# Multiple Random Variables

Henry D. Pfister

ECE 581 Random Signals and Noise

Lecture 5

Slides courtesy of Christ Richmond with slight modifications.

# Table of contents

# Remarks

- Up to now we've considered experiment outcomes that can be modeled as a single random variable, e.g. $X$, for both discrete and continuous events.
- Sometimes, experiment outcomes occur in $n$-tuples, e.g.
  - in pairs such as $(X, Y)$, e.g. two rolls of a die;
  - or in triplets such as $(X, Y, Z)$, e.g. position of ball #11 bouncing in a powerball machine;
  - or in quadruplets such as $(X, Y, W, Z)$;
  - or in $n$-dimensions such as $(X_1, X_2, \ldots, X_n)$.
- We need tools of probability theory to handle such multivariable observations, i.e. it is desired to extend the concepts of PMF / PDF, CDF, and expectations to multiple random variables.

# Joint Probability Mass Function (PMF)

- The joint probability mass function of paired discrete r.v.'s $(X, Y)$ is defined as

$$P_{X,Y}(x, y) = \Pr\left(\{X = x\} \cap \{Y = y\}\right) = \Pr(X = x, Y = y).$$

  - r.v.'s are indicated by subscript "$X, Y$" and hypothetical values dnoted by arguments "$x, y$".
  - Joint PMF specifies the probability of possible values of pairs $(X, Y)$

- The range of possible values pairs $(X, Y)$ can assume is denoted $S_{X,Y}$, i.e.

$$S_{X,Y} = \{(x, y) | P_{X,Y}(x, y) > 0\},$$ sometimes called the support of $P_{X,Y}$.

- By the axioms, the total probability of all possible pairs $(X, Y)$ is unity, i.e.

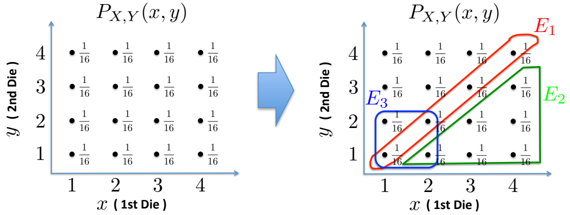
$$\sum_{(x,y)\in S_{X,Y}} P_{X,Y}(x, y) = \sum_{x\in S_X} \sum_{y\in S_Y} P_{X,Y}(x, y) = 1$$

  and $P_{X,Y}(x, y) \geq 0$ for all $(x, y)$.

- The joint PMF completely characterizes the probability of pairs $(X, Y)$ including any relationship, coupling, or statistical dependencies between $X$ and $Y$.

# Example: Rolling a Pair of Distinguishable “Fair” Four-Sided Dice

- Let the r.v.’s $(X, Y)$ denote values shown on the pair of dice, where $S_{X,Y} = \{(x, y) | x \in \{1, 2, 3, 4\},\ y \in \{1, 2, 3, 4\}\}$ and joint PMF $P_{X,Y}$:




  - The probability of an “event” is obtained by summing the probabilities of each outcome belonging to that event, i.e. event $\{(X, Y) \in B\}$ has probability

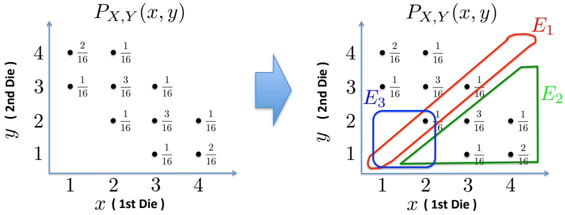
$$\Pr[B] = \sum_{(x,y)\in B} P_{X,Y}(x, y).$$

  - For example, consider events $E_1 = \{X = Y\}$, $E_2 = \{X > Y\}$, $E_3 = \{X \leq 2, Y \leq 2\}$; from above image on the right it is clear that

$$\Pr(E_1) = \frac{4}{16},\quad \Pr(E_2) = \frac{6}{16},\ \text{and}\ \Pr(E_3) = \frac{4}{16}. \tag{1}$$

# Example: Rolling of Pair of Magical Four-Sided Dice

- Consider rolling a magical pair of four-sided dice whose joint PMF $P_{X,Y}$ is:




  - These dice are unfair (i.e., not all outcomes have equal probability) and also dependent!
    - This is an unusual pair of dice since outcomes $(1,1),(1,2)$ and $(2,1)$ are impossible, and rolls where $Y = 5 - X$ are most likely.
  - The same events $E_1 = \{X = Y\}$, $E_2 = \{X > Y\}$, $E_3 = \{X \leq 2, Y \leq 2\}$ now yield

$$\Pr(E_1) = \frac{2}{16}, \quad \Pr(E_2) = \frac{7}{16}, \text{ and } \Pr(E_3) = \frac{1}{16}. \tag{2}$$

# Marginal PMFs Obtainable from Joint PMF

- The previous examples demonstrate that by summing probabilities $P_{X,Y}$ for all outcomes corresponding to $\{X = x\}$ for each $x \in S_X$, and by summing probabilities $P_{X,Y}$ for all outcomes corresponding to $\{Y = y\}$ for each $y \in S_Y$ it follows that

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x,y), \quad \text{for each } x \in S_X$$
$$P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x,y), \quad \text{for each } y \in S_Y.$$

- PMFs $P_X$ and $P_Y$ are sometimes referred to as the marginal PMFs to distinguish them from the joint PMF $P_{X,Y}$.

# Quiz 5.2

The joint PMF $P_{Q,G}(q,g)$ for random variables $Q$ and $G$ is given in the following table:

| $P_{Q,G}(q,g)$ | $g=0$ | $g=1$ | $g=2$ | $g=3$ |
|---|---|---|---|---|
| $q=0$ | 0.06 | 0.18 | 0.24 | 0.12 |
| $q=1$ | 0.04 | 0.12 | 0.16 | 0.08 |

Calculate the following probabilities:

(a) $\mathrm{P}[Q=0]$

(b) $\mathrm{P}[Q=G]$

(c) $\mathrm{P}[G>1]$

(d) $\mathrm{P}[G>Q]$

# Quiz 5.3

The probability mass function $P_{H,B}(h,b)$ for the two random variables $H$ and $B$ is given in the following table. Find the marginal PMFs $P_H(h)$ and $P_B(b)$.

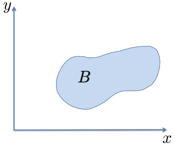
| $P_{H,B}(h,b)$ | $b=0$ | $b=2$ | $b=4$ |
|---|---|---|---|
| $h=-1$ | 0 | 0.4 | 0.2 |
| $h=0$ | 0.1 | 0 | 0.1 |
| $h=1$ | 0.1 | 0.1 | 0 |

(1)

# Joint Probability Density Function (PDF)

- The joint probability density function of paired continuous r.v.'s $(X, Y)$ is a function $f_{X,Y}(x, y)$ defined such that for any subset $B \subseteq S_{X,Y}$

$$\Pr[(X, Y) \in B] = \iint_{(x,y)\in B} f_{X,Y}(x, y)dxdy$$




  - Notation: subscript "$X, Y$" as r.v.'s, and arguments "$x, y$" as hypothetical values.

- The range of possible values pairs $(X, Y)$ can assume is denoted $S_{X,Y}$, i.e.

$$S_{X,Y} = \{(x, y) | f_{X,Y}(x, y) > 0\},$$ called the support of $f_{X,Y}$.

- By axioms of probability it follows that $f_{X,Y}(x, y) \geq 0$ for all $(x, y)$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1.$$

- The joint PDF completely characterizes the probability of pairs $(X, Y)$ including any relationship, coupling, or statistical dependencies between $X$ and $Y$.

# Joint PDF

- As an example, if $B = \{(x, y) | a \le x \le b,\ c \le y \le d\}$ i.e. set $B$ is a rectangular region in the $x, y$-plane, then

$$\Pr(a \le X \le b,\ c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy.$$

- Consider that

$$\Pr(a \le X \le a + \delta_x,\ c \le Y \le c + \delta_y) = \int_a^{a+\delta_x} \int_c^{c+\delta_y} f_{X,Y}(x, y) dx dy \simeq f_{X,Y}(a, c) \cdot \delta_x \cdot \delta_y \quad (3)$$

  where approximation improves as $\delta_x \to 0$, $\delta_y \to 0$.
  - Note $\delta_x \delta_y =$ (length) $\times$ (length) is a differential area.
  - Thus, joint PDF $f_{X,Y}$ provides a measure of the "probability per unit area."

# Marginal PDFs Obtainable from Joint PDF

- Note that for all sets $A \in S_X$ and $C \in S_Y$:

$$\Pr(X \in A) = \int_{x \in A} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = \int_{x \in A} f_X(x)dx$$
$$\implies \int_{x \in A} \left[ \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy - f_X(x) \right] dx = 0$$
$$\Pr(Y \in C) = \int_{y \in C} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_{y \in C} f_Y(y)dy$$
$$\implies \int_{y \in C} \left[ \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx - f_Y(y) \right] dy = 0$$

- Since the difference integral equal zero for arbitrary densities $f_{X,Y}$, $f_X$, and $f_Y$, the integrands in brackets $[\cdot]$ must be zero, i.e. the marginal PDFs $f_X$ and $f_Y$ are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

# Example: 2-D Uniform PDF

- Consider paired r.v.'s $(X, Y)$ having joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \gamma, & a \le x \le b, \quad c \le y \le d \\ 0, & \text{Otherwise} \end{cases} \tag{4}$$

Note that $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy \implies \gamma = \dfrac{1}{(b-a)(d-c)}$.

