EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 14 - Outline

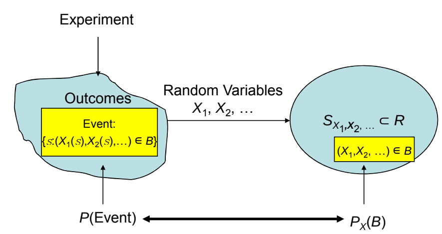
1. **Random Vectors**

Random Vectors

- Generalize concepts from 2 random variables on the same probability space (Chapter 4, “Pairs of Random Variables”)

 to

- N random variables on the same probability space (Chapter 5, “Random Vectors”)



Probability Models for N RVs

- Let $X_1, X_2, \ldots, X_n$ be n random variables defined on a sample space
- Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector (All vectors are assumed to be column vectors unless stated otherwise)
- Let $\mathrm{u} = (u_1, u_2, \ldots, u_n)$ be a real vector
- **Notation:** $\{\mathbf{X} \le \mathrm{u}\}$ denotes $\{X_1 \le u_1, X_2 \le u_2, \ldots, X_n \le u_n\}$, where, as before, the commas denote intersections,

$$\{\mathbf{X} \le \mathrm{u}\} = \{X_1 \le u_1\} \cap \{X_2 \le u_2\} \cap \ldots \cap \{X_n \le u_n\}$$

Multivariate Joint CDF

- The joint CDF of $X_1, X_2, \ldots, X_n$ or the CDF of the random vector $\mathbf{X}$ is defined as

  $F_{\mathbf{X}}(\mathrm{x}) = P\{\mathbf{X} \le \mathrm{x}\}$

  $= P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\}$

- $F_{\mathbf{X}}(\mathrm{x})$ is a real-valued function of n real variables (or of the n-vector x)
- $F_{\mathbf{X}}(\mathrm{x})$ always has value between 0 and 1
- $F_{\mathbf{X}}(\mathrm{x})$ is a non-decreasing, right-continuous function of each argument x_i

Joint CDF Properties

- $\lim_{x_j \to -\infty} F_{\mathbf{X}}(\mathrm{x}) = 0$
- If some of the $x_j \to +\infty$, the corresponding random variables X_j disappear and we get the joint CDF of the remaining variables
- Example: If $F_{X,Y,Z}(x,y,z)$ is the joint CDF of X, Y, Z, then $F_{X,Y,Z}(x,\infty,z) = F_{X,Z}(x,z)$ is the joint CDF of X and Z
- Even though this is still a joint CDF, it is nevertheless also a marginal CDF, since it describes a subset of the variables.

Discrete Random Variables: Joint PMF

- The joint PMF of $X_1, X_2, \ldots, X_n$ (the PMF of the random vector **X**) is defined as

  $p_{\mathbf{X}}(\mathrm{x})$ = $P\{\mathbf{X} = \mathrm{x}\}$ (column vector x is sample value of **X**)

  = $P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$
- $p_{\mathbf{X}}(\mathrm{x}) \geq 0$
- $\sum_{x_1} \sum_{x_2} \ldots \sum_{x_n} p_{\mathbf{X}}(\mathrm{x}) = 1$

Continuous Random Variables: Joint PDF

- The marginal PDF of any subset of $\{X_1, X_2, \ldots, X_n\}$ is obtained by summing over the unwanted variables

Example

- Quiz 5.2
  - Discrete random vectors **X**, **Y** (each with 3 components)
    - Related by **Y = AX**
    - Find joint PMF of **Y,** i.e., $P_{\mathbf{Y}}(\mathrm{y})$, if:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

$$p_{\underline{X}}(x) = \begin{cases} (1-p)p^{x_3} & x_1 < x_2 < x_3, x_i \in \{1,2,\ldots\} \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow y_1 = x_1, y_2 = x_2 - x_1, y_3 = x_3 - x_2$$

Example (continued)

- Range of **Y**: $\{1, 2, \ldots\}^3$

$$y_1 = x_1, y_2 = x_2 - x_1, y_3 = x_3 - x_2$$
$$\Rightarrow x_1 = y_1, x_2 = y_2 + y_1, x_3 = y_1 + y_2 + y_3$$
$$\Rightarrow \text{one} - \text{to} - \text{one map!}$$
$$p_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = p_{X_1,X_2,X_3}(y_1, y_2 + y_1, y_1 + y_2 + y_3)$$
$$= \begin{cases} (1-p)p^{y_1+y_2+y_3} & y_1, y_2, y_3 \in \{1,2,\ldots\} \\ 0 & \text{otherwise} \end{cases}$$

