## **21. Multivariate random variable** or **random vector**

$\vec{\xi} = [\xi_1, \xi_2, \ldots, \xi_n]$ random vector is composed of individual random variables

**Multivariate cumulative probability distribution**
or **joint cumulative probability distribution**: mapping $\mathbb{R}^n \mapsto [0,1]$

$$F(x_1, x_2, \ldots, x_n) = P\{(\xi_1 \leq x_1) \wedge (\xi_2 \leq x_2) \wedge \ldots \wedge (\xi_n \leq x_n)\}$$

**Multivariate probability density function**

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} f(u_1, u_2, \ldots, u_n)\, du_1\, du_2\, \ldots\, du_n$$

## Expected value for a function of a random vector (mapping $\mathbb{R}^n \rightarrow \mathbb{R}$)

$$E[h(\xi_1, \xi_2, \ldots, \xi_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(u_1, u_2, \ldots, u_n) f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n$$

## 22. Expected value of a linear combination of random variables

$$h(\xi_1, \xi_2, \ldots, \xi_n) = \sum_1^n a_i \xi_i + b$$

$$\begin{aligned} E\left[\sum_1^n a_i \xi_i + b\right] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_1^n a_i u_i + b\right] f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n \\ &= \sum_1^n\ a_i\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_i\ f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n \\ &\quad +b\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, u_2, \ldots, u_n)\ du_1\ du_2\ \ldots\ du_n = \\ &= \sum_1^n\ a_i\ E(\xi_i) + b \end{aligned}$$

## 23. Variance of a linear combination of random variables

$$\begin{aligned} D\left[\sum_{i=1}^{n} a_i\xi_i + b\right] &= E\left\{\left[\sum_{i=1}^{n} a_i\xi_i + b - E\left(\sum_{i=1}^{n} a_i\xi_i + b\right)\right]^2\right\} \\ &= E\left\{\left[\sum_{i=1}^{n} a_i\xi_i - \sum_{i=1}^{n} a_i E(\xi_i)\right]^2\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{cov}(\xi_i, \xi_j) \end{aligned}$$

## 24. Covariance

$$\mathrm{cov}(\xi_i, \xi_j) = E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\}$$

Auto-covariance of a random variable is the same as its variance

$$\mathrm{cov}(\xi_i, \xi_i) = E\left\{[\xi_i - E(\xi_i)]^2\right\} = D(\xi_i)$$

# Independent random variables

## Multivariate cumulative distribution of independent random variables

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= P\{(\xi_1 \leq x_1) \wedge (\xi_2 \leq x_2) \wedge \ldots \wedge (\xi_n \leq x_n)\} \\ &= P(\xi_1 \leq x_1) \cdot P(\xi_2 \leq x_2) \cdot \ldots \cdot P(\xi_n \leq x_n) \\ &= F(x_1) \cdot F(x_2) \cdot \ldots \cdot F(x_n) \end{aligned}$$

## Multivariate probability density function of independent random variables

$$f(x_1, x_2, \ldots, x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$$

## 25. Expected value of a product of independent random variables

$$\begin{aligned} E(\xi_1 \cdot \xi_2 \cdot \ldots \cdot \xi_n) &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u_1 \cdot u_2 \cdot \ldots \cdot u_n \; f(u_1, u_2, \ldots, u_n) \, du_1 \, du_2 \; \ldots \; du_n \\ &= E(\xi_1) \cdot E(\xi_2) \cdot \ldots \cdot E(\xi_n) \end{aligned}$$

## 26. Covariance of independent random variables

$$\operatorname{cov}(\xi_i, \xi_j) = E\left\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\right\} = E[\xi_i - E(\xi_i)] \cdot E[\xi_j - E(\xi_j)] = 0$$ for i≠j

## 27. Variance of a linear combination of independent random variables

$$D\left[\sum_{i=1}^{n} a_i \xi_i + b\right] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \operatorname{cov}(\xi_i, \xi_j) = \sum_{i=1}^{n} a_i^2 D(\xi_i)$$

## 28. Covariance matrix

$$C = \begin{pmatrix} \mathrm{cov}(\xi_1, \xi_1) & \cdots & \mathrm{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(\xi_n, \xi_1) & \cdots & \mathrm{cov}(\xi_n, \xi_n) \end{pmatrix}$$

- Symmetric by definition
- Only the main diagonal for **independent** random varaibles, cross-covariances are zero
- Main diagonal: auto-covariances $\mathrm{cov}(\xi_i, \xi_i) = D(\xi_i)$
- For linearly dependent random variables $(\xi_i = a_i \xi)$:

$$\mathrm{cov}(a_i\xi, a_j\xi) = E\left\{[a_i\xi - E(a_i\xi)][a_j\xi - E(a_j\xi)]\right\} = a_i a_j D(\xi) = \pm\sqrt{D(\xi_i)D(\xi_j)}$$

## 29. Coefficient of linear correlation

$$\rho(\xi_i, \xi_j) = \frac{\mathrm{cov}(\xi_i, \xi_j)}{\sqrt{D(\xi_i)D(\xi_j)}}$$

Independent random variables: $\rho(\xi_i, \xi_j) = 0$

Lineraly dependent random variables:

$\xi_i = a\xi_j, a > 0 : \rho(\xi_i, \xi_j) = 1$ $\quad$ $\xi_i = a\xi_j, a < 0 : \rho(\xi_i, \xi_j) = -1$

## 30. Pseudo-random sequences

- Observation: a realization of the random variable
- Computers are usually deterministic: how to obtain a random number?

