

EN 257: Applied Stochastic Processes

Problem Set 3

Douglas Lanman
dlanman@brown.edu
2 March 2007

Problem 4.28

Compute the Chernoff bound on $P[X \geq a]$ where X is a random variable that satisfies the exponential law $f_X(x) = \lambda e^{-\lambda x} u(x)$.

Recall, from Equation 4.6-4 on page 214 in [3], that the Chernoff bound for a continuous random variable X is given by

$$P[X \geq a] \leq \operatorname{argmin}_{t>0} \{e^{-at} \theta_X(t)\}, \quad (1)$$

where $\theta_X(t)$ is the moment-generating function

$$\theta_X(t) \triangleq E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \quad (2)$$

as defined by Equation 4.5-1 on page 211. Let's begin by finding the moment-generating function for an exponential random variable X . Substituting for $f_X(x)$ in Equation 2 gives the following expression.

$$\theta_X(t) = \lambda \int_0^{\infty} e^{tx} e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}, \text{ for } 0 < t < \lambda \quad (3)$$

Note that the integral in Equation 3 has a finite value only if $t < \lambda$ and, since $t > 0$, we conclude $0 < t < \lambda$. Substituting this result into Equation 1 gives the following form for the Chernoff bound.

$$P[X \geq a] \leq \operatorname{argmin}_{0 < t < \lambda} \left\{ \frac{\lambda e^{-at}}{\lambda - t} \right\}$$

In general, the minimum of the right-hand argument will occur where the first derivative, with respect to t , equals zero.

$$\frac{d}{dt} \left\{ \frac{\lambda e^{-at}}{\lambda - t} \right\} = \frac{[1 - a(\lambda - t)] e^{-at}}{(\lambda - t)^2} = 0 \quad \Rightarrow \quad t = \lambda - \frac{1}{a}$$

Since t must also satisfy the additional constraint $0 < t < \lambda$, we conclude that the Chernoff bound for an exponential random variable X has the following form (for all $\lambda > 0$).

$$P[X \geq a] \leq \begin{cases} a\lambda e^{1-a\lambda}, & \text{for } a > \frac{1}{\lambda} \\ 1, & \text{for } 0 < a \leq \frac{1}{\lambda} \end{cases}$$

Problem 5.6

Let \mathbf{X}_i , for $i = 1, \dots, n$, be n mutually orthogonal random vectors. Show that

$$E \left[\left\| \sum_{i=1}^n \mathbf{X}_i \right\|^2 \right] = \sum_{i=1}^n \mathbf{E} \left[\|\mathbf{X}_i\|^2 \right], \quad (4)$$

by using the definition $\|\mathbf{X}\|^2 \triangleq \mathbf{X}^T \mathbf{X}$.

Let's begin by expanding the argument of the right-hand side of Equation 4 using the identity $\|\mathbf{X}\|^2 \triangleq \mathbf{X}^T \mathbf{X}$.

$$E \left[\left\| \sum_{i=1}^n \mathbf{X}_i \right\|^2 \right] = E \left[\left(\sum_{i=1}^n \mathbf{X}_i \right)^T \left(\sum_{j=1}^n \mathbf{X}_j \right) \right]$$

Note that the argument of the expected value operator E can be written as follows.

$$\left(\sum_{i=1}^n \mathbf{X}_i \right)^T \left(\sum_{j=1}^n \mathbf{X}_j \right) = \left(\sum_{i=1}^n \mathbf{X}_i^T \right) \left(\sum_{j=1}^n \mathbf{X}_j \right) = \sum_{i=1}^n \sum_{j=1}^n \mathbf{X}_i^T \mathbf{X}_j$$

Substituting this result into the previous expression yields the following.

$$E \left[\left\| \sum_{i=1}^n \mathbf{X}_i \right\|^2 \right] = E \left[\sum_{i=1}^n \sum_{j=1}^n \mathbf{X}_i^T \mathbf{X}_j \right]$$

Recall that the expected value operator E is linear. As a result, we have

$$E \left[\left\| \sum_{i=1}^n \mathbf{X}_i \right\|^2 \right] = \sum_{i=1}^n \sum_{j=1}^n E \left[\mathbf{X}_i^T \mathbf{X}_j \right].$$

