Chapter 4
Random Variables and the Bayes formula

For the sake of clarity we consider the state space S (and also T) as the standard Euclidean space $\mathbb{R}^n$ for now. However, results developed below that do not involve probability densities (with respect to the Lebesgue measure) are also valid for general (e.g. infinite dimensional) state space S. We are in position to introduce the key player, the *random variable*.

Definition 4.1. A random variable m is a *measurable* map[1] from the sample space Ω to the state space S (with $\mathscr{S}$ as its σ-algebra)

$$m : \Omega \ni \omega \mapsto m(\omega) \in S.$$

We call $m(\omega)$ a random variable since we are uncertain about its outcome. In other words, we admit our ignorance about m by calling it a random variable. This ignorance is in turn a direct consequence of the uncertainty in the outcome of elementary event ω.

Definition 4.2. The *probability distribution* (or distribution or law for short) of a random variable m is defined as

$$\mu_m(A) \stackrel{\text{def}}{=} \mathbb{P}\left[m^{-1}(A)\right] = \mathbb{P}\left[\{m \in A\}\right], \quad \forall A \in \mathscr{S}, \tag{4.1}$$

where we have used the popular notation[2]

$$m^{-1}(A) \stackrel{\text{def}}{=} \{m \in A\} \stackrel{\text{def}}{=} \{\omega \in \Omega : m(\omega) \in A\}.$$

From the definition, we can see that the distribution is a probability measure[3] on S. In other words, the random variable m induces a probability measure, defined as μ_m, on the state space S. The key property of the induced probability measure μ_m is

Do you see why?

[1] A map is measurable if through its the inverse a measurable set in S is mapped into a measurable set in Ω.

[2] Rigorously, A must be a measurable subset of S.

[3] In fact, it is the push-forward measure by the random variable m.

the following. The probability for an event A in the state space to happen, denoted as $\mu_m(A)$, is defined as the probability for an event $B = m^{-1}(A)$ in the sample space to happen (see Figure 4.1 for an illustration). The distribution and the *probability density* (unless otherwise stated, density is understood with respect to the Lebesgue measure λ on S) π_m of m obey the following relation

$$\boxed{\mu_m(A) \stackrel{\text{def}}{=} \int_A \pi_m(m)\,\lambda(dm) \stackrel{\text{def}}{=} \mathbb{P}[\{m \in A\}], \quad \forall A \subset \mathscr{S}}. \tag{4.2}$$

Do you see this?

where the second equality of the definition is from (4.1). The meaning of random variable $m(\omega)$ can now be seen in Figure 4.1. It maps the event $B \in \Omega$ into the set $A = m(B)$ in the state space such that the "area" under the density function $\pi_M(m)$ and above A is exactly the probability that B happens.

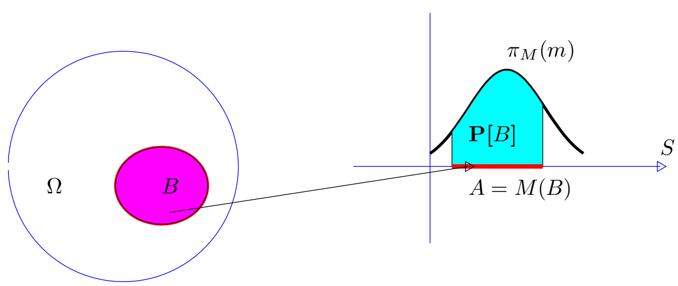

Fig. 4.1 Demonstration of random variable: $\mu_m(A) = \mathbb{P}[B]$.

We deduce the change of variable formula

$$\boxed{d\mu_m(m) \stackrel{\text{def}}{=} \mu_m(dm) = \pi(m)\,\lambda(dm) \stackrel{\text{def}}{=} \pi(m)\,d\lambda(m)},$$

where we write $\pi(m)$ instead of $\pi_m(m)$ if there is no ambiguity. *Note that $d\mu_m(m)$ is nothing more than the differential area under the curve $\pi(m)$ around m.* From an analogy to the differential calculus, we can formally rewrite the above formula as

$$\boxed{\frac{d\mu_m}{d\lambda}(m) = \pi(m)}, \tag{4.3}$$

which is an instance of the Radon-Nikodym derivative. In fact, one can show that ordinary derivative is a special case of the Radon-Nikodym derivative.

