

CH5350: Applied Time-Series Analysis

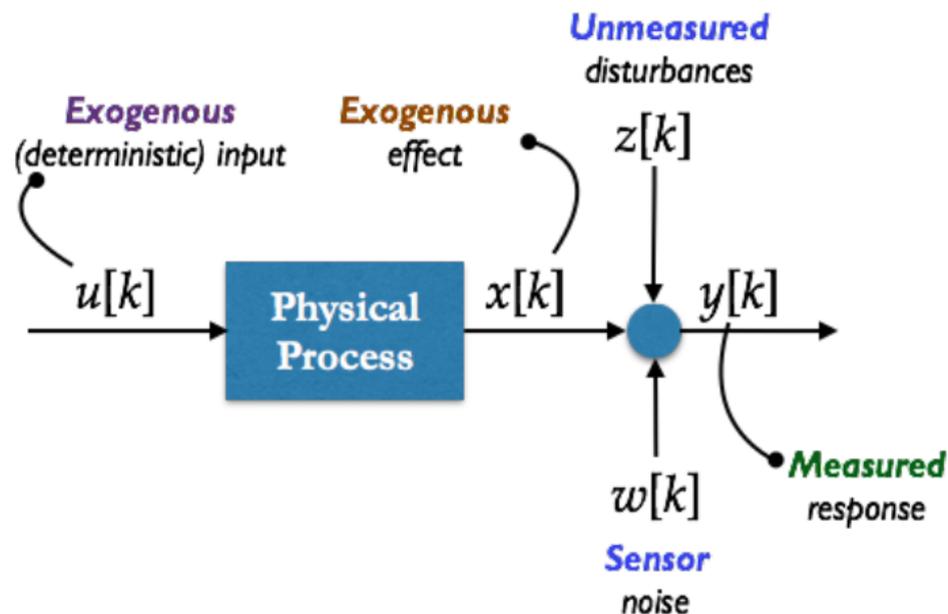
Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