To proceed, we recall that \mathbf{X}_i , for $i = 1, \dots, n$, are n mutually orthogonal random vectors. We can express this orthogonality condition as

$$\mathbf{X}_i^T \mathbf{X}_j = (\mathbf{X}_i^T \mathbf{X}_i) \delta_{ij} \triangleq \begin{cases} \mathbf{X}_i^T \mathbf{X}_i, & \text{for } i = j \\ 0, & \text{otherwise} \end{cases}$$

where δ_{ij} is the familiar Kronecker delta function. Substituting this result into the previous expression, and recalling that $\|\mathbf{X}\|^2 \triangleq \mathbf{X}^T \mathbf{X}$, yields the desired relationship.

$$\therefore E \left[\left\| \sum_{i=1}^n \mathbf{X}_i \right\|^2 \right] = \sum_{i=1}^n E \left[\mathbf{X}_i^T \mathbf{X}_i \right] = \sum_{i=1}^n \mathbf{E} \left[\|\mathbf{X}_i\|^2 \right]$$

(QED)

Problem 5.16

Two jointly Normal random variables X_1 and X_2 have a joint pdf given by

$$f_{X_1 X_2}(x_1, x_2) = \frac{2}{\pi\sqrt{7}} \exp \left[-\frac{8}{7} \left(x_1^2 + \frac{3}{2}x_1x_2 + x_2^2 \right) \right]. \quad (5)$$

Find a nontrivial transformation \mathbf{A} in

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

such that Y_1 and Y_2 are independent. Compute the joint pdf of Y_1 and Y_2 .

This problem can be solved by applying the method outlined in Example 5.6-2 on pages 274-276 of [3]. First, we recall that two zero-mean jointly Normal random variables X_1 and X_2 have a joint pdf given by

$$f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi[\det(\mathbf{K})]^{1/2}} \exp \left(-\frac{1}{2}(\mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}) \right), \quad (6)$$

for $\mathbf{x} \triangleq (x_1, x_2)^T$ and \mathbf{K} an arbitrary 2×2 positive definite covariance matrix. As demonstrated in the example, we notice that the quadratic in the exponent of Equation 5 can be written as follows.

$$ax_1^2 + (b+c)x_1x_2 + dx_2^2 = \mathbf{x}^T \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{x}$$

Comparing Equations 5 and 6, we find the following constraints on the covariance matrix \mathbf{K} .

$$(\mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}) = \mathbf{x}^T \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{x} = \frac{16}{7}x_1^2 + \frac{24}{7}x_1x_2 + \frac{16}{7}x_2^2 \quad \text{and} \quad \det(\mathbf{K}) = \frac{7}{16}$$

Recall that, for real X_1 and X_2 , both the covariance matrix and its inverse are real symmetric matrices. As a result, we must have $b = c$ in the previous expression. We conclude, by inspection, that the covariance matrix and its inverse are given by the following expressions.

$$\mathbf{K} = \frac{1}{4} \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{K}^{-1} = \frac{4}{7} \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$$

At this point we follow a modified version of the approach outlined in Example 5.6-2. We begin by finding the eigenvalues $\{\lambda_1, \lambda_2\}$ of the covariance matrix \mathbf{K} .

$$\det(\mathbf{K} - \lambda \mathbf{I}) = 0 \quad \Rightarrow \quad \lambda^2 - 2\lambda + \frac{7}{16} = 0 \quad \Rightarrow \quad \left\{ \lambda_1 = \frac{7}{4}, \lambda_2 = \frac{1}{4} \right\}$$

Next, we obtain the eigenvectors $\{\phi_1, \phi_2\}$ of \mathbf{K} as follows.

$$\begin{aligned} (\mathbf{K} - \lambda_1 \mathbf{I})\phi_1 &= \mathbf{0} \quad \text{with} \quad \|\phi_1\| = 1 \quad \Rightarrow \quad \phi_1 = (-1/\sqrt{2}, 1/\sqrt{2})^T \\ (\mathbf{K} - \lambda_2 \mathbf{I})\phi_2 &= \mathbf{0} \quad \text{with} \quad \|\phi_2\| = 1 \quad \Rightarrow \quad \phi_2 = (1/\sqrt{2}, 1/\sqrt{2})^T \end{aligned}$$

Now we define a similarity transformation $\mathbf{A} = \mathbf{U}^T$, whose rows are the scaled eigenvectors of \mathbf{K} , which diagonalizes the covariance matrix as follows.

