02403 Introduction to Mathematical Statistics

Lecture 3: Non-linear functions and the Multivariate normal distribution

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Agenda

1. Non-linear function of random variables
2. The multivariate normal distribution
3. Functions of Normal random variables
   - Normal distribution as a sample distribution
   - The χ^2-distribution

Overview

1. Non-linear function of random variables
2. The multivariate normal distribution
3. Functions of Normal random variables
   - Normal distribution as a sample distribution
   - The χ^2-distribution

Distribution by simulation

Assume that that we want to assess the distribution (or some properties) the random variable Y

$$Y = f(V_1, ..., V_m)$$

where $V_1, ..., V_m$ follow some specified dsitributions, and f is some (non-linear) function. The distribution of Y can be assessed by

- Simulate realizations k ($k >> 1$) of $V_1, ..., V_m$, we now have $v_{1,1}, ..., v_{m,1}, v_{1,2}, ..., v_{m,k}$
- Calculate $y_i = f(v_{1,i}, ..., v_{m,i})$ for $i \in \{1, ..., k\}$
- Calculate any summary statisics $\hat{\mu}_y, s_y, ...$, and so on, based on the realzations, or study the distribution from the histogram of the realizations

Example[1]

A simple predator-prey model is the Lotka-Volterra model

$$\frac{dx}{dt} = \alpha x - \beta xy$$
$$\frac{dy}{dt} = \delta xy - \gamma y,$$

where x is the size of the prey population and y is size of the predator population. The equation allows for a constant of motion (i.e. the quantity will stay constant through time for given initial conditions) given by

$$K = y^{\alpha} e^{-\beta y} x^{\gamma} e^{-\delta x}.$$

Assume that $\alpha = 2/3$, $\beta = 4/3$, $\gamma = \delta = 1$, and that

$$X \sim N\left(1, 1/8^2\right)$$
$$Y \sim N\left(1/2, 1/16^2\right)$$

further assume that X and Y are independent. Find (approximate)

- The mean, variance and standard deviation of K
- Sketch the distribution of K

[1] 2024 June

Error propagation

Assume that Y_i are random variables with $E(Y_i) = \mu_i$ og $V(Y_i) = \sigma_i^2$ og $Cov(Y_i, Y_j) = \sigma_{ij}$

We need to find:

$\sigma^2_{f(Y_1,\ldots,Y_n)} = \mathrm{Var}(f(Y_1,\ldots,Y_n))$

(Generalization of) Method 4.3: for non-linaer functions:

$$\sigma^2_{f(Y_1,\ldots,Y_n)} \approx \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_i}\right)^2 \sigma_i^2 + 2\sum_{i}\sum_{j>i} \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\sigma_{ij}$$

Where the derivatives of f are evaluated at $\mu_1, \ldots, \mu_n$.
Notice that if $Y_1, \ldots, Y_n$ are independent then (Method 4.3)

$$\sigma^2_{f(Y_1,\ldots,Y_n)} \approx \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_i}\right)^2 \sigma_i^2$$

Error propagation by Taylor expansion

Let $f(Y_1,\ldots,Y_n)$ be a non-linear function of the random variables $Y_1,...,Y_n$.
We make a first order Taylor expansion around $\mu = [E(Y_1),...,E(Y_n)]^T$

$$\begin{aligned} f(y_1,\ldots,y_n) =& f(\mu) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\bigg|_{\mathbf{y}=\mu} (y_i - \mu_i) + HOT \\ \approx& f(\mu) + \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu} (y_i - \mu_i). \end{aligned}$$

Hence we have

$$f(Y_1,\ldots,Y_n) \approx f(\mu) + \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu} (Y_i - \mu_i).$$

Error propagation by Taylor expansion - Cont.

Find the expectation

$$
\begin{aligned}
E[f(Y_1,\ldots,Y_n)] \approx & E[f(\mu)] + \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu} E[Y_i - \mu_i] \\
= & f(\boldsymbol{\mu})
\end{aligned}
$$

Hence

$$
f(\mathbf{Y}) - E[f(\mathbf{Y})] \approx \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu} (Y_i - \mu_i)
$$

We can now find the variance of $f(\mathbf{Y})$, $V[f(\mathbf{Y})] = E[(f(\mathbf{Y}) - E[f(\mathbf{Y})])^2]$.

Error propagation by Taylor expansion - Cont.

