ECE 8873
Data Compression & Modeling

Review of Probability and Random Processes

School of Electrical and Computer Engineering
Georgia Institute of Technology
Spring, 2004

Approaches to the Theory of Probability

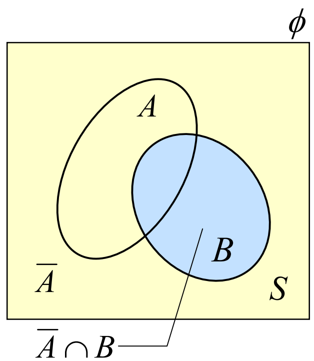
- Relative frequency approach
  - An event that occurs more frequently has higher probability and vice versa; intuitive but difficult to generalize
- Axiomatic approach
  - Probability is a real number between 0 and 1;
  - Built upon set theory;
  - A **Probability Space** consists of three components: the observation or sample space, S, the probability measure, Pr(), and the assignment of probability that satisfies a set of axioms.
  - **Observation or Sample Space** is a space whose elements are all the outcomes of an experiment.
  - **Events** are subsets of the observation space.
  - **Three axioms** to satisfy:

Non-negativity:	$\Pr(A) \geq 0$
Sure event & total probability:	$\Pr(S) = 1$
Exclusivity:	If $A \cap B = \phi$, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Important Corollaries

- 0-measure events:
  $\Pr(\phi) = 0$

  Theoretically, there could be other 0-measure events than the empty set – the axioms allow that. But in the relative frequency approach, they are usually either not treated or treated differently as "unseen events".



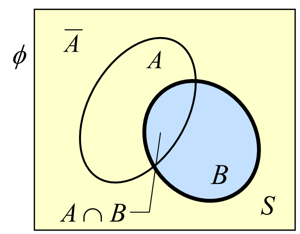
- Complementary events:
  $\Pr(A) = 1 - \Pr(\overline{A}) \leq 1$
- $\Pr(A \cup B) = \Pr(A) + \Pr(\overline{A} \cap B)$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \leq \Pr(A) + \Pr(B)$

Conditional Probability

- If event B is assumed to have non-zero probability, the conditional probability of A, given B, is

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(B)}, \qquad \Pr(B) > 0$$

- We say joint event (A, B) occurs if the outcome of a trial satisfies the definition of **both** events, in other words, when any of the common elements of A and B appears as the outcome of the trial. That is, $\Pr(A, B) = \Pr(A \cap B)$



$\Pr(B \mid B) = 1$

$\Pr(A \mid B)$ can be considered as the probability of A when B is the observation space.

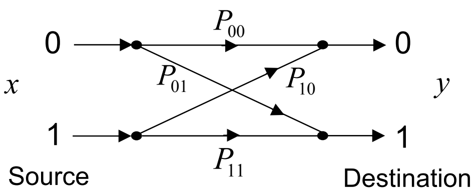
$(A \cap B) \cup (\overline{A} \cap B) = S \cap B = B,$

and $(A \cap B) \cap (\overline{A} \cap B) = \phi$

$\Rightarrow P(A, B) + P(\overline{A}, B) = P(B)$

Probability & Information Transmission

A binary channel



At the source,

$\Pr(x=0)=p$ and

$\Pr(x=1)=1-p$,

$\Pr(x=0)$ and $\Pr(x=1)$

are called **a priori** probabilities.

The channel, being non-ideal (causing confusions), is characterized by the four conditional probabilities: P_{00}, P_{01}, P_{10}, and P_{11}

$P_{ij}=\Pr(y=j\mid x=i)=$ (probability of j received at the destination, when i was actually sent by the source)

A posteriori probability –
The probability that i was sent at the source, given that j is received at the destination.

$$\Pr(x=i\mid y=j)=\frac{\Pr(y=j\mid x=i)\Pr(x=i)}{\Pr(y=j)}$$

Bayes Formula

Is used to find a posteriori probability

$$\Pr(x=i\mid y=j)=\frac{\Pr(y=j\mid x=i)\Pr(x=i)}{\Pr(y=j)}$$

More precisely,

$\Pr(x=i\mid y=j)=\Pr($symbol i was sent, given that j is received$)$

$$=\frac{\Pr(y=j\mid x=i)\Pr(x=i)}{\sum_{all\ i}\Pr(y=j\mid x=i)\Pr(x=i)}$$

$$=\frac{P_{ij}\Pr(x=i)}{\sum_{all\ i}P_{ij}\Pr(x=i)}=\frac{P_{ij}\Pr(x=i)}{\Pr(y=j)}=$$ the source portion of contributions that led to the reception of j

Probability Distribution Function

- $\Pr(X\le x)$ is a **function** of x.

  $F_X(x)=\Pr(X\le x)$ is the **probability distribution function** defined over all x.

