Decision Making under Uncertainty

Part 1: Introduction to probability

Christos Dimitrakakis

April 4, 2014

1 Probability

Two notions of probability

While probability is a simple mathematical construction, philosophically it has had at least two different meanings. In the classical sense, a probability distribution is a description for a truly random event. In the subjectivist sense, probabilty is merely a description for uncertainty which may or may not be due to randomness.

Classical Probability

- A *random experiment* is performed, with a given set $\mathcal{S}$ of possible outcomes. A simple example is the 2-slit experiment in physics, where a particle is generated and which can go through either one of two slits. According to our current understanding of quantum theory, it is impossible to predict which slit the particle will go through. There, the set of possible events correspond to the particle passing through one or the other slit.
- We care about the probability that the particle will go through one of the two slots in the experiment. Does it depend on where the other particles have passed through? In the 2-slit experiment, the probabilities of either event can be actually accurately calculated. However, which slit the particle will go through is fundamentally unpredictable.

Such quantum experiments are the only ones that are currently thought of as truly random (though some people disagree about that too). Any other procedure, such as tossing a coin or casting a die, is inherently deterministic and only *appears* random due to the difficulty in predicting the outcome. That is, modelling a coin toss as a random process is usually the best approximation we can make in practice, given our uncertainty about the complex dynamics involved.

Subjective Probability

- We assume that $\mathcal{S}$ is a set of possible *worlds* or realities, This set can be quite large and include anything imaginable. For example, it may include worlds where dragons are real. However, in practice one only cares about certain aspects of the world.
- We can interpret the probability of a world in $\mathcal{S}$ as a belief that it is the true world.

In such a setting there is an actual true world $\omega^* \in \mathcal{S}$, which is simply unknown. This could have been set by Nature to an arbitrary value deterministically. The probability only reflects our lack of knowledge.

1.1 Sets, experiments and sample spaces

Set theory definitions

A very useful way to describe a set A is as follows

$$A \triangleq \{x \mid x \text{ have property } Y\}$$

for example

$$B(c, r) \triangleq \{x \in \mathbb{R}^n \mid \|x - c\| \leq r\}$$

describes the set of points enclosed in an n-dimensional sphere of radius r with center $c \in \mathbb{R}^n$.

- If an element x *belongs* to a set A, we write $x \in A$.
- Let the *sample space* $\mathcal{S}$ be a set such that $x \in \mathcal{S}$ always.
- We say that A is a *subset* of B or that B *contains* A, and write $A \subset B$, iff, $x \in B$ for any $x \in A$.
- Let $B \setminus A \triangleq \{x \mid x \in B \wedge x \notin A\}$ be the set difference.
- Let $A \triangle B \triangleq (B \setminus A) \cup (A \setminus B)$ be the symmetric set difference.
- The *complement* of any $A \subset \mathcal{S}$ is $A^{\complement} \triangleq \mathcal{S} \setminus A$.
- The *empty set* is $\emptyset = \mathcal{S}^{\complement}$.
- The *union* of n sets: $A_1, \ldots, A_n$ is $\bigcup_{i=1}^n A_i = A_1 \cup \cdots \cup A_n$.
- The *intersection* of n sets $A_1, \ldots, A_n$ is $\bigcap_{i=1}^n A_i = A_1 \cap \cdots \cap A_n$.
- A and B are *disjoint* if $A \cap B = \emptyset$.

Experiments and sample spaces

Experiments

The set of possible experimental outcomes of an experiment is called the *sample space* $\mathcal{S}$.

- $\mathcal{S}$ must contain all possible outcomes.
- Each statistician i may consider a different $\mathcal{S}_i$ for the same experiment.

Example 1.1. *Experiment: give medication to a patient.*

- $\mathcal{S}_1 = \{$*Recovery within a day*, *No recovery after a day*$\}$.
- $\mathcal{S}_2 = \{$*The medication has side-effects*, *No side-effect*$\}$.
- $\mathcal{S}_3 =$ *all combinations of the above.*

Product spaces

- We perform n experiments.
- Assume that the i-th experiment has sample space $\mathcal{S}_i$.
- The *Cartesian product* or *product space* is defined as

$$\mathcal{S}_1 \times \cdots \times \mathcal{S}_n = \{(s_1, \ldots, s_n) \mid s_i \in \mathcal{S}_i, \forall i \in \{1, \ldots, n\}\} \tag{1.1}$$

  the set of all ordered n-tuples $(s_1, \ldots, s_n)$.
- The sample space $\prod_{i=1}^n \mathcal{S}_i$ can be thought of as a sample space of a *composite* experiment in which all n experiments are performed.

