Űrkommunikáció
Space Communication
2023/2.

Dr. János Bitó
bito.janos@vik.bme.hu
Dept. of Broadband Infocommunications and Electromagnetic Theory

The Entropy is bounded

Theorem: If the discrete random variable X has ***n*** possible values, then

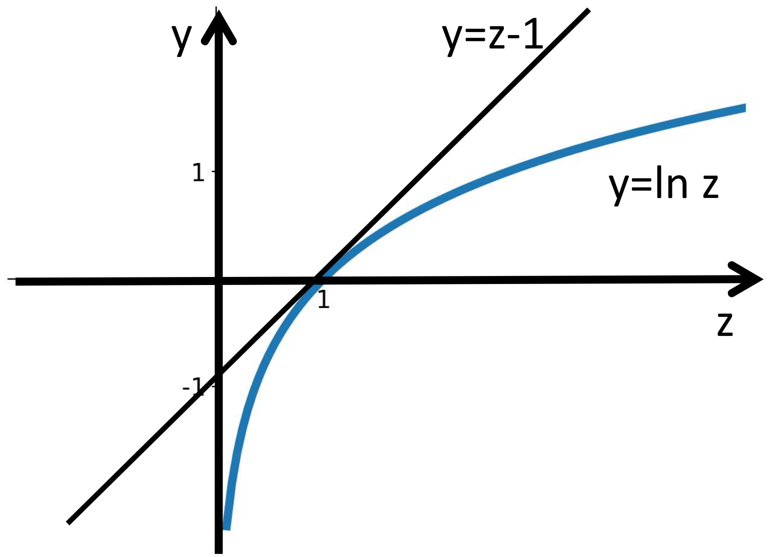
$$\mathbf{0 \le H(X) \le ld\, n = H_0(X)}$$

- Proof lower bound:

$$0 \le p(x_i) \le 1 \ \ \forall i$$

$$ld\, p(x_i) = \frac{1}{\ln 2} \ln p(x_i) \ \ \forall i$$

$$H(X) = -\sum_{i=1}^{n} p(x_i) \frac{1}{\ln 2} \ln p(x_i) \ge 0 \ \left[\frac{bit}{symbol}\right]$$



- Proof upper bound:

$$\mathrm{H}(X) \le ld\, n$$

$$\mathrm{H}(X) - ld\, n \le 0$$

$$\overbrace{\frac{1}{\ln 2}\sum_{i=1}^{n} p(x_i) \ln \frac{1}{p(x_i)}}^{H(x)} - \overbrace{\frac{1}{\ln 2}\sum_{i=1}^{n} p(x_i) \ln n}^{ld\, n} =$$

$$= \frac{1}{\ln 2}\sum_{i=1}^{n} p(x_i) \ln \underbrace{\frac{1}{n \cdot p(x_i)}}_{z} \le \frac{1}{\ln 2}\sum_{i=1}^{n} p(x_i)[z-1] = \frac{1}{\ln 2}\sum_{i=1}^{n} p(x_i)\left[\frac{1}{n \cdot p(x_i)} - 1\right] =$$

$$= \frac{1}{\ln 2}\left[\underbrace{\sum_{i=1}^{n} \frac{1}{n}}_{1} - \underbrace{\sum_{i=1}^{n} p(x_i)}_{1}\right] = 0$$

The Entropy ***H(X) has a maximum*** by $z = 1 = \frac{1}{n \cdot p(x_i)} \ \forall i$

$\boldsymbol{p(x_i) = 1/n}, \ \forall \boldsymbol{i}$ $\rightarrow$ ***Uniformly distributed random variable has maximum Entropy.***

Special case: Binary random variable

Binary random variable RV X, just two possibilities:

$$X = \{x_1 = 1, Yes, Black, True, \dots;\ x_2 = 0, No, White, False, \cdots\}$$

Discrete probability distribution function (PDF) is characterized by one parameter ***p***:

$$p(X) = \{p(x_1) = p;\ \ p(x_2) = 1 - p\}$$

Binary entropy function ***h(p):***

$$\mathrm{h}(p) = \sum_{i=1}^{2} p(x_i) \cdot ld\ \frac{1}{p(x_i)} = p \cdot ld\ \frac{1}{p} + (1-p) \cdot ld\ \frac{1}{(1-p)}\ \left[\frac{bit}{binary\ symbol}\right]$$

- Maximum of h(p) at uniform distribution:

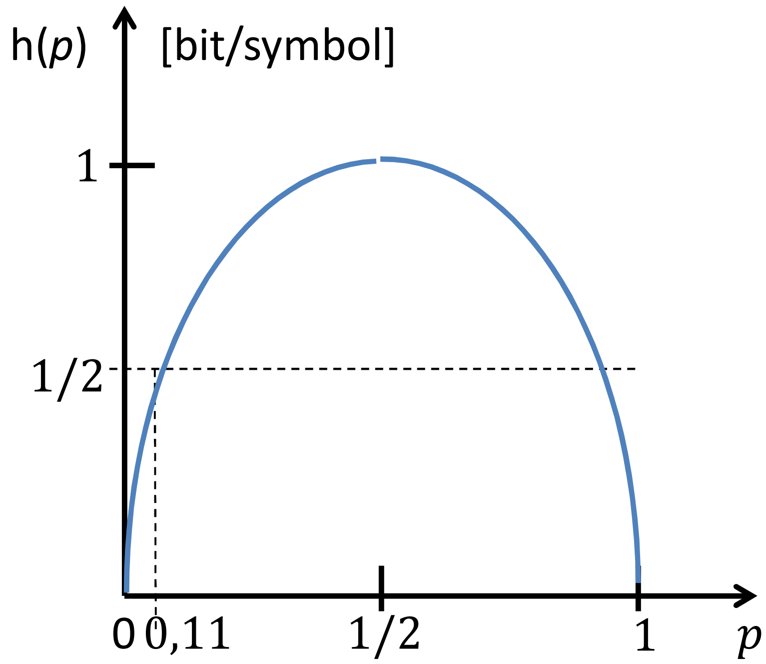
$$\mathrm{h}(p = 1/2) = 1\left[\frac{bit}{binary\ symbol}\right]$$

- If $p \to 0$

$$\lim_{p\to 0} p \cdot ld\ \frac{1}{p} = 0; \lim_{p\to 0}(1-p) \cdot ld\ \frac{1}{(1-p)} = ld\ 1 = 0$$

- If $p \to 1$

$$\lim_{p\to 1} p \cdot ld\ \frac{1}{p} = 0; \lim_{p\to 1}(1-p) \cdot ld\ \frac{1}{(1-p)} = 0$$



Recap: Probability Theory

- A branch of mathematics concerned with the analysis of ***random phenomena***.
- Def. ***Random Variable (RV)***: The **outcome** of a random event ***cannot be determined*** before it occurs, but it may be **any one of several** (could be infinite) **possible outcomes**.

Discrete Random Variable

$X = \{x_1, x_2, \cdots, x_n\}$

Continuous Random variable

$X = \{x \in [x_{min}, x_{max}]\}$

Cumulative Distribution Function (CDF)

$$P_X(x) = Prob(X \le x) = \sum_{i=1}^{x_i \le x} p(x_i)$$

$$F_X(x) = Prob(X \le x) = \int_{z=-\infty}^{x} f_X(z)\, dz$$

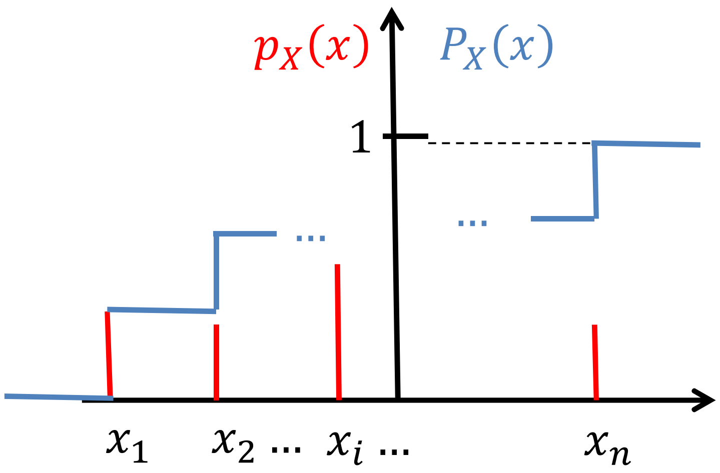
Probability Density Function (PDF)