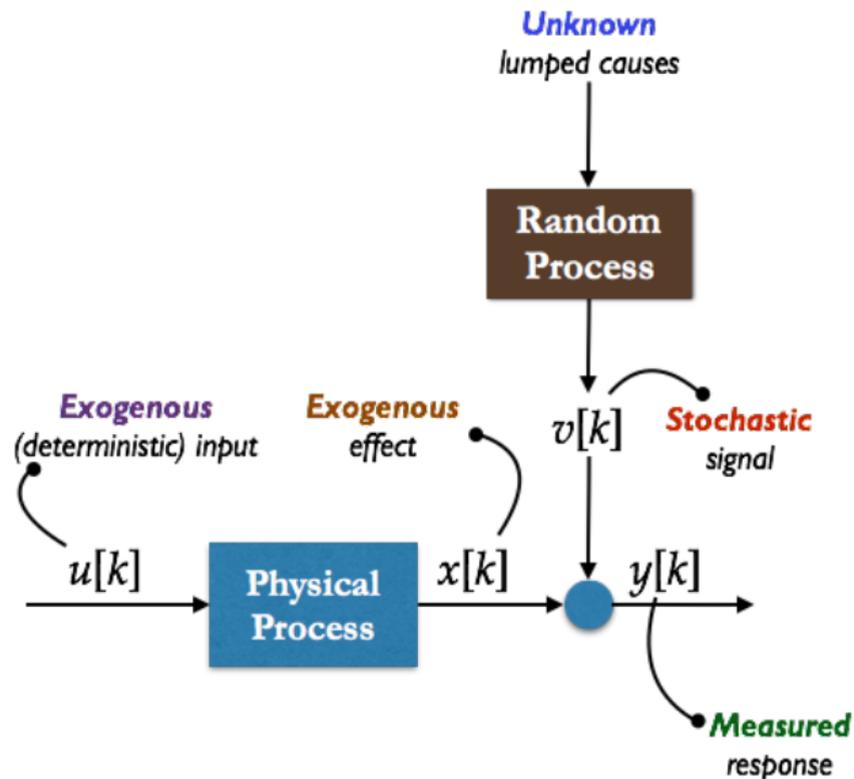


Probability & Statistics: Univariate case

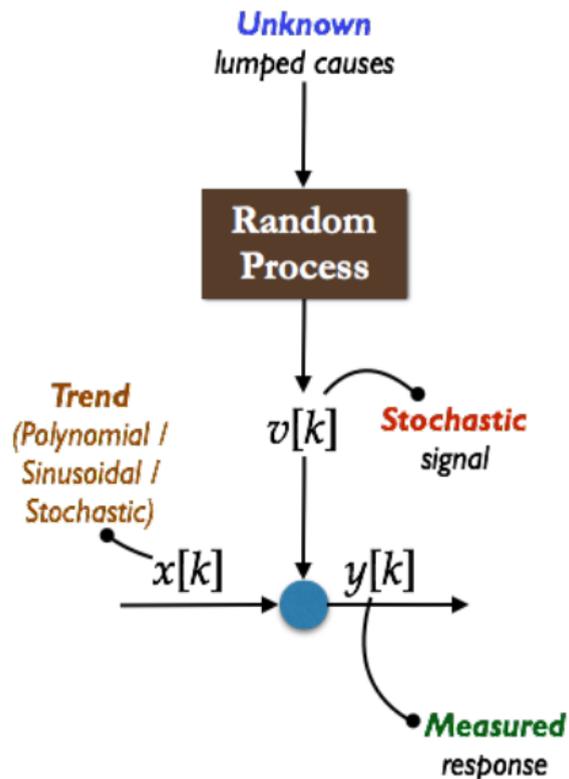
Recap: General process



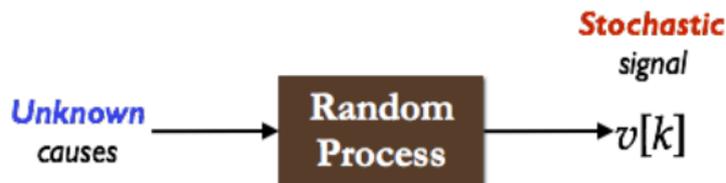
General process: Simplified



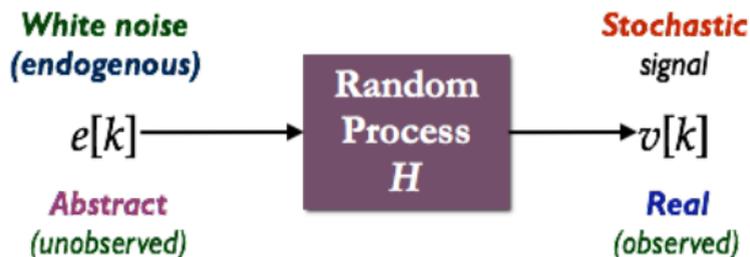
Trend-plus-random processes



Random process

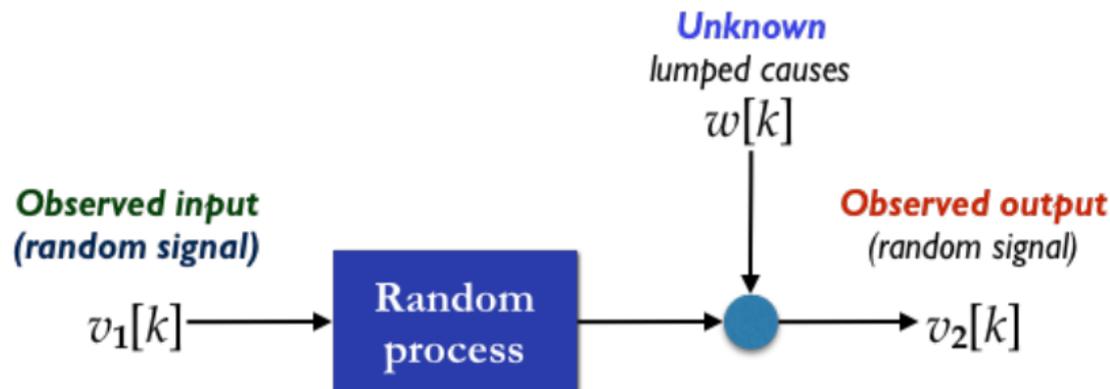


(Discrete-time, scalar-valued, lumped-cause)



(Discrete-time, scalar-valued, endogenously driven)

Bivariate random process



(Discrete-time, bivariate, scalar-valued, externally plus endogenously driven)

Framework

1. Univariate / bivariate
2. Linear random process
3. Stationary and non-stationarities (of certain types)
4. Discrete-time
5. Time- and frequency-domain analysis

The cornerstone of theory of random processes is the concept of a random variable and the associated probability theory.

