4.1 TWO RANDOM VARIABLES

Let \((\Omega, \mathcal{F}, P)\) be a probably space and consider two measurable mappings

\[
X : \Omega \rightarrow \mathbb{R}, \\
Y : \Omega \rightarrow \mathbb{R}.
\]

In other words, \(X\) and \(Y\) are two random variables defined on the common probability space \((\Omega, \mathcal{F}, P)\). In order to specify the random variables, we need to determine

\[
P\{(X, Y) \in A\} = P\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}
\]

for every Borel set \(A \subseteq \mathbb{R}^2\). By the properties of the Borel \(\sigma\)-field, it can be shown that it suffices to determine the probabilities of the form

\[
P\{a < X < b, c < Y < d\}, \quad a < b, c < d,
\]
or equivalently, the probabilities of the form

\[
P\{X \leq x, Y \leq y\}, \quad x, y \in \mathbb{R}.
\]

The latter defines their joint cdf

\[
F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\}, \quad x, y \in \mathbb{R},
\]

which is the shaded region in Figure 1.1.

The joint cdf satisfies the following properties:

1. \(F_{X,Y}(x, y) \geq 0\).
2. If \(x_1 \leq x_2\) and \(y_1 \leq y_2\), then

\[
F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2).
\]
3. Limits.

\[
\lim_{x,y \to \infty} F_{X,Y}(x, y) = 1, \\
\lim_{y \to -\infty} F_{X,Y}(x, y) = \lim_{x \to -\infty} F_{X,Y}(x, y) = 0.
\]
2 Pairs of Random Variables

![Diagram of joint pdf of X and Y](image)

**Figure 4.1.** An illustration of the joint pdf of X and Y.

4. **Marginal cdfs.**

\[
\lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x),
\]

\[
\lim_{x \to \infty} F_{X,Y}(x, y) = F_Y(y).
\]

The probability of any (Borel) set can be determined from the joint cdf. For example,

\[
P\{a < X \leq b, \ c < Y \leq d\} = F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c)
\]

as illustrated in Figure 4.2.

We say that X and Y are **statistically independent** or independent in short if for every

![Diagram of P[a < X \leq b, c < Y \leq d]](image)

**Figure 4.2.** An illustration of \(P[a < X \leq b, c < Y \leq d]\).
(Borel) $A$ and $B$,
\[ P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}. \]

Equivalently, $X$ and $Y$ are independent if
\[ F_{X,Y}(x, y) = F_X(x)F_Y(y), \quad x, y \in \mathbb{R}. \]

In the following, we focus on three special cases and discuss the joint, marginal, and conditional distributions of $X$ and $Y$ for each case.

- $X$ and $Y$ are discrete.
- $X$ and $Y$ are continuous.
- $X$ is discrete and $Y$ is continuous (mixed).

### 4.2 Pairs of Discrete Random Variables

Let $X$ and $Y$ be discrete random variables on the same probability space. They are completely specified by their joint pmf
\[ p_{X,Y}(x, y) = P\{X = x, Y = y\}, \quad x \in \mathcal{X}, \ y \in \mathcal{Y}. \]

By the axioms of probability,
\[ \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y) = 1. \]

We use the law of total probability to find the marginal pmf of $X$:
\[ p_X(x) = \sum_{y \in \mathcal{Y}} p(x, y), \quad x \in \mathcal{X}. \]

The conditional pmf of $X$ given $Y = y$ is defined as
\[ p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad p_Y(y) \neq 0, \ x \in \mathcal{X}. \]

Check that if $p_Y(y) \neq 0$, then $p_{X|Y}(x|y)$ is a pmf for $X$.

**Example 4.1.** Consider the pmf $p(x, y)$ described by the following table

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-3$</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>$y$</td>
<td>$\frac{1}{8}$</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{8}$</td>
<td>0</td>
</tr>
</tbody>
</table>
Pairs of Random Variables

Then,

\[
P_X(x) = \begin{cases} 
1/4 & x = 0, \\
3/8 & x = 1, \\
3/8 & x = 2.5, 
\end{cases}
\]

and

\[
P_{X|Y}(x|2) = \begin{cases} 
1/2 & x = 0, \\
1/2 & x = 1.
\end{cases}
\]

**Chain rule.** \(p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y|x) = p_Y(y)p_{X|Y}(x|y)\).

**Independence.** \(X\) and \(Y\) are independent if

\[
p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad x \in \mathcal{X}, \; y \in \mathcal{Y},
\]

which is equivalent to

\[
p_{X|Y}(x|y) = p_X(x), \quad x \in \mathcal{X}, \; p_Y(y) \neq 0.
\]

**Law of total probability.** For any event \(A\),

\[
P(A) = \sum_x p_X(x)P(A \mid X = x).
\]

**Bayes rule.** Given \(p_X(x)\) and \(p_{Y|X}(y|x)\) for every \((x, y) \in \mathcal{X} \times \mathcal{Y}'\), we can find \(p_{X|Y}(x|y)\) as

\[
P_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y|x)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y|x)}{\sum_{u \in \mathcal{X}'} p_{X,Y}(u, y)} = \frac{p_X(x)p_{Y|X}(y|x)}{\sum_{u \in \mathcal{X}'} p_X(u)p_{Y|X}(y|u)}.
\]

The final formula is entirely in terms of the known quantities \(p_X(x)\) and \(p_{Y|X}(y|x)\).

