Practical push-forward measure

The push-forward measure is one of those things which sounds horrible and complex when done in the abstract, but is pretty simple, intuitive, and really useful when actually used on things.

> **Problem 0.1 (Flipping coins with uniform r.v.).** *Suppose you want to flip a fair coin, but you don't have a coin – you only have access to $X \sim \mathrm{unif}[0,1]$. How would you do it? What if you wanted the coin to not be fair but instead have probability p to land heads?*

We can do this by using push-forward probability measures. The random variable we *have* operates on $([0,1], \mathcal{B}, \lambda)$ (so $\Omega_1 = [0,1]$); the random variable we *want* operates on $\Omega_2 = \{\mathsf{H}, \mathsf{T}\}$.

We now use a measurable function $f : \Omega_1 \to \Omega_2$ which you probably already figured out:

$$f(\omega) = \begin{cases} \mathsf{H} & \text{if } \omega \leq 1/2 \\ \mathsf{T} & \text{otherwise} \end{cases}$$

That this is measurable is obvious. We then note that $\mathbb{P}[\mathsf{H}] = \lambda(f^{-1}(\mathsf{H})) = 1/2$ and we're done.

To do this with a p probability of heads, just replace $1/2$ with p in the above.

> **Problem 0.2.** *What if you want $X \sim \mathrm{unif}[0,1]$, but you only have a fair coin? (You can flip it more than once and, as MIT students all do, you have infinite time).*

Now we have the probability space $(\{0,1\}^\infty, \mathcal{F}, \mathbb{P})$ of infinite coin-flips (we're using 0 and 1 now because it's easier to write the push-forward function) and we want to move it to $[0,1]$. Note that an infinite bit-string looks suspiciously like a real number written in binary. Thus, the proper function f to push this forward into $[0,1]$ is

$$f(\omega) = \sum_{i=1}^{\infty} \omega_i 2^{-i}$$

I'm not going to prove this works with all the rigorous bells and whistles, but the way to do it is to show that for any arbitrary interval $I = [a,b]$, the probability of getting an $X \in [a,b]$ is $b - a$ (which then shows that the probability matches on *every* Borel-measurable subset) – and the way to do *that* is to first consider only a, b of the form $m/2^n$ (so that we can decide the inclusion with only n bits, except for silly edge cases involving having all 0's after the nth bit) and then show that any $[a,b]$ can be approximated in this way.

Some MCT Example Problems

This is taken from last year's Week 7 recitation.

Problem 0.3. *Let $X_1, X_2, \ldots, X, Y$ be random variables on the same probability space. Then, we want to know:*

- *Suppose $0 \geq X_n \searrow X$. Does it necessarily hold that $\mathbb{E}[X_n] \to \mathbb{E}[X]$?*
- *Suppose $Y \leq X_n \nearrow X$, and $\mathbb{E}[|Y|] < \infty$. Does it necessarily hold that $\mathbb{E}[X_n] \to \mathbb{E}[X]$?*
- *Suppose $0 \leq X_n \searrow X$. Does it necessarily hold that $\mathbb{E}[X_n] \to \mathbb{E}[X]$?*
- *Suppose X_n is a continuous r.v. on $\mathbb{R}$ for all n, and has density function f_{X_n}; and suppose that $f_{X_n} \to f$ (pointwise). Then, is it necessarily true that:*

$$1.\ \text{Is} \int_{\mathbb{R}} f d\lambda \leq 1? \qquad 2.\ \text{Is} \int_{\mathbb{R}} f d\lambda = 1?$$

(Note that the last one is a *little* unfair because we haven't really discussed *probability density functions* yet - last year, pdf's were introduced before abstract integration - but probably many people already have some familiarity with them. If not, don't worry about that one.).

Solution: Our main tools here are the MCT and linearity of expectations.