Identical experiment sample spaces

- In many cases, $\mathcal{S}_i = \mathcal{S}$ for all i, i.e. the sample space is identical for all individual experiment (e.g. n coin tosses).
- We then write $\mathcal{S}^n = \prod_{i=1}^n \mathcal{S}$.

1.2 Events, measure and probability

Events and probability

Probability of a set

If A is a subset of $\mathcal{S}$, the probability of A is a measure of the chances that the outcome of the experiment will be an element of A.

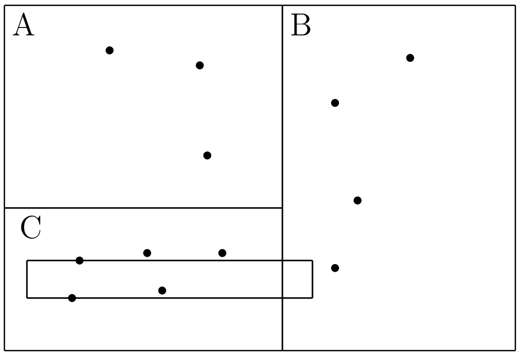

Figure 1: A fashionable apartment

Which sets?

Ideally, we would like to be able to assign a probability to *every subset of* $\mathcal{S}$. However, for technical reasons, this is not possible.

Example 1.2. *Let X be uniformly distributed on $[0, 1]$.*

- *What is the probability that X will be in $[0, 1/4)$?*
- *What is the probability that X will be in $[1/4, 1]$?*
- *What is the probability that X will be a rational number?*

Measure theory primer

Imagine that you have an apartment $\mathcal{S}$ composed of three rooms, A, B, C. There are some coins on the floor and a 5-meter-long red carpet. We can measure various things in this apartment.

Area

- A: $4 \times 5 = 20m^2$.
- B: $6 \times 4 = 24m^2$.
- C: $2 \times 5 = 10m^2$.

Coins on the floor

- A: 3.
- B: 4

- C: 5.

Length of red carpet

- A: $0m$
- B: $0.5m$
- C: $4.5m$.

Measure the sets: $\mathcal{F} = \{\emptyset, A, B, C, A \cup B, A \cup C, B \cup C, A \cup B \cup C\}$. It is easy to see that the union of any sets in $\mathcal{F}$ is also in $\mathcal{F}$. In other words, $\mathcal{F}$ is closed under union. Furthermore, $\mathcal{F}$ contains the whole space $\mathcal{S}$.

Note that all those measures have an *additive property*.

Measure and probability

As previously mentioned, the probability of $A \subset \mathcal{S}$ is a measure of the chances that the outcome of the experiment will be an element of A. Here we give a precise definition of what we mean by measure and probability.

Definition 1.1 (A field on $\mathcal{S}$). *A family $\mathcal{F}$ of sets, such that for each $A \in \mathcal{F}$, $A \subset \mathcal{S}$, is called a field on $\mathcal{S}$ if and only if*

1. $\mathcal{S} \in \mathcal{F}$
2. *if* $A \in \mathcal{F}$, *then* $A^{\complement} \in \mathcal{F}$.
3. *For any* $A_1, A_2, \ldots, A_n$ *such that* $A_i \in \mathcal{F}$, *it holds that:* $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$.

From the above definition, it is easy to see that $A_i \cap A_j$ is also in the field.

Definition 1.2 (σ-field on $\mathcal{S}$). *A family $\mathcal{F}$ of sets, such that $\forall A \in \mathcal{F}$, $A \subset \mathcal{S}$, is called a σ-field on $\mathcal{S}$ if and only if*

1. $\mathcal{S} \in \mathcal{F}$
2. *if* $A \in \mathcal{F}$, *then* $A^{\complement} \in \mathcal{F}$.
3. *For any sequence* $A_1, A_2, \ldots$ *such that* $A_i \in \mathcal{F}$, *it holds that:* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

It is easy to verify that the $\mathcal{F}$ given in the apartment example satisfies these properties.

