# Normal distributions on vector spaces

Kristian Wichmann

November 27, 2016

## 1 Affine transformations of euclidean spaces

Let $s : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. This means that the is an $m \times n$ matrix $A$ so $s(x) = Ax$.

An *affine* transformation $t$ is formed by following this linear map by a translation:

$$t : \mathbb{R}^n \to \mathbb{R}^m, t(x) = Ax + v \tag{1.1}$$

Here, $v \in \mathbb{R}^m$. Since translations are always bijective, we note that $t$ is bijective iff $A$ is invertible.

Each component of an affine transformation is composed from measurable function - is is understood that we mean with respect to the Borel algebras of each space) - so the affine transformation itself is measurable as well.

### 1.1 Transformation properties of the Lebesgue measure

Recall that the Lebesgue measure in $n$ dimensions $m_n$ is invariant under translation: If $t$ is a translation $t : \mathbb{R}^n \to \mathbb{R}^n, t(x) = x + x_0$, where $x_0 \in \mathbb{R}^n$ then:

$$t(m_n) = m_n \tag{1.2}$$

Also, if $s : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto Ax$ is an isomorphism, then:

$$s(m_n) = m_n |\det A^{-1}| \tag{1.3}$$

Combining the two, the formula for affine transformation is the same as for linear ones.

## 2 Orthogonal complement

Let $V$ be a finite-dimensional vector space with an inner product $\langle \cdot, \cdot \rangle$. Let $U$ be a subspace of $V$. Then we define the *orthogonal complement* of $U$ as:

$$U^\perp = \{v \in V | \forall u \in U : \langle u, v \rangle = 0\} \tag{2.1}$$

**Theorem 2.1.** *$U^\perp$ is a subspace of $V$.*

*Proof.* According to the subspace theorem, we need to show three things:

- $U^\perp$ is not empty: Clearly $0 \in U^\perp$.
- Closed under addition: If $v_1, v_2 \in U^\perp$, then for all $u \in U^\perp$:
$$\langle v_1 + v_2, u \rangle = \langle v_1, u \rangle + \langle v_2, u \rangle = 0 \tag{2.2}$$
- Closed under scalar multiplication: If $v \in U^\perp$ and $c \in \mathbb{R}$ then for all $u \in U^\perp$:
$$\langle cv, u \rangle = c\langle v, u \rangle = 0 \tag{2.3}$$

□

Since the only vector perpendicular to itself is 0, we further conclude that $U \cap U^\perp = \{0\}$.

**Theorem 2.2.** *If $e_1, e_2, \ldots, e_m$ is an orthonormal basis for $U$, then for any $v \in V$:*

$$v - \sum_{i=1}^{m} \langle v, e_i \rangle e_i \in U^\perp \tag{2.4}$$

*Proof.* Let $u \in U$. Then we can write $u = \sum_{j=1}^m \lambda_j e_j$ for some coefficients $\lambda_j$. Now calculate the inner product with the vector above:

$$\langle v - \sum_{i=1}^{m} \langle v, e_i \rangle e_i \ , \ \sum_{j=1}^{m} \lambda_j e_j \rangle = \sum_{i=j}^{m} \lambda_j \langle v, e_j \rangle - \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_j \langle v, e_i \rangle \langle e_i, e_j \rangle \tag{2.5}$$

Since $\langle e_i, e_j \rangle = \delta_{ij}$ this vanishes. □

This means that we may write any $v \in V$ as a sum of vectors from $U$ and $U^\perp$ respectively:

$$v = \underbrace{\sum_{i=1}^{m} \langle v, e_i \rangle e_i}_{\in U} \ + \ \underbrace{v - \sum_{i=1}^{m} \langle v, e_i \rangle e_i}_{\in U^\perp} \tag{2.6}$$