**Example 4.2 (Binary symmetric channel).** Consider the binary communication channel in Figure 4.3. The bit sent is \(X \sim \text{Bern}(p)\), \(p \in [0, 1]\), and the bit received is

\[
Y = (X + Z) \mod 2 = X \oplus Z,
\]

where the noise \(Z \sim \text{Bern}(\epsilon)\), \(\epsilon \in [0, 1/2]\), is independent of \(X\). We find \(p_{X|Y}(x|y)\), \(p_Y(y)\).
and $P[X \neq Y]$, namely, the probability of error. First, we use the Bayes rule

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{u \in X} p_{Y|X}(y|u)p_X(u)}.$$ 

We know $p_X(x)$, but we need to find $p_{Y|X}(y|x)$:

$$p_{Y|X}(y|x) = P[Y = y | X = x] = P[X \oplus Z = y | X = x]$$

$$= P[X = y \oplus x | X = x] = P[Y = y | X = x]$$

$$= p_Z(y \oplus x) \quad \text{since } Z \text{ and } X \text{ are independent}$$

Therefore

$$p_{Y|X}(0 | 0) = p_Z(0 \oplus 0) = p_Z(0) = 1 - \epsilon,$$

$$p_{Y|X}(0 | 1) = p_Z(0 \oplus 1) = p_Z(1) = \epsilon,$$

$$p_{Y|X}(1 | 0) = p_Z(1 \oplus 0) = p_Z(1) = \epsilon,$$

$$p_{Y|X}(1 | 1) = p_Z(1 \oplus 1) = p_Z(0) = 1 - \epsilon.$$

Plugging into the Bayes rule equation, we obtain

$$p_{X|Y}(0|0) = \frac{p_{Y|X}(0|0)p_X(0)}{p_{Y|X}(0|0)p_X(0) + p_{Y|X}(0|1)p_X(1)}p_X(0) = \frac{(1 - \epsilon)(1 - p)}{(1 - \epsilon)(1 - p) + \epsilon p}.$$ 

$$p_{X|Y}(1|0) = 1 - p_{X|Y}(0|0) = \frac{\epsilon p}{(1 - \epsilon)(1 - p) + \epsilon p}.$$ 

$$p_{X|Y}(0|1) = \frac{p_{Y|X}(1|0)}{p_{Y|X}(1|0)p_X(0) + p_{Y|X}(1|1)p_X(1)}p_X(0) = \frac{\epsilon(1 - p)}{(1 - \epsilon)p + \epsilon(1 - p)}.$$ 

$$p_{X|Y}(1|1) = 1 - p_{X|Y}(0|1) = \frac{(1 - \epsilon)p}{(1 - \epsilon)p + \epsilon(1 - p)}.$$ 

We already found $p_Y(y)$ as

$$p_Y(y) = p_{Y|X}(y|0)p_X(0) + p_{Y|X}(y|1)p_X(1)$$

$$= \begin{cases} 
(1 - \epsilon)(1 - p) + \epsilon p & \text{for } y = 0, \\
\epsilon(1 - p) + (1 - \epsilon)p & \text{for } y = 1. 
\end{cases}$$
Now to find the probability of error $P\{X \neq Y\}$, consider

$$P\{X \neq Y\} = p_{X,Y}(0, 1) + p_{X,Y}(1, 0)$$

$$= p_{Y|X}(1|0)p_X(0) + p_{Y|X}(0|1)p_X(1)$$

$$= \epsilon(1 - p) + \epsilon p$$

$$= \epsilon.$$

Alternatively, $P\{X \neq Y\} = P\{Z = 1\} = \epsilon$. An interesting special case is $\epsilon = 1/2$, whence $P\{X \neq Y\} = 1/2$, which is the worst possible (no information is sent), and

$$p_Y(0) = \frac{1}{2}p + \frac{1}{2}(1 - p) = \frac{1}{2} = p_Y(1).$$

Therefore $Y \sim \text{Bern}(1/2)$, regardless of the value of $p$. In this case, the bit sent $X$ and the bit received $Y$ are independent (check!).

