

SIMG-713

Homework 5 Solutions

Spring 2002

1. Photons strike a detector at an average rate of λ photons per second. The detector produces an output with probability β whenever it is struck by a photon. Compute the DQE of the detector.

Solution: The input X is a Poisson random variable with mean and variance $\mu_x = \sigma_x^2 = \lambda$. The output is also a Poisson random process, but now the rate is $\beta\lambda$, so the output has mean and variance $\mu_y = \sigma_y^2 = \beta\lambda$. The DQE is the ratio of the output SNR to the input SNR, or

$$DQE = \left(\frac{SNR_o}{SNR_I} \right)^2 = \frac{\mu_y^2 / \sigma_y^2}{\mu_x^2 / \sigma_x^2} = \frac{\beta\lambda}{\lambda} = \beta$$

2. This problem examines the correlation and covariance of two random variables \mathbf{X} and \mathbf{Y} . The covariance function has been defined as

$$\text{cov}(X, Y) = E[(\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y)]$$

where $\mu_x = E[\mathbf{X}]$ and $\mu_y = E[\mathbf{Y}]$. The correlation coefficient is defined to be the ratio

$$\rho_{xy} = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\sigma_x \sigma_y}$$

where, respectively, σ_x^2 and σ_y^2 are the variances of \mathbf{X} and \mathbf{Y} .

- (a) This is a result that will be needed in some other parts of this problem. Show that the function $v = au^2 + bu + c$ is positive for all values of u provided $c > 0$ and $b^2 < 4ac$. The quantity $d = b^2 - 4ac$ is called the *discriminant*. The condition that $v > 0$ for all values of u is equivalent to $d < 0$.

Solution: $v(0) = c$, so $c \geq 0$. To find the minimum point solve $v'(u) = 2au + b = 0$ which yields $u_{\min} = -b/2a$. Then $v(u_{\min}) = a(b^2/4a^2) - b(b/2a) + c = -b^2/4a + c \geq 0$ implies $d = b^2 - 4ac \leq 0$. Another method is to solve for the roots of v by the quadratic formula. This yields roots at $(-b \pm \sqrt{b^2 - 4ac})/2a$. The roots will be real, and therefore v will cross into negative territory, unless the term under the radical is negative. This requires $d \leq 0$.

- (b) Show that $|\text{cov}(\mathbf{X}, \mathbf{Y})| \leq \sigma_x \sigma_y$. Hint: Look at $E\left[\left((\mathbf{X} - \mu_x)u + (\mathbf{Y} - \mu_y)\right)^2\right]$ as a function of u . Make use of the discriminant.

Solution: Clearly $E\left[\left((\mathbf{X} - \mu_x)u + (\mathbf{Y} - \mu_y)\right)^2\right] \geq 0$. We now need to identify the discriminant. Expand the expression:

$$\begin{aligned} E\left[\left((\mathbf{X} - \mu_x)u + (\mathbf{Y} - \mu_y)\right)^2\right] &= E\left[(\mathbf{X} - \mu_x)^2\right]u^2 + 2E\left[(\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y)\right]u + E\left[(\mathbf{Y} - \mu_y)^2\right] \\ &= \sigma_x^2 u^2 + 2\text{cov}(\mathbf{X}, \mathbf{Y})u + \sigma_y^2 \end{aligned}$$

Identify the discriminant terms as $a = \sigma_x^2$, $b = 2\text{cov}(\mathbf{X}, \mathbf{Y})$ and $c = \sigma_y^2$. Since we know that $d \leq 0$ we have $4\text{cov}^2(\mathbf{X}, \mathbf{Y}) \leq 4\sigma_x^2\sigma_y^2$. The result follows.

- (c) Show that $|\rho_{xy}| \leq 1$.

Solution: Since $\rho = \text{cov}(\mathbf{X}, \mathbf{Y}) / \sigma_x \sigma_y$, this result follows directly from the above.