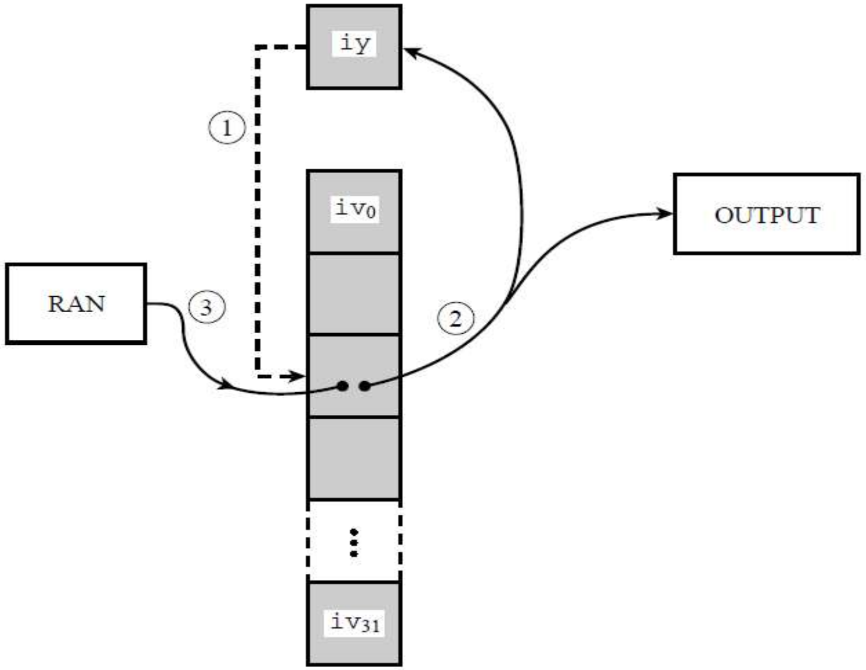
- Only **pseudo-random** – requirements depend on a particular purpose:
  - length of the sequence
  - internal bounds within a sequence of pseudo-random samples
  - Test of internal bounds: sequence of $k$ pseudo-random numbers defines a point in the $k$-dimensional space. These points usually form $k$-1 dimensional hyperplanes.

## 31. Uniform distribution – basis for all other distributions

- Linear congruential generators $I_{j+1} = (aI_j + c) \mod m \quad (*)$
- Period $\leq m$, $a, c, m$ must be carefully selected
- Simple example: 4 byte (32 bit) generator with

$$m = 2^{31} - 1 = 2147483647, \ c = 0, \ a = 7^5 = 16807$$

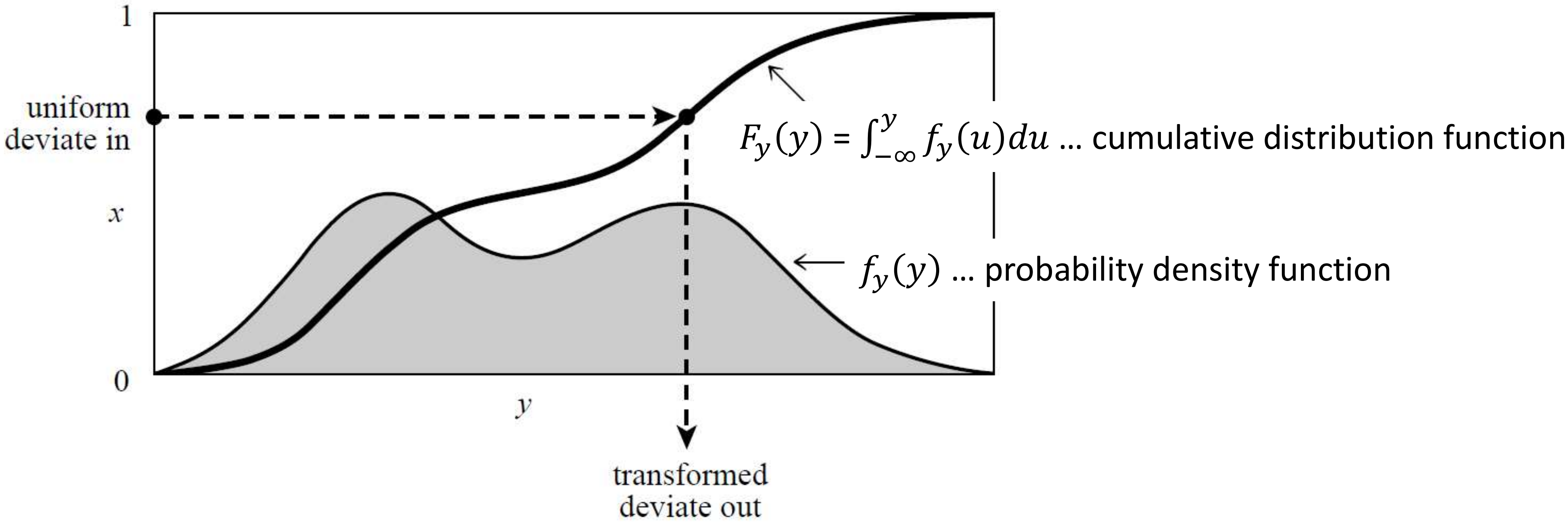
- Test: sequence of *k* pseudo-random numbers defines a point in the *k*-dimensional space. These points form a maximum number of $m^{1/k}$ *k*-1 dimensional hyperplanes.




