Probability: Definition

Definition - 1:
- If there are 'n' exhaustive equally likely elementary events in a trial and 'm' of them are favourable to an event 'A', then

\[ P(A) = \frac{m}{n} \]

where
- \( P(A) \) is the probability of event 'A',
- \( m \) is the number of outcomes favouring event 'A',
- \( n \) is the total number of possible outcomes.

Definition - 2:
- If a trial is conducted 'n' times and 'm' of them are favourable to an event 'A', then "relative frequency" \( R(A) \) is defined as

\[ R(A) = \frac{m}{n} \]

If \( \lim_{n \to \infty} R(A) \) exists, then the limit is called as the probability of 'A' i.e.,

\[ P(A) = \lim_{n \to \infty} R(A) = \lim_{n \to \infty} \left( \frac{m}{n} \right) \]
Sample Space and Event

**SAMPLE SPACE:**
The set of all possible outcomes in a trial is called as the sample space ‘S’ for the trial. The elements of S are called “Sample points”.

Examples:
1. Tossing of a coin : S = {H,T}
2. Tossing of two coins : S = {HH,HT,TH,TT}
3. Tossing of a die : S = {1,2,3,4,5,6}

**EVENT:**
Every subset of S is called an event.

Examples:
1. A = {1,3,5} is an event of S = {1,2,3,4,5,6}
2. A = {HH,TT} is an event of S = {HH,HT,TH,TT}

**NOTE:**
- The event Φ is called impossible event
- The event S is called Certain event

Disjoint, Exhaustive and Complementary Events

Two events A and B in a sample space S are called:
- **Mutually exclusive or Disjoint** if (A ∩ B) = Φ
- **Exhaustive** if (A U B) = S
- **Complementary** if (A U B) = S, (A ∩ B) = Φ

**Note:** Complement event of any event A is unique and is usually denoted by $\overline{A}$

i.e., $A \cup \overline{A} = S$, $A \cap \overline{A} = \phi$, $\overline{A} = A$

**Facts:**
- $P(\phi) = 0$
- $P(S) = 1$
- $P(A \cap B) = P(A).P(B)$
- $P(A \cup B) = P(A) + P(B)$, provided $A \cap B = \phi$
- If $A \subseteq B$, then $P(A) \leq P(B)$
Conditional Probability

The probability of outcome $A$, given an occurrence of outcome $B$ is called Conditional Probability of $A$ given $B$ and is defined as

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Random Variable
(associated with continuous signal)

- A random variable is essentially a ‘function’ that maps all points in a sample space to real numbers, the exact value of which is unknown
  
  Example: $x(t) : t \rightarrow x$ (expected position)

- A variable whose values are random but whose statistical distribution is known

**Note:** In case of continuous random variables, the probability of any single discrete event is Zero. Hence, there is a need to evaluate the probability of continuous events within a finite time interval.
Cumulative Distribution Function (CDF)

\[ F_X(x) \triangleq p(-\infty, x) \]

\( F_X(x) \) represents cumulative probability of the continuous random variable \( X \) for all events up to and including \( x \).

Properties:
(i) \( F_X(x) \to 0 \) as \( x \to -\infty \)
(ii) \( F_X(x) \to 1 \) as \( x \to +\infty \)
(iii) \( F_X(x) \) is a non-decreasing function of \( x \)

Probability Density Function (PDF)

\[ f_X(x) \triangleq \frac{d}{dx}[F_X(x)] \]

Properties:
(i) \( \int_{-\infty}^{\infty} f_X(x)dx = 1 \), (ii) \( f_X(x) \) is a non-negative function

Probability over interval \([a, b]\) is defined as:

\[ p_x[a, b] \triangleq F_X(b) - F_X(a) = \int_{a}^{b} \frac{d}{dx}[F_X(x)]dx - \int_{a}^{b} \frac{d}{dx}[F_X(x)]dx \]
\[ = \int_{a}^{b} \frac{d}{dx}[F_X(x)]dx + \int_{a}^{b} \frac{d}{dx}[F_X(x)]dx - \int_{a}^{b} f_X(x)dx \]
\[ = \int_{a}^{b} f_X(x)dx \]

Hence, sometimes \( p_x[a, b] = \int_{a}^{b} f_X(x)dx \) is taken as the definition.
Mean / Expected Value
(for Discrete Random Variables)

Let number of trials be $N$ and possible outcomes be $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively.

