Probability measures and expectations

All of the measure theory we've developed so far must of course apply to probability spaces. Here, we will build on that foundation by defining new terms and deriving some specific results that are essential to probability theory. You'll should find your intuition about probability and randomness helpful for the most part, but some results may be counter-intuitive. A probability space is an excellent mathematical structure for modeling uncertainty or "randomness." As such it is a crucial topic to study if you hope to draw sound conclusions about reality in light of your experience in this noisy world of ours.

1 Probability measures

In the context of probability spaces, we will often use special notation. Our coverage of finite measures is "limited" to probability measures. But this does not actually lose any generality, because *every finite measure can be thought of as a rescaled version of a probability measure.* In practice, rescale any and every finite measure that you're dealing with so that you can freely apply the toolbox and intuition of probability theory! Typically we will use the symbol $\mathbb{P}$ for a probability measure on Ω. We use capital letters (usually X, but otherwise in the range from U to Z) for vector-valued (or $\bar{\mathbb{R}}$-valued) functions on probability spaces. Breaking with De Finetti notation, we will denote the $\mathbb{P}$-integral by $\mathbb{E}$. I just can't help but prefer to distinguish $\mathbb{P}$robability and $\mathbb{E}$xpectation for the sake of my own intuition. If we write $\mathbb{E}_{X\sim P}g(X)$, we mean the $\mathbb{P}$-integral of $g \circ X$ where X has distribution P on its codomain; this is the same thing as the P-integral of g, so we could also represent it as Pg and avoid any reference to Ω or $\mathbb{P}$. Along these lines, $\mathbb{E}_{X\sim P}X$ would be denoted PI where I is the identity function.

WRITE this section.

2 Expectations

An essential intuition regarding any expectation operator is that it generalizes the idea of *weighted averaging.*

Theorem 2.1. *If X is a Pettis integrable function taking values in a real LCHS, then its expectation $\mathbb{E}X$ is in the closure of the convex hull of the range of X.*

1.1 Prove Theorem 2.1.

1.2 Suppose X maps almost every $\omega \in \Omega$ to a particular $v \in \mathbb{V}$. Assuming $\mathbb{V}'$ separates points, show that $\mathbb{E}X = v$.

1.3 Let $\mathbb{H}$ be a Hilbert space and let X be an $\mathbb{H}$-valued function that is Pettis integrable with respect to a probability measure $\mathbb{P}$. If the Banach space is separable, Pettis integrability implies Borel measurability, by the Pettis Measurability Theorem. Show that for any $h \in \mathbb{H}$, the BV-decomposition holds:

$$\mathbb{E}_L\|h - X\|^2 = \|h - \mathbb{E}X\|^2 + \mathbb{E}_L\|X - \mathbb{E}X\|^2$$

as long as the two squared-norms are measurable. It is sufficient for X to be measurable.

1.4* Let (X_n) be a sequence of Banach-space valued functions on a probability space. Suppose that for all $\epsilon > 0$, we are able to bound $\sum \mathbb{P}(\|X_n - X\| \geq \epsilon)$ by a finite number (which is allowed to depend on ϵ). Explain why the Borel-Cantelli Lemma lets us conclude that $X_n \overset{\text{a.s.}}{\to} X$.

Exercise **??** established that the integral is a linear operator on the space of integrable functions. In particular, this implies that if Ω is partitioned into n measurable subsets, then the integral of f is equal to the sum of the integrals taken with f and μ restricted to those subsets (if the integrals exist):

$$\begin{aligned}\mu f &= \mu\left(f\sum_{i=1}^n E_i\right)\\ &= \sum_{i=1}^n \mu\left(fE_i\right)\end{aligned}$$

In particular, when the measure in question is a PM, we often want to go a step further and express the expectation as an average of "conditional expectations," which we define implicitly in the following derivation:

$$\begin{aligned}\mathbb{E}X &= \mathbb{E}\left(X\sum_{i=1}^n E_i\right)\\ &= \sum_{i=1}^n \mathbb{E}(XE_i)\\ &= \sum_{i=1}^n (\mathbb{P}E_i)\int_{E_i}\tfrac{1}{\mathbb{P}E_i}X(\omega)d\mathbb{P}(\omega)\\ &= \sum_{i=1}^n (\mathbb{P}E_i)\int_{E_i}X(\omega)d\tfrac{\mathbb{P}}{\mathbb{P}E_i}(\omega)\\ &= \sum_{i=1}^n (\mathbb{P}E_i)\mathbb{E}_{E_i}X \qquad (1)\end{aligned}$$

We will have more to say on this topic in Section REF.

1.5 Suppose f is Bochner integrable. Explain why fE is also Bochner integrable for any $E \in \Sigma$.

1.6 Define $\tilde{X}$ to be the function that takes the constant value $\mathbb{E}_{E_i}X$ on E_i for $i \in \{1, \dots, n\}$ (assume these expectations exist). Show that $\mathbb{E}\tilde{X} = \mathbb{E}X$.

Let f map from a topological vector space to $\mathbb{R}$. We say that f is Jensen-convex if for every v in its domain there exists a continuous affine functional l such that $l(x) = f(x)$ and $l \leq f$.

1.7 Prove that every Jensen-convex function is convex. Show that every convex function with a finite-dimensional domain is Jensen-convex.

Convex functions play a key role in the study of optimization. Convexity is also important in probability theory due in part to Jensen's inequality.

Theorem 2.2 (Jensen's inequality)**.** *Let $\mathbb{V}$ be a topological vector space and X be a Pettis integrable $\mathbb{V}$-valued function. If f is Jensen-convex and $f \circ X$ is Borel measurable, then*

$$f(\mathbb{E}X) \leq \mathbb{E}_L f(X).$$

If f is strictly convex and X does not have a point-mass distribution, then the inequality is strict.