Exercise 4.1 (Ordinary derivative is a special case of Radon-Nikodym derivative). Take $f(x) = F'(x)$, where F is defined in Theorem 4.2. Show that

$$\frac{d\nu}{d\lambda}(x) = f(x) = F'(x),$$

where ν is the measure in Theorem 4.2, and λ is the standard Lesbegue measure on $\mathbb{R}$. •

Thus, *when we say $\pi(m)$ is the density of the random variable m, unless otherwise stated, we implicity mean that the Radon-Nikodym derivative of its probability measure with respect to the Lebesgue measure is $\pi(m)$.*

Definition 4.9 shows that if μ_r and λ obeys (4.3), then μ_r is *absolutely continuous* with respect to λ. *Do you see this?*

Remark 4.1. In theory, we introduce the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ in order to compute the probability of a subset[4] A in the state space S, and this is essentially the meaning of (4.1). However, once we know the probability density function $\pi_m(m)$, we can operate directly on the state space S without the need for referring back to probability space $(\Omega, \mathscr{F}, \mathbb{P})$, as shown in definition (4.2). This is the key observation, a consequence of which is that we simply ignore the underlying probability space in practice, since we don't need them in computation of probability in the state space. However, to intuitively understand the source of randomness, we need to go back to the probability space where the outcome of all events, except Ω, is uncertain. As a result, the pair $(S, \pi_m(m))$ contains complete information describing our ignorance about the outcome of random variable m. To the rest of this note, we shall work directly on the state space.

Remark 4.2. At this point, one may wonder what is the point of introducing the abstract probability space $(\Omega, \mathscr{F}, \mathbb{P})$ because it seems to unnecessarily make life more complicated? Well, its introduction is two fold. First, as discussed above, the probability space not only shows the origin of randomness but also provides the probability measure $\mathbb{P}$ for the computation of the randomness; it is also used to define random variables and furnishes a decent understanding about them. Second, the concepts of distribution and density in (4.1) and (4.2), which are introduced for random variable m, a measurable map from Ω to S, are valid for measurable maps[5] from an arbitrary space V to another space W. Here, W plays the role of S, and V the role of Ω on which we have a probability measure. For example, later in Section 5.1, we introduce the parameter-to-observable map $h(m): S \to \mathbb{R}^r$, then S plays the role of Ω and $\mathbb{R}^r$ of S in (4.1) and (4.2).

---

[4] Again, it needs to be measurable.

[5] Again, they must be measurable.

Definition 4.3. The *expectation* or the *mean* of a random variable m is the quantity

$$\mathbb{E}[m] \stackrel{\text{def}}{=} \int_S m\pi(m)\, dm = \overline{m}, \tag{4.4}$$

and the *variance* is

$$\mathbb{V}ar[m] \stackrel{\text{def}}{=} \mathbb{E}\left[(m-\overline{m})^2\right] \stackrel{\text{def}}{=} \int_S (m-\overline{m})^2\, \pi(m)\, dm.$$

Example 4.1. Let $\Omega = [-2,2]$ and define

$$\mathbb{P}[A \in \Omega] = \int_A \frac{1}{4}\, d\Omega,$$

where $d\Omega$ is the standard Lesbegue measure. Define a random variable $m : \Omega \to \mathbb{R}$ as

$$m(\omega) = \begin{cases} 2 & \text{if } \omega \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

We can compute the mean of m as

$$\begin{aligned} \mathbb{E}[m] \stackrel{\text{def}}{=} \int_{\mathbb{R}} m\pi(m)\, dm = \int_{\mathbb{R}} m\, \mu_m = \int_\Omega m(\omega)\, \mathbb{P}[d\Omega] \\ = \int_\Omega \frac{1}{4} m(\omega)\, d\Omega = \int_0^2 \frac{1}{4} \times 2\, d\Omega = 1. \end{aligned}$$

$\triangle$

This example shows that in general one *should not expect* the random variable under consideration can take its mean as a realization.

Exercise 4.2. A Gaussian random variable has the density given as

$$\pi(m) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(m-\overline{m})^2\right),$$

where $\overline{m}$ and σ are known as the mean and the deviation of the Gaussian distribution.
Show that $\mathbb{E}[m] = \overline{m}$ and $\mathbb{V}ar[m] = \sigma^2$. •

As we will see, the Bayes formula for probability densities is about the joint density of two or more random variables. So let us define the joint distribution and joint density of two random variables here.