In this situation, the number of occurrences of outcome $x_i = p_i N, \quad i = 1, 2, \ldots, N$
∴ The mean (or expected value) of the random variable $X$ is
\[
\mu_X = E(X) = \frac{(p_1N)x_1 + (p_2N)x_2 + \cdots + (p_nN)x_n}{N}
= \sum_{i=1}^{n} p_i x_i
\]

Note: (1) For function of random variables, the followings hold good

\[
E\left[ g(X) \right] \equalD \sum_{i=1}^{n} p_i g(x_i) \quad \text{(Discrete case)}
\]

\[
E\left[ g(X) \right] \equalD \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \quad \text{(Continuous case)}
\]

(2) Expected value is a linear operator, i.e.
\[
E(X_1 + X_2) = E(X_1) + E(X_2)
E(CX) = CE(X)
\]
**Statistical Moment**

When \( g(X) = X^k \), \( E[g(X)] = E(X^k) \), which is called as the \( k^{th} \) statistical moment of the continuous random variable \( X \).

i.e. \( E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) \, dx \)

Example:

\( E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx \) is called as the "Second Moment" of the random variable \( X \).

**Variance, Standard Deviation**

Taking \( g(X) = X - E(X) \) and operating the second moment, we get variance \( \sigma_x^2 \) of the continuous variable \( X \) about its mean.

\[
\sigma_x^2 = E\left(\left[ X - E(X) \right]^2 \right) \\
= E\left( X^2 - 2X E(X) + E(X)^2 \right) \\
= E(X^2) - 2\mu_x E(X) + \mu_x^2 \\
= E(X^2) - 2\mu_x^2 + \mu_x^2 \\
= E(X^2) - \mu_x^2
\]

Standard deviation of random variable \( X \) is defined as

\[
\sigma_x = \sqrt{\text{Variance of } X} = \sqrt{\sigma_x^2}
\]

**Note:**
Mean and variance are very useful statistical properties of random signals.
Normal / Gaussian Distribution

Reasons for popularity:
(i) Close to nature
(ii) Central limit theorem: Sum of random variables with any distribution tends towards normal distribution
(iii) Mathematically tractable and attractive

Probability density function (PDF):
Given a continuous random process $X \sim N(\mu, \sigma^2)$, the PDF for $X$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in (-\infty, \infty)$$

CDF:
$$F_X(x) = \int_{-\infty}^{x} f_X(v) dv$$

$p(\mu - \sigma < x \leq \mu + \sigma) = 68\%$
$p(\mu - 2\sigma < x \leq \mu + 2\sigma) = 95.5\%$
$p(\mu - 3\sigma < x \leq \mu + 3\sigma) = 99.73\%$
**Properties of Normal Distribution**

(1) Any linear function of a normally distributed random variable $X$ is also a normally distributed random variable. 

i.e. if $X \sim N\left(\mu, \sigma^2\right)$ and $Y = aX + b$, then the PDF for $Y$ is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}}$$

(2) If $X_1$ and $X_2$ are independent random variables, and

$$X_1 \sim N\left(\mu_1, \sigma_1^2\right), \quad X_2 \sim N\left(\mu_2, \sigma_2^2\right)$$

then $X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$

Then the PDF becomes:

$$f_{X_1 + X_2}(x_1 + x_2) = \frac{1}{\sqrt{2\pi} \left(\sigma_1^2 + \sigma_2^2\right)} e^{-\frac{(x_1 + x_2 - (\mu_1 + \mu_2))^2}{2 \left(\sigma_1^2 + \sigma_2^2\right)}}$$

**Independence and Conditional Probability**

Two continuous random variables $X$ and $Y$ are "Statistically Independent", if their joint PDF $f_{X,Y}(X,Y)$ is equal to the product of their individual PDFs, i.e. $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

**Conditional Probability (Baye’s Rule):**

Continuous-Continuous:

PDF of continuous $X$ given the presence of continuous $Y$: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Continuous-Discrete:

PDF of discrete $X$ given the presence of continuous $Y$: $p_{X|Y}(x|y) = \frac{f_{X,Y}(y|x)p_x(x)}{\sum_y f_{X,Y}(y|x)p_x(x)}$
Autocorrelation of a Time-varying Random Signal $X(t)$

Autocorrelation:

$$R_X(t_1, t_2) \triangleq E\left[ X(t_1) X(t_2) \right], \text{ where } t_1, t_2 \text{ are two sample times}$$

Theorem:

If a process is **stationary** (i.e. the PDF is invariant with time), then

$$R_X(t_1, t_2) = R_X(t_2 - t_1) = R_X(\tau), \quad \tau \triangleq (t_2 - t_1)$$

i.e. $R_X(\tau) = E\left[ X(t) X(t + \tau) \right]$

Autocorrelation

Compared to random signal $X_2$, random signal $X_1$ is relatively short and wide. As $|\tau|$ increases (as you move away from $\tau = 0$ at the center of the curve) the autocorrelation signal for $X_2$ drops off relatively quickly. This indicates that $X_2$ is less correlated with itself than $X_1$. 
**Autocorrelation: Spectral interpretation in Frequency Domain**

Clearly, the autocorrelation is a function of time, which means that it has a spectral interpretation in the frequency domain also. Again for a stationary process, there is an important temporal-spectral relationship known as the Wiener-Khintchine relation:

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

where \( \mathcal{F}[\cdot] \) indicates the Fourier transform, and \( \omega \) indicates the number of \( (2\pi) \) cycles per second. The function \( S_X(j\omega) \) is called the power spectral density of the random signal. As you can see, this important relationship ties together the time and frequency spectrum representations of the same signal.