1.8 Prove Jensen's inequality.

1.9 Assume that f is a convex on an open subset of a normed space, and that at every point in its domain, its directional derivative is bounded. Notice that the directional derivatives don't have to be uniformly bounded; the bound can depend on the point. Show that f is Jensen-convex.

1.10 Let X be an integrable random variable, and let f and g be convex functions from $\mathbb{R}$ to $\mathbb{R}$. Is it true that

$$\mathbb{E}_L f(g(X)) \geq f(g(\mathbb{E}X)) \quad ?$$

2.1 Generalized expectations

In the definition of the $L^p_{\mathbb{X}}$-norm, the norm of the function is taken to the p power, then an integration occurs, then the original transformation is "undone" by taking the $1/p$ power. This pattern of transforming, integrating, then inverse transforming is worth a bit of attention.

1.11 For the space of real sequences, considered as functions from $\mathbb{N}$ to $\mathbb{R}$ with counting measure μ on $\mathbb{N}$, the $L^p(\mu)$-norm is also called the l^p-norm and denoted $\|\cdot\|_p$. In other words, $s := (s_n)$ has l^p-norm

$$\|s\|_p = \left(\sum_i |s_i|^p\right)^{1/p}.$$

The space of sequences with finite l^p-norm is also called the l^p-space (or simply l^p). Show that $l^p \subseteq l^q$ for $p \leq q$.

If $X : \Omega \mapsto \mathbb{V}$ and $\psi : \mathbb{V} \mapsto \mathbb{W}$ is injective and continuous on the range of X, then $\mathbb{E}^\psi X := \psi^{-1}[\mathbb{E}\psi(X)]$ (if it exists) will be called the ψ-expectation of X. Any such expectation might also be called a *generalized expectation*, a *quasi-arithmetic expectation*, or a *Kolmogorov expectation*.

1.12 Explain why the ψ-expectation is equivalent to the $T \circ \psi$-expectation for any injective continuous affine operator T.

1.13 Show that every ψ-expectation has the *averaging property* described in Theorem 2.1 as well as the *value-preservation property* of Exercise **1.2**.

1.14 Show that ψ-expectations have a *partition property* akin to derivation (1) and Exercise **1.6**.

1.15 In the context of random variables, $\psi : \mathbb{R} \to \mathbb{R}$ must be strictly monotonic on $X(\Omega)$ in order to be continuous and injective. Show that if $\psi : \mathbb{R} \to \mathbb{R}$ is increasing on $X(\Omega)$, then the ψ-expectation operator is increasing, i.e. $X \leq Y$ implies $\mathbb{E}^\psi X \leq \mathbb{E}^\psi Y$.

When X is a random variable and the transformation has the form $\psi(t) = t^p$, we denote the ϕ-expectation (using the Lebesgue integral, so that $\pm\infty$ is allowed) by $\mathbb{E}^p$ and call it a power expectation. In particular, if X is $\mathbb{R}^+$-valued, then $\mathbb{E}^p X$ is well-defined for every $p \in \mathbb{R}$. Based on limiting behavior, $\mathbb{E}^0$ is defined to be

$$\mathbb{E}^0 X := \exp(\mathbb{E}_L \log X)$$

and is called the geometric expectation. Additionally, based on limiting behavior, $\mathbb{E}^{-\infty}$ is defined to be the essential infimum and $\mathbb{E}^\infty$ is defined to be the essential supremum. The $p = -\infty$ and $p = \infty$ case don't exactly fit the definition of ψ-expectations, but they do have many of the properties associated with ψ-expectations.

1.16 True or false: for any non-negative random variable X, $a \in \mathbb{R}$, and $p \in \bar{\mathbb{R}}$, the identity $\mathbb{E}^p(aX) = a\mathbb{E}^p X$ holds.

Theorem 2.3 (Power Expectation inequality)**.** *Let X be a non-negative random variable. Then for any $-\infty \leq p \leq q \leq \infty$,*

$$\mathbb{E}^p X \leq \mathbb{E}^q X.$$

In particular, the Power Expectation inequality implies that $\|X\|_{L^p_{\mathbb{X}}(\mathbb{P})} \leq \|X\|_{L^q_{\mathbb{X}}(\mathbb{P})}$ for $p \leq q$. Thus $L^p_{\mathbb{X}}(\mathbb{P})$-spaces have a "reversed" order of inclusion compared to l^p spaces.

1.17 Prove the Power Expectation inequality.

Three special cases of power expectations are called Pythagorean expectations: $p = 1$ (the [arithmetic] expectation), $p = 0$ (the geometric expectation), and $p = -1$ (the harmonic expectation). The more familiar terms Pythagorean/arithmetic/geometric/harmonic *mean* can be interpreted as the specific case of a mapping from $\{1, \dots, n\}$ to $\mathbb{R}$ or $\mathbb{R}^+$ with $1/n$ times counting measure on the domain.

EXERCISE using geometric expectation - reveal its role in rate problems - move the finance example here and include a lengthy discussion in the solution.

EXERCISE using harmonic expectation - reveal its role in rate problems, e.g. physics, finance.

EXERCISE introducing LogSumExp

1.18 $p, q \in [1, \infty]$ are called conjugate exponents if $1/p + 1/q = 1$, or in other words, if the harmonic mean of p and q is 2. Prove Young's inequality: for conjugate exponents p and q, and any $a, b \in \mathbb{R}^+$,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$