- The marginal PDF $f_X(x)$ for $a \le x \le b$ follows from

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{1}{(b-a)(d-c)}\int_c^d dy = \frac{1}{(b-a)(d-c)} \cdot \left[ y\Big|_c^d \right]$$
$$= \frac{1}{(b-a)(d-c)} \cdot (d-c) = \frac{1}{b-a}.$$

- The marginal PDF $f_Y(y)$ for $c \le y \le d$ follows from

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{1}{(b-a)(d-c)}\int_a^b dx = \frac{1}{(b-a)(d-c)} \cdot \left[ x\Big|_a^b \right]$$
$$= \frac{1}{(b-a)(d-c)} \cdot (b-a) = \frac{1}{d-c}.$$

# Example: More 2-D Uniform PDF

- For $a = 0, b = 1, c = 0$, and $d = 1$ in the previous example, we get

$$f_{X,Y}(x,y) = \begin{cases} 1, & 0 \le x \le 1, \ \ 0 \le y \le 1 \\ 0, & \text{Otherwise} \end{cases} \tag{5}$$

- Consider events $E_1 = \{X \le \frac{1}{2}, Y > \frac{3}{4}\}$, $E_2 = \{X > Y\}$:

$$\Pr(E_1) = \Pr(X \le \tfrac{1}{2}, Y > \tfrac{3}{4}) = \int_0^{\frac{1}{2}} dx \int_{\frac{3}{4}}^1 dy = \left[ x|_0^{\frac{1}{2}} \right] \cdot \left[ y|_{\frac{3}{4}}^1 \right] = \left( \frac{1}{2} - 0 \right) \cdot \left( 1 - \frac{3}{4} \right) = \frac{1}{8}$$

$$\begin{aligned} \Pr(E_2) = \Pr(X > Y) &= \int_0^1 dx \int_0^x dy = \int_0^1 dx \left[ y|_0^x \right] = \int_0^1 (x-0)dx \\ &= \frac{1}{2} x^2 \Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}. \end{aligned} \tag{6}$$

- Try computing probability for event $\{X^2 \le Y\}$, and show that it is $\frac{2}{3}$.

# Quiz 5.4

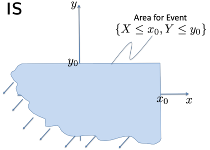
The joint probability density function of random variables $X$ and $Y$ is

$$f_{X,Y}(x,y) = \begin{cases} cxy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Find the constant $c$. What is the probability of the event $A = X^2 + Y^2 \le 1$?

# Joint Cumulative Distribution Function (CDF)

- The joint cumulative distribution function of paired r.v.'s $(X, Y)$ is




$$\begin{aligned} F_{X,Y}(x,y) &= \Pr(\{X \le x\} \cap \{Y \le y\}) \\ &= \Pr(X \le x, Y \le y) \end{aligned}$$

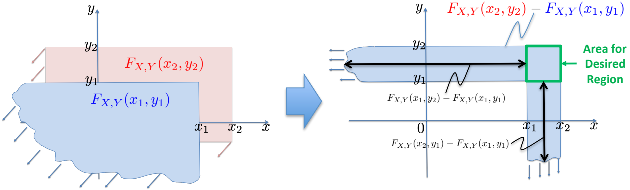
This definition applies to both discrete and continuous r.v.'s, but results in:

$$\begin{aligned} F_{X,Y}(x,y) &= \sum_{\substack{a \in S_X: \\ a \le x}} \sum_{\substack{b \in S_Y: \\ b \le y}} P_{X,Y}(a,b) \quad (\text{discrete r.v.}) \\ F_{X,Y}(x,y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv \quad (\text{continuous r.v.}) \end{aligned} \tag{7}$$

- The following properties of the CDF hold by inspection:
  - $0 \le F_{X,Y}(x,y) \le 1$, $F_{X,Y}(\infty,\infty) = 1$
  - $F_X(x) = \Pr(X \le x) = F_{X,Y}(x,\infty)$, $F_Y(y) = \Pr(Y \le y) = F_{X,Y}(\infty,y)$
  - $F_{X,Y}(x,-\infty) = 0$, $F_{X,Y}(-\infty,y) = 0$
  - If $x_1 \le x_2$ and $y_1 \le y_2$, then $F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$

# Joint Cumulative Distribution Function (CDF)

$$\begin{aligned}\Pr(x_1 < X \le x_2, y_1 < Y \le y_2) &= F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_1) \\ &\quad - [F_{X,Y}(x_1,y_2) - F_{X,Y}(x_1,y_1)] - [F_{X,Y}(x_2,y_1) - F_{X,Y}(x_1,y_1)] \\ &= F_{X,Y}(x_2,y_2) + F_{X,Y}(x_1,y_1) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1)\end{aligned}$$




- Note: Some probabilities are more easily obtained using the PMF / PDF, rather than the CDF.

- For continuous r.v.'s the fundamental theorem of calculus says that the joint PDF is given by

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv \implies f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

# Example: 2-D Uniform Distribution

- **Example:** Let $(X, Y)$ have joint PDF given in (4). The joint CDF follows from

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv = \begin{cases} 0, & x < a, \text{ or } y < c \\ \frac{1}{(b-a)(d-c)}\int_a^x\int_c^y dudv, & a \le x \le b,\ c \le y \le d \\ \frac{1}{(b-a)(d-c)}\int_a^x\int_c^d dudv, & a \le x \le b, y > d \\ \frac{1}{(b-a)(d-c)}\int_a^b\int_c^y dudv, & x > b,\ c \le y \le d \\ 1, & x > b,\ y > d. \end{cases}$$

- This simple integration leads to

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)dxdy = \begin{cases} 0, & x < a, \text{ or } y < c \\ \frac{(x-a)(y-c)}{(b-a)(d-c)}, & a \le x \le b,\ c \le y \le d \\ \frac{x-a}{b-a}, & a \le x \le b, y > d \\ \frac{y-c}{d-c}, & x > b,\ c \le y \le d \\ 1, & x > b,\ y > d. \end{cases}$$

# Independent Random Variables

- Recall that events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if and only if

$$\Pr(\mathcal{AB}) = \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}). \tag{8}$$

  - Recall that $\Pr(\mathcal{AB}) = \Pr(\mathcal{A})\Pr(\mathcal{B}|\mathcal{A}) = \Pr(\mathcal{B})\Pr(\mathcal{A}|\mathcal{B})$.
  - Thus, if events $\mathcal{A}, \mathcal{B}$ are independent, then $\Pr(\mathcal{B}|\mathcal{A}) = \Pr(\mathcal{B})$ and $\Pr(\mathcal{A}|\mathcal{B}) = \Pr(\mathcal{A})$.
- Consider paired discrete r.v.'s $(X, Y) \sim P_{X,Y}$ and events $\mathcal{A} = \{X = x\}$ and $\mathcal{B} = \{Y = y\}$.
  - If $\mathcal{A}, \mathcal{B}$ are independent, then

$$\begin{aligned} \Pr(\mathcal{AB}) &= \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}) \\ \Pr(\{X = x\} \cap \{Y = y\}) &= \Pr(\{X = x\}) \cdot \Pr(\{Y = y\}) \\ P_{X,Y}(x, y) &= P_X(x) \cdot P_Y(y) \end{aligned}$$

# Independent Random Variables

- For continuous r.v.'s $(X, Y) \sim f_{X,Y}$ and events $\mathcal{A} = \{x \mid x \in A\}$ and $\mathcal{B} = \{y \mid y \in B\}$,
  - If $\mathcal{A}, \mathcal{B}$ are independent, then

$$\begin{aligned}
\Pr(\mathcal{A}\mathcal{B}) &= \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}) \\
\Pr(\{x \in A\} \cap \{y \in B\}) &= \Pr(\{x \in A\}) \cdot \Pr(\{y \in B\}) \\
\int_{x \in A} \int_{y \in B} f_{X,Y}(x,y) dxdy &= \int_{x \in A} f_X(x) dx \cdot \int_{y \in B} f_Y(y) dy \Longrightarrow
\end{aligned}$$

$$\int_{x \in A} \int_{y \in B} \left[ f_{X,Y}(x,y) - f_X(x) \cdot f_Y(y) \right] dxdy = 0 \Longrightarrow$$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

- Thus, paired r.v.'s $(X, Y)$ are said to be independent if and only if

$$\begin{aligned}
P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y) \qquad &(\text{discrete r.v.'s}) \\
f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \qquad &(\text{continuous r.v.'s})
\end{aligned} \tag{9}$$

- To determine if paired r.v.'s $(X, Y)$ are independent one can test if condition (9) holds.

# Example: Roll of Pair of Four-Sided Die

- Recall this discrete r.v. example discussed earlier that resulted in equations (1) and (2) and resulted in the corresponding marginal PMFs specified below these equations.
  - The marginal PMFs $P_X$ and $P_Y$ below (1) show that $P_X(x) \cdot P_Y(y) = \frac{1}{16}$ for all $(x, y) \in S_{X,Y}$.
  - The Figure preceding (1) shows that $P_X(x) \cdot P_Y(y) = P_X(x, y)$.
    - Thus, the paired discrete r.v. $(X, Y)$ whose joint PMF resulted in (1) are independent.
  - The marginal PMFs $P_X$ and $P_Y$ below (2), however, have a product that is not equal to the joint PMF illustrated in the Figure preceding (2), i.e. $P_X(x) \cdot P_Y(y) \neq P_{X,Y}(x, y)$.
    - Thus, the paired discrete r.v. $(X, Y)$ whose joint PMF resulted in (2) are not independent.

# Expectation, Mean, Covariance, and Correlation

- Given paired r.v.'s $(X, Y) \sim P_{X,Y}$ or $f_{X,Y}$ the expected value of function $g(X, Y)$ is

$$E[g(X,Y)] = \sum_{(x,y)\in S_{X,Y}} g(x,y)\cdot P_{X,Y}(x,y)$$
$$= \sum_{x\in S_X} \sum_{y\in S_Y} g(x,y)\cdot P_{X,Y}(x,y) \quad \text{(discrete r.v.'s)}$$

$$E[g(X,Y)] = \iint_{(x,y)\in S_{X,Y}} g(x,y)\cdot f_{X,Y}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\cdot f_{X,Y}(x,y)\,dx\,dy \quad \text{(continuous r.v.'s)}$$

- This is an incredibly useful result.
- Indeed, the average of the r.v. $W = g(X, Y)$ can be determined without knowing PMF $P_W(w)$ (discrete case), or PDF $f_W(w)$ (continuous case).

# Expectation: Linear Combination of Functions

- Given paired r.v.'s $(X, Y) \sim P_{X,Y}$ (or $f_{X,Y}$), the expected value of function $g(X,Y) = a_1 g_1(X,Y) + a_2 g_2(X,Y) + \cdots + a_n g_n(X,Y)$ is:

$$\begin{aligned}
E[g(X,Y)] &= \sum_{(x,y)\in S_{X,Y}} g(x,y) \cdot P_{X,Y}(x,y) \\
&= \sum_{(x,y)\in S_{X,Y}} \Big[a_1 g_1(x,y) + a_2 g_2(x,y) + \cdots + a_n g_n(x,y)\Big] \cdot P_{X,Y}(x,y) \\
&= a_1 \sum_{(x,y)\in S_{X,Y}} g_1(x,y) P_{X,Y}(x,y) + \cdots + a_n \sum_{(x,y)\in S_{X,Y}} g_n(x,y) P_{X,Y}(x,y) \\
&= a_1 E[g_1(X,Y)] + a_2 E[g_2(X,Y)] + \cdots + a_n E[g_n(X,Y)]
\end{aligned}$$

- Thus, expected value of linear combinations is a linear combination of the expectations:

$$E[g(X,Y)] = a_1 E[g_1(X,Y)] + \cdots + a_n E[g_n(X,Y)] \tag{10}$$

and the same result holds for continuous r.v.'s via integration.

