Appendix A

Probability, Random Variables and Random Processes

In this appendix basic concepts from probability, random processes and signal theory are reviewed.

1. Probability and Random Variables

Probability Space $(\Omega, \mathcal{F}, P)$
Ω is the sample space or set of all possible outcomes.
$\mathcal{F}$ is a collection of events which are subsets of Ω (algebra, field)

$$A \in \mathcal{F}, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F};$$

$$\Omega \in \mathcal{F},$$
$$A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$$

P is a function from $\mathcal{F} \to [0,1]$ which satisfies

i) $0 \leq P(A) \leq 1, \quad A \in \mathcal{F}$

ii) $P(\Omega) = 1$

iii) If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B) \;\; A, B \in \mathcal{F}$

A random variable $X(w)$ is a function from Ω to R

$$X : \Omega \to R$$

that satisfies

$$\{w \in \Omega : X(w) \leq x\} \in \mathcal{F} \;\; \forall x \in R$$

The distribution function $F_X(x)$ of a random variable is defined as

$$F_X(x) = P\{X(w) \leq x\} = P\{w \in \Omega : X(w) \leq x\}$$

Properties of distribution functions

(i) $P\{a < X(w) \leq b\} = F_X(b) - F_X(a)$

(ii) $F_X(x)$ is continuous at x iff $P\{X(w) = x\} = 0$

(iii) $\lim_{y \downarrow x} F_X(y) = F_X(x)$ right continuous

(iv) $\lim_{x\to\infty} F_X(x) = 1, \lim_{x\to-\infty} F_X(x) = 0$

If $F_X(x)$ is continuous for all x then there exists a function $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

This function is called the density function. If a random variable has a density function we shall say the random variable is continuous.

Properties of density functions

(i) $P\{X(w) \in B\} = \int_B f_X(u)du \quad B \subset R$

(ii) $f_X(x) = F'_X(x) = \dfrac{dF_X(x)}{dx}$

If $F_X(x)$ is piecewise constant with a countable number of discontinuities then X is said to be a discrete random variable. For discrete random variables we will use their probability mass function

$$p_X(x) \triangleq P\{X = x\}$$

Expectation of a Random Variable

The expectation of a continuous random variable is

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

The expectation of a discrete random variable is

$$E[X] = \sum_{x:p_X(x)\neq 0} x p_X(x)$$

If $X_1, X_2, \ldots, X_n$ are random variables the joint distribution $F_n(x_1,\ldots,x_n)$ is defined as

$$F_n(x_1,\ldots,x_n) = P\{X_1 \le x_1,\ldots,X_n \le x_n\}$$

If these random variables are (jointly) continuous then their joint density $f_n(x_1,\ldots,x_n)$ is defined as

$$f_n(x_1,\ldots,x_n) = \frac{\partial^n F_n(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}$$

If these random variables are discrete then the joint probability mass function is

$$p_n(x_1,\ldots,x_n) = P\{X_1 = x_1,\ldots,X_n = x_n\}$$

A complex random variable is a function from Ω to $\mathcal{C}$ ($\mathcal{C}$ is the set of complex numbers)

$$X : \Omega \to \mathcal{C}$$

such that $\{\Re X \le x_r, \Im\, X \le x_i\} \in \mathcal{F} \quad \forall x_r, x_i \in R$

$$X(w) = \mathrm{Re}(X(w)) + j\ \mathrm{Im}(X(w)).$$

Useful Bounds

1) <u>Union Bound:</u>

$$\begin{aligned} P(A \cup B) &\le P(A) + P(B) \\ P\left(\bigcup_{i=1}^{M} A_i\right) &\le \sum_{i=1}^{M} P(A_i) \end{aligned}$$