- Modification of the linear congruential generator to decrease the strength of **internal bounds**:
  - Define an array of 32 numbers from $(*)$
  - *y*=0
    - (+) The *y*th element goes to the output
    - The *y*th element is refilled from $(*)$
    - The last 5 bits of the output = next *y*
    
    > (+)

## 32. Transformation method to generate an arbitrary distribution

- Probability of observing a value between $x$ and $x + dx$: $dP = f_x(x)\, dx$,
  where $f_x$ is the probability density function
- Transformed random variable $y(x)$ : $dP = f_x(x)dx = f_y(y)dy \rightarrow f_y(y) = f_x(x)\,\frac{dx}{dy}$
- Uniform distribution on $\langle 0,1 \rangle$: $f_x(x) = 1 \rightarrow f_y(y) = \frac{dx}{dy} \rightarrow x = F_y(y) \rightarrow y = F_y^{-1}(x)$



## 33. Pseudo-random sequence with an exponential distribution

**Probability density function of an exponential distribution:**

$$f_{exp}(x) = \frac{1}{\delta} \exp\left(-\frac{x}{\delta}\right)$$

**Cumulative distribution function of an exponential distribution:**

$$F_{exp}(x) = 1 - \exp\left(-\frac{x}{\delta}\right)$$

**Transformation of a uniform distribution**

$$x = {F_{exp}}^{-1}(y)$$

$$\boldsymbol{x} = -\,\delta \ln(1 - y) = \boldsymbol{-\,\delta \ln(u)}$$

$\boldsymbol{x}$ has an exponential distribution,

for a uniform distribution of $y$ or $\boldsymbol{u}$ on $\langle 0,1 \rangle$

## 34. Pseudo-random sequence with a normal distribution

**Multivariate (joint) probability density of a transformed random vector $y(x)$:**

$$dP = f_x(x_1, x_2, \dots, x_n)\, dx_1\, dx_2 \dots dx_n = f_y(y_1, y_2, \dots, y_n)\, dy_1\, dy_2 \dots dy_n$$

$$f_y(y_1, y_2, \dots, y_n) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} f_x(x_1, x_2, \dots, x_n)$$

where | … | is the determinant of the Jacobian matrix

George E. P. Box (1919–2013)

Mervin E. Muller (1928–2018)

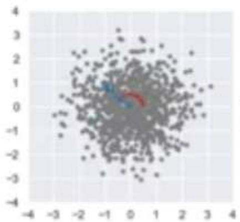
**Box-Muller method (1958):**

$x_1, x_2$ …uniform distribution on $\langle 0,1 \rangle$

$$y_1 = \sqrt{-2 \ln x_1} \cos 2\pi x_2$$

$$y_2 = \sqrt{-2 \ln x_1} \sin 2\pi x_2$$

$y_1, y_2$ …normal distribution $\mu = 0, \sigma = 1$



$\rightarrow$

$$x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$

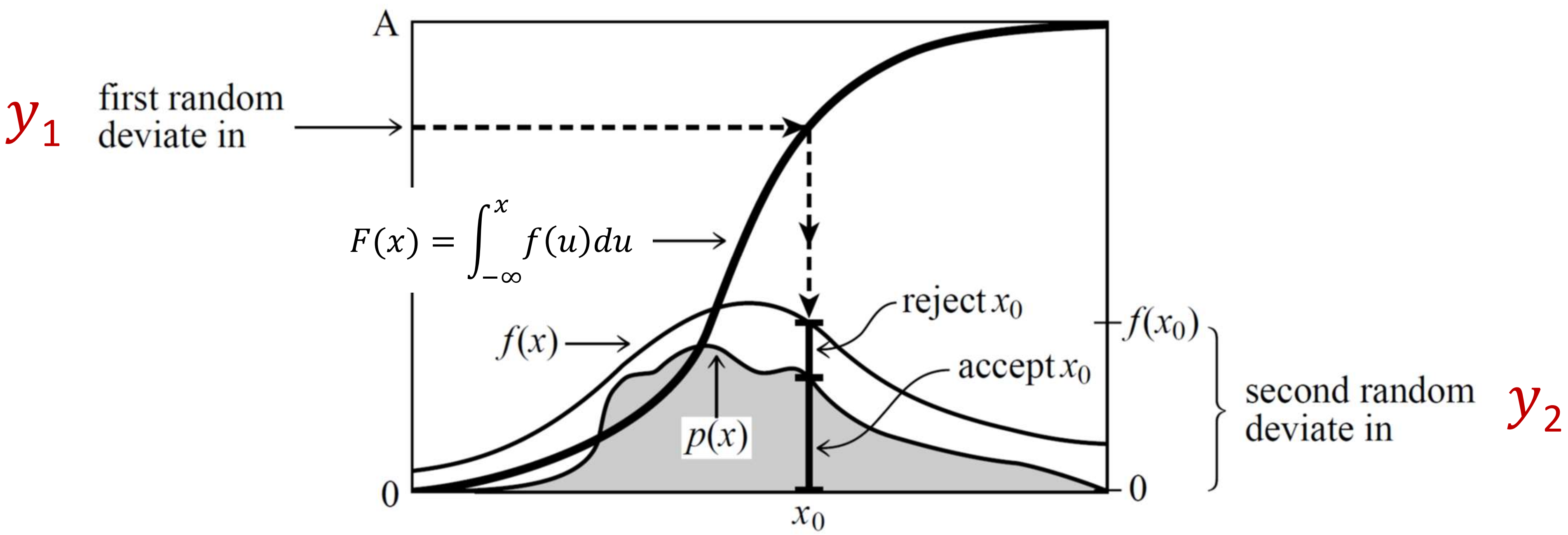
$$x_2 = \frac{1}{2\pi} \arctan \frac{y_2}{y_1}$$

$$| \dots | = -\left[\frac{1}{\sqrt{2\pi}} e^{-y_1^2/2}\right] \left[\frac{1}{\sqrt{2\pi}} e^{-y_2^2/2}\right]$$

