

Notes on Probability

Remarks: For academic use only.

Probability theory

These notes will take for granted some familiarity with abstract (Lebesgue) integration theory. For further reading, see Chung (2001).

1 Probability spaces and random variables

A **probability space** (Ω, \mathcal{F}, P) where Ω is a non-empty set, \mathcal{F} is a σ -algebra of subsets of Ω and $P : \mathcal{F} \rightarrow [0, 1]$ is a (positive) measure space such that

$$P(\Omega) = 1.$$

An **event** is a set $F \in \mathcal{F}$.

A **random variable** $X : \Omega \rightarrow \mathbb{R}$ is an \mathcal{F} -measurable real-valued function on Ω .

A random variable is said to have a property **almost surely** (P) if it has the property on a set F with $P(F) = 1$. We often abbreviate and write **a.s.**

The **expected value** $E[X]$ of a random variable X is defined via

$$E[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{\Omega} X(\omega) P(d\omega).$$

1.1 Distribution measures and distribution functions

The **distribution measure** of a random variable is a measure on the Borel σ -algebra of subsets of \mathbb{R} that tells you what the probability is that $X(\omega) \in B \subset \mathbb{R}$. That is, for any Borel set $B \subset \mathbb{R}$,

$$\mu(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}).$$

Remark. We will sometimes write $\{X \in B\}$ when we mean $X^{-1}(B)$.

The **distribution function** $F : \mathbb{R} \rightarrow [0, 1]$ of a random variable X is defined via

$$F_X(x) = \mu((-\infty, x]).$$

That is, $F_X(x)$ is the probability that $X(\omega) \leq x$.

If F_X is absolutely continuous, then it has a **density** $f : \mathbb{R} \rightarrow \mathbb{R}_+$ such that

$$F(x) = \int_{-\infty}^x f(y) dy.$$

In particular, if F_X is differentiable everywhere, then $f_X(x) = F'_X(x)$.

1.2 Information and σ -algebras

When considering σ -algebras $\mathcal{G} \subset \mathcal{F}$ one may interpret \mathcal{G} as the amount of available information. Intuitively, if our information is given by \mathcal{G} , we can distinguish between the events in \mathcal{G} in the sense that for any event $G \in \mathcal{G}$ we know with perfect certainty whether or not it has occurred. Given this, it makes sense to say that if $\mathcal{G} \subset \mathcal{H}$, then \mathcal{H} contains no less information than \mathcal{G} . Also, it is tempting to say that $\mathcal{G} = \sigma\{\text{singletons}\}$ corresponds to *full* information since it should enable us to tell exactly what ω has been drawn. But this turns out to be an awkward way of defining full information in general although admittedly it makes perfect sense when Ω is a finite set. Instead, we will define full information as $\mathcal{G} = \mathcal{F}$, since then our information enables us to forecast perfectly the realized value of every random variable. Finally, we will say that $\mathcal{G} = \{\Omega, \emptyset\}$ (the trivial σ -algebra) corresponds to *no* information.

Alternatively, we might tell the following story. Suppose our σ -algebra \mathcal{G} is generated by a finite partition \mathbb{P} .

(i) Someone (Tyche, the norns, the dean, or whoever it is) chooses an outcome $\omega \in \Omega$ without telling us which.

(ii) While we don't know which $\omega \in \Omega$ has been chosen, we *are*, however, told (by an oracle, Hugin & Munin, or the Gazette or whatever) in which component $P_k \in \mathbb{P}$ ω lies. In practice, this could be arranged by allowing us to observe a stochastic variable defined via

$$X(\omega) = \sum_{k=1}^n k I_{P_k}(\omega). \quad (1)$$

To flesh this out a little bit more, you may want to think that getting ‘more information’ in this context would correspond to having a ‘finer’ partition, where a partition \mathbb{Q} finer than \mathbb{P} arises from chopping up the components of \mathbb{P} . It follows, of course, that $\sigma(\mathbb{P}) \subset \sigma(\mathbb{Q})$, which was our original (and more general) definition of ‘more information’.

In any case, notice that the axioms that characterize a σ -algebra accord well with our intuitions about information. Obviously, we should know whether Ω , since it always occurs by definition. Also, if we know whether A , we should know whether not- A too. Similarly, if we know whether A and whether B , we should know whether $A \cup B$. Countable unions are perhaps a little trickier to motivate intuitively; they are there essentially for technical reasons. In particular, they allow us to prove various limit theorems which are part of the point of the Lebesgue theory.

In economic modelling, it is plausible to allow decisions to depend only upon the available information. Mathematically, this means that if the agent’s information is given by \mathcal{G} , then her decision must be a \mathcal{G} -measurable random variable. The interpretation of this is that the information in \mathcal{G} suffices to give us perfect knowledge of the decision. Thus when it is time for the agent to act, she knows precisely what to do.

