Summary comparison of Random Variables and Vectors

<table>
<thead>
<tr>
<th>Random Variables</th>
<th>Random Vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(\omega) = x, Y(\omega) = y$</td>
<td>$X_i(\omega) = x_i, i = 1, \ldots, d$</td>
</tr>
</tbody>
</table>

**Definitions**

- **Joint**: product of marginals
- **Independence**: $f_{XY}(x, y) = f_X(x)f_Y(y)$
- **Mutual independence**: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

**Moments**

- **1st-order moments**
  - $E[X], E[Y]$ scalar
  - $E[X^2], Var[X]$ scalar
  - $E[XY], Cov(X,Y)$ scalar

- **2nd-order moments**
  - $E[XX^t]$ d × d matrix
  - $C_{XY} = E[(X - m_X)(Y - m_Y)^t]$ d × k matrix

**Characteristic Function**

- $\Psi_X(w) = E_X[e^{jw^T X}]$ characteristic function

**Orthogonal**

- $R_{XY} = 0$ across vectors
- $R_X = \Lambda$ diagonal (within a vector)

**Uncorrelated**

- $C_{XY} = 0$ across vectors
- $C_X = \Lambda$ diagonal (within a vector)
Characteristic Functions for Random Vectors

\[ \Psi_X(w) = E[e^{jw^T X}] = E[e^{j(w_1X_1 + w_2X_2 + \cdots + w_dX_d)}] = \int \int \cdots \int e^{j(w_1x_1 + w_2x_2 + \cdots + w_dx_d)} f_X(x_1, x_2, \ldots, x_d) dx_1 dx_2 \cdots dx_d \]

Multi-dimensional Fourier Transform!

Special cases:

Mutually independent vector elements:

\[ \Psi_{\mathbf{X}}(w) = E_{\mathbf{X}}[e^{jw_1X_1}e^{jw_2X_2}\cdots e^{jw_dX_d}] = d^{\prod_i=1} \Psi_{X_i}(w_i) \]

Multivariate Gaussian: \( \mathbf{X} \sim N(\mathbf{m}_X, \Sigma_X) \) \( \Psi_X(w) = \exp(jw^T \mathbf{m}_X - \frac{1}{2}w^T \Sigma w) \)

CFs are useful for finding marginals:

- Given \( \Psi_X(w) \)
- Set \( w_i = 0 \) for all \( i \) not wanted in the marginal
- Take the inverse transform

Example: \( \mathbf{X} \) is 5-dimensional. Find the marginal for the subvector with elements 1, 2 and 5.

\[ f_{X_1X_2X_5}(x_1, x_2, x_5) = FT^{-1} \left\{ \Psi_X(w) \bigg| w_3 = 0, w_4 = 0 \right\} = E[e^{j(w_1X_1 + w_2X_2 + w_5X_5)}] \]

Show: Marginals of a joint Gaussian are Gaussian.

For \( d = 5 \)

\[ \Psi_{X}(\omega) = \exp \left[ j \sum_{i=1}^{5} \omega_i m_i - \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{5} \omega_i \Sigma_{ij} \omega_j \right] \]

\[ \left. \frac{\partial^2 \Psi_{X}(\omega)}{\partial \omega_i \partial \omega_j} \right|_{\omega_i = 0, \omega_j = 0} = \exp \left[ j \sum_{i=1}^{5} \omega_i m_i - \frac{1}{2} \sum_{i=2}^{5} \sum_{j=2}^{5} \omega_i \Sigma_{ij} \omega_j \right] \]

still has the form of a Gaussian CF

so it is a Gaussian

for any combination of \( \omega_i = 0 \)

Example: \( \mathbf{X} \sim N \left( \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \right) \)

\( \rho(X_1, X_2) \sim N \left( 0, \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \right) \) uncorrelated
Conditional distributions:
Random vectors have many different marginals, so also many different conditionals. Consider a 4-dimensional vector \( \mathbf{X} \). Some possibilities:

\[
p_{X_2|X_1,X_3,X_4}(x_2|x_1, x_3, x_4) = \frac{p_{\mathbf{X}}(x_1, x_2, x_3, x_4)}{p_{X_1,X_3,X_4}(x_1, x_3, x_4)} \quad \quad \quad \quad \quad \quad \quad \quad p_{X_1,X_2|X_3,X_4}(x_1, x_2| x_3, x_4) = \frac{p_{\mathbf{X}}(x_1, x_2, x_3, x_4)}{p_{X_2,X_3}(x_2, x_3)}
\]

The Chain Rule:

\[
p_{\mathbf{X}}(x_1, x_2, x_3, x_4) = p_{X_4|X_1,X_2,X_3}(x_4|x_1, x_2, x_3)p_{X_1,X_2,X_3}(x_1, x_2, x_3)
= p_{X_4|X_1,X_2,X_3}(x_4|x_1, x_2, x_3)p_{X_3|X_1,X_2}(x_3|x_1, x_2)p_{X_1,X_2}(x_1, x_2)
= p_{X_4|X_1,X_2}(x_4|x_1, x_2, x_3)p_{X_3|X_1,X_2}(x_3|x_1, x_2)p_{X_2|X_1}(x_2|x_1)p_{X_1}(x_1)
\]

