Appendix A - Probability and Statistics

Appendix A

Probability and Statistics

This appendix provides a summary of probability and random-process basics. The intent is a useful reference to assist with items used throughout the book, but not as a primary source for teaching probability and/or statistics, for which there are many excellent sources, including one for electrical engineers [1], a time-tested classic [2], an advanced treatment [**?**], and an introductory treatment [5].

Section A.1 samples and overviews prerequisite results from basic random variables and probability. Readers viewing this should have some previous exposure to random variables and basic systems analysis. This section is meant only as a refresher and notational consolidator. Section A.2 proceeds to random processes as time-indexed random variables, reviewing results in stationarity (both strict and wide-sense) and reviewing some basic linear-systems results. Section A.3 completes this appendix with specifics of the first two sections when considering complex passband (or baseband) signals.

A.1 Random Variables and Their Probability

Random variables values that are not specific, or are *not deterministic*. This means several possibilities exist, each with a certain likelihood (probability). This text's data transmission study has a few types of random variables of interest:

Random Variables in Data Transmission

Data messages $m \in \{0, ..., M-1\}$ or their consequent symbol representations $x_i \in C = \{x_0 \; ... \; x_{|C|-1}$ ($M = |C|$ when uncoded).

Noise $n \in \mathbb{C}$

channel gain $h \in \mathbb{C}$

user $u \in \{1, ..., U\}$

In general engineering and beyond, of course, there are many more examples, but these above are this text's focus. A good reference with more significant detail and proofs is [**?**]

A.1.1 The random variable, probability distributions and densities

The possibilities for a **discrete random variable** in some finite or countably infinite set C with $|C| \leq \infty$ elements are:

$$x \in C = x_i \in \{x_0 \; ... \; x_{|C|-1}\} \quad , \tag{A.1}$$

where i is an integer index $i \subseteq \mathbb{Z}^+$. In general, any label can apply to any finite set to characterize the random variable (not just a data symbol in $\mathbb{R}$ or $\mathbb{C}$) and the **probability mass function**, $p_x(i)$, specifies probabilities that each random-variable value may occur as:

$$p_x(i) \; i = 0, ..., |C| - 1 \quad , \tag{A.2}$$

where $0 \leq p_x(i) \leq 1 \; \forall \; i = 0, ..., |C| - 1$. The probability that the random variable is in subset $x' \subseteq C' = \{x_0 \; ... \; x_{|C|-1}\}$ with corresponding (possibly relabled) indices $j = 0, ..., |C'| - 1| \leq |C| - 1$ is

$$Pr\{C'\} = \sum_{j \in C'} p_x(i) \leq 1 \quad . \tag{A.3}$$

with $Pr\{C\} = 1$.

Example probability mass functions include: The discrete **uniform mass function** :

$$p_x(i) = \frac{1}{|C|} \quad , \tag{A.4}$$

and the **Poisson Distribution** that characterizes events occurring at rate R in time T

$$p_x(i) = \frac{(R \cdot T)^i \cdot e^{-R \cdot T}}{i!} \quad \forall \; i \; \geq 0 \text{ and } R > 0 \; , \; T \geq 0 \quad . \tag{A.5}$$

The **Bernoulli Distribution** has only two values $x = 1$ or $x = 0$ with respectively probability mass function values p and $q = 1 - p$ respectively, while the **geometric distribution** has countably infinite discrete values $i = 1, ..., \infty$ values with probabilities

$$p_x(i) = p \cdot (1 - p)^{i-1} \quad , \tag{A.6}$$

often corresponding to the first occurrence of a 1 on the i^{th} experimental sample from a Bernoulli distribution. Another somewhat unusual distribution has value i occurring with decaying probability $p_x(i) = \frac{1}{i(i+1)}$ but infinite average (see Subsection A.1.2).

A continuous real **probability density function**, $p_x(u) \geq 0$, measures a **continuous random variable**'s relatively likelihood . Continuous randcom variables take values as real numbers $x \in \mathbb{R}$, or as complex numbers $x \in \mathbb{C}$, with domain $x \in \mathscr{D}_x \subseteq \mathbb{C}$. The probability that the random variable is in continuous region (or set of such regions), $x' \subseteq \mathscr{D}'_x \subseteq \mathscr{D}_x$, is

$$Pr\{\mathscr{D}'_x\} = \int_{u \in \mathscr{D}'_x} p_x(u) \cdot du \leq 1 \quad . \tag{A.7}$$

with $Pr\{\mathscr{D}_x\} = 1$. Often the word **probability distribution** more generally describes either or both of probability mass function and probability density function. In such context, the integral notation most often appears and is equivalent to the sum if the integral is more broadly constructed as a Stieltjes Integral (https://en.wikipedia.org/wiki/Riemann\OT1\textendashStieltjes_integral).

The **cumulative distribution function** for both discrete and continuous real or complex random variables is

$$F_x(X) \triangleq Pr\{x \leq X\} \quad . \tag{A.8}$$

When x is continuous, then

$$p_x(u) = \frac{d}{du} F_x(u) \quad , \tag{A.9}$$

and when x is discrete, the probability distribution function's values are the successive differences between the cumulative distribution function's values.

Example probability densities include the continuous **uniform density**

$$p_x(u) = \frac{1}{d} \; \forall \; d \in [-d/2, d2] \tag{A.10}$$

and the **exponential distribution** (parametrized by λ)

$$p_x(u) = \begin{cases} \lambda \cdot e^{-\lambda u} & u \geq 0 \\ 0 & u < 0 \end{cases} \quad . \tag{A.11}$$

Functions of random variables are random variables. Thus, $f(x)$ is a random variable if x is a random variable. The new distribution is

$$p_y(y) = p_x(f^{-1}(y)) \cdot \frac{df^{-1}(y)}{dy} \quad . \tag{A.12}$$

A.1.2 Mean, Moments, and Variance

A discrete random variable's mean value is

$$\mu_x \triangleq \mathbb{E}[x] = \sum_{i=0}^{|C|-1} i \cdot p_x(i) \tag{A.13}$$

and for the continuous case,

$$\mu_x \triangleq \mathbb{E}[x] = \int_{\mathscr{D}_x} u \cdot p_x(u) \cdot du = \int_{\mathscr{D}_x} u dF_x(u) \quad , \tag{A.14}$$

where the last expression essentially reverses the axes in finding area under a curve. The mean is also the **first moment**; **higher order moments** (about zero) are for real variables:

$$\mathbb{E}[x^n] = \mathbb{E}[x] = \int_{\mathscr{D}_x} u^n \cdot p_x(u) \cdot du \tag{A.15}$$

(the integral is replaced by the sum for the discrete case). When $n = 2$, this is the **mean-square value** or the **autocorrelation** of x if generalized to complex variables as $\mathbb{E}[|x|^2] = \mathbb{E}[x \cdot x^*]$. The expectation operator $\mathbb{E}$ is linear in that $\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y]$. The **variance** is the second moment about the random variable's mean:

$$\sigma_x^2 \triangleq \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] = \int_{\mathscr{D}_x} (u - E[x])^2 \cdot p_x(u) \cdot du \quad . \tag{A.16}$$

The **standard deviation** is the (positive) square root of the variance, σ_x. indexstandard deviation

A.1.3 Moment Generating Functions

The **moment-generating** or **characteristic function** is the (frequency-reversed) Fourier Transform (See Appendix C) of the probability distribution function

$$\phi_x(r) = \mathbb{E}\left[e^{rx}\right] = \int_{u \in \mathscr{D}_{x,r}} p_x(u) \cdot e^{r \cdot u \cdot} \cdot du = P_x(r) \quad , \tag{A.17}$$

where $\mathscr{D}_{x,r}$ is a domain of convergence for the integral that depends on the distribution and choice of r. Often use $r = r = -\jmath\omega$ and then $\Phi_x(\omega)$ is the Fourier Transform of the probability density function.

The moment generating function helps generate moments according to (with $r = -\jmath\omega$)

$$\mathbb{E}\left[x^n\right] = (-\jmath)^n \cdot \frac{d^n\left(\phi_x(\omega = 0)\right)}{d\omega^n} \quad , \tag{A.18}$$

essentially replacing (A.15)'s integration with (often simpler) differentiation to generate the random variable's moments (see https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory)). There is also the **semi-invariant moment generating function**

$$\gamma_x(r) = \ln\left[\phi(r)\right] \tag{A.19}$$

where

$$\frac{d\gamma_x(r)}{dr} = \frac{1}{\phi(r)}\mathbb{E}\left[\frac{d\phi(r)}{dr}\right] = \frac{\mathbb{E}\left[e^{rx} \cdot x\right]}{\phi(r)} \quad , \tag{A.20}$$

so with $\prime$ denoting derivative

$$\gamma'(0) = \mathbb{E}\left[x\right] \quad . \tag{A.21}$$

Similarly

$$\frac{d^2\gamma(r)}{dr^2} = \frac{\phi(r) \cdot \mathbb{E}\left[x^2 \cdot e^{rx} + e^{rx}\right] - \mathbb{E}\left[x \cdot e^{rx}\right] \cdot \mathbb{E}\left[x \cdot e^{rx}\right]}{\phi^2(r)} \tag{A.22}$$

so that

$$\gamma_x''(0) = \sigma_x^2 \quad , \tag{A.23}$$

and $\gamma''(r) > 0$ so convex in r

Name	p_x	$P_x(-\omega)$
Bernoulli	$p \quad 1-p$	$1 - p + p \cdot e^{\jmath\omega}$
Binomial	p, N times	$(1 - p + p \cdot e^{\jmath\omega})^N$
Poisson	$\frac{(R \cdot T)^i \cdot e^{-R \cdot T}}{i!}$	$e^{RT(e^{\jmath\omega} - 1)}$
discrete uniform $b, a \in \mathbb{Z}$, $b > a$	$\frac{1}{b-a+1}$	$\frac{e^{\jmath a\omega} - e^{\jmath(b+1)\omega}}{(b-a+1)\cdot(1-e^{\jmath\omega})}$
continuous uniform	$\frac{1}{b-a}$	$\frac{e^{\jmath b\omega} - e^{\jmath a\omega}}{\jmath\omega(b-a)}$
exponential	$\lambda \cdot e^{-\lambda x}$	$\frac{1}{1 - \jmath\frac{\omega}{RT}}$
Gaussian	$\frac{e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}}{\sqrt{2\pi\sigma_x^2}}$	$e^{\jmath\omega\mu_x - \frac{1}{2}\cdot\sigma_x^2\omega^2}$

A.1.4 Joint random variables

Two random variables also have a joint distribution $p_{x,y}(u, v)$, where this subsection simply uses the continuous form, but the discrete forms follow with the obvious replacement of integrals by sums. The original **marginal** distributions are obtained by

$$p_x(u) = \int_{\mathscr{D}_y} p_{x,y}(u, v) \cdot dv \tag{A.24}$$

$$p_y(v) = \int_{\mathscr{D}_x} p_{x,y}(u, v) \cdot du \quad . \tag{A.25}$$

The probability that x and y are both in region $\mathscr{D}_{x,y}$ is

$$Pr\{(x,y) \in \mathscr{D}_{x,y}\} = \int_{\mathscr{D}_{x,y}} p_{x,y}(u,v) \cdot du \cdot dv \quad . \tag{A.26}$$

Extension to more than 2 random variables appears in Subsection A.1.6.

