APPENDIX

Hints from Probability and Statistics

A.1 TOTAL PROBABILITY AND THE BAYES RULE

Let $\mathcal{A}_i, i = 1, 2, \ldots, M$, be M events so that $\sum_{i=1}^{M} P(\mathcal{A}_i) = 1$. Then the probability of an arbitrary event $\mathcal{B}$ is given by

$$P(\mathcal{B}) = \sum_{i=1}^{M} P(\mathcal{B}|\mathcal{A}_i)P(\mathcal{A}_i) \tag{A.1}$$

where $P(\mathcal{B}|\mathcal{A})$ denotes the conditional probability of $\mathcal{B}$ assuming $\mathcal{A}$, which is defined as

$$P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{B}, \mathcal{A})}{P(\mathcal{A})} \tag{A.2}$$

and $P(\mathcal{B}, \mathcal{A})$ is the joint probability of the two events. Equation (A.1) is known as the *total probability theorem*. From the definition in (A.2) the Bayes rule is readily available

$$P(\mathcal{B}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B}) \tag{A.3}$$

These are easily extended to random variables or vectors described by probability density functions and we have

$$p(\boldsymbol{x}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{A}|\boldsymbol{x})p(\boldsymbol{x}) \tag{A.4}$$

and

$$p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \tag{A.5}$$

and finally

$$p(\boldsymbol{x}) = \sum_{i=1}^{M} p(\boldsymbol{x}|\mathcal{A}_i)P(\mathcal{A}_i) \tag{A.6}$$

A.2 MEAN AND VARIANCE

Let $p(x)$ be the probability density function (pdf) describing the random variable x. Its mean and variance are defined as

$$E[x] = \int_{-\infty}^{+\infty} xp(x)\,dx, \quad \sigma_x^2 = \int_{-\infty}^{+\infty} (x - E[x])^2 p(x)\,dx \tag{A.7}$$

A.3 STATISTICAL INDEPENDENCE

Two (or more) random variables x and y are statistically independent if and only if

$$p(x, y) = p_x(x)p_y(y) \tag{A.8}$$

It turns out that in this case $E[xy] = E[x]E[y]$. These are generalized to more than two variables.

A.4 MARGINALIZATION

Let $x_i,\ i = 1, 2, \ldots, l$, be a set of random variables with a joint probability density function $p(x_1, x_2, \ldots, x_l)$. It can be shown that by integrating the joint pdf with respect to some of the variables, over all possible values, the result is the joint pdf of the remaining variables. For example,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p(x_1, \ldots, x_l)dx_{k+1}dx_{k+2}\cdots dx_l = p(x_1, x_2, \ldots, x_k)$$

This calculation is known as *marginalization*. For discrete random variables, marginalization involves probabilities and summations, i.e.,

$$\sum_{x_{k+1}}\sum_{x_{k+2}}\cdots\sum_{x_l} P(x_1, \ldots, x_l) = P(x_1, \ldots, x_k)$$

where summations are over all possible values of the respective variables.

A.5 CHARACTERISTIC FUNCTIONS

Let $p(x)$ be the probability density function of a random variable x. The associated *characteristic function* is by definition the integral

$$\Phi(\Omega) = \int_{-\infty}^{+\infty} p(x)\exp(j\Omega x)\,dx \equiv E[\exp(j\Omega x)] \tag{A.9}$$

If $j\Omega$ is changed into s, the resulting integral becomes

$$\Phi(s) = \int_{-\infty}^{+\infty} p(x)\exp(sx)\,dx \equiv E[\exp(sx)] \tag{A.10}$$

and it is known as the *moment generating function*.

The function

$$\Psi(\Omega) = \ln \Phi(\Omega) \tag{A.11}$$

is known as the *second characteristic function* of x.

The joint characteristic function of l random variables is defined by

$$\Phi(\Omega_1, \Omega_2, \ldots, \Omega_l) = \int_{-\infty}^{+\infty} p(x_1, x_2, \ldots, x_l) \exp\left(j \sum_{i=1}^{l} \Omega_i x_i\right) d\boldsymbol{x} \tag{A.12}$$

The logarithm of the above is the second joint characteristic function of the l random variables.