  $\{X\le-\infty\}=\phi,\quad F_X(-\infty)=0$

  $\{X\le\infty\}$ is always true, a sure event, thus, $F_X(\infty)=1$

  $\{X\le-\infty\}\subset\{X\le x\}\subset\{X\le\infty\}$ for $-\infty<x<\infty$

  $\Rightarrow\quad 0\le F_X(x)\le 1$

  » If $x_1<x_2,\quad \{X\le x_1\}\subset\{X\le x_2\},\quad \{X\le x_1\}\cap\{X\le x_2\}=\{X\le x_1\}$

  $\{X\le x_2\}=\{X\le x_1\}\cup\{x_1<X\le x_2\}$, and $\{X\le x_1\}\cap\{x_1<X\le x_2\}=\phi$

  $\therefore\quad \Pr(x_1<X\le x_2)=F_X(x_2)-F_X(x_1)$

  $F_X(x)$ is a non-decreasing function of x.

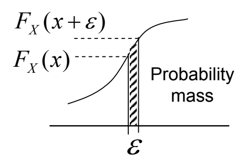
Probability Density Functions

- The slope of the probability distribution function at x represents the incremental probability at that point and thus gives the sense of how likely $X=x$ might be.

$$f_X(x)=\lim_{\varepsilon\to 0}\frac{F_X(x+\varepsilon)-F_X(x+\varepsilon)}{\varepsilon}=\frac{dF_X(x)}{dx}$$

$$f_X(x)dx=\Pr(x<X\le x+dx)$$

is the probability mass at x.



- The derivative is called the probability density function (pdf). Pdf is **non-negative**. In the case of discrete distributions, the pdf consists of delta functions at those realizable values, each having an area equal to the corresponding magnitude of probability.

Functions of Random Variable

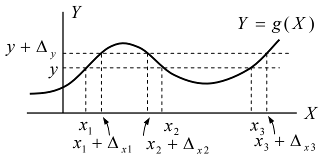
- X is a random variable with pdf $f_X(x)$.
- Y is a monotonic function of X; $Y = g(X)$. Find $f_Y(y)$.

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

Expressed in y with $g^{-1}(y) = x$, $\quad f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$

- If Y is non-monotonic function of X,

$$f_Y(y) = \sum_{\text{for all } x = g^{-1}(y)} f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$



Example: $\quad Y = X^2 \text{ or } X = \sqrt{Y}, \quad \left|\frac{dx}{dy}\right| = \frac{1}{2\sqrt{y}}, \quad \text{but for any } y > 0,\ x = \pm\sqrt{y}$

Therefore, $\quad f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right], \quad y \geq 0; \qquad = 0, \quad y < 0$

Mean Values and Moments

- Mean or expected value of a random variable

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

- Mean or expected value of a function of a random variable $\quad E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$
- N^{th} moments & central moments of a r.v.

$$\overline{X^n} = E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx \qquad \overline{(X-\overline{X})^n} = E[(X-\overline{X})^n] = \int_{-\infty}^{\infty} (x-\overline{X})^n f_X(x)dx$$

When $n = 2$,

$\overline{X^2} = E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx$ is called the mean-square value

and $\sigma^2 = E[(X-\overline{X})^2] = \int_{-\infty}^{\infty} (x-\overline{X})^2 f_X(x)dx$ is the variance.

$$\sigma^2 = E[(X-\overline{X})^2] = E[X^2 - 2X\overline{X} + \overline{X}^2] = E[X^2] - 2\overline{X}^2 + \overline{X}^2 = E[X^2] - \overline{X}^2$$

Conditional Probability Distribution

We define the conditional probability the same as before.

$$F_X(x\,|\,M) = \Pr(X \leq x\,|\,M) = \frac{\Pr(X \leq x, M)}{\Pr(M)}, \quad \Pr(M) > 0$$

If we use the event mapping concept, $\Pr(X \leq x, M)$ is the probability of all the outcomes which realize both events $X(\xi) \leq x$ and $\xi \in M$.

If $M = \{X \leq m\}$, $\; F_X(x\,|\,M) = \Pr(X \leq x\,|\,X \leq m) = \dfrac{\Pr(X \leq x, X \leq m)}{\Pr(X \leq m)}$

If $x \leq m$, $F_X(x\,|\,M) = \dfrac{\Pr(X \leq x)}{\Pr(X \leq m)} = \dfrac{F_X(x)}{F_X(m)}$ $\qquad$ If $x \geq m$, $F_X(x\,|\,M) = \dfrac{\Pr(X \leq m)}{\Pr(X \leq m)} = 1$

Conditional probability density function has all the properties of a usual pdf. $\quad f_X(x\,|\,M) = \dfrac{dF_X(x\,|\,M)}{dx}$

Characteristic Function

- Definition

$$\Phi_X(u) = E[e^{juX}] = \int_{-\infty}^{\infty} e^{jux} f_X(x)\,dx$$

that is, the characteristic function of a random variable can be viewed as the Fourier transform of its probability density function.

