Fundamentals of Digital Commun. Ch. 4: Random Variables and Random Processes

Klaus Witrisal
witrisal@tugraz.at

Signal Processing and Speech Communication Laboratory
www.spsc.tugraz.at
Graz University of Technology

November 26, 2015

Outline

- 4-1 Probability (brief review)
- 4-2 Random Variables
- 4-3 Random Processes
- 4-4 Noise – Example for a Random Process
4-1 Probability

- Fundamental concept:
  - random experiment with an uncertain outcome

- Definitions:
  - $\omega \ldots$ outcomes: $\omega \in \Omega$
  - $\Omega \ldots$ sample space: set of all outcomes
  - **Events**: subsets of sample space $\Omega$ for which probability can be defined
  - **Probability**: functional mapping of events onto reals: $0 \leq P(E) \leq 1$ for all events $E \in \mathcal{B}$
  - $\mathcal{B} \ldots$ collection of (all possible) events
  - $(\Omega, \mathcal{B}, P) \ldots$ probability space

Probability (cont’d)

- Probability is:
  - relative frequency of an event occurring in $n$ trials
  - $A, B \ldots$ events

\[
P(A) = \lim_{n \to \infty} \frac{n_A}{n}
\]

- $n_A \ldots$ number of occurancies of $A$
- $n \ldots$ number of trials

- definitions
  - sure event $E$: $P(E) = 1$
  - null event $A = 0$: $P(A) = 0$
Probability (cont’d)

Events can be represented as sets (Venn diagram)

- product of events “$A$ and $B$” (intersection)
  \[ C = A \cap B = AB \]
- mutually exclusive events $A \cap B = 0$, i.e. $P(AB) = 0$
- sum of events “$A$ or $B$” (union)
  \[ C = A \cup B = A + B \]
  \[ P(A + B) = P(A) + P(B) - P(AB) \]

- conditional probability
  \[ P(A|B) = \lim_{n \to \infty} \frac{n_{AB}}{n_B} = \frac{P(AB)}{P(B)} \]
  \(\text{probability that an event } A \text{ occurs, given that an event } B \text{ has also occurred}\)

- product theorem
  \[ P(AB) = P(A|B)P(B) = P(B|A)P(A) \]

- leads to the Bayes-Theorem
  \[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]
Probability (cont’d)

- **Total probability:**
  - event $B$ occurs jointly with one of the mutually exclusive events $A_1, A_2, ..., A_K$, where
  \[ \sum_{k=1}^{K} P(A_k) = 1 \]

- **Find the probability** $P(B)$, given $P(B|A_k)$ and $P(A_k)$:
  \[
P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_K)P(A_K)
  = \sum_{k=1}^{K} P(B|A_k)P(A_k) = \sum_{k=1}^{K} P(BA_k)
  \]

Def.: **a-priori probabilities**: $P(A_k)$

Def.: **a-posteriori probability**: $P(A_k|B)$

- **Assume:** event $B$ occurs jointly with one of the mutually exclusive events $A_1, A_2, ..., A_K$:
- **Find the probability for $A_k$, when $B$ has already occurred. ($B$ could be an observation of $A_k$)

  can be solved using the Bayes-theorem and the total probability

  \[
P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{K} P(B|A_i)P(A_i)}
  \]

  given $P(A_k)$ and $P(B|A_k)$ (likelihoods)
4-2 Random Variables

- sample space $\Omega$ (with outcomes $\omega_i, \nu_j$ (and events)) is mapped onto the set of reals $\mathbb{R}$
- the random variable is a function of the outcomes
  - notation: $X = X(\omega_i), Y = Y(\nu_j)$, short $X, Y$
  - upper-case symbols are used to distinguish from "defined" variables

- discrete random variables (RV)
  - e.g. dice $X \in \{1, 2, ..., 6\}$ (number of eyes);
    $P(X = x_m) = \frac{1}{6}$

- continuous RV
  - e.g. voltage of a battery $X = U$

Characterization of RVs

- cumulative distribution function – CDF
  (dt: Verteilungsfunktion)

  $F_X(a) = P(X \leq a) = \lim_{n \to \infty} \frac{n(X \leq a)}{n}$

  dimensionless

- probability density function – PDF
  (dt: Wahrscheinlichkeitsdichtefunktion)

  $f_X(x) = \frac{dF_X(x)}{dx} = \lim_{n \to \infty, \Delta x \to 0} \left[ \frac{1}{\Delta x} \left( \frac{n \Delta x}{n} \right) \right]$

  unit is $1/[x]$
Characterization of RVs (cont’d)