(d) Show that for any two random variables \mathbf{X} and \mathbf{Y} , $E^2[\mathbf{XY}] \leq E[\mathbf{X}^2] E[\mathbf{Y}^2]$.

Solution: Let $v = E[(\mathbf{X}u + \mathbf{Y})^2]$ and follow the method of part (b).

(e) Show that if $E[\mathbf{X}^2] = E[\mathbf{Y}^2]$ then $\mathbf{S} = \mathbf{X} + \mathbf{Y}$ and $\mathbf{D} = \mathbf{X} - \mathbf{Y}$ are orthogonal random variables.

Solution: $E[\mathbf{SD}] = E[(\mathbf{X} + \mathbf{Y})(\mathbf{X} - \mathbf{Y})] = E[\mathbf{X}^2] - E[\mathbf{Y}^2] = 0$.

(f) Show that if random variables \mathbf{X} and \mathbf{Y} are uncorrelated then $\sigma_s^2 = \sigma_x^2 + \sigma_y^2$, where $\mathbf{S} = \mathbf{X} + \mathbf{Y}$. Extend this to the sum of any number of uncorrelated random variables. Note that we once proved this for statistically independent random variables, but that this requirement is new because it is weaker.

Solution: First compute the mean value $\mu_s = \mu_x + \mu_y$. This is always true. Then $\sigma_s^2 = E[(\mathbf{S} - \mu_s)^2] = E[(\mathbf{X} + \mathbf{Y} - \mu_x - \mu_y)^2] = E[(\mathbf{X} - \mu_x)^2] - 2E[(\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y)] + E[(\mathbf{Y} - \mu_y)^2]$. If \mathbf{X} and \mathbf{Y} are uncorrelated then the middle term is zero, and the result follows.

3. If \mathbf{X} and \mathbf{Y} are jointly normal random variables, then so are random variables $\mathbf{S} = a\mathbf{X} + b\mathbf{Y}$ and $\mathbf{D} = c\mathbf{X} + d\mathbf{Y}$ for any real coefficients (a, b, c, d) .

(a) Given $\mu_x = 10$, $\mu_y = 0$, $\sigma_x = 2$, $\sigma_y = 1$, $\rho_{xy} = 0.5$ construct a contour plot of for $f_{XY}(x, y)$ showing the locus of constant probability points

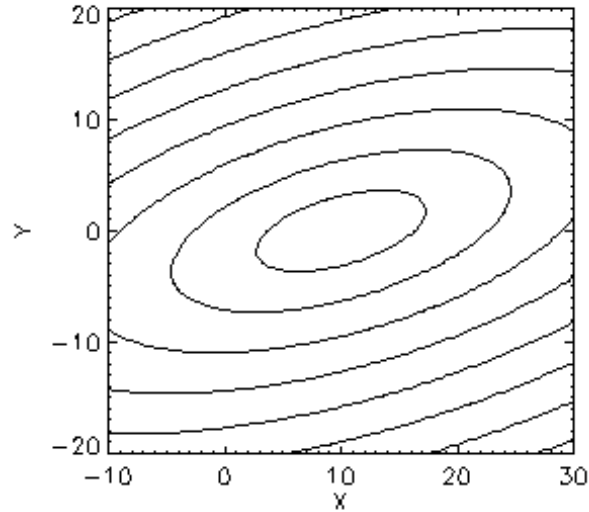
Solution: The joint probability density function is

$$\begin{aligned} f_{XY}(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \exp\left[-\frac{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2}{2(1-\rho^2)}\right] \\ &= \frac{1}{2\pi(2)(1)\sqrt{.75}} \exp\left[-\frac{\left(\frac{x-10}{2}\right)^2 - \left(\frac{x-10}{2}\right)(y) + y^2}{2(0.75)}\right] \end{aligned}$$