# Expectation of Sums of Random Variables

- For any two random variables $X$ and $Y$ we have $E[X+Y] = E[X] + E[Y]$.
  - Follows from (10) with $g(X,Y) = X + Y$, i.e. $n = 2$, $a_1 = 1, a_2 = 1$, $g_1(X,Y) = X$, and $g_2(X,Y) = Y$.
- Consider r.v.'s $X_1, X_2, \ldots, X_n$:
  - Let $\tilde{X}_1 = X_1$ and $\tilde{Y}_1 = X_2 + X_3 + \cdots + X_n$ then $E[\tilde{X}_1 + \tilde{Y}_1] = E[\tilde{X}_1] + E[\tilde{Y}_1]$, i.e.

$$E[X_1 + \cdots + X_n] = E[X_1] + E[X_2 + \cdots + X_n].$$

  - Let $\tilde{X}_2 = X_2$ and $\tilde{Y}_2 = X_3 + X_4 + \cdots + X_n$ then $E[\tilde{X}_2 + \tilde{Y}_2] = E[\tilde{X}_2] + E[\tilde{Y}_2]$, i.e.

$$E[X_1 + \cdots + X_n] = E[X_1] + E[X_2] + E[X_3 + \cdots + X_n]$$

  - And so on ... By induction, it follows that

$$E[X_1 + \cdots + X_n] = E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_{n-1}] + E[X_n]. \quad (11)$$

  i.e. the expected value of a sum is the sum of the expected values.

# Recall Previous Families of R.V.

- Recall the expected values of the Binomial, Pascal, and Erlang models.
- Binomial $(n, p)$ = sum of $n$ Bernoulli $(p)$ r.v.'s
  - Bernoulli $(p)$ r.v. $X$ has mean $E[X] = p$
  - Binomial $(n, p)$ r.v. $K$ has mean $E[K] = np$
- Pascal $(r, p)$ = sum of $r$ Geometric $(p)$ r.v.'s
  - Geometric $(p)$ r.v. $L$ has mean $E[L] = 1/p$
  - Pascal $(r, p)$ r.v. $L_r$ has mean $E[L_r] = r/p$
- Erlang $(r, \lambda)$ = sum of $r$ Exponential $(\lambda)$ r.v.'s
  - Exponential $(\lambda)$ has mean $E[X] = 1/\lambda$
  - Erlang $(r, \lambda)$ has mean $E[X] = r/\lambda$
- Equation (11) can be used to deduce expected value of these r.v.'s

# Covariance

- For any two random variables $X$ and $Y$ the variance of $W = X + Y$ is given by

$$\begin{aligned} \mathrm{Var}[W] &= E[(W - E[W])^2] \\ \mathrm{Var}[X+Y] &= E\left[[(X+Y) - (E[X] + E[Y])]^2\right] \\ &= E\left[[(X - E[X]) + (Y - E[Y])]^2\right] \\ &= E\left[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2\right] \\ &= \mathrm{Var}[X] + 2E[(X - E[X])(Y - E[Y])] + \mathrm{Var}[Y], \end{aligned} \tag{12}$$

  - The variance of the sum of two r.v.'s is the sum of the variances plus the cross term that quantifies the relationship/dependence (or covariance) between $X$ and $Y$.

- The covariance between two r.v.'s $X$ and $Y$ is defined as

$$\sigma_{X,Y} \triangleq \mathrm{Cov}(X,Y) \triangleq E[(X - E[X])(Y - E[Y])] = E[XY] - E[X] \cdot E[Y].$$

# Covariance Interpretation

- Covariance helps quantify the "spread" of pairs $(X, Y)$ around pair $(E[X], E[Y])$.
- If $\text{Cov}(X, Y) > 0$ then $(X - E[X])$ and $(Y - E[Y])$ have same sign "on average."
  - i.e. $(X - E[X])$ and $(Y - E[Y])$ move together more often than not
- If $\text{Cov}(X, Y) < 0$ then $(X - E[X])$ and $(Y - E[Y])$ have opposite sign "on average."
  - i.e. $(X - E[X])$ and $(Y - E[Y])$ move opposed more often than not
- If $\text{Cov}(X, Y) \simeq 0$ then $(X - E[X])$ and $(Y - E[Y])$ are asynchronous "on average."
  - $X$ and $Y$ are said to be uncorrelated if $\text{Cov}(X, Y) = 0$, i.e.

$$\text{Cov}(X, Y) = 0 \Longrightarrow E[XY] - E[X] \cdot E[Y] = 0$$
$$\Longrightarrow E[XY] = E[X] \cdot E[Y].$$

# Covariance Properties and Correlation

- Some useful properties of covariance include:
  - $\text{Cov}[X, X] = \text{Var}[X]$

$$\begin{aligned}\text{Cov}[X, aY + b] &= E[X(aY + b)] - E[X] \cdot E[aY + b] \\ &= E[aXY + bX] - E[X] \cdot (aE[Y] + b) \\ &= aE[XY] + bE[X] - aE[X] \cdot E[Y] - bE[X] \\ &= a(E[XY] - E[X] \cdot E[Y]) = a\text{Cov}(X, Y)\end{aligned}$$

  - Similarly, $\text{Cov}[X, a_1 Y_1 + a_2 Y_2] = a_1 \cdot \text{Cov}[X, Y_1] + a_2 \cdot \text{Cov}[X, Y_2]$

- The correlation between r.v. pair $(X, Y)$ is defined as $r_{X,Y} \triangleq E[XY]$.
  - Note that $\text{Cov}(X, Y) = r_{X,Y} - E[X] \cdot E[Y]$.
  - When $X = Y \implies r_{X,X} = E[X^2] = \text{Var}[X] + E[X]^2$.
  - If $r_{X,Y} = E[XY] = 0$ then r.v.'s $X$ and $Y$ are said to be orthogonal.
    - Orthogonality does not imply r.v.'s are uncorrelated.
    - Uncorrelatedness does not imply r.v.'s are orthogonal.
    - If $X$ and $Y$ are zero mean then they are orthogonal if and only if uncorrelated.

# Correlation Coefficient

- The (Pearson) correlation coefficient between r.v. pair $(X, Y)$ is

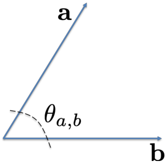
$$\rho_{X,Y} = E\left[\left(\frac{X - E[X]}{\sigma_X}\right)\left(\frac{Y - E[Y]}{\sigma_Y}\right)\right] = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[Y]}} = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y}. \quad (13)$$

  - $\rho_{X,Y}$ is the correlation between standardized values of $X$ and $Y$.
  - Note $\rho_{X,Y}$ is a unitless quantity due to normalization.

- One can establish that $-1 \le \rho_{X,Y} \le 1$. First, we review:
  - Recall that for two vectors $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$ the dot product is

$$\mathbf{a}\cdot\mathbf{b} = \sum_{i=1}^{n} a_i b_i = ||\mathbf{a}||\cdot||\mathbf{b}||\cdot\cos\theta_{a,b} \quad \text{where} \quad ||\mathbf{a}||^2 = \sum_{i=1}^{n} a_i^2$$

$$\text{and} \quad -1 \le \cos\theta_{a,b} \le 1 \Longrightarrow -1 \le \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2 \cdot \sum_{i=1}^{n} b_i^2}} \le 1, \quad (14)$$




where $\cos\theta_{a,b} = \pm 1$ if and only if $\mathbf{a} = c\cdot\mathbf{b}$ for some $c \in \mathbb{R}$, i.e. a linear relationship exists.

  - The last inequality in (14) is called the Cauchy-Schwarz inequality.

# Correlation Coefficient

- Proof cont.:
  - Given paired discrete r.v.'s $(X, Y) \sim P_{X,Y}$, consider the choice

$$a_{x,y} = (x - E[X]) \cdot \sqrt{P_{X,Y}(x,y)}, \quad b_{x,y} = (y - E[Y]) \cdot \sqrt{P_{X,Y}(x,y)} \Longrightarrow$$

$$\sum_{(x,y)\in S_{X,Y}} a_{x,y} b_{x,y} = E\{(X - E[X])(Y - E[Y])\} = \text{Cov}(X,Y),$$

$$\sum_{(x,y)\in S_{X,Y}} a_{x,y}^2 = E\{(X - E[X])^2\} = \text{Var}[X], \quad \sum_{(x,y)\in S_{X,Y}} b_{x,y}^2 = E\{(Y - E[Y])^2\} = \text{Var}[Y]$$

$$\Longrightarrow \frac{\sum_{(x,y)\in S_{X,Y}} a_{x,y} b_{x,y}}{\sqrt{\sum_{(x,y)\in S_{X,Y}} a_{x,y}^2 \cdot \sum_{(x,y)\in S_{X,Y}} b_{x,y}^2}} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}} = \rho_{X,Y}.$$

  Thus, by (14) we have established $|\rho_{X,Y}| \le 1$. ■.

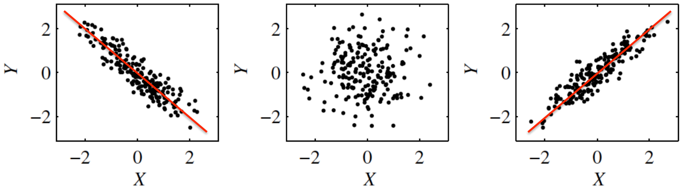
  - This proof illustrates that expectations can be interpreted geometrically as a type of dot product (or inner product in general) for random variables.
    - This is a powerful concept used extensively in signal processing.
  - A similar argument can be made for the continuous r.v. case.*

---

* A version of the Cauchy-Schwarz inequality for functions $a(x)$ and $b(x)$ of a continuous variable $x$ is $-1 \le \frac{\int a(x)b(x)dx}{\sqrt{\int a^2(x)dx \cdot \int b^2(x)dx}} \le 1$.