---

**White Noise**

If for a stationary random signal \( X \)

\[ R_X(\tau) = \delta(\tau) = \begin{cases} A, & \text{if } \tau = 0 \\ 0, & \text{otherwise} \end{cases}, \]

where \( A \) is a constant,

then the random variable \( X \) is a "white noise".

**Note:**

1. White noise is an important building block for random signal processing, including Kalman filter.

2. A standard way of handling coloured noise is to construct the coloured noise as output of another system with white noise being its input and the augmenting this system with original system (this introduces the concept of "shaping filter", which will be discussed later).
**White Noise**

\[ R_X(\tau) = \frac{\delta(\tau)}{\text{Dirac-delta function}} = \begin{cases} A, & \text{if } \tau = 0 \\ 0, & \text{otherwise} \end{cases} \]

- White noise in time domain
- White noise in frequency domain

\[ S_X(j\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega \tau} d\tau \] (for a stationary process)

---

**Cross-correlation and Cross-covariance**

Cross-correlation of two random variables \( X \) and \( Y \) is defined as

\[ R_{XY}(\tau) \triangleq E\left[ x(t) y(t+\tau) \right] \]

Note:

Crosscorrelation function is also only a function of \( \tau \) (and not a function of \( t \)) for stationary random process.

Cross-covariance of two random variables \( X \) and \( Y \) is defined as

\[ C_{XY}(\tau) \triangleq E\left[ \{x(t) - \mu_X\} \{y(t+\tau) - \mu_Y\} \right] \]
**Uncorrelated and Orthogonal Stochastic Processes**

Two stationary stochastic processes $X$ and $Y$ are uncorrelated if:

$$R_{xy}(\tau) \equiv E[x(t)y(t+\tau)] = E[x(t)]E[y(t+\tau)] = \mu_x \mu_y$$

or, equivalently, if:

$$C_{xy}(\tau) = E[(x(t)-\mu_x)(y(t+\tau)-\mu_y)]$$
$$= E[x(t)-\mu_x]E[y(t+\tau)-\mu_y]$$
$$= (\mu_x-\mu_x)(\mu_y-\mu_y)$$
$$= 0$$

Random variables $X$ and $Y$ are said to be "orthogonal", if $R_{xy}(\tau) = 0$

---

**Vector Stochastic Processes**

Let $V(t) = [v_1(t) \ \cdots \ \ v_n(t)]^T$, where $v_i(t)$ are $n$ scaler stochastic processes.

**Definitions:**

Mean: $\mu \equiv E(V(t)) = [E(v_1(t)) \ \cdots \ \ E(v_n(t))]^T = [\mu_1 \ \cdots \ \mu_n]^T$

Autocorrelation Matrix: $R(\tau) \equiv E[V(t) V^T(t+\tau)]$

Autocovariance Matrix: $C(\tau) \equiv E[(V(t)-\mu)(V(t+\tau)-\mu)^T]$

Variance Matrix: $\sum C(0) = E[(V(t)-\mu)(V(t)-\mu)^T]$
Summary of Continuous Time Kalman Filter

Prof. Radhakant Padhi
Dept. of Aerospace Engineering
Indian Institute of Science - Bangalore

Kalman Filter:
Information Required & Task

- **Information Required**
  - System model (linear/linearized model)
  - Measurements and their statistical behaviors
  - Statistical models characterizing the process and measurement noise (typically zero-mean uncorrelated white noise)
  - Initial condition information for the states

- **Task:**
  - To Estimate (filter) the state by processing the measurement data and using the system model
Kalman Filter

System dynamics

\[ \dot{X} = AX + BU + Gw \]
\[ Y = CX + v \]

\( w(t) \): Process noise that acts to disturb the plant
(e.g. Wind gusts, unmodelled high-frequency dynamics)

\( v(t) \): Measurement noise (sensor noise)

Assumptions of Kalman Filter

\( w(t), v(t) \): Zero-mean “white noise”

\( X(0) \): Unknown
\[ X(0) \sim (\hat{X}_0, P_0) \]