$$p_X(x) = \{p(x_1), p(x_2), \cdots, p(x_n)\}$$

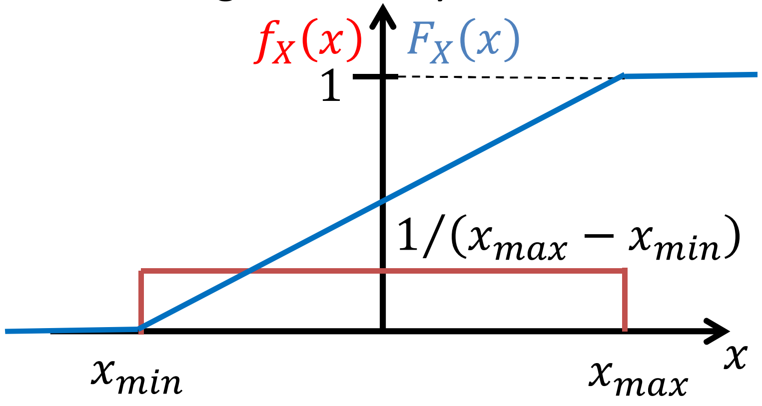
$$\frac{d}{dx} F_X(x) = f_X(x)$$

$$Prob(a < X \le b) = P_X(b) - P_X(a)$$

$$Prob(a < X \le b) = F_X(b) - F_X(a) = \int_{z=a}^{b} f_X(z)\, dz$$



e.g. Uniformly distributed RV



Recap: Probability Theory

Discrete Random Variable | **Continuous Random variable**

1st moment of a RV, **Expected value**, Mean value, $E\{X\} = \mu_1(X) = \mu_x$

$$\sum_{i=1}^{n} x_i \cdot p(x_i) \qquad\qquad \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

kth moment of a RV, $E\{X^k\} = \mu_k(X)$

$$\sum_{i=1}^{n} x_i^k \cdot p(x_i) \qquad\qquad \int_{-\infty}^{\infty} x^k \cdot f_X(x)\, dx$$

2nd central moment, **Variance**, $Var(X) = \sigma_x^2$

$$Var(X) = E\{(X - \mu_x)^2\} = E\{X^2\} - (E\{X\})^2 = \mu_2(X) - \mu_x^2 = \mu_2(X) - \mu_1^2(X)$$

$$\sum_{i=1}^{n} (x_i - \mu_x)^2 \cdot p(x_i) \qquad\qquad \int_{-\infty}^{\infty} (x - \mu_x)^2 \cdot f_X(x)\, dx$$

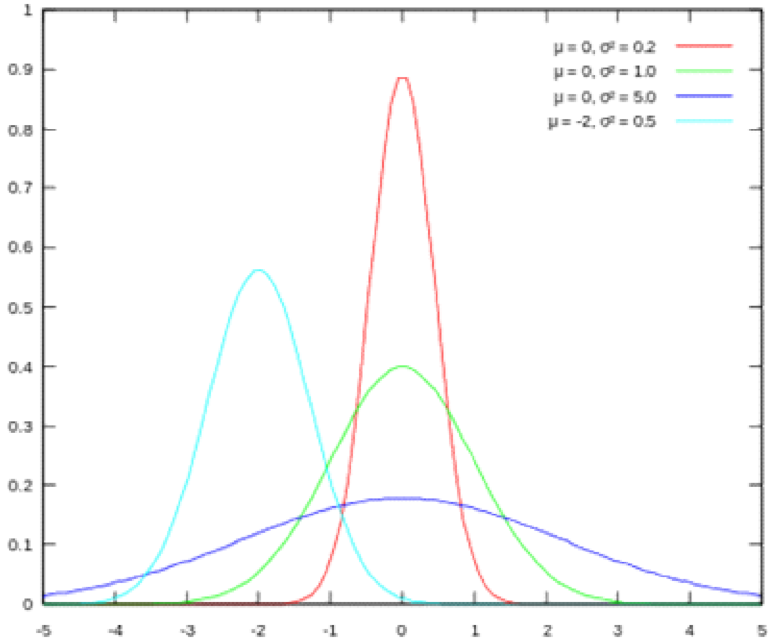
PDF Example: Normal (Gaussian) distribution

(first order = one dimensional):

$$f_x^{(1)}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)$$

Standard Gaussian distribution

$$G(\mu_x = 0, \sigma_x^2 = 1)$$



Entropy of Continuous Random Variable, Differential Entropy

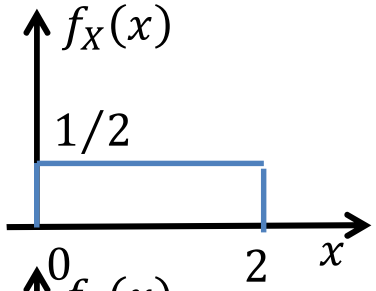
The Entropy ***H(X)*** of the continuous RV ***X*** with PDF $\boldsymbol{f_X(x)}$ is called as **Differential Entropy** and defined as:

$$H(X) = E\left\{ld\ \frac{1}{f_X(x)}\right\} = -E\{ld f_X(x)\} = -\int_{\mathcal{X}} f_X(x) \cdot ld f_X(x)\, dx$$

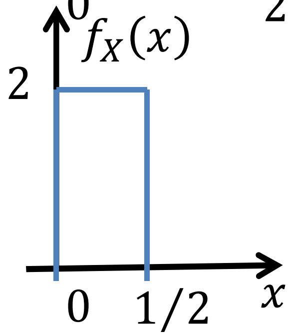
where $\mathcal{X}$ denotes the set of values for which $\boldsymbol{f_X(x) > 0}$.

