# 2. Reviews

- linear systems: state-space equations
- random (stochastic) processes
- hypothesis tests

# Continous-time systems

- an autonomous system

$$\dot{x}(t) = Ax(t), \quad y = Cx(t)$$

- a system with inputs

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y = Cx(t) + Du(t)$$

- $x \in \mathbf{R}^n$ is called state, $y \in \mathbf{R}^m$ is called output, and $u \in \mathbf{R}^p$ is the control input
- $A \in \mathbf{R}^{n \times n}$ is the dynamic matrix
- $B \in \mathbf{R}^{p \times n}$ is the input matrix
- $C \in \mathbf{R}^{m \times n}$ is the output matrix
- $D \in \mathbf{R}^{m \times p}$ is the direct forward term

# Solution of state-space equations

- an autonomous system

$$x(t) = e^{At}x(0), \quad y = Ce^{At}x(0)$$

$e^{At}$ is the state-transition matrix; can be computed analytically

- a system with inputs

$$\begin{aligned} x(t) &= e^{tA}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau, \\ y &= Ce^{At}x(0) + C\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau + Du(t) \end{aligned}$$

$x(t)$ consists of zero-input response and zero-state response

# Discrete-time systems

- an autonomous system

$$x(t+1) = Ax(t), \quad y(t) = Cx(t)$$

with solution

$$x(t) = A^t x(0)$$

- a system with inputs

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)$$

with solution

$$x(t) = A^t x(0) + \sum_{\tau=0}^{t-1} A^{t-1-\tau} Bu(\tau)$$

# Transfer function of linear systems

explains a relationship from $u$ to $y$

- continuous-time system: $Y(s) = H(s)U(s)$

$$H(s) = C(sI - A)^{-1}B + D$$

- discrete-time system: $Y(z) = H(z)U(z)$

$$H(z) = C(zI - A)^{-1}B + D$$

the inverse Laplace ($z$-) transform of $H$ is the impulse response, $h(t)$

# Important concepts of system analysis

- stability: if $x(t) \to 0$ when $t \to \infty$
  (eigenvalues of dynamic matrix, Lyapunov theory)
- controllability: how a target state can be acheived by applying a certain input
  (explained from $A$ and $B$)
- observability: how to estimate $x(0)$ from the measurement $y$
  (explained $A$ and $C$)

# Stochastic Signals

- stationary processes
- ergodic processes
- correlation and covariance function
- power spectral density
- independent and uncorrelated processes
- Gaussian or normal processes
- white noise
- linear process with stochastic signals

# Stochastic Processes

**stochastic process** is an entirely family (*ensemble*) of random time signals

$$\{x(t), \quad t \in T\}$$

$i.e.$, for each $t$ in the index set $T$, $x(t)$ is a random variable

- a signal realization $x(t)$ is called *sample function* or a *sample path*
- if $T$ is a countable set, $x(t)$ is called **discrete-time** stochastic process
- if $T$ is a continuum, $x(t)$ is called **continuous-time** process
- a process can be either discrete- or continuous-valued

# Joint distribution

let $x_1, \ldots, x_n$ be the $n$ random variables by sampling the process $x(t)$

$$x_1 = x(t_1), \quad x_2 = x(t_2), \ldots, \quad x_n = x(t_n)$$

a stochastic process is specified by the collection of joint cdf (depend on time)

$$F(x_1, x_2, \ldots, x_n) = P(x(t_1) \leq x_1, \quad x(t_2) \leq x_2, \ldots, \quad x(t_n) \leq x_n)$$

- continuous-valued process:

$$\begin{aligned} f(x_1, \ldots, x_n) dx_1 \cdots dx_n = \\ P(x_1 < x(t_1) \leq x_1 + dx_1, \ldots, x_n < x(t_n) \leq x_n + dx_n) \end{aligned}$$

- discrete-valued process:

$$p(x_1, x_2, \ldots, x_n) = P(x(t_1) = x_1, x(t_2) = x_2, \ldots, x(t_n) = x_n)$$

# Mean and variance of stochastic process

mean and variance function of a continous-time process are defined by

$$\begin{aligned}\mu(t) &= \mathbf{E}[x(t)] = \int_{-\infty}^{\infty} x f(x) dx \\ \mathbf{var}[x(t)] &= \int_{-\infty}^{\infty} (x - \mu(t))^2 f(x) dx\end{aligned}$$

- here $f$ is the pdf of $x(t)$ (depend on time)
- mean and variance are *deterministic* functions of time

# Correlation and Covariance

suppose $X, Y$ are random variables with means $\mu_x$ and $\mu_y$ resp.