The probability is constant on contours where the exponential term is constant. This requires that

$$\left(\frac{x-10}{2}\right)^2 - \left(\frac{x-10}{2}\right)(y) + y^2 = C$$

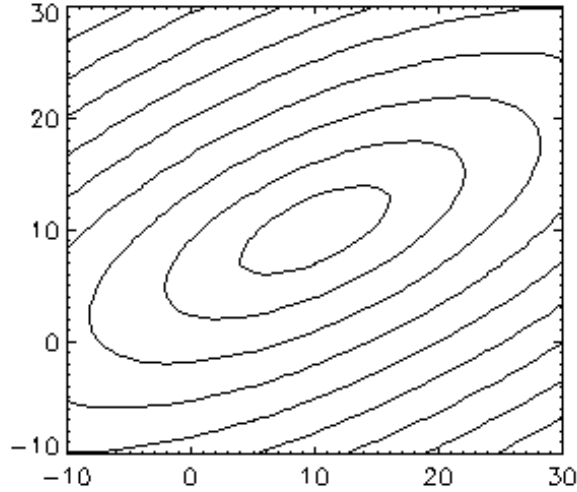
Contours of constant probabilities form ellipses as shown in the figure below.



- (b) Find the probability density function $f_{SD}(x, y)$ for the random variables $\mathbf{S} = \mathbf{X} + \mathbf{Y}$ and $\mathbf{D} = \mathbf{X} - \mathbf{Y}$.

Solution: We can find the parameters μ_s , μ_d , σ_s , σ_d and ρ_{sd} and then substitute into the general form for a bivariate normal distribution. We find $\mu_s = \mu_x + \mu_y = 10$, $\mu_d = \mu_x - \mu_y = 10$, $\sigma_s^2 = E[(\mathbf{S} - \boldsymbol{\mu}_s)^2] = E[(\mathbf{X} - \boldsymbol{\mu}_x) + (\mathbf{Y} - \boldsymbol{\mu}_y)]^2 = E[(\mathbf{X} - \boldsymbol{\mu}_x)^2] + 2E[(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)] + E[(\mathbf{Y} - \boldsymbol{\mu}_y)^2] = \sigma_x^2 + 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2 = 4 + 2 \cdot (2 \cdot 1 \cdot 0.5) + 1 = 7$. Similarly, $\sigma_d^2 = E[(\mathbf{D} - \boldsymbol{\mu}_d)^2] = E[(\mathbf{X} - \boldsymbol{\mu}_x) - (\mathbf{Y} - \boldsymbol{\mu}_y)]^2 = \sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2 = 4 - 2 \cdot (2 \cdot 1 \cdot 0.5) + 1 = 3$. $\text{cov}(\mathbf{S}, \mathbf{D}) = E[(\mathbf{S} - \boldsymbol{\mu}_s)(\mathbf{D} - \boldsymbol{\mu}_d)] = E[(\mathbf{X} - \boldsymbol{\mu}_x) + (\mathbf{Y} - \boldsymbol{\mu}_y)]((\mathbf{X} - \boldsymbol{\mu}_x) - (\mathbf{Y} - \boldsymbol{\mu}_y)) = E[(\mathbf{X} - \boldsymbol{\mu}_x)^2] - E[(\mathbf{Y} - \boldsymbol{\mu}_y)^2] = \sigma_x^2 - \sigma_y^2 = 3$. Finally, $\rho_{sd} = \text{cov}(\mathbf{S}, \mathbf{D}) / \sigma_s\sigma_d = 3 / \sqrt{3 \cdot 7} = 0.65$

(c) Construct a contour plot of for $f_{SD}(x, y)$ showing the locus of constant probability points.



4. Let X and Y be random variables with a joint probability density function $f_{XY}(x, y)$. Let $\hat{Y} = g(X)$ be a predictor of Y .

(a) Show that the mean-squared prediction error can be expressed as

$$E \left[(Y - \hat{Y})^2 \right] = E [Y^2] - 2E [g_0(X) g(X)] + E [g^2(X)]$$

where $g_0(X) = E [Y|X]$.