This is an ***extension of entropy for a discrete RV, however, lacks the same physical meaning (not guaranteed to be positive).*** Fortunately, mutual information I(X;Y) (See later) for continuous RV's X and Y can be considered as a measure of reduction of uncertainty.

Examples:



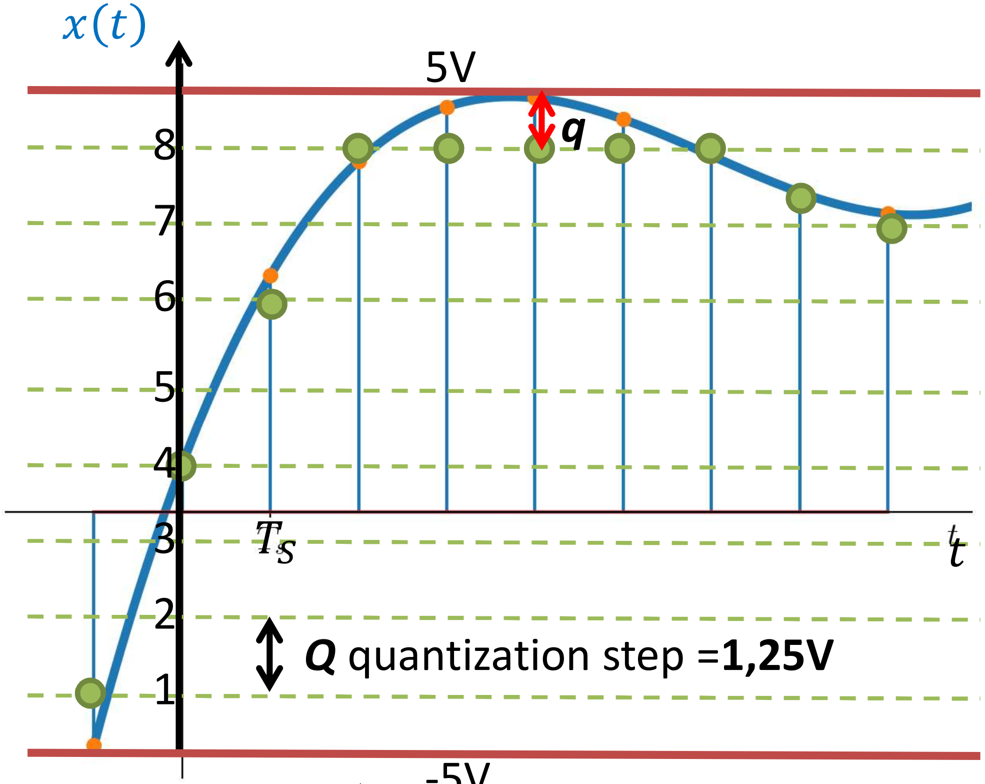
$$H(X) = \int_0^2 \frac{1}{2} \cdot ld2dx = \frac{1}{2}\int_0^2 1dx = \frac{1}{2} \cdot [x]_0^2 = 1$$

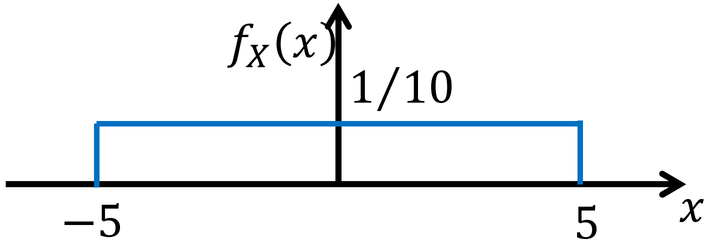


$$H(X) = \int_0^{1/2} 2 \cdot ld\frac{1}{2}dx = -2\int_0^{\frac{1}{2}} 1dx = -2 \cdot [x]_0^{1/2} = -1$$

Entropy of Continuous Random Variable, Differential Entropy

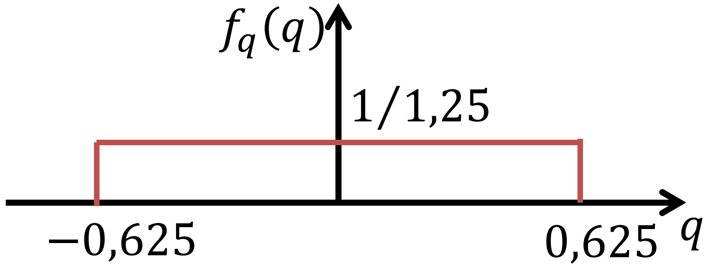
Example: **Linear quantization** of **analog random voltage function x(t)** with ***uniform*** value distribution in the range [-5V … +5V] applying **8 quantization levels**.





