In this document we discuss the meaning of conditioning on certain null events. We will have a fixed probability space, and a Polish space (Call it $S$) valued RV $X$, $x \in S$ and an event $A$. We will say that $X, x$ (and $S$) satisfy the nontrivial balls hypothesis (NBH) if $P(X \in B_r(x)) > 0$ for all $r > 0$. Notice that the presence or absence of this condition does not depend on the choice of metrization of $S$, so long as it is done in a way that is consistent with the topology. (The metrization chosen need not even enforce the completeness condition, and yet the meaning of this hypothesis remains the same.)

If $P(X = x) = 0$ then we cannot condition on the event $X = x$ in the normal way. However, sometimes it is still possible to assign meaning to the expression $P(A | X = x)$. Notice that one could at first be fooled into trying to proceed by recognizing $P(A | X = x)$ as a composition of some $h : S \to R$ with $X$, and then to try to compute $h(x)$ which sort of matches up with the intuitive idea of what $P(A | X = x)$ should mean. However, this is not correct in general because $P(A | X)$ manifests ambiguities on sets of measure 0, and thus the same is true of $h$ for sets of measure 0 via the pushforward of $P$ under $X$.

**Definition 1.** Call $A$ admissible for the data $X, x$ if NBH holds true and $\lim_{\epsilon \to 0} P(A | X\in B_\epsilon(x))$ exists. In case $A$ is admissible, define the singular conditional probability of $A$ conditioned on the event $\{X = x\}$, which we denote by $P(A | X = x)$, to be this unique limit.

We seek a general existence theorem here, but we cannot obtain it even if $X$ has an $L^1$ density, the problem being that the Lebesgue differentiation theorem only works a.e., and even were the limits to exist, they might be 0 at some point, even if we hypothesize that the density is nonzero, since the value at a point is meaningless.

On the other hand, we have the following theorem.

**Theorem 1.** Suppose that $S = \mathbb{R}^d$. If the pushforward of $P$ via $X$ is absolutely continuous wrt Lebesgue measure with a continuous density $f$ on $\mathbb{R}^d$ such that $f(x) > 0$, and $h(X)$ is a version of $P(A | X)$ with $h : \mathbb{R}^d \to R$ continuous, then $A$ is admissible (wrt $X, x$). Moreover, $P(A | X = x) = h(x)$.

If $Y$ is another $\mathbb{R}^n$ valued RV with $n$ and $d$ not necessarily equal and $Y$ has an $L^1$ density, then for all $\sigma$ algebras $\mathcal{G} \subset \mathcal{F}$ we hope to have the RCD of $Y$ wrt $\mathcal{G}$ is also AC wrt Lebesgue, but of course this fails, with for instance $\mathcal{G} = \sigma(Y)$.

There seems to be no reason to believe that the collection of admissible $A$ needs to form a $\sigma$ algebra, which means that one cannot really talk about $P(. | X = x)$ as if it were a measure, and thus we cannot speak of things like “the distribution of $Z$ (valued in any measurable space) conditioned on $X = x$.” Of course, we could speak of the distribution of $Z$ conditioned on $X$ when $Z$ takes values in a standard Borel space, which is just a layman’s codeword for the RCD of $Z$ wrt $\sigma(X)$ but we could not then go and evaluate at $x$ (which would mean to factor through, find a random measure on $S$ that takes values of probability measures in the Polish space that is the codomain of $Z$ and then to evaluate at $x$. (it’s ambiguous again.))
However, in situations where an entire subsigma algebra of the \( \sigma \) algebra of our probability space is a subset of the collection of admissible events, and the restriction of \( P(\cdot | X = x) \) from the collection of admissible events to this sigma algebra is a measure, we will then do this restriction and refer to \( P(\cdot | X = x) \) as a measure. (Although here by measure, I obviously mean probability measure, actually if it is any set function not necessarily even \( \geq 0 \) that is completely additive, then it is automatically a probability measure.) In particular, if this subsigma algebra is generated by \( Y \) which takes values in any measurable space, then we can talk about the distribution of \( Y \) conditioned on \( X = x \). In this setting, we may not even have distribution of \( Y \) conditioned on \( X \) because the target space of \( Y \) may not be Polish. But usually we will need some niceness of the target space because often times these strong regularity hypotheses can only be checked when everything is nice and Euclidean-smooth, at least in finite dimensional cross-sections.

Even so, situations where one is trying to interpret \( P(\cdot | X = x) \) as an entire measure are usually quite futile, and even in the sequel, significant restrictions and cautions shall apply. The better definition for this is outlined in AIMlog. There it is also discussed why these two definitions do not coincide in any reasonable way where both could be used, so context must be used to decide which is meant. Usually, we will mean the RCD variation (in AIMlog) when the properties of a whole distribution are needed, and sometimes we will mean this even if we only care about a single event. For the benefit of the reader, I will reproduce (in greater generality, and with added topics and details) the notes in my personal log here on the second definition.