- Yes it does. We note that $0 \leq -X_n \nearrow -X$. Then the MCT implies $\mathbb{E}[-X_n] \to \mathbb{E}[-X]$; and by linearity of expectation we can pull out the "$-$" and conclude $\mathbb{E}[X_n] \to \mathbb{E}[X]$.
- Yes it does. We write $Z_n := X_n - Y$ and $Z = X - Y$. Then $0 \leq Z_n \nearrow Z$; furthermore, because the integral of Y is absolutely convergent, $\mathbb{E}[Z_n]$ and $\mathbb{E}[Z]$ remain well-defined (exercise for those who want: construct a counterexample when this condition is removed). Therefore, by the MCT, $\mathbb{E}[Z_n] \to \mathbb{E}[Z]$.
- No it doesn't. We can construct a counterexample: suppose we have the probability space $\big([0,1], \mathcal{B}, \lambda\big)$ (Lebesgue) and $X_n = \frac{1}{n}\big(\frac{1}{\omega}\big)$. Then note that $X_n \searrow 0$ at all ω; but $\mathbb{E}[X_n] = \infty$ for all n (see note at bottom for why) but $\mathbb{E}[X] = 0$.
- (1) Yes. Here we use *Fatou's Lemma.* First, since f_{X_n} is a density function, $\int_{\mathbb{R}} f_{X_n} d\lambda = 1$. Second, by definition if $\lim_{n\to\infty} f_{X_n} = f$, then $\liminf_{n\to\infty} f_{X_n} = f$. But then by Fatou,

  $$1 = \liminf_{n\to\infty} \int_{\mathbb{R}} f_{X_n} d\lambda \geq \int_{\mathbb{R}} \big(\liminf_{n\to\infty} f_{X_n}\big) d\lambda = \int_{\mathbb{R}} f d\lambda$$

  (2) No. Consider $X_n \sim \text{unif}[n, n+1]$ (so $f_{X_n} = \mathbf{1}_{[n,n+1]}$. Then $f_{X_n}(x) \to 0$ at all x, so f is just 0. In which case, of course, $\int_{\mathbb{R}} f d\lambda = 0 \neq 1$.

Note: Why is $\int_0^1 t^{-1} dt = \infty$? The integral of t^{-1} is $\log(t)$; and $\log(t)$ is unbounded below as $t \to 0_+$ – so the definite integral ends up being $\log(1) - \log(0) = \infty$.

Fubini Failures: When does swapping order of summation change the result?

Here's an interesting thing to think about: suppose we have some ping-pong balls labeled $1, 2, \ldots,$ and a (really big) bucket. We make the following procedure: at each iteration n, we put in balls $10n - 9$ through $10n$ (ten balls at a time, starting with 1 through 10) and then remove ball n. Taking this to the limit - infinite iterations - how many balls are left in the bucket?

This is actually a well-known paradox. There are two perspectives:

- *Iterations view:* At each iteration, we put in 10 balls and remove 1, so we added 9 balls in total. So the number of balls grows toward ∞ and we have ∞ balls at the end.
- *Balls view:* Each ball m was put in at step $\lceil m/10 \rceil$ and then removed later at step m. So at the end, 0 balls can in the bucket.

This cute little story shows how Fubini's Theorem would break down if you failed to meet the absolute convergence condition. To make it more "mathematical", we can explicitly construct a double-sum whose value changes when the order of summation changes. The above helps to see how to do this - we'll make rows refer to iterations and columns to balls. The entry $f(m, n)$ will refer to what happened to ball n at iteration m: 1 if it was put in, -1 if it was taken out, and 0 (the vast majority of the time) when nothing happens. Formally,

$$f(m, n) = \begin{cases} 1 & \text{if } n = \lceil m/10 \rceil \\ -1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}$$

(technically there's one exception case for ball 1 which is put in and taken out at the same iteration, so we just use $f(1, 1) = 0$ because that's really what happened).

So then, summing by rows: each row has ten 1's in it and one -1 (except the first row which just has nine 1's) and so each sums to 9; so summing over all of them gives

$$\sum_m \sum_n f(m, n) = \infty$$

But summing by columns: each column has one 1 and one -1 (except the first row which is all 0's) and so each sums to 0; so summing over all of them gives

$$\sum_n \sum_m f(m, n) = 0$$

Hence the "paradox".

Why doesn't Fubini prevent this from happening? Because f is not absolutely convergent, nor is it nonnegative – it contains infinite 1's and infinite -1's.

Remark: A very similar example is in the lecture notes.

A weird integration counterexample

In general, Lebesgue integration is more general than Riemann integration – for example, $\int_{[0,1]} \mathbf{1}_{\mathbb{Q}}(x)dx$ is well-defined as a Lebesgue integral, but not as a Riemann integral. However, when dealing with *improper* Riemann integrals - which have an infinite domain or are unbounded around certain points - sometimes we get a function for which the Riemann integral converges but the Lebesgue integral is undefined.

Note: For this class, it is not required for you to understand this section; but it is something I thought we should mention.