**Theorem 2.3.** *The decomposition into elements from $U$ and $U^{\perp}$ from equation 2.6 is unique.*

*Proof.* Let $v = u_1 + u_1^{\perp}$ and $v = u_2 + u_2^{\perp}$ be two such decompositions. Then $u_1 + u_1^{\perp} = u_2 + u_2^{\perp}$ and hence $u_1 - u_2 = u_2^{\perp} - u_1^{\perp}$. But this means that this vector is a member of both $U$ and $U^{\perp}$, and hence it must be 0. This means $u_1 = u_2$ and $u_1^{\perp} = u_2^{\perp}$. □

## 2.1 The orthogonal projection

The previous section motivates the following:

**Definition 2.1.** *Let $V$ be a finite-dimensional inner product vector space and $U$ a subspace of $V$. The orthogonal projection from $V$ onto $U$ is the map $p : V \to V$ which satisfies:*

$$\forall v \in V : \quad p(v) \in U, \quad v - p(v) \in U^{\perp} \tag{2.7}$$

As we see, one could also define the co-domain of $p$ to be $U$. Usually, the distinction will not matter much.

**Theorem 2.4.** *The orthogonal projection operator is linear.*

*Proof.* We need to show additivity and homogeneity:

- Additivity: Let $v_1, v_2 \in V$. Then $p(v_1) + p(v_2) \in U$ and:

  $$v_1 - p(v_1) + v_2 - p(v_2) = v_1 + v_2 - (p(v_1) + p(v_2)) \in U^{\perp} \tag{2.8}$$

  Adding the two we get $v_1 + v_2$. So $p(v_1 + v_2) = p(v_1) + p(v_2)$.

- Homogeneity. Let $v \in V$ and $c \in \mathbb{R}$. Then $cp(v) \in U$ and $c(v - p(v)) = cv - cp(v) \in U^{\perp}$. Adding the two we get $cv$, so $p(cv) = cp(v)$.

□

**Theorem 2.5.** *The orthogonal projection operator $p : V \to V$ is idempotent. I.e. $p \circ p = p$.*

*Proof.* Let $v \in V$. Then $p(v) \in U$. But this means that the decomposition of $p(v)$ is $p(v) + 0$. So $p \circ p(v) = p(v)$. □

# 3 Lebesgue measures on vector spaces

## 3.1 Coordinate maps

Let $V$ be a finite-dimensional vector space of dimension $n$. Our question is, if we can turn $V$ into a measure space in a natural way. Since we know that $V$ is isomorphic to $\mathbb{R}^n$, it makes sense to tweak the usual Lebesgue measure in $N$ dimensions:

Let $e_1, e_2, \dots, e_n$ be a basis for $V$. Then we can define the *coordinate map* as follows:

$$\phi : \mathbb{R}^n \to V, \qquad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \sum_{i=1}^{n} x_i e_i \tag{3.1}$$

This is obviously an isomorphism. Specifically, it is invertible with inverse $\phi^{-1} : V \to \mathbb{R}^n$.

The coordinate map depends on the chosen basis. If we had chosen another basis $e_1^*, e_2^*, \dots, e_n^*$ we would get another isomorphism $\phi^*$.

## 3.2 Borel algebra on $V$

We can now use $\phi^{-1}$ to induce a $\sigma$-algebra on $V$. Set $\mathbb{B}_V$ to the smallest $\sigma$-algebra that makes $\phi^{-1}$ measurable when $\mathbb{R}^n$ is equipped with the Borel algebra $\mathbb{B}_n$. We call $\mathbb{B}_V$ the *Borel algebra on $V$*.

At first this object seems to depend of the choice of basis for $V$. But it turns out that the use of definite article in the definition is justified:

**Theorem 3.1.** *If $e_1, e_2, \dots, e_n$ and $e_1^*, e_2^*, \dots, e_n^*$ are bases for $V$, then the induced $\sigma$-algebra $\mathbb{B}_V$ and $\mathbb{B}_V^*$ is the same thing.*

*Proof.* We know that $\phi^{-1}$ is $\mathbb{B}_V - \mathbb{B}_n$ measurable by definition. We have:

$$(\phi^*)^{-1} = (\phi^*)^{-1} \circ \mathrm{id}_V = (\phi^*)^{-1} \circ (\phi \circ \phi^{-1}) = \big((\phi^*)^{-1} \circ \phi\big) \circ \phi^{-1} \tag{3.2}$$