$$\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A} \mathbf{K} \mathbf{A}^{-1} = \frac{1}{4} \begin{bmatrix} 7 & 0 \\ 0 & 1 \end{bmatrix}$$

Note that in the previous expression, we are free to scale \mathbf{A} by an arbitrary real constant since, for this problem, we are not required to “whiten” the covariance matrix (i.e., simply diagonalizing it is sufficient). In conclusion, Y_1 and Y_2 could have the following dependence on X_1 and X_2 .

$$\begin{aligned} Y_1 &= X_2 - X_1 \\ Y_2 &= X_1 + X_2 \end{aligned}$$

To complete our analysis, we recall that the joint pdf of a function of two random variables is given by

$$f_{Y_1 Y_2}(y_1, y_2) = \sum_{i=1}^n f_{X_1 X_2}(\mathbf{x}_i) / |J_i|, \quad (7)$$

where $\mathbf{x}_i \triangleq (x_1^{(i)}, x_2^{(i)})^T$, for $i = 1, \dots, n$, are the n solutions to $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$, and J_i is the associated Jacobian [3]. For the previous value of \mathbf{A} , there is only the following solution $\mathbf{x} = (x_1, x_2)^T$ to $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$.

$$\begin{aligned} x_1 &= \frac{y_2 - y_1}{2} \\ x_2 &= \frac{y_1 + y_2}{2} \end{aligned}$$

Furthermore, for $y_1 = g(x_1, x_2) = x_2 - x_1$ and $y_2 = h(x_1, x_2) = x_1 + x_2$, the associated Jacobian has the following form.

$$|J| = \left| \det \begin{pmatrix} \partial g / \partial x_1 & \partial g / \partial x_2 \\ \partial h / \partial x_1 & \partial h / \partial x_2 \end{pmatrix} \right| = 2$$

Substituting these results into Equation 7 gives the following form for the joint pdf of Y_1 and Y_2 .

$$f_{Y_1 Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1 X_2} \left(\frac{y_2 - y_1}{2}, \frac{y_1 + y_2}{2} \right)$$

Substituting Equation 5 for $f_{X_1 X_2}(x_1, x_2)$ gives the following solution for the joint pdf.

$$f_{Y_1 Y_2}(y_1, y_2) = \left[\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left(-\frac{y_1^2}{2\sigma_1^2} \right) \right] \cdot \left[\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left(-\frac{y_2^2}{2\sigma_2^2} \right) \right], \text{ for } \sigma_1 = \sqrt{\frac{7}{2}} \text{ and } \sigma_2 = \sqrt{\frac{1}{2}}$$

In conclusion, we find that the similarity transformation \mathbf{A} has diagonalized the covariance matrix \mathbf{K} such that $f_{Y_1 Y_2}(y_1, y_2)$ is the product of two univariate Normal distributions.

Problem 5.17

Show that if $\mathbf{X} = (X_1, \dots, X_n)^T$ has mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ and covariance

$$\mathbf{K} = \{K_{ij}\}_{n \times n},$$

then the scalar random variable Y given by

$$Y \triangleq p_1 X_1 + \dots + p_n X_n \tag{8}$$

has mean

$$E[Y] = \sum_{i=1}^n p_i \mu_i \tag{9}$$

and variance

$$\sigma_Y^2 = \sum_{i=1}^n \sum_{j=1}^n p_i p_j K_{ij}. \tag{10}$$

This problem examines the mean and variance of a linear combination of random variables X_i , for $i = 1, \dots, n$. Let's begin by proving that the mean of the linear combination is a weighted combination of the mean of each input variable X_i of the form given in Equation 9. First, we note that the expectation of Y can be written in the following form by substituting Equation 8.

$$E[Y] = E \left[\sum_{i=1}^n p_i X_i \right]$$

Since the expected value operator E is linear, we can simplify this expression as

$$E[Y] = \sum_{i=1}^n p_i E[X_i],$$

where $\{p_i\}$, for $i = 1, \dots, n$, are fixed constants. Finally, we recall that the expectation of X_i is μ_i . As a result, we conclude that Equation 9 is valid.

$$\therefore E[Y] = \sum_{i=1}^n p_i \mu_i$$

At this point, we turn our attention to proving Equation 10. Recall that σ_Y^2 is the variance of the random variable Y , which has the following well-known definition [3].

