# 02402 Statistik (Polyteknisk grundlag)

## Uge 3:
## Konkrete kontinuærte fordelinger og mere on stokastiske variable

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Dagsorden

1. Praktisk, Eksamen mm.
2. Konkrete kontinuærte fordelinger
   - Den Uniforme fordeling
   - Normalfordelingen
   - Lognormalfordelingen
   - Eksponentialfordelingen
   - Kontinuærte fordelinger i Python
3. Multivariate stokastiske variable
4. Regneregler for stokastiske variable
5. Særlige fordelinger vi skal bruge senere i kurset
   - $\chi^2$-fordeling
   - t-fordeling
   - F-fordeling

# 02402 Statistik (Polyteknisk grundlag)

# Praktisk, Eksamen mm.

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Praktisk information

- TestQuiz på hjemmesiden er ikke opdaterede
- Bogen er i trykken og kan snart købes i boghandleren
- Der kommer en ekstra TA (Sarah) i 306 foyerområde ved trappen

# Brug af Python til øvelser og eksamen

- Fremover vil vi forsøge at lave en tydelig angivelse af brug af Python til øvelser
- Der kommer et test-eksamenssæt lidt senere på semesteret

# Øvelser fra Uge 2

1.3 Tast data in i Python og brug Python til alle udregninger.

2.2 Opgaven kan laves uden Python. Prøv gerne at løse med lommeregner først og brug så Python til at finde samme resultat.

2.4 Opgaven kan laves uden Python. Prøv gerne at løse i hånden (evt med lommeregner) først og brug så Python til at finde samme resultat.

2.6 Opgaven kan laves uden Python. Prøv gerne at løse med lommeregner først og brug så Python til at finde samme resultat.

2.7 Lav opgaven **med Python**

2.8 Lav opgaven **med Python**

2.1 Opgaven indeholder Python kode, men kan besvares uden yderligere brug af Python (brug af lommeregner er ok)

1.4 Projekt start - brug gerne Python!

# Eksamensopgave med diskret fordeling

Eksamen fra Maj 2021, spørgsmål 19 og 20

https://02402.compute.dtu.dk/previousexams/2021may_02402_da.pdf

# Eksamensopgave

**Spørgsmål VII.2 (20)**

Hvad er sandsynligheden for at observere mindst tre ankomster i en tilfældigt valgt periode på 1 minut?

1 ☐ **stats.poisson.pmf(k=3, mu=3) = 0.2240**

2 ☐ **1 - stats.poisson.pmf(k=2, mu=1) = 0.8161**

3 ☐ **1 - stats.poisson.cdf(k=2, mu=3) = 0.5768**

4 ☐ **stats.poisson.cdf(k=3, mu=3) = 0.6472**

5 ☐ **1 - stats.poisson.cdf(k=1, mu=3) = 0.8009**

02402 Statistik (Polyteknisk grundlag)

# Konkrete kontinuærte fordelinger

- Den Uniforme fordeling
- Normalfordelingen
- Lognormalfordelingen
- Eksponentialfordelingen
- Kontinuærte fordelinger i Python

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# 02402 Statistik (Polyteknisk grundlag)

- Den Uniforme fordeling
- Normalfordelingen
- Lognormalfordelingen
- Eksponentialfordelingen
- Kontinuærte fordelinger i Python

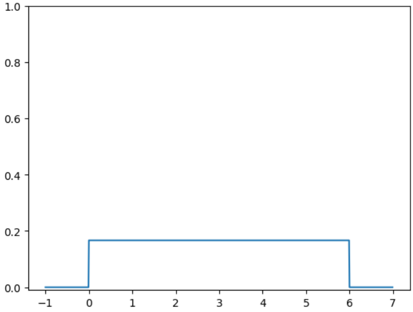
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Den Uniforme fordeling

Den Uniforme fordeling bruges til at beskrive en kontinuært stokastisk variabel $X$ som kun antager værdier i intervallet: $[\alpha;\beta]$. Fordelingen er symmestrisk og "flad".