$((\phi^*)^{-1} \circ \phi)$ is a linear operator on $\mathbb{R}^n$ and so according to section 1.1 is measurable. So $(\phi^*)^{-1}$ must be $\mathbb{B}_V - \mathbb{B}_n$-measurable. Since $\mathbb{B}_V^*$ is the smallest $\sigma$-algebra to make $(\phi^*)^{-1}$ $\mathbb{B}_V - \mathbb{B}_n$-measurable, we must have $\mathbb{B}_V^* \subseteq \mathbb{B}_V$.

But by a totally symmetric argument, we must also have $\mathbb{B}_V \subseteq \mathbb{B}_V^*$. Hence $\mathbb{B}_V = \mathbb{B}_V^*$. $\square$

It turns out, that $\phi$ must be measurable too. This is a direct consequence of the pipeline lemma.

**Theorem 3.2.** *Given two finite-dimensional vector spaces $V$ and $W$, then:*

$$\mathbb{B}_{V\times W} = \mathbb{B}_V \otimes \mathbb{B}_W \tag{3.3}$$

*Proof.* Let $e_1, e_2, \ldots, e_n$ be a basis for $V$ with corresponding coordinate map $\phi$. And $f_1, f_2, \ldots, f_m$ a basis for $W$ with corresponding coordinate map $\psi$. Then $(e_1, 0), (e_2, 0), \ldots, (e_n, 0), (0, f_1), (0, f_2), \ldots, (0, f_m)$ is a basis for $V \times W$. The corresponding coordinate map is:

$$\phi \times \psi : (x_1, x_2, \ldots, x_{n+m}) \mapsto \left( \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{m} x_{n+j} f_j \right) \tag{3.4}$$

The inverse is $(\phi \times \psi)^{-1} = \phi^{-1} \times \psi^{-1}$. Since $\phi^{-1}$ is $\mathbb{B}_V - \mathbb{B}_n$-measurable and $\psi^{-1}$ is $\mathbb{B}_W - \mathbb{B}_m$-measurable, $\phi^{-1} \times \psi^{-1}$ must be $\mathbb{B}_V \otimes \mathbb{B}_W - \mathbb{B}_n \otimes \mathbb{B}_n$-measurable. But $\mathbb{B}_n \otimes \mathbb{B}_m = \mathbb{B}_{n+m}$. Since $\mathbb{B}_{V\times W}$ is the smallest $\sigma$-algebra to make $\phi^{-1} \times \psi^{-1}$ measurable, we must have $\mathbb{B}_{V\times W} \subseteq \mathbb{B}_V \otimes \mathbb{B}_W$.

On the other hand, consider the projection operators:

$$\pi_V : V \times W \to V, (v, w) \mapsto v \tag{3.5}$$

$$\pi_n : \mathbb{R}^{n+m} \to \mathbb{R}^n, (x_1, \ldots, x_{n+m}) \mapsto (x_1, \ldots, x_n) \tag{3.6}$$

Now consider $\pi_V \circ (\phi \times \psi)$. Applied to an $x \in \mathbb{R}^{n+m}$ we have:

$$\pi_V \circ (\phi \times \psi)(x) = \pi_v \left( \left( \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{m} x_{n+j} f_j \right) \right) = \sum_{i=1}^{n} x_i e_i \tag{3.7}$$

But this is the same as:

$$\phi \circ \pi_1(x) = \phi\left((x_1, \ldots x_n)\right) = \sum_{i=1}^{n} x_i e_i \tag{3.8}$$

So $\pi_V \circ (\phi \times \psi) = \phi \circ \pi_1$. Now apply $\phi^{-1} \times \psi^{-1}$ from the right to get:

$$\pi_V = \phi \circ \pi_1 \circ (\phi^{-1} \times \psi^{-1}) \tag{3.9}$$

Since all the three functions on the right side are measurable, $\pi_V$ must be $\mathbb{B}_{V\times W} - \mathbb{B}_V$-measurable. By a similar argument the corresponding projection operator $\pi_W : V \times W \to W$ is $\mathbb{B}_{V\times W} - \mathbb{B}_W$-measurable. Since $\mathbb{B}_V \otimes \mathbb{B}_W$ is the smallest $\sigma$-algebra to make both $\pi_V$ and $\pi_W$ measurable, we must have: $\mathbb{B}_V \otimes \mathbb{B}_W \subseteq \mathbb{B}_{V\times W}$. □