When the probability distribution factors as

$$p_{x,y}(u,v) = p_x(u) \cdot p_y(v)) \ , \tag{A.27}$$

the two random variables are **independent**. The cross correlation of two random variables is

$$r_{xy} \triangleq \mathbb{E}\left[x \cdot y^*\right] \quad . \tag{A.28}$$

With independent random variables, trivially $r_{xy} = \mathbb{E}\left[x\right] \cdot \mathbb{E}\left[y^*\right]$. The converse does not necessarily imply independence (but does for Gaussian random variables in Subsection A.1.7. The covariance is

$$\sigma_{xy}^2 = \mathbb{E}\left[(x - \mathbb{E}[x]) \cdot (y - \mathbb{E}[y])\right]^* = r_{xy} - \mathbb{E}\left[x\right] \cdot \mathbb{E}\left[y^*\right] \quad . \tag{A.29}$$

The sum of two random variables has variance

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2 \cdot \Re r_{xy} \quad . \tag{A.30}$$

Thus, the variance of a sum is the sum of the variances, so $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$ **if** x **and** y **are independent**. Also, the probability density for the sum of two **independent** random variables is the convolution of their probability densities ([1])

$$p_{x+y}(u) = p_x(u) * p_y(u) \quad . \tag{A.31}$$

A.1.5 Conditional Probability Distributions

The conditional probability distribution furthers the joint distribution in providing a distribution for one random variable y, given a specific value's observation, $x = u$, for the other random variable. This conditional distribution is clearly the ratio of the joint distribution, $p_{x,y}(u,v)$ at the specific value $x = u$ to the marginal distribution $p_y(v)$.

$$p_{y/x}(u,v) = \frac{p_{x,y}(u,v)}{p_y(v)} \quad . \tag{A.32}$$

(A.32)'s inherent symmetry then leads to

$$p_{y/x}(u,v) \cdot p_y(v) = p_{x,y}(u,v) = p_{x/y}(u,v) \cdot p_x(u) \quad , \tag{A.33}$$

sometimes known as **Bayes Theorem** :

Theorem A.1.1 *[Bayes Theorem] The conditional probability distributions satisfy*

$$p_{y/x}(u,v) = \frac{p_{x/y}(u,v) \cdot p_x(u)}{p_y(v)} \quad . \tag{A.34}$$

proof; *Rewrite (A.32).* **QED.**

Also,

$$p_y(v) = \int_{\mathscr{D}_x} p_{y/x}(u,v) \cdot p_x(u) \cdot du \quad . \tag{A.35}$$

Expectations apply to any distribution, and whence to conditional distributions. The implied expectation variable may be written as a subscript, for instance

$$\mathbb{E}_y\left\{\mathbb{E}_x\left[x/y\right]\right\} = \mathbb{E}\left[x\right] = \mathbb{E}_x\left[x\right] \ . \tag{A.36}$$

A.1.6 The random vector

The random vector

$$\boldsymbol{x} = \begin{bmatrix} x_N \\ \vdots \\ x_1 \end{bmatrix} \tag{A.37}$$

has N random-variable components. These N random variables' joint distribution is $p_{\boldsymbol{x}}(\boldsymbol{u})$. The components are independent if $p_{\boldsymbol{x}}(\boldsymbol{u}) = \prod_{n=1}^{N} p_{x_n}(u_n)$. If any two of the components are independent individually in the joint margin distribution p_{x_i,x_j}, for $i \neq j$, the components are said to be **pairwise independent**. Pairwise independence does not necessarily imply independent of all components, but independence necessarily includes pairwise independence. The concept of mean expands to the N-dimensional **mean vector**

$$\boldsymbol{\mu}_{\boldsymbol{x}} \triangleq \mathbb{E}[\boldsymbol{x}] = \begin{bmatrix} \mathbb{E}[x_N] \\ \vdots \\ \mathbb{E}[x_1] \end{bmatrix} \quad , \tag{A.38}$$

while the second moment concept expands to the $N \times N$ **autocorrelation matrix**:

$$R_{\boldsymbol{x}\boldsymbol{x}} = \mathbb{E}[\boldsymbol{x} \cdot \boldsymbol{x}^*] \quad . \tag{A.39}$$

A **covariance matrix** subtracts the mean vectors so that

$$\Sigma_{\boldsymbol{x}\boldsymbol{x}} = \mathbb{E}[(\boldsymbol{x} - \mu_{\boldsymbol{x}}) \cdot (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})^*] = R_{\boldsymbol{x}\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}} \cdot \boldsymbol{\mu}_{\boldsymbol{x}}^* \quad . \tag{A.40}$$

Similarly there is a **cross-correlation matrix** between two random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$$R_{\boldsymbol{x}\boldsymbol{y}} = \mathbb{E}[\boldsymbol{x} \cdot \boldsymbol{y}^*] \quad , \tag{A.41}$$

which is the all zeroes matrix when the two random vectors are uncorrelated. Again, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent, then $p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u},\boldsymbol{v}) = p_{\boldsymbol{x}}(\boldsymbol{u}) \cdot p_{\boldsymbol{y}}(\boldsymbol{v})$. When the vectors are jointly Gaussian (Subsection A.1.7), uncorrelated and independence are the same thing, but not in general.

The **Chain Rule** of probability recursively extends Bayes Theorem to random vectors so that

$$p_{\boldsymbol{x}}(\boldsymbol{u}) = \prod_{n=1}^{N} p_{x_n/[x_{n-1}\ldots x_1]}(u_n, u_{n-1}, \ldots u_1) \quad , \tag{A.42}$$

with $p_{x_1/x_0} \triangleq p_{x_1}$. The vector elements in the chain rule can be in any order.

A.1.7 The Gaussian and the Q-function

The **Q Function** is used to evaluate error probability in digital communication - It is the integral of a zero-mean unit-variance Gaussian random variable's probabiliy density from some specified argument to ∞:

Definition A.1.1 [Q Function]

$$Q(x) = F_x(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du \tag{A.43}$$

The Q-function appears in Figures A.3, A.1, and A.2 for very low SNR (-10 to 0 dB), low SNR (0 to 10 dB), and high SNR (10 to 16 dB) using a very accurate approximation (less than 1% error) formula from Leon-Garcia[1]:

$$Q(x) \approx \left[\frac{1}{\frac{\pi-1}{\pi}x + \frac{1}{\pi}\sqrt{x^2 + 2\pi}} \right] \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \quad . \tag{A.44}$$

For the mathematician at heart, $Q(x) = .5 \cdot erfc(x/\sqrt{2})$, where erfc is known as the complimentary error function by mathematicians.

The integral cannot be evaluated in closed form for arbitrary x, but $Q(0) = 0.5$ and $Q(-\infty) = 1$. Matlab's q.m function evaluates it directly.

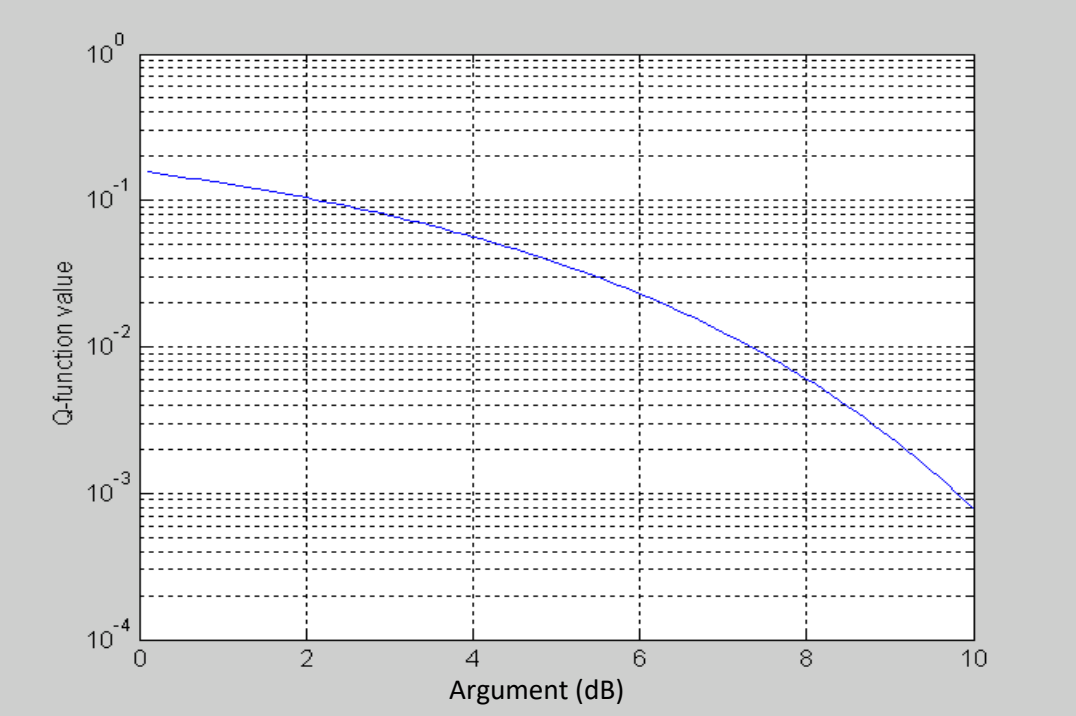

Figure A.1: **Low SNR Q-Function Values**

Figures A.1, A.2, and A.3 have their horizongal axes in dB ($20\log_{10}(x)$). $Q(-x) = 1 - Q(x)$, so valuation need only be for positive arguments. The following bounds apply

$$(1 - \frac{1}{x^2})\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi x^2}} \leq Q(x) \leq \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi x^2}} \tag{A.45}$$

The readily computed upper bound in (A.45) is easily seen to be a very close approximation for $x \geq 3$.

Computation of the probability that a Gaussian random variable u with mean μ_x and variance σ_x^2 exceeds some value d then uses the Q-function as follows:

$$P\{u \geq d\} = Q(\frac{d - \mu_x}{\sigma_x}) \tag{A.46}$$

A.1.8 The Central Limit Theorem

The Central Limit Theorem (see also Wikipedia).

> **Theorem A.1.2 (Central Limit Theorem)** *A scaled* [1] *sample average sum of random variables independently drawn from the same distribution*
>
> $$\sqrt{N} \cdot < x >= \frac{1}{N}\sum_{n=0}^{\infty} x_n \tag{A.47}$$

[1] The scaling here prevents the sum's variance from going to zero.

