Appendix C

Improper Normal Priors

Most of the prior densities considered in this book are improper. The fact that they are improper has no negative consequences because our likelihood and therefore our posterior distribution is always proper. What is relevant is the reason they are improper. Roughly speaking, the reason is that we have only partial prior knowledge: that is, we know certain things but are ignorant about others. This implies that our prior density is flat over unbounded subsets of its support, which causes it to fail to be integrable.

C.1 Definitions

In this section, we describe the mathematical tools needed to deal with the kinds of improper prior densities we use. These densities have the following form:

$$\mathcal{P}(x;\theta) \propto \exp\left(-\frac{1}{2}\theta x' D x\right) \ , \quad x \in \mathbb{R}^d, \tag{C.1}$$

where $\theta > 0$ and D is a symmetric, $d \times d$ positive semidefinite matrix with $\text{rank}(D) = r < d$, and $\text{nullity}(D) = n = d - r$. The obvious way to see why this density is improper is to notice that its covariance, which would be calculated by taking the inverse of the matrix D, does not exist because D does not have full rank. We now develop a richer understanding of improperness through a geometric interpretation. The key to understanding and manipulating improper densities of the form C.1 is noting that, because D is symmetric, it can be diagonalized. That is, every such D can be uniquely decomposed in the following (see appendix B.2.2, page 231):

$$D = RLR', \tag{C.2}$$

where R is an orthogonal matrix (so that $R^{-1} = R'$ and $\det(R) = 1$) and L is diagonal. This implies that we can always put ourselves in a reference system where the density in equation C.1 is the product of one dimensional densities. In fact, if we make the change of variables $x = Rz$, the preceding density can be written as a function of z:

$$\mathcal{P}(z;\theta) \propto \exp\left(-\frac{1}{2}\theta z' L z\right) = \prod_{i=1}^{d} \exp\left(-\frac{1}{2}\theta \ell_i z_i^2\right), \tag{C.3}$$

where ℓ_i are the diagonal elements of L. This observation suggests that everything we need to know about improper densities of the type C.1 can be learned analyzing the simpler densities like that in equation C.3. For this reason we next analyze in detail a simple but highly informative example in three dimensions.

C.2 An Intuitive Special Case

Here we take $\theta = 1$ and consider a diagonal matrix D:

$$D \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix is positive semidefinite, between its eigenvalues, which coincide with the elements on the diagonal, are greater than or equal to zero. The rank of this matrix is 2, and its nullity is $1 = 3 - 2$. The range of D is the (x_1, x_2) plane: $\text{range}(D) \equiv \{(x_1, x_2, 0) \in \mathbb{R}^3\}$, while its null space is the x_3-axis: $\mathfrak{N}(D) \equiv \{(0, 0, x_3) \in \mathbb{R}^3\}$. Because the matrix is symmetric, its range coincides with the orthogonal complement of its null space: $\text{range}(D) = \mathfrak{N}(D)_\perp$. In the following we refer to the null space of D and its orthogonal complement as $\mathfrak{N}$ and $\mathfrak{N}_\perp$.

The density corresponding to this matrix is

$$\mathcal{P}(x_1, x_2, x_3) \propto \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \tag{C.4}$$

and is not integrable, because we have

$$\int_{\mathbb{R}^3} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) = \int_{\mathbb{R}} dx_1 \exp\left(-\frac{1}{2}x_1^2\right) \int_{\mathbb{R}} dx_2 \exp\left(-\frac{1}{2}x_2^2\right) \int_{\mathbb{R}} dx_3 = +\infty.$$

This density has two important features:

1. It does not carry any information about the probability of the realization of values of x_3: it is ignorant with respect to x_3. More formally, we have:

   $$\mathcal{P}(x_1, x_2, x_3) = \mathcal{P}(x_1, x_2, x_3'), \quad \forall x_3, x_3' \in \mathbb{R}.$$