Definition 1.3 (Measure). *A measure λ on $(\mathcal{S}, \mathcal{F})$ is a function $\lambda : \mathcal{F} \to \mathbb{R}^+$ such that*

1. $\lambda(\emptyset) = 0$.
2. $\lambda(A) \geq 0$ *for any* $A \in \mathcal{F}$.

3. *For any collection of subsets $A_1, \ldots, A_n$ with $A_i \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$.*

$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda(A_i) \tag{1.2}$$

It is easy to verify that the floor area, the number of coins, and the length of the red carpet are all measures. In fact, the area and length correspond to what is called a *Lebesgue measure* and the number of coins to a *counting measure.*

Definition 1.4 (Probability measure)**.** *A probability measure P on $(\mathcal{S}, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that:*

1. $P(\mathcal{S}) = 1$
2. $P(\emptyset) = 0$
3. $P(A) \geq 0$ *for any* $A \in \mathcal{F}$.
4. *If $A_1, A_2, \ldots$ are disjoint then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \tag{union}$$

$(S, \mathcal{F}, P)$ *is called a* probability space.

So, probability is just a special type of measure.

1.2.1 The Lebesgue measure

Definition 1.5 (Outer measure)**.** *Let $(\mathcal{S}, \mathcal{F}, \lambda)$ be a measure space. The outer measure of a set $A \subset \mathcal{S}$ is:*

$$\lambda^* \triangleq \inf A \subset \bigcup_k B_k \sum_k \lambda(B_k). \tag{1.3}$$

Definition 1.6 (Inner measure)**.** *Let $(\mathcal{S}, \mathcal{F}, \lambda)$ be a measure space. The outer measure of a set $A \subset \mathcal{S}$ is:*

$$\lambda_* \triangleq \lambda(\mathcal{S}) - \lambda(\mathcal{S} \setminus A). \tag{1.4}$$

Definition 1.7 (Lebesgue measurable sets)**.** *A set A is (Lebesgue) measurable if the outer and inner measures are equal.*

$$\lambda^*(A) = \lambda_*(B). \tag{1.5}$$

The common value of the inner and outer measure is called the Lebesgue measure[1] $\bar{\lambda} = \lambda^*(A)$.

[1]It is easy to see that $\bar{\lambda}$ is a measure.