Alle værdier i intervallet er "*lige sandsynlige*"

$$X \sim U(\alpha,\beta)$$




Kahoot!
(x1)

# Den Uniforme fordelings tæthedsfunktion

**Definition 2.35 Uniform distribution**

Let $X$ be a uniform distributed random variable

$$X \sim U(\alpha, \beta), \tag{2-46}$$

where $\alpha$ and $\beta$ defines the range of possible outcomes. It has the *pdf*

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}. \tag{2-47}$$

The uniform *cdf* is

$$F(x) = \begin{cases} 0 & \text{for } x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{for } x \in [\alpha, \beta) \\ 1 & \text{for } x \geq \beta \end{cases}. \tag{2-48}$$

# Den Uniforme fordelings middelværdi og varians

**Theorem 2.36 Mean and variance of the uniform distribution**

The mean of a uniform distributed random variable $X$ is

$$\mu = \frac{1}{2}(\alpha + \beta), \tag{2-49}$$

and the variance is

$$\sigma^2 = \frac{1}{12}(\beta - \alpha)^2. \tag{2-50}$$

# 02402 Statistik (Polyteknisk grundlag)



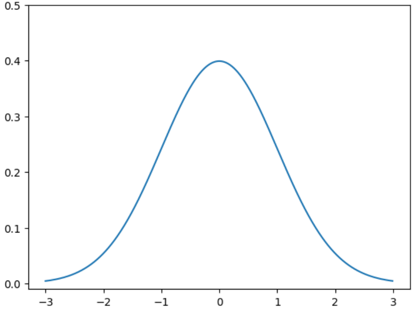

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Normalfordelingen

Normalfordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$ som følger en karakteristisk klokkeformet fordeling. Fordelingen er symmestrisk omkring gennemsnittet $\mu$ og har spredning $\sigma$.

*Værdier tæt på gennemsnittet er mest sandsynlige*

$$X \sim N(\mu, \sigma^2)$$



# Normalfordelingens tæthedsfunktion

## Definition 2.37 Normal distribution

Let $X$ be a normal distributed random variable

$$X \sim N(\mu, \sigma^2), \tag{2-51}$$

where $\mu$ is the mean and $\sigma^2$ is the variance (remember that the standard deviation is $\sigma$). Note that the two parameters are actually the mean and variance of $X$.
It follows the normal *pdf*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2-52}$$

and the normal *cdf*

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\, du. \tag{2-53}$$

# Normalfordelingens middelværdi og varians

**Theorem 2.38 Mean and variance**

The mean of a Normal distributed random variable is

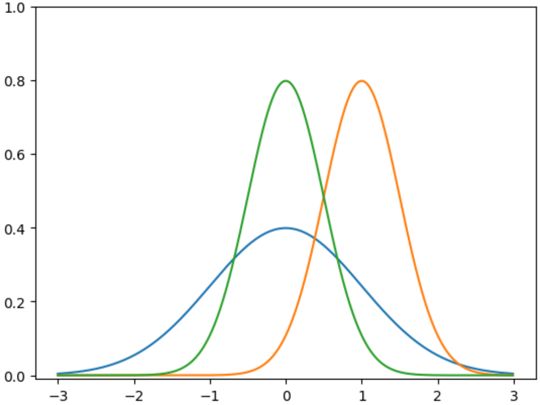
$$\mu, \tag{2-54}$$

and the variance is

$$\sigma^2. \tag{2-55}$$

Hence simply the two parameters defining the distribution.

# Eksempler på Normalfordelinger

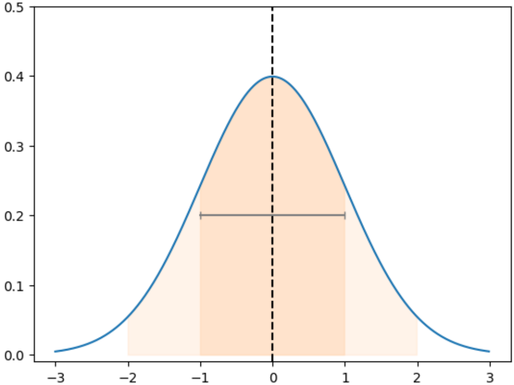



Hvilke to fordelinger har forskellig middelværdi ($\mu$)?

Hvilke to fordelinger har forskellig spredning ($\sigma$)?

Kahoot!
(x2)

# Normalfordelingen: Et par tal der er værd at huske




Indenfor $\mu \pm \sigma$ ligger ca 68.3% af sandsynlighedsmassen
Indenfor $\mu \pm 2\sigma$ ligger ca 95.4% af sandsynlighedsmassen
95% af sandsynlighedsmassen ligger i intervallet $\mu \pm 1.96 \cdot \sigma$

# Eksempel med Normalfordelingen 1

Det antages, at folkeskolelæreres løn kan beskrives med en normalfordeling med middelværdi $\mu = 290$ (i 1000 DKK) og standardafvigelse $\sigma = 4$ (1000 DKK).