**cross correlation**

$$R_{xy} = \mathbf{E}[XY^T]$$

**autocorrelation**

$$R = \mathbf{E}[XX^T]$$

**cross covariance**

$$C_{xy} = \mathbf{E}\left[(X - \mu_x)(Y - \mu_y)^T\right]$$

**autocovariance**

$$C = \mathbf{E}\left[(X - \mu_x)(X - \mu_x)^T\right]$$

**correlation = covariance** when considering zero mean

# Correlation and Covariance functions

suppose $x(t), y(t)$ are random processes

**cross correlation**

$$R_{xy}(t_1, t_2) = \mathbf{E}x(t_1)y(t_2)^T$$

**autocorrelation**

$$R(t_1, t_2) = \mathbf{E}x(t_1)x(t_2)^T$$

**cross covariance**

$$C_{xy}(t_1, t_2) = \mathbf{E}\left[(x(t_1) - \mu_x(t_1))(y(t_2) - \mu_y(t_2))^T\right]$$

where $\mu_x(t) = \mathbf{E}x(t)$ and $\mu_y(t) = \mathbf{E}y(t)$

**autocovariance**

$$C(t_1, t_2) = \mathbf{E}\left[(x(t_1) - \mu(t_1))(x(t_2) - \mu(t_2))^T\right]$$

# Stationary processes

a process is called **strictly stationary** if the joint cdf of

$$x(t_1), x(t_2), \ldots, x(t_n)$$

is *the same as* that of

$$x(t_1 + \tau), x(t_2 + \tau), \ldots, x(t_n + \tau)$$

for *all time shifts* $\tau$ and for *all choices of sample times* $t_1, \ldots, t_k$

- first-order cdf of a stationary process must be independent of time

$$F_{x(t)}(x) = F_{x(t+\tau)}(x) = F(x), \quad \forall t, \tau$$

  implication: mean and variance are **constant** and **independent** of time

# Wide-sense stationary Process

a process is **wide-sense** stationary if the two conditions hold:

1. $\mathbf{E}x(t) =$ constant for all $t$
2. $R(t_1, t_2) = R(t_1 - t_2)$ (only depends on the time gap)

the correlation/covariance functions are simplifed to

$$R(\tau) = \mathbf{E}x(t+\tau)x(t)^T, \qquad R_{xy}(\tau) = \mathbf{E}x(t+\tau)y(t)^T$$

$$C(\tau) = \mathbf{E}x(t+\tau)x(t)^T - \mu_x\mu_x^T, \qquad C_{xy}(\tau) = \mathbf{E}x(t+\tau)y(t)^T - \mu_x\mu_y^T$$

# Example

determine the mean and the autocorrelation of a random process

$$x(t) = A\cos(\omega t + \phi)$$

where the random variables $A$ and $\phi$ are independent and $\phi$ is uniform on $(-\pi, \pi)$

since $A$ and $\phi$ are independent, the mean is given by

$$\mathbf{E}x(t) = \mathbf{E}[A]\mathbf{E}[\cos(\omega t + \phi)]$$

using the uniform distribution in $\phi$, the last term is

$$\mathbf{E}\cos(\omega t + \phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\omega t + \phi)d\phi = 0$$

therefore, $\mathbf{E}x(t) = 0$

using trigonometric identities, the autocorrelation is determined by

$$\mathbf{E}x(t_1)x(t_2) = \frac{1}{2}\mathbf{E}A^2\mathbf{E}[\cos\omega(t_1 - t_2) + \cos(\omega t_1 + \omega t_2 + 2\phi)]$$

since

$$\mathbf{E}[\cos(\omega t_1 + \omega t_2 + 2\phi)] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(\omega t_1 + \omega t_2 + 2\phi)d\phi = 0$$

we have

$$R(t_1, t_2) = (1/2)\mathbf{E}[A^2]\cos\omega(t_1 - t_2)$$

hence, the random process in this example is wide-sense stationary

# Power Spectral Density

**Wiener-Khinchin Theorem**: if a process is wide-sense stationary, the autocorrelation function and the power spectral density form a Fourier transform pair:

$$S(\omega) = \int_{-\infty}^{\infty} e^{-\mathrm{i}\omega\tau} R(\tau) d\tau \qquad \text{continuous}$$

$$S(\omega) = \sum_{k=-\infty}^{k=\infty} R(k) e^{-\mathrm{i}\omega k} \qquad \text{discrete}$$

the autocorrelation function at $\tau = 0$ indicates the average power:

$$R(0) = \mathbf{E}[x(t)x(t)^T] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega$$