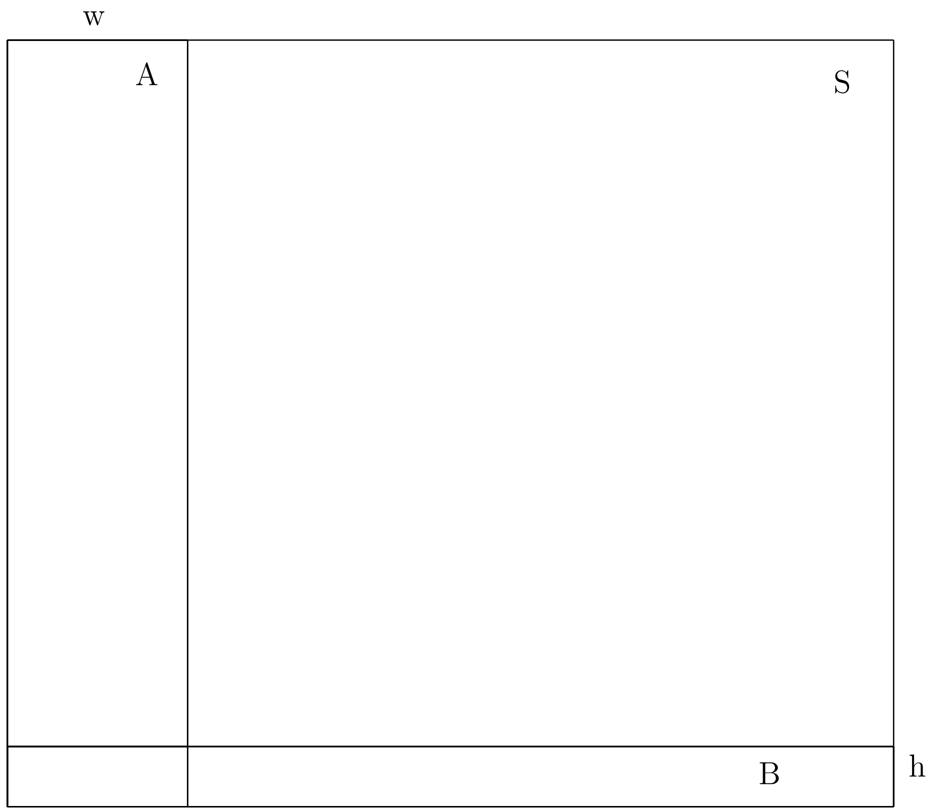

Figure 2: In the above case, S is a unit square and taking P to be the Lebesgue measure, we see that $P(S) = 1 \cdot 1$, $P(A) = 1 \cdot w$, $P(B) = h \cdot 1$ and $P(A \cap B) = wh$, so A and B are independent.

1.3 Conditioning and independence

Independent events and conditional probability

Events correspond to sets. Thus, the probability of the event that a draw from $\mathcal{S}$ is in A is equal to the probability measure of A, $P(A)$.

Definition 1.8 (Independent events)**.** *Two events A, B are independent if $P(A \cap B) = P(A)P(B)$. The events in a family $\mathcal{F}$ of events are independent if for any sequence $A_1, A_2, \ldots$ of events in $\mathcal{F}$,*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) \tag{independence}$$

Definition 1.9 (Conditional probability)**.** *The conditional probability of A when B, s.t. $P(B) > 0$, is given is:*

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}. \tag{1.6}$$

Of course, $P(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B)$ *even if* A, B *are not independent.*

Bayes' theorem

The following theorem trivially follows from the above discussion. However, versions of it shall be used repeatedly throughout. For this reason we present it here together with a detailed proof.

Theorem 1.1 (Bayes' theorem)**.** *Let* $A_1, A_2, \ldots$ *be a (possibly infinite) sequence of disjoint events such that* $\bigcup_{i=1}^{\infty} A_i = \mathcal{S}$ *and* $P(A_i) > 0$ *for all* i*. Let* B *be another event with* $P(B) > 0$*. Then*

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)P(A_j)} \tag{1.7}$$

Proof. From (1.6), $P(A_i \mid B) = P(A_i \cap B)/P(B)$ and also $P(A_i \cap B) = P(B \mid A_i)P(A_i)$. Thus

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)},$$

and we continue analyzing the denominator $P(B)$. First, due to $\bigcup_{i=1}^{\infty} A_i = \mathcal{S}$ we have $B = \bigcup_{j=1}^{\infty}(B \cap A_j)$. Since A_i are disjoint, so are $B \cap A_i$. Then from the union property of probability distributions we have

$$P(B) = P\left(\bigcup_{j=1}^{\infty}(B \cap A_j)\right) = \sum_{j=1}^{\infty} P(B \cap A_j) = \sum_{j=1}^{\infty} P(B \mid A_j)P(A_j),$$

which finishes the proof. ☐

Binomial coefficients

Binomial coefficients appear in a lot of different distributions. They are especially useful for combinatorial problems.

$$\binom{x}{n} \triangleq \frac{\prod_{i=0}^{n-1}(x-i)}{n!}, \qquad x \in \mathbb{R}, n \in \mathbb{N}, \tag{1.8}$$

and $\binom{x}{0} = 1$. It follows that

$$\binom{k}{n} = \frac{k!}{n!(k-n)!} \qquad k, n \in \mathbb{N}, k \geq n. \tag{1.9}$$

2 Random variables

Random variables

A random variable X is a special kind of random quantity, defined as a real function of outcomes in $\mathcal{S}$. Thus, it also defines a mapping from a probability measure P on $(\mathcal{S}, \mathcal{F})$ to a probability measure P_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. More precisely, we define the following.

