1. Measure Theory.

The question that is always asked in a course like this is how much measure theory do we need. Not much. Measure theory is mostly language. There are a few technical facts and we will deal with them as we need them. Right now we will talk about the language.

The (sample) space. This is fairly arbitrary. One visualizes a random phenomenon and every possible outcome is a distinct point of our space. This should be large enough to accommodate the complexity of the model we have, but small enough that we can work with it. Most of the time the choice is natural. If we want to model the results of a single toss of a coin the space Ω will have just two points in it. T and H. If the coin is tossed three times the space will be the eight words of length three with letters T, H. If we want to model an unending stream of tosses then Ω will be the space of all infinite sequences consisting of H or T.

Events. These are questions that can be asked about the outcome that can be answered by "yes" or "no". They correspond to subsets of Ω. The set corresponds to outcomes that result in the answer "yes". Complement is negation while union is "or" and intersection is "and". Not all subsets of Ω need to be events. Some are. They constitute a privileged collection $\mathcal{F}$. Satisfies some properties. Contains the whole space Ω and the empty set $\emptyset$ and is closed under complementation and finite or **countable** unions and intersections. Such collections are called σ-fields.

Measure or Probability. Probability P is a set-function with values $P(A)$ defined for sets $A \in \mathcal{F}$. A model is specified by assigning $P(A)$ for events A. It has some properties. $0 \leq P(A) \leq 1$. $P(\Omega) = 1$. $P(\emptyset) = 0$. Additive for (**countable**) union of disjoint sets.

All the technical aspects have to do with **countable additivity**. The hard part is the construction such P. There are theorems that do that. Typically they show that given a P on a class $\mathcal{A}$ of subsets of Ω with certain properties then it extends uniquely as a countably additive P to the σ-field generated by $\mathcal{A}$.

Example. Given a non-decreasing right continuous function $F(x)$ on the real line R such that $F(-\infty) = 0$ and $F(\infty) = 1$, we can define for an interval $A = (a, b]$, $P(A) = F(b) - F(a)$. This extends uniquely as a measure to the Borel σ-field, one generated by intervals. The function

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

yields the Lebesgue measure on $[0, 1]$.

Random Variable. This is just a function $\Omega \to R$. $X = X(\omega)$ associates a real value to every outcome $\omega \in \Omega$ in our sample space. One assumes that this is measurable, i.e for any interval $(a, b]$ the inverse image $X^{-1}[(a, b]] = \{\omega : X(\omega) \in (a, b]\}$ is in $\mathcal{F}$. This is so we can define

$$F(b) = P(X^{-1}[(-\infty, b]])$$

and get a measure on R which will be the distribution of X under our model P.

Expectations and Integrals. Given a random variable $f : \Omega \to R$ its expectation or integrals written as $\int f(\omega)dP$ is defined for $f(\omega) = \mathbf{1}_A(\omega)$ as $P(A)$ and then extended by linearity to simple functions and by a limiting procedure to all bounded measurable functions and then to absolutely integrable functions. If a function is badly unbounded the integral may not converge. But if it does, it does so absolutely. For non-negative measurable functions

$$\int f(\omega)\, dP = \sup_{\substack{g:0\le g\le f \\ g \text{ bounded}}} \int g(\omega)\, dP$$

And for general $f = f^+ - f^-$,

$$\int f dP = \int f^+\, dP - \int f^-\, dP$$

New measures from old. If f is non-negative integrable function with $\int f(\omega)dP = 1$ then

$$Q(A) = \int_A f(\omega)dP = \int \mathbf{1}_A(\omega)f(\omega)dP$$

defines a new probability measure. f is called the density or Radon Nikodym derivative of Q with respect to P. In symbols $f = \frac{dQ}{dP}$. Such Q's are characterized by the property $Q(A) = 0$ when ever $P(A) = 0$. A Q that satisfies the above condition is said to be **absolutely continuous** with respect to P. The Radon-Nikodym theorem proves that for such a Q, f exists and is unique.

It is useful to imagine the Radon-Nikodym derivative as

$$\lim_{A\downarrow\{\omega\}} \frac{Q(A)}{P(A)}$$

where $\{\omega\}$ is the single point set consisting of ω. In general $P(\{\omega\}) = 0$. Otherwise $f(\omega)$ would be just $\frac{Q(\{\omega\})}{P(\{\omega\})}$.