**Theorem 2.** Let \( S, S' \) be Polish, \( X, Y \) be RVs with values in \( S \) and \( S' \) respectively, defined on the same probability triple \( (\Omega, \mathcal{F}, P) \). Suppose that \( x \in S, X \in S \) satisfy NBH, and assume that there exists \( \mu \) a random probability measure defined on \( S \) with values in \( P(S') \) that is continuous, for which \( \mu \circ X = \nu \) where \( \nu \) is the RCD for \( Y \) wrt \( X \). Then there is exactly one member \( \Phi \) of \( P(S') \) that arises as the evaluation of any such \( \mu \) at the point \( x \in S \).

We call this probability measure \( \Phi \) the distribution of \( Y \) given \( X = x \).

Notice that we do not assert that there is a probability measure \( P(\cdot | X = x) \) on even just \( (\Omega, \sigma(Y)) \) for which the distribution of \( Y \) is equal to \( \Phi \). Notice that the NBH is important for establishing well-definedness of \( \Phi \), so we are not really straying that far away from classical conditioning, in that although the conditional event may be null, it is not null upon a fattening.

The object type of this definition and of the previous definition attempting to rigiorize the same intuition are different. However, this isn’t really the full extent of just how badly the two definitions would be seen to disagree to a reasonable mind. Consider \( X \) being \( U(0,1) \), and \( Y = X \), with \( x = 1/2 \). Then \( A := \{ Y > x \} \) is admissible wrt the given data, while yet \( P(A | X = x) = 1/2 \). We wanted the answer 0, which affirms the remark above that the first definition is not really that trustworthy. The second definition technically punts. One would be tempted to say that we have agreement if the second definition evaluates \( A \) the same way the first does, but the second definition has no idea how to
evaluate \( A \) because the object obtained is a measure on \((0,1)\). We do not have the ability to pullback measures, and so the ability to do so depends on the specific structure of our probability triple, presumably, but is in any case not an interesting question. If we suspend this disbelief and pretend that we could just evaluate \( \Phi \) on \((x, \infty)\) and consider this as a “second definition type representation of the correct intuitive idea,” then we would get the answer 0, since \( \Phi \) is the point mass at \( x \).

\( \Phi \) is the distribution of \( Y \) relative to \( P(\cdot | X = x) \) when \( P(X = x) > 0 \). Additionally, whenever we have \( S \), \( x \in S \), \( X \in S \), \( A \in F \) for which \( P(X = x) > 0 \) then \( A \) is admissible and \( P(A | X = x) \) is given by the usual undergraduate definition.

**Theorem 3.** If \( X, Y \) are independent, and valued in \( S, S' \) respectively which are Polish and \( x \in S \), \( X \) satisfy NBH, then \( S \times S' \) is Polish. Moreover, the RCD of \((X,Y)\) can be expressed as \( \delta_X \times P_Y \) where \( P_Y \) is the (constant) distribution of \( Y \) and \( \delta_X \) is the random measure that is the pointmass located at \( X \), which is a valid continuous representation. Hence \( \Phi = \delta_X \times P_Y \). We have \( \forall f \) bounded continuous on \( S \times S' \)

\[ E(f(X,Y) | X = x)^\prime := \int_{S \times S'} f(s,s')d\Phi(s,s') = E(f(x,Y)). \]

The scare quotes are there because remember, there is no measure that we know of to put on our probability triple that “behaves like \( P(\cdot | X = x) \).

As an example of the previous concepts, we will give two constructions of Brownian Bridge and verify that they are the same. In either case, we will check that the two definitions are equivalent, consider a finite dimensional \( (\mathbb{R}^2 \times \mathbb{R})^{2n} \) for \( 0 \leq t_1 < \cdots < t_n \leq 1 \), \( t_i \) dyadic rationals. To do this, we obtain \( \Phi \) for \( Y = (B_{t_1}, \ldots, B_t) \), \( X = B_1, x = 0, S = \mathbb{R}, S' = \mathbb{R}^n \). And then we impose these on the \( \mathbb{R}^Q \) in the obvious way. Kolmogorov consistency follows from consistency of the FDDS for BM, and then we use Kolmogorov’s continuity theorem to extend to \( C[0,1] \). Call this measure \( \nu \).

Call \( X_t \) the coordinates on \( C[0,1] \).

To check that the two definitions are equivalent, consider a finite dimensional distribution for coordinates \( 0 \leq t_1 < \cdots < t_n \leq 1 \), and \( f : \mathbb{R}^n \to \mathbb{R} \) bounded continuous. Let \( \Phi \) correspond to the choices above made for the FDD for coordinates \( t_1, \ldots, t_n, 1 \). Then \( \int_{C[0,1]} f(X_{t_1}, \ldots, X_{t_n})d\mu = \int_{\mathbb{R}^{n+1}} f(x_1, \ldots, x_n)d\Phi = \int_{\mathbb{R}^{n+1}} f(x_1 - x_{n+1} + x_{n+1}, \ldots, x_n - x_{n+1} + x_{n+1})d\Phi = E(f(B_{t_1} - B_1, \ldots, B_{t_n} - B_1)) = \int_{C[0,1]} f(X_{t_1}, \ldots, X_{t_n})d\nu \). This suffices to see that \( \mu = \nu \).

This shows that Brownian Bridge is clamped down at time 1, i.e. that \( X_t = 1\mu = \nu \) a.s. Furthermore, this shows that Brownian Bridge has the FDDS that are multivariate Gaussian. (I won’t call it a Gaussian process since it is only on \([0,1]\).