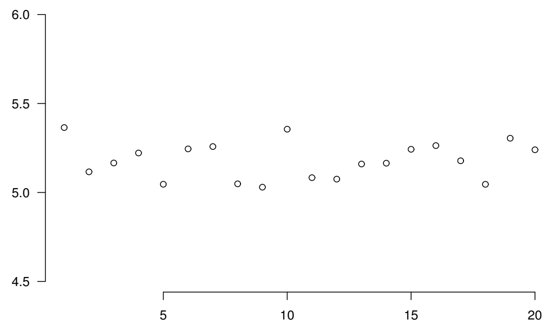
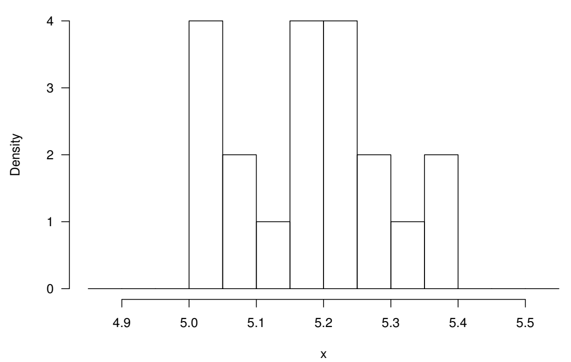
Högskolan i Halmstad

Kontinuerliga variabler

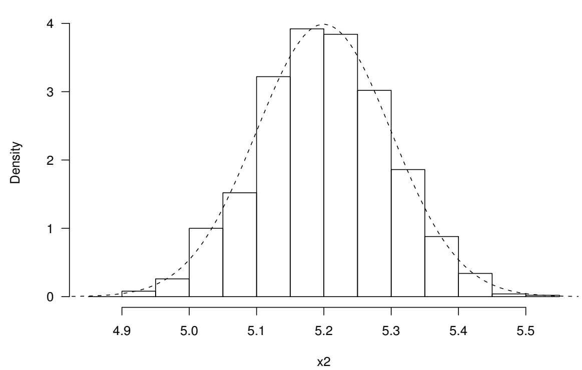
Antag att vi kunde mäta med oändligt stor noggrannhet hur stor strömstyrka en viss typ av motstånd klarar. Inga mätningar skulle då vara *exakt* lika. Om vi mätte 20 motstånd och plottade resultaten mot respektive observationsnummer skulle det kunna se ut någonting i den här stilen:



Om man indelar utfallsrummet i delintervall och räknar relativa frekvenser inom respektive intervall skulle man få något liknande följande histogram.



Antag nu att vi mäter ytterligare 980 motstånd och bildar motsvarande histogram. Då ser man hur staplarna följer en mer jämn kurva.

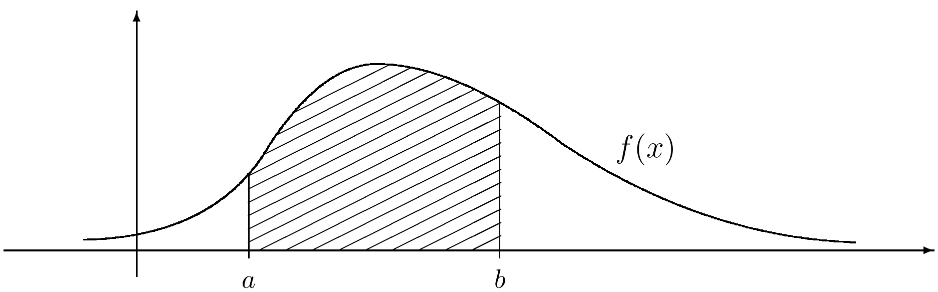


Denna jämna kurva (den streckade kurvan i figuren) kan sägas vara “början och slutet”: “slutet” för den är gränsvärdeshistogrammet med allt finare intervallindelning, "början" för att den teoretiskt beskriver sannolikhetsmassan (täthetsfunktionen) då man ska börja göra observationerna.

Vid diskreta variabler kunde staplarna i histogrammet svarat mot sannolikheten att få värdet stapeln representerar. Vid kontinuerliga variabler representer kurvan som beskriver täthetsfunktionen för variabeln X sannolikheten på så sätt att *arean* av ytan under kurvan på intervallet (x_1, x_2) är sannolikehten att $P(x_1 \leq X \leq x_2)$.