A.6 MOMENTS AND CUMULANTS

Taking the derivative of $\Phi(s)$ in Eq. (A.10) we obtain

$$\frac{d^n \Phi(s)}{ds^n} \equiv \Phi^{(n)}(s) = E[x^n \exp(sx)] \tag{A.13}$$

and hence for $s = 0$

$$\Phi^{(n)}(0) = E[x^n] \equiv m_n \tag{A.14}$$

where m_n is known as the nth-order moment of x. If the moments of all orders are finite, the Taylor series expansion of $\Phi(s)$ near the origin exists and is given by

$$\Phi(s) = \sum_{n=0}^{+\infty} \frac{m_n}{n!} s^n \tag{A.15}$$

Similarly, the Taylor expansion of the second generating function results in

$$\Psi(s) = \sum_{n=1}^{+\infty} \frac{\kappa_n}{n!} s^n \tag{A.16}$$

where

$$\kappa_n \equiv \frac{d^n \Psi(0)}{ds^n} \tag{A.17}$$

and are known as the *cumulants* of the random variable x. It is not difficult to show that $\kappa_0 = 0$. For a zero mean random variable, it turns out that

$$\kappa_1(x) = E[x] = 0 \tag{A.18}$$

$$\kappa_2(x) = E[x^2] = \sigma^2 \tag{A.19}$$

$$\kappa_3(x) = E[x^3] \tag{A.20}$$

$$\kappa_4(x) = E[x^4] - 3\sigma^4 \tag{A.21}$$

That is, the first three cumulants are equal to the corresponding moments. The fourth-order cumulant is also known as *kurtosis*. *For a Gaussian process all cumulants of order higher than two are zero*. The kurtosis is commonly used as a measure of the non-Gaussianity of a random variable. For random variables described by (unimodal) pdfs with spiky shape and heavy tails, known as leptokurtic or super-Gaussian, κ_4 is positive, whereas for random variables associated with pdfs with a flatter shape, known as platykurtic or sub-Gaussian, κ_4 is negative. Gaussian variables have zero kurtosis. The opposite is not always true, in the sense that there exist non-Gaussian random variables with zero kurtosis; however, this can be considered rare.

Similar arguments hold for the expansion of the joint characteristic functions for multivariate pdfs. For zero mean random variables, x_i, $i = 1, 2, \ldots, I$, the cumulants of order up to four are given by

$$\kappa_1(x_i) = E[x_i] = 0 \tag{A.22}$$

$$\kappa_2(x_i, x_j) = E[x_i x_j] \tag{A.23}$$

$$\kappa_3(x_i, x_j, x_k) = E[x_i x_j x_k] \tag{A.24}$$

$$\kappa_4(x_i, x_j, x_k, x_r) = E[x_i x_j x_k x_r] - E[x_i x_j]E[x_k x_r] \tag{A.25}$$

$$- E[x_i x_k]E[x_j x_r] - E[x_i x_r]E[x_j x_k] \tag{A.26}$$

Thus, once more, the cumulants of the first three orders are equal to the corresponding moments. If all variables coincide, we talk about *auto-cumulants*, and otherwise about *cross-cumulants*, that is,

$$\kappa_4(x_i, x_i, x_i, x_i) = \kappa_4(x_i)$$

that is, the auto-cumulant of x_i is identical to its kurtosis. It is not difficult to see that if the zero mean random variables are mutually independent, their cross-cumulants are zero. *This is also true for the cross-cumulants of all orders*.

A.7 EDGEWORTH EXPANSION OF A PDF

Taking into account the expansion in Eq. (A.16), the definition given in Eq. (A.11), and taking the inverse Fourier of $\Phi(\Omega)$ in Eq. (A.9) we can obtain the following expansion of $p(x)$ for a zero mean unit variance random variable x:

$$p(x) = g(x)\left(1 + \frac{1}{3!}\kappa_3(x)H_3(x) + \frac{1}{4!}\kappa_4(x)H_4(x) + \frac{10}{6!}\kappa_3^2(x)H_6(x) + \frac{1}{5!}\kappa_5(x)H_5(x) + \frac{35}{7!}\kappa_3(x)\kappa_4(x)H_7(x) + \ldots\right) \tag{A.27}$$

where $g(x)$ is the unit variance and zero mean normal pdf, and $H_\kappa(x)$ is the Hermite polynomial of degree k. The rather strange ordering of terms is the outcome of a

specific reordering in the resulting expansion, so that the successive coefficients in the series decrease uniformly. This is very important when truncation of the series is required. The Hermite polynomials are defined as

$$H_k(x) = (-1)^k \exp(x^2/2)\frac{d^k}{dx^k}\exp(-x^2/2) \tag{A.28}$$

and they form a complete orthogonal basis set in the real axis, that is,

$$\int_{-\infty}^{+\infty} \exp(-x^2/2)H_n(x)H_m(x)\,dx = \begin{cases} n!\sqrt{2\pi} & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases} \tag{A.29}$$

The expansion of $p(x)$ in Eq. (A.27) is known as the *Edgeworth expansion*, and it is actually an expansion of a pdf around the normal pdf [Papo 91].