- The pdf is then the inverse Fourier transform of the characteristic function

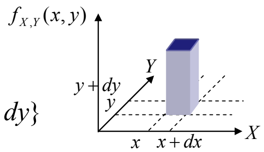
$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jux}\Phi_X(u)\,du$$

$$\frac{d}{du}\Phi(u) = \int_{-\infty}^{\infty}\left(\frac{d}{du}e^{jux}\right) f_X(x)\,dx = \int_{-\infty}^{\infty} jxe^{jux} f_X(x)\,dx$$

$$\left.\frac{d}{du}\Phi(u)\right|_{u=0} = \int_{-\infty}^{\infty} jx\,f_X(x)\,dx = j\overline{X} \qquad \left.\frac{d^n}{du^n}\Phi(u)\right|_{u=0} = j^n E[X^n] = j^n\,\overline{X^n}$$

Joint Probability Density Functions

Definition $f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x\, \partial y}$



$$f_{X,Y}(x,y)\, dx\, dy = \Pr\{x < X \le x+dx, y < Y \le y+dy\}$$

1. $f_{X,Y}(x,y) \ge 0, \quad -\infty < x < \infty, -\infty < y < \infty$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$
3. $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du$
4. $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$ → marginals
5. $\Pr\{x_1 < X \le x_2, y_1 < Y \le y_2\} = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\, dy\, dx$

Expectation of Functions of Two R.V.s

- Similar definition as in single random variable case

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

- When $g(X,Y) = X^n Y^k$

$E[X^n Y^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^n y^k f_{X,Y}(x,y)\, dx\, dy$ is called the ***joint moment of*** X ***and*** Y. When $n=k=1$, it is called correlation. $n+k$ is the order of the moment.

$$E[(X-\overline{X})(Y-\overline{Y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\overline{X})(y-\overline{Y}) f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy - \overline{X}y - x\overline{Y} + \overline{X}\overline{Y}) f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\, dx\, dy - \overline{X}\overline{Y} = \overline{XY} - \overline{X}\overline{Y}$$

Probability Distributions of Functions of Several Random Variables

- The probability space is defined on a hyperspace that contains the random variables.
- Function $Y = g(X_1, X_2, \cdots, X_N)$

$$F_Y(y) = \Pr(Y \le y) = \Pr[g(X_1, X_2, \cdots, X_N) \le y]$$
$$= \Pr(\{\xi : g(x_1 = X_1(\xi), x_2 = X_2(\xi), \cdots, x_N = X_N(\xi)) \le y\})$$
$$F_Y(y) = \Pr[g(X_1, X_2, \cdots, X_N) \le y] =$$
$$\int\cdots\int_{\{g(x_1,x_2,\cdots,x_N)\le y\}} f_{X_1,X_2,\cdots,X_N}(x_1,x_2,\cdots,x_N)\, dx_1 dx_2 \cdots dx_N$$

⟹ Direct integration

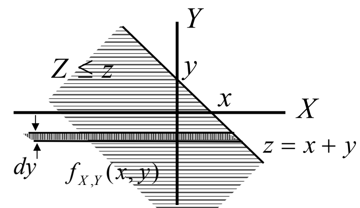
- Another method is through transformation of variables.

Sum of Two Random Variables

$$Z = X + Y$$

At any y, the small stripe has a probability mass of

$$dy \int_{-\infty}^{z-y} f_{X,Y}(x,y) dx$$



Thus, the shaded area which is $\Pr\{Z \le z\}$ can be obtained by

$$\Pr\{Z \le z\} = F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_{X,Y}(x,y)\, dx\, dy$$

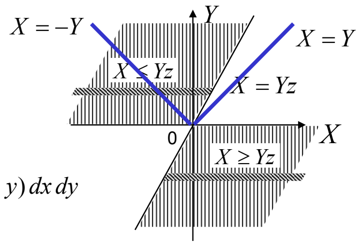
If X and Y are independent, $F_Z(z) = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{z-y} f_X(x)\, dx\, dy$

And, $f_Z(z) = \dfrac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy$ Use Leibniz's rule

The pdf of the sum of two statistically independent random variables is the convolution of their individual pdf's.

Probability Density Functions of Two R.V.s

$Z = X / Y$



Given z, the function $x=yz$ is a line going through the origin.

$$F_Z(z) = \Pr\{Z \le z\} = \Pr\left\{\frac{X}{Y} \le z\right\} = \iint_{\text{shaded area}} f_{X,Y}(x,y)\,dx\,dy$$

because if $y > 0$, then $x \le yz$; if $y < 0$, then $x \ge yz$

$$F_Z(z) = \int_0^\infty \int_{-\infty}^{yz} f_{X,Y}(x,y)\,dx\,dy + \int_{-\infty}^0 \int_{yz}^\infty f_{X,Y}(x,y)\,dx\,dy$$

$$f_Z(z) = \int_0^\infty y\, f_{X,Y}(yz,y)\,dy - \int_{-\infty}^0 y f_{X,Y}(yz,y)\,dy = \int_{-\infty}^\infty |\,y\,| f_{X,Y}(yz,y)\,dy$$

If $Y = g(X)$, the integral for $F_Z(z)$ becomes a line integral. A special case is when $Y = |\,X\,|$.