# Correlation Coefficient

- Consider data samples of r.v. pairs $(X_i, Y_i) \sim f_{X,Y}$ with $E[X] = E[Y] = 0$, $\text{Var}[X] = \text{Var}[Y] = 1$ and cases of $\rho_{X,Y}$ plotted (Yates/Goodman p. 188):




(a) $\rho_{X,Y} = -0.9$ (b) $\rho_{X,Y} = 0$ (c) $\rho_{X,Y} = 0.9$

- Note cases (a) and (c) show a cloud of data points roughly falling on a line.
  - Case (a) shows that $\rho_{X,Y} \simeq -1 \implies X \simeq -Y$
  - Case (c) shows that $\rho_{X,Y} \simeq +1 \implies X \simeq Y$
  - Such paired r.v.'s are said to be highly correlated
- Case (b) for $\rho_{X,Y} \simeq 0$ shows no apparent linear relationship between $X$ & $Y$

# Correlation Coefficient and Linear Relationships

- If paired r.v.'s $(X, Y)$ satisfy $Y = aX + b$ for $a, b \in \mathbb{R}$, then $\rho_{X,Y} = \text{sign}(a)$:

$$\rho_{X,Y} = \frac{\text{Cov}[X, aX+b]}{\sqrt{\text{Var}[X] \cdot \text{Var}[aX+b]}} = \frac{a\text{Cov}[X,X]}{\sqrt{\text{Var}[X] \cdot a^2\text{Var}[X]}} = \frac{a}{|a|} \cdot \frac{\text{Var}[X]}{\text{Var}[X]}$$
$$= \text{sign}(a) = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & a < 0. \end{cases} \quad (15)$$

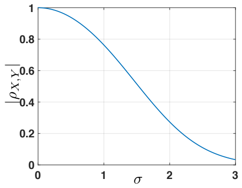
  - Recall that $-1 \le \rho_{X,Y} \le 1$.
  - Cauchy-Schwarz inequality (14), and equation (15) above show that to obtain $\rho_{X,Y} = \pm 1$, a linear relationship between $X$ and $Y$ is necessary.
  - Thus, correlation coefficient $\rho_{X,Y}$ is a measure of linear relationship.

# Example: Correlation Coefficient and Nonlinear Relationships

- Let $X \sim N(\mu, \sigma^2)$ and $Y = e^{-X} + b$ where $b \in \mathbb{R}$ is constant.
  - Clearly, paired r.v.'s $(X, Y)$ are dependent/related but via a nonlinear relationship.
- Use the MGF $\phi_X(s) = E[e^{sX}] = e^{s\mu + \frac{s^2\sigma^2}{2}}$ to show the following*

$$\mathrm{Cov}(X,Y) = -\sigma^2 e^{-\mu + \frac{\sigma^2}{2}}, \quad \mathrm{Var}[Y] = e^{-2\mu+\sigma^2}(e^{\sigma^2} - 1) \Longrightarrow$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}} = \frac{-\sigma}{\sqrt{e^{\sigma^2} - 1}} \tag{16}$$




- This example highlights that correlation is only a measure of linear dependence.
  - Thus, $\rho_{X,Y} \simeq 0$ does not mean $X$ and $Y$ unrelated; only that relationship is not linear.

---

*Hint: Convince yourself that $E[X^n e^{aX}] = \frac{d^n \phi_X(s)}{ds^n}\Big|_{s=a}$ and use this to find desired expectations.

# Bivariate Gaussian Random Variables

- Recall univariate Gaussian random variable $X \sim N(\mu, \sigma^2)$ has PDF and MGF:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \overset{\mathcal{L}}{\longleftrightarrow} e^{s\mu + \frac{s^2\sigma^2}{2}} = \phi_X(s) \tag{17}$$

- The column vector $\mathbf{W} = [X, Y]^T$ is a bivariate Gaussian if its joint density is

$$f_{\mathbf{W}}(\mathbf{w}) = |2\pi \mathbf{C}_W|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu}_W)^T \mathbf{C}_W^{-1}(\mathbf{w} - \boldsymbol{\mu}_W)\right], \text{ where}$$

$$\boldsymbol{\mu}_W = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \text{ and } \mathbf{C}_W = \begin{bmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{X,Y}\sigma_X\sigma_Y \\ \rho_{X,Y}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}, \tag{18}$$

and $\mu_X = E[X]$, $\mu_Y = E[Y]$, $\sigma_X^2 = \text{Var}[X]$, $\sigma_Y^2 = \text{Var}[Y]$, $\sigma_{X,Y} = \text{Cov}(X, Y)$, $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$, $\mathbf{w} = [x, y]^T$, and the support $S_{\mathbf{W}}$ is the entire $xy$-plane.

---

$^*$ Notation: italics indicate scalars, as in $A$; lower case boldface indicate column vectors, as in $\mathbf{a}$; upper case boldface indicate matrices, as in $\mathbf{A}$. Complex conjugation is indicated by a superscript $*$ as in $A^*$. Matrix transpose is indicated by a superscript $T$ as in $\mathbf{A}^T$, and complex conjugate plus matrix transpose is indicated by a superscript $H$ as in $\mathbf{A}^H = (\mathbf{A}^T)^*$. Absolute value of a complex scalar indicated as $|A|$; and the matrix determinant indicated as $|\mathbf{A}|$ and $\det(\mathbf{A})$.

# Bivariate Gaussian Random Variables*

- Intuitive value of expressing bivariate Gaussian PDF as in (18) is its analogous form to univariate Gaussian PDF in (17) and similar interpretations:
  - Parameter vector $\boldsymbol{\mu}_W$ is mean vector for random vector $\mathbf{W} = [X, Y]^T$.
  - Parameter matrix $\mathbf{C}_W$ is covariance matrix of random vector $\mathbf{W} = [X, Y]^T$.
  - We denote this distribution as $\mathbf{W} \sim N(\boldsymbol{\mu}_W, \mathbf{C}_W)$.
- Note that (18) can be written exclusively in terms of the pair $(X, Y)$ via $(x, y)$.
  - To see this clearly, first recall from linear algebra that

$$|\mathbf{C}_W| = \sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2, \text{ and } \mathbf{C}_W^{-1} = \begin{bmatrix} \sigma_Y^2 & -\sigma_{X,Y} \\ -\sigma_{X,Y} & \sigma_X^2 \end{bmatrix} \cdot \frac{1}{\sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2};$$

Also, $|2\pi\mathbf{C}_W| = (2\pi)^2|\mathbf{C}_W|$, $|2\pi\mathbf{C}_W|^{-1/2} = \left(2\pi\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2}\right)^{-1}$, and $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y} \Longrightarrow$

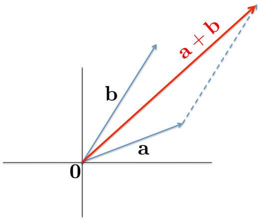
$$\mathbf{C}_W^{-1} = \begin{bmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho_{X,Y}}{\sigma_X\sigma_Y} \\ -\frac{\rho_{X,Y}}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{bmatrix} \cdot \frac{1}{1-\rho_{X,Y}^2}, \text{ and } |2\pi\mathbf{C}_W|^{-1/2} = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}\right)^{-1}.$$

  - Thus, it is seen that

$$f_{\mathbf{W}}(\mathbf{w}) = f_{X,Y}(x,y) = \frac{\exp\left[-\frac{1}{2}\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho_{X,Y}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{1-\rho_{X,Y}^2}\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}.$$

# Bivariate Gaussian: Vector Interpretations

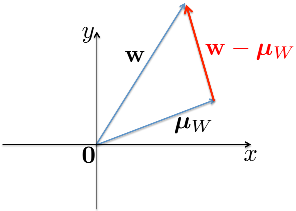
- (18) expresses $f_{X,Y}$ in vector form and is perhaps more intuitive.
  - Recall from linear algebra that vectors add graphically from head-to-tail. This is illustrated below:




where vector **b** is redrawn as a dashed-line arrow starting from the head of vector **a**. The sum vector $\mathbf{a} + \mathbf{b}$ is obtained by drawing an arrow from the tail of **a** to the head of the re-drawn **b** (this is illustrated in red).

# Bivariate Gaussian: Vector Interpretations*

- Regarding (18), note that the argument of the exponential is a quadratic in terms of the vector $\mathbf{w} - \boldsymbol{\mu}_W$.
  - This vector difference is illustrated below with head-to-tail addition:




  - Clearly, the tail of $\mathbf{w} - \boldsymbol{\mu}_W$ begins at the head of $\boldsymbol{\mu}_W$, and the head of $\mathbf{w} - \boldsymbol{\mu}_W$ ends at the head of $\mathbf{w}$ (this is illustrated in red).
    - Note graphically the head-to-tail addition is consistent, i.e. $\boldsymbol{\mu}_W + (\mathbf{w} - \boldsymbol{\mu}_W) = \mathbf{w}$.
  - Thus, regarding (18) we should treat $\boldsymbol{\mu}_W$ as the new origin of the coordinate system that matters here, and view all vectors $\mathbf{w}$ relative to $\boldsymbol{\mu}_W$ via the difference $\mathbf{w} - \boldsymbol{\mu}_W$.