Solution: Expand the expectation and substitute for $\hat{Y} = g(X)$. $E \left[(Y - \hat{Y})^2 \right] = E [Y^2] - 2E [Y\hat{Y}] + E [\hat{Y}^2] = E [Y^2] - 2E [Yg(X)] + E [g^2(X)]$. We only need to show that the middle term yields $E [Yg(X)] = E [g_0(X)g(X)]$. Write out the expectation in terms of the joint probability function:

$$\begin{aligned} E [Yg(X)] &= \sum_x \sum_y Yg(X) P(X, Y) \\ &= \sum_x g(X) P(X) \sum_y YP(Y|X) \\ &= \sum_x P(X) (g(X) E [Y|X]) \\ &= E [g(X)g_0(X)] \end{aligned}$$

where $E [Y|X] = g_0(X) = \sum_y YP(Y|X)$ is a function of X obtained by computing the mean value of the conditional probability function $P(Y|X)$. This function is equal to the mean value of the conditional probability function on Y when X is given and shown below to be the best estimate based on the given information about X .

- (b) Let X and Y be random variables with a joint probability density function $f_{XY}(x, y)$. Let $\hat{Y}_0 = g_0(X) = E[Y|X]$ be a predictor of Y . Show that the mean-squared prediction error can be expressed as

$$E \left[(Y - \hat{Y}_0)^2 \right] = E [Y^2] - E [g_0^2(X)]$$

Solution: $E \left[(Y - \hat{Y}_0)^2 \right] = E [Y^2] - 2E [Yg_0(X)] + E [g_0^2(X)]$. The middle term now becomes

$$\begin{aligned} E [Yg_0(X)] &= \sum_x \sum_y Yg_0(X) P(X, Y) \\ &= \sum_x g_0(X) P(X) \sum_y YP(Y|X) \\ &= \sum_x P(X) (g_0(X) E[Y|X]) \\ &= E [g_0^2(X)] \end{aligned}$$

and therefore combines with the last term. This yields $E \left[(Y - \hat{Y}_0)^2 \right] = E [Y^2] - E [g_0^2(X)]$

- (c) Show that the previous two problems establish that

$$E \left[(Y - \hat{Y})^2 \right] \geq E \left[(Y - \hat{Y}_0)^2 \right]$$

for any prediction function $\hat{Y} = g(X)$. That is, the conditional expectation of Y given X gives the least-mean-square prediction of Y . This shows that $g_0(X) = E[Y|X]$ is the “optimum” predictor function and provides a tool to find the optimum predictor.

Solution: By subtraction of the above results we find

$$\begin{aligned} E \left[(Y - \hat{Y})^2 \right] - E \left[(Y - \hat{Y}_0)^2 \right] &= E [g^2(X)] - 2E [g(X)g_0(X)] + E [g_0^2(X)] \\ &= E [(g(X) - g_0(X))^2] \geq 0 \end{aligned}$$

5. Suppose that $Y = \alpha X + \beta + Z$ where Z is a random variable statistically independent of X with mean $E[Z] = \mu_z$. What is the optimum predictor function $\hat{Y} = g(X)$, based on the above results?

Solution: $g_0(X) = E[\alpha X + \beta + Z|X] = \sum_z (\alpha X + \beta + Z) P(Z|X) = \sum_z (\alpha X + \beta + Z) P(Z)$ because Z is statistically independent of X . Then $\hat{Y}_0 = g_0(X) = \alpha X + \beta + \mu_z$.

6. In this problem we will construct a formulation of the probability density function for the bivariate normal distribution based on the covariance matrix and mean values. This approach extends to any number of dimensions and is very useful in constructing algorithms. We begin by assuming that $\mathbf{X} = [X_1, X_2]^T$ is a column vector whose elements are statistically independent normal random variables.