*a) Hvad er sandsynligheden for, at en tilfældigt udvalgt lærer tjener mere end 300.000 kr?*

*b) Specificér et løninterval (som er symmetrisk omkring middelværdien), som dækker 95% af lærernes lønninger*

Hvad skal repræsenteres som en stokastiske variabel $X$? (og hvad er fordelingen af $X$)

---

a) Hvilken sandsynlighed skal udregnes?

---

b) Hvilket interval skal udregnes? Kan vi også skitsere dette med en tegning af pdf (eller cdf)?

---

# Lineære kombinationer af normalfordelte stokastiske variable

**Theorem 2.40 Linear combinations of normal random variables**

Let $X_1, \ldots, X_n$ be independent normal random variables, then any linear combination of $X_1, \ldots, X_n$ will follow a normal distribution, with mean and variance given in Theorem 2.56.

Hvis både den stokastiske variabel $X$ og stokastiske variabel $Y$ følger en normalfordlinger (med hver sin middelværdi og varians), så følger lineære kombinationer af $X$ og $Y$ også en normalfordeling.

Eksempler:
$Z = X + Y$
$U = X - Y$
$W = 2X + 3Y$

(man kan også kombinere flere end 2)

# Standard Normalfordelingen

### Definition 2.42 Standard normal distribution

The standard normal distribution is the normal distribution with zero mean and unit variance

$$Z \sim N(0,1), \tag{2-61}$$

where $Z$ is the standardized normal random variable.

### Theorem 2.43 Transformation to the standardized normal random variable

A normal distributed random variable $X$ can be transformed into a standardized normal random variable by

$$Z = \frac{X - \mu}{\sigma}. \tag{2-62}$$

# Eksempel med Normalfordelingen 2

En given vægt har en målefejl (målt i gram), $Z$, som kan beskrives med en standardnormalfordeling, $Z \sim N(0, 1^2)$.

Dvs. at den gennemsnitlige målefejl er $\mu_Z = 0$ gram og standardafgivelsen er $\sigma_Z = 1$ gram

*a) Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram større end den sande vægt af produktet?*

*b) Hvad er sandsynligheden for, at vægten giver et resultat, som er mindst 2 gram mindre end den sande vægt af produktet?*

*c) Hvad er sandsynligheden for, at vægten har en afvigelse på højst $\pm 1$ gram?*

Hvilke sandsynligheder skal udregnes? Kan vi skitsere dem med en tegning af pdf (eller cdf)?

# 02402 Statistik (Polyteknisk grundlag)

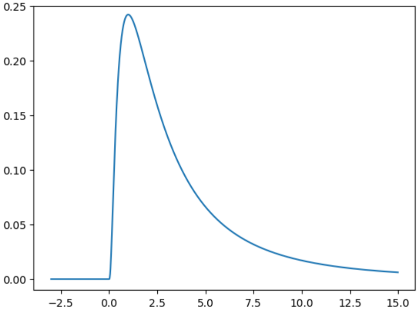


DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Lognormalfordelingen

Lognormalfordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$ som følger en karakteristisk skæv fordeling.

$$X \sim LN(\alpha, \beta^2)$$



# Transformation fra Lognormal til Normal

En lognormalfordelt stokastisk variabel $X \sim LN(\alpha, \beta^2)$ kan *transformeres* til en normalfordelt stokastisk variabel $Y$ ved:

$$Y = \ln(X).$$

Her er $Y$ normalfordelt med middelværdi $\alpha$ og varians $\beta^2$, dvs. $Y \sim N(\alpha, \beta^2)$.

Hvis **data** $\{x_i\}$ er Lognormaltfordelt, vil man ofte *transformere* data og beregne $y_i = \log(x_i)$.
De **transformerede data** $\{y_i\}$ vil da følge en normalfordeling.