$$\sigma_Y^2 = \text{Var}[Y] \triangleq E[(Y - E[Y])^2] = E[Y^2] - E[Y]^2$$

Substituting Equations 8 and 9 into this expression gives the following result.

$$\sigma_Y^2 = E \left[\left(\sum_{i=1}^n p_i X_i \right)^2 \right] - \left(\sum_{i=1}^n p_i \mu_i \right)^2 \tag{11}$$

As an aside, we note that the square of a summation can be written in the following form.

$$\left(\sum_{i=1}^n p_i X_i\right)^2 = \left(\sum_{i=1}^n p_i X_i\right) \cdot \left(\sum_{j=1}^n p_j X_j\right) = \sum_{i=1}^n \sum_{j=1}^n p_i p_j X_i X_j \quad (12)$$

Substituting Equation 12 into Equation 11 gives

$$\sigma_Y^2 = E \left[\sum_{i=1}^n \sum_{j=1}^n p_i p_j X_i X_j \right] - \sum_{i=1}^n \sum_{j=1}^n p_i p_j \mu_i \mu_j.$$

Once again, we note that the expectation is linear and $\{p_i\}$, for $i = 1, \dots, n$, are fixed constants. As a result, the previous expression has the following simple form.

$$\begin{aligned} \sigma_Y^2 &= \sum_{i=1}^n \sum_{j=1}^n p_i p_j E[X_i X_j] - \sum_{i=1}^n \sum_{j=1}^n p_i p_j \mu_i \mu_j \\ &= \sum_{i=1}^n \sum_{j=1}^n p_i p_j (E[X_i X_j] - \mu_i \mu_j) \end{aligned} \quad (13)$$

Now we expand the expectation of $X_i X_j$ as

$$\begin{aligned} E[X_i X_j] &= E[\{(X_i - \mu_i) + \mu_i\} \{(X_j - \mu_j) + \mu_j\}] \\ &= E[(X_i - \mu_i)(X_j - \mu_j)] + \mu_i E[X_j - \mu_j] + \mu_j E[X_i - \mu_i] + \mu_i \mu_j \\ &= E[(X_i - \mu_i)(X_j - \mu_j)] + \mu_i \mu_j, \end{aligned} \quad (14)$$

since $E[X_i - \mu_i] = E[X_i] - \mu_i = 0$. Substituting Equation 14 into Equation 13 yields the following.

$$\sigma_Y^2 = \sum_{i=1}^n \sum_{j=1}^n p_i p_j E[(X_i - \mu_i)(X_j - \mu_j)]$$

At this point, we recall that the elements of the covariance matrix are given by

$$K_{ij} \triangleq E[(X_i - \mu_i)(X_j - \mu_j)].$$

In conclusion, we find that Equation 10 is also valid.

$$\therefore \sigma_Y^2 = \sum_{i=1}^n \sum_{j=1}^n p_i p_j K_{ij}$$

(QED)

Problem 5.22

Let X_1 and X_2 be two zero-mean jointly Normal random variables with covariance matrix

$$\mathbf{K}_{\mathbf{X}} = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix}. \quad (15)$$

Find a whitening transformation for $\mathbf{X} = (X_1, X_2)^T$. Write a MATLAB program to show a scatter diagram of x_2 versus x_1 (i.e., the realizations of X_2 and X_1). Make a similar plot for the whitened variables as well. Choose between a hundred and a thousand realizations.

Recall, from pages 257 and 259 in [3], that the whitening transformation \mathbf{D} is a linear transformation of a random vector \mathbf{X} that reduces the covariance matrix to an identity matrix. Written in matrix-vector form, the whitening transformation is given by

$$\mathbf{Y} = \mathbf{D}\mathbf{X}, \text{ where } \mathbf{K}_{\mathbf{Y}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \triangleq \mathbf{I}.$$

In order to find \mathbf{D} , we follow the procedure described in Example 5.6-1 on pages 273 and 274. First, we begin by finding the eigenvalues $\{\lambda_1, \lambda_2\}$ of the covariance matrix $\mathbf{K}_{\mathbf{X}}$.

