Lecture 6 - Random Vectors

A. Vectors and Tensors

A vector is defined as components x_i with unit directions $\underline{e}_i$,

$$\underline{x} = \sum_i x_i \underline{e}_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1, x_2, \dots, x_n]^T$$

← transpose

← "column vector"

A tensor is defined similarly, except its components are indexed in two dimensions A_{ij} and the directions are a unit dyad $\underline{e}_i \underline{e}_j$

$$\underline{\underline{A}} = \sum_i \sum_j A_{ij} \underline{e}_i \underline{e}_j = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \\ A_{n1} & & & A_{nn} \end{bmatrix}$$

A (2nd rank) tensor is a matrix with directions.

We will need some basic operations for these objects including sums, transpose, inner/dot product, trace, and norm. These are reviewed in the appendix.

B. Distributions

Just as we saw with the case with two variables, we can define multivariate joint cdfs, pdfs, and pmfs.

Joint cdf:

$$F(x_1, x_2, \dots, x_n) = F(\underline{x}) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

Joint pdf:

$$P(\underline{x} \in V) = \underset{V}{\iint \cdots \int} f(x_1, x_2, \dots x_n)\, dx_1\, dx_2 \dots dx_n = \int_V f(\underline{x})\, d\underline{x}$$

← some "hypervolume" V

note that as before:

$$f(\underline{x}) = \frac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n}$$

Relationship between CDF and PDF

Joint pmf:

$$P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = p(x_1, x_2, \dots x_n) = p(\underline{x})$$

If we want to eliminate one (or more) variables from the distribution, the process is called marginalization.

cdf marginalization:

$$F(x_1, x_2, \dots, x_{n-1}) = F(x_1, x_2, \dots, x_{n-1}, \infty)$$

pdf marginalization:

$$f(x_1, x_2, \dots, x_{n-1}) = \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_{n-1}, x_n)\, dx_n$$

pmf marginalization:

$$p(x_1, x_2, \dots, x_{n-1}) = \sum_{i=1}^{\infty} p(x_1, x_2, \dots, x_{n-1}, x_n^i)$$

Notes:

- We can marginalize ("integrate out") any variable, not just the last.
- If we marginalize everything, we will get one (by axiom).

We can also have multivariate conditional distributions

$$f(x_1, x_2, \dots, x_n) = \underbrace{f(x_n \mid x_1, x_2, \dots, x_{n-1})}_{\text{conditional pdf}}\, f(x_1, x_2, \dots, x_{n-1})$$

Note: I'm omitting the discrete case. It is nearly identical to the continuous one.

we can decompose a pdf into successive conditional probabilities. This is called the chain rule. (no relation to calculus rule)

$$f(x_1, x_2, \dots, x_n) = f(x_1) \prod_{i=2}^{n} f(x_i \mid x_1, x_2, \dots, x_{i-1})$$

Example: $n=3$

$$f(x_1, x_2, x_3) = f(x_3 \mid x_1, x_2)\, f(x_1, x_2)$$ definition of conditional

$$= f(x_3 \mid x_1, x_2)\, \underbrace{f(x_2 \mid x_1)\, f(x_1)}$$ now do these two

$$= f(x_1)\, f(x_2 \mid x_1)\, f(x_3 \mid x_1, x_2)$$ re-arrange

Independence means:

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

This is a pretty big assumption in multivariate cases. Sometimes less drastic assumptions such as pairwise independence are used.

Example: Full vs. Pairwise Independence for $n=3$

Full: $f(x_1, x_2, x_3) = f(x_1) f(x_2) f(x_3)$

Pairwise: $f(x_1, x_2, x_3) = f(x_3 \mid x_1, x_2)\, f(x_1, x_2)$

$$= f(x_3 \mid x_1, x_2)\, \underbrace{f(x_2 \mid x_1)}_{f(x_2 \mid x_1) = f(x_2)}\, f(x_1)$$

$$= f(x_1) f(x_2)\, \underbrace{f(x_3 \mid x_1, x_2)}_{\text{This one stays}}$$

C. Expectation and moments

The expectation operator is a straightforward generalization.

$$E[g(\underline{x})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\underline{x})\, f(\underline{x})\, d\underline{x}$$ $\leftarrow dx_1\, dx_2 \ldots dx_n$

mean:

$$E[\underline{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \underline{m}$$ $\leftarrow$ mean vector