# Lognormalfordelingens tæthedsfunktion

**Definition 2.46 Log-Normal distribution**

A log-normal distributed random variable

$$X \sim LN(\alpha, \beta^2), \tag{2-63}$$

where $\alpha$ is the mean and $\beta^2$ is the variance of the normal distribution obtained when taking the natural logarithm to $X$.
The log-normal *pdf* is

$$f(x) = \frac{1}{x\sqrt{2\pi}\beta} e^{-\frac{(\ln x - \alpha)^2}{2\beta^2}}. \tag{2-64}$$

# Lognormalfordelingens middelværdi og varians

**Theorem 2.47 Mean and variance of log-normal distribution**

Mean of the log-normal distribution

$$\mu = e^{\alpha+\beta^2/2}, \tag{2-65}$$

and variance

$$\sigma^2 = e^{2\alpha+\beta^2}(e^{\beta^2} - 1). \tag{2-66}$$

# 02402 Statistik (Polyteknisk grundlag)



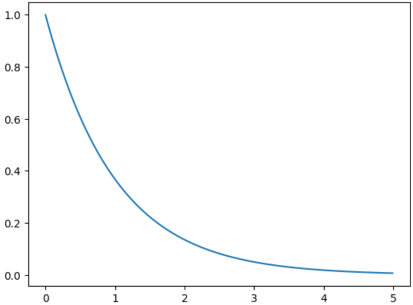
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Eksponentialfordelingen

Eksponentialfordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$ som følger en eksponentielt aftagende fordeling og bruges ofte til at beskrive *levetider og ventetider*.

Eksponentialfordelingen kan bruges til at beskrive (vente)tiden mellem events i en Poisson process.

$$X \sim Exp(\lambda)$$

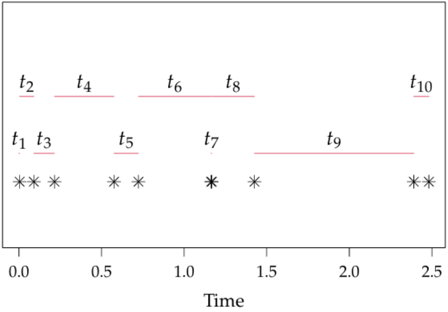



Kahoot!
(x3)

# Eksponentialfordeling og Poissonfordelingen

Eksponentialfordelingen relateret til Poissonfordelingen:

- Eksponentialfordelingen beskriver tiden mellem tilfældige events. Dette er en kontinuært størrelse
- Poissonfordelingen beskriver antallet af tilfældige events der sker indenfor et fast tidsrum. Dette er en diskret størrelse.




Nogle opgaver vil kunne løses med både eksponentialfordelingen of poissonfordelingen.

# Eksponentialfordelingens tæthedsfunktion

**Definition 2.48 Exponential distribution**

Let $X$ be an exponential distributed random variable

$$X \sim Exp(\lambda), \tag{2-67}$$

where $\lambda$ is the average rate of events.

It follows the exponential *pdf*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}. \tag{2-68}$$

# Eksponentialfordelingens middelværdi og varians

**Theorem 2.49 Mean and variance of exponential distribution**

Mean of an exponential distribution is

$$\mu = \frac{1}{\lambda}, \tag{2-69}$$

and the variance is

$$\sigma^2 = \frac{1}{\lambda^2}. \tag{2-70}$$

# Eksempel med Eksponentialfordelingen

Det antages, at der i gennemsnit bliver indlagt 0.3 patienter pr. dag på københavnske hospitaler som følge af luftforurening.

*En patient indlægges på hospitalet som følge af luftforurening. Hvad er sandsynligheden for at der går mere end 10 dage før der indlægges endnu en patient som følge af luftforurening?*

Hvad skal repræsenteres af den stokastiske variabel $X$?

---

Hvad er fordelingen af $X$ (og hvad er parametrene)?

---

Hvilken sandsynlighed skal udregnes?

---

# 02402 Statistik (Polyteknisk grundlag)

- Den Uniforme fordeling
- Normalfordelingen
- Lognormalfordelingen
- Eksponentialfordelingen
- Kontinuærte fordelinger i Python

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Kontinuærte fordelinger i Scipy.stats

```
Scipy.stats.uniform
Scipy.stats.normal
Scipy.stats.lognormal
Scipy.stats.expon
```

Generelle 'methods' for kontinuærte fordelinger:

| | |
|---|---|
| `.rvs` | 'random variates' (simulér tilfældige tal) |
| `.pdf` | 'probability mass function' (pdf/tæthedsfunktion) |
| `.cdf` | 'cumulative distribution function' (fordelingsfunktion) |
| `.ppf` | 'percent point function' (invers cdf / fraktilfunktion) |
| `.mean /.var /.std` | 'mean'/'variance'/'standard deviation' |

Se også Appendix A.2.1 i bogen.