### 3.3 Lebesgue measures on $V$

We now want to define a measure on the measurable space $(V, \mathbb{B}_V)$. If $e_1, e_2, \ldots, e_n$ is a basis for $V$, we will use the associated coordinate map $\phi$ to define a measure:

$$\lambda_V = \phi(m_n) \tag{3.10}$$

Here, $m_n$ is the usual Lebesgue measure in $n$ dimensions. The problem is, that this measure depends on the chosen basis! Consider another basis $e_1^*, e_2^*, \ldots, e_n^*$ and associated coordinate map $\phi^*$. Then the measure is:

$$\lambda_V^* = \phi^*(m_n) = (\phi \circ \phi^{-1}) \circ \phi^*(m_n) = \phi \circ (\phi^{-1} \circ \phi^*(m_n)) \tag{3.11}$$

Now $\phi^{-1} \circ \phi^*$ is an isomorphism $\mathbb{R}^n \to \mathbb{R}^n$, so according to section 1.1, there is a constant $c$ such that $(\phi^{-1} \circ \phi^*(m_n)) = cm_n$. So:

$$\lambda_V^* = c\phi(m_n) = c\lambda_V \tag{3.12}$$

So while there are many Lebesgue measures on $V$ they only differ from each other by a constant factor. This means that they all agree on what constitutes a null set, and on which functions are integrable. They disagree on the integral, but agree on whether it is finite or not. The also agree on whether a measure $\mu$ has a density with respect to $\lambda_V$ or not.

# 4 Random vectors

In this section, we consider vectors of random variables. So a random vector of dimension $n$ is:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \tag{4.1}$$

Here, each $X_i$ is a random variable.

### 4.1 Variance

The variance of a $n$-dimensional vector is the $n \times n$ matrix:

$$\mathrm{Var}(X) = E[(X - \mu_X)(X - \mu_X)^t] \tag{4.2}$$

Here $\mu_X = E[X]$, i.e. the vector of expectation values of the $X_i$'s. From the usual definitions of variances and covariances between random variables, we see that the diagonal of $\mathrm{Var}(X)$ contains the variances of each $X_i$, while the off diagonal elements are the covariances between variables:

$$[\mathrm{Var}(X)]_{ij} = \mathrm{Cov}(X_i, X_j) \tag{4.3}$$

Due to the symmetry of covariance, this means that $\mathrm{Var}(X)$ is a symmetric matrix.

### 4.1.1 Variance calculation rules

Similarly to ordinary random variables, we might calculate the variance matrix as follows:

$$\begin{aligned}\mathrm{Var}(X) =& E[(X-\mu_X)(X-\mu_X)^t] = && (4.4)\\ & E(XX^t) - E(X)\mu_X^t - \mu_X E(X)^t + \mu_X\mu_X^t = && (4.5)\\ & E(XX^t) - \mu_X\mu_X^t && (4.6)\end{aligned}$$

Here, we've used the linearity of the expectation value and the definition of $\mu_X$.

Adding a constant vector $b$ does not change the variance, since $E[X+b] = \mu_X + b$:

$$\mathrm{Var}(X+b) = E[(X+b-(\mu_X+b))(X+b-(\mu_X+b))^t] = E[(X-\mu_X)(X-\mu_X)^t] \tag{4.7}$$

This is just the variance of $X$.