### 4.3 Pairs of Continuous Random Variables

#### 4.3.1 Joint and Marginal Densities

We say that $X$ and $Y$ are jointly continuous random variables if their joint cdf is continuous in both $x$ and $y$. In this case, we can define their joint pdf, provided that it exists, as the function $f_{X,Y}(x, y)$ such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) \, du \, dv, \quad x, y \in \mathbb{R}.$$

If $F_{X,Y}(x, y)$ is differentiable in $x$ and $y$, then

$$f_{X,Y}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

$$= \lim_{\Delta x, \Delta y \to 0} \frac{F_{X,Y}(x + \Delta x, y + \Delta y) - F_{X,Y}(x + \Delta x, y) - F_{X,Y}(x, y + \Delta y) + F_{X,Y}(x, y)}{\Delta x \Delta y}$$

$$= \lim_{\Delta x, \Delta y \to 0} \frac{P\{x < X \leq x + \Delta x, y < Y \leq y + \Delta y\}}{\Delta x \Delta y}.$$

The joint pdf $f_{X,Y}(x, y)$ satisfies the following properties:

1. $f_{X,Y}(x, y) \geq 0$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \, dy = 1$.
3. The probability of any set $A \subset \mathbb{R}^2$ can be calculated by integrating the joint pdf over $A$:

$$P\{(X, Y) \in A\} = \int_{(x,y) \in A} f_{X,Y}(x, y) \, dx \, dy.$$
The marginal pdf of $X$ can be obtained from the joint pdf via the law of total probability:

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

To see this, recall that

$$g(x) = \frac{d}{dx} \int_{-\infty}^{x} g(u) \, du$$

and consider

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, y) \, dy \, du$$

with $g(u) = \int_{-\infty}^{\infty} f_{X,Y}(u, y) \, dy$. We then differentiate $F_X(x)$ with respect to $x$. Alternatively, we can consider the following figure 

![Graph](image)

and note that

$$f_X(x) = \lim_{\Delta x \to 0} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{1}{\Delta x} \lim_{\Delta y \to 0} \sum_{n=-\infty}^{\infty} P\{x < X \leq x + \Delta x, n\Delta y < Y \leq (n+1)\Delta y\}$$

$$= \lim_{\Delta x \to 0} \frac{1}{\Delta x} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \, \Delta x$$

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

**Example 4.3.** Let $(X, Y) \sim f(x, y)$, where

$$f(x, y) = \begin{cases} 
  c & x \geq 0, \ y \geq 0, \ x + y \leq 1, \\
  0 & \text{otherwise.}
\end{cases}$$

We find $c$, $f_Y(y)$, and $P\{Y \leq 2X\}$. Note first that

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = \int_{0}^{1} \int_{0}^{1-y} c \, dx \, dy = c \int_{0}^{1} (1 - y) \, dy = \frac{1}{2} c.$$
Hence, \( c = 2 \). To find \( f_Y(y) \), we use the law of total probability
\[
f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx
\]
\[
= \begin{cases} 
  \int_0^{(1-y)} 2 \, dx & 0 \leq y \leq 1, \\
  0 & \text{otherwise}
\end{cases}
\]
\[
= \begin{cases} 
  2(1 - y) & 0 \leq y \leq 1, \\
  0 & \text{otherwise}.
\end{cases}
\]

To find the probability of the set \( \{Y \leq 2X\} \), we first sketch the set

From the figure we find that
\[
P \left\{ X \geq \frac{1}{2} Y \right\} = \int_{\{(x,y):x=\frac{1}{2}y\}} f_{X,Y}(x, y) \, dx \, dy
\]
\[
= \int_0^2 \int_{(1-y)}^2 dx \, dy = \frac{2}{3}.
\]
Recall that \( X \) and \( Y \) are independent if \( f_{X,Y}(x, y) = f_X(x) f_Y(y) \) for every \( x, y \). But \( f_{X,Y}(0, 1) = 2 \) and \( f_X(0) f_Y(1) = 0 \), so \( X \) and \( Y \) are not independent. Alternatively, consider
\[
f_{X,Y}(1/2, 1/2) = 2 \neq 1 \cdot 1 = f_X(1/2) f_Y(1/2)
\]
to see that \( X \) and \( Y \) are dependent.
4.3.2 Conditional Densities

Let $X$ and $Y$ be continuous random variables with joint pdf $f_{X,Y}(x, y)$. We wish to define

$$F_{Y|X}(y \mid X = x) = P\{Y \leq y \mid X = x\}.$$ 

We cannot define the above conditional probability as

$$P\{Y \leq y, X = x\} \over P\{X = x\}$$

since both the numerator and the denominator are equal to zero. Instead, we define the conditional probability for continuous random variables as a limit

$$F_{Y|X}(y \mid x) = \lim_{\Delta x \to 0} \frac{P\{x < Y \leq y \mid X = x\}}{P\{x \leq x + \Delta x\}}$$

$$= \lim_{\Delta x \to 0} \frac{\int_y^{x+\Delta x} f_{X,Y}(u, v) du dv}{f_X(x) \Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\int_y^{x+\Delta x} f_{X,Y}(u, v) du}{f_X(x) \Delta x}$$

$$= \int_{-\infty}^{y} \frac{f_{X,Y}(x, u)}{f_X(x)} du.$$

By differentiating w.r.t. $y$, we thus define the conditional pdf as

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{if } f_X(x) \neq 0.$$ 

It can be readily checked that $f_{Y|X}(y \mid x)$ is a valid pdf and

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} f_{Y|X}(u \mid x) du$$

is a valid cdf for every $x$ such that $f_X(x) > 0$. Sometimes we use the notation

$$Y \mid \{X = x\} \sim f_{Y|X}(y \mid x)$$

to denote $Y$ has a conditional pdf $f_{Y|X}(y \mid x)$ given $X = x$.