- (a) Show that

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi \det(\Gamma)^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_x)^T \Gamma^{-1} (\mathbf{x} - \mathbf{m}_x)}$$

where $\mathbf{m} = [E[X_1], E[X_2]]^T$ is a column vector of the mean values and Γ is the covariance matrix

$$\Gamma = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

In reality, this is just a compact way to express the equation

$$f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[- \left[\left(\frac{x_1 - m_1}{\sqrt{2}\sigma_1} \right)^2 + \left(\frac{x_2 - m_2}{\sqrt{2}\sigma_2} \right)^2 \right] \right]$$

Solution: The inverse is

$$\Gamma^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix}$$

as can be verified by computing $\Gamma\Gamma^{-1} = I$. Then

$$\begin{aligned} (\mathbf{x} - \mathbf{m}_x)^T \Gamma^{-1} (\mathbf{x} - \mathbf{m}_x) &= [X_1 - m_1 \quad X_2 - m_2] \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} X_1 - m_1 \\ X_2 - m_2 \end{bmatrix} \\ &= [(X_1 - m_1)/\sigma_1^2 \quad (X_2 - m_2)/\sigma_2^2] \begin{bmatrix} X_1 - m_1 \\ X_2 - m_2 \end{bmatrix} \\ &= (X_1 - m_1)^2 / \sigma_1^2 + (X_2 - m_2)^2 / \sigma_2^2 \end{aligned}$$

which is just the exponential term for the bivariate distribution when X_1 and X_2 are statistically independent. The term $\det(\Gamma)^{1/2} = \sigma_1\sigma_2$ which is the required factor in the denominator for the multiplier of the exponential function. When the substitutions are made the result is the known bivariate pdf above.

- (b) Let \mathbf{G} be a square matrix with $\det(\mathbf{G}) \neq 0$ and let $\mathbf{Y} = \mathbf{G}\mathbf{X}$. That is, \mathbf{Y} is a vector of random variables formed by a linear combination of elements of \mathbf{X} . The only restriction we are making is that the transformation should have an inverse. Show that $\mathbf{m}_y = \mathbf{G}\mathbf{m}_x$. This means that $\mathbf{Y} - \mathbf{m}_y = \mathbf{G}(\mathbf{X} - \mathbf{m}_x)$.

Solution: Each component of \mathbf{Y} is just a function of the elements of \mathbf{X} . For example, $Y_1 = g_{11}X_1 + g_{12}X_2$. This is just a function of two random variables, so that we can calculate the mean value by the usual method: $E[Y_1] = g_{11}E[X_1] + g_{12}E[X_2] = g_{11}m_{x1} + g_{12}m_{x2}$. Similarly, $E[Y_2] = g_{21}E[X_1] + g_{22}E[X_2] = g_{21}m_{x1} + g_{22}m_{x2}$. The same results are produced by multiplying \mathbf{m}_x by \mathbf{G} .

- (c) Show that the covariance matrix for \mathbf{Y} is

$$\Lambda = \mathbf{G}\Gamma\mathbf{G}^T$$

Solution:

$$\begin{aligned} \text{var}(Y_1) &= E[(Y_1 - m_{y1})^2] \\ &= E[(g_{11}(X_1 - m_{x1}) + g_{12}(X_2 - m_{x2}))^2] \\ &= g_{11}^2 E[(X_1 - m_{x1})^2] + 2g_{11}g_{12}E[(X_1 - m_{x1})(X_2 - m_{x2})] + g_{12}^2 E[(X_2 - m_{x2})^2] \\ &= g_{11}^2 \text{var}(X_1) + 2g_{11}g_{12} \text{cov}(X_1, X_2) + g_{12}^2 \text{var}(X_2) \end{aligned}$$

Similarly, $\text{var}(Y_2) = g_{21}^2 \text{var}(X_1) + 2g_{21}g_{22} \text{cov}(X_1, X_2) + g_{22}^2 \text{var}(X_2)$. Finally,

$$\begin{aligned} \text{cov}(Y_1, Y_2) &= E[(Y_1 - m_{y1})(Y_2 - m_{y2})] \\ &= E[(g_{11}(X_1 - m_{x1}) + g_{12}(X_2 - m_{x2}))(g_{21}(X_1 - m_{x1}) + g_{22}(X_2 - m_{x2}))] \\ &= g_{11}g_{21} \text{var}(X_1) + (g_{11}g_{22} + g_{12}g_{21}) \text{cov}(X_1, X_2) + g_{12}g_{22} \text{var}(X_2) \end{aligned}$$

Therefore,

$$\Lambda = \begin{bmatrix} \text{var}(Y_1) & \text{cov}(Y_1, Y_2) \\ \text{cov}(Y_1, Y_2) & \text{var}(Y_2) \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) \end{bmatrix} \begin{bmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{bmatrix}$$

as can be verified by multiplying out the matrix expression on the right.