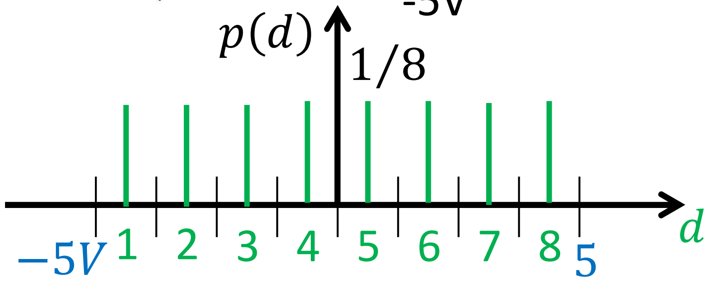
$$H(X) = \int_{-5}^{5} \frac{1}{10} \cdot ld10dx = = \frac{ld10}{10} \cdot \int_{-5}^{5} 1dx$$

$$= ld10 \approx 3{,}322$$



$$H(q) = \int_{-0{,}625}^{0{,}625} \frac{1}{1{,}25} \cdot ld1{,}25dx =$$

$$= \frac{ld1{,}25}{1{,}25} \cdot \int_{-0{,}625}^{0{,}625} 1dx = ld1{,}25 \approx 0{,}322$$



$$\mathrm{H}(X) = \sum_{i=1}^{8} p(d_i) \cdot ld\,\frac{1}{p(d_i)} = \sum_{i=1}^{8} \frac{1}{8} \cdot ld\,8 = ld\,8 = 3\,\left[\frac{bit}{Symbol}\right]$$

Entropy of Continuous Random Variable, Differential Entropy

Arkhimédész (with Greek-letters: Αρχιμήδης)
„Heuréka!", got it!

Really?

uniform value distribution in the range [-5V … +5V] $\mathrm{H}(X) = ld10 = 3{,}32192809...$

Linear quantization applying ***n*** quantization levels:

n	$H_d(X) = ld\, n \ \frac{bit}{symbol}$	$Q = 10/n$	$H(q) = ld\ Q$	$H_d(X) + H(q)$
4	2	2,5	1,32192809...	3,32192809...
8	3	1,25	0,32192809...	3,32192809...
16	4	0,625	-0,6780719...	3,32192809...

For discrete RV	For continuous RV	
$0 < p(x_i) \leq 1 \ \ \forall i$	$0 < f_X(x) \leq 1$	$1 < f_X(x)$
$1 \leq \frac{1}{p(x_i)} < \infty \ \ \forall i$	$1 \leq \frac{1}{f_X(x)} < \infty$	$0 \leq \frac{1}{f_X(x)} < 1$
$0 \leq ld \frac{1}{p(x_i)} \ \ \forall i$	$0 \leq ld \frac{1}{f_X(x)}$	$ld \frac{1}{f_X(x)} < 0$

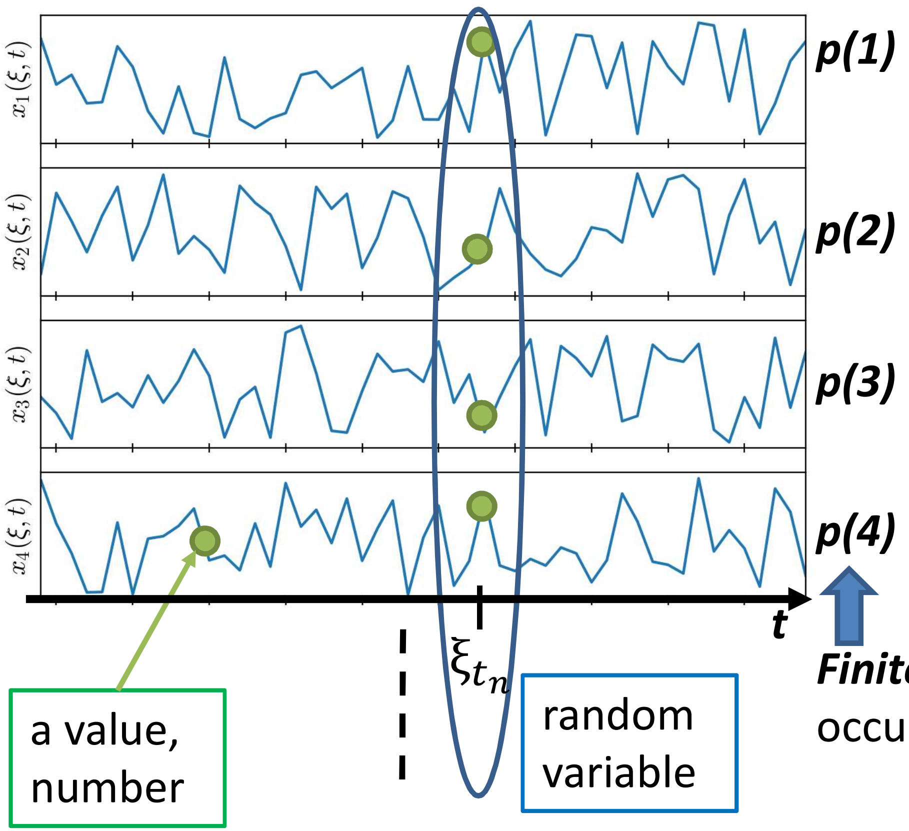
Fortunately, mutual information for continuous RV's can be considered as a measure!