Chain rule. $f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y \mid x) = f_Y(y) f_{X|Y}(x \mid y)$.

Independence. $X$ and $Y$ are independent iff

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad x, y \in \mathbb{R},$$

or equivalently,

$$f_{X|Y}(x \mid y) = f_X(x), \quad x \in \mathbb{R}, \ f_Y(y) \neq 0.$$
**Law of total probability.** For any event $A$, 

$$P(A) = \int_{-\infty}^{\infty} f_X(x) P(A|X = x) \, dx.$$ 

**Bayes rule.** Given $f_X(x)$ and $f_{Y|X}(y|x)$, we can find 

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) \, f_X(x)}{f_Y(y)} = \frac{\int_{-\infty}^{\infty} f_{X,Y}(u,y) \, du}{\int_{-\infty}^{\infty} f_X(u) \, f_{Y|X}(y|u) \, du}.$$ 

**Example 4.4.** As in Example 4.3, let $f(x, y)$ be defined as 

$$f(x, y) = \begin{cases} 2 & x \geq 0, \ y \geq 0, \ x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

We already know that 

$$f_Y(y) = \begin{cases} 2(1 - y) & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore 

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{1}{1 - y} & 0 \leq y < 1, \ 0 \leq x \leq 1 - y, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $X|\{Y = y\} \sim \text{Unif}[0, 1 - y].$
Example 4.5. Let \( \Lambda \sim \text{Unif}[0, 1] \) and the conditional pdf of \( X \) given \( \{\Lambda = \lambda\} \) be
\[
f_{X \mid \Lambda}(x \mid \lambda) = \lambda e^{-\lambda x}, \quad 0 < \lambda \leq 1,
\]
i.e., \( X \mid \{\Lambda = \lambda\} \sim \text{Exp}(\lambda) \). Then, by the Bayes rule,
\[
f_{\Lambda \mid X}(\lambda \mid 3) = \frac{f_{X \mid \Lambda}(3 \mid \lambda) f_{\Lambda}(\lambda)}{\int_0^1 f_{\Lambda}(u) f_{X \mid \Lambda}(3 \mid u) \, du} = \begin{cases} \frac{\lambda e^{-3 \lambda}}{3(1 - 4e^{-3})} & 0 < \lambda \leq 1, \\ 0 & \text{otherwise.} \end{cases}
\]

4.4 PAIRS OF MIXED RANDOM VARIABLES

Let \( X \) be a discrete random variable with pmf \( p_X(x) \). For each \( x \) with \( p_X(x) \neq 0 \), conditioned on the event \( \{X = x\} \), let \( Y \) be a continuous random variable with conditional cdf \( F_{Y \mid X}(y \mid x) \) and conditional pdf
\[
f_{Y \mid X}(y \mid x) = \frac{\partial}{\partial y} F_{Y \mid X}(y \mid x),
\]
provided that the derivative is well-defined. Then, by the law of total probability,
\[
F_Y(y) = \mathbb{P}\{Y \leq y\} = \sum_x \mathbb{P}\{Y \leq y \mid X = x\} p_X(x)
= \sum_x F_{Y \mid X}(y \mid x) p_X(x)
= \sum_x \int_{-\infty}^y f_{Y \mid X}(u \mid x) \, du p_X(x),
\]
and hence
\[ f_Y(y) = \frac{d}{dy} F_Y(y) = \sum_x f_{Y|X}(y|x)p_X(x). \]

The conditional pmf of \( X \) given \( Y = y \) can be defined as a limit:
\[
P_{X|Y}(x|y) = \lim_{\Delta y \to 0} \frac{P(X = x, y < Y \leq y + \Delta y)}{P(y < Y \leq y + \Delta y)}
= \lim_{\Delta y \to 0} \frac{P_X(x)P\{y < Y \leq y + \Delta y|X = x\}}{P\{y < Y \leq y + \Delta y\}}
= \lim_{\Delta y \to 0} \frac{p_X(x)f_{Y|X}(y|x)\Delta y}{f_Y(y)\Delta y}
= \frac{f_{Y|X}(y|x)}{f_Y(y)}p_X(x).
\]

**Chain rule.** \( p_X(x)f_{Y|X}(y|x) = f_Y(y)p_{X|Y}(x|y). \)

**Bayes rule.** Given \( p_X(x) \) and \( f_{Y|X}(y|x) \),
\[
P_{X|Y}(x|y) = \frac{p_X(x)f_{Y|X}(y|x)}{\sum_u p_X(u)f_{Y|X}(y|u)}.
\]

**Example 4.6 (Additive Gaussian noise channel).** Consider the communication channel in Figure 4.4. The signal transmitted is a binary random variable
\[
X = \begin{cases} 
+\sqrt{P} & \text{w.p. } p, \\
-\sqrt{P} & \text{w.p. } 1 - p.
\end{cases}
\]