(d) One can make a change of variables in n dimensions by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{G}^{-1}\mathbf{y}) |\det(\mathbf{G}^{-1})|$$

The exponent is transformed by

$$\begin{aligned} (\mathbf{x} - \mathbf{m}_x)^T \Gamma^{-1} (\mathbf{x} - \mathbf{m}_x) &= [\mathbf{G}^{-1}(\mathbf{y} - \mathbf{m}_y)]^T (\mathbf{G}^{-1} \Lambda \mathbf{G}^{-T})^{-1} [\mathbf{G}^{-1}(\mathbf{y} - \mathbf{m}_y)] \\ &= (\mathbf{y} - \mathbf{m}_y)^T \mathbf{G}^{-T} \mathbf{G}^T \Lambda^{-1} \mathbf{G} \mathbf{G}^{-1} (\mathbf{y} - \mathbf{m}_y) \\ &= (\mathbf{y} - \mathbf{m}_y)^T \Lambda^{-1} (\mathbf{y} - \mathbf{m}_y) \end{aligned}$$

Also, $\det(\Gamma) = \det(\mathbf{G}^{-1} \Lambda \mathbf{G}^{-T}) = \det(\Lambda) \det^2(\mathbf{G}^{-1})$. When all this is substituted back we find

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi \det(\Lambda)^{1/2}} e^{-(\mathbf{y} - \mathbf{m}_y)^T \Lambda^{-1} (\mathbf{y} - \mathbf{m}_y)/2}$$

This is exactly the same form, but now it accommodates random variables that are not uncorrelated. This is a demonstration that a linear transformation of normal random variables produces another set of normal random variables. Assume that we are working in 2D and that X_1 and X_2 are statistically independent normal random variables. Find expressions for m_{y1} , m_{y2} , σ_{y1} , σ_{y2} , and ρ in terms of m_{x1} , m_{x2} , σ_{x1} , σ_{x2} and the elements of the transformation matrix

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

Assume $\det \mathbf{G} \neq 0$. Write an expression for $f_{\mathbf{Y}}(\mathbf{y})$ in terms of m_{y1} , m_{y2} , σ_{y1} , σ_{y2} , and ρ .

Solution: Above we computed expressions for

$$\begin{aligned} m_{y1} &= g_{11}m_{x1} + g_{12}m_{x2} \\ m_{y2} &= g_{21}m_{x1} + g_{22}m_{x2} \\ \sigma_{y1}^2 &= g_{11}^2\sigma_{x1}^2 + 2g_{11}g_{12} \text{cov}(X_1, X_2) + g_{12}^2\sigma_{x2}^2 \\ \sigma_{y2}^2 &= g_{21}^2\sigma_{x1}^2 + 2g_{21}g_{22} \text{cov}(X_1, X_2) + g_{22}^2\sigma_{x2}^2 \\ \text{cov}(Y_1, Y_2) &= g_{11}g_{21}\sigma_{x1}^2 + (g_{11}g_{22} + g_{12}g_{21}) \text{cov}(X_1, X_2) + g_{12}g_{22}\sigma_{x2}^2 \end{aligned}$$

We can compute the correlation coefficient

$$\rho_y = \frac{\text{cov}(Y_1, Y_2)}{\sigma_{y1}\sigma_{y2}}$$

using the above components. From this the probability density function can be assembled. Incidentally, the above equations can be used to check the results of Problem 3(b).