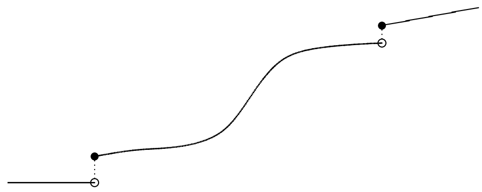

Figure 3: A distribution function F

Definition 2.1 (Measurable function)**.** *Let $\mathcal{F}$ on $\mathcal{S}$ be a σ-field. A function $g : \mathcal{S} \to \mathbb{R}$ is said to be measurable with respect to $\mathcal{F}$, or $\mathcal{F}$-measurable, if, for any $x \in \mathbb{R}$,*

$$\{s \in \mathcal{S} \mid g(s) \leq x\} \in \mathcal{F}.$$

Definition 2.2 (Random variable)**.** *Let $(\mathcal{S}, \mathcal{F}, P)$ be a probability space. A random variable $X : \mathcal{S} \to \mathbb{R}$ is a real-valued, $\mathcal{F}$-measurable function.*

The distribution of X

Every random variable X induces a probability measure P_X on $\mathbb{R}$. For any $B \subset \mathbb{R}$ we define

$$P_X(B) = \mathbb{P}(X \in B) = P(\{s \mid X(s) \in B\}). \tag{2.1}$$

Thus, the probability that X is in B is equal to the P-measure of the points $s \in \mathcal{S}$ such that $X(s) \in B$ and also equal to the P_X-measure of B.

Here $\mathbb{P}$ is used as a *short-hand* notation.

Exercise 1. *$\mathcal{S}$ is the set of 52 playing cards. $X(s)$ is the value of each card ($1, 10$ for the ace and figures respectively). What is the probability of drawing a card s with $X(s) > 7$?*

(Cumulative) Distribution functions

Definition 2.3 ((Cumulative) Distribution function)**.** *The distribution function of a random variable X is the function $F : \mathbb{R} \to \mathbb{R}$:*

$$F(t) = \mathbb{P}(X \leq t). \tag{2.2}$$

Properties

- If $x \leq y$, then $F(x) \leq F(y)$.
- F is right-continuous.
- At the limit,

$$\lim_{t \to -\infty} F(t) = 0, \qquad \lim_{t \to \infty} F(t) = 1.$$

2.1 Discrete and continuous random variables

Types of distributions

On the real line, there are two types of distributions for a random variable. Here, once more, we employ the $\mathbb{P}$ notation as a shorthand for the probability of general events involving random variables, so that we don't have to deal with the measure notation. The two following examples should give some intuition.

Discrete distributions

$X : \mathcal{S} \to \{x_1, \ldots, x_n\}$ takes n discrete values (n can be infinite). The probability function of X is

$$f(x) \triangleq \mathbb{P}(X = x),$$

defined for $x \in \{x_1, \ldots, x_n\}$. For any $B \subset \mathbb{R}$:

$$P_X(B) = \sum_{x_i \in B} f(x_i).$$

In addition, we write $\mathbb{P}(X \in B)$ to mean $P_X(B)$.

Continuous distributions

X has a continuous distribution if there exists a *probability density function* f s.t. $\forall B \subset \mathbb{R}$:

$$P_X(B) = \int_B f(x) \, \mathrm{d}x.$$

It is possible that X has neither a continuous, nor a discrete distribution.

2.2 Random vectors

Generalisation to $\mathbb{R}^m$

We can generalise to random *vectors* in a Euclidean space. Once more, there are two special cases of distributions for the random vector $X = (X_1, \ldots, X_n)$.

Discrete distributions

$$\mathbb{P}(X_1 = x_1, \ldots, X_m = x_m) = f(x_1, \ldots, x_m)$$

Continuous distributions

For $B \subset \mathbb{R}^m$

$$\mathbb{P}\left\{(X_1, \ldots, X_m) \in B\right\} = \int_B f(x_1, \ldots, x_m) \, \mathrm{d}x_1 \cdots \mathrm{d}x_m$$

Measure-theoretic notation

The previously seen special cases can be handled with a unified notation if we take advantage of the fact that probability is only a particular type of measure. As a first step, we note that summation can also be seen as integration with respect to the counting measure and that Riemann integration is integration with respect to the Lebesgue measure.

Integral with respect to a measure μ

Introduce the common notation $\int \cdots \mathrm{d}\mu(x)$, where μ is a measure. Let some real function $g : \mathcal{S} \to \mathbb{R}$. Then for any subset $B \subset \mathcal{S}$ we can write

- Discrete case: f is the probability function and we choose the *counting measure* for μ, so:

$$\sum_{x \in B} g(x) f(x) = \int_B g(x) f(x) \, \mathrm{d}\mu(x)$$

  Roughly speaking, the counting measure $\mu(\mathcal{S})$ is equal to the number of elements in $\mathcal{S}$.