(similarly, use discrete inverse Fourier transform for discrete systems)

# Properties

- $R(-t) = R(t)^T$ (if the process is scalar, then $R(-t) = R(t)$)
- non-negativity: that is for any $a_i, a_j \in \mathbf{R}^n$, with $i, j = 1, \ldots N$, we have

$$\sum_i^N \sum_j^N a_i^T R(i-j) a_j \geq 0,$$

which follows from

$$\sum_i^N \sum_j^N a_i^T R(i-j) a_j = \sum_i^N \sum_j^N \mathbf{E}[a_i^T x(i) x(j)^T a_j] = \mathbf{E}\left[\left(\sum_i^N a_i^T x(i)\right)^2\right] \geq 0.$$

- $S(\omega)$ is self-adjoint, *i.e.*, $S(\omega)^* = S(\omega)$ for all $\omega$
- diagonals of $S(\omega)$ are real-valued

# Ergodic Processes

a stochastic process is *ergodic* if

$$\mathbf{E}[x(t)] = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)dt \qquad \text{(continuous)}$$

$$\mathbf{E}[x(t)] = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k) \qquad \text{(discrete)}$$

(time average $=$ ensemble average)

- one typically gets statistical information from emsemble averaging
- ergodic hypothesis means this information can also be obtained from averaging a single sample $x(t)$ over time

with ergodic assumption,

**continous time**

$$
\begin{aligned}
R(\tau) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t+\tau)x(t)^T dt \\
R_{xy}(\tau) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t+\tau)y(t)^T dt
\end{aligned}
$$

**discrete time**

$$
\begin{aligned}
R(\tau) &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k+\tau)x(k)^T \\
R_{xy}(\tau) &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} x(k+\tau)y(k)^T
\end{aligned}
$$

# Independent and Correlated Processes

stationary processes $x(t)$ and $y(t)$ are called **independent** if

$$f(x, y) = f(x)f(y)$$

(the joint pdf is equal to the product of marginals)

and are called **uncorrelated** if

$$C_{xy}(\tau) = 0, \quad \forall \tau$$

- independent processes are always uncorrelated
- the opposite may not be true

# White noise

a zero-mean process with the following properties:

**continuous time**

$$R(\tau) = S_0 \delta(\tau), \qquad S(\omega) = \int_{-\infty}^{\infty} S_0 \delta(\tau) e^{-i\omega\tau} d\tau = S_0$$

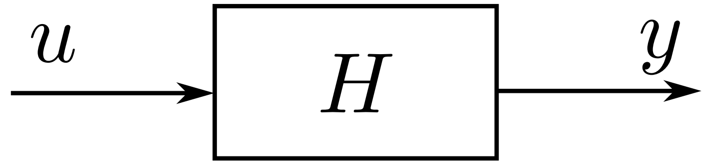
**discrete time**

$$R(k) = S_0 \delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}, \qquad S(\omega) = \sum_{-\infty}^{\infty} S_0 \delta(\tau) e^{-i\omega k} d\tau = S_0$$

(constant spectrum)

# Linear systems with random input

let $y$ be the response to input $u$ under a linear causal system $H$




**Facts:** if $u(t)$ is a wide-sense stationary process and $H$ is stable then

- $y(t)$ is also a wide-sense stationary process
- spectrum of $u$ and $y$ are related by