## 35. Rejection method

**General** method to generate an arbitrary distribution with a probability density function $p(x)$
Does not need the inverse cumulative probability function

- Choose a "nice" comparison function $f(x) > p(x)$, with a finite $A = \int_{-\infty}^{\infty} f(u)du$
- Calculate the inverse function $F^{-1}(y)$ to $F(x) = \int_{-\infty}^{x} f(u)du$
- Generate a random number $y_1$ uniformly distributed on $\langle 0, A \rangle$ $\rightarrow$ $x_0 = F^{-1}(y_1)$
- Generate a random number $y_2$ uniformly distributed on $\langle 0, f(x_0) \rangle$ $\rightarrow$ **reject or accept $x_0$**



## 36. Mathematical statistics: application of probability theory

- **Statistical population:** set of objects (existing or hypothetical), the latter can be infinite
- **Sample:** a finite subset of a given statistical population, selected by a known procedure
- **Sample size:** a finite number of elements of the sample
- **Observations**, a.k.a sample points, sample units: elements of the sample
- **Random sample:** a sample with defined (e.g., equal) selection probabilities for all elements of the population;

  For a hypothetical **infinite** statistical population generated by a random process with a **given probability distribution**, the random sample with a sample size $N$ is a set of realizations of $N$ independent, identically distributed (i.i.d.) random variables.
- **A sample statistic:** a quantity calculated from elements of a sample;

  for a random sample, every statistic is a **random variable**
- **An estimator** $\widehat{\boldsymbol{\vartheta}}$: a sample statistic is a **random variable** designed to estimate a parameter $\boldsymbol{\vartheta}$ of the population

## 37. Bias of an estimator $\hat{\vartheta}$

$$b(\hat{\vartheta}) = \mathrm{E}(\hat{\vartheta}) - \vartheta$$

$b(\hat{\vartheta}) = 0$ for an **unbiased estimator**

**Example: sample mean value** obtained as the **arithmetic average**

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Is it an unbiased estimator of the expected value (population mean) $\mu$ of $N$ independent, identically distributed (i.i.d.) random variables $X_i$ ?

$$b(\hat{\mu}) = \mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \mu = \frac{1}{N}\sum_{i=1}^{N} E(X_i) - \mu = \frac{N\mu}{N} - \mu = 0$$

**BUT:** taking just the last value $X_N$ (and trash $X_1 \dots X_{N-1}$): $\hat{\mu}_L = X_N$ is also unbiased

## 38. Asymptotically consistent estimator $\hat{\boldsymbol{\vartheta}}_N$

converges to the true value $\boldsymbol{\vartheta}$ for the sample size $\boldsymbol{N} \to \infty$:

$$\forall \varepsilon > 0: \quad \lim_{N\to\infty} P\left(\left|\hat{\vartheta}_N - \vartheta\right| > \varepsilon\right) = 0 \qquad \text{(convergence in probability)}$$

**Example:** $\hat{\mu}_L = X_N$: no convergence, while the sample mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$ converges to $\mu$

**But:** an estimator can be biased and still asymptotically consistent

$$\hat{\mu}_B = \frac{1}{N}\left(333 + \sum_{i=1}^{N} X_i\right) \quad \to \quad b(\hat{\mu}_B) = \frac{333}{N} \neq 0 \quad \text{while} \quad \lim_{N\to\infty} \hat{\mu}_B = \mu$$

**Conclusion:** We need BOTH unbiased AND asymptotically consistent estimators

## 39. Variance of an estimator: $\mathrm{D}(\hat{\vartheta}) = \mathrm{E}\left\{\left[\hat{\vartheta} - \mathrm{E}(\hat{\vartheta})\right]^2\right\}$ $\to \lim_{N\to\infty} \mathrm{D}(\hat{\vartheta}) = 0 \;\Rightarrow$ consistency

**Example:**

$$\mathrm{D}(\hat{\mu}) = \mathrm{E}\left\{\left[\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \mu\right]^2\right\} = \frac{1}{N^2}\mathrm{E}\left\{\left[\sum_{i=1}^{N}(X_i - \mu)\right]^2\right\} =$$

$$= \frac{1}{N^2}\mathrm{E}\left\{\sum_{i=1}^{N}[X_i - \mu]^2\right\} + \frac{1}{N^2}\mathrm{E}\left\{\sum\sum_{i\neq j}\left[(X_i - \mu)(X_j - \mu)\right]\right\} =$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\{[X_i - \mu]^2\} + \frac{1}{N^2}\sum\sum_{i\neq j}\Big[\underbrace{\mathrm{E}(X_i - \mu)\mathrm{E}(X_j - \mu)}_{0\,\cdot\,0}\Big] = \frac{1}{N^2}N\sigma^2 = \frac{1}{N}\boldsymbol{\sigma}^2$$

$\mathrm{E}(\Sigma X_i) = \Sigma\,\mathrm{E}(X_i)$

$(X_i - \mu)$ and $(X_j - \mu)$ are independent r. v.