In general:

\[
p_{\mathbf{X}}(x_1, x_2, \ldots, x_d) = p(x_1) \prod_{i=2}^{d} p(x_i|x_1, \ldots, x_{i-1})
\]

Two types of independence:

- \( X_1 \) is \textit{independent of} \( X_2 \) and \( X_3 \) \( \Rightarrow \)
  \[p(x_1, x_2, x_3) = p(x_1)p(x_2, x_3) \quad \text{and} \quad p(x_1|x_2, x_3) = p(x_1) \quad \text{and} \quad p(x_1|x_3) = p(x_1)\]

- \( X_1 \) is \textit{conditionally independent} of \( X_3 \) given \( X_2 \) \( \Rightarrow \)
  \[p(x_1, x_3|x_2) = p(x_1|x_2)p(x_3|x_2) \quad \text{and} \quad p(x_1|x_2, x_3) = p(x_1|x_2)\]
  but NOT \( p(x_1|x_3) = p(x_1) \)

Consider \( X_i \) is conditionally independent of \( X_j \) for \( j < i - 1 \) given \( X_{i-1} \)

\[
p_{\mathbf{X}}(x_1, x_2, \ldots, x_d) = p(x_1) \prod_{i=2}^{d} p(x_i|x_1, \ldots, x_{i-1}) = p(x_1) \prod_{i=2}^{d} p(x_i|x_{i-1})
\]
Joint Gaussians have Gaussian conditional distributions

Let $X$ and $Y$ be jointly Gaussian vectors with

$$
\begin{bmatrix}
X \\
Y
\end{bmatrix} \sim N\left(\begin{bmatrix}
m_X \\
m_Y
\end{bmatrix}, \begin{bmatrix}
\Sigma_X & \Sigma_{XY} \\
\Sigma_{YX} & \Sigma_Y
\end{bmatrix}\right)
$$

Then $X|Y = y \sim N(m, \Sigma)$ where

$$
m = m_X + \Sigma_{XY} \Sigma_Y^{-1} (y - m_Y) \quad \text{and} \quad \Sigma = \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX}
$$

When $X$ and $Y$ are scalar, this simplifies to

$$
m = m_X + C_{XY} \frac{\sigma_Y}{\sigma_Y} (y - m_Y) = m_X + \rho \frac{\sigma_X}{\sigma_Y} (y - m_Y)
$$

$$
\sigma^2 = \sigma_X^2 - C_{XY} C_{YX} \frac{\sigma_Y^2}{\sigma_Y^2} = \sigma_X^2 (1 - \rho^2)
$$

Class Assignment

$$
X \sim N\left(\begin{bmatrix}
2 \\
-3
\end{bmatrix}, \begin{bmatrix}
2 & \frac{1}{2} \\
\frac{1}{2} & 1 \\
\frac{1}{2} & 1
\end{bmatrix}\right)
$$

Find $p(x_1, x_2 | x_3)$
Finding moments with the characteristic function

Consider a 3-dimensional vector for simplifying notation.

\[ \Psi_{X_1X_2X_3}(w) = E[e^{j(w_1X_1 + w_2X_2 + w_3X_3)}] = E[e^{jw_1X_1}e^{jw_2X_2}e^{jw_3X_3}] \]

\[ = E \left( \sum_{l=0}^{\infty} \frac{(jw_1X_1)^l}{l!} \right) \left( \sum_{m=0}^{\infty} \frac{(jw_2X_2)^m}{m!} \right) \left( \sum_{n=0}^{\infty} \frac{(jw_3X_3)^n}{n!} \right) \]

\[ = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} E[X_1^lX_2^mX_3^n] \frac{(jw_1)^l}{l!} \frac{(jw_2)^m}{m!} \frac{(jw_3)^n}{n!} \]

\[ E[X_1^lX_2^mX_3^n] = \frac{1}{j^{l+m+n}} \frac{\partial^l \partial^m \partial^n}{\partial w_1^l \partial w_2^m \partial w_3^n} \Psi_{X_1X_2X_3}(w) \]

\[ \bigg|_{w_1 = 0} \bigg|_{w_2 = 0} \bigg|_{w_3 = 0} \]

**Ex:** \( X \sim \exp(\alpha) \) Find \( E[X^k] \).

\[ \Psi_X(w) = \frac{\alpha}{\alpha - j\omega} \]

\[ E[X^k] = \frac{1}{j^k} \frac{d^k}{dw^k} \Psi_X(w) \bigg|_{w=0} = \frac{1}{j^k} \frac{(-1)^k(-j)^k\alpha^k}{(\alpha - jw)^{k+1}} \bigg|_{w=0} = k! \frac{\alpha^k}{\alpha^k} \]