$$S_y(\omega) = H(\omega) S_u(\omega) H(\omega)^*$$

where $H(\omega)^*$ is the complex conjugate transpose of $H(\omega)$

# Random walk

a process $x(t)$ is a random walk if

$$x(t) = x(t-1) + w(t-1)$$

where $w(t)$ is a white noise with covariance $\Sigma$

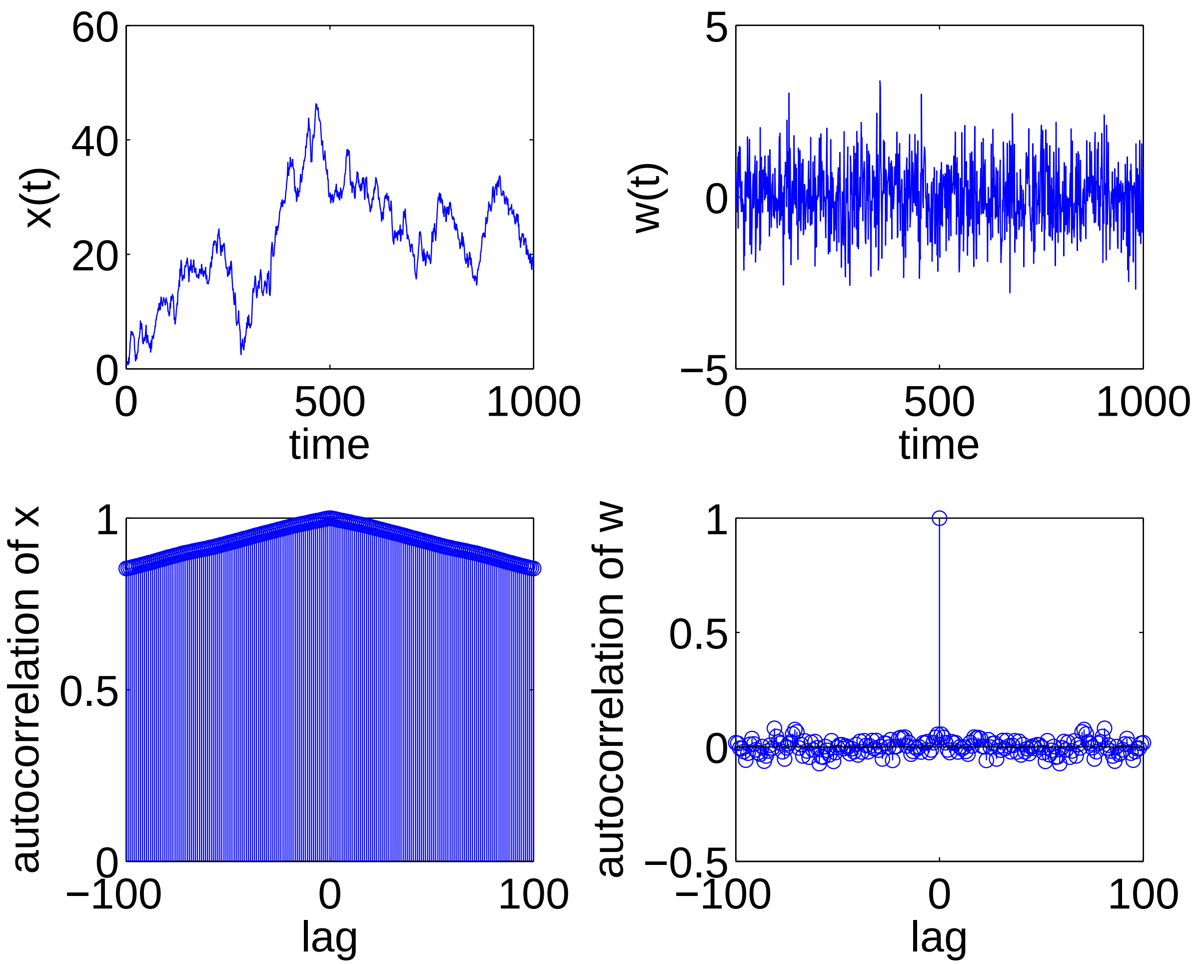
- $x(t)$ obeys a linear (unstable) system with a random input
- with back substitution, we can express $x(t)$ as

$$x(t) = w(t) + w(t-1) + \cdots + w(1)$$

- $x(t)$ is *non-stationary* because $R(t, t+\tau)$ depends on $t$

$$R(t, t+\tau) = \mathbf{E}[x(t)x(t+\tau)^T] = t\Sigma$$

time plot of random walk and its normalized autocorrelation (correlogram)




correlogram of $x$ gradually decays

# Reviews

- linear systems: state-space equations
- random (stochastic) processes
- **hypothesis tests**

# Hypothesis tests

elements of statistical tests

- null hypothesis, alternative hypothesis
- test statistics
- rejection region
- type of errors: type I and type II errors
- confidence intervals, $p$-values

examples of hypothesis tests:

- hypothesis tests for the mean, and for comparing the means
- hypothesis tests for the variance, and for comparing variances

# Testing procedures

a test consists of

- providing a statement of the hypotheses ($H_0$ (null) and $H_1$ (alternative))
- giving a rule that dictates if $H_0$ should be rejected or not

the decision rule involves a test statistic calculated on observed data

the Neyman-Pearson methodology partitions the sample space into two regions

the set of values of the test statistic for which:

| | |
|---|---|
| the null hypothesis is rejected | **rejection region** |
| we fail to reject the null hypothesis | **acceptance region** |