   The fact that the density is uniform in the direction of x_3 is what causes the density not to be integrable. The key observation here is that the direction along which the density is uniform coincides with the null space $\mathfrak{N}$.
2. Because the density C.4 does not depend on x_3, we can always set x_3 to 0 (or any other value) in its argument and look at it as a density over $\mathbb{R}^2$ rather than $\mathbb{R}^3$. In this case, the density is informative and proper:

   $$\int_{R^2} dx_1 dx_2 \mathcal{P}(x_1, x_2, 0) < +\infty. \tag{C.5}$$

   The set of points in $\mathbb{R}^3$ of the form $(x_1, x_2, 0)$ is $\mathfrak{N}_\perp$, the orthogonal complement of the null space.

We conclude from this simple analysis that the support of the density C.4, that is, $\mathbb{R}^3$, can be decomposed in two orthogonal parts: one is the null space $\mathfrak{N}$, over which the density is flat and which represents our ignorance; and the other is $\mathfrak{N}_\perp$, the orthogonal complement of $\mathfrak{N}$, over which the density is proper and carries information.

More formally, every vector $x \in \mathbb{R}^3$ can be decomposed as follows:

$$x = (x_1, x_2, x_3) = (x_1, x_2, 0) + (0, 0, x_3) = P_\perp x + P_\circ x,$$

where $P_\perp$ is the projector onto $\mathfrak{N}_\perp$, and $P_\circ = I - P_\perp$ is the projector onto the null space $\mathfrak{N}$. In our case, we have

$$P_\perp \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Using this notation, we can summarize what we have learned so far as:

$$\mathcal{P}(P_\circ x) = \text{ is uniform on } \mathfrak{N} \qquad \mathcal{P}(P_\perp x) = \mathcal{P}(x) \text{ is proper on } \mathfrak{N}_\perp.$$

We conclude from these observations that while expected values of functions of x cannot be computed under the density C.4, because x as a random variable is ill defined, expected values of functions of $P_\perp x$ are well defined, as long as we remember that, in order to compute them, we must restrict the support of the density to $\mathfrak{N}_\perp$. For example, while the covariance matrix of x does not exist, the covariance matrix of $P_\perp x$ is well defined, because it is the covariance matrix of the density $\mathcal{P}(x_1, x_2) \propto \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right)$, which is the identity.

C.3 The General Case

We now move to the general case, where the null space is a generic subspace of $\mathbb{R}^d$. As we have seen, meaningful expectations can be computed if we restrict the support of $\mathcal{P}$ to $\mathfrak{N}_\perp$. The only difficulty here is that computing integrals over the subspace $\mathfrak{N}_\perp$ is a bit more complicated. As an exercise, and in order to convince ourselves that the density C.1 is indeed well defined on $\mathfrak{N}_\perp$, we compute its integral and show that it is finite. Therefore, we want to compute the quantity:

$$K = \int_{\mathfrak{N}_\perp} dx \exp\left(-\frac{1}{2}\theta x' D x\right).$$

In the following, we denote by r the rank of D, and by $n = d - r$ the dimensionality of its null space. We adopt the convention that the eigenvalues of D, that is, the diagonal elements of L in equation C.2, are sorted in descending order, so that $\ell_1 \geq \ell_2 \geq \ldots \ell_r > \ell_{r+1} = \ell_{r+2} = \cdots = \ell_{r+n} = 0$. We start by substituting equation C.2 in the preceding equation:

$$K = \int_{\mathfrak{N}_\perp} dx \exp\left(-\frac{1}{2}\theta (R'x)' L R' x\right).$$

This expression suggests that we perform the change of variable $R'x = y$, which makes the integrand a product of independent terms. The Jacobian of this transformation is one since R is a rotation matrix. The only thing left to do is to figure out how the domain of integration, $\mathfrak{N}_\perp$, changes under this transformation. Remember that $\mathfrak{N}_\perp$ is the subspace

of vectors which are orthogonal to the null space, that is, such that $P_\circ x = 0$. Because $P_\circ = I - P_\perp$, we have

$$\mathfrak{N}_\perp = \{x \in \mathbb{R}^d \mid x = P_\perp x\}.$$

This can be rewritten in terms of the variable y as

$$\mathfrak{N}_\perp = \{y \in \mathbb{R}^d \mid Ry = P_\perp Ry\} = \{y \in \mathbb{R}^d \mid y = R'P_\perp Ry\}.$$