2^{nd} moments give a correlation matrix.

$$E[\underline{x}\underline{x}] = \begin{bmatrix} E[x_1x_1] & E[x_1x_2] & \cdots & E[x_1x_n] \\ E[x_2x_1] & E[x_2x_2] & \cdots & E[x_2x_n] \\ \vdots & & \ddots & \vdots \\ E[x_nx_1] & E[x_nx_2] & \cdots & E[x_nx_n] \end{bmatrix} = \underline{\underline{R}}$$

Centered second moments give a covariance matrix.

$$E[(\underline{x}-\underline{m})(\underline{x}-\underline{m})] = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_n) \\ cov(x_2,x_1) & cov(x_2,x_2) & \cdots & cov(x_2,x_n) \\ \vdots & & \ddots & \vdots \\ cov(x_n,x_1) & cov(x_n,x_2) & \cdots & cov(x_n,x_n) \end{bmatrix} = \underline{\underline{C}}$$

One can also compute a cross-covariance matrix from two different random vectors

$$C_{xy} = E[(\underline{x}-\underline{m}_x)(\underline{y}-\underline{m}_y)]$$

Notes:

- Both $\underline{\underline{R}}$ and $\underline{\underline{C}}$ are symmetric matrices.
- Diagonal elements of $\underline{\underline{C}}$ are the variance
- If $C_{ij}=0$ for $i \neq j$ (diagonal), then x_i and x_j are uncorrelated.

Finally, one can calculate the characteristic function using an n-dimensional Fourier transform.

$$\phi(\underline{\omega}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\underline{\omega}\cdot\underline{x}} f(\underline{x})\, d\underline{x} \qquad \underline{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}$$

$$f(\underline{x}) = \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-i\underline{\omega}\cdot\underline{x}} \phi(\underline{\omega})\, d\underline{\omega}$$

The mean and components of the correlation matrix can be found using derivatives of $\phi(\underline{\omega})$.

$$\frac{\partial}{\partial \omega_i}\phi(\underline{\omega}=0) = i\, E[\omega_i]$$

$$\frac{\partial^2}{\partial \omega_i \partial \omega_j}\phi(\underline{\omega}=0) = -E[\omega_i \omega_j]$$

D. n-Component Normal Distribution

The multivariate normal pdf is given by

$$f(\underline{x}) = \frac{1}{[(2\pi)^n \det C]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{m})\cdot \underline{\underline{C}}^{-1}\cdot(\underline{x}-\underline{m})\right]$$

determinant of covariance matrix

inverse of covariance matrix

The characteristic function is given by:

$$\phi(\underline{\omega}) = \exp\left(i\underline{x}\cdot\underline{m} - \frac{1}{2}\underline{x}\cdot\underline{\underline{C}}\cdot\underline{x}\right)$$

- For a Gaussian random vector, uncorrelated = independent
- Level sets of the density are ellipsoids centered at m.

Appendix: Vector and Matrix Operations

There are numerous algebraic and calculus operations performed on vectors and tensors. We only need to review a few.

For these examples, lets consider

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \underline{\underline{A}} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \underline{\underline{B}} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Transpose:

$$(A_{ij})^T = A_{ji} \qquad \underline{\underline{A}}^T = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}$$

← swap rows and columns

sums:

$$\underline{x} + \underline{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

element wise addition

same for matrix / tensor

inner prodcut:

$$\underline{\underline{A}} \cdot \underline{\underline{B}} = \sum_{k=1}^{n} A_{ik} B_{kj} = \begin{bmatrix} A_{11} B_{11} + A_{12} B_{21} & A_{11} B_{12} + A_{12} B_{22} \\ A_{21} B_{11} + A_{22} B_{21} & A_{21} B_{12} + A_{22} B_{22} \end{bmatrix}$$

"matrix multiplication"

$$\underline{x} \cdot \underline{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2$$

"dot product"

Trace:

$$Tr(A) = \sum_{i=1}^{n} A_{ii} = A_{11} + A_{22} \qquad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

sum of the diagonal elements

Norm:

$$\| \underline{x} \| = (\underline{x} \cdot \underline{x})^{1/2} = (x_1^2 + x_2^2)^{1/2}$$

"distance formula"