För *kontinuerliga stokastiska variabler* (dvs variabler som antar värden ur ett överuppräkneligt värderum) har man, precis som för diskreta, en **fördelningsfunktion** $F(x)$ definierad av $F(x) = P(X \leq x)$. Emellertid kan man inte tala om någon sannolikhetsfunktion. Motsvarigheten är istället **täthetsfunktionen** (kallad *frekvensfunktion* i Vännman). Täthetsfunktionen $f(x)$ har dock inte tolkningen som en sannolikhet (som är fallet med sannolikhetsfunktionen) utan är bara definerad av $f(x) = \frac{d}{dx}P(X \leq x)$ (den streckade kurvan ovan).

När det gäller stokastiska variabler allmänhet och kontinuerliga variabler i synnerhet tjänar man ibland på att tänka på täthetsfunktionen (och därmed fördelningsfunktionen som integrerad bit) ritad i ett koordinatsystem. Förtjänsten vid resonemang om sannolikheter är ungefär densamma som att ta Venn-diagram till hjälp vid resonemang om mängder.



$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Definition 1 *Om $P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$ för alla x i värdemängden, så kallas X* **kontinuerlig stokastisk variabel** *och f dess* **täthetsfunktion** *(eller* **frekvensfunktion***).*

Eftersom P är ett sannolikhetsmått måste $0 \leq P(a \leq X \leq b) \leq 1$ för alla $a, b \in \mathbb{R}$ vilket innebär att $P(X \leq b) - P(X \leq a) \geq 0 \;\Rightarrow\; \int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt \geq 0$

$$\Rightarrow \int_a^b f(t)\,dt \geq 0 \text{ för alla } a \leq b \in \mathbb{R}$$

$$\Rightarrow \boxed{f(x) \geq 0 \text{ för alla } x \in \mathbb{R}}.$$

Vidare är $P(\Omega) = 1$ varmed $\lim_{b\to\infty} P(X \leq b) = 1$ så $\boxed{\int_{\mathbb{R}} f(x)\,dx = 1}$.

Eftersom $F(x)$ är fördelningsfunktion är $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$. Därmed är $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.

Till skillnad mot det diskreta fallet har $<$ och $\leq$ samma betydelse, och dito $>$

och $\geq$. Dessutom, till skillnad mot diskreta fallet, är $P(X = a) = \int_a^a f(x)\,dx = 0$. Detta understryker att $f(x) \neq P(X = x)$, dvs man har *inte* tolkningen av täthetsfunktionen $f(x)$ som en sannolikhet!

Detta sammanfattas i Sats 4A – läs den!

Vanliga kontinuerliga fördelningar

$\boxed{\text{Rektangulärfördelning}}$

Betecknas $X \in R(a, b)$ eller $X \in U(a, b)$.

$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{annars} \end{cases}$

Detta är den kontinuerliga motsvarigheten till (diskret) likformig fördelning (och kallas ibland istället just "likformig fördelning"). Rektangulärfördelningen fördelar sannolikhetsmassan jämnt över hela värdemängden (intervallet $[a, b]$). Denna fördelning är det man kan anta om man ej har någon information som indikerar att "något delintervall är mer sannolikt än ett annat". (Ex. Tid kvar till nästa buss då man anländer till en busshållplats.)

$\boxed{\text{Exponentialfördelning}}$

Betecknas $X \in Exp(\lambda)$.

$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$

X är t.ex. livslängden hos många elektroniska komponenter såsom glödlampor, transistorer etc.

Övning Beräkna fördelningsfunktionenerna för en rektangulärfördelad resp. en exponentialfördelad variabel.

$\boxed{\text{Normalfördelning}}$

Betecknas $X \in N(\mu, \sigma^2)$.

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

P.g.a. ett av de viktigaste resultaten i hela sannolikhetsläran (den s.k. *centrala gränsvärdessatsen* som vi kommer till så småningom) är normalfördelningen approximativ fördelning för en mängd naturliga fenomen. Normalfördelningen är den i särklass viktigaste fördelningen och därför kommer vi ägna den extra mycket uppmärksamhet.

Exempel Antag att vi anländer till en busshållplats där bussarna går var tjugonde minut. Vad är sannolikheten att bussen kommer inom 10 minuter men inte förr än efter 3 minuter?