A.8 KULLBACK–LEIBLER DISTANCE

The Kullback–Leibler (KL) distance is a measure of the distance between two probability density functions $p(\boldsymbol{x})$ and $\hat{p}(\boldsymbol{x})$ and is defined as

$$L = -\int p(\boldsymbol{x}) \ln \frac{\hat{p}(\boldsymbol{x})}{p(\boldsymbol{x})} d\boldsymbol{x} \tag{A.30}$$

Sometimes it is referred to as *cross or relative entropy*. The KL distance can be shown to be always nonnegative but it is not a true distance measure, from a mathematical point of view, since it is not symmetric. Sometimes it is referred as the KL divergence.

The KL distance is closely related to the *mutual information* measure, I, between l scalar random variables, $x_i, i = 1, 2, \ldots, l$. Indeed, let us compute the KL distance between the joint pdf $p(\boldsymbol{x})$ and the pdf resulting from the product of the corresponding marginal probability densities, that is,

$$\begin{aligned} L &= -\int p(\boldsymbol{x}) \ln \frac{\prod_{i=1}^{l} p_i(x_i)}{p(\boldsymbol{x})} d\boldsymbol{x} \\ &= \int p(\boldsymbol{x}) \ln p(\boldsymbol{x}) d\boldsymbol{x} - \sum_{i=1}^{l} \int p(\boldsymbol{x}) \ln p_i(x_i) d\boldsymbol{x} \\ &= -H(\boldsymbol{x}) - \sum_{i=1}^{l} \int p(\boldsymbol{x}) \ln p_i(x_i) d\boldsymbol{x} \end{aligned} \tag{A.31}$$

Carrying out the integrations on the right-hand side it is straightforward to see the KL distance is equal to the mutual information, I, defined as

$$I(x_1, x_2, \ldots, x_l) = -H(\boldsymbol{x}) + \sum_{i=1}^{l} H(x_i) \tag{A.32}$$

where $H(x_i)$ is the associated entropy of x_i, defined as ([Papo 91])

$$H(x_i) = -\int p_i(x_i) \ln p_i(x_i)\, dx_i \tag{A.33}$$

It is now easy to see that if the variables $x_i, i = 1, 2, \ldots, l$, are statistically independent their mutual information I is zero. Indeed, in this case $\Pi_{i=1}^{l} p_i(x_i) = p(\boldsymbol{x})$, hence $L = I(x_1, x_2, \ldots, x_l) = 0$.

A.9 MULTIVARIATE GAUSSIAN OR NORMAL PROBABILITY DENSITY FUNCTION

This is defined as a generalization of the univariate normal pdf

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big), \tag{A.34}$$

where $\boldsymbol{\mu}$ is the mean vector, that is, $E\left[[x_1, x_2, \ldots, x_l]^T\right] = [\mu_1, \mu_2, \ldots, \mu_l]^T$ and Σ the covariance matrix

$$\Sigma = E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] \tag{A.35}$$

and we say that $\boldsymbol{x}$ is normally distributed as $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. For the one dimensional, $l = 1$, case the covariance matrix becomes the variance σ^2 and the Gaussian density function takes the form

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

Figure A.1 shows the plots of two Gaussians for the same mean and different variances. For the general l-dimensional case the covariance matrix has the form

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1l} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{l1} & \sigma_{l2} & \cdots & \sigma_l^2 \end{bmatrix} \tag{A.36}$$

where $\sigma_i^2 = E[(x_i - \mu_i)^2], \sigma_{ij} = \sigma_{ji} = E[(x_i - \mu_i)(x_j - \mu_j)]$. Thus, the main matrix diagonal consists of the respective variances of the elements of the random vector and the off-diagonal elements are the respective covariances between the elements of the random vector. Note that if the random variables x_i are statistically independent, then the mean of the product equals the product of the means, that is, $E[(x_i - \mu_i)(x_j - \mu_j)] = E[(x_i - \mu_i)]E[(x_j - \mu_j)] = 0$, and the covariance matrix is diagonal. However, a diagonal covariance matrix does not, in general, mean that the variables are statistically independent. In the case, though, of multivariate Gaussian