# Eksempel med Normalfordelingen i Python

Lad os simulere den stokastiske variabel $X$, derfølger en Normalfordeling med middelværdi 5 og standardafvigele 2:

$$X \sim N(5, 2^2)$$

I Python bruger vi Scipy.stats' `norm.rvs` function:

```
stats.norm.rvs(loc=5, scale=2)
```

- Gå til Python notebook "simulation_normal.ipynb" i VS Code

Visual Studio Code

Kahoot!
(x1)

# 02402 Statistik (Polyteknisk grundlag)

# Multivariate stokastiske variable

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Joint pdf

**Definition 2.65 Joint *pdf* of two-dimensional discrete random variables**

The *pdf* of a two-dimensional discrete random variable $[X, Y]$ is

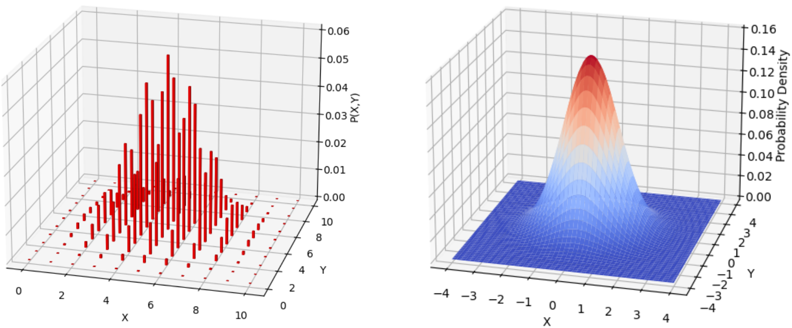
$$f(x, y) = P(X = x, Y = y), \tag{2-79}$$

with the properties

$$f(x, y) \geq 0 \text{ for all } (x, y), \tag{2-80}$$

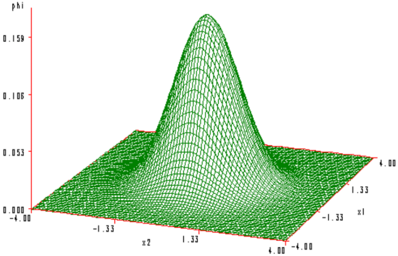
$$\sum_{\text{all } x} \sum_{\text{all } y} f(x, y) = 1. \tag{2-81}$$

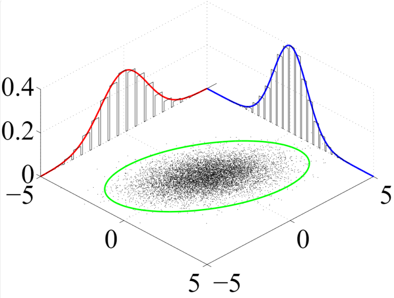
# Joint pdf



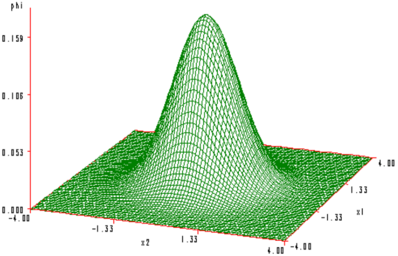

Eksempler på to-dimensionel pdf: diskret (venstre) og kontinuær (højre)

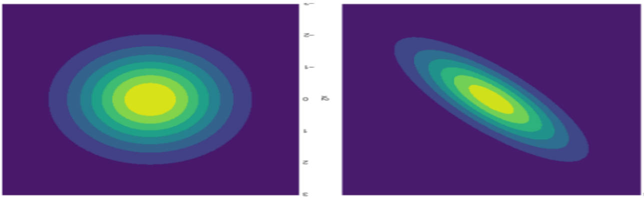
# Ekstra: Flerdimensionelle stokastiske variable



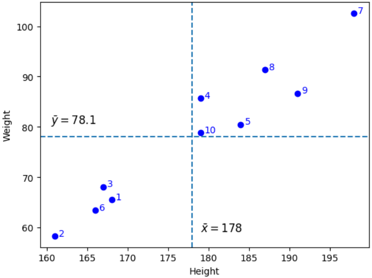



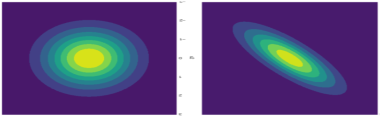
# Ekstra: Flerdimensionelle stokastiske variable