$$\det(\mathbf{K}_{\mathbf{X}} - \lambda\mathbf{I}) = 0 \quad \Rightarrow \quad \lambda^2 - 4\lambda + \frac{7}{4} = 0 \quad \Rightarrow \quad \left\{ \lambda_1 = \frac{7}{2}, \lambda_2 = \frac{1}{2} \right\}$$

Next, we obtain the eigenvectors $\{\phi_1, \phi_2\}$ of $\mathbf{K}_{\mathbf{X}}$ as follows.

$$\begin{aligned} (\mathbf{K}_{\mathbf{X}} - \lambda_1\mathbf{I})\phi_1 &= \mathbf{0} \text{ with } \|\phi_1\| = 1 &\Rightarrow &\phi_1 = (1/\sqrt{2}, -1/\sqrt{2})^T \\ (\mathbf{K}_{\mathbf{X}} - \lambda_2\mathbf{I})\phi_2 &= \mathbf{0} \text{ with } \|\phi_2\| = 1 &\Rightarrow &\phi_2 = (1/\sqrt{2}, 1/\sqrt{2})^T \end{aligned}$$

At this point, we recall from page 273 that the whitening transformation is given by

$$\mathbf{D} = \mathbf{Z}\mathbf{U}^T, \text{ where } \mathbf{Z} = \begin{bmatrix} \lambda_1^{-1/2} & 0 \\ 0 & \lambda_2^{-1/2} \end{bmatrix}$$

and \mathbf{U} is the matrix whose columns are the eigenvectors of $\mathbf{K}_{\mathbf{X}}$. Substituting the previous results into this expression gives the following form for the whitening transformation \mathbf{D} .

$$\mathbf{D} = \begin{bmatrix} \frac{1}{\sqrt{7}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{7}} & -\frac{1}{\sqrt{7}} \\ 1 & 1 \end{bmatrix}$$

The whitening transformation is demonstrated using the attached MATLAB script `prob5.m`. Note that there are two applications of \mathbf{D} . First, we can apply \mathbf{D}^{-1} (i.e., the inverse whitening transformation) to a random vector \mathbf{Y} (with a covariance matrix equal to the identity matrix) to create a new random vector \mathbf{X} with covariance matrix $\mathbf{K}_{\mathbf{X}}$. This procedure is applied on lines 20-28 of `prob5.m`. In addition, we note that the built-in MATLAB function `mvnrnd` can be used to directly generate correlated samples from a bivariate Normal distribution. Afterwards, we can apply \mathbf{D} to transform to uncorrelated samples \mathbf{Y} . This procedure is applied on lines 30-33 of `prob5.m`. The resulting scatter plots of \mathbf{X} and \mathbf{Y} are shown in Figure 1.

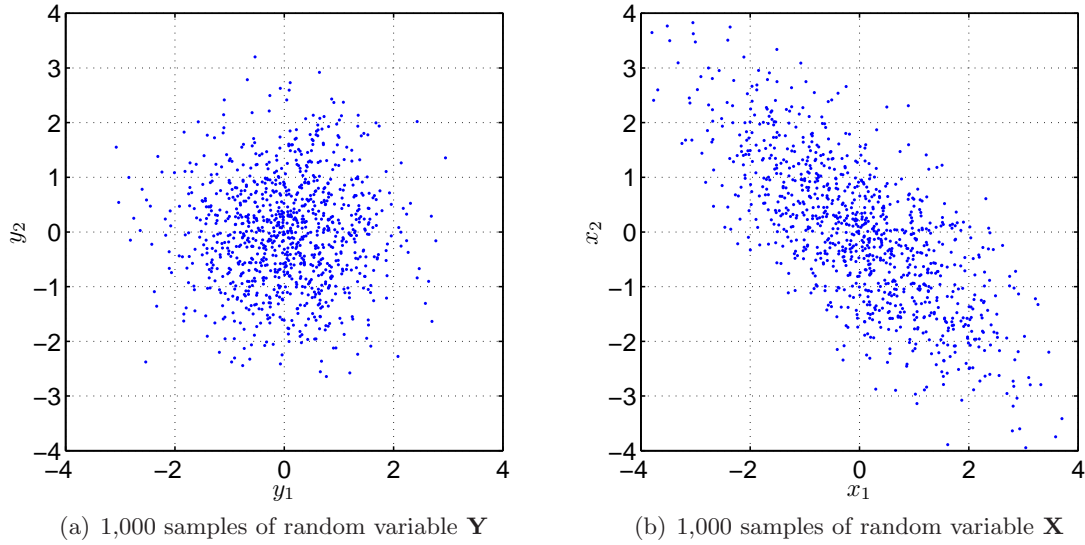


Figure 1: Whitening transformation. (a) 1,000 uncorrelated samples of a bivariate Normal distribution with identity covariance. (b) 1,000 samples of a bivariate Normal distribution with covariance \mathbf{K}_X , given in Equation 15, produced by applying the inverse whitening transformation \mathbf{D}^{-1} to \mathbf{Y} .