Transformation of Multiple Random Variables

$Z = \varphi_1(X,Y)$ and $W = \varphi_2(X,Y)$ are two functions of r.v. X and Y.

Both φ_1 and φ_2 are continuous functions with corresponding inverse functions $X = \psi_1(Z,W)$ and $Y = \psi_2(Z,W)$, respectively.

Since all the events that map to $\{x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\}$ would also map to $\{z_1 < Z(\xi) \le z_2, w_1 < W(\xi) \le w_2\}$, we have

$$\Pr\{x_1 < X \le x_2, y_1 < Y \le y_2\} = \Pr\{z_1 < Z \le z_2, w_1 < W \le w_2\}.$$

or $\int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dydx = \int_{z_1}^{z_2}\int_{w_1}^{w_2} f_{Z,W}(z,w)\,dwdz$

But $\int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dydx = \int_{z_1}^{z_2}\int_{w_1}^{w_2} f_{X,Y}(\psi_1(z,w),\psi_2(z,w))\,|\,J\,|\,dwdz$

by way of change of variables with

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \frac{\partial \psi_1}{\partial z} & \frac{\partial \psi_1}{\partial w} \\ \frac{\partial \psi_2}{\partial z} & \frac{\partial \psi_2}{\partial w} \end{vmatrix}$$

J is the Jacobian that relates the incremental area $dzdw$ to $dxdy$.

Transformation of Multiple R.V. (cont'd)

$$\int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\,dydx = \int_{z_1}^{z_2}\int_{w_1}^{w_2} f_{X,Y}(\psi_1(z,w),\psi_2(z,w))\,|\,J\,|\,dwdz$$

$$= \int_{z_1}^{z_2}\int_{w_1}^{w_2} f_{Z,W}(z,w)\,dwdz$$

Therefore, $f_{Z,W}(z,w) = |\,J\,|\, f_{X,Y}[\psi_1(z,w),\psi_2(z,w)]$

Example: $Z = XY,\quad W = X \;\Rightarrow\; X = W,\quad Y = Z/W$

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ w^{-1} & -\frac{z}{w^{-2}} \end{vmatrix} = -w^{-1}$$

Thus, $f_{Z,W}(z,w) = \frac{1}{|\,w\,|} f_{X,Y}\left(w, \frac{z}{w}\right)$

The mariginals, $f_Z(z) = \int_{-\infty}^\infty \frac{1}{|\,w\,|} f_{X,Y}\left(w,\frac{z}{w}\right)dw$

$$f_W(w) = \int_{-\infty}^\infty \frac{1}{|\,w\,|} f_{X,Y}\left(w,\frac{z}{w}\right)dz$$

Characteristic Function of Sum of R.V.s

- The characteristic function of a random variable is the Fourier transform of its pdf.
- Very useful in deriving the moments of the r.v.

$\Phi_X(u) = \int_{-\infty}^\infty f_X(x)e^{jux}dx$ and $f_X(x) = (2\pi)^{-1}\int_{-\infty}^\infty \Phi_X(u)e^{-jux}du$

$\Phi_Y(u) = \int_{-\infty}^\infty f_Y(y)e^{juy}dy$ and $f_Y(y) = (2\pi)^{-1}\int_{-\infty}^\infty \Phi_Y(u)e^{-juy}du$

- Now, $Z = X + Y$ where X and Y are independent,

The pdf of the sum of two statistically independent random variables is the convolution of their individual pdf's.

$$\Phi_Z(u) = \Phi_X(u)\Phi_Y(u)$$

$$f_Z(z) = (2\pi)^{-1}\int_{-\infty}^\infty \Phi_Z(u)e^{-juz}du = (2\pi)^{-1}\int_{-\infty}^\infty \Phi_X(u)\Phi_Y(u)e^{-juz}du$$

Characteristic Function of A Gaussian R.V.

$$\Phi_X(u) = E[e^{juX}] = \int_{-\infty}^{\infty} e^{jux} f_X(x)\,dx$$

$$\Phi_X(u) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp(jux) \exp\left[\frac{-(x-\overline{X})^2}{2\sigma^2}\right] dx$$

The exponential term:

$$jux - \frac{(x-\overline{X})^2}{2\sigma^2} = \frac{jux\,2\sigma^2 - (x-\overline{X})^2}{2\sigma^2} = -\frac{\left[x-(\overline{X}+ju\sigma^2)\right]^2}{2\sigma^2} + ju\overline{X} - \frac{u^2\sigma^2}{2}$$

$$\Phi_X(u) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left[x-(\overline{X}+ju\sigma^2)\right]^2}{2\sigma^2} + ju\overline{X} - \frac{u^2\sigma^2}{2}\right] dx$$

$$= \left\{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left[x-(\overline{X}+ju\sigma^2)\right]^2}{2\sigma^2}\right] dx\right\} \exp\left(ju\overline{X} - \frac{u^2\sigma^2}{2}\right) = \exp\left(ju\overline{X} - \frac{u^2\sigma^2}{2}\right)$$

because $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left[x-(\overline{X}+ju\sigma^2)\right]^2}{2\sigma^2}\right] dx = 1$

Use Cauchy theorem for contour integral – net result is that it is just the same as integrating along the real axis.