# Ekstra: Flerdimensionelle stokastiske variable



Stikprøvekovarians:

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Kovarians:

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])]$$

# covariance

**Definition 2.58 Covariance**

Let $X$ and $Y$ be two random variables, then the covariance between $X$ and $Y$, is

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] \,. \tag{2-75}$$

# correlation

### Definition 2.62 Correlation

Let $X$ and $Y$ be two random variables with $\mathrm{V}(X) = \sigma_x^2$, $\mathrm{V}(Y) = \sigma_y^2$, and $\mathrm{Cov}(X, Y) = \sigma_{xy}$, then the correlation between $X$ and $Y$ is

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{2-78}$$

# Regneregler for stokastiske variable

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# Beregninger hvor der indgår stokastiske variable

Når man udfører beregninger hvor der indgår en stokastisk variabel, får man en ny stokastisk variabel.

Eksempler:

Lad $X$ være en stokastisk variabel.

Hvis $Y = 2X + 10$, så er $Y$ også en stokastisk variabel.

Hvis $Z = 3X^2$, så er $Z$ også en stokastisk variabel.

Hvis $W = 5\cos(X)$, så er $W$ også en stokastisk variabel.

# Regneregler for stokastiske variable 1

Regneregel for lineære funktioner $Y = aX + b$:
(Dette gælder kun for det simpleste eksempel fra forrige slide: $Y = 2X + 10$)

**Theorem 2.54 Mean and variance of linear functions**

Let $Y = aX + b$ then

$$\mathrm{E}(Y) = \mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b, \tag{2-71}$$

and

$$\mathrm{V}(Y) = \mathrm{V}(aX + b) = a^2\,\mathrm{V}(X). \tag{2-72}$$

Disse regneregler gælder både for kontinuerte og diskrete stokastiske variable.

# Eksempel med $Y = aX + b$ (2.54)

Temperaturen i et kølerum, må helst ikke svinge for meget og specifikt ønskes det at standardafvigelsen er maks $1^\circ$C (Celcius).
Det oplyses at temperaturen (målt hver time) i gennemsnit er $41^\circ$F (Fahrenheit) og har en standardafvigelse på $1.5^\circ$F.

Fahrenheit kan omregnes til celcius med følgende formel: $C = \frac{F-32}{1.8}$

Lad $F$ være en stokastisk variabel der angiver temperaturen målt i $^\circ$F.
Hvad er $\mu_F = E[F]$ og $\sigma_F^2 = V[F]$?

---

Lad $C$ være en stokastisk variabel der angiver temperaturen målt i $^\circ$C.
Hvad er $\mu_C = E[C]$ og $\sigma_C^2 = V[C]$?

---

Er standardafvigelsen af temperaturen mindre end det ønskede $1^\circ$C?

---

# Beregninger hvor der indgår flere stokastiske variable

Vi kan også forestille os beregninger hvor der indgår flere stokastiske variable:

Eksempler:

Lad X = én stokastisk variabel og Y = en anden stokastisk variabel.

Hvis $Z = 2X - Y$, så er Z også en stokastisk variabel.

Hvis $W = 5X/Y$, så er W også en stokastisk variabel.

Hvis $U = X \cdot (Y - 3)^2$, så er U også en stokastisk variabel.

# Regneregler for stokastiske variable 2

Regneregler for lineære kombinationer $Z = aX + bY + \ldots$:
(Dette gælder kun for det simpleste eksempel fra forrige slide: $Z = 2X - Y$):

**Theorem 2.56 Mean and variance of linear combinations**

The mean of a linear combination of independent random variables is

$$\mathrm{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1\,\mathrm{E}(X_1) + a_2\,\mathrm{E}(X_2) + \cdots + a_n\,\mathrm{E}(X_n), \tag{2-73}$$

and the variance

$$\mathrm{V}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2\,\mathrm{V}(X_1) + a_2^2\,\mathrm{V}(X_2) + \cdots + a_n^2\,\mathrm{V}(X_n). \tag{2-74}$$

Disse regneregler gælder både for kontinuerte og diskrete stokastiske variable.