At this stage it is worth thinking about what it means for a stochastic variable X to be \mathcal{G} -measurable. Intuitively, it means that the information in \mathcal{G} suffices in order to know the value $X(\omega)$. To make this more concrete, suppose that \mathcal{G} is generated by a partition \mathbb{P} . Then for X to be \mathcal{G} -measurable, X has to be constant on each element $P_k \in \mathbb{P}$. It follows that knowing which element P_k has occurred is enough

to be able to tell what the value of $X(\omega)$ must be.

As a further illustration of the fact that σ -algebras do a good job in modelling information, we have the following result.

Definition. Let $\{X_\alpha, \alpha \in I\}$ be a family of random variables. Then the σ -algebra *generated* by $\{X_\alpha, \alpha \in I\}$, denoted by $\sigma\{X_\alpha, \alpha \in I\}$ is the smallest σ -algebra \mathcal{G} such that all the random variables in $\{X_\alpha, \alpha \in I\}$ are \mathcal{G} -measurable.

Remark. Such a σ -algebra exists. (Recall the proof: consider the intersection of all σ -algebras on Ω such that $\{X_\alpha, \alpha \in I\}$ are measurable.)

Proposition. Let $\mathbb{X} = \{X_1, X_2, \dots, X_n\}$ be a finite set of random variables. Let Z be a random variable. Then Z is $\sigma\{\mathbb{X}\}$ -measurable iff there exists a Borel measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that, for all $\omega \in \Omega$,

$$Z(\omega) = f(X_1(\omega), X_2(\omega), \dots, X_n(\omega)). \quad (2)$$

Proof. The case when $\sigma\{\mathbb{X}\}$ is generated by a finite partition (i.e. when the mapping $T : \Omega \rightarrow \mathbb{R}^n$ defined via $T(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ is \mathcal{F} -simple) is not too hard and is left as an exercise. For the rest, see Williams (1991). ■

2 The conditional expectation

Intuitively, the conditional expectation is the best predictor of the realization of a random variable given the available information. By “best” we will mean the one that minimizes the mean square error.

A formal definition, that works for square integrable random variables, is given by the following.

Definition Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra and let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Then the conditional expectation $Y = E[X|\mathcal{G}]$ is the projection of X onto $L^2(\Omega, \mathcal{G}, P)$.

Remark. By the Hilbert space projection theorem, the conditional expectation solves

$$Y = \min_{Z \in \mathcal{L}^2(\Omega, \mathcal{G}, P)} E[(X - Z)^2]. \quad (3)$$

Remark. The conditional expectation is itself a random variable. Its value is uncertain because it depends on precisely which events $G \in \mathcal{G}$ actually occur. In other words, it is a (contingent) forecasting *rule* whose output (the forecast) depends on the content of the information revealed. For example, suppose our information set is such that we know whether the president has been shot. Then our actions may depend on whether he is or is not shot.

The projection-based definition is intuitively the most appealing one, but unfortunately it only applies to *square* integrable stochastic variables. One way to extend the definition to merely *integrable* stochastic variables is to note that \mathcal{L}^2 is dense in \mathcal{L}^1 and define $E[X|\mathcal{G}]$ as the limit of the sequence $\{E[X_n|\mathcal{G}]\}$ where $X_n \in \mathcal{L}^2$ and $X_n \rightarrow X$ (in \mathcal{L}^1). Another way is the following.

Proposition. Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra and let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Then there is an a.s. (P) unique integrable random variable Z such that

1. Z is \mathcal{G} -measurable and

$$2. \int_G X dP = \int_G Z dP \quad \text{for each } G \in \mathcal{G}.$$

Using this result, we define $E[X|\mathcal{G}] = Z$.

Proof. The Radon-Nikodym theorem. ■

Remark. Since the conditional expectation is only a.s. (P) unique, most of the equations below strictly speaking need a qualifying ‘a.s. (P) ’ appended to them to be true. But since this is a bit tedious, we adopt instead the convention that the statement $X = Y$ means $P(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1$. If two random variables W and Z both qualify as the conditional expectation $E[X|\mathcal{G}]$, then we will sometimes call them *versions* of $E[X|\mathcal{G}]$.

This \mathcal{L}^1 -based definition can be intuitively motivated independently of the projection-based definition in the following way. On events such that we know whether they have occurred, our best guess of X should track X perfectly.

In any case, it had better be true that our two definitions of the conditional expectation coincide when they both apply, i.e. on $\mathcal{L}^1 \cap \mathcal{L}^2 = \mathcal{L}^2$. They do. You may want to try to prove this for yourself.