We can now find the variance of $f(\mathbf{X})$, $V[f(\mathbf{X})] = E[(f(\mathbf{X}) - E[f(\mathbf{X})])^2]$.

$$\begin{aligned}
V[f(\mathbf{Y})] \approx & E\left[\left(\sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\boldsymbol{\mu}} (Y_i - \mu_i)\right)^2\right] \\
= & \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu}\right)^2 E\left[(Y_i - \mu_i)^2\right] + \\
& \sum_{i \neq j} \left(\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\right)\bigg|_{\mathbf{y}=\mu} E\left[(Y_i - \mu_i)(Y_j - \mu_j)\right] \\
= & \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_i}\bigg|_{\mathbf{y}=\mu}\right)^2 \sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j>i} \left(\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\right)\bigg|_{\mathbf{y}=\mu} \sigma_{ij}
\end{aligned}$$

Error propagation - Matrix formulation

Assume that $\boldsymbol{Y} \in \mathbb{R}^n$ is a random variable with

$$\begin{aligned} \boldsymbol{\mu} =& E[\boldsymbol{Y}] \\ \boldsymbol{\Sigma} =& V[\boldsymbol{Y}], \end{aligned}$$

the mean and variance covariance of the function $f(\boldsymbol{Y}) \in \mathbb{R}^l$ is approximated by

$$\begin{aligned} E[f(\boldsymbol{Y})] \approx& f(\boldsymbol{\mu}) \\ V[f(\boldsymbol{Y})] \approx& \boldsymbol{J}_f(\boldsymbol{\mu})\boldsymbol{\Sigma}\boldsymbol{J}_f(\boldsymbol{\mu})^T, \end{aligned}$$

where $\boldsymbol{J}_f(\boldsymbol{\mu})$ is the Jacobian of $f(y)$.

Example[2]

A simple predator-prey model is the Lotka-Volterra model

$$\frac{dx}{dt} = \alpha x - \beta xy$$
$$\frac{dy}{dt} = \delta xy - \gamma y,$$

where x is the size of the prey population and y is size of the predator population. The equation allows for a constant of motion (i.e. the quantity will stay constant through time for given initial conditions) given by

$$K = y^{\alpha} e^{-\beta y} x^{\gamma} e^{-\delta x}.$$

Assume and that $E[X] = \mu_x$, $E[Y] = \mu_y$, $V[X] = \sigma_x^2$, $V[Y] = \sigma_y^2$, and $Cov(X,Y) = \sigma_{xy}$. Find the approximation of

- The mean, variance and standard deviation of K
- With $\alpha = 2/3, \beta = 4/3$, $\gamma = \delta = 1$, $\sigma_x^2 = 1/8^2$, $\sigma_y^2 = 1/16^2$, $\sigma_{xy} = 0$, and $\mu_y = 1/2$, sketch the mean and variance of K as a function of μ_x.

[2] Adapted from 2024 June

Overview

The multivariate normal distribution - first definition

A random variable $\boldsymbol{Y} \in \mathbb{R}^n$ with pdf given by

$$f_Y(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})},$$

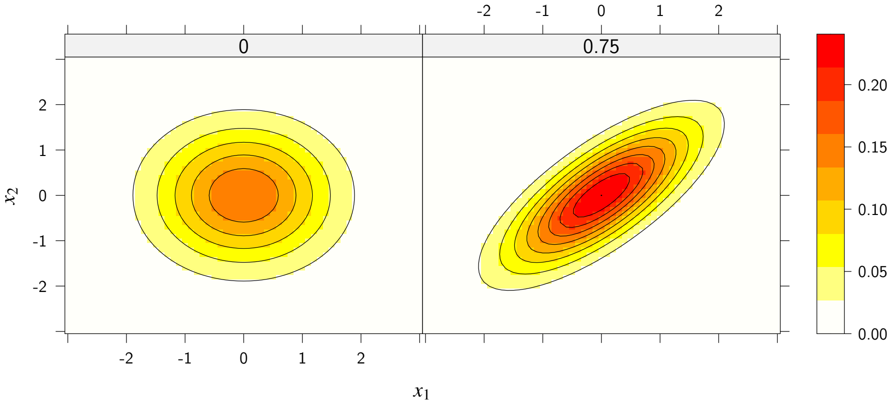
is said to follow a multivariate normal distribution, and we write $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The expected value and the variance-covariance is

$$E[\boldsymbol{Y}] = \boldsymbol{\mu}$$
$$V[\boldsymbol{Y}] = \boldsymbol{\Sigma}.$$

As usual we can write $\boldsymbol{\Sigma} = \boldsymbol{\sigma} \boldsymbol{R} \boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ is a diagonal matrix and $\boldsymbol{R}_{ij}$ is the correlation between Y_i and Y_j.