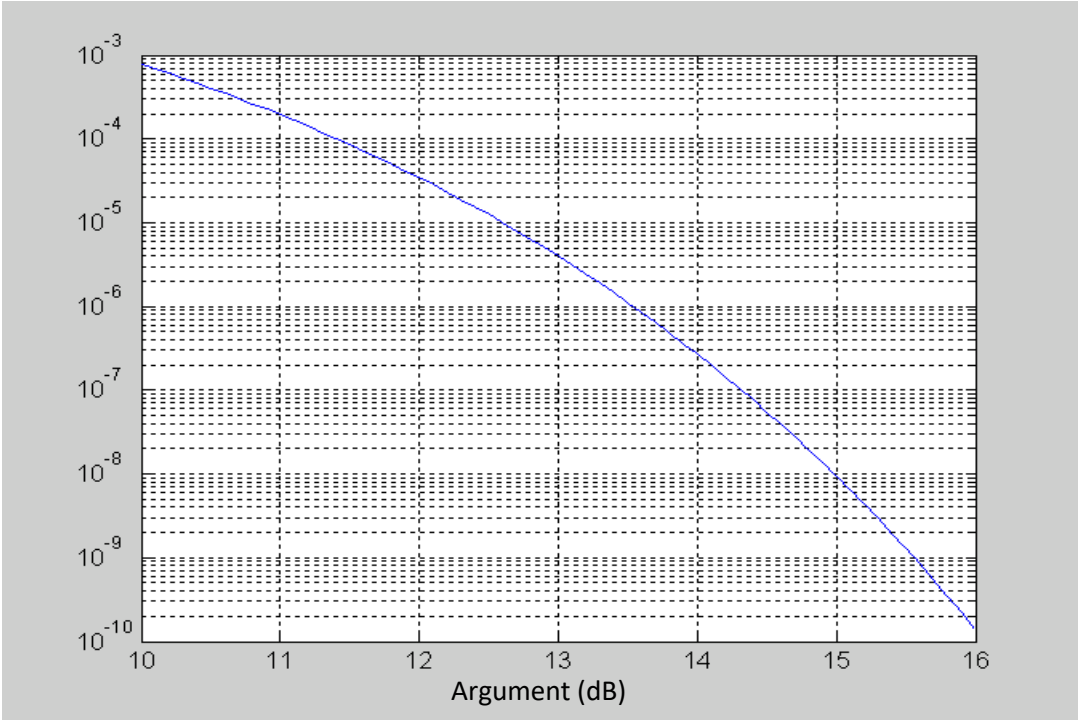

Figure A.2: **High SNR Q-Function Values**

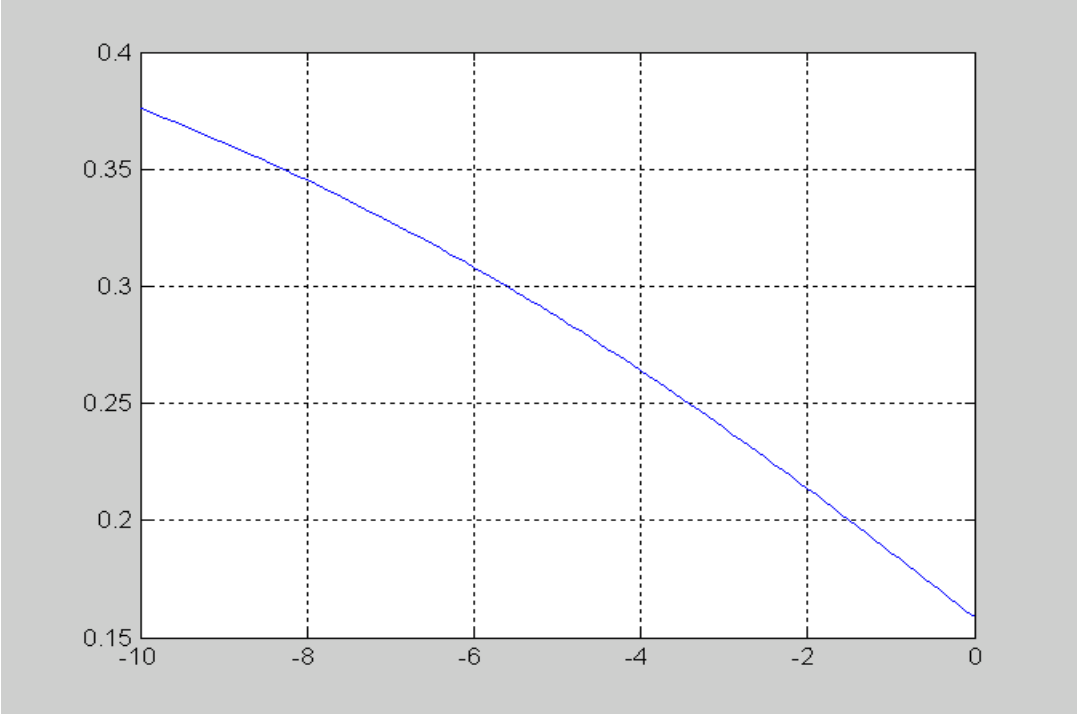

Figure A.3: **Very Low SNR Q-Function Values**

approaches a Gaussian probability density with variance $\frac{1}{N} \cdot \sigma_x^2$ *and mean* $\sqrt{N} \cdot \mu_x$.

Proof: The sum trivially has mean $\mathbb{E}[< x >] = \frac{N}{N} \cdot \mu_x = \mu_x$, and similarly since the random variables are independent variance $\sigma^2_{<x>} = \frac{N}{N^2} \cdot \sigma_x^2 = \frac{1}{N}\sigma_x^2$, so then the variance of $\sqrt{N} \cdot < x >$ is σ_x^2. The Gaussian probability density follows from recognizing the convolution of probability densities corresponds to the product of their characteristic functions, N times. When a random variable is scaled by $a > 0$, the probability density (see A.12) is $\frac{1}{a} \cdot p_x\left(\frac{u}{a}\right)$. This means the terms $x' = \frac{1}{\sqrt{N}} \cdot x_n$ in the sum have probability densities $p_{x'}(u) = p_x(\sqrt{N} \cdot u)$, and thus characteristic functions $\phi_x\left(\frac{\omega}{\sqrt{N}}\right)$. Further, the mean of μ_x can be subtracted from each term in the sum (leaving overall zero mean) for the entire sum in this next proof segment. With independent samples added, the characteristic functions (which are all the same) multiply, so

$$\phi_{<x>}(\omega) = \cdot \left[\phi_x(\frac{\omega}{\sqrt{N}})\right]^N = \left[\phi_x(\frac{\omega}{\sqrt{N}})\right]^N \quad . \tag{A.48}$$

Using a **Taylor Series Expansion** (see Appendix B), and recalling that $\phi_x(\omega) = \mathbb{E}(e^{\jmath\omega})$ and the exponential function has simple Taylors series expansion (the $\omega/\sqrt{N}$ term zeros because the mean is zero).

$$\phi_x(\frac{\omega}{\sqrt{N}}) = 1 - \frac{\omega^2}{2N} + o(\frac{\omega^2}{2N}) \tag{A.49}$$

where the $o(\frac{\omega^2}{2N}$ term decays more rapidly than $\frac{\omega^2}{2N}$ as $N \to \infty$ and the Fourier Transform presumably does not increase with frequency (as it would correspond to infinite energy, see the Paley Weiner Criterion in Appendix D). Then

$$\lim_{N\to\infty}\left[1 - \frac{\omega^2}{2N} + o(\frac{\omega^2}{2N})\right]^N = \left[1 - \frac{\omega^2}{2N})\right]^N = e^{-\frac{1}{2}\omega^2} \quad . \tag{A.50}$$

The latter is recognized as the characteristic function of a unit-variance Gaussian probability density. **QED.**

A.1.9 The Law of Large Numbers

The **Markov Inequality**, for nonnegative random variable x, relates that

$$Pr\{x > x_0\} \leq \frac{\mathbb{E}[x]}{x_0} \quad , \tag{A.51}$$

which implies that larger values are less probable. For instance a value twice the mean, has probability less than half of occurring. This is trivially follows from

$$\mathbb{E}[x] = \underbrace{\mathbb{E}[x/x \leq x_0]}_{\geq x_0} \cdot Pr\{x \geq x_0\} + \underbrace{\mathbb{E}[x/x < x_0] \cdot Pr\{x < x_0\}}_{>0} \tag{A.52}$$

$$\geq x_0 \cdot Pr\{x > x_0\} \tag{A.53}$$

$$\frac{\mathbb{E}[x]}{x_0} \geq Pr\{x > x_0\} \quad . \tag{A.54}$$

Chebychev's inequality also follows from conditional expectations and relates the probability that a random variable deviates more than k standard deviations from its mean to the variance and k.

Lemma A.1.1 *[Chebyshev's Inequality]*

$$Pr\{|x - \mu_x| > k \cdot \sigma_x\} \leq \frac{1}{k^2} \quad . \tag{A.55}$$

Proof:

$$\begin{aligned}
\sigma_x^2 &= \mathbb{E}\left[(x-\mu_x)^2\right] \\
&= \mathbb{E}\left[(x-\mu_x)^2/k\cdot\sigma_x \le |x-\mu_x|\right]\cdot Pr\{k\cdot\sigma_x \le |x-\mu_x|\} + \\
&\quad \mathbb{E}\left[(x-\mu_x)^2/k\cdot\sigma_x > |x-\mu_x|\right]\cdot Pr\{k\cdot\sigma_x > |x-\mu_x|\} \\
&\ge (k\sigma_x^2)\cdot Pr\{k\cdot\sigma_x \le |x-\mu_x|\} + 0\cdot Pr\{k\cdot\sigma_x > |x-\mu_x|\} \\
Pr\{k\cdot\sigma_x \le |x-\mu_x|\} &\le \frac{1}{k^2}
\end{aligned}$$

QED.

Chebyshev's inequality helps prove the law of large numbers (LLN), which basically says that the sample average converges to the mean if the samples are independent and drawn from the same distribution.

Theorem A.1.3 *[Law of Large Numbers]*

Weak Form*:*

$$\lim_{N\to\infty} Pr\{|< x > -\mu_x| < \epsilon\} = 1 \ . \tag{A.56}$$

Strong Form:

$$\lim_{N\to\infty} Pr\{< x >= \mu_x\} = 1 \ . \tag{A.57}$$

Proof: *Already established are* $\mathbb{E}[< x >_N] = \mu_x$ *and* $\sigma^2_{<x>_N} = \frac{\sigma_x^2}{N}$*. Chebychev's equality as* $N \to \infty$ *establishes the weak form. The strong form follows from*

$$Pr\{|< x >_N - \mu_x| < \epsilon\} = 1 - Pr\{|< x >_N - \mu_x| \ge \epsilon\} \ge 1 - \frac{\sigma_x^2}{N\cdot\epsilon^2} \ , \tag{A.58}$$

which tends to 1 *as* N *becomes large.* **QED.**

Estimating probability distribution values by counting (binning) the number of occurrences of a specific value (or small ranges of values in continuous case) computes the **relative frequency**.. Such calculation corresponds to each such calculation to a sample mean of the indicator function (1 if the value occurs, 0 otherwise). Indeed, then the strong LLN implies that these relative frequencies converge with probability 1 to the actual probability values. Thus, computing relative frequencies (see for instance Chapter 4) can be an accurate means for computing probability distributions when models themselves may not be accurate.

The **Chernoff bound** follows directly from Markov's inequality

$$Pr\{x \ge x_0\} = Pr\{e^{rx} \ge e^{rx_0}\} \le \frac{\mathbb{E}\left[e^{rx}\right]}{e^{rx_0}} = \frac{\phi_x(r)}{e^{rx_0}} \ , \tag{A.59}$$

which can be made tight by finding the r value that makes the bound as low as possible:

$$Pr\{x \ge x_0\} \le \min_{r\ge 0} e^{-rx_0}\cdot\phi_x(r) \ . \tag{A.60}$$

This bound when optimized (see [3]) is found to be exact for the sample mean $< x >$ as $N \to \infty$. Chapter 2 largely uses the LLN in constructing the AEP, which essentially avoids much of the need for the other bounds in this Section, but they appear here for completeness.

A.1.10 Memoryless Random Variable

A memoryless random variable is such that

$$Pr\{x > t + u\} = Pr\{x > u\} \cdot Pr\{x > t\} \quad . \tag{A.61}$$

The only probability density that satisfies this property is the exponential distribution, for which $Pr\{x > u\} = F_x(u) = e^{-\lambda \cdot u}$.

A.2 Random Processes

Special thanks are due to Dr. James Aslanis who wrote a version of this section when a Professor at the University of Hawaii.

Random processes add a time-index to random variables, aggregating into a vector or time series multiple samples from the random variable's density/distribution at different times. In the most general case, the probability density/distribution p_x may vary at the different sampling times, leading to a **time-variant random process** as opposed to a **stationary random process** when the distribution is invariant to time shifts.