Definition 4.4. Denote μ_{my} and π_{my} as the joint distribution and density, respectively, of two random variables m with values in S and y with values in T (with $\mathscr{T}$ as its σ-algebra) defined on the same probability space, then the joint distribution function and the joint probability density, in the light of (4.2), satisfy

$$\mu_{my}(\{m \in A\}, \{y \in B\}) \stackrel{\text{def}}{=} \int_{A\times B} \pi_{my}(m,y)\, dmdy, \quad \forall A \times B \subset \mathscr{S} \times \mathscr{T}, \tag{4.5}$$

where the notation $A \times B \subset \mathscr{S} \times \mathscr{T}$ simply means that $A \in \mathscr{S}$ and $B \in \mathscr{T}$.

We say that m and y are independent if

$$\mu_{my}\left(\{m \in A\}, \{y \in B\}\right) = \mu_m(A)\,\mu_y(B), \quad \forall A \times B \subset \mathscr{S} \times \mathscr{T},$$

or if

$$\pi_{my}(m,y) = \pi_m(m)\,\pi_y(y).$$

Definition 4.5. The *marginal* density of m is the probability density of m when y may take on any value, i.e.,

$$\pi_m(m) = \int_T \pi_{my}(m,y)\,dy.$$

Similarly, the marginal density of y is the density of y regardless m, namely,

$$\pi_y(y) = \int_S \pi_{my}(m,y)\,dm.$$

Before deriving the Bayes formula, we define conditional density $\pi(m|y)$ in the same spirit as (4.2) as

$$\mu_{m|y}\left(\{m \in A\}|y\right) = \int_A \pi(m|y)\,dm.$$

Let us prove the following important result.

Theorem 4.1. *The conditional density of m given $y = y$ is given by*

$$\boxed{\pi(m|y) = \frac{\pi(m,y)}{\pi(y)}}.$$

Proof. From the definition of conditional probability (3.3), we have

$$\begin{aligned}
\mu_{m|y}\left(\{m \in A\}|y\right) &= \mathbb{P}\left[\{m \in A\}|y = y\right] && \text{(definition (4.1))}\\
&= \lim_{\Delta y \to 0} \frac{\mathbb{P}\left[\{m \in A\}, y \le y' \le y + \Delta y\right]}{\mathbb{P}\left[y \le y' \le y + \Delta y\right]} && \text{(definition (3.3))}\\
&= \lim_{\Delta y \to 0} \frac{\int_A \pi(m,y)\,dm\Delta y}{\pi(y)\,\Delta y} && \text{(definitions (4.2), (4.5))}\\
&= \int_A \frac{\pi(m,y)}{\pi(y)}\,dm,
\end{aligned}$$

which ends the proof.

By symmetry, we have

$$\pi(m,y) = \pi(m|y)\,\pi(y) = \pi(y|m)\,\pi(m),$$

from which the well-known Bayes formula for finite dimensional state spaces follows

$$\boxed{\pi(m|y) = \frac{\pi(y|m)\,\pi(m)}{\pi(y)}}. \tag{4.6}$$

Exercise 4.3. Prove directly the Bayes formula for conditional density (4.6) using the Bayes formula for conditional probability (3.4). •

Definition 4.6 (Likelihood). We call $\pi(y|m)$ the likelihood. It is the probability density of y given $m = m$.

Definition 4.7 (Prior). We call $\pi(m)$ the prior. It is the probability density of m regardless y. The prior encodes, in the Bayesian framework, all information before any observations/data are made.

Definition 4.8 (Posterior). The density $\pi(m|y)$ is called the *posterior*, the distribution of parameter m given the measurement $y = y$, and it is the solution of the Bayesian inverse problem under consideration.

4.1 Appendix

Theorem 4.2. *Suppose $F : \mathbb{R} \to \mathbb{R}$ is: i) non-decreasing, and 2) right continuous. There is a unique measure ν on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that*

$$\nu((a,b]) = F(b) - F(a), \quad \forall a, b,$$

where $\mathscr{B}(\mathbb{R})$ is the Borel algebra on $\mathbb{R}$.

Definition 4.9 (Absolutely continuous of measures). We say ν is absolutely continuous with respect to μ, denoted by $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$.