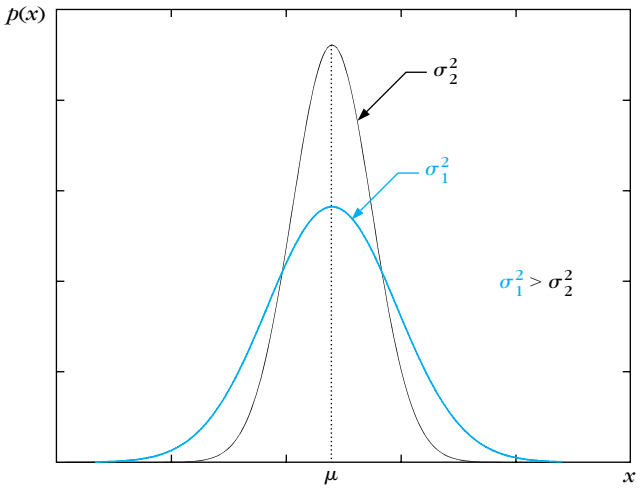

FIGURE A.1

Two Gaussians with the same mean μ and different variances.

densities the opposite is also valid. Indeed, if the covariance matrix is diagonal then it is straightforward to see that

$$p(\boldsymbol{x}) = \prod_{i=1}^{l} p_i(x_i) \tag{A.37}$$

where

$$p_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$

which is the univariate Gaussian describing the ith variable (why?). Thus, the joint probability density is the product of the individual (marginal) ones, which is the definition of statistical independence.

A.10 TRANSFORMATION OF RANDOM VARIABLES

Let $X = \{x_1, x_2, \ldots, x_l\}$ be a set of random variables, which are jointly distributed according to the joint pdf $p_X(x_1, x_2, \ldots, x_l)$. We form a new set of random variables by the following transformations

$$y_1 = g_1(x_1),\ y_2 = g_2(x_2), \ldots,\ y_l = g_l(x_l) \tag{A.38}$$

It can be shown [Papo 91] that the joint pdf describing the set Y is given by

$$p_Y(y_1, y_2, \ldots, y_l) = \frac{p_X(x_1, x_2, \ldots, x_l)}{|J(x_1, x_2, \ldots, x_l)|} \tag{A.39}$$

where $|\cdot|$ denotes the determinant of a matrix and $J(x_1, x_2, \ldots, x_l)$ is the Jacobian matrix of the transformation, defined as

$$J(x_1, x_2, \ldots, x_l) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_l} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_l} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_l}{\partial x_1} & \frac{\partial y_l}{\partial x_2} & \cdots & \frac{\partial y_l}{\partial x_l} \end{bmatrix}$$

provided that the system of equations given in (A.38) has a solution with respect to $x_1, x_2, \ldots, x_l$ and it is ***unique***.

A widely used transformation is the linear one and the set in (A.38) is compactly written as

$$\boldsymbol{y} = A\boldsymbol{x} \tag{A.40}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_l]^T$ and $\boldsymbol{x} = [x_1, x_2, \ldots, x_l]^T$ and the matrix A is invertible. Then the system of equations has a unique solution, $\boldsymbol{x} = A^{-1}\boldsymbol{y}$, and the Jacobian is easily shown to be

$$J(x_1, x_2, \ldots, x_l) = A \tag{A.41}$$

The above can be used, for example, in the MATLAB framework to generate jointly Gaussian variables with mean value $\boldsymbol{\mu}$ and covariance matrix Σ, utilizing the normalized Gaussian generator $\mathcal{N}(\mathbf{0}, I)$. The latter is described by

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{l/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right) \tag{A.42}$$