The natural mathematical construct to describe a noisy communication signal is a random process. Random processes are simply a generalization of random variables. Consider a single sample of a random process, which is described by a random variable x with probability density $p_x(u)$. (Often the random process is assumed to be Gaussian but not always). This random variable only characterizes the random process for a single instant of time. For a finite set of time instants, a random vector $\boldsymbol{x} = [x_{t_1}, \ldots, x_{t_n}]$ with a joint probability density function $P_{\boldsymbol{x}}(\boldsymbol{u})$ describes the random process. Extension to a countably infinite set of random variable samples, indexed by n, defines a discrete-time random process.

Definition A.2.1 (Discrete-Time Random Process) *A Discrete-Time Random Process is a countably infinite (or finite), indexed set of random variables described by a joint density function $p_{\boldsymbol{x}}(\boldsymbol{u})$ where $\boldsymbol{x} = [x_{t_i}, \ldots, x_{t_{i+n}}]$, where i is an integer and n is a positive integer. $\boldsymbol{u}$ contains particular sample values as inputs to the probability density/distribution.*

The random variables in a random process need not be independent nor identically distributed, although the random variables all sample the same domain. Similarly, the mean $E[f(X_t)]$ is a function of the index. The mean, also known as an **ensemble average**, should not be confused with averaging over the time index (sometimes referred to as the **sample mean**). For example the mean value of a random variable associated with the tossing of a die is approximated by averaging the values over many independent tosses.

When the sample mean converges to the ensemble average, the random process is **mean ergodic**). In a random process, each random variable in the collection of random variables may have a different probability density function. Thus, time averaging, or the sample mean, over successive samples may not yield any information about the ensemble averages.

Random processes are classified by statistical properties that their density functions obey, the most important of which is stationarity.

A.2.1 Stationarity

Definition A.2.2 [Strict Sense Stationarity] *A discrete andom process x_t is called* **strict sense stationary (SSS)** *if the joint probability density function*

$$p_{x_{t_1},\ldots,x_{t_n}}(u_{t_1}, \ldots, u_{t_n}) = p_{x_{t_1+t},\ldots,x_{t_n+t}}(u_{t_1+t}, \ldots, u_{t_n+t}) \; \forall \; n, t, \{t_1, \ldots, t_n\} \quad . \quad \text{(A.62)}$$

Roughly speaking, the statistics of X_t are invariant to a time shift; i.e. the placement of the origin $t = 0$ is irrelevant.

This text next considers commonly calculated functions of a random process: These functions encapsulate properties of the random processes, and a linear-systems analysis sometimes calculates these functions for processes without knowing the exact process, probability density. Certain random processes, such as stationary Gaussian processes, are completely described by a collection of these functions.

Definition A.2.3 (Mean) *The mean of a random process X_t is*

$$\mathbb{E}(x_t) = \int_{-\infty}^{\infty} u_t \cdot p_{x_t}(u_t) \cdot du_t \triangleq \mu_X(t) \quad . \tag{A.63}$$

A second-moment-like quanity is the autocorrelation function

Definition A.2.4 [Autocorrelation Function] *A random process x_t has an autocorrelation function:*

$$\mathbb{E}(x_{t_1} \cdot x^*_{t_2}) = \int_{-\infty}^{\infty} u_{t_1} \cdot u^*_{t_2} \cdot p_{x_{t_1} x_{t_2}}(u_{t_1}, u_{t_2}) \cdot du_{t_1} \cdot du_{t_2} \triangleq r_x(t_1, t_2) \tag{A.64}$$

In general, the autocorrelation is a two-dimensional function of the pair $\{t_1, t_2\}$. For a stationary process, the autocorrelation is a one-dimensional function of the time difference $\tau \triangleq t_1 - t_2$ only:

$$\mathbb{E}(x_{t_1} \cdot x^*_{t_2}) = r_x(t_1 - t_2) = r_x(\tau) \tag{A.65}$$

The stationary process' autocorrelation also satisfies a Hermitian property $r_x(\tau) = r^*_x(-\tau)$.

Using the mean and autocorrelation functions, also known as the first- and second-order moments, designers often define a weaker form of stationarity.

Definition A.2.5 [Wide Sense Stationarity] *A random process X_t is called* **wide sense stationary (WSS)** *if*

1. $\mathbb{E}(x_t) =$ *constant,*
2. $\mathbb{E}(x_{t_1} \cdot x^*_{t_2}) = r_x(t_1 - t_2) = r_x(\tau)$*, i.e. a function of the time difference only.*

While SSS $\Rightarrow$ WSS, WSS $\not\Rightarrow$ SSS. Often, random processes' analysis only considers their first- and second-order statistics. Such results do not reveal anything about the random process' higher-order statistics; however, the Gaussian needs only the lower-order statistics. In particular,

Definition A.2.6 [Gaussian Random Process] *The joint probability density function of a stationary real* **Gaussian random process** *for any set of n indices $\{t_1, \ldots, t_n\}$ is*

$$p_{\boldsymbol{x}}(\boldsymbol{u}) = \frac{1}{(2\pi)^{n/2}(|\Sigma_{\boldsymbol{xx}}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{u} - \boldsymbol{\mu}_x)\Sigma^{-1}_{\boldsymbol{xx}}(\boldsymbol{x} - \boldsymbol{\mu}_x)'\right] \quad . \tag{A.66}$$

A complex Gaussian random variable has independent Gaussian random variables in both the real and imaginary parts, both with the same variance, which is half the variance of the complex random variable. Then, the distribution is

$$p_{\boldsymbol{x}}(\boldsymbol{u}) = \frac{1}{(\pi)^{n}(|\Sigma_{\boldsymbol{xx}}|} \exp\left[-(\boldsymbol{x} - \boldsymbol{\mu}_x)\Sigma^{-1}_{\boldsymbol{xx}}(\boldsymbol{x} - \boldsymbol{\mu}_x)'\right] \quad . \tag{A.67}$$

For a Gaussian random process, the set of random variables $\{x_{t_1}, \ldots, x_{t_n}\}$ are jointly Gaussian. A Gaussian random process also satisfies the following two important properties:

1. The output response of a linear time-invariant system to a Gaussian input is also a Gaussian random process.

2. A WSS, real-valued, Gaussian random process is SSS.

Much of this textbook's analysis will considers Gaussian random processes passed through linear time-invariant systems. As a result of the properties listed above, the designer only requires these processes' mean and autocorrelation functions for complete characterization. Fortunately, the designer can calculate the effect of linear time-invariant systems on the randcom process without explicitly using the probability densities/distributions.

In particular, for a linear time-invariant system defined by an impulse response $h(t)$, the mean of the output random process Y_t is

$$\mu_y(t) = \mu_x(t) * h(t) \quad . \tag{A.68}$$

The output autocorrelation function is

$$r_y(\tau) = h(\tau) * h^*(-\tau) * r_x(\tau) \quad . \tag{A.69}$$

In addition, many analyses use the correlation between the input and output random processes:

Definition A.2.7 [Cross-correlation] *The* **cross-correlation** *between the random processes X_t and Y_t is given by*

$$E(x_{t_1} y^*_{t_2}) \triangleq r_{xy}(t_1, t_2) \quad . \tag{A.70}$$

For a jointly WSS random processes, the cross-correlation only depends on the time difference

$$r_{xy}(t_1, t_2) = r_{xy}(t_1 - t_2) = r_{xy}(\tau) \quad . \tag{A.71}$$

The cross-correlation $r_{XY}(\tau)$ does not satisfy the Hermitian property that the autocorrelation obeys, but

$$r_{xy}(\tau) = r^*_{yx}(-\tau) \quad . \tag{A.72}$$

Further

$$r_{xy}(\tau) = r_x(\tau) * h^*(-\tau) \tag{A.73}$$

$$r_{yx}(\tau) = r_x(\tau) * h(\tau) \tag{A.74}$$

A more general form of stationarity is **cyclostationarity**, wherein the random process' density/distribution is invariant only to specific index shifts.

Definition A.2.8 [Strict-Sense Cyclostationarity] *A random process is* **strict sense cyclostationary** *if the joint probability density function satisfies*

$$p_{x_{t_1},\ldots,x_{t_n}}(u_{t_1}, \ldots, u_{t_n}) = p_{x_{t_1+T},\ldots,x_{t_n+T}}(u_{t_1+T}, \ldots, u_{t_n+T}) \; \forall \; n, \{t_1, \ldots, t_n\}, \tag{A.75}$$

where T is called the **period** *of the process.*

That is, x_{t+kT} is statistically equivalent to $x_t \; \forall \; t, k$. Cyclostationarity accounts for the regularity in communication transmissions that repeat a particular operation at specific time intervals; however, within a particular time interval, the statistics are allowed to vary arbitrarily. As with stationarity, a weaker form for cyclostationarity that depends only on the first and second order statistics of the random process is.

Definition A.2.9 [Wide Sense Cyclostationarity] *A random process is* **wide sense cyclostationary** *if*

1. $\mathbb{E}(x_t) = \mathbb{E}(x_{t+kT})\ \forall\ t, k.$
2. $r_x(t+\tau, t) = r_x(t+\tau+kT, t+kT)\ \forall\ t, \tau, k.$

Thus, the mean and autocorrelation functions of a WS cyclostationary process are periodic functions with period T. Many random signals in communications, such as an ensemble of modulated waveforms, satisfy the WS cyclostationarity properties.

The periodicity of a WS cyclostationary random process would complicate the study of modulated signals without use of the following convenient property. Given a WS cyclostationary random process x_t with period T, the random process $x_{t+\theta}$ is WSS if θ is a uniform random variable over the interval $[0, T]$. Thus, analysis often shall include (or assume) a random phase θ to yield a WSS random process.

Alternatively for a WS cyclostationary random process, there is a time-averaged autocorrelation function.

Definition A.2.10 [Time-Averaged Autocorrelation] *The* **time-averaged autocorrelation** *of a WS cyclostationary random process x_t is*

$$\bar{r}_x(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} r_x(t+\tau, t)\cdot dt \quad . \tag{A.76}$$

Since the autocorrelation function $r_X(t+\tau, t)$ is periodic, integration could be over any closed interval of length T in A.76.

As in the study of deterministic signals and systems, frequency-domain descriptions are often useful for analyzing random processes. First, this appendix continues with the definitions for deterministic signals.

Definition A.2.11 [Energy Spectral Density] *The* **Energy Spectral Density** *of a finite energy deterministic signal $x(t)$ is $|X(\omega)|^2$ where*

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\cdot e^{-j\omega t} dt \triangleq \mathcal{F}\{x(t)\} \tag{A.77}$$

is the Fourier transform of $x(t)$. Thus, the energy is calculable as

$$\mathcal{E}_x \triangleq \int_{-\infty}^{\infty} |x(t)|^2 \cdot dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2 \cdot d\omega < \infty \tag{A.78}$$

If the finite-energy signal $x(t)$ is nonzero for only a finite time interval, say T, then the time average power in the signal equals $P_x = \mathcal{E}_x/T$.

Communication signals are usually modeled as repeated patterns extending from $(-\infty, \infty)$, in which case the energy is infinite, although the time average power may be finite.

Definition A.2.12 [Power Spectral Density] *The* **power spectral density** *of a finite power signal defined as*

$$S_x(\omega) = \lim_{T\to\infty} \frac{|X_T(\omega)|^2}{T} \tag{A.79}$$

where $X_T(\omega) = \mathcal{F}(x_T(t))$ is the Fourier transform of the truncated signal

$$x_T(t) = \begin{matrix} x(t) & |t| < \frac{T}{2} \\ 0 & \text{otherwise} \end{matrix} \tag{A.80}$$

Thus, the time average power is calculable as $P_x \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |x(t)|^2 \cdot dt = \int_{-\infty}^{\infty} S_x(\omega) \cdot d\omega < \infty$

For deterministic signals, the power P_x is a time-averaged quantity.