- **properties of the CDF**
  - $0 \leq F_X(x) \leq 1$, nondecreasing
  - $F_X(-\infty) = 0$
  - $F_X(\infty) = 1$
  - following the def.: discrete points at $X = x$ are included

- probability that $X$ lies in some interval:
  $$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F_X(x_2) - F_X(x_1)$$

- **properties of the PDF**
  - PDF is **nonnegative** $f_X(x) \geq 0$
  - area under the PDF $\int_{-\infty}^{\infty} f_X(x) dx = 1$

- computation of the CDF from the PDF
  $$F_X(x) = \int_{-\infty}^{x+\epsilon} f_X(\lambda)d\lambda = P(X \leq x)$$

- probability that $X$ lies in some interval:
  $$P(x_1 < X \leq x_2) = \int_{x_1+\epsilon}^{x_2+\epsilon} f_X(\lambda)d\lambda$$
Ensemble Averages and Moments

- mean value, expected value, first-order moment, ensemble average
  
  \[ E\{X\} = m_X = \int_{-\infty}^{\infty} x f_X(x) \, dx \]

  (center of gravity of the PDF)

- \(k\)-th order moment
  
  \[ E\{X^k\} = \int_{-\infty}^{\infty} x^k f_X(x) \, dx \]

- centralized \(k\)-th-order moment (moment about mean)
  
  \[ E\{(X - m_X)^k\} = \int_{-\infty}^{\infty} (x - m_X)^k f_X(x) \, dx \]

Ensemble Averages and Moments (cont’d)

- variance (centralized second moment)
  
  \[ \sigma_X^2 = \text{var}\{X\} = E\{(X - m_X)^2\} = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x) \, dx \]

- standard deviation
  
  \[ \sigma_X = \sqrt{\text{var}\{X\}} \]

- generalized: mean of a function \(g(X)\) of an RV
  
  \[ E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \]
Ensemble Averages and Moments (cont’d)

- definition through events (for \( \lim n \to \infty \))

\[
E\{X\} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X(A_i)
\]

\[
E\{g(X)\} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g[X(A_i)]
\]

- can be used to estimate moments
- will be an approximation for a limited sample \( n < \infty \)

Extension to two Random Variables

- bivariate statistics: joint characterization of two RVs
  - for > 2 variables: multivariate statistics

- CDF

\[
F_{XY}(x, y) = P(X \leq x, Y \leq y)
\]

- PDF

\[
f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}
\]
Extension to two Random Variables (cont’d)

- **correlation**
  \[ R_{XY} = E\{XY\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{XY}(x, y) \, dx \, dy \]

- **uncorrelated RVs**
  \[ R_{XY} = E\{XY\} = E\{X\}E\{Y\} = m_X m_Y \]

- **orthogonal RVs**
  \[ R_{XY} = E\{XY\} = 0 \]

- **covariance**
  \[ K_{XY} = E\{(X - m_X)(Y - m_Y)\} = R_{XY} - m_X m_Y \]
4-3 Random processes

- $X(A_i, t)$ ... function of **two** variables:
  - $A_i$ ... event, realization
  - $t$ ... time (usually)
- for an event $A_i$: sample function $x_i(t)$
- ensemble of sample functions: $\{x_i(t)\}, \forall i$
  - random process: $X(t)$
    (also in lower case)
- illustration of an ensemble of sample functions:
  (blackboard)

Characterization of random processes

- ensemble averages
  - mean (at time $t$)

$$m_X(t) = \mathbb{E}\{X(t)\} = \int_{-\infty}^{\infty} x f_X(t)(x) \, dx$$

-can be time-variant (e.g. sine-wave plus noise)
- autocorrelation (function – ACF)

$$R_X(t_1, t_2) = \mathbb{E}\{X(t_1)X(t_2)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_X(t_1)X(t_2)(x_1, x_2) \, dx_1 \, dx_2$$

$f_X(t_1)X(t_2)(x_1, x_2)$ ... joint PDF for $X(t_1)$ and $X(t_2)$
realignments 1 till 5 of an ensemble of 1000 trials

Mean value: first moment

Second moment (= autocorrelation)

Power Spectrum Magnitude (dB)

Scatter plot of two samples of \( (t) \)

Scatter plot of two samples of \( (t) \)

joint PDF of \( (t_1) \) and \( (t_2) \)

joint PDF of \( (t_1) \) and \( (t_2) \)
Characterization of random processes (cont’d)