Problem 5.23

Let $Y_k = \sum_{j=1}^n a_{kj} X_j$, for $k = 1, \dots, n$, where $\{a_{kj}\} \in \mathbb{R}$, the matrix $\mathbf{A} = \{a_{kj}\}_{n \times n}$ is nonsingular, and the $\{X_j\}$, for $j = 1, \dots, n$, are random variables. Let $\mathbf{B} = \mathbf{A}^{-1}$. Show that the pdf of \mathbf{Y} , $f_{\mathbf{Y}}(y_1, \dots, y_n)$ is given by

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = |\det \mathbf{B}| f_{\mathbf{X}}(x_1^*, \dots, x_n^*), \text{ where } x_i^* = \sum_{k=1}^n b_{ik} y_k \text{ for } i = 1, \dots, n.$$

In this problem we will derive the pdf for a linear transformation of a random vector. First, consider the random vector $\mathbf{Y} = (y_1, \dots, y_n)^T$ that is a linear transformation of the random vector $\mathbf{X} = (x_1, \dots, x_n)^T$ such that $Y_k = \sum_{j=1}^n a_{kj} X_j$. Written as a matrix-vector product, we have

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \text{ where } \mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}.$$

Recall, from Equation 7 and Equation 3.4-11 on page 156 in [3], that the joint pdf $f_{\mathbf{Y}}(y_1, \dots, y_n)$ of a function of a random vector \mathbf{X} is given by

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = \sum_{i=1}^m |\tilde{J}_i| f_{\mathbf{X}}(\mathbf{x}_i), \quad (16)$$

where $\mathbf{x}_i \triangleq (x_1^{(i)}, \dots, x_n^{(i)})^T$, for $i = 1, \dots, m$, are the m solutions to $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$, and \tilde{J}_i is the associated Jacobian defined on the inverse mapping. Since \mathbf{A} is nonsingular, then we can apply the

the right inverse, given by $\mathbf{B} = \mathbf{A}^{-1}$, to obtain a closed-form expression for the the only solution \mathbf{x}^* to $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$.

$$\mathbf{X}^* = \mathbf{B}\mathbf{Y}, \text{ where } \mathbf{B} = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix}$$

By inspection, we conclude that the only solution \mathbf{x}^* to $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$ is given by

$$x_i^* = \sum_{k=1}^n b_{ik}y_k \text{ for } i = 1, \dots, n. \quad (17)$$

Substituting this result into Equation 16 gives the following preliminary result.

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = |\tilde{J}| f_{\mathbf{X}}(x_1^*, \dots, x_n^*), \text{ where } x_i^* = \sum_{k=1}^n b_{ik}y_k \text{ for } i = 1, \dots, n \quad (18)$$

At this point, all that remains is to prove that $|\det \mathbf{B}| = |\tilde{J}|$. In other words, we need to show that the Jacobian for the inverse mapping is given by the determinant of \mathbf{B} . Recall, from Equation 3.4-10 and Appendix C in [3], that the magnitude of the multi-dimensional Jacobian is given by

$$|\tilde{J}| = \left| \det \begin{pmatrix} \partial\phi_1(\mathbf{y})/\partial y_1 & \dots & \partial\phi_1(\mathbf{y})/\partial y_n \\ \vdots & \ddots & \vdots \\ \partial\phi_n(\mathbf{y})/\partial y_1 & \dots & \partial\phi_n(\mathbf{y})/\partial y_n \end{pmatrix} \right|, \quad (19)$$

where $x_i^* = \phi_i(\mathbf{y})$. Note that we have previously determined the inverse mapping $x_i^* = \phi_i(\mathbf{y})$ in Equation 17. As a result, we can calculate the elements of the Jacobian as follows.