Stochastic processes ξ

- We need to transmit/store not just one outcome of a random variable, but a series of such outcomes.
- Our ***information sources*** generate the realizations of ***stochastic, random time/space functions*** called as ***Stochastic Processes*** ξ .

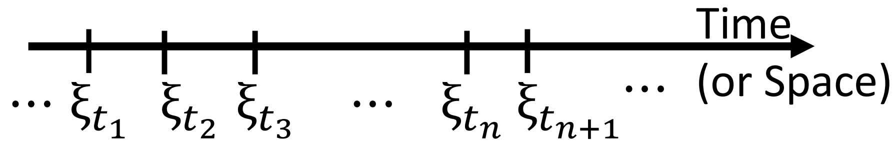
There are two common ***interpretations*** (and a third one):

The whole (infinite) ***set of realizations***.



Finite set of possible realizations occurs with given probabilities.

Infinite series of RV
ordered in time (or space).



Mathematical description of stochastic processes

- Cumulative Distribution Function (CDF) n-th order

$$F_{\xi}^{(n)}(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = F_{\xi}^{(n)}(\bar{x}, \bar{t}) = Prob\left(\xi_{t_1} \le x_1, \xi_{t_2} \le x_2, \dots, \xi_{t_n} \le x_n\right)$$

Joint probability

- Probability Density Function (PDF) n-th order

$$f_{\xi}^{(n)}(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = f_{\xi}^{(n)}(\bar{x}, \bar{t}) = \frac{\delta^n}{\delta x_1 \delta x_2 \dots \delta x_n} F_{\xi}^{(n)}(\bar{x}, \bar{t})$$

- Expected value function - ensemble averages not time averages (if exist)

$$m_{\xi}(t) = E\{\xi_t\} = \int_{-\infty}^{\infty} x \cdot f_{\xi}^{(1)}(x, t) dx$$

- Instantaneous Power: $P_{\xi}(t) = E\left\{{\xi_t}^2\right\} = \int_{-\infty}^{\infty} x^2 \cdot f_{\xi}^{(1)}(x, t) dx$
- Autocorrelation: $R_{\xi}(t_1, t_2) = E\left\{\xi_{t_1} \cdot \xi_{t_2}\right\} = \iint x_1 \cdot x_2 \cdot f_{\xi}^{(2)}(x_1, x_2, t_1, t_2) dx_1 dx_2$
- Covariance:

$$K_{\xi}(t_1, t_2) = E\left\{(\xi_{t_1} - m_{\xi}(t_1)) \cdot (\xi_{t_2} - m_{\xi}(t_2))\right\} =$$

$$= \iint (x_1 - m_{\xi}(t_1)) \cdot (x_2 - m_{\xi}(t_2)) \cdot f_{\xi}^{(2)}(x_1, x_2, t_1, t_2) dx_1 dx_2$$

Remark: If $m_{\xi}(t_1) = m_{\xi}(t_2) = 0$ then $R_{\xi}(t_1, t_2) = K_{\xi}(t_1, t_2)$

Mathematical description of stochastic processes

- **D=n** dimensional **expected value vector** $\bar{\boldsymbol{m}}_{\xi}(\bar{\boldsymbol{t}})$ defined at a given time point vector $\bar{t} = [t_1, t_2, \dots, t_n]$:

$$\bar{m}_{\xi}(\bar{t}) = \left[m_{\xi}\ (t_1), m_{\xi}\ (t_2), \dots, m_{\xi}\ (t_n)\right]$$

- **Covariance matrix**: quadratic n x n matrix of covariance $K_{\xi}\ (t_i, t_j)$ values of any two time (or space) points t_i, t_j from a time point vector $\bar{t}$:

$$\overline{\overline{K_{\xi}}}(\bar{t}) = \begin{bmatrix} K_{\xi}\ (t_1, t_1) & \cdots & K_{\xi}\ (t_1, t_n) \\ \vdots & \ddots & \vdots \\ K_{\xi}\ (t_n, t_1) & \cdots & K_{\xi}\ (t_n, t_n) \end{bmatrix}$$

Main diagonal: Instantaneous power $\boldsymbol{P}_{\xi}\ (\boldsymbol{t}) = E\left\{{\xi_t}^2\right\}$ values at the time point vector $\bar{t}$ (for processes without direct components that is the expected values $m_{\xi}\ (t)$ at any time point of $\bar{t}$ are zero i.e. the expected value vector $\bar{\boldsymbol{m}}_{\xi}(\bar{\boldsymbol{t}})$ is the $\bar{\mathbf{0}}$ vector.