In order to understand the form $R'P_\perp Ry$, denote by r_i the columns of R. As shown in section B.2.2, the first r columns form a basis for range$(D) = \mathfrak{N}_\perp$, and the last n columns form a basis for $\mathfrak{N}$. Denoting by y_i the components of y, we have

$$Ry = \sum_{i=1}^{d} y_i r_i \ , \quad \Rightarrow P_\perp Ry = \sum_{i=1}^{d} y_i P_\perp r_i = \sum_{i=1}^{r} y_i r_i,$$

where we have used the fact that $P_\perp r_i = r_i$ if $r_i \in \mathfrak{N}_\perp$ and $P_\perp r_i = 0$ if $r_i \in \mathfrak{N}$. Therefore,

$$R'P_\perp Ry = \sum_{i=1}^{r} y_i R'r_i = \sum_{i=1}^{r} y_i e_i = (y_1, y_2, \ldots, y_r, 0, \ldots, 0),$$

where e_i is the vector $(0, \ldots, 0, 1, 0, \ldots 0)$, with the 1 in the i-th place, and we have used the fact that the columns of R are mutually orthogonal. We conclude that the only vectors y such that $y = R'P_\perp Ry$ are those whose last n components are zero, and therefore:

$$\mathfrak{N}_\perp = \{y \in \mathbb{R}^d \mid y_{r+1} = y_{r+2} = \cdots = y_{r+n} = 0\}.$$

This last expression implies that when we integrate over the variable y, we should keep the values of the last n components of y fixed at 0, while the first r components of y are unconstrained. Because integrating over a variable held constant is equivalent to not integrating over that variable, we summarize this finding as

$$\int_{\mathfrak{N}_\perp} dx = \int_{\mathbb{R}^d} dy \prod_{i=r+1}^{d} \delta(y_i), \quad x = Ry, \tag{C.6}$$

where $\delta(y_i)$ stands for the probability density (as a function of y_i) with unit mass at the origin. Now we complete our change of variables, and rewrite

$$K = \int_{\mathbb{R}^d} dy \prod_{i=r+1}^{d} \delta(y_i) \exp\left(-\frac{1}{2}\theta y'Ly\right).$$

The exponent in this expression does not depend on $y_{r+1}, \ldots, y_{r+n}$, and so simplifies to

$$K = \int_{\mathbb{R}^r} dy_1 \ldots dy_r \exp\left(-\frac{1}{2}\theta \sum_{i=1}^{r} \ell_i y_i^2\right) = \prod_{i=1}^{r} \left[\int_{\mathbb{R}} dy_i \exp\left(-\frac{1}{2}\theta \ell_i y_i^2\right)\right].$$

From the Gaussian integral identity in equation B.28, we have

$$\int_{\mathbb{R}} dy_i \exp\left(-\frac{1}{2}\theta \ell_i y_i^2\right) = \left(\frac{2\pi}{\theta \ell_i}\right)^{\frac{1}{2}}.$$

Putting everything together, we obtain

$$K = \left(\frac{2\pi}{\theta}\right)^{\frac{r}{2}} \frac{1}{\sqrt{\prod_{i=1}^{r} \ell_i}}.$$

Defining the quantity:

$$\det D_\perp \equiv \prod_{i=1}^{r} \ell_i,$$

we can finally rewrite the preceding expression as the following identity:

$$\int_{\mathfrak{N}_\perp} dx \exp\left(-\frac{1}{2}\theta x' D x\right) = \left(\frac{2\pi}{\theta}\right)^{\frac{r}{2}} \frac{1}{\sqrt{\det D_\perp}}. \tag{C.7}$$