Notation

- Random variable: UPPERCASE e.g., X ; Outcomes: lowercase e.g., x .
- Probability distribution and density functions: $F(x)$ and $f(x)$, respectively.
- Scalars: lowercase x , θ , etc.
- Vectors: lowercase **bold faced** e.g., \mathbf{x} , \mathbf{v} , $\boldsymbol{\theta}$, etc.
- Matrices: Uppercase **bold faced** \mathbf{A} , \mathbf{X} .
- Expectation operator: $E(\cdot)$
- Discrete-time random signal and process: $v[k]$ (or $\{v[k]\}$) (scalar-valued)
- White-noise: $e[k]$
- Backward / forward shift-operator: q^{-1} and q s.t. $q^{-1}v[k] = v[k - 1]$.
- Angular and cyclic frequencies: ω and f , respectively.
- ...

Random Variable

Definition

A **random variable** (RV) is one whose value set contains at least two elements, i.e., it draws one value from at least two possibilities. The space of possible values is known as the **outcome space or sample space**.

Examples: Toss of a coin, roll of a dice, outcome of a game, atmospheric temperature.

Formal definition

Outcomes of random phenomena can be either qualitative and/or quantitative. In order to have a unified mathematical treatment, RVs are defined to be quantitative.

Definition (Priestley (1981))

A random variable X is a mapping from the sample space \mathbb{S} onto the real line s.t. to each element $s \in \mathbb{S}$ there corresponds a unique real number.

- ▶ In the study of RVs, the time (or space) dimension does not come into picture. Instead they are analysed only in the outcome space.

Two broad classes of RVs

- ▶ When the set of possibilities contains a single element, the randomness vanishes to give rise to a **deterministic variable**.
- ▶ Two classes of random variables exist:
 - Discrete-valued RV:** **discrete** set of possibilities (e.g., roll of a dice)
 - Continuous-valued RV:** **continuous-valued** RV (e.g., ambient temperature)

Focus of this course: **Continuous-valued random variables** (with occasional digression to discrete-valued RVs).

Do random variables actually exist?

The tag of randomness is given to any variable or a signal which is not accurately predictable, i.e., the outcome of the associated event is not predictable with zero error.

In reality, there is no reason to believe that the true process behaves in a “random” manner. It is merely that since we are unable to predict its course, i.e., due to lack of sufficient understanding or knowledge that any process becomes random.

Randomness is, therefore, not a characteristic of a process, but is rather a reflection of our (lack of) knowledge and understanding of that process

Probability Distribution

The natural recourse to dealing with uncertainties is to list all possible outcomes and assign a chance to each of those outcomes

Examples:

- ▶ Rainfall in a region: $\Omega = \{0, 1\}$, $P = \{0.3, 0.7\}$
- ▶ Face value from the roll of a die: $\Omega = \{1, 2, \dots, 6\}$, $P(\omega) = \{1/6\} \forall \omega \in \Omega$

The specification of the outcomes and the associated probabilities through what is known as **probability distribution** completely characterizes the random variable.

Probability Distribution Functions

Probability distribution function

Also known as the **cumulative distribution function**,

$$F(x) = \Pr(X \leq x)$$

- ▶ Probability distribution functions can be either continuous or piecewise-continuous (step-like) depending on whether the RV is continuous- or discrete-valued, respectively.
- ▶ They are known either a priori (through physics or postulates) or by means of experiments

Probability density functions

When the density function exist, i.e., for continuous-valued RVs,

1. The density function is such that the area under the curve gives the probability,

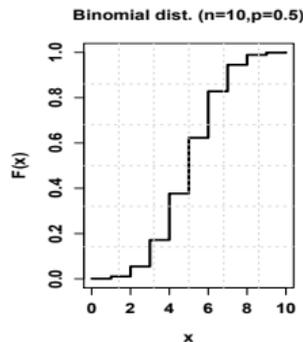
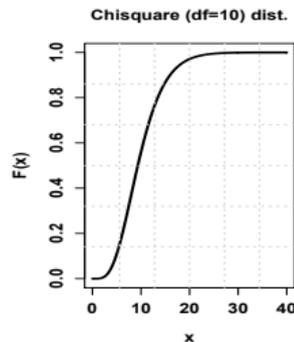
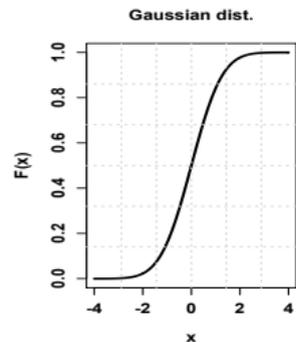
$$\Pr(a < x < b) = \int_a^b f(x) dx \quad \implies \int_{-\infty}^{\infty} f(x) dx = 1 \quad (1)$$