Jointly Continuous Random Vectors

- $X_1, X_2, \ldots, X_n$ are called jointly continuous random variables if
  - **X** = $(X_1, X_2, \ldots, X_n)$ takes on all possible values in a region of *nonzero volume* in n-dimensional space, and
  - The probabilistic behavior is described by the n-variate joint PDF

$$f_{\mathbf{X}}(\mathrm{x}) = f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}$$

More properties

- **X** jointly continuous
  - Let A be an event expressed in terms of the random vector **X**

$$P[A] = \int \cdots \int_A f_{\underline{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

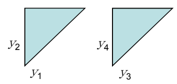
- Example: Quiz 5.1: $Y_1, \ldots, Y_4$ distributed as

$$f_{\underline{Y}}(y_1, \ldots, y_4) = \begin{cases} 4 & 0 \leq y_1 \leq y_2 \leq 1, 0 \leq y_3 \leq y_4 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let C be event that $\{\max_i Y_i < 1/2\}$. Find $P[C]$?

Example (Solution)

- Note that C implies
  $Y_i < 1/2$ for all i



$$
\begin{aligned}
P[C] &= \int_0^{0.5} dy_2 \int_0^{y_2} dy_1 \int_0^{0.5} dy_4 \int_0^{y_4} 4 dy_3 \\
&= 4\left(\int_0^{0.5} dy_2 \int_0^{y_2} dy_1\right)\left(\int_0^{0.5} dy_4 \int_0^{y_4} dy_3\right) \\
&= \frac{1}{16}
\end{aligned}
$$

Marginal Probabilities

- Given any discrete random vector **X** with PMF function $P_{\mathbf{X}}(\mathrm{x})$, get marginals for any subset of RVs in **X** by:
  - Summing over all RVs not in subset
  - e.g. given $X_1, \ldots, X_n$

$$P_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} P_{\underline{X}}(x_1, \ldots, x_n)$$

$$P_{X_i,X_j}(x_i, x_j) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} P_{\underline{X}}(x_1, \ldots, x_n)$$

Marginal Probabilities: Continuous Random Vectors

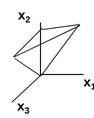
- Given any continuous random vector **X** with PDF function $f_{\mathbf{X}}(\mathrm{x})$, get marginal PDFs for any subset of RVs in **X** by:
  - Integrating over all RVs not in subset
  - e.g. given $X_1, \ldots, X_n$

$$f_{X_i}(x_i) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_n} f_{\underline{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

$$f_{X_i,X_j}(x_i, x_j) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_{j-1}} \int_{x_{j+1}} \cdots \int_{x_n} f_{\underline{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_{j-1} dx_{j+1} \ldots dx_n$$

Example

- Quiz 5.3: three-component **X** with PDF



$$f_{\underline{X}}(\underline{x}) = \begin{cases} 6 & 0 \le x_1 \le x_2 \le x_3 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$
\begin{aligned}
f_{X_1X_2}(x_1, x_2) &= \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, x_2, x_3) dx_3 \\
&= 6 \int_{x_2}^{1} dx_3 = 6(1 - x_2),\ 0 \le x_1 \le x_2 \le 1 \\
f_{X_1X_3}(x_1, x_3) &= 6 \int_{x_1}^{x_3} dx_2 = 6(x_3 - x_1),\ 0 \le x_1 \le x_3 \le 1 \\
f_{X_2X_3}(x_1, x_3) &= 6 \int_{0}^{x_2} dx_1 = 6x_2,\ 0 \le x_2 \le x_3 \le 1 \\
f_{X_1}(x_1) &= \int_{x_1}^{1} 6(1 - x_2) dx_2 = 3(1 - x_1)^2,\ 0 \le x_1 \le 1 \\
f_{X_3}(x_3) &= \int_{0}^{x_3} 6x_2 dx_2 = 3x_3^2,\ 0 \le x_3 \le 1
\end{aligned}
$$

Independence

- Random variables $X_1, \ldots, X_n$ are independent if and only if
  - Discrete: $P_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} P_{X_i}(x_i)$
  - Continuous: $f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Example