2) <u>Chebyshev Bound:</u> Let $m_X = E[X]$ and $\sigma_X^2 = E[(X-m_X)^2]$ then

$$P\{|X-m_X| \geq \delta\} \leq \frac{\sigma_X^2}{\delta^2}.$$

Proof:

$$\begin{aligned}
\sigma_x^2 &= \int_{-\infty}^{\infty} (x-m_X)^2 f_X(x)dx \\
&\geq \int_{|x-m_X|\geq\delta} (x-m_X)^2 f_X(x)dx \\
&\geq \delta^2 \int_{|x-m_X|\geq\delta} f_X(x)dx \\
&= \delta^2 P\{|X-m_X| \geq \delta\}
\end{aligned}$$

3) <u>Chernoff Bound:</u>

$$P\{X \geq u\} \leq e^{-su}\, E[e^{sX}], \quad s \geq 0$$

Proof:

$$\text{Let} \quad g(x) = \begin{cases} 1, & x \geq u \\ 0, & x < u. \end{cases}$$

Since $s \geq 0$

$$g(x) \leq e^{s(x-u)}.$$

Thus

$$\begin{aligned}
P\{X \geq u\} = \int_u^{\infty} f_X(x)dx &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\
&= E[g(X)] \\
&\leq E[e^{s(X-u)}] = e^{-su}\, E[e^{sX}].
\end{aligned}$$

Example: Let $X_1,\ldots,X_n$ be random variables. Let H_0 and H_1 be two events. Let $p_0(x_1,\ldots,x_n)$ be the conditional density function of $X_1,\ldots,X_n$ given H_0 and $p_1(x_1,\ldots,x_n)$ be the conditional density function of $X_1,\ldots,X_n$ given H_1. Find a bound on

$$\begin{aligned}
P_e &= P\{p_1(X_1,\ldots,X_n) \geq p_0(X_1,\ldots,X_n)|H_0\} \\
&= P\{\frac{p_1(X_1,\ldots,X_n)}{p_0(X_1,\ldots,X_n)} \geq 1 \ |H_0\} \\
&= P\{\ln\left(\frac{p_1(X_1,\ldots,X_n)}{p_0(X_1,\ldots,X_n)}\right) \geq 0 \ |H_0\}
\end{aligned}$$

Let $Y = \ln \frac{p_1(\underline{X})}{p_0(\underline{X})}$

$$\begin{aligned}
P_e = P\{Y \geq 0|H_0] &\leq E[e^{sY}|H_0] \\
&= \int_{R^n} \exp\left[s \ln \frac{p_1(\underline{x})}{p_0(\underline{x})}\right] p_0(\underline{x})d\underline{x}
\end{aligned}$$

A random variable is Gaussian if the density function is

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

where μ is the mean and σ^2 is the variance. The characteristic function of a random variable X is defined as $\phi_X(s) = E[e^s X]$. For a Gaussian random variable the characteristic function is

$$\phi_X(s) = e^{\frac{s^2\sigma^2}{2}+\mu s}$$

Def: A function $g : R^N \to R$ is said to be concave (convex $\cap$) if for any $\underline{x}_1 \in R^n, \underline{x}_2 \in R^n$ and $0 < \theta < 1$

$$\Theta g(\underline{x}_1) + (1-\Theta) g(\underline{x}_2) \le g(\Theta \underline{x}_1 + (1-\Theta)\underline{x}_2)$$

where the vector addition is component-wise addition.
A function $g : R^n \to R$ is said to be convex (convex $\cup$) if

$$\Theta g(\underline{x}_1) + (1-\theta) g(\underline{x}_2) \ge g(\Theta \underline{x}_1 + (1-\Theta)\underline{x}_2)$$