For a random process, strictly speaking, the Fourier transform does not exactly specify the power spectral density. Even if the random process is well-behaved, the result would be another random process. Instead, ensemble averages are required for frequency-domain analysis.

For a random process x_t, the ensemble average power, $P_{x_t} \triangleq \mathbb{E}[|x_t|^2]$, may vary instantaneously over time. For a WSS random process, however, $P_{x_t} = P_x$ is a constant.

Definition A.2.13 [Power Spectral Density] *For a WSS continuous-time random process X_t, the* **power spectral density** *is*

$$S_X(\omega) = \mathcal{F}\{r_x(\tau)\}. \tag{A.81}$$

Also, $P_x = \int_{-\infty}^{\infty} S_X(\omega) \cdot d\omega$.

For a WS cyclostationary random process x_t the autocorrelation function $r_x(t+\tau, t)$, for a fixed time lag τ, is periodic in t with period T. Consequently the autocorrelation function can be expanded using a Fourier series.

$$r_x(t+\tau, t) = \sum_{n=-\infty}^{\infty} \gamma_n(\tau) \cdot e^{j2\pi nt/T}, \tag{A.82}$$

where $\gamma_n(\tau)$ are the Fourier coefficients

$$\gamma_n(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} r_x(t+\tau, t) \cdot e^{-j2\pi nt/T} \cdot dt \quad . \tag{A.83}$$

The time average autocorrelation function is then

$$\frac{1}{T} \int_{-T/2}^{T/2} r_x(t+\tau, t) \cdot e^{-j2\pi nt/T} \cdot dt = \gamma_0(\tau). \tag{A.84}$$

The average power for the WS cyclostationary random process is

$$P_{X_t} = \frac{1}{T} \int_{-T/2}^{T/2} E[|X_t|^2] \cdot dt = \bar{r}_X(0) = \gamma_0(0) = \int_{-\infty}^{\infty} G_0(f) \cdot df \tag{A.85}$$

where $G_0(f)$ is the Fourier transform of the $n = 0$ Fourier coefficient

$$\mathcal{F}\{\gamma_0(\tau)\} = G_0(f) = \mathcal{F}\{\bar{r}_X(\tau)\}. \tag{A.86}$$

The function $G_0(f)$ is the power spectral density of the WS cyclostationary random process X_t associated with the time average autocorrelation.

For a nonstationary random process, the average power must be calculated by both time and ensemble averaging, i.e. $P_{X_t} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} E[|X_t|]^2 dt$.

A.2.2 Random Processes and Linear Systems

Followinbg (A.69), the frequency-domeain equivalent is

$$S_{yy}(\omega) = |H(\omega)|^2 \cdot S_{xx}(\omega) \quad . \tag{A.87}$$

These same relations hold in discrete time also with $\omega \to e^{-\jmath\omega}$. For more on Fourier Transforms and well-behaved functions, see Appendix D.

A.3 Passband Processes

This appendix investigates properties of the correlation functions for a WSS passband random process $x(t)$ in its several representations. For a brief introduction to the definitions of random processes see Section A.2.

A.3.1 The Hilbert Transform

The Hilbert Transform finds frequent use in passband signal processing, and is is a linear operator that shifts the phase of a sinusoid by 90^o:

Definition A.3.1 (Hilbert Transform) *The Hilbert Transform of $x(t)$ is denoted $\check{x}(t)$ and is given by*

$$\check{x}(t) = \hbar(t) * x(t) = \int_{-\infty}^{\infty} \frac{x(u)}{\pi(t-u)} \cdot du \quad , \tag{A.88}$$

where

$$\hbar(t) = \begin{cases} \frac{1}{\pi t} & t \neq 0 \\ 0 & t = 0 \end{cases} \quad . \tag{A.89}$$

The Fourier Transform of $\hbar(t)$ is

$$\begin{aligned} \mathcal{H}(\omega) &= \int_{-\infty}^{\infty} \frac{e^{-\jmath\omega t}}{\pi t} \cdot dt & \text{(A.90)} \\ &= \int_{-\infty}^{\infty} \frac{-\jmath \sin(\omega t)}{\pi t} \cdot dt & \text{(A.91)} \\ &= \int_{-\infty}^{\infty} \frac{-\jmath \sin(\pi u)}{\pi u} \text{sgn}(\omega) \cdot du & \text{(A.92)} \\ &= -\jmath \left[\int_{-\infty}^{\infty} \text{sinc}(u) \cdot du \right] \cdot \text{sgn}(\omega) & \text{(A.93)} \\ &= -\jmath \cdot \text{sgn}(\omega) & \text{(A.94)} \end{aligned}$$

Equation (A.94) shows that a frequency component at a positive frequency is shifted in phase by -90^o, while a component at a negative frequency is shifted by $+90^o$. Summarizing

$$\check{X}(\omega) = -\jmath \cdot \text{sgn}(\omega) X(\omega) \quad . \tag{A.95}$$

Since $|\mathcal{H}(\omega)| = 1 \; \forall \; \omega \neq 0$, then $|X(\omega)| = |\check{X}(\omega)|$, assuming $X(0) = 0$. This text only considers passband signals with no energy present at DC ($\omega = 0$). Thus, the Hilbert Transform only affects the phase and not the magnitude of a passband signal.

A.3.1.1 Examples

Let

$$x(t) = \cos(\omega_c t) = \frac{1}{2} \cdot (e^{\jmath\omega_c t} + e^{-\jmath\omega_c t}) \quad , \tag{A.96}$$

then

$$\check{x}(t) = \frac{1}{2} \cdot (-\jmath e^{\jmath\omega_c t} + \jmath e^{-\jmath\omega_c t}) = \frac{1}{2\jmath} \cdot (e^{\jmath\omega_c t} - e^{-\jmath\omega_c t}) = \sin(\omega_c t) \quad . \tag{A.97}$$

Let

$$x(t) = \sin(\omega_c t) = \frac{1}{2\jmath} \cdot (e^{\jmath\omega_c t} - e^{-\jmath\omega_c t}) \quad , \tag{A.98}$$

then

$$\check{x}(t) = \frac{1}{2j} \cdot (-\jmath e^{\jmath\omega_c t} - \jmath \cdot e^{-\jmath\omega_c t}) = -\frac{1}{2}(e^{\jmath\omega_c t} + e^{-\jmath\omega_c t}) = -\cos(\omega_c t) \quad . \tag{A.99}$$

Note

$$\check{\check{x}}(t) = \hbar(t) * \hbar(t) * x(t) = -x(t) \quad . \tag{A.100}$$

since $(-j \cdot \text{sgn}(\omega))^2 = -1 \ \forall \ \omega \neq 0$. A correct interpretation of the Hilbert transform is that every sinusoial component is passed with the same amplitude, but with its phase reduced by 90 degrees.

A.3.1.2 Inverse Hilbert

The inverse Hilbert Transform is easily specified in the frequency domain as

$$\mathcal{H}^{-1}(\omega) = \jmath \cdot \text{sgn}(\omega) \quad , \tag{A.101}$$

or then

$$\hbar^{-1}(t) = -\hbar(t) = \begin{cases} -\frac{1}{\pi t} & t \neq 0 \\ 0 & t = 0 \end{cases} \quad . \tag{A.102}$$

A.3.2 Hilbert Transform of Passband Signals

Given a passband signal $x(t)$, form the quadrature decomposition

$$x(t) = x_I(t) \cdot \cos(\omega_c t) - x_Q(t) \cdot \sin(\omega_c t) \tag{A.103}$$

and transform $x(t)$ into the frequency domain

$$X(\omega) = \frac{1}{2} \cdot [X_I(\omega + \omega_c) + X_I(\omega - \omega_c)] - \frac{1}{2\jmath} \cdot [X_Q(\omega - \omega_c) - X_Q(\omega + \omega_c)] \quad . \tag{A.104}$$

Equation A.104 shows that if $X(\omega) = 0 \ \forall \ |\omega| > 2\omega_c$ then $X_I(\omega) = 0$ and $X_Q(\omega) = 0 \ \forall \ |\omega| > \omega_c$.[2] Using this fact the Hilbert transform $\check{X}(\omega)$ is given by

$$\check{X}(\omega) = \frac{\jmath}{2} \cdot [X_I(\omega + \omega_c) - X_I(\omega - \omega_c)] + \frac{1}{2} \cdot [X_Q(\omega - \omega_c) + X_Q(\omega + \omega_c)] \ . \tag{A.105}$$

The inverse Fourier Transform of $\check{X}(\omega)$ then yields

$$\check{x}(t) = x_I(t) \cdot \sin(\omega_c t) + x_Q(t) \cdot \cos(\omega_c t) = \Im\{x_A(t)\} \ , \tag{A.106}$$

where $\Im$ denotes the imaginary part.

A.3.2.1 Hilbert Transform of a Random Process

Let $x(t)$ be a WSS real-valued random process and $\check{x}(t) = \hbar(t) * x(t)$ be its Hilbert transform. By Equation (A.69), the autocorrelation of $\check{x}_t$ is

$$r_{\check{x}}(\tau) = \hbar(\tau) * \hbar^*(-\tau) * r_x(\tau) = r_x(\tau). \tag{A.107}$$

Since $|\mathcal{H}(\omega)|^2 = 1 \quad \forall \ \omega \neq 0$, and assuming $S_X(0) = 0$, then

$$S_{\check{x}}(\omega) = S_x(\omega) \quad . \tag{A.108}$$

Thus, a WSS random process and its Hilbert Transform have the same autocorrelation function and the same power spectral density.

By Equation (A.74)

$$r_{\check{x},x}(\tau) = \hbar(\tau) * r_x(\tau) = \check{r}_x(\tau) \tag{A.109}$$

$$r_{x,\check{x}}(\tau) = \hbar^*(-\tau) * r_x(\tau) = -\hbar(\tau) * r_x(\tau) = -\check{r}_x(\tau) \quad . \tag{A.110}$$

[2] Recall $x_I(t)$ and $x_Q(t)$ are real signals.

The cross correlation between the random process $x(t)$ and its Hilbert transform $\check{x}(t)$ is the Hilbert transform of the autocorrelation function of the random process $x(t)$.

Thus also,

$$r_{\check{x},x}(\tau) = \hbar(\tau) * r_x(\tau) = \hbar^*(\tau) * r_x(\tau) = \hbar^*(\tau) * r_x(-\tau) = r^*_{x,\check{x}}(-\tau) = r_{x,\check{x}}(-\tau) \quad . \tag{A.111}$$

By using Equations (A.109), (A.110) and (A.111),

$$r_{\check{x},x}(\tau) = \check{r}_x(\tau) = -r_{x,\check{x}}(\tau) = -r_{\check{x},x}(-\tau) \quad . \tag{A.112}$$

Equation (A.112) implies that $r_{\check{x},x}(\tau)$ is an odd function, and thus

$$r_{\check{x},x}(0) = 0 \; . \tag{A.113}$$

That is, a real-valued random process and its Hilbert Transform are uncorrelated at any particular point in time.