# Eksempel med $Y = a_1X_1 + a_2X_2 + a_3X_3 + \ldots$ (2.56)

Planlægning for flyselskab:
Et fly, der kan tage 55 passagerer, må max. lastes med 4000 kg (kun passagernes vægt betragtes her som last).

Den individuelle vægt af passagerer på en flytur, $X$, antages at være normalfordelt så $X \sim N(70, 10^2)$.

Beregn sandsynligheden for, at flyet bliver overbelastet.

Lad $Y$ være en stokastisk variabel, der beskriver den totale vægt af 55 passagerer. Opskriv et udtryk for $Y$. Hvad er $E[Y]$ og $V[Y]$?

---

Hvilken fordeling følger $Y$?

---

Hvad er sandsynligheden for at flyet overbelastes? (Brug Python til dette spørgsmål)

---

# Særlige fordelinger vi skal bruge senere i kurset

- $\chi^2$-fordeling
- t-fordeling
- F-fordeling

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# 02402 Statistik (Polyteknisk grundlag)

- $\chi^2$-fordeling
- t-fordeling
- F-fordeling

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# $\chi^2$-fordelingen

$\chi^2$-fordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$, der er lig med *summen af $\nu$ kvadrerede uafhængige standardnormalfordelte variable*:

$$X \sim \chi^2(\nu) \qquad \text{hvor } X = \sum_{i=1}^{\nu} Z_i^2 \text{ og } Z_i \sim N(0, 1^2)$$

**Theorem 2.79**

Let $Z_1, \ldots, Z_\nu$ be independent random variables following the standard normal distribution, then

$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi^2(\nu). \tag{2-94}$$

$\chi^2$-fordelingen er kun defineret for ikke-negative værdier og er "højre-skæv", men formen afhænger af parametren $\nu$.

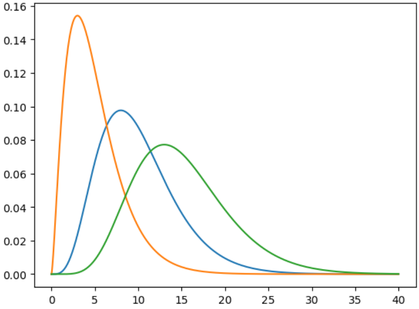
# $\chi^2$-fordelingens tæthedsfunktion

**Definition 2.78**

Let $X$ be $\chi^2$ distributed, then its *pdf* is

$$f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}; \quad x \geq 0, \tag{2-93}$$

where $\Gamma\left(\frac{\nu}{2}\right)$ is the $\Gamma$-function and $\nu$ is the degrees of freedom.




Kahoot!
(x1)

# 02402 Statistik (Polyteknisk grundlag)

- $\chi^2$-fordeling
- t-fordeling
- F-fordeling

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# t-fordelingen

t-fordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$, der er lig med *en standardnormalfordelt variabel (Z) divideret med kvadratroden af en uafhængig $\chi^2$-fordelt variabel (Y) skaleret med parametren $\nu$*:

$$X \sim t(\nu)$$

**Theorem 2.87**

Let $Z \sim N(0,1)$ and $Y \sim \chi^2(\nu)$, then

$$X = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu). \tag{2-99}$$

t-fordelingen er symmetrisk ligner standard Normalfordelingen (især hvis parametren $\nu$ er meget stor), men t-fordelingen har lidt *tykkere haler*.

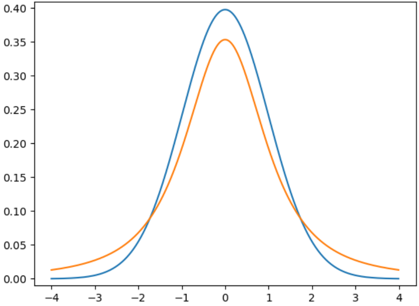
# t-fordelingens tæthedsfunktion

**Definition 2.86**

The $t$-distribution *pdf* is

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad (2\text{-}98)$$

where $\nu$ is the degrees of freedom and $\Gamma()$ is the Gamma function.