Example: Bivariate Normal distribution

Normal density for different values of the correlation.



Independece and correlation

Theorem (Independence of normal random variables)

If $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T]^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *and*

$$Cov[\boldsymbol{Y}_1, \boldsymbol{Y}_2] = \boldsymbol{0},$$

then $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$ *are independent.*

Hence in general zero correlation does not imply independence, but for the multivariate normal it does.

Normalization

Theorem (Normalization of normal random vectors)

If $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with the pdf of $\boldsymbol{Y}$ as defined in slide 13 (implying that $\boldsymbol{\Sigma}$ is positive definite), then

$$\boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim N(\boldsymbol{0}, \boldsymbol{I})$$

with $\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}$ (implying that $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{T}{2}} = \boldsymbol{\Sigma}$), where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{\Sigma}$ in the diagonal and $\boldsymbol{V}$ is the corresponding eigenvectors.

From linear algebra: A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with n linear independent eigenvectors can be factored as

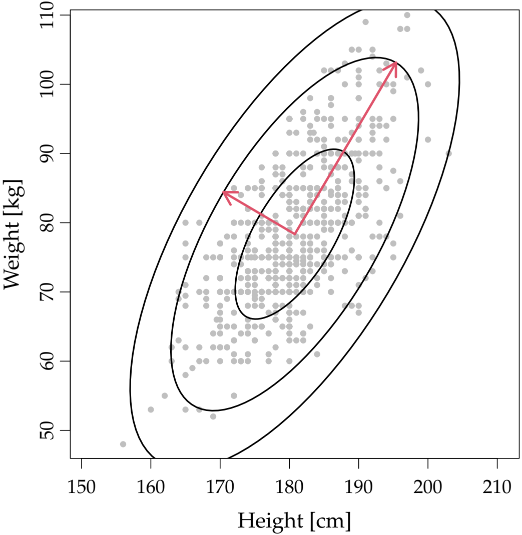
$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues along the diagonal, and $\boldsymbol{V}$ the collection of eigenvectors. If in addition $\boldsymbol{A}$ is symmetrix then also

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{T}$$

i.e. $\boldsymbol{V}^T = \boldsymbol{V}^{-1}$.

Example: Distribution of height and weight



Multivariate normal: general definition

Definition (Multivariate normal distribution)

Let Z_i, $i = 1, ..., n$, be iid. standard normal random variables, s.t. $(\boldsymbol{Z} = [Z_1, ..., Z_n]^T)$

$$\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}).$$

Then the random vector $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}$, with $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, follow an m-dimensional multivariate normal distribution with

$$E[\boldsymbol{Y}] = \boldsymbol{b}$$
$$V[\boldsymbol{Y}] = \boldsymbol{A}\boldsymbol{A}^T,$$

this holds also when $\boldsymbol{A}\boldsymbol{A}^T$ is not positive definite.

Example

You put an item (with a known weight) onto to a scale and record the error made by the scale. The procedure is repeated two times and the mean and standard error of the scale is assumed to be 0 g, and 1 g respectively. The outcome are denoted Z_1 and Z_2, Now define

$$r_i = Z_i - \bar{Z}.$$

- what is the distribution of $\boldsymbol{r} = [r_1, r_2]$?
- what is the distribution of $[\boldsymbol{r}^T, \bar{Z}]$?

Overview

1. Non-linear function of random variables
2. The multivariate normal distribution
3. Functions of Normal random variables
   - Normal distribution as a sample distribution
   - The χ^2-distribution

The framework of *statistisk inferens*

From eNote, Chapter 1:

- An *observational unit* is the single entity/level about which information is sought (e.g. a person) (**Observationsenhed**)
- The *statistical population* consists of all possible "measurements" on each *observational unit* (**Population**)
- The *sample* from a statistical population is the actual set of data collected. (**Stikprøve**)

Language and concepts:

- μ and σ are parameters that describe the populationen
- $\bar{x}$ is an *estimate* of μ (an actual outcome, a number)
- $\bar{X}$ and S^2 are *estimatorerers* of μ and σ^2 (these are random variables)
- The concept *'statistic(s)'* is used for both

The aim

In lecture 1 we saw a number of summary statistics, we now assume that

$$Y_i \sim N(\mu, \sigma^2), \quad \text{and iid.}$$

In this and the next lecture we will answer the following questions

- What is the distribution of $\bar{Y}$?
- What is the distribution of S^2?
- What is the distribution of $\frac{\bar{Y}-\mu}{S^2/\sqrt{n}}$?
- If we calculated observed variances from two different groups, what is then the distribution of $\frac{S_1^2}{S_2^2}$?