Having defined the conditional expectation with respect to a σ -algebra, we now define the conditional expectation with respect to a family of stochastic variables.

Definition. Let $Y \in L^1(\Omega, \mathcal{F}, P)$ and let $\{X_\alpha, \alpha \in I\}$ be a family of random variables. Then the conditional expectation $E[Y|\{X_\alpha, \alpha \in I\}]$ is defined as $E[Y|\sigma\{X_\alpha, \alpha \in I\}]$

Since $E[Y|X]$ is a $\sigma(X)$ -measurable random variable, there is a Borel function f such that $E[Y|X] = f(X)$. Sometimes we use the notation $f(x) = E[Y|X = x]$

where the expression on the right hand side is defined by the left hand side.

Definition. Let $Y \in L^1(\Omega, \mathcal{F}, P)$ and let X be a stochastic variable. Then the function $E[Y|X = x]$ is defined as any Borel function $f : \mathbb{R} \rightarrow \mathbb{R}$ with the property that $f(X)$ is a version of $E[Y|X]$. Note that $E[Y|X = x]$ is not always uniquely defined, but that this does not matter in practice.

Having defined the conditional expectation, we now note some of its properties. Let the given probability space be (Ω, \mathcal{F}, P) .

Proposition. Let $\mathcal{G} = \{\Omega, \emptyset\}$. Then $E[X|\mathcal{G}] = E[X]$.

Proof. Exercise. ■

Proposition. Let X and Y be integrable random variables, let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra and let α, β be scalars. Then

$$E[\alpha X + \beta Y|\mathcal{G}] = \alpha E[X|\mathcal{G}] + \beta E[Y|\mathcal{G}] \quad (4)$$

Proof. Exercise. ■

Proposition [Law of iterated expectations]. Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and let $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$ be σ -algebras. Then

$$E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}] \quad (5)$$

Proof. We check that the left hand side satisfies the conditions for being the conditional expectation of X with respect to \mathcal{G} . Clearly it is \mathcal{G} -measurable. Now let $G \in \mathcal{G}$ and we have, since $\mathcal{G} \subset \mathcal{H}$ and consequently $G \in \mathcal{H}$,

$$\int_G E[E[X|\mathcal{H}]|\mathcal{G}] dP = \int_G E[X|\mathcal{H}] dP = \int_G X dP. \quad (6)$$

□

Corollary. Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra. Then $E[E[X|\mathcal{G}]] = E[X]$.

Proposition. Let X and Y be random variables such that XY is integrable. Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra and suppose X is \mathcal{G} -measurable. Then

1. $E[X|\mathcal{G}] = X$ and
2. $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$.

Proof. (1) is trivial. To prove (2), note first that, the right hand side is \mathcal{G} -measurable (why?). To show that the right hand side integrates to the right thing, suppose $X = I_G$ where $G \in \mathcal{G}$. Let $F \in \mathcal{G}$. Then

$$\begin{aligned}
 \int_F X E[Y|\mathcal{G}] dP &= \int_F I_G E[Y|\mathcal{G}] dP = \int_{G \cap F} E[Y|\mathcal{G}] dP = \\
 &= \{(G \cap F) \in \mathcal{G}\} = \int_{G \cap F} Y dP = \int_F I_G Y dP = \quad (7) \\
 &= \int_F XY dP
 \end{aligned}$$

To show the more general case, show it for simple functions and then use the Monotone Convergence Theorem.

■

We end the discussion of the conditional expectation by defining the conditional probability of an event. We then note with satisfaction that our formal definition substantiates our claim above that if our information is given by \mathcal{G} , then we know, for all events $G \in \mathcal{G}$ whether or not they have occurred.

Definition. Let (Ω, \mathcal{F}, P) be a probability space and let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra. Let $A \in \mathcal{F}$. Then the conditional $P(A|\mathcal{G})$ probability of A given \mathcal{G} is defined via

$$P(A|\mathcal{G}) = E[I_A|\mathcal{G}] \tag{8}$$

It follows from this definition (why?) that if $G \in \mathcal{G}$ then $P(A|\mathcal{G}) = 1$ when G occurs and $P(A|\mathcal{G}) = 0$ when it does not.

2.1 Stochastic processes

Let (Ω, \mathcal{F}, P) be a probability space. A stochastic process in discrete time is a mapping $X : \mathbb{Z}_+ \times \Omega \rightarrow \mathbb{R}$ such that, for each fixed $t \in \mathbb{Z}_+$, the mapping $\omega \rightarrow X(t, \omega)$ is a random variable. For each fixed $\omega \in \Omega$, the mapping $t \rightarrow X(t, \omega)$ is called a **trajectory**.

The definition of a stochastic process in continuous time is the same, except that \mathbb{Z}_+ is replaced by \mathbb{R}_+ .