Kahoot!
(x1)

# 02402 Statistik (Polyteknisk grundlag)

- $\chi^2$-fordeling
- t-fordeling
- F-fordeling

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

# F-fordelingen

F-fordelingen bruges til at beskrive en kontinuært stokastisk variabel $X$, der er lig med *forholdet mellem to uafhængige $\chi^2$-fordelte variable (der hvér er skaleret deres parameter $\nu$)*:

$$X \sim F(\nu_1, \nu_2)$$

**Theorem 2.96**

Let $U \sim \chi^2(\nu_1)$ and $V \sim \chi^2(\nu_2)$, be independent then

$$F = \frac{U/\nu_1}{V/\nu_2} \sim F(\nu_1, \nu_2). \quad (2\text{-}105)$$

F-fordelingen er kun defineret for ikke-negative værdier og er "højre-skæv", men formen afhænger af parametrene $\nu_1$ og $\nu_2$.

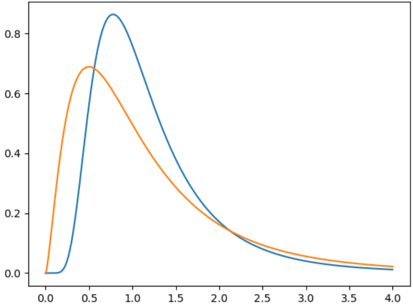
# F-fordelingens tæthedsfunktion

**Definition 2.95**

The $F$-distribution *pdf* is

$$f_F(x) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}}, \tag{2-104}$$

where $\nu_1$ an $\nu_2$ are the degrees of freedom and $B(\cdot,\cdot)$ is the Beta function.



# Særlige fordelinger i Scipy.stats

```
Scipy.stats.chi2
Scipy.stats.t
Scipy.stats.f
```

Generelle 'methods' for kontinuærte fordelinger:

| | |
|---|---|
| `.rvs` | 'random variates' (simulér tilfældige tal) |
| `.pdf` | 'probability mass function' (pdf/tæthedsfunktion) |
| `.cdf` | 'cumulative distribution function' (fordelingsfunktion) |
| `.ppf` | 'percent point function' (invers cdf / fraktilfunktion) |
| `.mean /.var /.std` | 'mean'/'variance'/'standard deviation' |

Se også Appendix A.2.1 i bogen.

Kahoot!
(x1)

# Tjekliste

Efter i dag skal du kunne:

- TEORI: Kende og beskrive de kendte kontinuærte sandsynlighedsfordelinger: Uniform, Normal-, Lognormal- og Exponential-fordelingen
- DATA: Visuelt kunne afgøre om data i en stikprøve der ud til at følge en bestemt fordeling
- TEORI: Anvende regneregler for lineære kombinationer af stokastiske variable
- TEORI: Opskrive stokastiske variable og herpå udføre beregninger passende til en given kontekst
- TEORI: Kende og beskrive de særlige kontinuærte sandsynlighedsfordelinger: $\chi^2$-, $t$- og $F$-fordelingen.
- TEORI m. PYTHON: Benytte Python til at finde diverse størrelser relateret til kontinuærte sandsynlighedsfordelinger (fx sandsynligheder, fraktiler og lign.)
- TEORI m. PYTHON: Simulere tilfældige tal fra en given (kontinuær) sandsynlighedfordeling.

# Øvelser

- Bygning 306 1. sal.
  - 105 (øvelseslokale 96). Hjælpelærer: Nuria
  - 122 (øvelseslokale 98). Hjælpelærer: Ali (KID students)
  - 119 (øvelseslokale 99). Hjælpelærer: Alfred (KID students, overflow)
  - 108A. Hjælpelærer: Afonso
  - 108B. Hjælpelærer: Uffe
  - Man kan også sidde i foyerområdet ved trappen. Hjælpelærer: Sarah
- Bygning 324, stueetagen.
  - 060. Hjælpelærer: Phd-student Kyril
  - 040. Hjælpelærer: Phd-student Thea
  - 050. Hjælpelærer: Jakob
  - 020. Hjælpelærer: Drin
  - 030. Hjælpelærer: Maliha
  - Man kan også sidde i foyerområde 003, 004, 005 og 008.

# Øvelser Uge 3

2.9 Opgaven indeholder Python kode, men kan besvares uden yderligere brug af Python. Tegn gerne i hånden (ikke perfekt).

2.10 Lav opgave (b) **med Python**

2.11 Svær opgave. Lav opgaven **med Python** (men du skal ikke bruge Python særlig meget)

2.12 Svær opgave. Lav opgaven **med Python** (men du skal ikke bruge Python særlig meget)

2.13 OK at bruge Python, men opgaven *kan* laves uden (kræver dog lommeregner).