The received signal, also called the *observation*, is
\[
Y = X + Z,
\]
where \( Z \sim N(0, N) \) is additive noise and independent of \( X \). First note that \( Y | \{X = x\} \sim \)
\[
Z \sim N(0, N)
\]

![Figure 4.4. Additive Gaussian noise channel.](image-url)
N(x, N). To see this, consider
\[
P\{Y \leq y \mid X = x\} = P\{X + Z \leq y \mid X = x\} = P\{Z \leq y - x \mid X = x\} = P\{Z \leq y - x\},
\]
which, by taking derivative w.r.t. y, implies that
\[
f_{Y|X}(y|x) = f_Z(y - x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-x)^2}{2N}}.
\]
In other words,
\[
Y \mid \{X = +\sqrt{P}\} \sim N(+\sqrt{P}, N) \quad \text{and} \quad Y \mid \{X = -\sqrt{P}\} \sim N(-\sqrt{P}, N).
\]
Next by the law of total probability,
\[
f_Y(y) = p_X(+)\sqrt{P} f_{Y|X}(y|+\sqrt{P}) + p_X(-\sqrt{P}) f_{Y|X}(y|-\sqrt{P})
\]
\[
= p\frac{1}{\sqrt{2\pi N}} e^{-\frac{(y+\sqrt{P})^2}{2N}} + (1-p)\frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-\sqrt{P})^2}{2N}}.
\]
Finally, by the Bayes rule,
\[
P_{X\mid Y}(+)\sqrt{P} | y) = \frac{p}{\sqrt{2\pi N}} e^{-\frac{(y+\sqrt{P})^2}{2N}} + \frac{(1-p)}{\sqrt{2\pi N}} e^{-\frac{(y-\sqrt{P})^2}{2N}}
\]
\[
= \frac{pe^{\frac{\sqrt{P}}{N}}}{pe^{\frac{\sqrt{P}}{N}} + (1-p)e^{-\frac{\sqrt{P}}{N}}}
\]
and
\[
P_{X\mid Y}(-\sqrt{P} | y) = \frac{(1-p)e^{\frac{\sqrt{P}}{N}}}{pe^{\frac{\sqrt{P}}{N}} + (1-p)e^{-\frac{\sqrt{P}}{N}}}.
\]

4.5 Signal Detection

Consider the general digital communication system depicted in Figure 4.5 where the signal sent is \(X \sim p_X(x), x \in \mathcal{X} = \{x_1, x_2, \ldots, x_n\}\), and the observation (received signal) is
\[
Y \mid \{X = x\} \sim f_{Y|X}(y|x).
\]
A detector is a mapping \(d : \mathbb{R} \to \mathcal{X}\) that generates an estimate \(\hat{X} = d(Y)\) of the input signal \(X\). We wish to find an optimal detector \(d^*(y)\) that minimizes the probability of error
\[
P_e := P(\hat{X} \neq X) = P(d(Y) \neq X).
Suppose that there is no observation and we would like to find the best guess \( \hat{d} \) of \( X \), that is,
\[
d^* = \arg \min_d P[d \neq X].
\]
Clearly, the best guess is the one with the largest probability, that is,
\[
d^* = \arg \max_x p_X(x).
\quad (4.1)
\]
This simple observation continues to hold with the observation \( Y = y \). Define the maximum a posteriori probability (MAP) detector as
\[
d^*(y) = \arg \max_x p_{X|Y}(x|y).
\quad (4.2)
\]
Then, by (4.1)
\[
P[d^*(y) \neq X | Y = y] \leq P[d(y) \neq X | Y = y]
\]
for any detector \( d(y) \). Consequently,
\[
P[d^*(Y) \neq X] = \int P[d^*(y) \neq X | Y = y]
\leq \int P[d(y) \neq X | Y = y]
= P[d(Y) \neq X],
\]
and the MAP detector minimizes the probability of error \( P_e \).
When \( X \) is uniform on \( \{x_1, x_2, \ldots, x_n\} \), that is, \( p_X(x_1) = p_X(x_2) = \cdots = p_X(x_n) = \frac{1}{n} \),
\[
p_{X|Y}(x|y) = \frac{p_X(x) f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)}{n f_Y(y)},
\]
which depends on \( x \) only through the likelihood \( f_{Y|X}(y|x) \) for a given \( y \). Then, the MAP detection rule in (4.2) simplifies as the maximum likelihood (ML) detection rule
\[
d^*(y) = \arg \max_x f_{Y|X}(y|x).
\]
Example 4.7. We revisit the additive Gaussian noise channel in Example 4.6 with signal

\[ X = \begin{cases} +\sqrt{P} & \text{w.p. } \frac{1}{2} \\ -\sqrt{P} & \text{w.p. } \frac{1}{2} \end{cases} \]

independent noise \( Z \sim N(0, N) \), and output \( Y = X + Z \). The MAP detector is

\[
d^*(y) = \begin{cases} +\sqrt{P} & \text{if } p_{X|Y}(+\sqrt{P}|y) > p_{X|Y}(-\sqrt{P}|y), \\ -\sqrt{P} & \text{otherwise}. \end{cases}
\]