# Bivariate Gaussian: Covariance Eigenvector Decomposition

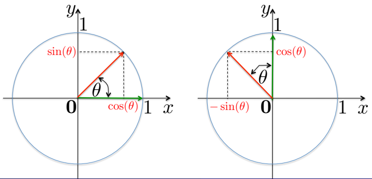
- Later we will prove covariance matrix $\mathbf{C}_W$ is symmetric, i.e. $\mathbf{C}_W = \mathbf{C}_W^T$, and it is positive definite, i.e. $\mathbf{a}^T \mathbf{C}_W \mathbf{a} > 0$ for all $2 \times 1$ vectors $\mathbf{a} = [a_x, a_y]^T \neq \mathbf{0}$.
  - Recall from linear algebra such matrices have an eigen-decomposition such that

$$\mathbf{C}_W = \mathbf{Q} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{Q}^T, \text{ where } \mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \lambda_1 > 0,\ \lambda_2 > 0.$$

  - Matrix $\mathbf{Q}$ is orthogonal or orthonormal and rotates/reflects vectors they multiply.
  - The form of $\mathbf{Q}$ that rotates vectors by angle $\theta$ is:

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \Longrightarrow \mathbf{Q} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \text{ and } \mathbf{Q} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}$$

  - i.e., standard basis $\mathbf{e}_1 = [1, 0]^T \mapsto [\cos(\theta), \sin(\theta)]^T$; $\mathbf{e}_2 = [0, 1]^T \mapsto [-\sin(\theta), \cos(\theta)]^T$.
  - These rotations are illustrated here:



# Bivariate Gaussian: Covariance Eigenvector Decomposition

- Let $\mathbf{w}_0 = \mathbf{w} - \boldsymbol{\mu}_W$ and note the inverse covariance can be written as

$$\mathbf{C}_W^{-1} = \mathbf{Q}\begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix}\mathbf{Q}^T \implies (\mathbf{w}-\boldsymbol{\mu}_W)^T \mathbf{C}_W^{-1}(\mathbf{w}-\boldsymbol{\mu}_W) = \mathbf{w}_0^T \mathbf{Q}\begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix}\mathbf{Q}^T \mathbf{w}_0. \quad (19)$$

- Define the new variables $\mathbf{z} = \mathbf{Q}^T \mathbf{w}_0$ whose origin is $\boldsymbol{\mu}_W$ in the $\mathbf{w}$ plane.
  - Note that

$$\mathbf{w}_0^T \mathbf{Q}\begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix}\mathbf{Q}^T \mathbf{w}_0 = \mathbf{z}^T \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix}\mathbf{z} = \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} \implies$$

$$f_{\mathbf{W}}(\mathbf{w}) \propto \exp\left[-\frac{1}{2}\left(\frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2}\right)\right].$$

  - Contours of constant density given by ellipses in variables $z_1$ and $z_2$, i.e.

$$f_{\mathbf{W}}(\mathbf{w}) = \text{constant} \implies \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} = \gamma^2 \ (\text{constant} \geq 0) \quad (20)$$

which is clearly the equation for an ellipse in $[z_1, z_2]^T$, whose origin is $\boldsymbol{\mu}_W$.

# Bivariate Gaussian: Change of Variables Interpretation*

- To interpret new variables $\mathbf{z}$, first recall that in the original coordinate system for $\mathbf{w}$ that if coordinates $\mathbf{w} = [x, y]^T$, then we interpret this as meaning

$$\mathbf{w} = x \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{i.e. each coordinate represents a coefficient for a basis vector.} \tag{21}$$

- Transformation via multiplication by $\mathbf{Q}^T$ defines new coordinates $\mathbf{z} = \mathbf{Q}^T \mathbf{w}_0 = [z_1, z_2]^T$:
  - Let $\mathbf{Q} = [\mathbf{q}_1 | \mathbf{q}_2]$ where $\mathbf{q}_i$, $i = 1, 2$ are column vectors. Since $\mathbf{Q}$ is full rank, then for any vector $\mathbf{w}_0 \implies$

$$\mathbf{w}_0 = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 \text{ where coefficients } \alpha_i \text{ can be found by noting}$$

$$\mathbf{q}_1^T \mathbf{w}_0 = \alpha_1 \mathbf{q}_1^T \mathbf{q}_1 + \alpha_2 \mathbf{q}_1^T \mathbf{q}_2 = \alpha_1 \cdot 1 + \alpha_2 \cdot 0 = \alpha_1$$

$$\mathbf{q}_2^T \mathbf{w}_0 = \alpha_1 \mathbf{q}_2^T \mathbf{q}_1 + \alpha_2 \mathbf{q}_2^T \mathbf{q}_2 = \alpha_1 \cdot 0 + \alpha_2 \cdot 1 = \alpha_2.$$

  - Noting $\mathbf{z} = \mathbf{Q}^T \mathbf{w}_0 = \begin{bmatrix} \mathbf{q}_1^T \\ \hline \mathbf{q}_2^T \end{bmatrix} \mathbf{w}_0 = \begin{bmatrix} \mathbf{q}_1^T \mathbf{w}_0 \\ \hline \mathbf{q}_2^T \mathbf{w}_0 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$, and recalling interpretation (21), it is clear that $\mathbf{z} = [z_1, z_2]^T$ are the coordinates of the new variables, and the basis vectors defining the coordinates of the new systems are given by $\mathbf{q}_1$ and $\mathbf{q}_2$.

# Bivariate Gaussian: Change of Variables Interpretation*

- Now, we put these pieces together to visualize the bivariate Gaussian PDF .
- Consider plotting contours of constant density, i.e. set of points with $f_{\mathbf{W}}(\mathbf{w}) = A$.
  - By (18) this requires $(\mathbf{w} - \boldsymbol{\mu}_W)^T \mathbf{C}_W^{-1} (\mathbf{w} - \boldsymbol{\mu}_W) = -2 \ln \left[ |2\pi \mathbf{C}_W|^{1/2} \cdot A \right]$ .
  - By (19)–(20) these contours are given by

$$\frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} = -2 \ln \left[ |2\pi \mathbf{C}_W|^{1/2} \cdot A \right] \triangleq \gamma^2 \tag{22}$$

  i.e. ellipses where the major and minor axes are along the basis vectors $\mathbf{q}_1$ and $\mathbf{q}_2$.
    - Which axes is minor or major depends on the values of $\lambda_1$ and $\lambda_2$.
  - Note that the set of all points satisfying (22) can be parameterized as

$$\left. \begin{aligned} z_1 &= \gamma\sqrt{\lambda_1} \cdot \cos(\phi) \\ z_2 &= \gamma\sqrt{\lambda_2} \cdot \sin(\phi) \end{aligned} \right\} \Longrightarrow$$

$$\frac{[\gamma\sqrt{\lambda_1} \cdot \cos(\phi)]^2}{\lambda_1} + \frac{[\gamma\sqrt{\lambda_2} \cdot \sin(\phi)]^2}{\lambda_2} = \frac{\gamma^2\lambda_1 \cdot \cos^2(\phi)}{\lambda_1} + \frac{\gamma^2\lambda_2 \cdot \sin^2(\phi)}{\lambda_2}$$
$$= \gamma^2[\cos^2(\phi) + \sin^2(\phi)] = \gamma^2 \cdot 1.$$

    - An ellipse is obtained as the desired contour by varying $\phi$ from 0 to $2\pi$.

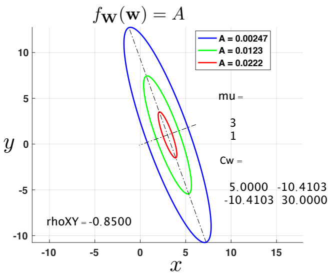
# Bivariate Gaussian: Contours of Constant Density

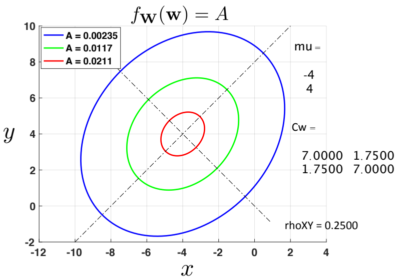
- These parameterized points in $[z_1, z_2]^T$ for a contour of constant density can be transformed back into the original coordinates $\mathbf{w}$ via:

$$\mathbf{z} = \mathbf{Q}^T(\mathbf{w} - \boldsymbol{\mu}_W) \Longrightarrow \mathbf{Q}\mathbf{z} = \mathbf{Q}\mathbf{Q}^T(\mathbf{w} - \boldsymbol{\mu}_W) = \mathbf{I}_2(\mathbf{w} - \boldsymbol{\mu}_W) \Longrightarrow$$

$$\mathbf{w} = \boldsymbol{\mu}_W + \mathbf{Q}\mathbf{z}.$$

- To demonstrate such contours consider two examples illustrated below:



# Contours of Constant Density: Example 1

- The first example in the left image is for a bivariate Gaussian with parameters:

$$\boldsymbol{\mu}_W = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \; \mathbf{C}_W = \begin{bmatrix} 5 & -10.4013 \\ -10.4013 & 30 \end{bmatrix}, \text{ where } \rho_{X,Y} = -0.85.$$

  - Notice that contours of constant density are centered on the mean $[3,1]^T$.
  - The axes of the new coordinate system are indicated by the cross-hair black dash-dot lines.
    - These cross-hairs are colinear with basis vectors $\mathbf{q}_1$ and $\mathbf{q}_2$.
  - The spread of the PDF about these axes is determined by $\lambda_1$ and $\lambda_2$.
  - Three level density contours are illustrated: one for 0.1, 0.5, and 0.9 times the maximum value the density assumes (namely, $|2\pi\mathbf{C}_W|^{-1/2}$).
  - Note that a correlation coefficient of $\rho_{X,Y} = -0.85$ suggests a strong linear dependence with points clustering in the "quadrants" II and IV. This interpretation is consistent with "quadrants" of the coordinate axes of the original system when relocated to center on $\boldsymbol{\mu}_W$.

# Contours of Constant Density: Example 2

- The second example in the right image is for a bivariate Gaussian with parameters:

$$\boldsymbol{\mu}_W = \begin{bmatrix} -4 \\ 4 \end{bmatrix}, \ \mathbf{C}_W = \begin{bmatrix} 7 & 1.75 \\ 1.75 & 7 \end{bmatrix}, \text{ where } \rho_{X,Y} = 0.25.$$

  - The contours of constant density are centered on the mean $[-4, 4]^T$.
  - The axes of the new coordinate system are indicated by the cross-hair black dash-dot lines.
    - These cross-hairs are colinear with basis vectors $\mathbf{q}_1$ and $\mathbf{q}_2$.
  - The spread of the PDF about these axes is determined by $\lambda_1$ and $\lambda_2$.
  - Three level density contours are illustrated: one for 0.1, 0.5, and 0.9 times the maximum value the density assumes.
  - A correlation coefficient of $\rho_{X,Y} = 0.25$ suggests a weak linear dependence with points clustering more so in the “quadrants” I and III. This interpretation is consistent with “quadrants” of the coordinate axes of the original system when relocated to center on $\boldsymbol{\mu}_W$.

# MGF of Bivariate Gaussian*

- We want to determine the MGF for $\mathbf{W} \sim N(\boldsymbol{\mu}_W, \mathbf{C}_W)$. Recall that

$$1 = \int_{S_\mathbf{W}} f_\mathbf{W}(\mathbf{w}) d\mathbf{w} \Longrightarrow \int_{S_\mathbf{W}} \exp\left[-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu}_W)^T \mathbf{C}_W^{-1}(\mathbf{w}-\boldsymbol{\mu}_W)\right] d\mathbf{w} = |2\pi \mathbf{C}_W|^{1/2}. \tag{23}$$

  - Interestingly, this integral converges no matter the actual value of $\boldsymbol{\mu}_W$.