# 40. Unbiased estimator of variance, Bessel's correction


**Friedrich Wilhelm Bessel**
Accountant in a Bremen trade company; navigation → astronomy. Orbit calculations of Halley's comet. Director of the Königsberg observatory. First measured the distance of a star (61 Cyg) using its parallax. Introduced personal equation. Geodesy: Earth´s ellipsoid.

Average squared deviations from the sample mean: $\hat{d} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{\mu})^2$ with $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$

$$\mathrm{E}(\hat{d}) = \mathrm{E}\left\{\frac{1}{N}\sum_{i=1}^{N}[(X_i - \mu) - (\hat{\mu} - \mu)]^2\right\} =$$

$$= \frac{1}{N}\mathrm{E}\left\{ \sum_i[(X_i - \mu)^2] - 2\sum_i\left[(X_i - \mu)\left(\frac{1}{N}\sum_{j=1}^{N} X_j - \mu\right)\right] + \sum_i\left[\left(\frac{1}{N}\sum_{j=1}^{N} X_j - \mu\right)^2\right] \right\} =$$

$$= \frac{1}{N}\left\{ \sum_i \mathrm{E}[(X_i - \mu)^2] - \frac{2}{N}\sum_i\sum_j \mathrm{E}[(X_i - \mu)(X_j - \mu)] + N\,\mathrm{E}\left[\left(\frac{1}{N}\sum_{j=1}^{N} X_j - \mu\right)^2\right] \right\} =$$

$$= \frac{1}{N}\left\{ N\sigma^2 - \frac{2}{N}\sum_{i=1}^{N}\mathrm{E}\{[X_i - \mu]^2\} - \frac{2}{N}\sum\sum_{i\neq j}\Big[\underbrace{\mathrm{E}(X_i - \mu)\mathrm{E}(X_j - \mu)}_{0\ \cdot\ 0}\Big] + N\,\mathrm{D}[\hat{\mu}] \right\} =$$

$\mathrm{E}(\Sigma X_i) = \Sigma\,\mathrm{E}(X_i)$

$(X_i - \mu)$ and $(X_j - \mu)$ are independent r. v.

$$= \frac{1}{N}\left(N\sigma^2 - \frac{2}{N}N\sigma^2 - 0 + N\frac{\sigma^2}{N}\right) = \sigma^2\frac{N-1}{N} \Rightarrow$$ $\hat{d}$ is biased

Unbiased estimator of variance $\hat{\sigma}^2$ using Bessel's correction $\frac{N}{N-1}$:

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu})^2$$

$$\Rightarrow \hat{\sigma}_{\hat{\mu}}^2 = \frac{1}{N(N-1)}\sum_{i=1}^{N}(X_i - \hat{\mu})^2$$

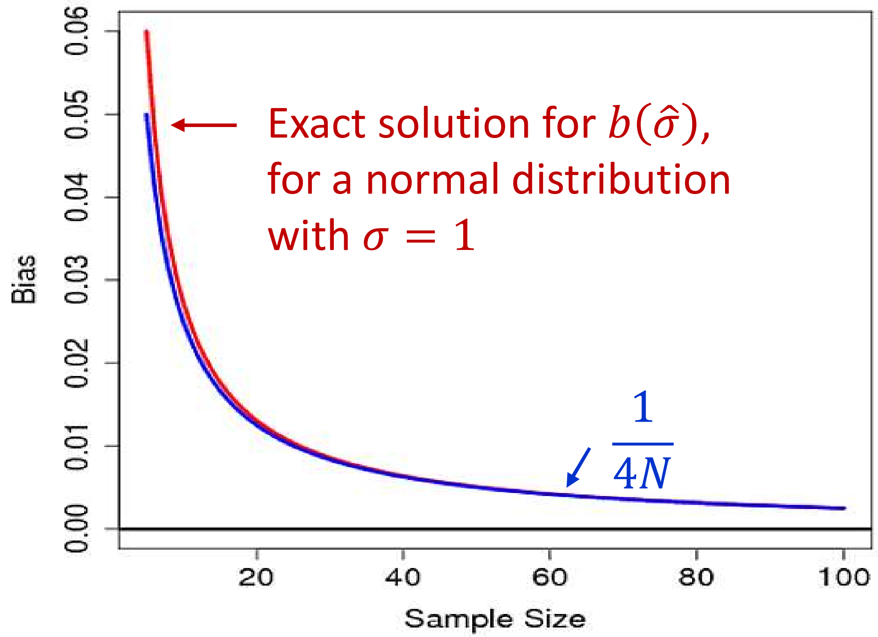
## 41. Sample standard deviation

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}\text{, where } \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu})^2$$

**BUT:** $E\left(\sqrt{\hat{\sigma}^2}\right) \neq \sqrt{E(\hat{\sigma}^2)} \quad \Rightarrow \quad b(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = 0 \;\rightarrow$ **unbiased estimator of variance**

and $b(\hat{\sigma}) = b\left(\sqrt{\hat{\sigma}^2}\right) = E\left(\sqrt{\hat{\sigma}^2}\right) - \sqrt{\sigma^2} \neq 0 \rightarrow$ **biased estimator of standand deviation**

Bias depends on the probability distribution of $X_i$. For a normal distribution $b(\hat{\sigma}) \approx \frac{\sigma}{4N}$



## 42. Practical calculation of the sample variance

**Single pass**

accumulation of $S_X = \sum_{i=1}^{N} X_i$ and $S_{XX} = \sum_{i=1}^{N} {X_i}^2$

then

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu})^2 = \frac{1}{N-1}\sum_{i=1}^{N}({X_i}^2 - 2X_i\hat{\mu} + \hat{\mu}^2) = \frac{1}{N-1}\left(S_{XX} - \frac{1}{N}S_X^2\right)$$