Since the two signals are equally likely, the MAP detection rule reduces to the ML detection rule

\[
d^*(y) = \begin{cases} +\sqrt{P} & \text{if } f_{Y|X}(y + \sqrt{P}) > f_{Y|X}(y - \sqrt{P}), \\ -\sqrt{P} & \text{otherwise}. \end{cases}
\]

From the figure above and using the Gaussian pdf, the MAP/ML detector reduces to the minimum distance detector

\[
d^*(y) = \begin{cases} +\sqrt{P} & (y - \sqrt{P})^2 < (y - (-\sqrt{P}))^2, \\ -\sqrt{P} & \text{otherwise}, \end{cases}
\]

which further simplifies to

\[
d^*(y) = \begin{cases} +\sqrt{P} & y > 0, \\ -\sqrt{P} & y \leq 0. \end{cases}
\]

Now the minimum probability of error is

\[
P_e^* = P\{d^*(Y) \neq X\}
\]

\[
= P\{X = \sqrt{P}\} P\{d^*(Y) = -\sqrt{P} | X = \sqrt{P}\} + P\{X = -\sqrt{P}\} P\{d^*(Y) = \sqrt{P} | X = -\sqrt{P}\}
\]

\[
= \frac{1}{2} P\{Y \leq 0 | X = \sqrt{P}\} + \frac{1}{2} P\{Y > 0 | X = -\sqrt{P}\}
\]

\[
= \frac{1}{2} P\{Z \leq -\sqrt{P}\} + \frac{1}{2} P\{Z > \sqrt{P}\}
\]

\[
= Q\left(\frac{\sqrt{P}}{\sqrt{N}}\right).
\]

Note that the probability of error is a decreasing function of \( P/N \), which is often called the signal-to-noise ratio (SNR).
4.6 FUNCTIONS OF TWO RANDOM VARIABLES

Let \((X, Y) \sim f(x, y)\) and let \(g(x, y)\) be a differentiable function. To find the pdf of \(Z = g(X, Y)\), we first find the inverse image of \(\{Z \leq z\}\) to compute its probability expressed as a function of \(z\), i.e., \(F_Z(z)\), and then take the derivative.

Example 4.8. Suppose that \(X \sim f_X(x)\) and \(Y \sim f_Y(y)\) are independent. Let \(Z = X + Y\). Then

\[
F_Z(z) = P[Z \leq z] = P[X + Y \leq z] = \int_{-\infty}^{\infty} P[X + Y \leq z \mid X = x] f_X(x) \, dx
\]

\[
= \int_{-\infty}^{\infty} P[X \leq z - y] f_Y(y) \, dy
\]

\[
= \int_{-\infty}^{\infty} F_X(z - y) f_Y(y) \, dy.
\]

Taking derivative w.r.t. \(z\), we have

\[
f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) \, dx,
\]

which is the convolution of \(f_X(x)\) and \(f_Y(y)\). For example, if \(X \sim N(\mu_X, \sigma_X^2)\) and \(Y \sim N(\mu_Y, \sigma_Y^2)\) are independent, then it can be readily checked that

\[
Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).
\]

A similar result also holds for the sum of two independent discrete random variables (replacing pdfs with pmfs and integrals with sums). For example, if \(X \sim \text{Poisson}(\lambda_1)\) and \(Y \sim \text{Poisson}(\lambda_2)\) are independent, then \(Z = X + Y \sim \text{Poisson}(\lambda_1 \ast \lambda_2) = \text{Poisson}(\lambda_1 + \lambda_2)\). The property that the sum of two independent random variables with the same distribution has the same distribution, which is obeyed by Gaussian and Poisson random variables, is referred to as infinite divisibility. For example, a Poisson(\(\lambda\)) r.v. can be written as the sum of any number of independent Poisson(\(\lambda_i\)) r.v.s, as long as \(\sum_i \lambda_i = \lambda\).

It is sometimes easier to work with cdfs first to find the pdf of a function of \(X\) and \(Y\) (especially when \(g(x, y)\) is not differentiable).

Example 4.9 (Minimum and maximum). Let \(X \sim f_X(x)\) and \(Y \sim f_Y(y)\) be independent. Define

\[
U = \max\{X, Y\} \quad \text{and} \quad V = \min\{X, Y\}.
\]
To find the pdf of $U$, we first find its cdf
\[ F_U(u) = P[U \leq u] = P[X \leq u, Y \leq u] = F_X(u)F_Y(u). \]

Using the product rule for derivatives,
\[ f_U(u) = f_X(u)F_Y(u) + f_Y(u)F_X(u). \]

Now to find the pdf of $V$, consider
\[ P[V > v] = P[X > v, Y > v] \implies 1 - F_V(v) = (1 - F_X(v))(1 - F_Y(v)). \]

Thus
\[ f_V(v) = f_X(v) + f_Y(v) - f_X(v)F_Y(v) - f_Y(v)F_X(v). \]

More generally, the joint pdf of $(U, V)$ can be found by similar arguments; see Problem ??.