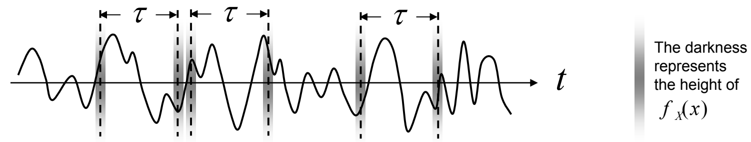
Wide Sense Stationary Processes

$$R_X(t_1,t_2) = R_X(t_1+T, t_2+T) = E[X(t_1+T)X(t_2+T)]$$

Set $T = -t_1 \quad \Rightarrow \quad R_X(t_1,t_2) = R_X(t_1-T, t_2-T) = R_X(0, t_2-t_1)$

For a wide sense stationary process, the autocorrelation function does not depend on the absolute time origin. The first argument 0 is thus arbitrary and the autocorrelation is a function of only the time difference, $t_2 - t_1 = \tau$.

$$R_X(\tau) = R_X(0, t_2-t_1) = E[X(t_1)X(t_1+\tau)] = E[X(t)X(t+\tau)]$$



If we look at any two time instances of a wide sense stationary process, their correlation is only a function of their time difference, no matter where they are. In subsequent discussions, wide sense stationarity is always assumed.

Properties of Autocorrelation Functions (I)

- $R_X(0) = \overline{X^2} \ge 0$
- Symmetry: $R_X(\tau) = R_X(-\tau)$

$$R_X(\tau) = E[X(t)X(t+\tau) = E[X(t-\tau)X(t)] = R_X(-\tau)$$

- $|R_X(\tau)| \le R_X(0)$

$$E[(X_1 \pm X_2)^2] = E[X_1^2 \pm 2X_1X_2 + X_2^2] \ge 0$$

$$E[X_1^2 + X_2{}^2] = 2R_X(0) \ge |E[2X_1X_2]| = 2|R_X(\tau)|$$

- If $X(t)$ has a constant component, say, $X(t) = A + V(t)$, where $V(t)$ has zero mean, the autocorrelation function has a constant component.

$$E\{[A+V(t)][A+V(t+\tau)]\} = E[A^2 + AV(t) + AV(t+\tau) + V(t)V(t+\tau)]$$

$$= A^2 + AE[V(t)] + AE[V(t+\tau)] + E[V(t)V(t+\tau)] = A^2 + R_V(\tau)$$

Properties of Autocorrelation Functions (II)

- If $X(t)$ has a periodic component, then the autocorrelation function has a periodic component.

$$X(t) = A\cos(\omega t + \Theta)$$

where A and ω are constant and Θ a r.v. uniformly distributed over (0,2π); i.e. $f_\Theta(\theta) = (2\pi)^{-1}, 0 \le \theta < 2\pi; = 0,$ elsewhere.

$$R_X(\tau) = E[A\cos(\omega t + \Theta) A\cos(\omega t + \omega\tau + \Theta)]$$

$$= E\left[\frac{A^2}{2}\cos(2\omega t + \omega\tau + 2\Theta) + \frac{A^2}{2}\cos(\omega\tau)\right]$$

$$= \frac{A^2}{2}\cos(\omega\tau) + \frac{A^2}{2}\int_0^{2\pi} \frac{1}{2\pi}\cos(2\omega t + \omega\tau + 2\theta)\,d\theta = \frac{A^2}{2}\cos(\omega\tau)$$

- If $X(t)$ is ergodic and zero-mean, and has no periodic components,

$$\lim_{|\tau|\to\infty} R_X(\tau) = 0$$

That is, time samples far apart tend to behave statistically independently.