BUT: has a large accumulated round-off error (cancellation effect)

**Two-pass procedure**

1. Sample mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$
2. Sample variance with round-off correction

$$\hat{\sigma}^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N}(X_i - \hat{\mu})^2 - \frac{1}{N}\left[\sum_{i=1}^{N}(X_i - \hat{\mu})\right]^2\right\}$$

**Recursive procedure**

Used when adding new data to improve the $\hat{\mu}$ and $\hat{\sigma}^2$ estimators → recurrence relation

Welford's algorithm (1962)

Sample mean: $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} X_i = \hat{\mu}_{N-1} + \frac{1}{N}(X_N - \hat{\mu}_{N-1})$

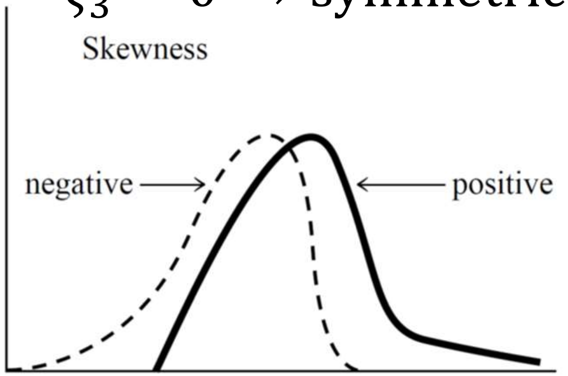
Sample variance: $\hat{s}^2{}_N = \sum_{i=1}^{N}(X_i - \hat{\mu})^2 = \hat{s}^2{}_{N-1} + (X_N - \hat{\mu}_{N-1})(X_N - \hat{\mu}_N)$

$$\hat{\sigma}^2{}_N = \frac{\hat{s}^2{}_N}{N-1}$$

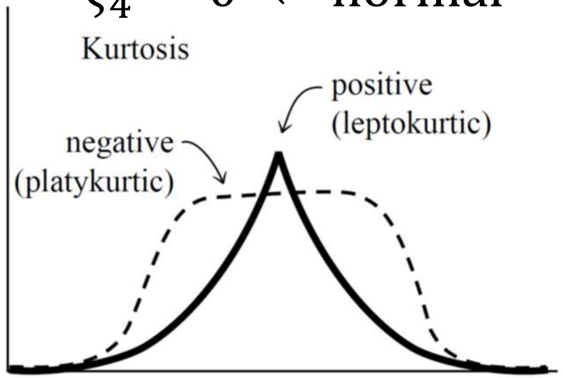
## 43. Estimators of higher moments

Nondimensional quantities:

$\zeta_3 = 0 \Rightarrow$ symmetrical




$\zeta_4 = 0 \Leftarrow$ normal




**Skewness** $$\zeta_3 = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}} = \frac{\mu_3}{\sigma^3}$$

**Kurtosis** $$\zeta_4 = \frac{E[(X-\mu)^4]}{E[(X-\mu)^2]^2} = \frac{\mu_4}{\sigma^4}$$

**Excess Kurtosis** $$\zeta_4 = \zeta_4 - 3$$

**Sample skewness** $$\hat{\zeta}_3 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right)^3$$

$X_i$ normal $\Rightarrow$ variance $D(\hat{\zeta}_3) \approx \frac{6}{N}$

**Sample (excess) kurtosis** $$\hat{\zeta}_4 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right)^4 - 3$$

$X_i$ normal $\Rightarrow$ variance $D(\hat{\zeta}_4) \approx \frac{24}{N}$

## 44. Sample median, average absolute deviation

The **sample median $\widehat{M}$** as an estimator of the population median $M$ makes use of only one or two of the middle values out of the entire sample of $N$ values $X_1 \le \cdots \le X_N$, and is thus not affected by extremes (a robust statistic):

$$\widehat{M} = X_{(N+1)/2} \qquad \text{for odd } N$$

$$\widehat{M} = \frac{1}{2}\left(X_{N/2} + X_{N/2+1}\right) \qquad \text{for even } N$$

The distribution of $\widehat{M}$ from a population with a probability density function $f(x)$ is asymptotically normal with the expected value $M$ and variance $D(\widehat{M}) = \dfrac{1}{4Nf(M)^2}$

For samples with a normal distribution: $f(M) = 1/\sqrt{2\pi\sigma^2}$ and $D(\widehat{M}) = \dfrac{\pi\sigma^2}{2N}$


Pierre-Simon, marquis de Laplace
1749 – 1827

The **sample average absolute deviation** $\widehat{D}_{\widehat{M}} = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - \widehat{M}\right|$

- Using $\widehat{M}$ minimizes $\widehat{D}_{\widehat{M}}$: $\frac{1}{N}\left(\widehat{M} - X_1\right) + \cdots + \frac{1}{N}\left(X_N - \widehat{M}\right)$, $\partial\widehat{D}_{\widehat{M}}/\partial\widehat{M} = \underbrace{\frac{1}{N} + \cdots + \frac{1}{N}}_{\lfloor N/2 \rfloor} \underset{\text{for odd } N}{(+0)} - \underbrace{\frac{1}{N} - \cdots - \frac{1}{N}}_{\lfloor N/2 \rfloor} = 0$
- Using $\hat{\mu}$ minimizes $\hat{\sigma}^2$: $\partial\hat{\sigma}^2/\partial\hat{\mu} = \frac{1}{(N-1)}\sum_i \partial(X_i - \hat{\mu})^2/\partial\hat{\mu} = \frac{2}{(N-1)}\left[\sum_i X_i - N\hat{\mu}\right] = 0$