Let Σ be the covariance matrix of the multivariate Gaussian, which describes the variables to be generated. We know that Σ is a symmetric matrix, $\Sigma = \Sigma^T$, and therefore it can be diagonalized (Appendix B) as

$$\Sigma = P\Lambda P^T$$

where Λ is a diagonal matrix having the eigenvalues of Σ as its elements and P is a unitary matrix ($P^{-1} = P^T$) having as columns the corresponding eigenvectors of Σ. Define, now, the linear transformation

$$\boldsymbol{y} = P\Lambda^{1/2}\boldsymbol{x} \tag{A.43}$$

where $\Lambda^{1/2}$ is the square root of Λ. Assuming the transformation to be invertible (i.e., Σ is invertible) then, recalling (A.39), (A.41), (A.42) and (A.43) we obtain

$$p(\boldsymbol{y}) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T P\Lambda^{-1/2}\Lambda^{-1/2}P^T\boldsymbol{y}\right) \tag{A.44}$$

$$= \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y}\right) \tag{A.45}$$

where we have used the fact that the determinant of the Jacobian matrix is $|P\Lambda^{1/2}| = |P\Lambda^{1/2}\Lambda^{1/2}P^T|^{1/2} = |\Sigma|^{1/2}$. Thus, in order to generate a set of random variables described by the multivariate Gaussian $\mathcal{N}(\mathbf{0}, \Sigma)$, it suffices to generate a vector, $\boldsymbol{x}$, using the normalized Gaussian, $\mathcal{N}(\mathbf{0}, I)$, and then transform it, that is, $\boldsymbol{y} = P\Lambda^{1/2}\boldsymbol{x}$. Finally, a further shift by $\boldsymbol{\mu}$, that is, $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\mu}$, suffices to produce multivariate Gaussian variables described by the Gaussian $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

A.11 THE CRAMER–RAO LOWER BOUND

Let $p(\boldsymbol{x}; \boldsymbol{\theta})$ be the pdf of a random vector, parameterized in terms of an r-dimensional vector parameter $\boldsymbol{\theta}$. If X is the set of N observations $\boldsymbol{x}_i$, $i = 1, 2, \ldots, N$, the log likelihood function is the logarithm of the joint pdf of the observations $\ln p(X; \boldsymbol{\theta}) = L(\boldsymbol{\theta})$. The Fisher matrix is defined so that its (i, j) element equals

$$J_{ij} = -E\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right], \quad i, j = 1, 2, \ldots, r \tag{A.46}$$

It can be shown that the ith element of *any* unbiased estimate $\hat{\boldsymbol{\theta}}$, of the parameter $\boldsymbol{\theta}$, based on the observations set X satisfies

$$E[(\hat{\theta}_i - \theta_i)^2] \geq J_{ii}^{-1} \tag{A.47}$$

In other words, its variance is *lower bounded by the* (i, i) *element of the inverse Fisher matrix*. This is known as the Cramer–Rao bound. If the relation is valid with equality, the corresponding estimator is called *efficient*.

A.12 CENTRAL LIMIT THEOREM

Let $x_1, x_2, \ldots, x_N$ be N independent random variables, with mean and variances μ_i, σ_i^2, respectively. We form the new random variable

$$z = \sum_{i=1}^{N} x_i \tag{A.48}$$

Its mean and variance are given by $\mu = \sum_{i=1}^{N} \mu_i$ and $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2$. The central limit theorem states that as $N \to \infty$, and under certain general conditions, the pdf of the variable

$$q = \frac{z - \mu}{\sigma} \tag{A.49}$$

approaches $\mathcal{N}(0, 1)$, irrespective of the pdfs of the summands [Papo 91]. Thus in practice, for large enough N we can consider z as approximately Gaussian with mean μ and variance σ^2.

A.13 CHI-SQUARE DISTRIBUTION

Let $x_i, i = 1, 2, \ldots, N$, be samples of a Gaussian $\mathcal{N}(0, 1)$ random variable x. The sum of squares variable

$$\chi^2 = x_1^2 + x_2^2 + \cdots + x_N^2 \equiv y \tag{A.50}$$

is a chi-square distributed variable with N degrees of freedom. Its probability density function is given by [Papo 91]

$$p_y(y) = \frac{1}{2^{N/2}\Gamma(N/2)} y^{N/2-1} \exp(-y/2)u(y) \tag{A.51}$$

where

$$\Gamma(b+1) = \int_0^\infty y^b \exp(-y)\, dy \quad b > -1 \tag{A.52}$$

where $u(y)$ is the step function (1 for $y > 0$ and 0 for $y < 0$). Recalling the respective definitions, it is easy to show that $E[y] = N, \sigma_y^2 = 2N$.