Properties of Autocorrelation Functions (III)

- Autocorrelation functions cannot have arbitrary shape – they must correspond to some power spectrum which must be non-negative over the entire frequency range. More discussions later.

$$\mathcal{F}\,[R_X(\tau)] = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau = \text{power spectrum of } X(t)$$

$$\mathcal{F}\,[R_X(\tau)] = S_X(\omega) \geq 0 \quad \text{for all } \omega$$

Example: An ergodic random process has an autocorrelation function of the form $R_X(\tau) = \dfrac{4\tau^2+6}{\tau^2+1}$

Find the mean-square value, mean value, and variance of the process.

$$R_X(\tau) = \frac{4\tau^2+6}{\tau^2+1} = 4 + \frac{2}{\tau^2+1} \qquad \overline{X^2} = R_X(0) = 6$$

$$\overline{X}^2 = 4 \;\Rightarrow\; \overline{X} = \pm 2 \qquad \sigma_X^2 = \overline{X^2} - (\overline{X})^2 = 6 - 4 = 2$$

Sinusoid Plus Noise

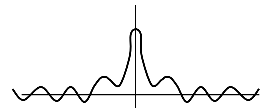
Let $X(t) = A\cos(\omega t + \Theta)$ where Θ is a random variable uniformly distributed over $(0, 2\pi)$. $R_X(\tau) = \frac{1}{2}A^2\cos\omega\tau$

Let $V(t)$ be a zero mean noise process, statistically independent of $X(t)$, the signal, with autocorrelation function:

$$R_V(\tau) = B^2 e^{-\alpha|\tau|}$$

The observed process is $Z(t) = A\cos(\omega t + \Theta) + V(t)$ which has an autocorrelation function:

$$R_Z(\tau) = \frac{1}{2}A^2\cos\omega\tau + B^2 e^{-\alpha|\tau|}$$



Note that $R_V(\tau) = B^2 e^{-\alpha|\tau|} \to 0$, as $\tau \to \infty$.

It is thus possible to recover a sinusoid from noise contamination as long as we measure the autocorrelation at sufficiently long time lags.

Spectral Density

$$S_X(\omega) = \lim_{T\to\infty} \frac{E[|F_{X_T}(\omega)|^2]}{2T} \qquad \overline{X^2} = R_X(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)d\omega$$

Since $S_X(\omega)$ is an average over time, it is thus usually called a **power density spectrum**. When $S_X(\omega)$ is integrated over the entire frequency range, we obtain the average power of the signal, which is equal to the mean-square value of the wide-sense stationary process.

Example: $S_X(\omega) = \dfrac{2a}{\omega^2 + a^2} \qquad S_X(0) = \dfrac{2a}{0^2 + a^2} = \dfrac{2}{a}$

$$\overline{X^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{2a}{\omega^2+a^2}d\omega = \frac{1}{2\pi}\left[\frac{2a}{a}\tan^{-1}\left(\frac{\omega}{a}\right)\right]_{-\infty}^{\infty} = \frac{2}{2\pi}\left[\frac{\pi}{2}+\frac{\pi}{2}\right] = 1$$

Useful integral:

$$\frac{1}{2\pi}\int_p^q \frac{2a}{\omega^2+a^2}d\omega = \frac{1}{\pi}\left[\tan^{-1}\left(\frac{\omega}{a}\right)\right]_p^q \qquad \int_p^q \frac{2a}{(2\pi f)^2+a^2}df = \frac{1}{\pi}\left[\tan^{-1}\left(\frac{2\pi f}{a}\right)\right]_p^q$$

ω in radian/s $\qquad$ f in cycle/s or Hz

Spectral Density of Constant or Periodic Signals

Consider a process $X(t) = A + B\cos(2\pi f_0 t + \Theta)$

where *A, B* and f_0 are constant and Θ is a random variable uniformly distributed over $(0,2\pi)$.

Let $X_T(t)$ be a truncated version of $X(t)$ over $(-T, T)$.

$$S_X(\omega) = \lim_{T\to\infty} \frac{E[|F_{X_T}(\omega)|^2]}{2T}$$

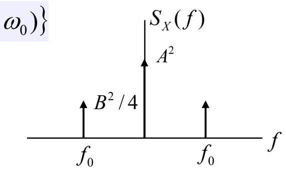
$$F_{X_T}(f) = \int_{-T}^{T}[A + B\cos(2\pi f_0 + \Theta)]e^{-j2\pi ft}dt$$

$$\mathbf{F}[X(t)] = \int_{-\infty}^{\infty}[A + B\cos(2\pi f_0 + \Theta)]e^{-j2\pi ft}dt$$

$$= A\delta(f) + (B/2)\left[\delta(f+f_0)e^{-j\Theta} + \delta(f-f_0)e^{j\Theta}\right]$$

$$S_X(f) = A^2\delta(f) + (B^2/4)\{\delta(f+f_0) + \delta(f-f_0)\}$$

$$S_X(\omega) = 2\pi A^2\delta(\omega) + (\pi B^2/2)\{\delta(\omega+\omega_0) + \delta(\omega-\omega_0)\}$$



The spectral density thus consists of three spikes (delta function) at DC (with height A^2) and at $\pm f_0$ (with height $B^2/4$), respectively.