Problem 0.4. *Let $f(x) = \frac{\sin(x)}{x}$ (where $f(0) = 1$, which keeps it continuous). Consider:*

$$\int_0^\infty f(x)dx := \lim_{t\to\infty} \int_0^t f(x)dx$$

We want to show that this is Riemann-integrable but that the integral $\int_{\mathbb{R}_+} f(x)dx$ is not *Lebesgue-integrable.*

We're not going to formally prove this, but we'll give a sketch. We need the following:

Fact 0.1 (Facts about the Harmonic Series)**.**

- *The* Harmonic Series $h_n := \sum_{i=1}^n \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \dots$ *diverges, i.e.* $\lim_{n\to\infty} h_n = \infty$.
- *The* Alternating Harmonic Series $a_n := \sum_{i=1}^n \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \dots$ *converges to a finite result (which happens to be* $\log(2)$ *but honestly it doesn't matter for this).*

Note the the first fact means that the Alternating Harmonic Series is *not* absolutely convergent.

Note that $f(x) = \frac{\sin(x)}{x}$ oscillates about 0, crossing at $k\pi$ for all $k > 0$. Let s_k be the area between $f(x)$ and 0 over the interval $[(k-1)\pi, k\pi]$. Note that $\sin(x)/x \leq 1$ and so $s_1 \leq \pi$.

As $k \to \infty$, these oscillations get closer and closer to being scaled-down versions of $\sin(x)$; and therefore there is a constant c such that $s_k \approx c/k$ for large values of k (there is a rigorous def'n of this but we're not going to worry about it, this is for intuition only).

Then, because the Riemann integral is a limit as $t \to \infty$, it is explicitly being summed in order: $s_1 - s_2 + s_3 - \dots$ which converges to a finite value; on the other hand, the Lebesgue integral splits off the prositive from the negative parts and therefore has the order

$$(s_1 + s_3 + s_5 + \dots) - (s_2 + s_4 + s_6 + \dots) = \infty - \infty, \text{ which results in an undefined integral}$$

(note that the two sums are both more than half the Harmonic Series and therefore diverge).

EXTRA: The probabilistic method for max-cuts

The basic idea of the *probabilistic method* is as follows: Suppose we want to show that some structure exists, or that some structure of a particular size exists – say, an independent set of size k in some graph. But building the thing and showing the build works is hard. So instead we build it randomly and show that the expected size is $\geq k$; then there must be some outcome which actually achieves that size (or greater), thus showing existence.

Remark: Note that this doesn't show, at all, how to build it. However, there is a way to (sometimes) convert this kind of proof into a (deterministic) algorithm for actually constructing whatever it is. This is called the *method of conditional expectations*, and interested students are encouraged to look it up. I promise it's really cool.

Problem 0.5. *Let $G = (V, E)$ be an undirected graph where $E \neq \emptyset$ (at least one edge exists). Show that V can be partitioned into $(S, V - S)$ in such a way so that* strictly more *than half the edges are between S and $V - S$ (as opposed to being internal in S or internal in $V - S$).*

Solution: As with all probabilistic method proofs, we put vertices into S at random – specifically, $v \in S$ with probability 1/2, independent of all other vertices. Then every edge (u, v) has a 1/2 probability of being across S to $V - S$, because whichever set u ends up in, v has a 1/2 chance of being in the other one. Denote the set of edges crossing $S, V - S$ to be $E(S, V - S)$, and denote S^* to be the subset maximizing $|E(S^*, V - S^*)|$

Therefore, if $|E| = m > 0$, then $\mathbb{E}\big[|E(S, V - S)|\big] = m/2$, and so

$$E(S^*, V - S^*) \geq \mathbb{E}\big[|E(S, V - S)|\big] = m/2$$

But this isn't exactly what we wanted - we wanted a *strict* inequality. Luckily, we have the following fact (easy to verify):

Fact 0.2. *If X is a random variable, then* $\max_\omega X(\omega) = \mathbb{E}[X]$ only if $X = \mathbb{E}[X]$ almost everywhere. *Furthermore, if we're in a* discrete probability space *(like the graph-cutting example here) and every ω has a positive probability, then*

$$\max_\omega X(\omega) = \mathbb{E}[X] \text{ only if } X \text{ is constant.}$$

Specifically, if we find some ω for which $X(\omega) < \mathbb{E}[X]$ then we've shown that $\max_\omega X(\omega) > \mathbb{E}[X]$. Here we use the event that $S = \emptyset$, for which

$$|E(\emptyset, V)| = 0 < m/2 = \mathbb{E}[|E(S, V - S)|]$$

This completes the proof.

Exercise in precision: If we remove the assumption that $E \neq \emptyset$, is the theorem still true? What part of the proof breaks down?