If $A$ is a constant $m\times n$ matrix and $X$ is an $n$-dimensional random vector, then:

$$\begin{aligned}\mathrm{Var}(AX) =& E[(AX-A\mu_X)(AX-A\mu_X)^t] = && (4.8)\\ & E[(A(X-\mu_X))(A(X-\mu_X))^t] = && (4.9)\\ & E[A(X-\mu_X)(X-\mu_X)^t A^t] = && (4.10)\\ & A[(X-\mu_X)(X-\mu_X)^t]A^t && (4.11)\end{aligned}$$

So we have $\mathrm{Var}(AX) = A\ \mathrm{Var}(X)A^t$.

## 4.2 Covariance

The covariance matrix between two variable vectors $X$ and $Y$ is:

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)^t] \tag{4.12}$$

If $X$ has dimension $m$ and $Y$ dimension $n$, then $\text{Cov}(X, Y)$ has dimension $m \times n$. Here, the matrix elements reduce to ordinary covariances between $X_i$'s and $Y_j$s:

$$[\text{Cov}(X, Y)]_{ij} = \text{Cov}(X_i, Y_j) \tag{4.13}$$

We note, that the variance could have been defined as a special case of covariance, since $\text{Var}(X) = \text{Cov}(X, X)$.

### 4.2.1 Covariance calculation rules

Similarly to the rule for variances, we have:

$$\text{Cov}(X, Y) = E[XY^t] - \mu_X \mu_Y^t \tag{4.14}$$

The proof is essentially the same.

If $A$ and $B$ are constant matrices of appropriate dimesion, we also have:

$$\text{Cov}(AX, BY) = A\ \text{Cov}(X, Y)B^t \tag{4.15}$$

Again, the proof is entirely analogous to the corresponding variance formula.

The covariance is bilinear:

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z) \tag{4.16}$$

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z) \tag{4.17}$$

This follows from the bilinearity of the ordinary covariance.

### 4.2.2 Addiational variance formulas

Since we noted that $\text{Var}(X) = \text{Cov}(X, X)$, we may use these rules to derive further properties of variances.

For instance, the variance of a sum:

$$\text{Var}(X + Y) = \text{Cov}(X + Y, X + Y) = \tag{4.18}$$

$$\text{Cov}(X, X) + \text{Cov}(X, Y) + \text{Cov}(Y, X) + \text{Cov}(Y, Y) = \tag{4.19}$$

$$\text{Var}(X) + \text{Var}(Y) + \text{Cov}(X, Y) + \text{Cov}(Y, X) \tag{4.20}$$

This mirrors the formula for the covariance of sums ordinary random variables, but is complicated by the fact that the vector covariance is not symmetric.

## 4.3 Quadratic forms

If $X$ is an $n$-dimensional random variable and $A$ an $n \times n$ matrix, then the corresponding quadratic form is $Q = X^t AX$. I.e. a scalar. What is the expectation value of the quadratic form? We can use a trick here. Since $Q$ is a scalar, we can trivially write this as a trace:

$$Q = X^t AX = \mathrm{tr}(X^t AX) = \mathrm{tr}(AXX^t) \tag{4.21}$$

Here, we've used the cyclic property of traces. Now, the expectation value is:

$$E[Q] = E[\mathrm{tr}(AXX^t)] = \mathrm{tr}(E[AXX^t]) = \mathrm{tr}(A\ E[XX^t]) \tag{4.22}$$

But we know, that $\mathrm{Var}(X) = E(XX^t) - \mu_X \mu_X^t$, so $E[XX^t] = \mathrm{Var}(X) + \mu_X \mu_X^t$:

$$E[Q] = \mathrm{tr}(A(\mathrm{Var}(X) + \mu_X \mu_X^t)) = \mathrm{tr}(A\ \mathrm{Var}(X)) + \mathrm{tr}(A\mu_X \mu_X^t) \tag{4.23}$$

The last term may be rewritten:

$$\mathrm{tr}(A\mu_X \mu_X^t) = \mathrm{tr}(\mu_X^t A\mu_X) = \mu_X^t A\mu_X \tag{4.24}$$

In the last step we've used that the contents of the parenthesis is a scalar. So all in all:

$$E[X^t AX] = \mathrm{tr}(A\ \mathrm{Var}(X)) + \mu_X^t A\mu_X \tag{4.25}$$