- Example: **D=n dimensional normal (Gaussian)** distribution (n-th order PDF) of a process with $\bar{\boldsymbol{m}}_{\xi}(\bar{\boldsymbol{t}})$ and $\overline{\overline{\boldsymbol{K}_{\xi}}}(\bar{\boldsymbol{t}})$ – notation determinant $\|M\|$, inverse M^{-1}, transpose M^T of matrix M:

$$f_{\xi}^{(n)}(\bar{x}, \bar{t}) = \frac{1}{\sqrt[2]{(2\pi)^n \left\|\overline{\overline{K_{\xi}}}(\bar{t})\right\|}} exp\left(-\frac{1}{2} \cdot \left[\bar{x} - \bar{m}_{\xi}(\bar{t})\right] \cdot \overline{\overline{K_{\xi}}}(\bar{t})^{-1} \cdot \left[\bar{x} - \bar{m}_{\xi}(\bar{t})\right]^T\right)$$

D dimensional normal (Gaussian) stochastic processes

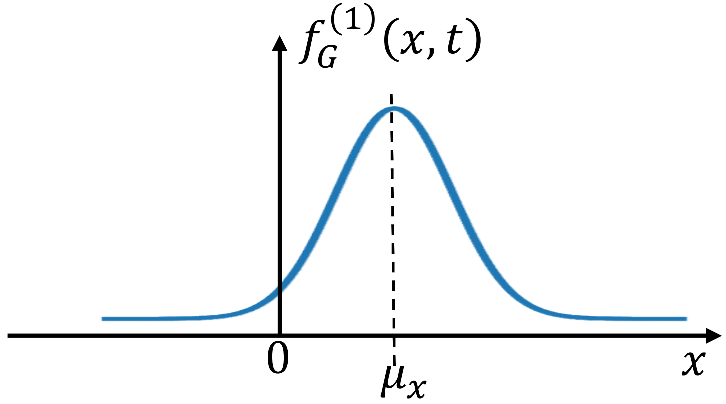
- ***D=n***: $f_{\xi}^{(n)}(\bar{x},\bar{t}) = \frac{1}{\sqrt[2]{(2\pi)^n \|\overline{\overline{K_\xi}}(\bar{t})\|}} exp\left(-\frac{1}{2}\cdot[\bar{x}-\bar{m}_\xi(\bar{t})]\cdot\overline{\overline{K_\xi}}(\bar{t})^{-1}\cdot[\bar{x}-\bar{m}_\xi(\bar{t})]^T\right)$

- ***D=1***, we investigate the process at a given time point ***t***

Expected value: $\mu_x = m_\xi\ (t) = E\{\xi_t\}$

Variance: $\sigma_x^2 = E\{(\xi_t - \mu_x)^2\} = K_\xi\ (t,t) \overset{\mu_x=0}{\Longrightarrow} P_\xi\ (t)$

PDF: $f_G^{(1)}(x,t) = \frac{1}{\sqrt{2\pi\sigma_x^2}} exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)$



- ***D=2***, we investigate the process at given time points $\bar{t} = [\boldsymbol{t_1}, \boldsymbol{t_2}]$ (notation $\bar{x} = [x_1, x_2]$):

Expected value vector: $\bar{m}_\xi(\bar{t}) = [m_\xi\ (t_1), m_\xi\ (t_2)]$

✓ If the ensemble average time invariant and zero (no direct component): $\boldsymbol{\bar{m}_\xi(\bar{t}) = \bar{0}}$

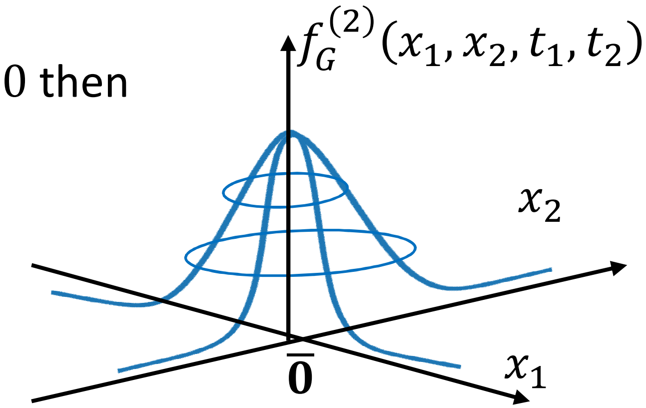
✓ If the instantaneous power time invariant:

$P_\xi\ (t_1) = P_\xi\ (t_2) = K_\xi\ (t_1,t_1) = K_\xi\ (t_2,t_2) = \sigma_x^2,$

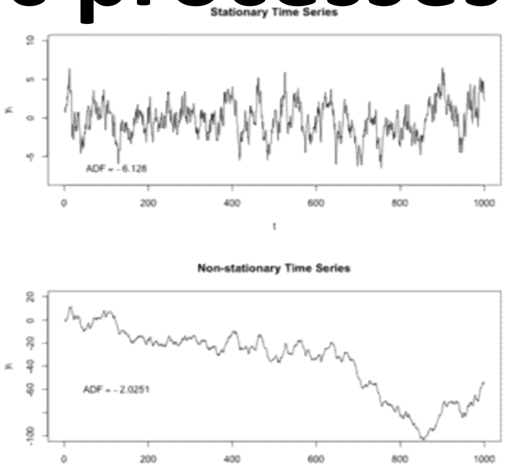
✓ and the process is not correlated $K_\xi\ (t_1,t_2) = K_\xi\ (t_2,t_1) = 0$ then

Covariance matrix: $\overline{\overline{K_\xi}}(\bar{t}) = \begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_x^2\end{bmatrix}$

PDF: $f_G^{(2)}(\bar{x},\bar{t}) = \frac{1}{2\pi\sigma_x^2} exp\left(-\frac{|\bar{x}|^2}{2\sigma_x^2}\right)$



Stationarity of stochastic processes



In the most intuitive sense, stationarity means that the statistical properties of a process generating a time series do not change over time.

- **Stationarity in n-th order**

$$F_{\xi}^{(n)}(x_1, x_2, \ldots, x_n, t_1, t_2, \ldots, t_n) = F_{\xi}^{(n)}(x_1, x_2, \ldots, x_n, t_1 + \Delta t, t_2 + \Delta t, \ldots, t_n + \Delta t)$$

$$F_{\xi}^{(n)}(\bar{x}, \bar{t}) = F_{\xi}^{(n)}(\bar{x}, \bar{t} + \Delta t) \text{ for any } \Delta t \text{ and any set of } \bar{t} = [t_1, t_2, \ldots, t_n]$$

This means that the distribution of a finite sub-sequence of random variables of the stochastic process remains the same as we shift it along the time index axis.

- **Strict-sense Stationarity** or strong-sense Stationarity (or simply Stationarity)

If the process is stationary in n-th order for any n, even if $\mathrm{n} \rightarrow \infty$

$$F_{\xi}^{(n)}(\bar{x}, \bar{t}) = F_{\xi}^{(n)}(\bar{x}, \bar{t} + \Delta t) \; holds \; for \; \forall \, \Delta t, \forall \, \bar{t}, \forall \, n$$

- **Wide Sense Stationarity (WSS)** or weak-sense stationarity, covariance stationarity

WSS only requires the shift-invariance (in time) of the first moment and the cross moment. This means the process has the same mean at all time points, and that the covariance between the values at any two time points, depend only on the difference between the two times.

Stationarity of stochastic processes

- **Wide Sense Stationarity (WSS)** or weak-sense stationarity, covariance stationarity

✓ Expected value function is constant

$$m_\xi(t) = m_\xi(t_0) = m_\xi \; holds \; for \; \forall \, t$$

✓ Covariance:

$$K_\xi(t, t+\tau) = E\left\{(\xi_t - m_\xi) \cdot (\xi_{t+\tau} - m_\xi)\right\} = K_\xi(t+\Delta t, t+\Delta t+\tau) = K_\xi(\tau)$$
$$holds \; for \; \forall \, t, \forall \, \Delta t$$

Or similarly because m_ξ constant:

✓ Autocorrelation:

$$R_\xi(t, t+\tau) = E\{\xi_t \cdot \xi_{t+\tau}\} = R_\xi(t+\Delta t, t+\Delta t+\tau) = R_\xi(\tau)$$
$$holds \; for \; \forall \, t, \forall \, \Delta t$$

- Remark: Second order stationarity vs. WSS:

2nd order Stationarity **=>** WSS however WSS ≠**>** 2nd order Stationarity

$$R_\xi(t_1, t_2) = E\left\{\xi_{t_1} \cdot \xi_{t_2}\right\} = \iint x_1 \cdot x_2 \cdot f_\xi^{(2)}(x_1, x_2, t_1, t_2) dx_1 dx_2$$

Examples: http://www.hit.bme.hu/~dallos/hirkelm/Sztfoly_exmp.pdf

Stationarity of stochastic processes