Note that this result is consistent with the first and second order moments that we already know: \( E[X] = 1/\alpha \) and \( E[X^2] = 2/\alpha^2 \).
Ex: $X \sim N(m, \Sigma)$ where $X = [X_1 \ X_2]^t$. Find $E[X_1^2X_2]$.

\[
\begin{align*}
\Psi_X(w) &= \exp(jw^t m - 0.5w^t \Sigma w) \\
&= \exp(j(w_1m_1 + w_2m_2) - 0.5(w_1^2\Sigma_{11} + 2w_1w_2\Sigma_{12} + w_2^2\Sigma_{22}))
\end{align*}
\]

\[
E[X_1^2X_2] = \frac{1}{j^3} \frac{\partial^3}{\partial w_1^2 \partial w_2} \Psi_X(w) \bigg|_{w_1 = 0 \atop w_2 = 0}
\]

\[
= j \frac{\partial^2}{\partial w_1^2} (jm_2 - w_1\Sigma_{12} - w_2\Sigma_{22}) \Psi_X(w) \bigg|_{w_1 = 0 \atop w_2 = 0}
\]

\[
= j \frac{\partial}{\partial w_1} [-\Sigma_{12} + (jm_1 - w_1\Sigma_{11} - w_2\Sigma_{12})(jm_2 - w_1\Sigma_{12} - w_2\Sigma_{22})] \Psi_X(w) \bigg|_{w_1 = 0 \atop w_2 = 0}
\]

\[
= j [-jm_1\Sigma_{12} - jm_2\Sigma_{11} + 2w_1\Sigma_{11}\Sigma_{12} + (jm_1 - w_1\Sigma_{11} - w_2\Sigma_{12})(jm_2 - w_1\Sigma_{12} - w_2\Sigma_{22})] \Psi_X(w) \bigg|_{w_1 = 0 \atop w_2 = 0}
\]

\[
= m_1\Sigma_{12} + m_2\Sigma_{11} - m_1(-\Sigma_{12} - m_1m_2) = 2m_1\Sigma_{12} + m_2\Sigma_{11} + m_1^2m_2
\]

\[
\Psi_X(0) =
\]

Notice that even though you only need 1st and 2nd order moments to characterize a Gaussian, you can still have non-zero higher order moments, but they are described in terms of $m$ and $\Sigma$.

Answer check: real RV's have real moments, so $j$ terms must cancel.
Multinomial Distribution

Reminder of Binomial: $n$ trials of binary outcomes, where probability of 1 is $q$ and 0 is $1 - q$

$$P(X = k) = \frac{n!}{k!(n-k)!} q^k (1 - q)^{n-k}$$

Multinomial: $n$ trials of $m$-ary outcomes, where probability of outcome $i$ is $q_i$ for $i = 1, \ldots, m$ and $\sum_i q_i = 1$

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) = \frac{n!}{x_1!x_2! \cdots x_m!} q_1^{x_1} q_2^{x_2} \cdots q_m^{x_m} \sum_{i=1}^{m} x_i = n$$

Examples:
- $n$ rolls of a die, $m = 6$
- counts of word types in a document of length $n$, $m$ is the number of words in the vocabulary
- $n$ packets arriving in $m$ different ports

Properties of multinomials:
1st-order marginals are binomial with parameters $n$ and $q_i$

$$E[X_i] = nq_i \quad Var[X_i] = nq_i(1 - q_i)$$

Characteristic function

$$\Psi_X(w) = (\sum_{i=1}^{m} q_i e^{iw_i})^n$$

You can use the CF to show that

$$R_{lk} = E[X_l X_k] = n(n - 1)q_k q_l \quad C_{lk} = Cov(X_l, X_k) = -nq_k q_l$$

$$E[X_l X_k] = \frac{1}{j^2} \frac{d}{d\omega_l} \frac{d}{d\omega_k} \psi_X(w) \bigg|_{\bar{w} = \bar{0}}$$

$$= \frac{1}{j^2} \frac{d}{d\omega_l} \left( n^j q_k e^{j\omega_k} \left( \sum_{i=1}^{m} q_i e^{j\omega_i} \right)^{n-1} \right) \bigg|_{\bar{w} = \bar{0}}$$

$$= \frac{1}{j^2} \left( n(n-1) \right)^j q_k q_l e^{j\omega_k} e^{j\omega_l} \left( \sum_{i=1}^{m} q_i e^{j\omega_i} \right)^{n-2} \bigg|_{\bar{w} = \bar{0}}$$

$$= \frac{1}{j^2} \left( n(n-1) \right)^j q_k q_l \left( \sum_{i=1}^{m} q_i e^{j\omega_i} \right)^{n-2} \bigg|_{\bar{w} = \bar{0}}$$

$$= n(n-1) q_k q_l$$