Outer Product

The outer product of two vectors x and y is xy^T:

$$xy^T = \begin{pmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_m \\ x_2y_1 & x_2y_2 & \dots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \dots & x_ny_m \end{pmatrix}$$

The outer product of a vector with itself xx^T is the multi-dimensional generalization of the square of a number

Random Vectors

Covariance Matrix

The covariance Σ of a random vector X is defined as:

$$\Sigma = \text{Var}\left[X\right] = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right] = \mathbb{E}\left[XX^T\right] - \mu\mu^T$$

The ij th entry of Σ is $\text{Cov}\left(X_i, X_j\right)$

- The inverse of the covariance matrix is sometimes called the **precision** matrix

Linear Transformations

Let $a \in \mathbb{R}^k$ be a deterministic vector. Then $a^TX \in \mathbb{R}$ with:
- $\mathbb{E}\left[a^TX\right] = a^T\mu$
- $\text{Var}\left[a^TX\right] = a^T\Sigma a$

Let $A \in \mathbb{R}^{k\times l}$ be a deterministic matrix and $b \in \mathbb{R}^\ell$ be a deterministic vector. Then $A^TX + b \in \mathbb{R}^\ell$ with:
- $\mathbb{E}\left[A^TX+b\right] = A^T\mu + b$
- $\text{Var}\left[A^TX+b\right] = A^T\Sigma A$

Proof of First Statement:
- Expectations is relatively straightforward via linearity
- For computing the variance, we make use of the fact that for some real number x, $x = x^T$
- We first compute the expectation $\mathbb{E}\left[(a^TX)^2\right]$:
  - $\mathbb{E}\left[(a^TX)^2\right] = \mathbb{E}\left[a^TXX^Ta\right] = a^T\mathbb{E}\left[XX^T\right]a$
- Then we have:
  - $\left(\mathbb{E}\left[a^TX\right]\right)^2 = a^T\mu\mu^Ta$
- Subtracting these and factoring gives us $a^T\Sigma a$

Multivariate Distributions

Probability Density Function

The PDF of a random vector X is a function $f: \mathbb{R}^k \to \mathbb{R}$ such that $f(x) \geq 0$ and $\int_{\mathbb{R}^k} f(x)dx = 1$

We also call this the **join** density of each of the individual variables in X (i.e. f is the joint density of $X_1, \dots, X_k$)

Marginal Density

The marginal density of X_j is obtained by integrating with respect to all variables except the jth one:

$$f_j(x_j) = \iint\cdots\int f(x_1,\dots,x_k)dx_1\dots dx_{j-1}dx_{j+1}\dots dx_k$$

Specifying joint PDFs can be complicated since it has to account for all possible dependencies between variables. Covariances capture pairwise, but there could be triples / quadruples.

Independent Random Variables

If $X_1, \ldots, X_k$ are independent, then the joint density takes the following product form:

$$f(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_i(x_i)$$

Conditional Distribution

Conditional PDF

The conditional density of X_k given the values for all other indices is:

$$f(x_k | X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = \frac{f(x_1, \ldots, x_k)}{\int_{\mathbb{R}^k} f(x_1, \ldots, x_k) dx_k}$$