Why do we answer those questions?

In statistics we typically have a number of summary statistics (e.g. $\bar{y}$ and s^2) from a sample.

- We want to make statements about the population parameters (e.g. μ and σ^2)
- In general this require distribution assumption about the population (e.g. the normal distribution)
- In order to quantify unceartainties we need distributions of the summary statistics.

Using the normal assumption we will study these distribututions in todays and the next lecture.

Example: Average and variance of normal sample

Assume that we plan a study with 5 independent observations. We also assume that the mean and variance in the population is $\mu = 10$ and $\sigma^2 = 2$, what is the distribution of the average $(\bar{Y})$ and the empirical variance S^2 under these assumptions?

Some general concepts

Central Estimator:
An estimator, $\hat{\theta}$, is central (or non-biased), if and only if, the mean value of the estimator equals θ

Consistent Estimator
A central estimator, $\hat{\theta}$, that converge in probability is called a consistent estimator (you can think of this as $V(\theta_n) \to 0$).

Efficient Estimator
An estimator $\hat{\theta}_1$ is a more efficient estimator for θ than $\hat{\theta}_2$ if:

1. $\hat{\theta}_1$ and $\hat{\theta}_2$ both are central estimators of θ
2. The variance of $\hat{\theta}_1$ is less than the variance of $\hat{\theta}_2$

Estimate
When we have the actual sample and have calculated the summary statistic, we have an estimate (this is not a random variabele)

Example

If $Y_1,..,Y_n$ are iid. $N(\mu,\sigma^2)$ random variables, then

- $\overline{Y} = \hat{\mu}$ is a central estimator for μ ($E[\overline{Y}] = \mu$).
- $\overline{Y}$ is also a consistent estimator for μ ($V[\overline{Y}] = \frac{\sigma^2}{n} \to 0$, $n \to \infty$).
- $\bar{y}$ is an estimate of μ.
- The median is also a central and consistent estimator for μ, but the median is less efficient.

Distribution of the average

The (sampling) distribution for $\bar{Y}$

Assume that $Y_1, \ldots, Y_n$ are independent and identically normally distributed random variables, $Y_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$, then:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Find d such that

$$P(\bar{Y} - d\sigma < \mu < \bar{Y} + d\sigma) = 1 - \alpha$$

Example

Let $Y_1 \sim N(\mu, \sigma^2)$ and $Y_2 \sim N(\mu, \sigma^2)$ be independent random variables, what is the mean value of

$$Q = (Y_1 - \overline{Y})^2 + (Y_2 - \overline{Y})^2$$

Example: Expected value of variance estimator

Let $Y_1, \ldots, Y_n$ be iid random variables with mean values $E[Y_i] = \mu$ and variance $V[Y_i] = \sigma^2$, and let Q be

$$Q = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

What is the expected value of Q?

Example: Expected value of variance estimator

Let $Y_1, \ldots, Y_n$ be iid random variables with mean values $E[Y_i] = \mu$ and variance $V[Y_i] = \sigma^2$, and let Q be

$$Q = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

What is the expected value of Q?

$$\begin{aligned}
E[Q] &= \sum_{i=1}^{n} E[(Y_i - \overline{Y})^2] \\
&= \sum_{i=1}^{n} E[(Y_i - \mu + \mu - \overline{Y})^2] \\
&= \sum_{i=1}^{n} E[(Y_i - \mu)^2 + (\mu - \overline{Y})^2 + 2(Y_i - \mu)(\mu - \overline{Y})]
\end{aligned}$$

Example: Expected value of variance estimator

$$
\begin{aligned}
E[Q] =& \sum_{i=1}^{n} E[(Y_i-\mu)^2] + E[(\mu-\overline{Y})^2] - 2Cov[Y_i,\overline{Y}] \\
=& n\sigma^2 + \sigma^2 - 2\sum_{i=1}^{n}\frac{1}{n}Cov\left(Y_i, \sum_{j=1}^{n} Y_j\right) \\
=& (n+1)\sigma^2 - 2\sum_{i=1}^{n}\frac{1}{n}Cov(Y_i,Y_i) \\
=& (n+1)\sigma^2 - 2\sigma^2 \\
=& (n-1)\sigma^2
\end{aligned}
$$

Hence $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$ is a central estimator for σ^2.

The χ^2-distribution

Definition

If $Y_1, .., Y_n$ is *iid* N(0,1) then

$$Q = \sum_{i=1}^{n} Y_i^2$$

follow a χ^2-distribution with n-degrees of freedom, we write $Q \sim \chi^2(n)$.