- The bivariate MGF for vector $\mathbf{W} = [X, Y]^T$ is defined as

$$\phi_{X,Y}(s_x, s_y) = E\{e^{s_x X + s_y Y}\} = E\{e^{\mathbf{s}^T \mathbf{W}}\} = \phi_\mathbf{W}(\mathbf{s})$$

  where the frequency vector $\mathbf{s} = [s_x, s_y]^T$.

- Consider

$$\phi_\mathbf{W}(\mathbf{s}) = \int_{S_\mathbf{W}} e^{\mathbf{s}^T \mathbf{w}} f_\mathbf{W}(\mathbf{w}) d\mathbf{w} = |2\pi \mathbf{C}_W|^{-1/2} \int_{S_\mathbf{W}} \exp[\mathrm{Arg}_1] d\mathbf{w} \qquad \text{where,}$$

$$\begin{aligned}
\mathrm{Arg}_1 &= \mathbf{s}^T \mathbf{w} - \frac{1}{2}\left[\mathbf{w}^T \mathbf{C}_W^{-1} \mathbf{w} - \mathbf{w}^T \mathbf{C}_W^{-1} \boldsymbol{\mu}_W - \boldsymbol{\mu}_W^T \mathbf{C}_W^{-1} \mathbf{w} + \boldsymbol{\mu}_W^T \mathbf{C}_W^{-1} \boldsymbol{\mu}_W\right] \\
&= -\frac{1}{2}\left[\mathbf{w}^T \mathbf{C}_W^{-1} \mathbf{w} - 2\left(\boldsymbol{\mu}_W^T \mathbf{C}_W^{-1} + \mathbf{s}^T\right) \mathbf{w} + \boldsymbol{\mu}_W^T \mathbf{C}_W^{-1} \boldsymbol{\mu}_W\right] \\
&= -\frac{1}{2}\left[\mathbf{w}^T \mathbf{C}_W^{-1} \mathbf{w} - 2\left(\boldsymbol{\mu}_W^T + \mathbf{s}^T \mathbf{C}_W\right) \mathbf{C}_W^{-1} \mathbf{w} + \boldsymbol{\mu}_W^T \mathbf{C}_W^{-1} \boldsymbol{\mu}_W\right].
\end{aligned}$$

# MGF of Bivariate Gaussian*

- Now by completing the square in $\mathbf{w}$ we obtain

$$\begin{aligned}\text{Arg}_1 = -\frac{1}{2}\Big\{ & [\mathbf{w}-(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})]^T\,\mathbf{C}_W^{-1}\,[\mathbf{w}-(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})] \\ & -(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})^T\mathbf{C}_W^{-1}(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})+\boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\boldsymbol{\mu}_W\Big\}.\end{aligned}$$

- Thus, integral for MGF can be written as

$$\begin{aligned}\phi_{\mathbf{w}}(\mathbf{s}) = & |2\pi\mathbf{C}_W|^{-1/2}\exp\left[\frac{1}{2}(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})^T\mathbf{C}_W^{-1}(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})-\frac{1}{2}\boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\boldsymbol{\mu}_W\right] \\ & \times\int_{S_{\mathbf{w}}}\exp\left(-\frac{1}{2}\left\{[\mathbf{w}-(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})]^T\,\mathbf{C}_W^{-1}\,[\mathbf{w}-(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})]\right\}\right)d\mathbf{w}.\end{aligned}$$

- This integral can be evaluated using the integral identity in (23) $\Longrightarrow$

$$\phi_{\mathbf{w}}(\mathbf{s}) = |2\pi\mathbf{C}_W|^{-1/2}\exp\left[\frac{1}{2}\text{Arg}_2\right]\times|2\pi\mathbf{C}_W|^{1/2}, \qquad \text{where}$$

$$\begin{aligned}\text{Arg}_2 &= (\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})^T\mathbf{C}_W^{-1}(\boldsymbol{\mu}_W+\mathbf{C}_W\mathbf{s})-\boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\boldsymbol{\mu}_W \\ &= \boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\boldsymbol{\mu}_W+\boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\mathbf{C}_W\mathbf{s}+\mathbf{s}^T\mathbf{C}_W\mathbf{C}_W^{-1}\boldsymbol{\mu}_W+\mathbf{s}^T\mathbf{C}_W\mathbf{C}_W^{-1}\mathbf{C}_W\mathbf{s}-\boldsymbol{\mu}_W^T\mathbf{C}_W^{-1}\boldsymbol{\mu}_W \\ &= 2\mathbf{s}^T\boldsymbol{\mu}_W+\mathbf{s}^T\mathbf{C}_W\mathbf{s}.\end{aligned}$$

# MGF of Bivariate Gaussian*

- Thus, we conclude that the MGF for the bivariate Gaussian is

$$\phi_{\mathbf{W}}(\mathbf{s}) = \exp\left[\mathbf{s}^T \boldsymbol{\mu}_W + \frac{1}{2}\mathbf{s}^T \mathbf{C}_W \mathbf{s}\right] \qquad \text{for all } \mathbf{s} \in \mathbb{C}^2 \tag{24}$$

- The form of $\phi_{\mathbf{W}}(\mathbf{s})$ is similar to the univariate MGF in (17), but represents a bivariate extension.
- We will discover that (24) is the general form of the MGF for a multivariate Gaussian distribution that we will soon define;
  - namely, when we consider $n$ joint random variables $\mathbf{W} = [X_1, X_2, \ldots, X_n]^T$ with mean vector $\boldsymbol{\mu}_W = [\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}]^T$ and $n \times n$ covariance matrix $\mathbf{C}_W$, and frequency vector $\mathbf{s} = [s_1, s_2, \ldots, s_n]^T$.

# Example: 2-D Uniform PDF

- Recall joint PDF example in (4) where we showed that the marginal PDFs are

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{Otherwise.} \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{d-c}, & c \le y \le d \\ 0, & \text{Otherwise.} \end{cases}$$

  From this, we also know that the product is $f_X(x) \cdot f_{X,Y}(x,y)$

- If pair of r.v.'s $(X, Y)$ are independent, then they are uncorrelated:

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f_X(x) \cdot f_Y(y)dxdy$$
$$= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \cdot \int_{-\infty}^{\infty} y \cdot f_Y(y)dy = E[X] \cdot E[Y]$$
$$\Longrightarrow \text{Cov}(X,Y) = E[XY] - E[X] \cdot E[Y] = 0.$$

- If pair of r.v.'s $(X, Y)$ are uncorrelated, then they are not necessarily independent.
  - Such was illustrated in the example of equation (16).
  - Independence is a much stronger statement than uncorrelatedness.

# Sums of Independent Random Variables

- If pair of r.v.'s $(X, Y)$ are independent, then $\text{Cov}(X, Y) = 0$ and by (12) it follows:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]. \tag{25}$$

  - Consider $n$ independent r.v.'s $X_1, X_2, \ldots, X_n$. By (25) we can establish

$$\text{Var}[X_1 + X_2 + \cdots + X_n] = \text{Var}[X_1] + \text{Var}[X_2] + \cdots + \text{Var}[X_{n-1}] + \text{Var}[X_n] \tag{26}$$

    i.e. the variance of a sum of independent r.v.'s is the sum of the variances.

- If r.v.'s $X$ and $Y$ are independent, then given functions $g(X)$ and $q(Y) \Longrightarrow$

$$\begin{aligned} E[g(X) \cdot q(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot q(y) \cdot f_{X,Y}(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot q(y) \cdot f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx \int_{-\infty}^{\infty} q(y) \cdot f_Y(y)\, dy \\ &= E[g(X)] \cdot E[q(Y)]. \end{aligned} \tag{27}$$

# Sums of Independent Random Variables

- Consider $W = X + Y$ where r.v. pair $(X, Y) \sim f_{X,Y} = f_X \cdot f_Y$.
  - The MGF for $W$ follows from

$$\phi_W(s) = E[e^{sW}] = E[e^{s(X+Y)}] = E[e^{sX} \cdot e^{sY}] = E[e^{sX}] \cdot E[e^{sY}] = \phi_X(s) \cdot \phi_Y(s)$$
$$= \int_{-\infty}^{\infty} f_X(x) e^{sx}\, dx \cdot \int_{-\infty}^{\infty} f_Y(y) e^{sy}\, dy = \int_{-\infty}^{\infty} f_X(x)\, dx \int_{-\infty}^{\infty} f_Y(y) e^{s(x+y)}\, dy,$$

  where the fourth equality follows from (27), i.e. independence between $X$ and $Y$;
  - Change variables from $y$ to $w = x + y$ for fixed $x$; thus, $dy = dw$, $y = w - x \Longrightarrow$

$$\phi_W(s) = \int_{-\infty}^{\infty} f_X(x)\, dx \int_{-\infty}^{\infty} f_Y(w - x) e^{sw}\, dw$$
$$= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx \right] e^{sw}\, dw = \int_{-\infty}^{\infty} f_W(w) e^{sw}\, dw = E[e^{sW}]$$

  - Since the Laplace transform is unique, $f_W(w)$ is equal to integrand in brackets $[\cdot]$.
  - Thus, PDF of sum of two independent r.v.'s is given by convolution of their PDFs:

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx.$$

  - Try finding $f_W$ when $X \sim \text{box}_x(0,1)$, $Y \sim \text{box}_x(0,1)$, i.e. both uniform on $[0,1]$ and independent.

# Conditioning on Discrete Random Variables

- Conditional probabilities naturally update a model based on new information.
- Conditional probability of event $\mathcal{A}$ given that $\mathcal{B}$ has occurred is

$$\Pr(\mathcal{A}|\mathcal{B}) = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})}, \quad \text{for } \Pr(\mathcal{B}) \neq 0. \tag{28}$$

  - It follows that

$$\Pr(\mathcal{A} \cap \mathcal{B}) = \Pr(\mathcal{B}) \Pr(\mathcal{A}|\mathcal{B}). \tag{29}$$

- Given discrete r.v. $X$ and event $\mathcal{B} \subseteq S_X$ with $\Pr(\mathcal{B}) > 0$, one gets a conditional PMF:
  - Consider event $\mathcal{A} = \{X = x\}$; by (28) it follows that

$$P_{X|\mathcal{B}}(x) = \Pr(X = x|\mathcal{B}) = \frac{\Pr(\{X = x\} \cap \{x \in \mathcal{B}\})}{\Pr(\mathcal{B})} = \begin{cases} \frac{P_X(x)}{\Pr(\mathcal{B})}, & x \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases}$$

  - Notational convention for a conditional PMF given event $\mathcal{B}$ is $P_{X|\mathcal{B}}$
  - Note that updated model satisfies all axioms, e.g. $\sum_{x \in S_X} P_{X|\mathcal{B}}(x) = 1$.