- Continuous case: f is the probability density function and we choose the *Lebesgue measure* for μ, so:

$$\int_B g(x) f(x) \, \mathrm{d}x = \int_B g(x) f(x) \, \mathrm{d}\mu(x)$$

  Roughly speaking, the Lebesgue measure $\mu(S)$ is equal to the volume of S.

In fact, since probability is a measure in itself, we do not need to complicate things by using f and μ at the same time! This allows us to use the following notation.

Lebesgue-Stiletjes notation

If P is a probability measure on $(\mathcal{S}, \mathcal{F})$ and $B \subset \mathcal{S}$, and g is $\mathcal{F}$-measurable, we write the probability that $g(x)$ takes the value B can be written equivalently as:

$$\mathbb{P}(g \in B) = P_g(B) = \int_B g(x) \, \mathrm{d}P(x) = \int_B g \, \mathrm{d}P. \tag{2.3}$$

Intuitively, $\mathrm{d}P$ is related to densities in the following way. If P is a measure on $\mathcal{S}$ and is absolutely continuous with respect to another measure μ, then $p \triangleq \frac{\mathrm{d}P}{\mathrm{d}\mu}$ is the (Radon-Nikodyn) derivative of P with respect to μ. We write the integral as $\int gp \, \mathrm{d}\mu$. If μ is the Lebesgue measure, then p coincides with the probability density function.

Marginal distributions and independence

Although this is a straightforward outcome of the set-theoretic definition of probability, we also define the marginal explicitly for random vectors.

Marginal distribution

The marginal distribution of $X_1, \ldots, X_k$ from a set of variables $X_1, \ldots, X_m$, is

$$\mathbb{P}(X_1, \ldots, X_k) \triangleq \int \mathbb{P}(X_1, \ldots, X_k, X_{k+1} = x_{k+1}, \ldots, X_m = x_m) \, \mathrm{d}\mu(x_{k+1}, \ldots, x_m). \tag{2.4}$$

In the above, $\mathbb{P}(X_1, \ldots X_k)$ can be thought of as the probability measure for any events related to the random vector $(X_1, \ldots, X_k)$. Thus, it defines a probability measure over $\left(\mathbb{R}^k, \mathfrak{B}\left(\mathbb{R}^k\right)\right)$. In fact, let $Y = (X_1, \ldots, X_k)$ and $Z = (X_{k+1}, \ldots, X_m)$ for simplicity. Then define $Q(A) \triangleq \mathbb{P}(Z \in A)$, with $A \subset \mathbb{R}^{m-k-1}$. Then the above can be re-written as:

$$\mathbb{P}(Y \in B) = \int_{\mathbb{R}^{m-k-1}} \mathbb{P}(Y \in B \mid Z = z) \, \mathrm{d}Q(z).$$

Similarly, $\mathbb{P}(Y \mid Z = z)$ can be thought of as a function mapping from values of Z to probability measures. Let $P_z(B) \triangleq \mathbb{P}(Y \in B \mid Z = z)$ be this measure corresponding to a particular value of z. Then we can write

$$\mathbb{P}(Y \in B) = \int_{\mathbb{R}^{m-k-1}} \left(\int_B \mathrm{d}P_z(y) \right) \mathrm{d}Q(z).$$

Independence

If X_i is independent of X_j for all $i \neq j$:

$$\mathbb{P}(X_1, \ldots, X_m) = \prod_{i=1}^{M} \mathbb{P}(X_i), \qquad f(x_1, \ldots, x_m) = \prod_{i=1}^{M} g_i(x_i) \tag{2.5}$$

2.3 Moments

There are some simple properties of the random variable under consideration which are frequently of interest in statistics. Two of those properties are *expectation* and *variance*.

Expectation

Definition 2.4. *The expectation* $\mathbb{E}(X)$ *of any random variable* $X : \mathcal{S} \to R$*,where* R *is a vector space, with distribution* P_X *is defined by*

$$\mathbb{E}(X) \triangleq \int_R t \, \mathrm{d}P_X(t), \tag{2.6}$$

as long as the integral exists.

Furthermore,

$$\mathbb{E}[g(X)] = \int g(t) \, \mathrm{d}P_X(t),$$

for any function g.