# Test errors

since a test statistic is random, the same test can lead to different conclusions

- **type I error**: the test leads to *reject* $H_0$ when it is *true*
- **type II error:** the test *fails* to reject $H_0$ when it is *false*; sometimes called false alarm

probabilities of the errors:

- let $\beta$ be the probability of type II error
- the **size** of a test is the probability of a type I error and denoted by $\alpha$
- the **power** of a test is the probability of rejecting a false $H_0$ or $(1-\beta)$

$\alpha$ is known as **significance level** and typically controlled by an analyst

for a given $\alpha$, we would like $\beta$ to be as small as possible

# Some common tests

- normal test
- $t$-test
- $F$-test
- Chi-square test

e.g. a test is called a $t$-test if the test statistic follows $t$-distribution

two approaches of hypothesis test

- critical value approach
- $p$-value approach

# Critical value approach

**Definition:** the critical value (associated with a significance level $\alpha$) is the value of the known distribution of the test statistic such that the probability of type I error is $\alpha$

steps involved this test

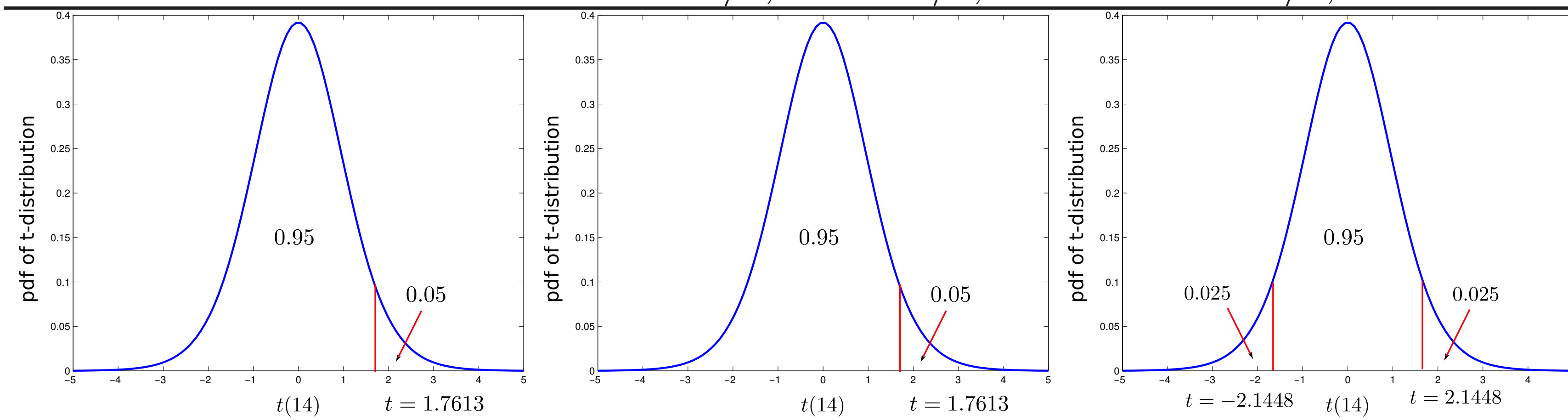
1. define the null and alternative hypotheses.
2. assume the null hypothesis is true and calculate the value of the test statistic
3. set a small significance level (typically $\alpha = 0.01, 0.05$, or $0.10$) and determine the corresponding critical value
4. compare the test statistic to the critical value

| condition | decision |
|---|---|
| the test statistic is more extreme than the critical value | reject $H_0$ |
| the test statistic is less extreme than the critical value | accept $H_0$ |

**example:** hypothesis test on the population mean

- samples $N = 15$, $\alpha = 0.05$
- the test statistic is $t^* = \frac{\bar{x}-\mu}{s/\sqrt{N}}$ and has $t$-distribution with $N-1$ df

| test | $H_0$ | $H_1$ | critical value | reject $H_0$ if |
|---|---|---|---|---|
| right-tail | $\mu = 3$ | $\mu > 3$ | $t_{\alpha,N-1}$ | $t^* \geq t_{\alpha,N-1}$ |
| left-tail | $\mu = 3$ | $\mu < 3$ | $-t_{\alpha,N-1}$ | $t^* \leq -t_{\alpha,N-1}$ |
| two-tail | $\mu = 3$ | $\mu \neq 3$ | $-t_{\alpha/2,N-1}, t_{\alpha/2,N-1}$ | $t^* \geq t_{\alpha/2,N-1}$ or $t^* \leq -t_{\alpha/2,N}$ |