2. The density function is the derivative (w.r.t. x) of the distribution function

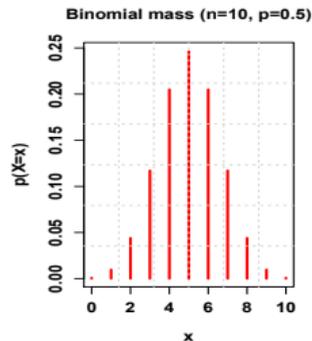
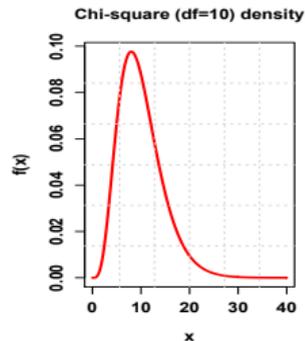
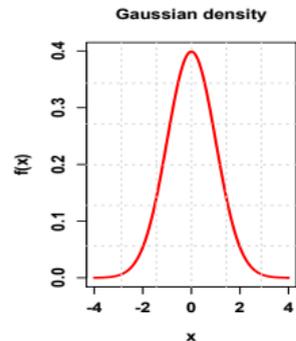
$$f(x) = \frac{dF(x)}{dx} \quad (2)$$

- For discrete-valued RVs, a probability mass function (p.m.f.) is used

Examples: c.d.f. and p.d.f.



The type of distribution for a random phenomenon depends on its nature.



Density Functions

1. **Gaussian density function:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right)$$

2. **Uniform density function:**

$$f(x) = \frac{1}{b - a}, \quad a \leq x \leq b$$

3. **Chi-square density:**

$$f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

Commands in R

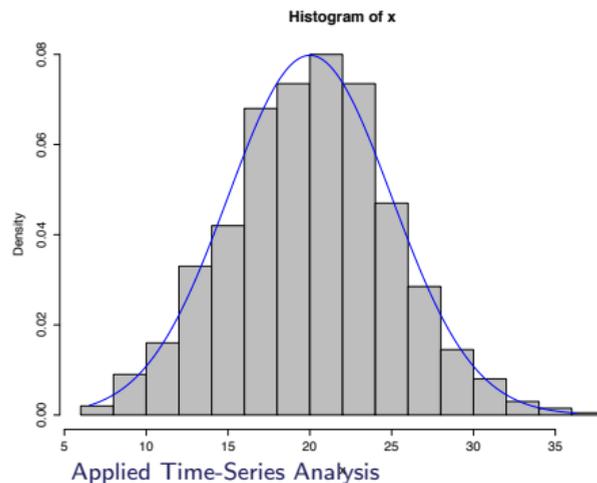
Every distribution that R handles has four functions for *probability*, *quantile*, *density* and *random variable* (value), and has the same root name, but prefixed by p, q, d and r respectively

Few relevant functions:

Commands	Distribution
<code>rnorm</code> , <code>pnorm</code> , <code>qnorm</code> , <code>dnorm</code>	Gaussian
<code>rt</code> , <code>pt</code> , <code>qt</code> , <code>dt</code>	Student's- <i>t</i>
<code>rchisq</code> , <code>pchisq</code> , <code>qchisq</code> , <code>dchisq</code>	Chi-square
<code>runif</code> , <code>punif</code> , <code>qunif</code> , <code>dunif</code>	Uniform distribution
<code>rbinom</code> , <code>pbinom</code> , <code>qbinom</code> , <code>dbinom</code>	Binomial

Sample usage

```
1 x <- rnorm(1000, mean=20, sd=5)
2 hist(x, probability=TRUE)
3 xseq <- seq(min(x), max(x), length=200, col='grey')
4 lines(xseq, dnorm(xseq, mean=20, sd=5), col='blue', lwd=2)
```



Practical Aspects

The p.d.f. of a RV allows us to compute the probability of X taking on values in an infinitesimal interval, *i.e.*, $\Pr(x \leq X \leq x + dx) \approx f(x)dx$

Note: Just as the way the density encountered in mechanics cannot be interpreted as mass of the body at a point, the probability density should never be interpreted as the probability at a point. In fact, for continuous-valued RVs, $\Pr(X = x) = 0$

In practice, knowing the p.d.f. theoretically is seldom possible. One has to conduct experiments and then try to fit a known p.d.f. that best explains the behaviour of the RV.

Practical Aspects: Moments of a p.d.f.

- ▶ It may not be necessary to know the p.d.f. in practice!
- ▶ What is of interest in practice is (i) **the most likely value and/or the expected outcome** (**mean**) and (ii) **how far the outcomes are spread** (**variance**)

The useful **statistical properties**, namely, mean and variance are, in fact, the first and second-order (central) moments of the p.d.f. $f(x)$ (similar to the moments of inertia).