Mean-Square Value and Total Power

$X(t) = A + B\cos(\omega_0 t + \Theta)$ Θ uniformly distributed in $(0,2\pi)$

$S_X(\omega) = 2\pi A^2\delta(\omega) + (\pi B^2/2)\{\delta(\omega+\omega_0) + \delta(\omega-\omega_0)\}$

Total power:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\{2\pi A^2\delta(\omega) + (\pi B^2/2)[\delta(\omega+\omega_0) + \delta(\omega-\omega_0)]\}d\omega$$

$$= A^2 + \frac{B^2}{4} + \frac{B^2}{4} = A^2 + \frac{B^2}{2}$$

Mean-Square Value of the process

$$E_\Theta[X^2(t)] = E_\Theta\{[A + B\cos(\omega_0 t+\Theta)]^2\} = E_\Theta\{A^2 + 2AB\cos(\omega_0 t+\Theta) + B^2\cos^2(\omega_0 t+\Theta)\}$$

$$= A^2 + E_\Theta\{2AB\cos(\omega_0 t+\Theta)\} + B^2 E_\Theta[\cos^2(\omega_0 t+\Theta)]$$

$E_\Theta\{2AB\cos(\omega_0 t+\Theta)\} = 0$ for $\Theta \sim U(0,2\pi)$

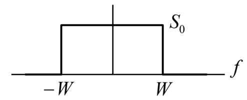
$$E_\Theta\{\cos^2(\omega_0 t+\Theta)\} = E_\Theta\left\{\frac{\cos(2\omega_0 t + 2\Theta)+1}{2}\right\} = \frac{1}{2}$$

Therefore, $E_\Theta[X^2(t)] = \overline{X^2} = A^2 + \frac{B^2}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)d\omega$

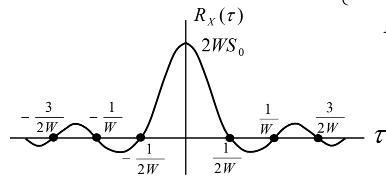
Autocorrelation of Bandlimited White Noise

A more useful concept is the bandlimited white noise whose spectral density is a constant over a finite bandwidth and zero outside the frequency range. For example:

$$S_X(\omega) = \begin{cases} S_0, & |\omega| \le 2\pi W \\ 0, & |\omega| > 2\pi W \end{cases}$$



$$R_X(\tau) = \mathbf{F}^{-1}\{S_X(f)\} = \mathbf{F}^{-1}\left\{S_0 \mathrm{rect}\left(\frac{f}{2W}\right)\right\} = 2WS_0\mathrm{sinc}(2W\tau)$$



$R_X(\tau) = 0$ at $\tau = n/2W$, $n = \pm 1, \pm 2, \cdots$

Random variables from a band-limited white noise are uncorrelated if they are separated in time by any multiple of $1/2W$ seconds.

Therefore, if a continuous time bandlimited white noise process is sampled at twice the maximum frequency limit ($2W$), then the resultant samples of the discrete time sequence are uncorrelated.

Spectral Density of Binary Processes

$$S_X(\omega) = \lim_{T\to\infty}\frac{E[|F_{X_T}(\omega)|^2]}{2T} = \frac{A^2|F_P(\omega)|^2}{t_1}$$

For $p(t) = p_r(t) = \mathrm{rect}(t/t_1)$,

$F_P(\omega) = t_1\mathrm{sinc}(t_1\omega/2\pi)$ and thus $|F_P(\omega)|^2 = t_1^2\mathrm{sinc}^2(t_1\omega/2\pi)$

Therefore, $S_X(\omega) = \frac{A^2|F_P(\omega)|^2}{t_1} = A^2 t_1 \mathrm{sinc}^2(t_1\omega/2\pi)$

For $p(t) = p_c(t) = \frac{1}{2}\left(1 + \cos\frac{2\pi t}{t_1}\right), \quad |t| \le \frac{t_1}{2}; \ = 0, \quad |t| > \frac{t_1}{2},$

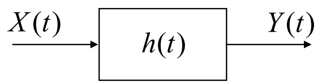
$$F_P(\omega) = \frac{1}{2}\int_{-t_1/2}^{t_1/2}\left(1+\cos\frac{2\pi t}{t_1}\right)e^{-j\omega t}dt = \frac{t_1}{2}\left[\frac{\sin(\omega t_1/2)}{(\omega t_1/2)}\right]\left[\frac{\pi^2}{\pi^2 - (\omega t_1/2)^2}\right]$$

$$S_X(\omega) = \frac{A^2 t_1}{4}\left[\frac{\sin(\omega t_1/2)}{(\omega t_1/2)}\right]^2\left[\frac{\pi^2}{\pi^2 - (\omega t_1/2)^2}\right]^2$$

$$S_X(f) = \frac{A^2 t_1}{4}\mathrm{sinc}^2(t_1 f)\left[\frac{1}{1-(t_1 f)^2}\right]^2$$

Note that in both cases, max $S_X(\omega)$ occurs at $\omega = 0$

Random Input to A System



$x(t) \Rightarrow X(t)$ $y(t) \Rightarrow Y(t)$

The input is no longer a fixed function.