\( w(t) \sim (0, Q), \quad Q \geq 0 \)
\( v(t) \sim (0, R), \quad R > 0 \)
Kalman Filter

Estimator
\[ \dot{\hat{X}} = A\hat{X} + BU + K_c (Y - \hat{Y}) \]

Where
\[ \hat{Y} = E(CX + v) \]
\[ = E(CX) + E(v) \]
\[ = CE(X) \]
\[ = C\hat{X} \]

Kalman Filter

Error Covariance Matrix
\[ P(t) = E[\hat{X} \hat{X}^T] \]
where
\[ \hat{X} \triangleq X(t) - \hat{X}(t) \]

Note:
1. \( P(t) \) is a measure of uncertainty in the estimate
2. If the observer dynamics is asymptotically stable, and \( w(t), v(t) \) are stationary processes, the error will eventually reach a steady state

Key:
The gain \( K_c \) is chosen so that it minimizes the steady-state error covariance. The optimal gain will be a “constant matrix”
Kalman Filter (Mechanization)

Initialization
\[ \hat{X}(0) = \hat{X}_0 \]

Kalman Gain
\[ K_e = P C^T R^{-1}, \quad P > 0 \]

Error Covariance ARE
\[ PA^T + AP - PC^T R^{-1} CP + G Q G^T = 0 \]

Estimator (Filter) Dynamics
\[ \hat{X} = A \hat{X} + B U + K_e (Y - C \hat{X}) \]

Summary of Extended Kalman Filter (EKF):
A Continuous-Discrete Implementation

Prof. Radhakant Padhi
Dept. of Aerospace Engineering
Indian Institute of Science - Bangalore
Extended Kalman Filter (EKF)

- For nonlinear system models, Kalman Filter is not applicable
- How to predict the state vector and its error covariance?
- EKF is extension of Kalman filter via linearization
- It processes all available sensor measurements in estimating the value of states of interest using:
  - Knowledge of system and sensor dynamics.
  - Statistical models reflecting uncertainty in process noise and sensor noise
  - Some information regarding initial condition.

Nonlinear System Dynamics and EKF Design

System Dynamics
\[ \dot{X}(t) = f(X(t), U(t), t) + G(t) w(t) \quad E[w(t)w^T(\tau)] = Q(t)\delta(t - \tau) \]

Output dynamics
\[ Y(t_i) = h(X(t_i), t_i) + v(t_i) \quad E[v_i^T v_j] = R \delta_{ij} \]

It works in two step:

I. Time Update (‘Prediction’).
II. Measurement Update (‘Correction’).
Step I: Prediction from $t_{k-1}^+$ to $t_k^-$

- The optimal estimated states and $P$ are propagated, based on the previous values, the system dynamics, and the previous control inputs and errors of the actual system.

- Propagate the state equation (by numerical integration)
  $$\dot{\hat{X}}(t) = f[\hat{X}(t), U(t), t]$$

- Propagate the error Covariance matrix
  $$\dot{\hat{P}}(t) = PA + A^T P + Q$$
  where, $$A(t) \triangleq \frac{\partial f}{\partial \hat{X}}$$

Step II: Filtering from $t_k^-$ to $t_k^+$

- Compute the filter gain
  $$K_{\epsilon_k} = P_k^- C_k^{-T} \left[C_k^- P_k^- C_k^{-T} + R_k \right]^{-1}, \quad C_k^- \triangleq \frac{\partial h}{\partial \hat{X}}_{\hat{X}^-}$$

- Update the state vector and error covariance matrix
  $$\hat{X}_k^+ = \hat{X}_k^- + K_{\epsilon_k} \left[Y_k - h(\hat{X}_k^-) \right]$$
  $$P_k^+ = \left[I - K_{\epsilon_k} C_k^- \right] P_k^-$$
Advantages/Limitations of EKF

- **Advantages:**
  - It works for a wide variety of practical problems
  - Its computationally very efficient

- **Limitations:**
  - Linearization can introduce significant error
  - No general convergence guarantee
  - Works in general; but in some cases its performance can be surprisingly bad
  - Unreliable for colour noise

- **Issues:**
  - Optimal measurement schedules
  - Parameter/Modeling uncertainties
  - Computational errors
  - Noise model (e.g. Non-Gaussian PDF)

Beyond EKF

- **Need**
  - Nonlinear systems
  - Non-Gaussian noises/inputs
  - Correlated noises

- **Characteristics of Such Filters**
  - Are often approximate
  - Sacrifices theoretical accuracy in favour of practical constraints and considerations like robustness, adaptation, numerical feasibility
  - Attempt to cover the limitations of EKF
Thanks for the Attention....!!

questions ... ??