# Conditioning on Multiple Discrete Random Variables*

- The conditional expectation of function $g(X)$ given event $\mathcal{B}$ can be defined as

$$E[g(X)|\mathcal{B}] = \sum_{x \in S_X} g(x) \cdot P_{X|\mathcal{B}}(x). \tag{30}$$

i.e. the same as unconditional definition, but using conditional PMF.

- This idea extends to paired r.v. $(X, Y) \sim P_{X,Y}$ by choosing the event $\mathcal{A} = \{X = x, Y = y\}$
  - Conditional joint PMF is also obtained using (28):

$$P_{X,Y|\mathcal{B}}(x, y) = \Pr(X = x, Y = y|\mathcal{B}) = \frac{\Pr(\{X = x, Y = y\} \cap \{(x, y) \in \mathcal{B}\})}{\Pr(\mathcal{B})}$$
$$= \begin{cases} \dfrac{P_{X,Y}(x, y)}{\Pr(\mathcal{B})}, & (x, y) \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases} \tag{31}$$

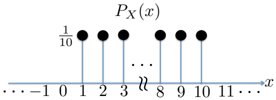
  - Conditional expectation of function $g(X, Y)$ given event $\mathcal{B}$ defined as

$$E[g(X, Y)|\mathcal{B}] = \sum_{(x,y) \in S_{X,Y}} g(x, y) \cdot P_{X,Y|\mathcal{B}}(x, y).$$

# Example: Uniform PMF

- Consider a discrete r.v. $X \sim P_X(x)$ such that

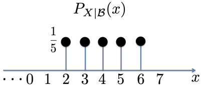
$$P_X(x) = \begin{cases} \frac{1}{10}, & x = 1, 2, 3, \ldots, 10 \\ 0, & \text{Otherwise.} \end{cases}$$




and consider events $\mathcal{A} = \{X = x\}$ and $\mathcal{B} = \{|X - 4| \le 2\}$.

- Note that $\Pr(\mathcal{A}) = P_X(x)$ and $\Pr(\mathcal{B}) = \Pr(2 \le X \le 6) = \sum_{x=2,3,4,5,6} P_X(x) = \frac{5}{10}$
- The conditional probability of event $\mathcal{A}$ given event $\mathcal{B}$ is

$$\begin{aligned} \Pr(\mathcal{A}|\mathcal{B}) &= \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})} \\ \Pr(X = x \mid |X - 4| \le 2) &= \frac{\Pr(\{X = x\} \cap \{|X - 4| \le 2\})}{\Pr(|X - 4| \le 2)} \\ P_{X|\mathcal{B}}(x) &= \begin{cases} \frac{P_X(x)}{\Pr(\mathcal{B})} = \frac{\frac{1}{10}}{\frac{5}{10}} = \frac{1}{5}, & 2 \le x \le 6 \\ 0, & \text{Otherwise} \end{cases} \end{aligned}$$



# Example: Uniform PMF Cont.

- The conditional expectation of $X$, $X^2$ and its variance given $\mathcal{B}$:

$$E[X|\mathcal{B}] = \sum_{x \in S_X} x \cdot P_{X|\mathcal{B}}(x) = (2+3+4+5+6) \cdot \frac{1}{5} = \frac{20}{5} = 4$$

$$E[X^2|\mathcal{B}] = \sum_{x \in S_X} x^2 \cdot P_{X|\mathcal{B}}(x) = (2^2+3^2+4^2+5^2+6^2) \cdot \frac{1}{5} = \frac{90}{5} = 18$$

$$\text{Var}[X|\mathcal{B}] = \sum_{x \in S_X} (x - E[X|\mathcal{B}])^2 \cdot P_{X|\mathcal{B}}(x) = E[X^2|\mathcal{B}] - E[X|\mathcal{B}]^2 = \frac{90-80}{5} = \frac{10}{5} = 2.$$

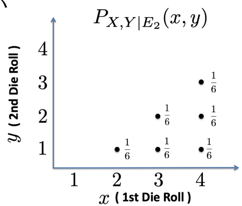
# Example: Rolling a Pair of Four-Sided Dice

- Recall this discrete r.v. example discussed earlier that resulted in equation (1).
- Consider events $\mathcal{A} = \{(X = x) \cap (Y = y)\}$ and $\mathcal{B} = E_2 = \{X > Y\}$
- Conditional probablity of event $\mathcal{A}$ given event $\mathcal{B}$ follows from (31):

$$\Pr(\mathcal{A}|\mathcal{B}) = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})}$$

$$\Pr(X = x, Y = y \mid X > Y) = \frac{\Pr(\{X = x, Y = y\} \cap \{X > Y\})}{\Pr(X > Y)}$$

$$P_{X,Y|E_2}(x,y) = \begin{cases} \dfrac{P_{X,Y}(x,y)}{\Pr(E_2)} = \dfrac{\frac{1}{16}}{\frac{6}{16}} = \dfrac{1}{6}, & (x,y) \in E_2 \\ 0, & \text{Otherwise} \end{cases}$$




- The marginal PMFs easily follow:
  - $P_{X|E_2}(x) = \sum_{y \in S_Y} P_{X,Y|E_2}(x,y) = (\frac{1}{6}, \frac{2}{6}, \frac{3}{6})$ for $x = 2, 3, 4$ respectively; otherwise $P_{X|E_2}(x) = 0$.
  - $P_{Y|E_2}(y) = \sum_{x \in S_X} P_{X,Y|E_2}(x,y) = (\frac{3}{6}, \frac{2}{6}, \frac{1}{6})$ for $y = 1, 2, 3$ respectively; otherwise $P_{Y|E_2}(y) = 0$.

# Example: Rolling a Pair of Four-Sided Dice Cont.

- Conditional expectations can be determined by using the conditional PMFs:

$$E[X|E_2] = \sum_{x \in S_X} x \cdot P_{X|E_2}(x) = 2 \cdot \frac{1}{6} + 3 \cdot \frac{2}{6} + 4 \cdot \frac{3}{6} = \frac{2+6+12}{6} = \frac{20}{6} = \frac{10}{3}$$

$$E[Y|E_2] = \sum_{y \in S_Y} y \cdot P_{Y|E_2}(y) = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{3+4+3}{6} = \frac{10}{6} = \frac{5}{3}$$

$$E[XY|E_2] = \sum_{(x,y) \in S_{X,Y}} xy \cdot P_{X,Y|E_2}(x,y) = (2+3+4+6+8+12) \cdot \frac{1}{6} = \frac{35}{6}$$

and the conditional covariance follows from

$$\mathrm{Cov}[X,Y|E_2] = E[XY|E_2] - E[X|E_2] \cdot E[Y|E_2] = \frac{35}{6} - \frac{50}{9} = \frac{105-100}{18} = \frac{5}{18}$$

# Conditioning on Continuous Random Variables

- Given continuous r.v. $X$ and event $\mathcal{B} \subseteq \mathbb{R}$ with $\Pr(\mathcal{B}) > 0$, one gets a conditional PDF:
  - Consider event $\mathcal{A} = \{x \leq X \leq x + \delta\}$; by (28) it follows that

$$\Pr(x \leq X \leq x + \delta | \mathcal{B}) = \frac{\Pr(\{x \leq X \leq x + \delta\} \cap \mathcal{B})}{\Pr(\mathcal{B})} = \begin{cases} \frac{f_X(x) \cdot \delta}{\Pr(\mathcal{B})}, & x \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases}$$

  Thus, it follows that the conditional PDF given event $\mathcal{B}$ is

$$f_{X|\mathcal{B}}(x) = \begin{cases} \frac{f_X(x)}{\Pr(\mathcal{B})}, & x \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases}$$

  - Notational convention for a conditional PDF given event $\mathcal{B}$ is $f_{X|\mathcal{B}}$.
  - Note that updated model satisfies all axioms, e.g. $\int_{-\infty}^{\infty} f_{X|\mathcal{B}}(x)dx = 1$.
- The conditional expectation of function $g(X)$ given event $\mathcal{B}$ can be defined as

$$E[g(X)|\mathcal{B}] = \int_{-\infty}^{\infty} g(x) \cdot f_{X|\mathcal{B}}(x)dx. \quad (32)$$

  - Same as unconditional definition, but using conditional PDF

# Example: Exponential PDF

- Consider r.v. $X \sim f_X(x) = 2e^{-2x}$, $x \geq 0$, and conditioning event $B = \{X > 1\}$.

$$\Pr(B) = \int_{x \in B} f_X(x)dx = \int_1^{\infty} 2e^{-2x}dx = 2\int_{-2}^{-\infty} e^u \frac{du}{-2} = \int_{-\infty}^{-2} e^u du = e^u\big|_{-\infty}^{-2} = e^{-2}$$

where third equality uses change of variables $u = -2x$, $du = -2dx$. Thus,

$$f_{X|B}(x) = \begin{cases} \dfrac{f_X(x)}{\Pr(B)}, & x \in B \\ 0, & \text{Otherwise} \end{cases} = \begin{cases} 2e^{2-2x}, & x > 1 \\ 0, & \text{Otherwise.} \end{cases}$$

- The conditional mean of $X$ given event $\{X > 1\}$ can now be determined:

$$E[X|B] = \int_{-\infty}^{\infty} x \cdot f_{X|B}(x)dx = 2e^2 \int_1^{\infty} x \cdot e^{-2x}dx$$

$$= 2e^2 \left[ \frac{-x}{2} e^{-2x} \Big|_1^{\infty} - \int_1^{\infty} \frac{-1}{2} e^{-2x}dx \right] = 2e^2 \left[ \frac{1}{2}e^{-2} + \frac{1}{4}e^{-2} \right] = \frac{3}{2}$$

where we use integration by parts with $u = x$, $dv = e^{-2x}dx$, $v = \frac{-1}{2}e^{-2x}$, $du = dx$.