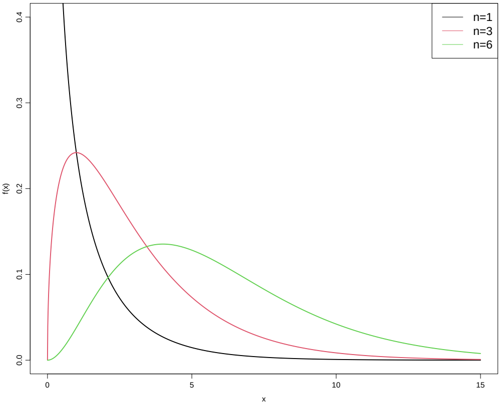
Theorem

The probability function of a χ^2-distribution is given by

$$f(y) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\, y^{\frac{n}{2}-1} e^{-\frac{y}{2}}; \quad x \geq 0.$$

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

The χ^2-distribution



Properties of the χ^2-distribution

If $Q \sim \chi^2(n)$ then

$$E(Q) = n$$
$$V(Q) = 2n$$

If $Q_1 \sim \chi^2(n_1)$ and $Q_2 \sim \chi^2(n_2)$ are independent then

$$Q = Q_1 + Q_2 \sim \chi^2(n_1 + n_2)$$

Example

If $Y_1,\ldots,Y_{10}$ are iid $N(\mu,\sigma^2)$ and

$$Q = \frac{1}{\sigma^2}\sum_{i=1}^{10}(Y_i - \mu)^2$$

What is $P(Q > 10)$ then?

Distribution of variance estimator

If $Y_1, \ldots, Y_n$ are iid. $N(\mu, \sigma^2)$, with $\overline{Y}$, S^2 the average and empirical variance. Then

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{Y} - \mu}{\sigma / \sqrt{n}} \right)^2$$

and it follows that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Proof (sketch)

1: $\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2 \sim \chi^2(n)$ and $\frac{(\bar{Y}-\mu)^2}{\sigma^2/n} \sim \chi^2(1)$

2: $\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2 = \frac{(n-1)S^2}{\sigma^2} + \frac{(\bar{Y}-\mu)^2}{\sigma^2/n}$

3: $Cov(\bar{Y}, \bar{Y} - Y_i) = 0 \Rightarrow S^2$ and $(\bar{Y}-\mu)^2$ independent

4: If $Q_1 \sim \chi^2(n_1)$ and $Q_2 \sim \chi^2(n_2)$ independent, then $Q_1 + Q_2 \sim \chi^2(n_1 + n_2)$

Example

Find $E(S^2)$ and $V(S^2)$.

Example

Find $E(S^2)$ and $V(S^2)$. Answer:

$$E[S^2] = \frac{\sigma^2}{n-1} E\left[\frac{n-1}{\sigma^2} S^2\right]$$
$$V[S^2] = \left(\frac{\sigma^2}{n-1}\right)^2 V\left[\frac{n-1}{\sigma^2} S^2\right]$$

Since $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$ it follows that

$$E[S^2] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$
$$V[S^2] = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

This imply that S^2 er en central and consistent estimator for σ^2.

Example: Pooled variance

Let $Y_{1,1},...,Y_{1,n_1}$ og $Y_{2,1},...,Y_{2,n_2}$ be independent variables, with $Y_{1,i} \sim N(\mu_1, \sigma^2)$, and $Y_{2,i} \sim N(\mu_2, \sigma^2)$. With $a \in [0,1]$ find a such that $V[S_p^2]$ is minimised with

$$S_P^2 = aS_1^2 + (1-a)S_2^2$$

where S_1^2 and S_2^2 is the sample variance for Y_1 and Y_2. Find a and $V[S_P(a)^2]$.

The aim

In lecture 1 we saw a number of summary statistics, we now assume that

$$Y_i \sim N(\mu, \sigma^2), \quad \text{and iid.}$$

In this and the next lecture we will answer the following questions

- What is the distribution of $\bar{Y}$? DONE!
- What is the distribution of S^2? DONE!
- What is the distrubution of $\frac{\bar{Y}-\mu}{S^2/\sqrt{n}}$? TBD
- If we calculated observed variances from two different groups, what is then the distribution of $\frac{S_1^2}{S_2^2}$? TBD

Summary of today

1. Non-linear function of random variables
2. The multivariate normal distribution
3. Functions of Normal random variables
    - Normal distribution as a sample distribution
    - The χ^2-distribution