A.3.2.2 Quadrature Decomposition

The quadrature decomposition for any real-valued WSS passband random process and its Hilbert transform is

$$\begin{aligned} x(t) &= x_I(t) \cdot \cos\omega_c t - x_Q(t) \cdot \sin\omega_c t & \text{(A.114)} \\ \check{x}(t) &= x_I(t) \cdot \sin\omega_c t + x_Q(t) \cdot \cos\omega_c t. & \text{(A.115)} \end{aligned}$$

The baseband equivalent complex-valued random process is

$$x_{bb}(t) = x_I(t) + \jmath \cdot x_Q(t) \tag{A.116}$$

and the analytic equivalent complex-valued random process is

$$x_A(t) = x(t) + \jmath \cdot \check{x}(t) = x_{bb}(t) \cdot e^{j\omega_c t} \; . \tag{A.117}$$

The original random process can be recovered as

$$x(t) = \Re\left\{x_A(t)\right\} \quad . \tag{A.118}$$

The autocorrelation of $x_A(t)$ is

$$\begin{aligned} r_A(\tau) &= E\left\{x_A(t)x_A^*(t-\tau)\right\} & \text{(A.119)} \\ &= 2 \cdot (r_x(\tau) + \jmath\check{r}_x(\tau)) \quad . & \text{(A.120)} \end{aligned}$$

The right hand side of Equation (A.120) is twice the analytic equivalent of the autocorrelation function $r_x(\tau)$. The power spectral density is

$$S_A(\omega) = 4 \cdot S_x(\omega) \quad \omega > 0 \; . \tag{A.121}$$

The functions in the quadrature decomposition of $x(t)$ also have autocorrelation functions:

$$\begin{aligned} r_I(\tau) &\triangleq E\left\{x_I(t) \cdot x_I^*(t-\tau)\right\} & \text{(A.122)} \\ r_Q(\tau) &\triangleq E\left\{x_Q(t) \cdot x_Q^*(t-\tau)\right\} & \text{(A.123)} \\ r_{IQ}(\tau) &\triangleq E\left\{x_I(t) \cdot x_Q^*(t-\tau)\right\} & \text{(A.124)} \end{aligned}$$

Then also,

$$\begin{aligned} r_x(\tau) &= E\left\{x(t) \cdot x^*(t-\tau)\right\} & \text{(A.125)} \\ &= r_I(\tau) \cdot \cos(\omega_c t)\cos(\omega_c(t-\tau)) \\ &- r_{IQ}(\tau) \cdot \cos(\omega_c t) \cdot \sin(\omega_c(t-\tau)) \\ &- r_{QI}(\tau) \cdot \sin(\omega_c t) \cdot \cos(\omega_c(t-\tau)) \\ &+ r_Q(\tau) \cdot \sin(\omega_c t) \cdot \sin\omega_c(t-\tau) \end{aligned}$$

Standard trigonometric identities simplify (A.125) to

$$\begin{aligned} r_x(\tau) &= \frac{1}{2}\cdot[r_I(\tau)+r_Q(\tau)]\cdot\cos(\omega_c\tau) \\ &+\frac{1}{2}\cdot[r_{IQ}(\tau)-r_{QI}(\tau)]\cdot\sin(\omega_c\tau) \\ &-\frac{1}{2}\cdot[r_Q(\tau)-r_I(\tau)]\cdot\cos\omega_c(2t-\tau) \\ &-\frac{1}{2}\cdot[r_{IQ}(\tau)+r_{QI}(\tau)]\cdot\sin(\omega_c(2t-\tau)) \end{aligned} \tag{A.126}$$

Strictly speaking, most modulated waveforms are WS cyclostationary with period $T_c = \frac{2\pi}{\omega_c}$, i.e. $\mathbb{E}\left[x(t)\cdot x^*(t-\tau)\right] = r_x(t,t-\tau) = r_x(t+T_c,t+T_c-\tau)$. For cyclostationary random processes a time-averaged autocorrelation function of one variable τ can be defined by $r_x(\tau) = 1/T_c\int_{-T_c/2}^{T_c/2} r_x(t,t-\tau)dt$, and this new time-averaged autocorrelation function will satisfy the properties derived thus far in this section. The next set of properties require that the random process to be WSS, not WS cyclostationary – or equivalently the time-averaged autocorrelation function $r_x(\tau) = 1/T_c\int_{-T_c/2}^{T_c/2} r_x(t,t-\tau)dt$ can be used. An example of a WSS random process is AWGN. Modulated signals often have equal energy inphase and quadrature components with the inphase and quadrature signals derived independently from the incoming bit stream; thus, the modulated signal is then WSS.

For $x(t)$ to be WSS, the last two terms in (A.126) must equal zero. Thus, $r_I(\tau) = r_Q(\tau)$ and $r_{IQ}(\tau) = -r_{QI}(\tau) = -r_{IQ}(-\tau)$. The latter equality shows that $r_{IQ}(\tau)$ is an odd function of τ and thus $r_{IQ}(0) = 0$. For $x(t)$ to be WSS, the inphase and quadrature components of $x(t)$ have the same autocorrelation and are uncorrelated at any particular instant in time. Substituting back into Equation A.126,

$$r_x(\tau) = r_I(\tau)\cdot\cos(\omega_c\tau) - r_{QI}(\tau)\cdot\sin(\omega_c\tau) \tag{A.127}$$

Equation A.127 expresses the autocorrelation $r_x(\tau)$ in a quadrature decomposition and thus

$$\check{r}_x(\tau) = r_I(\tau)\cdot\sin(\omega_c\tau) + r_{QI}(\tau)\cdot\cos(\omega_c\tau) \tag{A.128}$$

Further algebra leads to

$$r_{bb}(\tau) = E\left\{x_{bb}(t)\cdot x_{bb}^*(t-\tau)\right\} \tag{A.129}$$

$$= 2\cdot(r_I(\tau)+jr_{QI}(\tau)) \tag{A.130}$$

$$r_A(\tau) = r_{bb}(\tau)\cdot e^{j\omega_c\tau} \tag{A.131}$$

The power spectral density is

$$S_A(\omega) = S_{bb}(\omega-\omega_c) \quad . \tag{A.132}$$

If $\mathcal{S}_x(\omega)$ is symmetric about ω_c, then $S_{bb}(\omega)$ is symmetric about $\omega = 0$ (recall that the spectrum $S_A(\omega)$ is a scaled version of the positive frequencies of $S_X(\omega)$ and $S_{bb}(\omega)$ is $S_A(\omega)$ shifted down by ω_c). In this case $r_{bb}(\tau)$ is real, and using Equation A.130, $r_{QI}(\tau) = 0$. Equivalently, the inphase and quadrature components of a random process are uncorrelated at any lag τ (not just $\tau = 0$) if the power spectral density is symmetric about the carrier frequency. Finally,

$$r_x(\tau) = \frac{1}{2}\cdot\Re\left\{r_{bb}(\tau)e^{j\omega_c\tau}\right\} = \frac{1}{2}\cdot\Re\left\{r_A(\tau)\right\} \quad . \tag{A.133}$$

If $x(t)$ is a random modulated waveform, by construction it is usually true that $r_I(\tau) = r_Q(\tau)$ and $r_{IQ}(\tau) = -r_{QI}(\tau) = 0$, so that the constructed $x(t)$ is WSS. For AWGN, $n(t)$ is usually WSS so that $r_I(\tau) = r_Q(\tau)$ and $r_{IQ}(\tau) = -r_{QI}(\tau) = 0$. When a QAM waveform is such that $r_{IQ}(\tau) \neq -r_{QI}(\tau)$ or $r_I(\tau) \neq r_Q(\tau)$, then $x(t)$ is WS cyclostationary with period π/ω_c.

A.4 Markov Processes

Markov processes often characterize the distribution dynamics of time-varying processes, but have more general mathematical distribution, ss in Section 1.6. While the Markov Process is time-varying, it often has a stationary distribution that characterizes the likelihood of being in any one of a number, $|\mathscr{A}|$, of different states. The Markov process specifically has a probability of being in a next state that only depends on the previous state:

$$p_{k/i} = p_{k/i,i-1,\ldots 0} \quad . \tag{A.134}$$

The discrete (special case of a) Markov model has an $|\mathscr{A}| \times |\mathscr{A}|$ probability-transition matrix with entries $p_{k/j}$ representing the probability that the state k will occur when the current/last state is j. The probability matrix is then

$$P \triangleq \begin{bmatrix} p_{|\mathscr{A}|-1/|\mathscr{A}|-2} & p_{|\mathscr{A}|-1/|\mathscr{A}|-3} & \cdots & p_{|\mathscr{A}|-1/0} \\ \vdots & \vdots & \ddots & \vdots \\ p_{0/|\mathscr{A}|-2} & p_{1/|\mathscr{A}|-3} & \cdots & p_{0/0} \end{bmatrix} \quad . \quad . \tag{A.135}$$

Every non-zero conditional-probability entry in P corresponds to a transition possibility in Figure A.4's directed graph of the state machine. Often channel uses order in time, but of course order could also most generally include any path though sets of used dimensions in space, time, and/or frequency. Zero probability entries have no transition shown.

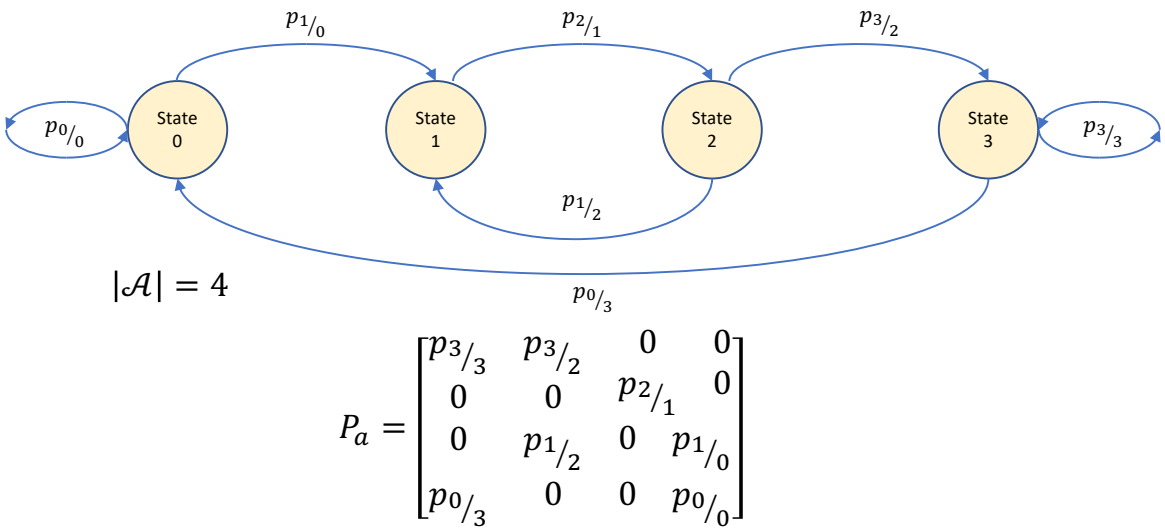

Figure A.4: **Markov State-Machine Model, with non-zero transition probabilities shown.**