**Practical calculation of the median estimator $\widehat{M}$**

1. Sorting the entire sample of $N$ values so that $X_1 \leq \cdots \leq X_N$
   obtains not only median but all the quantiles - Quicksort, Heapsort
   $\rightarrow$ number of operations scales as $N \ln(N)$
2. Direct selection of $(N/2)$th largest value
   $\rightarrow$ number of operations scales as $N$
3. Iterative procedure

$$\sum_{i=1}^{N} \frac{X_i - \widehat{M}}{\left|X_i - \widehat{M}\right|} = 0 \quad \Rightarrow \quad \widehat{M}_k = \frac{\sum_{i=1}^{N} \frac{X_i}{\left|X_i - \widehat{M}_{k-1}\right|}}{\sum_{i=1}^{N} \frac{1}{\left|X_i - \widehat{M}_{k-1}\right|}}$$

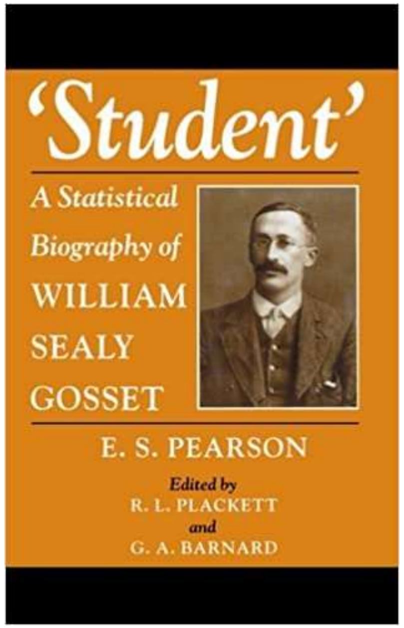
## 45. Are sample means of two sets of data significantly different from each other?

2 samples: $X_i \quad i = 1 \ldots N_X \quad \rightarrow$ sample mean $\hat{\mu}_X$
$\phantom{2 samples:}$ $Y_i \quad i = 1 \ldots N_Y \quad \rightarrow$ sample mean $\hat{\mu}_Y$

**Student's *t*:**

$$t = \frac{\hat{\mu}_X - \hat{\mu}_Y}{s_D}, \quad s_D = \sqrt{\frac{\sum_{i=1}^{N_X}(X_i - \hat{\mu}_X)^2 + \sum_{i=1}^{N_Y}(Y_i - \hat{\mu}_Y)^2}{N_X + N_Y - 2}\left(\frac{1}{N_X} + \frac{1}{N_Y}\right)}$$

$t$ measures the difference of the two sample means.




**William Sealy Gosset**
1876 – 1937
Guinness Head Brewer

Probability of obtaining a value of $|t|$ or larger randomly, with otherwise equal expected values of the underlying distributions of the two populations (null hypothesis):

$$P = 1 - A(t, N_X + N_Y - 2)$$

where $A(t, \nu)$ is the CDF of the Student's distribution with $\nu$ degrees of freedom.
Small $P < \alpha$ means that $\hat{\mu}_X$ and $\hat{\mu}_Y$ are significantly different
$\rightarrow$ reject the null hypothesis at the significance level of $\alpha$ (typically 0.05 or lower).

**Student's distribution**

Cumulative distribution function $A(t,\nu)$, one parameter $\nu$ – number of degrees of freedom:

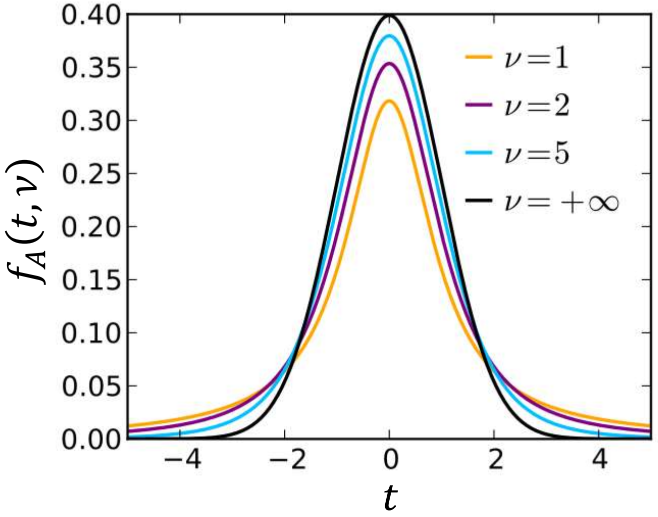
$$A(t,\nu) = \frac{1}{\sqrt{\nu}\, B\left(\frac{1}{2},\frac{\nu}{2}\right)} \int_{-t}^{t} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx$$

where $B(z,w)$ is the beta function

$$B(z,w) = \int_0^1 t^{z-1}\,(1-t)^{w-1} dt$$

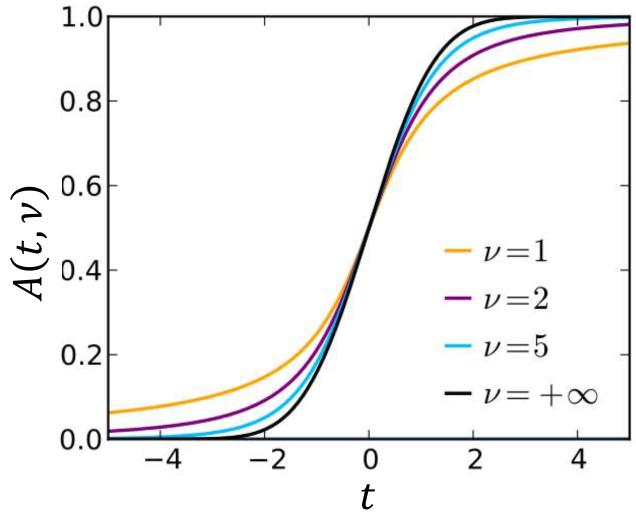
Probability density function:

$\nu \rightarrow \infty$ normal




Cumulative distribution function:

Finite $\nu \rightarrow$ heavy tail



## 46. Pearson's correlation coefficient: estimator of the linear correlation coefficient

**Recall:** $\rho(\xi_i, \xi_j) = \dfrac{\mathrm{cov}(\xi_i, \xi_j)}{\sqrt{D(\xi_i) D(\xi_j)}}$ for independent random variables: $\rho(\xi_i, \xi_j) = 0$

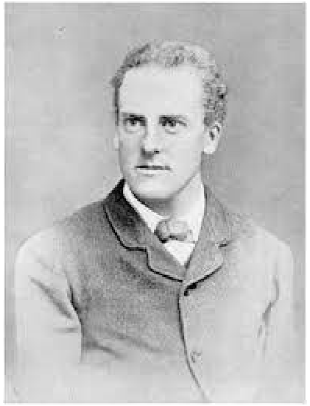
$\xi_i = a\xi_j, a > 0 : \rho(\xi_i, \xi_j) = 1 \qquad a < 0 : \rho(\xi_i, \xi_j) = -1$

2 samples of the same size: $X_i \quad i = 1 \dots N \quad \rightarrow$ sample mean $\hat{\mu}_X$

$Y_i \quad i = 1 \dots N \quad \rightarrow$ sample mean $\hat{\mu}_Y$

**Pearson's correlation coefficient:**

$$r = \frac{\sum_{i=1}^{N} (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)}{\sqrt{\sum_{i=1}^{N} (X_i - \hat{\mu}_X)^2 \ \sum_{i=1}^{N} (Y_i - \hat{\mu}_Y)^2}}$$


Karl Pearson
1857- 1936

Which value of $r$ means that we can reject the null hypothesis of independent populations $X$ and $Y$ ? At which significance level? **Approximate answer:** For large $N$ and normally distributed $X_i$ and $Y_i$ the values of $r$ are ≈ normally distributed with $\mathrm{E}(r) = 0, \quad \mathrm{D}(r) \approx 1/N$

**Less approximate answer for lower N:** $t = r\sqrt{\frac{N-2}{1-r^2}}$ has Student's t-distribution with $\nu = N - 2$.

## 47. Spearman's correlation coefficient: non-parametric estimator

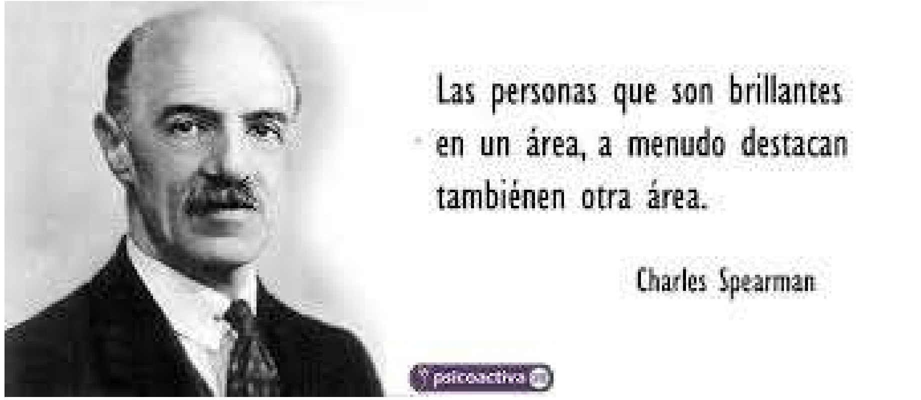
2 samples of the same size

$X_i \quad i = 1 \dots N$ → convert to ranks $R_i$. Example: $X$=[0.34,0.29,2.85] → $R$=[2,1,3]

$Y_i \quad i = 1 \dots N$ → ranks $S_i$. Randomly decide cases of equal values → $\Sigma = \frac{1}{2}N(N+1)$

**Spearman's correlation coefficient (robust, tests a monotonic relation) :**

$$r_s = \frac{\sum_{i=1}^{N}\left(R_i - \frac{N+1}{2}\right)\left(S_i - \frac{N+1}{2}\right)}{\sqrt{\sum_{i=1}^{N}\left(R_i - \frac{N+1}{2}\right)^2 \sum_{i=1}^{N}\left(S_i - \frac{N+1}{2}\right)^2}}$$



Charles Edward Spearman 1863- 1945, psychologist

For an arbitrarily distributed $X_i$ and $Y_i$, $t = r_s\sqrt{\frac{N-2}{1-{r_s}^2}}$ has ≈ Student's t-distribution with the number of degrees of freedom $\nu = N - 2$.

If $P = 1 - A(t, N-2) < \alpha$ then we can reject the null hypothesis of independent populations at the significance level of $\alpha$. $A(t,\nu)$ is the CDF of the Student's distribution.