Information and σ-fields. One can think abstractly of "information" as all possible questions for which one can answer yes or no. Then one can see that natural candidates are σ-fields. If $\mathcal{F}$ represents complete information in the model, "partial information" corresponds to various sub-σ-fields. With the trivial σ-field $\{\Omega, \emptyset\}$ representing no information. If the information is just answer to one question, that corresponds to the $\sigma-$field $\{\emptyset, A, A^c, \Omega\}$. Corresponds to a **partition** of Ω into A, A^c. One can think more generally of partitioning Ω into disjoints sets $A_1, A_2, \ldots, A_n$ with $\cup_j A_j = \Omega$. This corresponds to the σ-field consisting of sets that are finite unions of some of the A_i's. Think of A_i as atoms that we cannot split. A general sub-σ-field is essentially a continuous version of this. Think of Ω as R^2 . If we only have information about the first coordinate then the partition is $\omega = \cup_x\{L_x\}$ where each atom is a line parallel to the y-axis.

Conditional Expectation. An important notion is that of conditional expectation. Expectation is the average and it is what we "expect". If we have partial information then our "expectation" will change. Imagine the information we have is given by a partition $\mathcal{P} = \{A_1, \cdots, A_n\}$ of Ω. Consider the random variable $X(\omega) = i$ on A_i. Then the conditional expectation of f given $\mathcal{P}$ is

$$g(\omega) = E[f(\omega)|\mathcal{P}] = g_i = E[f(\omega)|X(\omega) = i] = \frac{1}{P(A_i)} \int_{A_i} f(\omega)\, dP \text{ for } \omega \in A_i$$

Although $g(\omega)$ is a function it is really only different constants on different A_i. In other words it is a function measurable with respect to the partition $\mathcal{P}$. The values $\{g_i\}$ are averages adjusted for the information we have, i.e. conditional expectation. One way to obtain the conditional expectation is to define a new measure (assume $f \geq 0$ as we can always deal with $f^{\pm}$ separately)

$$Q(A) = \int_A f(\omega) dP$$

Q is absolutely continuous with respect to P, on $\mathcal{F}$ and therefore on $\mathcal{P}$ as well. The Radon-Nkodym derivatives $\frac{dQ}{dP}$ are different. On $\mathcal{F}$ it is f. But on $\mathcal{P}$ it is g. If we replace $\mathcal{P}$ by an arbitrary sub-σ-field $\mathcal{G}$ then the Radon-Nikodym derivative $\frac{dQ}{dP}$ on $\mathcal{G}$ is the conditional expectation of f given $\mathcal{G}$ and is $\mathcal{G}$ measurable. Its value only depends on the information contained in $\mathcal{G}$. Formally it is defined by two properties

$$\int_A f\, dP = \int_A g\, dP \text{ for all } A \in \mathcal{G} \; ; \quad g \text{ is } \mathcal{G} \text{ measurable}$$

g is denoted by $g = E[f \,|\, \mathcal{G}]$. It has some properties, if $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{F}$ and $g_1 = E[f \,|\, \mathcal{G}_1]$, then

$$E[f \,|\, \mathcal{G}_2] = E[g_1 \,|\, \mathcal{G}_2]$$

If h is $\mathcal{G}$ measurable then

$$E[hf \,|\, \mathcal{G}] = h\, E[f \,|\, \mathcal{G}]$$

Given $\mathcal{G}$ h is a constant and not "random".

Conditional Probability. Conditional probability is no different from conditional expectation. If we take f to be the indicator function of a set $A \in \mathcal{F}$, then $Q(\omega, A)$ is defined by

$$Q(\omega, A) = P(A|\mathcal{G}) = E[\mathbf{1}_A(\omega)|\mathcal{G}]$$

The defining relation becomes

$$\int_B Q(\omega, A) dP = P(A \cap B) \text{ for } B \in \mathcal{G}, A \in \mathcal{F}$$

with the condition that $Q(\omega, A)$ is $\mathcal{G}$ measurable for every $B \in \mathcal{F}$. More over one tries to choose $Q(\omega, A)$ so that for almost all ω it is a countably additive measure on $\mathcal{F}$. The last point is technical and needs some assumptions on the nature of $(\Omega, \mathcal{F})$.

Disintegration. Suppose $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ are two spaces and $f : X \to Y$ is a measurable map. Then if P is a probability measure on X it induces Q on Y by $Q = Pf^{-1}$, i.e.

$$Q(A) = P[x : f(x) \in A]$$

One can think of conditional probability as writing

$$P = \int \mu_y dQ(y)$$

where μ_y is supported on $X_y = \{x : f(x) = y\}$. The sub-σ-field of sets of the form $B = \{x : f(x) \in A\} = f^{-1}(A)$ as A varies over $\mathcal{G}$ is the natural σ-field with respect to which the conditioning is being done. The measure μ_y supported on X_y can be viewed as a measure on X and $Q(x, A) = \mu_{f(x)}(A) = \mu_{f(x)}(A \cap X_{f(x)})$ is the conditional probability.

Conversely given Q on $(Y, \mathcal{G})$ and μ_y on X_y one can define P on $(X, \mathcal{F})$ by integrating.