The n^{th} moment of a p.d.f. is defined as

$$M_n(X) = \int_{-\infty}^{\infty} x^n f(x) dx \quad (3)$$

Linear random process and moments

It turns out that for linear processes, predictions of random signals and estimation of model parameters it is sufficient to have the knowledge of **mean**, **variance** and **covariance** (to be introduced shortly), *i.e.*, it is sufficient to know the first and second-order moments of p.d.f.

First Moment of a p.d.f.: Mean

The mean is defined as the **first moment** of the p.d.f. (analogous to the center of mass). It is also the **expected value** (outcome) of the RV.

Mean

The mean of a RV, also the **expectation** of the RV, is defined as

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x) dx \quad (4)$$

Remarks

- ▶ The integration in (4) is across the **outcome space** and NOT across any time space.
- ▶ Applying the **expectation operator** E to a random variable produces its “average” or expected value.
- ▶ Prediction perspective:

The mean is the best prediction of the random variable in the minimum mean square error sense, *i.e.*,

$$\mu = \min_c E(X - \hat{X})^2 \quad \text{s.t.} \quad \hat{X} = c$$

where \hat{X} denotes the prediction of X .

Expectation Operator

- ▶ For any constant, $E(c) = c$.
- ▶ The expectation of a function of X is given by

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx \quad (5)$$

- ▶ It is a **linear** operator:

$$E\left(\sum_{i=1}^k c_i g_i(X)\right) = \sum_{i=1}^k c_i E(g_i(X)) \quad (6)$$

Examples: Computing expectations

Example

Problem: Find the expectation of a random variable $y[k] = \sin(\omega k + \phi)$ where ϕ is uniformly distributed in $[-\pi, \pi]$.

Solution:

$$\begin{aligned} E(y[k]) &= E(\sin(\omega k + \phi)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(\omega k + \phi) d\phi \\ &= \frac{1}{2\pi} (-\cos(\omega k + \phi)|_{-\pi}^{\pi}) \\ &= \frac{1}{2\pi} (\cos(\omega k - \pi) - \cos(\omega k + \pi)) = 0 \end{aligned}$$

Variance / Variability

An important statistic useful in decision making, error analysis of parameter estimation, input design and several other prime stages of data analysis is the **variance**.

Variance

The variance of a random variable, denoted by σ_X^2 is the average spread of outcomes around its mean,

$$\sigma_X^2 = E((X - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx \quad (7)$$

Points to note

- ▶ As (7) suggests, σ_X^2 is the **second central moment** of $f(x)$. Further,

$$\sigma_X^2 = E(X^2) - \mu_X^2 \quad (8)$$

- ▶ The variance definition is in the space of outcomes. **It should not be confused with the widely used variance definition for a series or a signal (sample variance).**
- ▶ Large variance indicates far spread of outcomes around its statistical center. Naturally, in the limit as $\sigma_X^2 \rightarrow 0$, X becomes a **deterministic** variable.

Mean and Variance of scaled RVs

- ▶ Adding a constant to a RV simply shifts its mean by the same amount. The variance remains unchanged.

$$E(X + c) = \mu_X + c, \quad \text{var}(X + c) = \text{var}(X) = \sigma_X^2 \quad (9)$$

- ▶ **Affine transformation:**

$$Y = \alpha X + \beta, \alpha \in \mathcal{R} \implies \mu_Y = \alpha \mu_X + \beta \quad (10)$$

$$\sigma_Y^2 = \alpha^2 \sigma_X^2 \quad (11)$$

- ▶ Properties of non-linearly transformed RV depend on the non-linearity involved

Properties of Normally distributed variables

The normal distribution is one of the most widely assumed and studied distribution for two important reasons:

- ▶ It is completely characterized by the mean and variance
- ▶ Central Limit Theorem
- ▶ If x_1, x_2, \dots, x_n are uncorrelated normal variables, then $y = a_1x_1 + a_2x_2 + \dots + a_nx_n$ is also a normally distributed variable with mean and variance

$$\mu_y = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n$$

$$\sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2$$

Central Limit Theorem

Let X_1, X_2, \dots, X_m be a sequence of independent identically distributed random variables each having finite mean μ and finite variance σ^2 . Let

$$Y_N = \sum_{i=1}^N X_i, \quad N = 1, 2, \dots$$

Then, as $N \rightarrow \infty$, the distribution of

$$\frac{Y_N - N\mu}{\sigma\sqrt{N}} \rightarrow \mathcal{N}(0, 1)$$

One of the popular applications of the CLT is in deriving the distribution of sample mean.