Jensen's Inequality: If $f(x)$ is a concave (convex $\cap$) function mapping $R^n \to R$ then

$$E[f(\underline{X})] \le f(E[\underline{X}]).$$

If $f(x)$ is a convex (convex $\cup$) function mapping $R^n \to R$ then

$$E[f(\underline{X})] \ge f(E[\underline{X}]).$$

Proof for discrete random variables: (By induction) Let $\underline{X}$ take on values $\underline{x}_1, \underline{x}_2$, with nonzero probability

$$\begin{aligned} E[f(\underline{X})] &= p(\underline{x}_1) f(\underline{x}_1) + p(\underline{x}_2) f(\underline{x}_2) \\ &\le f(p(\underline{x}_1)\underline{x}_1 + p(\underline{x}_2)\underline{x}_2) \\ &= f(E[\underline{X}]) \end{aligned}$$

where the first inequality is due to the definition of convexity.
Assume if $\underline{X}$ is discrete taking values

$$\underline{x}_1, \ldots, \underline{x}_{n-1} \text{ then } \qquad \sum_{i=1}^{n-1} p(\underline{x}_i) = 1$$

$$\sum_{i=1}^{n-1} p(\underline{x}_i) f(\underline{x}_i) \quad \le \quad f\left(\sum_{i-1}^{n-1} p(\underline{x}_i) f(x_i)\right)$$

Now let $\underline{X}$ take values $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$

$$\begin{aligned} E[f(\underline{X})] &= \sum_{i=1}^{n} p(\underline{x}_i) f(\underline{x}_i) = \sum_{i=1}^{n-1} p(\underline{x}_i) f(\underline{x}_i) + p(\underline{x}_n) f(\underline{x}_n) \\ \text{Let } \alpha &= \sum_{j=1}^{n-1} p(\underline{x}_j) \\ E[f(\underline{X})] &= \alpha \sum_{i=1}^{n-1} \frac{p(\underline{x}_i)}{\alpha} f(\underline{x}_i) + p(\underline{x}_n) f(\underline{x}_n) \\ & \sum_{i=1}^{n-1} \frac{p(\underline{x}_i)}{\alpha} = 1 \\ E[f(\underline{X})] &\le \alpha f\left(\sum_{i=1}^{n-1} \frac{p(\underline{x}_i)}{\alpha} \underline{x}_i\right) + p(\underline{x}_n) f(\underline{x}_n) \\ &\le f\left(\sum_{i=1}^{n-1} p(\underline{x}_i)\underline{x}_i + p(\underline{x}_n)\underline{x}_n\right) \\ &= f\left(\sum_{i=1}^{n} p(\underline{x}_i)\underline{x}_i\right) \end{aligned}$$

Let $X_1, \ldots, X_n$ be a random vector. The covariance matrix of $X_1, \ldots, X_n$ is defined to be

$$K_X = \begin{bmatrix} K_{1,1} & K_{1,2} & \ldots & K_{1,n} \\ K_{2,1} & & & \ddots \\ & \ddots & & \\ K_{n,1} & & & K_{n,n} \end{bmatrix}.$$

where

$$K_{i,j} = E[(X_i - \mu_i)(X_j - \mu_j)^*].$$

and

$$\mu_i = E[X_i]$$

Def: A $n \times n$ matrix is said to be nonnegative definite if for any vector $(a_1, \ldots, a_n)$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i k_{i,j} a_j^* \geq 0 \quad \text{and} \quad \text{real}$$

i.e.,

$$\mathbf{a} K_X \mathbf{a}^T \geq 0 \text{ and } \quad \text{real}.$$

(positive definite if strict inequality holds).
Claim: The covariance matrix is always nonnegative definite.
Proof:

$$\begin{aligned}
& \sum_{i=1}^{n} \sum_{j=1}^{n} a_i k_{ij} a_j^* \\
& = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i E[(X_i - \mu_i)(X_j - \mu_j)^*] a_j^* \\
= \quad & E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_i (X_i - \mu_i) a_j^* (X_j - \mu_j)^*\right] \\
= \quad & E\left[\sum_{i=1}^{n} a_i (X_i - \mu_i) \sum_{j=1}^{n} a_j^* (X_j - \mu_j)^*\right] \\
= \quad & E\left[\left|\sum_{i=1}^{n} a_i (X_i - \mu_i)\right|^2\right] \geq 0
\end{aligned}$$