- Ex. 5.7, Q. 5.4
  - **W** = (W_1, W_2, W_3, W_4) $\quad f_{\underline{W}}(\underline{w}) = \begin{cases} 1 & 0 \le w_i \le 1 \\ 0 & \text{otherwise} \end{cases}$
  - Clearly independent!
  - **Y** defined as $y_1 = w_1$, $y_2 = w_1 + w_2$, $y_3 = w_3$, $y_4 = w_3 + w_4$
  - Not independent individually, but may have good properties
    - Clearly, the part of the experiment that generates y_1, y_2 is independent of the part that generates y_3, y_4

$$
\begin{aligned}
F_{Y_2Y_4}(y_2, y_4) &\equiv P(Y_2 \le y_2, Y_4 \le y_4) \\
&= P(w_1 + w_2 \le y_2, w_3 + w_4 \le y_4) \\
&= P(w_1 + w_2 \le y_2)P(w_3 + w_4 \le y_4) \text{ independence of } w_i \\
&= F_{Y_2}(y_2)F_{Y_4}(y_4)
\end{aligned}
$$

Covariance and Cross-Correlation

- Definition: Covariance of two RVs, *X* and *Y*, is

$$Cov[X,Y] \equiv E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]$$

- Definition: Correlation of *X* and *Y* is $r_{X,Y} = E[XY]$
- Identities

$$Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$$
$$Cov[X,Y] = r_{X,Y} - E[X]E[Y]$$
$$X = Y \Rightarrow Cov[X,Y] = Var[X], r_{X,Y} = E[X^2]$$

EC381/MN308 – 2007/2008 19

Covariance Matrix (2-Vectors)

- Form vector $\underline{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$
- Expected value $E[\underline{X}] = \begin{pmatrix} E[X] \\ E[Y] \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$
- Covariance matrix

$$\Sigma_{\underline{X}} = E[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T]$$
$$= \begin{pmatrix} \sigma_X^2 & Cov[X,Y] \\ Cov[X,Y] & \sigma_Y^2 \end{pmatrix}$$

EC381/MN308 – 2007/2008 20

Key Statistics

- Expected Value

$$E(\underline{X}) = E\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} \equiv \underline{\mu}_X$$

- Covariance Matrix for arbitrary vectors:

$$\Sigma_{\underline{x}} = E[(\underline{X} - \underline{\mu}_X)(\underline{X} - \underline{\mu}_X)^T]$$
$$= E\Big[\begin{pmatrix} X_1 - \mu_{X_1} \\ \vdots \\ X_n - \mu_{X_n} \end{pmatrix} \begin{pmatrix} X_1 - \mu_{X_1} & \cdots & X_n - \mu_{X_n} \end{pmatrix}\Big]$$

EC381/MN308 – 2007/2008 21

More Statistics

- Correlation matrix for arbitrary random vectors:

$$R_{\underline{x}} = E[(\underline{X})(\underline{X})^T]$$
$$= E\Big[\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \begin{pmatrix} X_1 & \cdots & X_n \end{pmatrix}\Big]$$
$$= \begin{pmatrix} E(X_1)^2 & E[X_1X_2] & \cdots & E[X_1X_n] \\ E[(X_2X_1] & E[X_2^2] & \cdots & E[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \cdots & E[X_n^2] \end{pmatrix}$$

EC381/MN308 – 2007/2008 22

Covariance Matrix

Expressed entirely in terms of pairs of components of **X**

$$\Sigma_{\underline{X}} = E[(\underline{X} - \underline{\mu}_X)(\underline{X} - \underline{\mu}_X)^T]$$
$$= E\Big[\begin{pmatrix} (X_1 - \mu_{X_1})^2 & (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) & \cdots & (X_1 - \mu_{X_1})(X_n - \mu_{X_n}) \\ (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) & (X_2 - \mu_{X_2})^2 & \cdots & (X_n - \mu_{X_n})(X_2 - \mu_{X_2}) \\ \vdots & \vdots & \ddots & \vdots \\ (X_1 - \mu_{X_1})(X_n - \mu_{X_n}) & (X_n - \mu_{X_n})(X_2 - \mu_{X_2}) & \cdots & (X_n - \mu_{X_n})^2 \end{pmatrix}\Big]$$
$$= \begin{pmatrix} E(X_1 - \mu_{X_1})^2 & E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] & \cdots & E[(X_1 - \mu_{X_1})(X_n - \mu_{X_n})] \\ E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] & E[(X_2 - \mu_{X_2})^2] & \cdots & E[(X_n - \mu_{X_n})(X_2 - \mu_{X_2})] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_1 - \mu_{X_1})(X_n - \mu_{X_n})] & E[(X_n - \mu_{X_n})(X_2 - \mu_{X_2})] & \cdots & E(X_n - \mu_{X_n})^2] \end{pmatrix}$$
$$= \begin{pmatrix} \sigma_{X_1}^2 & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & \sigma_{X_2}^2 & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & \sigma_{X_n}^2 \end{pmatrix}$$