Distant past is unimportant in Markov processes, only the last state matters. The selection of a next state may be associated with a random variable with probability distribution characterized by an input that determines to which state the Markov process next proceeds. That input has a distribution that corresponds to a column of P. When, as always the case in this text, that input is stationary, P is a constant matrix of nonnegative entries. Each column also sums to unity value. There may be a random value associated with each state that is observed and not directly the state itself, and where that value is a function of the state and perhaps jointly with another random variable that is independent of the state, which then crease a **Hidden Markov Process**. This latter HMP often characterizes a time-varying channel where the H varies according to P but independent noise is added to create a channel output. In this case, the transition between states does not depend on an input. However, in other situations

the channel H may be fixed and the transmitted symbol causes the state to change; the number of states $|\mathscr{A}| = M$ (number of symbol constellation values). This second example often corresponds leads to maximum-likelihood sequence decoders and again the noise is independent of the channel input. Yet another (not hidden) example uses a Markov process to characterize the interarrival times between message packets/symbols at the channel input, where the arrival of the next packet only depends on the time since the last arrival. Markov processes are often also characterized a state-probability distribution $\boldsymbol{p}_k$ where

$$\boldsymbol{p}_k = P \cdot \boldsymbol{p}_{k-1} \quad . \tag{A.136}$$

The matrix P has all nonnegative entries, as does $\boldsymbol{p}$. Because of the possibility of zero entries in P, non-degenerate Markov processes satisfy an additional condition that

$$P^n_{k/i} > 0 \text{ for some } n > 0 \text{ and } P^{n'}_{i/k} > 0 \text{ for some } n' > 0 \ , \tag{A.137}$$

which means that any state will eventually be reached from any other state with nonzero probability[3]. That is, the Markov Process is **irreducible**. A periodic state i occurs when $P^n_{i/i} \neq 0$ only if $n = k \cdot d$ for $k, d \in \mathbb{Z}^+$, and $d > 1$ is the period. If $d = 1$, it is aperiodic. If no states are periodic, the Markov Process is aperiodic.. If the MP is both irreducible and aperiodic, then for large enough number of successive transitions, occurring starting at any point in time, some integer power of the matrix P will contain all positive (nonzero) entries. Then by Perron-Frobenious theory[4], such a matrix has a largest positive-real eigenvalue that dominates $P^{n\to\infty}$ and a corresponding all-positive-entry eigenvector. This eigenvector $\boldsymbol{\pi}$ satisfies also

$$P \cdot \boldsymbol{\pi} = \boldsymbol{\pi} \ , \tag{A.138}$$

and is the **stationary vector** or probability distribution (when normalized so that it sums to one). When a stationary distribution exists, the Markov Process is balanced so that the sum of the probabilities of going to other states out of any state j equals the probability of arriving in state j from all other states, thus satisfying the **balance equation**:

$$\left(\sum_i p_{i/j} \right) \cdot p_j = \sum_i P_{j/i} \cdot p_i \quad . \tag{A.139}$$

The balance equation simplifies for many arrival-process situations as in Section A.5.

A.4.1 Birth-Death Processes

Birth-Death Markov Processes are a special case of interest to queuing. They always have a stationary distribution and satisfy the properties:

$$P_{i/j} = 0 \text{ in } |i-j| > 1 \text{ and} \tag{A.140}$$
$$P_{i/i+1} > 0 \tag{A.141}$$
$$p_{i+1/i} > 0 \tag{A.142}$$

Birth-death processes well model arrival processes necessarily also satisfy

$$P \cdot \boldsymbol{p}_0 = P^t \cdot \boldsymbol{p}_0 \quad . \tag{A.143}$$

[3]Equivalently for some large $n'' \geq \max(n, n')$, then $P^{n''} \succeq 0$, so it is as if the transition matrix for n'' transitions is all positive and then satisfies Peron Frobenious condition of all positive entries.

[4]See the Wikipedia page at https://en.wikipedia.org/wiki/Perron\OT1\textendashFrobenius_theorem.

A.5 Queuing Theory

In this text, renewal processes are basically random processes that attempt to model the arrival or departure of messages. There is an **interarrival interval** that becomes random, essentially signalling a burst of messages to be sent. Often these messages are queued with perhaps a steady output flow (perhaps augmented by dummy packets) into a modulator that transmits b bits/symbol. The depth of the queue indicates roughly the need for use/service by the channel. For this text's purposes, the queue can be possibly infinite in length/delay to avoid consideration of "buffer overflow" errors that typically find solution at higher layers. Instead, this appendix simply investigates random modeling of the arrivals, by the common Poisson distribution or other potential distributions. The book by Gallager [?] (Chapter 2) is an excellent more detailed reference on this subject. See also Giambene.

A.5.1 Arrival Processes

An **arrival process** is discrete that models with a random process the times at which a message arrives (more generally when an event occurs), $t_i > 0$, $i = 1, ..., \infty$ and $t_j > t_i$ if $j > i$. The stationary random-variable distribution from which this processes samples are selected permits only positive values. These "times" have separations or **interarrival** times

$$\tau_i = t_i - t_{i-1} > 0 \quad , \tag{A.144}$$

which is also a random process with only positive values. Clearly,

$$t_n = \sum_{i=1}^{n} \tau_i \quad . \tag{A.145}$$

The instants t_n theoretically need not be integer multiples of some basic time unit T (like a sampling clock's zero crossings), but in practice any system would likely approach that integer multiple if arrivals are slightly delayed until the start of the next symbol as would be necessary in most practice. A continuous-time random process can count the number of arrivals that have occurred before or at any time $t > 0$, so

$$\mathcal{B}(t) \quad = \quad n \;\; t \in [t_n, t_{n+1}) \;\; \text{or in terms of sets} \tag{A.146}$$

$$t_n \leq t \quad \Rightarrow \quad \mathcal{B}(t) \geq n \quad . \tag{A.147}$$

So, $t_1 > t$ implies that $\mathcal{B}(t) = 0$.

A.5.2 Renewal and the Poisson Distribution

A **renewal process** is an arrival process where the arrival-time random-process samples t_i are independent of each other. A **Poisson** renewal process has the arrival increments distributed exponentially with **arrival rate** $\lambda \in \mathbb{R}^+$, so

$$p_\tau(u) = \lambda \cdot e^{\lambda \cdot u} \;\; u \geq 0 \quad , \tag{A.148}$$

or equivalently

$$Pr\{\tau > \tau'\} = e^{-\lambda \cdot \tau'} \quad . \tag{A.149}$$

As earlier, the exponential distribution is memoryless so

$$Pr\{\tau > u + v\} = Pr\{\tau > u\} \cdot Pr\{\tau > v\} \quad . \tag{A.150}$$

Since $t_n = \sum_{i=1}^{n} \tau_n$ is a sum of independent random variables, then the distribution of t_n derives as the convolution of the exponential distribution $n - 1$ times or the **Erlang density**function (a continuous random variable)

$$p_{t_n}(t) = \frac{\lambda^n \cdot t^{n-1} \cdot e^{-\lambda \cdot t}}{(n-1)!} \quad . \tag{A.151}$$

The discrete-valued distribution for $\mathcal{B}(t)$ derives from looking at the integral of the Erlang density over the infinitessimmally small interval $(t < t_{n+1} < t + \delta)$ as in [?], which then provides the Poisson probability mass function

$$p_{\mathcal{B}(t)}(n) = \frac{(\lambda \cdot t)^n \cdot e^{-\lambda \cdot t}}{n!} \quad . \tag{A.152}$$

A.5.3 Queue Modelling

Queuing convention describes (known as Kendall's notation https://en.wikipedia.org/wiki/Kendall%27s_notation) queues according to a notation, which is simplified here for this text's purposes, arrival-distribution/channel-processing-distribution/U/queue-depth. This text uses only the letters M, G, or D for the arrival and channel-processing distributions, where M corresponds to Poisson arrivals (so exponential inter-arrival or channel-processing times), G for general distribution, and D for deterministic (non random). Very common in this text is the notation M/D/U queue, which means Poisson arrival process, the channel immediately processes its inputs without delay, and U users. The simplest queue is the $M/M/1$ queue (which will behave like the $M/G/1$ queue for this text's purposes, leaving some formulas for $M/M/1$ that simplify for $M/D/1$.

The Poisson processes' exponentially distributed inter-arrival time has probability for some measured τ since last arrival $Pr\{t > \tau\} = 1 = e^{-\lambda \cdot \tau}$. Description by a state-machine or Markov model, with matrix entires $p_{n/n\pm1}$ requires looking at the instantaneous probability because multiple arrivals could occur in any finite time. For infinitely small time increment dt, the probability becomes $\lambda \cdot dt$, and similarly for the channel-entry times, $b \cdot dt$.

A.5.3.1 M/M/1 Queue:

Figure A.5 illustrates the corresponding state machine.

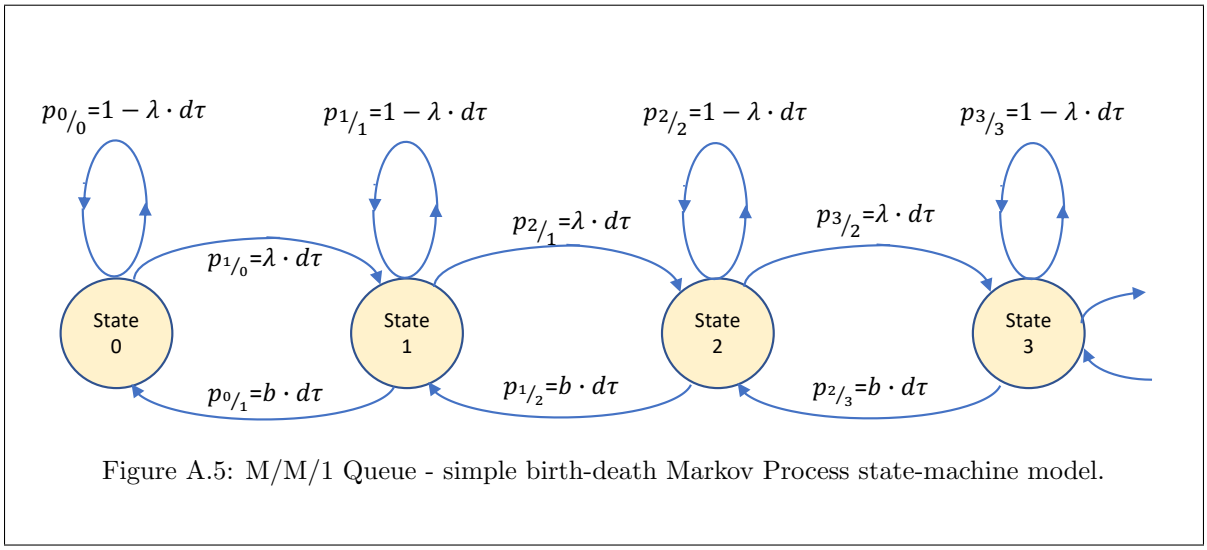

Figure A.5: M/M/1 Queue - simple birth-death Markov Process state-machine model.

The balance equation simplifies in this case to

$$\lambda \cdot \pi_n = b \cdot \pi_{n+1} \tag{A.153}$$

Defining $\rho = \lambda/b$, it is clear that $\lambda \leq b$ for queue stability, and

$$\rho_{n+1} = \rho^n \cdot (1 - \rho) \tag{A.154}$$

This leads to calculation of the delays and number of symbols ($\mathcal{B}$) as

$$\mathcal{B} \quad = \quad \frac{\rho}{1-\rho} = \frac{\lambda}{b-\lambda} \tag{A.155}$$

$$\Delta = \frac{1}{b-\lambda} \tag{A.156}$$

$$\Delta_q = \frac{\rho}{b-\lambda} \tag{A.157}$$

$$\mathcal{B}_q = \frac{\rho^2}{1-\rho} \ . \tag{A.158}$$

The variance of the delay depends on scheduling policy and can be complicated to compute. The M/M/1 queue is sometimes called a **statistical multiplexer** because as the users arrive randomly they are served by (typically) a high-speed channel. To understand this better, the input may be viewed as U users' messages/symbols all multiplexed to one channel. Each of the users may have an average data rate of λ/U, which of course sum over all users to λ. The aggregate as well as the individual message streams are Poisson processes, just the users are slower individually. A **deterministic multiplexer** instead equally apportions b/U of the channel bandwidth to each user (presuming tacitly the capacity region allows such equal apportionment). It has then delay

$$\Delta_{det} = \frac{1}{\lambda/U - b/U} = \frac{U}{b-\lambda} = U \cdot \Delta_{stat} \ . \tag{A.159}$$

The statistical multiplexer thus has less delay than would have say a TDM (or OFDM - see Chapter 4) system. However, a slightly different queue is the **D/D/1 queue where** each user is presumably streaming at data rate λ/U and matched to its turn in the channel that runs at $b \geq \lambda$. The D/D/1 system has constant delay of one symbol and no variation, and no chance of buffer overflow. While the M/M/1 has significantly lower average delay for $U > 1$, very large delays can have nonzero probability. Thus, the trade-off to use statistical multiplexing (over deterministic) is not so simple, eventually depending on the relative user rates and the scheduling method and maximum tolerable delay. Systems with many users continually streaming (e.g., video), especially when the aggregate data rate approaches the channel sum-capacity, are better served by deterministic scheduling. Other systems, with more infrequent use (whether low data rate or high data rate for any particular user) may be better served by a statistical multiplexer. A method intermediate to Chapter 2's QPS (queue-proportional scheduling) that considers the channel capacity region, an intermediate queue of interest investigates a system for which there are U parallel channels each carrying b bits/symbol, as in the next subsection.