The same linear system concept applies.

Example: $h(t) = \begin{cases} 5e^{-3t}, & t \ge 0 \\ 0, & t < 0 \end{cases}$

$X(t) = M + 4\cos(2t+\Theta)$ where M is a random variable and Θ is an independent random variable, uniformly distributed in $(0,2\pi)$.

$$Y(t) = \int_{-\infty}^{t}[M + 4\cos(2\lambda+\Theta)]5e^{-3(t-\lambda)}d\lambda$$

$$= \frac{5}{3}M + \frac{20}{13}[3\cos(2t+\Theta) + 2\sin(2t+\Theta)]$$

$Y(t)$ is also a random process whose statistical properties can be derived from the distributions of the random variables, M and Θ.

Mean of System Output

As demonstrated in the previous examples,

$$\overline{Y} = E[Y(t)] = E\left[\int_{-\infty}^{\infty} X(t-\lambda)h(\lambda)d\lambda\right]$$

In general, for $E\left[\int_{t_1}^{t_2} Z(t)f(t)dt\right] = \int_{t_1}^{t_2} E[Z(t)]f(t)dt$, it requires that

1. $\int_{t_1}^{t_2} E[|Z(t)|]\,|f(t)|\,dt$
2. $Z(t)$ is bounded on the interval (t_1, t_2) .

In most cases, we assume these conditions are satisfied.

And if $X(t)$ is wide sense stationary with $E[X(t)] = \overline{X}$, then

$$\overline{Y} = \int_{-\infty}^{\infty} E[X(t-\lambda)]h(\lambda)d\lambda = \overline{X}\int_{-\infty}^{\infty} h(\lambda)d\lambda$$

Note that $\int_{-\infty}^{\infty} h(\lambda)d\lambda$ is the dc gain of the system; the dc component of the output is thus equal to the dc component of the input times the dc gain of the system.

Mean-Square Value of System Output

$$\overline{Y^2} = E[Y^2(t)] = E\left[\int_0^{\infty} X(t-\lambda_1)h(\lambda_1)d\lambda_1 \bullet \int_0^{\infty} X(t-\lambda_2)h(\lambda_2)d\lambda_2\right]$$

$$= E\left[\int_0^{\infty} d\lambda_1 \int_0^{\infty} X(t-\lambda_1)X(t-\lambda_2)h(\lambda_1)h(\lambda_2)d\lambda_2\right]$$

$$= \int_0^{\infty} d\lambda_1 \int_0^{\infty} E[X(t-\lambda_1)X(t-\lambda_2)]h(\lambda_1)h(\lambda_2)d\lambda_2$$

But, $E[X(t-\lambda_1)X(t-\lambda_2)] = R_X(t-\lambda_1-t+\lambda_2) = R_X(\lambda_2-\lambda_1)$

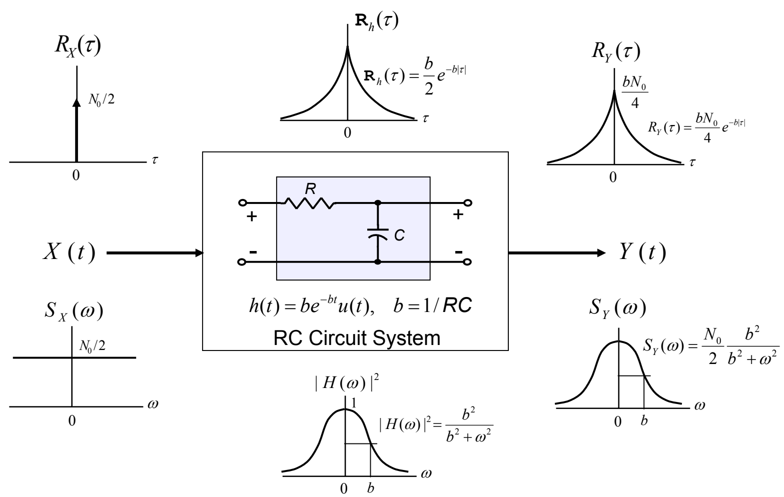
Therefore, $\overline{Y^2} = \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_X(\lambda_2-\lambda_1)h(\lambda_1)h(\lambda_2)d\lambda_2$

Example: For white noise input with $R_X(\tau) = \frac{N_0}{2}\delta(\tau)$, the output "noise" power is

$$\overline{Y^2} = \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_X(\lambda_2-\lambda_1)h(\lambda_1)h(\lambda_2)d\lambda_2$$

$$= \int_0^{\infty} d\lambda_1 \int_0^{\infty} \frac{N_0}{2}\delta(\lambda_2-\lambda_1)h(\lambda_1)h(\lambda_2)d\lambda_2 = \frac{N_0}{2}\int_0^{\infty} h^2(\lambda)d\lambda$$

RC Circuit – Input Output Relationship



Finite-Time Integrator – Input Output Relationship