# Conditioning on Multiple Continuous Random Variables*

- Given cont. r.v.'s $(X, Y)$ and $\mathcal{B} \subseteq S_{X,Y}$ with $\Pr(\mathcal{B}) > 0$, define a conditional joint PDF:
  - Consider event $\mathcal{A} = \{x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y\}$; by (28) we have

$$\Pr(x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y | \mathcal{B}) = \frac{\Pr(\{x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y\} \cap \mathcal{B})}{\Pr(\mathcal{B})}$$
$$= \begin{cases} \dfrac{f_{X,Y}(x,y) \cdot \delta_x \delta_y}{\Pr(\mathcal{B})}, & (x,y) \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases}$$

  Thus, it follows that the conditional joint PDF given event $\mathcal{B}$ is

$$f_{X,Y|\mathcal{B}}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{\Pr(\mathcal{B})}, & (x,y) \in \mathcal{B} \\ 0, & \text{Otherwise} \end{cases}$$

  - Notational convention for a conditional PDF given event $\mathcal{B}$ is $f_{X,Y|\mathcal{B}}$
  - Updated model satisfies all axioms, e.g. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y|\mathcal{B}}(x,y) dx dy = 1$.
- The conditional expectation of function $g(X, Y)$ given event $\mathcal{B}$ can be defined as

$$E[g(X,Y)|\mathcal{B}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y|\mathcal{B}}(x,y) dx dy.$$

  - Same as unconditional definition, but using conditional PDF

# Example: 2-D Uniform Joint PDF

- Recall 2-D uniform r.v. described in (5), and consider finding for event $E_2 = \{X > Y\}$ the conditional joint PDF $f_{X,Y|E_2}(x)$.

$$f_{X,Y|E_2}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{\Pr(E_2)}, & (x,y) \in E_2 \\ 0, & \text{Otherwise} \end{cases} = \begin{cases} 2, & 0 \leq y < x \leq 1 \\ 0, & \text{Otherwise} \end{cases}$$

where $\Pr(E_2)$ was calculated in (6).

- The conditional marginal PDFs given $E_2$ are obtained as

$$f_{X|E_2}(x) = \int_{-\infty}^{\infty} f_{X,Y|E_2}(x,y)dy = \begin{cases} 2\displaystyle\int_0^x dy = 2x, & 0 \leq x \leq 1 \\ 0, & \text{Otherwise} \end{cases}$$

$$f_{Y|E_2}(y) = \int_{-\infty}^{\infty} f_{X,Y|E_2}(x,y)dx = \begin{cases} 2\displaystyle\int_y^1 dx = 2(1-y), & 0 \leq y \leq 1 \\ 0, & \text{Otherwise} \end{cases}$$

# Example: 2-D Uniform Joint PDF Cont.

- Conditional expectations given $E_2$ follow as

$$\begin{aligned} E[X|E_2] &= \int_{-\infty}^{\infty} x \cdot f_{X|E_2}(x)dx = \int_0^1 x \cdot 2xdx = 2\frac{x^3}{3}\Big|_0^1 = \frac{2}{3}. \\ E[Y|E_2] &= \int_{-\infty}^{\infty} y \cdot f_{Y|E_2}(y)dy = \int_0^1 y \cdot 2(1-y)dy = 2\int_0^1 (y-y^2)dy \\ &= 2\left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_0^1 = 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{2}{6} = \frac{1}{3}. \\ E[XY|E_2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f_{X,Y|E_2}(x,y)dxdy = \int_0^1 dx \int_0^x dy \cdot xy \cdot 2 \\ &= \int_0^1 2xdx \cdot \frac{y^2}{2}\Big|_0^x = \int_0^1 x^3 dx = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4}. \end{aligned}$$

- The conditional covariance follows from

$$\text{Cov}[X,Y|E_2] = E[XY|E_2] - E[X|E_2] \cdot E[Y|E_2] = \frac{1}{4} - \frac{2}{9} = \frac{9-8}{36} = \frac{1}{36}$$

# Conditioning on a Discrete Random Variable

- Consider discrete paired r.v.'s $(X, Y) \sim P_{X,Y}(x, y)$ and the events
  - $\mathcal{A} = \{X = x\}$ and $\mathcal{B} = \{Y = y\}$ where $(x, y) \in \mathcal{S}_{X,Y}$

$$\Pr(\mathcal{A}|\mathcal{B}) = \Pr(\{X = x\} \mid \{Y = y\}) = \frac{\Pr(\{X = x\} \cap \{Y = y\})}{\Pr(Y = y)}$$

$$= \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} = \begin{cases} \frac{P_{X,Y}(x,y)}{P_Y(y)}, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

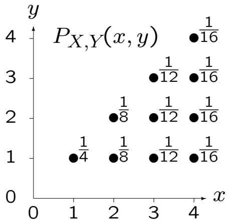
- Thus, the conditional PMF of $X$ given event $Y = y$ is

$$P_{X|Y}(x|y) = \begin{cases} \dfrac{P_{X,Y}(x, y)}{P_Y(y)}, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

- Similarly, the conditional PMF of $Y$ given event $X = x$ is

$$P_{Y|X}(y|x) = \begin{cases} \frac{P_{X,Y}(x,y)}{P_X(x)}, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

# Example 7.15 Problem




Random variables $X$ and $Y$ have the joint PMF $P_{X,Y}(x,y)$, as given in Example 7.11 and repeated in the accompanying graph. Find the conditional PMF of $Y$ given $X = x$ for each $x \in S_X$.

# Conditioning on a Continuous Random Variable

- Consider continuous paired r.v.'s $(X, Y) \sim f_{X,Y}(x, y)$ and the events
  - $\mathcal{A} = \{x \leq X \leq x + \delta_x\}$ and $\mathcal{B} = \{y \leq Y \leq y + \delta_y\}$ where $(x, y) \in \mathcal{S}_{X,Y}$

$$\Pr(\mathcal{A}|\mathcal{B}) = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})} = \Pr(\{x \leq X \leq x + \delta_x\} \mid \{y \leq Y \leq y + \delta_y\})$$

$$= \frac{\Pr(\{x \leq X \leq x + \delta_x\} \cap \{y \leq Y \leq y + \delta_y\})}{\Pr(y \leq Y \leq y + \delta_y)} = \frac{\Pr(x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y)}{\Pr(y \leq Y \leq y + \delta_y)}$$

$$= \begin{cases} \frac{f_{X,Y}(x,y) \cdot \delta_x \delta_y}{f_Y(y) \cdot \delta_y}, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise} \end{cases} = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} \cdot \delta_x, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

- Thus, the conditional PDF of $X$ given event $Y = y$ is

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_Y(y)}, & (x, y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

# Conditioning on a Continuous Random Variable*

- Similarly, the conditional PDF of $Y$ given event $X = x$ is

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_X(x)}, & (x,y) \in \mathcal{S}_{X,Y} \\ 0, & \text{Otherwise.} \end{cases}$$

- **Example: (Prob. 7.4.6. Yates/Goodman)** Consider $(X, Y) \sim f_{X,Y}(x,y)$ where

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{4x+2y}{3}, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Determine $f_{Y|X}(y|x)$ and $E[Y|X]$.

# Some Useful Relationships for PMFs and PDFs

- Recall that the chain rule says $\Pr(\mathcal{AB}) = \Pr(\mathcal{A}) \Pr(\mathcal{B}|\mathcal{A})$ for events $\mathcal{A}, \mathcal{B}$.
- For a joint PMF or PDF, the chain rule implies that:

$$P_{X,Y}(x,y) = P_X(x)P_{Y|X}(y|x) = P_Y(y)P_{X|Y}(x|y),$$
$$f_{X,Y}(x,y) = f_X(x)f_{Y|X}(y|x) = f_Y(y)f_{X|Y}(x|y),$$

  that follow from the definitions of conditional PMFs and PDFs.
- Comparing these to (9), it follows that if $X$ and $Y$ are independent if and only if

$$\begin{array}{lll} P_{Y|X}(y|x) = P_Y(y), & P_{X|Y}(x|y) = P_X(x) & \text{(discrete r.v.'s)} \\ f_{Y|X}(y|x) = f_Y(y), & f_{X|Y}(x|y) = f_X(x) & \text{(continuous r.v.'s)} \end{array}$$

# Expectations Conditioned on a Random Variable

- From (30) and (32) with $\mathcal{B} = \{Y = y\}$, the conditional mean of $g(X, Y)$ given $Y = y$ is

$$E[g(X,Y)|Y=y] = \sum_{x \in S_X} g(x,y) P_{X|Y}(x|y), \qquad E[g(X,Y)|Y=y] = \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y) dx$$

  - Note that if $g(x, y) = x$, then

$$E[X|Y=y] = \sum_{x \in S_X} x \cdot P_{X|Y}(x|y), \quad E[X|Y=y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx.$$

- Similarly, for $\mathcal{B} = \{X = x\}$ that the conditional mean $g(X, Y)$ given $X = x$ is

$$E[g(X,Y)|X=x] = \sum_{y \in S_Y} g(x,y) P_{Y|X}(y|x), \; E[g(X,Y)|X=x] = \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y|x) dy \qquad (33)$$

  - Note that if $g(x, y) = y$, then

$$E[Y|X=x] = \sum_{y \in S_Y} y \cdot P_{Y|X}(y|x), \qquad E[Y|X=x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) dy.$$

- Note $E[g(X, Y)|X = x]$ is a function of $x$; $\quad E[g(X, Y)|Y = y]$ is a function of $y$.

# Chain Rule of Expectation / Law of Total Expectation

- For continuous r.v.'s the expectation of a function $g(X,Y)$ can be written

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_X(x) f_{Y|X}(y|x)dxdy$$

$$= \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y|x)dy\right]dx = \int_{-\infty}^{\infty} f_X(x) E[g(X,Y)|X=x]dx$$

and for discrete r.v.'s we have

$$E[g(X,Y)] = \sum_{x\in S_X}\sum_{y\in S_Y} g(x,y) P_{X,Y}(x,y) = \sum_{x\in S_X}\sum_{y\in S_Y} g(x,y) P_X(x) P_{Y|X}(y|x)$$

$$= \sum_{x\in S_X} P_X(x)\left[\sum_{y\in S_Y} g(x,y) P_{Y|X}(y|x)\right] = \sum_{x\in S_X} P_X(x) E[g(X,Y)|X=x].$$

- Both expression demonstrate that in general that

$$E[g(X,Y)] = E\left[E[g(X,Y)|X]\right]; \qquad E[g(X,Y)] = E\left[E[g(X,Y)|Y]\right].$$

# References


[1] Yates and Goodman, *Probability and Stochastic Processes*, $3^{rd}$ Ed., Wiley, 2014.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Sci., 2002.

[3] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill Inc., 1967.