EC381/MN308 – 2007/2008 23

Properties of Covariance Matrix

- Covariance matrix $\Sigma_{\mathbf{X}}$:
  - Symmetric matrix
  - Positive semi-definite: For any nonzero vector $\underline{a}$,

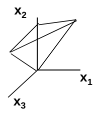
$$\underline{a}^T \Sigma_{\underline{X}} \underline{a} \geq 0$$

  - Has all eigenvalues real-valued, non-negative
  - Has a complete set of distinct eigenvectors (n of them)

EC381/MN308 – 2007/2008 24

Example

- Quiz 5.6: **X** with PDF



$$f_{\underline{X}}(\underline{x}) = \begin{cases} 6 & 0 \le x_1 \le x_2 \le x_3 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X_1] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_1 dx_1 dx_2 dx_3 = 0.25$$
$$E[X_2] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_2 dx_1 dx_2 dx_3 = 0.5$$
$$E[X_3] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_3 dx_1 dx_2 dx_3 = 0.75$$
$$E[X_1X_2] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_1x_2 dx_1 dx_2 dx_3 = 0.15$$
$$E[X_1X_3] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_1x_3 dx_1 dx_2 dx_3 = 0.20$$
$$E[X_1^2] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_1^2 dx_1 dx_2 dx_3 = 0.10$$

Example 2

- Correlation and Covariance

$$E[X_2^2] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_2^2 dx_1 dx_2 dx_3 = 0.3$$
$$E[X_2X_3] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_2x_3 dx_1 dx_2 dx_3 = 0.40$$
$$E[X_3^2] = \int_0^1 \int_0^{x_3} \int_0^{x_2} 6x_1^2 dx_1 dx_2 dx_3 = 0.6$$
$$R_{\underline{X}} = \begin{pmatrix} .1 & .15 & .2 \\ .15 & .3 & .4 \\ .2 & .4 & .6 \end{pmatrix}$$
$$\Sigma_{\underline{X}} = \begin{pmatrix} .1 & .15 & .2 \\ .15 & .3 & .4 \\ .2 & .4 & .6 \end{pmatrix} - \begin{pmatrix} 0.25 \\ 0.5 \\ 0.75 \end{pmatrix} \cdot \begin{pmatrix} 0.25 & 0.5 & 0.75 \end{pmatrix}$$

Linear Transformations

- **Y = AX + b**, for known *m*-by-*n* matrix **A**, *m*-vector **b**
- *E*[**Y**] = *E*[**AX + b**] = **A** *E*[**X**] + **b**
  - Expectations are linear operations!
- Covariance of **Y**: $\Sigma_Y = A\,\Sigma_X A^T$

$$\begin{aligned}\Sigma_{\underline{Y}} &= E[(\underline{Y} - E[\underline{Y}])(\underline{Y} - E[\underline{Y}])^T] \\ &= E[(A\underline{X} + \underline{b} - AE[\underline{X}] - \underline{b})(A\underline{X} + \underline{b} - AE[\underline{X}] - \underline{b})^T] \\ &= E[A(\underline{X} - E[\underline{X}])(A(\underline{X} - E[\underline{X}]))^T] \\ &= E[A(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T A^T] \\ &= AE[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T]A^T = A\Sigma_{\underline{X}}A^T\end{aligned}$$

Functions of Random Vectors

- Given **X**, can define a derived random variable *W* = *g*(**X**) or a random vector **Y** = *g*(**X**)
  - Dimension of **Y** can be smaller or greater than dimension of **X**
  - Implicit distribution on **Y** defined by map *g*()
- Can take expectations

$$\begin{aligned}E[W] &= \int_{\underline{x}} g(\underline{x}) f_{\underline{X}}(\underline{x}) d\underline{x} \quad \text{(continuous)} \\ &= \sum_{\underline{x}} g(\underline{x}) P_{\underline{X}}(\underline{x}) \quad \text{(discrete)}\end{aligned}$$