The chi-square distribution possesses the *additive property*. Let χ_1^2 and χ_2^2 be independent random variables of chi-square distribution with N_1, N_2 degrees of freedom, respectively. Then the random variable

$$\chi^2 = \chi_1^2 + \chi_2^2 \tag{A.53}$$

is a chi-square variable with $N_1 + N_2$ degrees of freedom. Based on these properties, we can show that the variance estimate of Eq. (5.13) is described by a chi-square distribution with $N-1$ degrees of freedom, provided x is Gaussian and the samples x_i are independent. The proof is simple and interesting [Fras 58]. Define the following transformation:

$$\begin{aligned}
y_1 &= \sqrt{N}\bar{x} = \frac{x_1 + \cdots + x_n}{\sqrt{N}} \\
y_2 &= \frac{1}{\sqrt{2}}(x_2 - x_1) \\
&\vdots \\
y_n &= \frac{1}{\sqrt{n(n-1)}}[(n-1)x_n - (x_1 + \cdots + x_{n-1})], \quad n = 2, 3, \ldots, N
\end{aligned}$$

It is easy to show that this transformation is an orthogonal one (Problem 5.5). Thus, the random variables y_i are also Gaussian, statistically independent, and with the same variance σ^2 as x (Problem 5.6). This transformation easily results in

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} x_i^2 \tag{A.54}$$

and of course

$$y_1^2 = N\bar{x}^2 \tag{A.55}$$

Subtracting the two, we obtain

$$\sum_{i=2}^{N} y_i^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 \equiv (N-1)\hat{\sigma}^2 \tag{A.56}$$

Furthermore, $E[y_i] = 0, i = 2, \ldots, N$. Thus the variable

$$z = \frac{N-1}{\sigma^2}\hat{\sigma}^2 = \sum_{i=2}^{N} \frac{y_i^2}{\sigma^2} \tag{A.57}$$

is a chi-square with $N-1$ degrees of freedom.

A.14 *t*-DISTRIBUTION

Let x and z be two independent random variables with x being $\mathcal{N}(0,1)$ and z a chi-square with N degrees of freedom. Then it can be shown [Papo 91] that the variable

$$q = \frac{x}{\sqrt{z/N}} \tag{A.58}$$

is a so-called t-distributed variable with probability density function given by

$$p_q(q) = \frac{\gamma_1}{\sqrt{(1 + q^2/N)^{N+1}}}, \quad \gamma_1 = \frac{\Gamma((N+1)/2)}{\sqrt{\pi N}\Gamma(N/2)}$$

where $\Gamma(\cdot)$ was defined in Eq. (A.44). Thus, from the test statistic in Eqs. (5.14) and (A.57) we have

$$q = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{N}}}{\sqrt{z/N - 1}} \tag{A.59}$$

Since z is a chi-square with $N-1$ degrees of freedom, q is a t-distributed variable with $N-1$ degrees of freedom. In a similar way we can show that the test statistic in Eq. (5.18) is t-distributed with $2N-2$ degrees of freedom.

A.15 BETA DISTRIBUTION

A random variable follows the Beta distribution with parameters a and $b\,(a, b > 0)$, if its probability density function is defined as

$$p(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & 0 < x < 1 \\ 0, & \text{Otherwise} \end{cases} \tag{A.60}$$

where

$$B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,du \tag{A.61}$$

Its mean and variance are equal to $a/(a+b)$ and $ab/((a+b)^2(a+b+1))$, respectively.

A.16 POISSON DISTRIBUTION

A Poisson distributed random variable X, with parameter a, takes the values $k = 0, 1, 2, \ldots$, with probabilities

$$P(X = k) = e^{-a}\frac{a^k}{k!} \tag{A.62}$$

A *Poisson process* scatters vectors in a Euclidean space in such a way that the random variable X, denoting the number of vectors in a region of volume V, has a Poisson distribution with parameter λV, that is,

$$P(X = k) = e^{-\lambda V}\frac{(\lambda V)^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{A.63}$$

The parameter λ is called the *intensity* of the process and equals the expected number of vectors per unit volume.

A.17 GAMMA FUNCTION

The Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

If α is an integer, integrating by parts we get

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)!$$

REFERENCES

[Digg 83] Diggle P.J. *Statistical Analysis of Spatial Point Processes*, Academic Press, 1983.

[Fras 58] Fraser D.A.S. *Statistics: An Introduction*, John Wiley, 1958.

[Papo 91] Papoulis A. *Probability Random Variables and Stochastic Processes*, 3rd ed. McGraw Hill, 1991.

[Spie 75] Spiegel M.R. *Schaum's Outline of Theory and Problems of Probability and Statistics*, McGraw Hill, 1975.