Lösning: Eftersom vi inte vet någonting om när vi anlänt i förhållande till när bussarna kommer, endast att vi inte ska behöva vänta mer än 20 minuter, är det enda vi kan anta att sannolikheten för bussens ankomst till hållplatsen är jämnt fördelad, dvs att X (tiden vi behöver vänta på bussen) är fördelad $R(0, 20)$. Därmed är P(bussen kommer inom 10 min. men ej förr än om 3 min.) $= P(3 \leq X \leq 10) =$ $= \int_3^{10} \frac{dx}{20} = [\frac{x}{20}]_3^1 0 = \frac{10}{20} - \frac{3}{20} = \frac{7}{20} = 0.35.$ □

Betingning med kontinuerliga variabler

Om X och Y är kontinuerliga stokastiska variabler så är $P(X \leq x, Y \leq y) =$ $= \int_{-\infty}^{y} P(X \leq x \,|\, Y = t) f_Y(t)\, dt = \int_{-\infty}^{x} P(Y \leq y \,|\, X = t) f_X(t)\, dt.$

Exempel I ett klassrum finns 10 parallellkopplade oberoende lysrörslampor. För varje lampa är sannolikheten att den lyser mer är 1000 timmar 0.8. Vad är sannolikheten att det inte är helt mörkt i klassrummet då det ges en kvällskurs där efter 10 000 timmars brinntid?

Lösning: Låt X_i vara livslängden av den i:te lampan. Då är $P(X_i \leq x) = 1 - e^{-\lambda x}$. $0.8 = P$(lampa i lyser i mer än 1000 timmar) $= P(X_i > 1000) = 1 - (1 - e^{-\lambda 1000}) =$ $= e^{-1000\lambda} \Rightarrow \lambda = -\frac{\ln 0.8}{1000} = 0.000223\ldots$

Därmed är P(ej helt mörkt efter 10 000 timmar) =
= $1 - P$(alla lampor slut efter 10 000 timmar) =
$= 1 - P(\{X_1 \leq 10\,000\} \cap \{X_2 \leq 10\,000\} \cap \cdots \cap \{X_{10} \leq 10\,000\} =$
$= 1 - P(X_1 \leq 10\,000)^{10} = 1 - (1 - e^{-0.000223 \cdot 10\,000})^{10} = 1 - (1 - e^{2.23})^{10} = 0.68.$ □

Exponentialfördelningens glömska

Antag att vi vet att en glödlampa lyst i a timmar. Vad är då den betingade sannolikheten att den fortsätter lysa i ytterligare b timmar?

Låt X vara livsängden hos glödlampan i timmar. Då är $X \in Exp(\lambda)$, dvs $P(X \leq x) = F(x) = 1 - e^{-\lambda x}$ där $x \geq 0$ och $\lambda \in \mathbb{R}$ är en konstant specifik för denna sorts lampor. Vi vill veta vad sannolikheten är att lampan fungerar i $a + b$ timmar givet att den redan har lyst i a timmar.

$$
\begin{aligned}
P(X > a + b \,|\, X > a) &= \frac{P(\{X > a + b\} \cap \{X > a\})}{P(X > a)} \\
&= \frac{P(X > a + b)}{P(X > a)} \\
&= \frac{1 - P(X \leq a + b)}{1 - P(X \leq a)} \\
&= \frac{1 - F(a + b)}{1 - F(a)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1-(1-e^{-\lambda(a+b)})}{1-e^{-\lambda a}} \\
&= \frac{e^{-\lambda a}e^{-\lambda b}}{e^{-\lambda a}} \\
&= e^{-\lambda b} \\
&= 1-(1-e^{-\lambda b}) \\
&= 1-F(b) \\
&= P(X>b)
\end{aligned}
$$

Detta innebär att sannolikheten att den ska fortsätta lysa i b timmar då vi *vet* att den redan lyst i a timmar är densamma som sannolikheten att den lyser i b timmar utan vetskap om hur länge den lyst innan!! På detta vis menar man att lampan "glömt" att den levat i a timmar redan så det ej påverkar sannolikheten att den lever i ytterligare b. $\square$

(**Obs** Behöver ej läsa om Weibullfördelning!)

Definition 2 *Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion $f(x)$.* **Väntevärdet** *för X är $\mu = E(X) = \int_{\mathbb{R}} x f(x)\,dx$.* **Variansen** *för X är $\sigma^2 = V(X) = \int_{\mathbb{R}} (x-\mu)^2 f(x)\,dx$.*

Räknereglerna för väntevärde och varians av kontinuerliga är desamma som reglerna för diskreta variabler med undantaget att om X är kontinuerlig så är

$$E(g(X)) = \int_{\mathbb{R}} g(x)f(x)\,dx$$