Special Case (Theorem 5.11)

- If **A** is an invertible matrix ➔ 1-1 correspondence between **X**, **Y** so that **Y = AX + b** can be written
  X = A^{-1}(**Y** – **b**)
- Change of variable formula yields PDF and CDF:

$$f_{\underline{Y}}(\underline{y}) = \frac{1}{|\det A|} f_{\underline{X}}(A^{-1}(\underline{Y} - \underline{b}))$$

$$\begin{aligned}F_{\underline{Y}}(\underline{y}) = P(\underline{Y} \le \underline{y}) &= \int_{\underline{x}:A\underline{x}+\underline{b}\le\underline{y}} f_{\underline{X}}(\underline{x}) d\underline{x} \\ &= \int_{\underline{y}' \le \underline{y}} \frac{1}{|\det A|} f_{\underline{X}}(A^{-1}(\underline{y}' - \underline{b})) d\underline{y}'\end{aligned}$$

Special Classes of Functions

- Special case: **X** is a random vector of n i.i.d. random variables $X_1, \ldots, X_n$
  - CDF of X_i is $F_X(x)$
- Define $Z = \max_i \{X_1, \ldots, X_n\}$. Find $F_Z(z)$

Hint: $P(Z \le z) \equiv P(X_1 \le z, \ldots, X_n \le z)$

$$\begin{aligned}F_Z(z) \equiv P(Z \le z) &= P(X_1 \le z, \ldots, X_n \le z) \\ &= P(X_1 \le z) \cdots P(X_n \le z) \quad \text{independence} \\ &= F_X(z) \cdots F_x(z) \quad \text{identically distributed} \\ &= F_X(z)^n\end{aligned}$$

Another Special Case

- Define $Z = \min_i \{X_1, \ldots, X_n\}$. Find $F_Z(z)$

Hint: $P(Z \geq z) \equiv P(X_1 \geq z, \ldots, X_n \geq z)$

$$\begin{aligned} F_Z(z) &\equiv 1 - P(Z \geq z) = 1 - P(X_1 \geq z, \ldots, X_n \geq z) \\ &= 1 - P(X_1 \geq z) \cdots P(X_n \geq z) \text{ independence} \\ &= 1 - (1 - F_X(z)) \cdots (1 - F_x(z)) \text{ identically distributed} \\ &= 1 - (1 - F_X(z))^n \end{aligned}$$

Example

- Quiz 5.5
  - Testing light bulbs yields 3 outcomes: *G(ood)*, *A(verage)*, *B(ad)*
  - Each light bulb has *P*[*G*] = 0.3, *P*[*A*] = 0.5, *P*[*B*] = 0.2, independently
  - Experiment: test 4 light bulbs.
    RV X_1 = # of *G*, X_2 = # of *A*, X_3 = # of *B*
  - Find PMF of $\underline{X}$, marginal PMFs of X_i, and PMF of $W = \max_i(X_i)$

$$\text{Observe}: X_1 + X_2 + X_3 = 4, X_i \in \{0, \ldots, 4\}$$
$$P(X_1 = j, X_2 = k) = \binom{4}{k}\binom{4-k}{j} 0.3^j 0.5^k 0.2^{4-k-j}$$

Notation

- Can work with pairs of random vectors **X**, **Y**
  - Extend notation for pairs of random variables
- Joint PMF for pairs of discrete random vectors

$$P_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) \equiv P(\underline{X} = \underline{x}, \underline{Y} = \underline{y})$$

- Joint PDF for pairs of jointly continuous random vectors

$$f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})\Delta_{\underline{x}}\Delta_{\underline{y}} \equiv P[\underline{X} \in (\underline{x}, \underline{x} + \Delta_{\underline{x}}), \underline{Y} \in (\underline{y}, \underline{y} + \Delta_{\underline{y}})])$$

Statistics for Pairs of Random Vectors

- **X**, **Y** random vectors
- Vector Cross-Correlation

$$R_{\underline{X}\underline{Y}} = E[\underline{X} \cdot \underline{Y}^T] = R^T_{\underline{Y}\underline{X}}$$

- Vector Cross-Covariance

$$\Sigma_{\underline{X}\underline{Y}} = E[(\underline{X} - \underline{\mu}_X)(\underline{Y} - \underline{\mu}_Y)^T] = \Sigma^T_{\underline{Y}\underline{X}}$$