**PROBLEMS**

4.1. *Geometric with conditions.* Let $X$ be a geometric random variable with pmf
\[ p_X(k) = p(1 - p)^{k-1}, \quad k = 1, 2, \ldots. \]

Find and plot the conditional pmf $p_X(k|A) = P[X = k|X \in A]$ if:
(a) $A = \{X > m\}$ where $m$ is a positive integer.
(b) $A = \{X < m\}$.
(c) $A = \{X\text{ is an even number}\}$.

Comment on the shape of the conditional pmf of part (a).

4.2. *Conditional cdf.* Let $A$ be a nonzero probability event $A$. Show that
(a) $P(A) = P(A|X \leq x)F_X(x) + P(A|X > x)(1 - F_X(x))$.
(b) $F_X(x|A) = \frac{P(A|X \leq x)}{P(A)}F_X(x)$.

4.3. *Joint cdf or not.* Consider the function
\[ G(x, y) = \begin{cases} 1 & \text{if } x + y \geq 0, \\ 0 & \text{otherwise}. \end{cases} \]

Can $G$ be a joint cdf for a pair of random variables? Justify your answer.

4.4. *Time until the $n$-th arrival.* Let the random variable $N(t)$ be the number of packets arriving during time $(0, t]$. Suppose $N(t)$ is Poisson with pmf
\[ p_N(n) = \frac{(\lambda t)^n}{n!}e^{-\lambda t} \quad \text{for } n = 0, 1, 2, \ldots. \]

Let the random variable $Y$ be the time to get the $n$-th packet. Find the pdf of $Y$. 
4.5. **Diamond distribution.** Consider the random variables $X$ and $Y$ with the joint pdf

$$f_{X,Y}(x, y) = \begin{cases} c, & \text{if } |x| + |y| \leq 1/\sqrt{2}, \\ 0, & \text{otherwise,} \end{cases}$$

where $c$ is a constant.

(a) Find $c$.

(b) Find $f_X(x)$ and $f_{X|Y}(x|y)$.

(c) Are $X$ and $Y$ independent random variables? Justify your answer.

(d) Define the random variable $Z = (|X| + |Y|)$. Find the pdf $f_Z(z)$.

4.6. **Coin with random bias.** You are given a coin but are not told what its bias (probability of heads) is. You are told instead that the bias is the outcome of a random variable $P \sim U[0, 1]$. To get more information about the coin bias, you flip it independently 10 times. Let $X$ be the number of heads you get. Thus $X \sim \text{Binom}(10, P)$. Assuming that $X = 9$, find and sketch the a posteriori probability of $P$, i.e., $f_{P|X}(p|9)$.

4.7. **First available teller.** Consider a bank with two tellers. The service times for the tellers are independent exponentially distributed random variables $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$, respectively. You arrive at the bank and find that both tellers are busy but that nobody else is waiting to be served. You are served by the first available teller once he/she is free.

(a) What is the probability that you are served by the first teller?

(b) Let the random variable $Y$ denote your waiting time. Find the pdf of $Y$.

4.8. **Optical communication channel.** Let the signal input to an optical channel be given by

$$X = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 10 & \text{with probability } \frac{1}{2} \end{cases}.$$  

The conditional pmf of the output of the channel $Y|[X = 1] \sim \text{Poisson}(1)$, i.e., Poisson with intensity $\lambda = 1$, and $Y|[X = 10] \sim \text{Poisson}(10)$.

(a) Show that the MAP rule reduces to

$$D(y) = \begin{cases} 1, & y < y^* \\ 10, & \text{otherwise.} \end{cases}$$

(b) Find $y^*$ and the corresponding probability of error.

4.9. **Iocane or Sennari.** An absent-minded chemistry professor forgets to label two identically looking bottles. One bottle contains a chemical named “Iocane” and the other bottle contains a chemical named “Sennari”. It is well known that the radioactivity level of “Iocane” has the $U[0, 1]$ distribution, while the radioactivity level of “Sennari” has the $\text{Exp}(1)$ distribution.
(a) Let $X$ be the radioactivity level measured from one of the bottles. What is the optimal decision rule (based on the measurement $X$) that maximizes the chance of correctly identifying the content of the bottle?

(b) What is the associated probability of error?

4.10. Independence. Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be two independent discrete random variables.

(a) Show that any two events $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$ are independent.

(b) Show that any two functions of $X$ and $Y$ separately are independent; that is, if $U = g(X)$ and $V = h(Y)$ then $U$ and $V$ are independent.

4.11. Family planning. Alice and Bob choose a number $X$ at random from the set $\{2, 3, 4\}$ (so the outcomes are equally probable). If the outcome is $X = x$, they decide to have children until they have a girl or $x$ children, whichever comes first. Assume that each child is a girl with probability $1/2$ (independent of the number of children and gender of other children). Let $Y$ be the number of children they will have.