$$\frac{\partial\phi_i(\mathbf{y})}{\partial y_j} = \frac{\partial}{\partial y_j} \sum_{k=1}^n b_{ik}y_k = b_{ij}$$

Substituting this result into Equation 19 gives the following expression for the Jacobian of the inverse mapping.

$$|\tilde{J}| = \left| \det \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} \right| = |\det \mathbf{B}|$$

Substituting this result into Equation 18 gives the desired expression for the pdf $f_{\mathbf{Y}}(y_1, \dots, y_n)$ for a linear transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ of a random vector \mathbf{X} .

$$\therefore \boxed{f_{\mathbf{Y}}(y_1, \dots, y_n) = |\det \mathbf{B}| f_{\mathbf{X}}(x_1^*, \dots, x_n^*), \text{ where } x_i^* = \sum_{k=1}^n b_{ik}y_k \text{ for } i = 1, \dots, n}$$

(QED)

Problem 6.1

Prove the chain rule for the probability of the intersection of N events $\{A_i\}$, for $i = 1, \dots, N$. For example, for $N = 3$ we have

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2|A_1]P[A_3|A_1 \cap A_2].$$

Interpret this result for joint probability distributions and joint probability densities.

To begin our analysis, we let $P\left[\bigcap_{i=1}^N A_i\right]$ denote the probability of the event that $\{A_i\}$, for $i = 1, \dots, N$, occur simultaneously. In general, we can express this quantity using conditional probabilities using the so-called chain rule.

$$P\left[\bigcap_{i=1}^N A_i\right] = P[A_1]P[A_2|A_1]P[A_3|A_1 \cap A_2] \dots P\left[A_N \mid \bigcap_{i=1}^{N-1} A_i\right]$$

In more compact notation, the chain rule is given by the following expression.

$$P\left[\bigcap_{i=1}^N A_i\right] = \prod_{i=1}^N P\left[A_i \mid \bigcap_{j=1}^{i-1} A_j\right] \quad (20)$$

We note that this equation can be verified by induction. First, consider the probability $P[A_1 \cap A_2]$. Using conditional probabilities we must have

$$P[A_1 \cap A_2] = P[A_1]P[A_2|A_1].$$

Now consider $P[A_1 \cap A_2 \cap A_3]$. Proceeding in an identical manner, and applying the previous result, we have

$$P[A_1 \cap A_2 \cap A_3] = P[A_1 \cap A_2]P[A_3|A_1 \cap A_2] = P[A_1]P[A_2|A_1]P[A_3|A_1 \cap A_2].$$

By induction we conclude that the general form of the chain rule is given by Equation 20. To complete our proof, we recall the definition of conditional probability give by Equation 1.6-3 on page 16 in [3]. For two events A and B , we have

$$P[A|B] \triangleq \frac{P[A \cap B]}{P[B]}, \text{ for } P[B] > 0.$$

Substituting this definition into Equation 20 yields the following result.

$$P\left[\bigcap_{i=1}^N A_i\right] = P[A_1] \left(\frac{P[A_1 \cap A_2]}{P[A_1]}\right) \left(\frac{P[A_1 \cap A_2 \cap A_3]}{P[A_1 \cap A_2]}\right) \dots \left(\frac{P\left[\bigcap_{i=1}^N A_i\right]}{P\left[\bigcap_{i=1}^{N-1} A_i\right]}\right) = P\left[\bigcap_{i=1}^N A_i\right]$$

Since the left-hand and right-hand sides of this expression are identical, we conclude that Equation 20 is the correct form for the chain rule expressing the intersection of N events.

Now let's consider how the chain rule can be applied to better understand the behavior of joint probability distributions. Recall, from Equation 5.1-1 on page 245 in [3], that the joint probability distribution function (PDF) $F_{\mathbf{X}}(\mathbf{x})$ of a random vector \mathbf{X} is given by

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq P[\{X_1 \leq x_1\} \cap \dots \cap \{X_N \leq x_N\}],$$

where $\mathbf{x} = (x_1, \dots, x_n)^T$. Substituting this general form of a joint PDF into Equation 20 yields the following expression.