Notice the complete analogy between this expression and the identity in equation B.28. The preceding number can be thought of as the normalization constant for the density C.1 restricted to $\mathfrak{N}_\perp$, which we present here for completeness:

$$\mathcal{P}(x;\theta) = \left(\frac{\theta}{2\pi}\right)^{\frac{r}{2}} \sqrt{\det D_\perp} \exp\left(-\frac{1}{2}\theta x' D x\right), \quad x \in \mathbb{R}^d \ , \quad x \in \mathfrak{N}_\perp. \tag{C.8}$$

Using a similar technique, we can now compute the expected values of several quantities of interest. In the following, we denote by $\mathrm{E}_\perp[\cdot]$ the expected value with respect to the density C.8. It is possible to show that the covariance matrix is

$$\mathrm{E}_\perp[xx'] = \frac{1}{\theta} D^+, \tag{C.9}$$

where D^+ is the pseudo-inverse of D, as defined in section B.2.5 (page 235). Using this result, we can easily see that

$$\mathrm{E}_\perp[x' D x] = \frac{r}{\theta}. \tag{C.10}$$

This last result can also be derived by first computing the distribution of the quantity $H = x' D x$ and then computing its expected value. The quantity H, which in general is interpreted as a smoothness functional, has the following distribution:

$$\mathcal{P}(H) = K H^{\frac{r}{2}-1} e^{-\frac{1}{2}\theta H}. \tag{C.11}$$

In the notation of chapter 9, we would write $H \sim \mathcal{G}(r, \theta)$.

C.4 Drawing Random Samples

Here we consider the problem of sampling from the improper normal density of equation C.1. We have seen in the previous sections that this density is not proper, but it becomes proper if we restrict its argument to the r-dimensional subspace $\mathfrak{N}_\perp$, that is, the portion of $\mathbb{R}^d$ on which the prior is informative. This restriction means that we should think of the argument x of the density as a linear combination of r independent elements $\mathfrak{N}_\perp$. We know from results about the eigenvectors-eigenvalue decomposition (section B.2.2, page 231) that a basis for $\mathfrak{N}_\perp$ is given by the first r columns of R (where $D = RLR'$), which we collect as columns of a $d \times r$ matrix $R_\perp$. Therefore, the restriction of the domain of the improper density $P(x;\theta)$ is simply expressed as $x = R_\perp z,\ \ z \in \mathbb{R}^r$. The role of the matrix $R_\perp$ here is simply to take $\mathbb{R}^r$ and rotate it to obtain $\mathfrak{N}_\perp$. We notice that this transformation is invertible ($z = R'_\perp x$) and that, because it is a rotation, its Jacobian is 1. Therefore, the density of the variable z is

$$\mathcal{P}(z;\theta) \propto \exp\left(-\frac{1}{2}\theta(R_\perp z)' D R_\perp z\right) = \exp\left(-\frac{1}{2}\theta z' R'_\perp D R_\perp z\right).$$

This probability density is now well defined, and therefore all we have to do in order to sample from the improper normal of equation C.1 is the following:

$$x = R_\perp z\ , \quad z \sim \mathcal{N}\left(0, \frac{1}{\theta}\left(R'_\perp D R_\perp\right)^{-1}\right). \tag{C.12}$$

The covariance of the variable z in the preceding expression is unnecessarily complicated and can be greatly simplified. To this end, it suffices to remember that, from equation B.19, the matrix D can be written as $D = R_\perp \ell R'_\perp$, where ℓ is the $r \times r$ diagonal matrix whose diagonal elements are the nonzero eigenvalues of D. Substituting this formula in equation C.12, and remembering that $R'_\perp R_\perp = I$, we arrive at the following simple formula:

$$x = R_\perp z\ , \quad z \sim \mathcal{N}\left(0, \frac{1}{\theta}\ell^{-1}\right). \tag{C.13}$$