(a) Find the conditional pmf $p_{Y|X}(y|x)$ for all possible values of $x$ and $y$.

(b) Find the pmf of $Y$.

4.12. Radar signal detection. The signal for a radar channel $S = 0$ if there is no target and a random variable $S \sim N(0, P)$ if there is a target. Both occur with equal probability. Thus

$$S = \begin{cases} 0, & \text{with probability } \frac{1}{2} \\ X \sim N(0, P), & \text{with probability } \frac{1}{2}. \end{cases}$$

The radar receiver observes $Y = S + Z$, where the noise $Z \sim N(0, N)$ is independent of $S$. Find the optimal decoder for deciding whether $S = 0$ or $S = X$ and its probability of error? Provide your answer in terms of intervals of $y$ and provide the boundary points of the intervals in terms of $P$ and $N$.

4.13. Ternary signaling. Let the signal $S$ be a random variable defined as follows:

$$S = \begin{cases} -1 & \text{with probability } \frac{1}{3} \\ 0 & \text{with probability } \frac{1}{3} \\ +1 & \text{with probability } \frac{1}{3}. \end{cases}$$

The signal is sent over a channel with additive Laplacian noise $Z$, i.e., $Z$ is a Laplacian random variable with pdf

$$f_Z(z) = \frac{\lambda}{2} e^{-\lambda|z|}, \quad -\infty < z < \infty.$$ 

The signal $S$ and the noise $Z$ are assumed to be independent and the channel output is their sum $Y = S + Z$. 
(a) Find \( f_{Y|M}(y|s) \) for \( s = -1, 0, +1 \). Sketch the conditional pdfs on the same graph.

(b) Find the optimal decoding rule \( D(Y) \) for deciding whether \( S \) is \(-1, 0 \) or \(+1\). Give your answer in terms of ranges of values of \( Y \).

(c) Find the probability of decoding error for \( D(y) \) in terms of \( \lambda \).

4.14. Signal or no signal. Consider a communication system that is operated only from time to time. When the communication system is in the "normal" mode (denoted by \( M = 1 \)), it transmits a random signal \( S = X \) with

\[
X = \begin{cases} 
+1, & \text{with probability } 1/2, \\
-1, & \text{with probability } 1/2.
\end{cases}
\]

When the system is in the "idle" mode (denoted by \( M = 0 \)), it does not transmit any signal (\( S = 0 \)). Both normal and idle modes occur with equal probability. Thus

\[
S = \begin{cases} 
X, & \text{with probability } 1/2, \\
0, & \text{with probability } 1/2.
\end{cases}
\]

The receiver observes \( Y = S + Z \), where the ambient noise \( Z \sim U[-1, 1] \) is independent of \( S \).

(a) Find and sketch the conditional pdf \( f_{Y|M}(y|1) \) of the receiver observation \( Y \) given that the system is in the normal mode.

(b) Find and sketch the conditional pdf \( f_{Y|M}(y|0) \) of the receiver observation \( Y \) given that the system is in the idle mode.

(c) Find the optimal decoder \( d^*(y) \) for deciding whether the system is normal or idle. Provide the answer in terms of intervals of \( y \).

(d) Find the associated probability of error.

4.15. Two independent uniform random variables. Let \( X \) and \( Y \) be independently and uniformly drawn from the interval \([0, 1]\).

(a) Find the pdf of \( U = \max(X, Y) \).

(b) Find the pdf of \( V = \min(X, Y) \).

(c) Find the pdf of \( W = U - V \).

(d) Find the pdf of \( Z = (X + Y) \mod 1 \).

(e) Find the probability \( P(|X - Y| \geq 1/2) \).

4.16. Two independent Gaussian random variables. Let \( X \) and \( Y \) be independent Gaussian random variables, both with zero mean and unit variance. Find the pdf of \( |X - Y| \).
4.17. Functions of exponential random variables. Let $X$ and $Y$ be independent exponentially distributed random variables with the same parameter $\lambda$. Define the following three functions of $X$ and $Y$:

$$U = \max(X, Y), \quad V = \min(X, Y), \quad W = U - V.$$ 

(a) Find the joint pdf of $U$ and $V$.

(b) Find the joint pdf of $V$ and $W$. Are they independent?

Hint: You can solve part (b) either directly by finding the joint cdf or by expressing the joint pdf in terms of $f_{U,V}(u, v)$ and using the result of part (a).

4.18. Maximal correlation. Let $(X, Y) \sim F_{X,Y}(x, y)$ be a pair of random variables. 

(a) Show that

$$F_{X,Y}(x, y) \leq \min\{F_X(x), F_Y(y)\}.$$ 

Now let $F(x)$ and $G(y)$ be continuous and invertible cdfs and let $X \sim F(x)$.

(b) Find the cdf of

$$Y = G^{-1}(F(X)).$$

(c) Show that

$$F_{X,Y}(x, y) = \min\{F(x), G(y)\}.$$