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^N P \left[\{X_i \leq x_i\} \mid \bigcap_{j=1}^{i-1} \{X_j \leq x_j\} \right]$$

We note that the right-hand side of this expression is simply the product of a set of conditional PDF's, such that

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^N F_{X_i|\{X_j\}_{j=1}^{i-1}}(x_i|\{x_j\}_{j=1}^{i-1}). \quad (21)$$

As a result, we conclude that the chain rule allows one to separate a joint PDF into a product of conditional PDF's using Equation 21. For example, for $N = 3$ we have

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = F_{X_1}(x_1) F_{X_2|X_1}(x_2|x_1) F_{X_3|X_1 X_2}(x_3|x_1, x_2).$$

Finally, we note that the chain rule can also be used to understand the general behavior of joint probability densities. Recall, from Equation 5.1-4 on page 245 in [3], that the joint probability density function (pdf) $f_{\mathbf{X}}(\mathbf{x})$ of a random vector \mathbf{X} is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq \frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_n},$$

where $\mathbf{x} = (x_1, \dots, x_n)^T$. Applying this result to Equation 21 gives the so-called chain rule of probability from Equation 6.5-1 on page 362 in [3].

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^N f_{X_i|\{X_j\}_{j=1}^{i-1}}(x_i|\{x_j\}_{j=1}^{i-1}). \quad (22)$$

As a result, we conclude that the chain rule allows one to separate a joint pdf into a product of conditional pdf's using Equation 22.

References

- [1] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes (Third Edition)*. Oxford University Press, 2001.
- [2] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
- [3] Henry Stark and John W. Woods. *Probability and Random Processes with Applications to Signal Processing (Third Edition)*. Prentice-Hall, 2002.

```
1
2 % EN 257, Problem Set 3, Problem 5.22
3 %
4 % Whitening Transformation Demo
5 %     Applies the inverse of the whitening transformation D
6 %     to a bivariate Normal random variable (with identity
7 %     covariance) to generate a sample from a bivariate
8 %     Normal distribution with user-defined covariance K.
9 %
10 % Douglas Lanman, Brown University, Feb. 2007
11
12 % Reset Matlab environment.
13 clear; clc;
14
15 % Define number of samples, covariance matrix K, and whitening transform D.
16 n = 1e3; % number of random samples
17 K = [2 -1.5; -1.5 2]; % covariance matrix
18 D = [1/sqrt(7) -1/sqrt(7); 1 1]; % whitening transform
19
20 % Generate zero-mean jointly Gaussian samples with identity covariance.
21 % Note: This simply involves drawing two sets of samples
22 %     from a univariate Gaussian distribution.
23 Y = randn(2,n);
24
25 % Apply the inverse whitening transformation.
26 % Note: This obtains samples with covariance matrix K,
27 %     using the procedure described on page 274.
28 X = inv(D)*Y;
29
30 % Generate samples from jointly Normal distribution with covariance K.
31 % Note: This confirms the previous results with {X,Y}.
32 A = mvnrnd([0 0]',K,n)';
33 B = D*A;
34
35 % Display estimated covariance matrices.
36 disp('cov(Y) = ');
37 disp(cov(Y'));
38 disp('cov(X) = ');
39 disp(cov(X'));
40 disp('cov(B) = ');
41 disp(cov(B'));
42 disp('cov(A) = ');
43 disp(cov(A'));
44
45 % Display a scatter plot for the uncorrelated samples Y.
46 figure(1); clf;
47 plot(Y(1,:),Y(2,:),'.');
48 axis([-4 4 -4 4]); axis square; grid on;
49 set(gca, 'LineWidth',1, 'FontSize',16, 'YDir', 'normal');
50 set(get(gca, 'XLabel'), 'FontSize',18, 'Interpreter', 'LaTeX');
```

```
51 set(get(gca, 'YLabel'), 'FontSize', 18, 'Interpreter', 'LaTeX');
52 xlabel('$y_1$');
53 ylabel('$y_2$');
54
55 % Display a scatter plot for the correlated samples X.
56 figure(2); clf;
57 plot(X(1,:), X(2,:), '.');
58 axis([-4 4 -4 4]); axis square; grid on;
59 set(gca, 'LineWidth', 1, 'FontSize', 16, 'YDir', 'normal');
60 set(get(gca, 'XLabel'), 'FontSize', 18, 'Interpreter', 'LaTeX');
61 set(get(gca, 'YLabel'), 'FontSize', 18, 'Interpreter', 'LaTeX');
62 xlabel('$x_1$');
63 ylabel('$x_2$');
```