- Identity

$$\Sigma_{\underline{X}\underline{Y}} = R_{\underline{X}\underline{Y}} - \underline{\mu}_X \cdot \underline{\mu}^T_Y$$

Independence of Random Vectors

- Two random vectors **X**, **Y** are said to be independent if

$$P_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) \equiv P_{\underline{X}}(\underline{x})P_{\underline{Y}}(\underline{y}) \text{ (discrete)}$$
$$f_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) \equiv f_{\underline{X}}(\underline{x})f_{\underline{Y}}(\underline{y}) \text{ (continuous)}$$

➔ Any component of **X** is independent of any component of **Y**

Sums of independent random variables

- Recall *X*, *Y* joint dependent RVs, and *Z* = *X* + *Y* ➔

$$P_Z(z) = \sum_{x \in S_X, z-x \in S_Y} P_{X,Y}(x, z-x) \text{ (discrete)}$$
$$f_Z(z) = \int_{x \in S_x, z-x \in S_y} f_{X,Y}(x, z-x)dx \text{ (continuous)}$$

- If *X*, *Y* independent ➔

$$P_Z(z) = \sum_{x \in S_X, z-x \in S_Y} P_X(x)P_Y(z-x) \text{ (discrete)}$$
$$f_Z(z) = \int_{x \in S_x, z-x \in S_y} f_X(x)f_Y(z-x)dx \text{ (continuous)}$$

- Convolution! PMF and PDF of sum of 2 independent RVs is convolution of their individual PMFs or PDFs
  - Generalizes to n independent RVs by induction

Conditional Probability for Random Vectors

- Discrete Random Vectors: PMF of **X** given **Y** = y

$$P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{P_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{P_{\underline{Y}}(\underline{y})}$$

- Total Probability Theorem:

$$\begin{aligned} P_{\underline{X}}(\underline{x}) &= \sum_{\underline{y}} P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) P_{\underline{Y}}(\underline{y}) \\ &= \sum_{\underline{y}} \frac{P_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{P_{\underline{Y}}(\underline{y})} P_{\underline{Y}}(\underline{y}) \\ &= \sum_{\underline{y}} P_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) \end{aligned}$$

Conditional Probability - 2

- Bayes' Rule: Discrete Random Vectors

$$\begin{aligned} P_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) &= \frac{P_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{P_{\underline{Y}}(\underline{y})} \\ &= \frac{P_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) P_{\underline{X}}(\underline{x})}{P_{\underline{Y}}(\underline{y})} \\ &= \frac{P_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) P_{\underline{X}}(\underline{x})}{\sum_{\underline{x}'} P_{\underline{X},\underline{Y}}(\underline{x}',\underline{y})} \end{aligned}$$

Conditional Probability - 3

- Continuous Random Vectors: PDF of **X** given **Y** = y

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{f_{\underline{Y}}(\underline{y})}$$

- Total Probability Theorem:

$$\begin{aligned} f_{\underline{X}}(\underline{x}) &= \int_{\underline{y}} f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \int_{\underline{y}} \frac{f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{f_{\underline{Y}}(\underline{y})} f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \int_{\underline{y}} f_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) d\underline{y} \end{aligned}$$

Conditional Probability - 4

- Bayes' Rule: Continuous Random Vectors

$$\begin{aligned} f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) &= \frac{f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{f_{\underline{Y}}(\underline{y})} \\ &= \frac{f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) f_{\underline{X}}(\underline{x})}{f_{\underline{Y}}(\underline{y})} \\ &= \frac{f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) f_{\underline{X}}(\underline{x})}{\int_{\underline{x}'} f_{\underline{X},\underline{Y}}(\underline{x}',\underline{y}) d\underline{x}'} \end{aligned}$$

Gaussian Random Vectors

- n random variables **X** = $[X_1, \ldots, X_n]^T$ are jointly continuous Gaussian if their joint PDF is

$$f_{\underline{X}}(\underline{x}) = Ce^{-Q(\underline{x}-\underline{\mu}_X)}$$

  – Constant: $C = \frac{1}{(2\pi)^{n/2} \det(\Sigma_{\underline{X}})^{1/2}}$

  – Quadratic exponent: $Q(\underline{x}) = \frac{1}{2}\underline{x}^T \Sigma_{\underline{X}}^{-1} \underline{x}$