Let $X_1, \ldots, X_n$ be a real random vector. The characteristic function of $X_1, X_2, \ldots X_n$ is defined as

$$\Psi_{X_1,\ldots,X_n}(\nu_1, \ldots, \nu_n) = E\left[\exp\left(j \sum_{i=1}^{n} \nu_i X_i\right)\right].$$

Def: The random vector $X_1, \ldots, X_n$ is said to be jointly Gaussian if the characteristic function of $X_1, \ldots, X_n$ is

$$\Psi_{X_1,\ldots,X_n}(\nu_1, \ldots, \nu_n) = \exp\{j \underline{\nu}^T \underline{\mu} - \frac{1}{2} \underline{\nu}^T K \underline{\nu}\}$$

where $\nu^T = (\nu_1, \ldots, \nu_n)$, $\mu^T = (\mu_1, \ldots, \mu_n)$ and K is a real symmetric nonnegative definite $n \times n$ matrix. If K is *positive* definite then the joint density of $X_1, \ldots, X_n$ is

$$p(\mathbf{x}) = (2\pi)^{-1/2} (\det K)^{-1/2} \exp\{-1/2(\mathbf{x} - \underline{\mu})^T K^{-1} (\underline{x} - \mu)\}.$$

Fact: Let $\mathbf{X}$ be a random n vector. Then $\mathbf{X}$ is jointly Gaussian iff $\mathbf{X}$ can be expressed as $W\mathbf{Y} + \mu$ where $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n, W$ is and $n \times n$ matrix and $Y_1, \ldots, Y_n$ are independent mean zero Gaussian random variables (the matrix W can be taken to be orthogonal, i.e. the rows of W are orthogonal).

$$K_x = W K_Y W^T.$$

Now let $\mathbf{X}$ be a jointly Gaussian random vector (of length n) with mean μ covariance matrix K. Let F be a n by n matrix. Consider the random variable

$$Y = \mathbf{X} F \mathbf{X}^T$$

We would like to be able to determine the density function of this random variable. Instead, we will determine the characteristic function of this random variable. The characteristic function is

$$\begin{aligned}\Psi_Y(\nu) &= E[\exp(j\nu Y)] \\ &= \frac{\exp\{j\nu\mu^T(F^{-1}-2j\nu K)^{-1}\mu\}}{\det(I-2j\nu KF)}.\end{aligned}$$

For example, let $n=1$, then $K=\sigma^2$ and

$$\Psi_Y(\nu) = \frac{\exp\{j\nu\mu^2 F/(1-2j\nu\sigma^2 F)\}}{(1-2j\nu\sigma^2 F)}.$$

Inverting this yields the Rician distributed random variable. For $\nu = -js$, $F=1$ the characteristic function becomes

$$E[\exp(sY)] = E[\exp(sX^2)] = \frac{\exp\{s\mu^2/(1-2s\sigma^2)\}}{(1-2s\sigma^2)}.$$

provided that $\mathrm{Re}[s] < 1/2\sigma^2$.

2. Random Processes

Def: A random process $\{X(t); t\in T\}$ is an indexed collection of random variables (i.e. for each $t\in T$, the index set, $X(t)$ is a random variable).
Def: The covariance function of a random process $\{X(t); t\in T\}$ is defined as

$$K(s,t) = E[(X(s)-\mu(s))(X(t)-\mu(t))^*]$$

where $\mu(t) = E[X(t)]$.
Def: A function $K(s,t): R\times R \to R$ is said to be nonnegative definite if for any $n\geq 1$ and time instants $t_1,\ldots,t_n$ and any function $a(t)$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a(t_i)K(t_i,t_j)a^*(t_j) \geq 0 \text{ (and is real)}$$

(positive definite if strict equality holds). Equivalently $\int\int a(t)K(t,s)a^*(s)dtds \geq 0$ and is real.
Claim: The covariance function is a nonnegative definite function.
Def: A random process is said to be Gaussian if for any n and time instances $t_1,\ldots,t_n$, $X(t_1),\ldots,X(t_n)$ is jointly Gaussian.