Variance

Definition 2.5. *The variance $\mathbb{V}(X)$ of any random variable $X : \mathcal{S} \to \mathbb{R}$ with distribution P_X is defined by*

$$\begin{aligned} \mathbb{V}(X) &\triangleq \int_{-\infty}^{\infty} [t - \mathbb{E}(X)]^2 \, \mathrm{d}P_X(t) \\ &= \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} \\ &= \mathbb{E}(X^2) - \mathbb{E}^2(X). \end{aligned} \tag{2.7}$$

When $X : \mathcal{S} \to R$ with R an arbitrary vector space, the above becomes the *covariance matrix*:

$$\begin{aligned} \mathbb{V}(X) &\triangleq \int_{-\infty}^{\infty} [t - \mathbb{E}(X)] \, [t - \mathbb{E}(X)]^\top \, \mathrm{d}P_X(t) \\ &= \mathbb{E}\left\{[X - \mathbb{E}(X)] \, [X - \mathbb{E}(X)]^\top\right\} \\ &= \mathbb{E}(XX^\top) - \mathbb{E}(X) \, \mathbb{E}(X)^\top. \end{aligned} \tag{2.8}$$

Divergences
One useful idea is KL-divergences on measures.

Definition 2.6. *KL-Divergence*

$$D\,(P \parallel Q) \triangleq \int \frac{\mathrm{d}P}{\mathrm{d}Q} \, \mathrm{d}P. \tag{2.9}$$

Empirical distributions

Definition 2.7. *Let $x^n = (x_1, \dots, x_n)$ drawn from a product measure $x^n \sim P^n$ on the measurable space $(\mathcal{X}^n, \mathfrak{F}_n)$. Let $\mathfrak{S}$ be any σ-field on $\mathcal{X}$. Then empirical distribution of x^n is defined as*

$$\hat{P}_n(B) \triangleq \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}\left\{x_t \in B\right\}. \tag{2.10}$$

3 Conclusion

Recommended further reading
Most of this material is based on [2]. See [3] for a really clear exposition of measure, starting from rectangle areas (developed from course notes in 1957). Also see [4] for a verbose, but interesting and rigorous introduction to subjective probability. More technical books, such as [1] are not very approachable by non-math graduates.

Summary

- *Sample space* $\mathcal{S}$ contains all possible *outcomes* of an experiment.
- σ-field $\mathcal{F}$ s.t. $\forall A, B \in \mathcal{F}$, $A \subset \mathcal{S}$, $A \cup B \in \mathcal{F}$, $\mathcal{S} \in \mathcal{F}$.
- Measurable space $(\mathcal{S}, \mathcal{F})$, measure space $(\mathcal{S}, \mathcal{F}, \mu)$.
- *Measure* $\mu : \mathcal{F} \to \mathbb{R}$ such that $\mu(\emptyset) = 0$, and $\mu(A_i) \geq 0$ for any $A_i \in \mathcal{F}$. For *disjoint* A_i, $\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$.
- Probability space $(\mathcal{S}, \mathcal{F}, P)$, with *probability measure* P such that $P(\mathcal{S}) = 1$.
- *Probability* that $x \in A$:

$$\mathbb{P}(x \in A) \triangleq P(A) = \int_A \mathrm{d}P(t), \qquad A \subset \mathcal{S}$$

- *Expectation* of $X : \mathcal{S} \to Z$

$$\mathbb{E}(X) \triangleq \int_{\mathcal{S}} X(t) \, \mathrm{d}P(t) = \int_Z u \, \mathrm{d}P_X(u)$$

- *Conditional probability*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}, \quad P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

- *Marginal* distribution

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B, A = i), \quad \sum_i \mathbb{P}(A = i) = 1,$$

$$P(B) = \sum_i P(B \cap A_i), \quad \bigcup_i A_i = \mathcal{S}.$$

- If A, B are *independent*

$$\mathbb{P}(A, B) = \mathbb{P}(A) \, \mathbb{P}(B), \quad P(A \cap B) = P(A)P(B).$$

References

[1] Robert B. Ash and Catherine A. Doleéans-Dade. *Probability & Measure Theory*. Academic Press, 2000.

[2] Morris H. DeGroot. *Optimal Statistical Decisions*. John Wiley & Sons, 1970.

[3] AN Kolmogorov and SV Fomin. *Elements of the theory of functions and functional analysis*. Dover Publications, 1999.

[4] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, 1972.