Properties of Gaussian Random Vectors

- If **X** is a Gaussian random vector, and **A** is a known *m*-by-*n* matrix, and **b** is a known *m*-vector,
  ➔**Y** = **AX** + **b** is a Gaussian random vector

  – New mean: $\underline{\mu}_Y = A\underline{\mu}_X + \underline{b}$

  – New Covariance: $\Sigma_{\underline{Y}} = A\Sigma_{\underline{X}} A^T$

- Notation: $\underline{Y} \sim N(\underline{\mu}_Y, \Sigma_{\underline{Y}})$

Pairs of Gaussian Random Vectors

- **X**, **Y** jointly Gaussian random vectors
- **Y** Notation: Cross-Covariance Matrix $\Sigma_{X,Y}$

$$\Sigma_{\underline{X},\underline{Y}} = E[(\underline{X}-\underline{\mu}_X)(\underline{Y}-\underline{\mu}_Y)^T] \\ = \Sigma^T_{\underline{Y},\underline{X}}$$

- Joint Covariance
  - Joint Vector $\underline{Z} = \begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix}$
  - Joint Covariance $\Sigma_{\underline{Z}} = \begin{pmatrix} \Sigma_{\underline{X}} & \Sigma_{\underline{XY}} \\ \Sigma_{\underline{YX}} & \Sigma_{\underline{Y}} \end{pmatrix}$

Conditional Probability for Gaussian Random Vectors

- PDF of **X** given **Y** = y is also Gaussian!

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{f_{\underline{Y}}(\underline{y})} \\ = \frac{f_{\underline{Z}}(\underline{z})}{f_{\underline{Y}}(\underline{y})}$$

  - Numerator is exponent of negative quadratic in **x**, **y**
  - Denominator is exponent of negative quadratic in **y**
  - ➔ Ratio is exponent of negative quadratic in **x** !!!

Conditional Probability: Gaussians

- $$E[\underline{X}|\underline{Y}] = \underline{\mu}_X + \Sigma_{\underline{XY}}\Sigma_{\underline{Y}}^{-1}(\underline{y}-\underline{\mu}_Y)$$
  - Scalar case: $E[X|Y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y) = \mu_X + \frac{Cov[X,Y]}{\sigma_Y^2}(y-\mu_Y)$
- Covariance of **X** given **Y** = **y**:

$$\Sigma_{\underline{X}|\underline{Y}} = \Sigma_{\underline{X}} - \Sigma_{\underline{XY}}\Sigma_{\underline{Y}}^{-1}\Sigma_{\underline{YX}}$$

Example - 1

- Quiz 5.7 **Z** is 2-D standard Normal (pair of independent, 0-mean, 1 variance RVs)
  - $X_1 = 2\,Z_1 + Z_2 + 2$; $X_2 = Z_1 - Z_2$
  - Calculate mean and variance:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \qquad b = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$E[\underline{X}] = AE[\underline{Z}] + \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\Sigma_{\underline{X}} = A\Sigma_{\underline{Z}}A^T = AA^T = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$$

Example - 2

- Compute cross-covariance between **Z** and X_1:

$$\Sigma_{\underline{Z}X_1} = E[\underline{Z}(X_1-2)] = \begin{pmatrix} E[Z_1(2Z_1+Z_2)] \\ E[Z_2(2Z_1+Z_2)] \end{pmatrix} \\ = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ \{independence, standard normal\}}$$

- Compute E[**Z**|X_1=x_1]

$$E[\underline{Z}|X_1 = x_1] = \underline{\mu}_Z + \Sigma_{\underline{Z}X_1}\Sigma_{X_1}^{-1}(x_1-2) \\ = \begin{pmatrix} \frac{2}{5}(x_1-2) \\ \frac{1}{5}(x_1-2) \end{pmatrix}$$

Example – 3

- Compute covariance of estimation error: Covariance of **Z** given X_1

$$\Sigma_{\underline{Z}|X_1} = \Sigma_{\underline{Z}} - \Sigma_{\underline{Z}X_1}\Sigma_{X_1}^{-1}\Sigma_{X_1\underline{Z}} \\ = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 2 & 1 \end{pmatrix} \\ = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{pmatrix}$$

- Eigenvalues 0, 1 (nonnegative: it is a covariance)
  - Not invertible ➔ conditional density of **Z** given X_1 is not jointly continuous
  - **Z** has 2 degrees of freedom in uncertainty. Observing 1 of them reduces it to 1 degree of freedom… Hence, one zero eigenvalue