A.5.3.2 M/M/U Queue

The M/M/U system views a MAC as having U users on a single channel (common transmitter and receiver,), each of rate $b_u = b$, so the sum rate is $b_{sum} = U \cdot b$. This does not directly match any multiuser physical-layer channel, any of which have either or both of physically isolated transmitters and/or receivers. Nonetheless, the theory naturally follows the M/M/1 queue and so appears here. The state machine has return to state $u-1$ from state u that can relieve up to $u \leq U$ users needs simultaneously, which the $u \cdot b \cdot dt$ in the lower transitions represents. The quantity ρ remains as λ/b, but now can have a value as large as U. Otherwise, analysis proceeds similar to the M/M/1 case and the balance equation becomes:

$$\lambda \cdot \pi_{n-1} = \begin{cases} n \cdot b \cdot \pi_n & n \leq U \\ U \cdot b \cdot \pi_n & n > U \end{cases} \tag{A.160}$$

The solution, using much algebra, for the steady state probabilities is

$$\pi_n = \begin{cases} \left[\sum_{n=0}^{U-1} \frac{\rho^n}{n} + \frac{U \cdot \rho^U}{U(U-\rho)}\right]^{-1} & n = 0 \\ \pi_0 \cdot \frac{\rho^n}{n} & 0 \leq n \leq U \\ \pi_0 \cdot \frac{\rho^n}{U} & 0 \leq n \leq U \end{cases} \ . \tag{A.161}$$

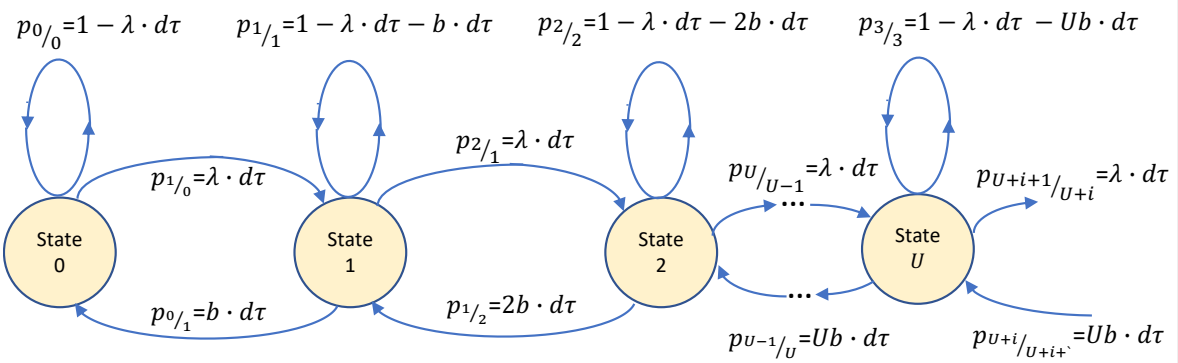

Figure A.6: M/M/U Queue - U-user birth-death Markov Process state-machine model.

This leads to calculation of the delays and number of symbols ($\mathcal{B}$) as

$$\mathcal{B} = \rho + \frac{\pi_0 \cdot \rho^U}{(U-1)! \cdot (U-\rho)^2} \tag{A.162}$$

$$\Delta = \frac{1}{b} + \frac{1}{\lambda} \cdot \frac{\pi_0 \cdot \rho^{U+1}}{(U-1)! \cdot (U-\rho)^2} \tag{A.163}$$

$$\Delta_q = \frac{1}{\lambda} \cdot \frac{\pi_0 \cdot \rho^{U+1}}{(U-1)! \cdot (U-\rho)^2} \tag{A.164}$$

$$\mathcal{B}_q = \frac{\pi_0 \cdot \rho^{U+1}}{(U-1)! \cdot (U-\rho)^2} \quad . \tag{A.165}$$

Another related quantity of interest is the probability that a user has to wait in queue

$$P_Q = \sum_{n=m}^{\infty} \cdot \pi_n = \frac{\pi_0 \cdot \rho^U}{(U-1)! \cdot (U-\rho)} \quad , \tag{A.166}$$

which is known as the **Erlang C** formula [4]. The M/M/U queue also has an interesting large-number-of-users property that the system essentially creates a Poisson arrival process to whatever comes next:

$$\lim_{U\to\infty} \pi_n = \pi_0 \frac{1}{n!} \cdot \left(\frac{\lambda}{b}\right)^n \tag{A.167}$$

$$= \frac{e^{-\lambda/b}}{n!} \cdot \left(\frac{\lambda}{b}\right)^n \tag{A.168}$$

$$\lim_{U\to\infty} \mathcal{B} = \rho \tag{A.169}$$

A.5.3.3 Finite Queues and Blocking

Any implementable queue must have finite memory and can only handle a finite number of users, so the M/M/U queue then adjusts to the closely related M/M/U/U queue that blocks user inputs when in state U, as in Figure A.7.

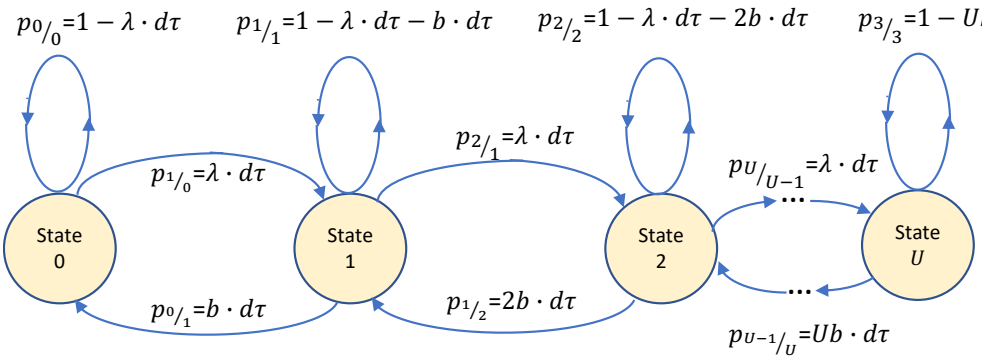

Figure A.7: M/M/U Queue - U-user birth-death Markov Process state-machine model.

The relationship $\rho < U$ (or $\rho \leq 1$ for a single user, meaning channel must run faster than inputs on average) need no longer hold, but there is a nonzero probability that messages are simply not sent, which is π_U. The balance equations provide in this case

$$\pi_n = \pi_0 \cdot \frac{\rho^n}{n!} \quad , \tag{A.170}$$

which then yields (probabilities for states 0 to U must add to 1)

$$\pi_0 = \frac{1}{\sum_{n=0}^{U} \rho^n / n!} \quad . \tag{A.171}$$

The blocking probability is then

$$P_B = \pi_U = \pi_0 \cdot \frac{\rho^U}{U!} \quad , \tag{A.172}$$

which is the **Erlang B** formula. This formula can be recursively computed in the number of users according to[5]

$$P_B^{-1}(\rho, U) = P_B^{-1}(\rho, U-1) + \frac{U}{\rho \cdot P_B(\rho, U-1)} \quad . \tag{A.173}$$

Further algebra leads to

$$\mathcal{B} = \rho \cdot (1 - P_b) \tag{A.174}$$

$$\Delta = \frac{1}{b} \quad , \tag{A.175}$$

so the delay is basically the channel processing time for a symbol, but that may occur with unacceptable blocking probability.

[5]The inverse form has less dynamic range in quantities so is easier to calculate in finite precision.

A.6 Gram-Schmidt Orthonormalization Procedure

This appendix illustrates the construction of a set of orthonormal basis functions $\varphi_n(t)$ from a set of modulated waveforms $\{x_i(t),\ i = 0, ..., M-1\}$. The process for doing so, and achieving minimal dimensionality is called **Gram-Schmidt Orthonormalization**.

Step 1:

Find a signal in the set of modulated waveforms with nonzero energy and call it $x_0(t)$. Let

$$\varphi_1(t) \triangleq \frac{x_0(t)}{\sqrt{\mathcal{E}_{\boldsymbol{x}_0}}} \quad , \tag{A.176}$$

where $\mathcal{E}_{\boldsymbol{x}} = \int_{-\infty}^{\infty}[x(t)]^2 dt$. Then $\boldsymbol{x}_0 = \left[\sqrt{\mathcal{E}_{\boldsymbol{x}_0}}\ 0\ ...\ 0\right]$.

Step i for $i = 2, ..., M$:

- Compute $x_{i-1,n}$ for $n = 1, ..., i-1$ ($x_{i-1,n} \triangleq \int_{-\infty}^{\infty} x_{i-1}(t)\varphi_n(t)dt$).
- Compute

$$\theta_i(t) \triangleq x_{i-1}(t) - \sum_{n=1}^{i-1} x_{i-1,n}\varphi_n(t) \tag{A.177}$$

- if $\theta_i(t) = 0$, then $\varphi_i(t) = 0$, skip to step $i+1$.
- If $\theta_i(t) \neq 0$, compute

$$\varphi_i(t) = \frac{\theta_i(t)}{\sqrt{\mathcal{E}_{\theta_i}}} \quad , \tag{A.178}$$

where $\mathcal{E}_{\theta_i} = \int_{-\infty}^{\infty}[\theta_i(t)]^2 dt$. Then $\boldsymbol{x}_{i-1} = \left[x_{i-1,1}\ ...\ x_{i-1,i-1}\ \sqrt{\mathcal{E}_{\theta_i}}\ 0\ ...\ 0\right]'$.

Final Step:

Delete all components, n, for which $\varphi_n(t) = 0$ to achieve minimum dimensional basis function set, and reorder indices appropriately.

Bibliography

[1] Alberto Leon-Garcia. *"Probability, Statistics, and Random Processes for Electrical Engineering", 3rd Edition.* Pearson Education, USA, 2007.

[2] Anthanasios Papoulis. *"Probability, Random Variables, and Stochastic Processes".* McGraw-Hill, New York, 1965.

[3] Robert G. Gallager *"Stochastic Processes: Theory for Application".* Cambridge Press, UK, 2013. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/

[4] Dimitris Bertsekas and Robert Gallager *"Data Networks, Second Edition".* Prentice Hall, Englewood Cliffs,NJ, 1992.

[5] Sheldon Ross *"Introduction to Probability and Statistics for Engineers and Scientists", 6th Edition.* Elsevier, Academic Press, 2020.

Index

1

2

3

4

5

6

7

8