- stationarity (time-invariance)
  - first-order stationarity (stationary w.r.t. mean)
    \[ E\{X(t)\} = m_X = \text{const.} \]
  - second-order stationarity (stationary w.r.t. ACF)
    \[ R_X(t_1, t_2) = R_X(\tau) = E\{X(t)X(t + \tau)\} \]
    \[ \text{where } \tau = t_1 - t_2 \]
    i.e.: no variability w.r.t. time shift

- exact definition:
  Stationarity (of \(n\)-th order): PDF (of \(n\)-th order) is invariant w.r.t. time shifts
Characterization of random processes (cont’d)

Important definitions:

- **Strict-Sense-Stationary (SSS)** (Stationär im strengen Sinn)
  - PDFs of any order are stationary

- **Wide-Sense-Stationary (WSS)** (Stationär im weiteren Sinn)
  - Mean and autocorrelation function (ACF) are stationary (i.e. first and second-order moments)
  - for Gaussian processes: WSS implies SSS (hence the importance of the first two moments)

Characterization (cont’d); ACF of WSS Random Processes

- **Properties**
  - $R_X(\tau) = R_X(-\tau)$ ... even symmetry
  - $R_X(0) \geq |R_X(\tau)|$
  - $R_X(0) = \mathbb{E}\{X^2(t)\}$ ... second moment; $\propto$ power

- **Power Spectral Density – PSD** (Leistungsdichtespektrum)
  - $X(t)$ ... **power signal**, if stationary
  - PSD is Fourier transform of the ACF

\[ \text{Def. : } R_X(\tau) \overset{\mathcal{F}}{\longrightarrow} S_X(j\omega) \]

- PSD is power content in frequency domain
Power Spectral Density (PSD) (cont’d)

- definition using the Fourier transform of $X(t)$:
  - needs time limitation of the sample functions:

$$X_T(A_i, j\omega) = \int_{-T/2}^{T/2} x(A_i, t)e^{-j\omega t} \, dt$$

- Wiener-Kinchin theorem

$$S_X(j\omega) = \lim_{T \to \infty} \frac{1}{T} E_{A_i}\{|X_T(A_i, j\omega)|^2\}$$

$$= \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau} \, d\tau = \mathcal{F}[R_X(\tau)]$$

Proof at blackboard!
realizations 1 till 5 of an ensemble of 1000 trials

Mean value; first moment

Second moment (= autocorrelation)

Power Spectrum Magnitude (dB)

Characterization (cont’d);
Ergodic Random Proc.

■ Definition: (all) time and ensemble averages are identical

♦ necessary condition: stationarity
(otherwise, ensemble average would be time-variant)

♦ example: ensemble average and DC-value
it must hold:

\[ x_{DC} \equiv m_X , \text{where} \]
\[ x_{DC} = \langle x_i(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t) \, dt \]
\[ m_X = E\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x) \, dx = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i(t) \]

■ equivalence must hold for all averages → hard to proof

■ Example: ACF

\[ R_X(\tau) = E\{X^*(t)X(t+\tau)\} \equiv \langle x_i^*(t)x_i(t+\tau) \rangle \]

■ consequence of ergodicity: ensemble averages (moments) can be estimated from a single sample function!
4-4 Noise

- Example for a random process
- Physics: freely moving electrons $e^-$
- PSD, ACF

$$S_X(j\omega) = S_X(f) = \frac{N_0}{2} \leftrightarrow R_X(\tau) = \frac{N_0}{2} \delta(\tau)$$

$N_0/2$ ... double-sided noise PSD
- thermal noise $N_0 = kT$
- PDF: Gaussian PDF = normal distribution
  - Power: $P_N = N_0 B = \sigma^2$; $B$ ... equivalent bandwidth

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

Why is noise Gaussian distributed?

- Central Limit Theorem (Zentraler Grenzwertsatz): a sum of (many) random processes has Gaussian PDF
- illustration:
  - PDF of one dice: uniform (discrete)
  - two dice: triangular PDF
  - many dice: PDF approaches Gaussian PDF
- generally (without proof): PDF of the sum of two (independent) random variables corresponds to the convolution of their PDFs
Ein Würfel
Wahrscheinlichkeit

Zwei Würfel
Wahrscheinlichkeit

Drei Würfel
Wahrscheinlichkeit

Augenzahl

(discrete PDFs)

Simulation: Ein Würfel, 1000 Würfe

Simulation: Zwei Würfel, 1000 Würfe

Simulation: Drei Würfel, 1000 Würfe

Simulation: Ein Würfel, 100000 Würfe

Simulation: Zwei Würfel, 100000 Würfe

Simulation: Drei Würfel, 100000 Würfe

Augenzahl

Ereignisse

Ereignisse

Ereignisse

Ereignisse