# $p$-value approach

**Definition:** the $p$-value is the probability that we observe a more extreme test statistic in the direction of $H_1$

steps involved this test

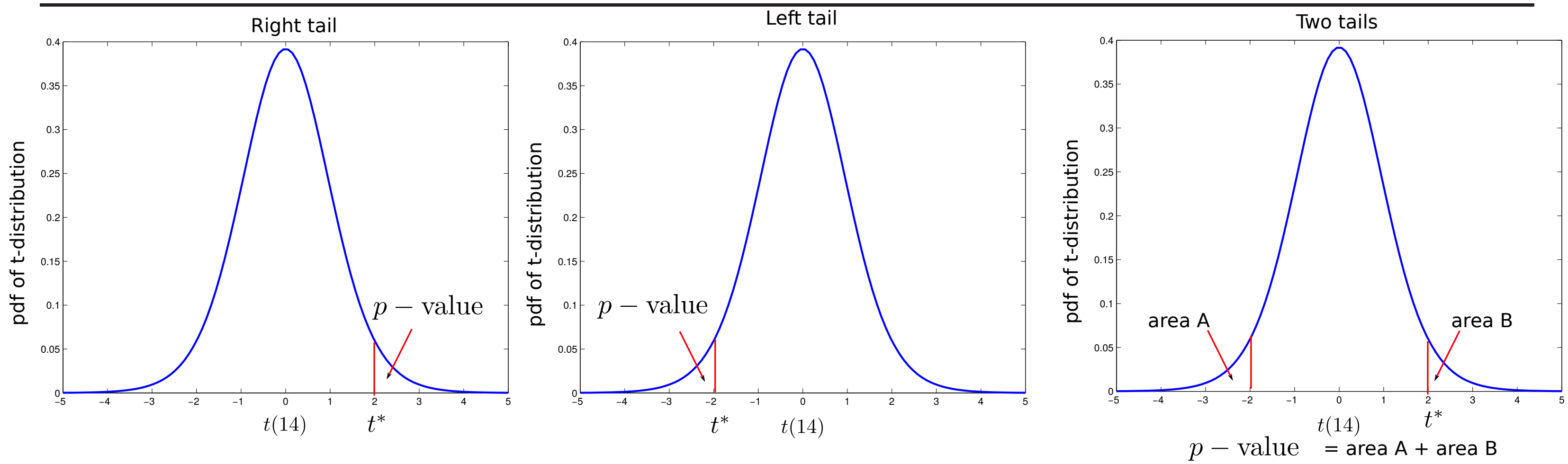
1. define the null and alternative hypotheses.
2. assume the null hypothesis is true and calculate the value of the test statistic
3. calculate the $p$-value using the known distribution of the test statistic
4. set a significance level $\alpha$ (small value such as $0.01, 0.05$)
5. compare the $p$-value to $\alpha$

| condition | decision |
|---|---|
| $p$-value $\leq \alpha$ | reject $H_0$ |
| $p$-value $\geq \alpha$ | accept $H_0$ |

**example:** hypothesis test on the population mean (same as on page 2-32)

- samples $N = 15$, $\alpha = 0.01$ (have only a $1\%$ chance of making a Type I error)
- suppose the test statistic (calculated from data) is $t^* = 2$

| test | $H_0$ | $H_1$ | $p$-value expression | $p$-value |
|---|---|---|---|---|
| right-tail | $\mu = 3$ | $\mu > 3$ | $P(t_{14} \geq 2)$ | $0.0127$ |
| left-tail | $\mu = 3$ | $\mu < 3$ | $P(t_{14} \leq -2)$ | $0.0127$ |
| two-tail | $\mu = 3$ | $\mu \neq 3$ | $P(t_{14} \geq 2) + P(t_{14} \leq -2)$ | $0.0255$ |




right-tail/left-tail tests: reject $H_0$, two-tail test: accept $H_0$

the two approaches assume $H_0$ were true and determine

| $p$-value | critical value |
|---|---|
| the probability of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed | whether or not the observed test statistic is more extreme than would be expected (called critical value) |

the null hypothesis is rejected if

| $p$-value | critical value |
|---|---|
| $p - \text{value} \leq \alpha$ | test statistic $\geq$ critical value |

# References

Chapter 9 in

A. Leon-Garcia, *Probaility, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson, 2009

Chapter 3 in
T. Söderström and P. Stoica, *System Identification*, Prentice Hall, 1989