

# EE 5110: Probability Foundations

## RESPONSES TO REVIEWER COMMENTS

### Reviewer 1:

*Here is a list of my notes from reviewing the lectures. Many of these items are simply suggestions, a few point out minor errors. These errors can easily be addressed via an 'errata' accompanying the video lectures.*

*1. Lecture 5: Before pointing out when a measure is finite/infinite, it might have been worth highlighting the following: It follows easily from the definition of a measure that for any measurable set  $A$ ;  $\mu(A) \leq \mu(\Omega)$ . Only once this is stated is it clear that under a probability measure, each measurable set has measure 1:*

Response: I agree with the reviewer. The following remark can be inserted.

Action item: (31:00) From the axioms, it is clear that for any measurable set  $A$ ;  $\mu(A) \leq \mu(\Omega)$ . Specifically, the probability of any event is less than or equal to one.

*2. Lecture 7: While stating the requirements of a uniform measure on  $[0; 1]$  (just before the impossibility theorem), the following edits are needed:*  
*i)  $\mu(a; b) > 0$  (if the measures are not required to be positive, then the trivial measure assigning zero measure to all subsets can be constructed, contradicting the impossibility theorem)*  
*ii) For condition (ii), need to say  $x$  belongs to  $[0, 1]$*

Action item: (36:45) Before the impossibility theorem, we need  $\mu(a; b) > 0$  in condition (i) and  $0 \leq x \leq 1$  in condition (ii).

*3. Lecture 10, Minute 16: The line following the claim can be sharpened...*

Action item: (16:00) The claim can be stated more precisely: If we take  $(\Omega, \mathcal{F}_n)$  to be the measurable space, then not all elements of  $\mathcal{F}_m$ ;  $m \geq n$  can be measurable.

*4. Lecture 16: After stating the theorem regarding extension of a measure specified on a  $\pi$  system, it might be worth clarifying the distinction of this result with the Caratheodory extension theorem:*  
*In the latter result, a 'pseudo-measure' defined on an algebra can (under certain conditions) be extended to a construct a unique measure on the sigma-algebra generated by the algebra. In contrast, in the former*

*result, assuming the existence of a measure on the sigma-algebra generated by the the pi-system, it is uniquely specified by its values over the pi-system.*

Response: I thank the reviewer for pointing out this subtlety. The following remark can be added.

Action item: (3:14) I would like to point out a subtle difference between the results about the uniqueness of measures specified on a  $\pi$ -system, and the Caratheodory extension theorem. In the latter, a 'pseudo-measure' defined on an algebra can (under certain conditions) be extended to a construct a unique measure on the sigma-algebra generated by the algebra. In contrast, in the former result, assuming the existence of a measure on the sigma-algebra generated by the the  $\pi$ -system, it is uniquely specified by its values over the  $\pi$ -system.

*5. Lecture 20: When giving examples of sigma algebras generated by random variables, there is a subtle issue in the case of a constant random variable. The lecturer describes a constant random variable  $X$  as one that takes a certain value with probability 1. This itself does not imply that  $\sigma(X) = (\emptyset, \Omega)$ ; since the random variable may take other values over a zero-measure set, which will lead to additional sets being added to  $\sigma(X)$ : So one must for this example consider a random variable that \*always\* takes the same fixed value.*

Response: Both reviewers have pointed out this minor error, which can be corrected with the following remark:

Action item: (7:26) The statement '*If the random variable  $X$  is constant with probability 1, the sigma algebra generated by it is the trivial sigma algebra*' is not quite correct; it is correct if the random variable always maps to a constant value for *all* points in the sample space.

*6. Lecture 21: When talking about joint PMF of two discrete random variables, it seems imprecise to write the sum for any Borel set  $B$  in  $\mathbb{R}^2$ : Since  $B$  may in general be uncountable, the sum is, in general, not well posed.*

Both reviewers have pointed out this slip. This can be rectified by adding the following remark.

Action item: (38:32) The summation over all  $(x,y)$  belonging to the Borel set  $B$  is not well-defined as stated in the lecture. What I really mean is the following: Let  $E$  be a countable subset of  $\mathbb{R}^2$  over which the discrete random variables take

values with positive probability. Then, the sum should run over  $E \cap B$  (i.e., the intersection of the countable set  $E$  with the Borel set  $B$ ).

7. Lecture 27: When discussing the Jacobian formula in  $n$  dimensions, the lecturer says  $Y_i = g_i(X_i)$ ; but the formula is valid when  $Y_i = g_i(X)$ ; where  $X = (X_1; X_2; \dots; X_n)$ .

Action item: (16:23) The Jacobian formula is valid even when each  $Y_i = g_i(X_1, X_2, \dots, X_n)$  – this is more general than what is stated in the lecture.

8. Lecture 46: This is an editing issue: Around minute 26, the audio blanks out for some time.

Action item: Fix audio glitch if possible.

9. Lecture 49, Minute 17: The expression for the joint PDF of a bivariate Gaussian is missing a factor in the exponent.

Action item: (17:00) The expression for the joint PDF of a bivariate Gaussian is missing a factor  $1 - \rho^2$  in the denominator inside the exponent.

10. Lecture 49, Minute 30: In the counter-example showing that a vector of marginally Gaussian random variables is not necessarily jointly Gaussian, it might be worth mentioning that the vector constructed is actually jointly continuous. So a joint PDF exists here, it is just not a jointly Gaussian PDF.

Action item: (30:00) Note that in the counter-example showing that a vector of marginally Gaussian random variables is not necessarily jointly Gaussian, the vector constructed is actually jointly continuous. So a joint PDF exists here, it is just not a jointly Gaussian PDF.

## Reviewer 2:

----

Lec 1 (31:00) Would it be better to state: "This paradox arises because there are 3 different sample spaces together with different probability measures."?

Response: I agree with the reviewer

Action item: Please insert at 31:00 : Remark: This paradox essentially arises because there are 3 different sample spaces together with different probability measures."

Lec 2:

(02:15) "x is THE pre-image of  $f(x)$ " is technically incorrect

Response: I agree with the reviewer

Action item: At 2:15, "x is a pre-image," not "the pre-image."

*(33:30) I think students may be confused by the claim of a bijection between  $(Q \setminus [0,1])$  and  $N$ . It could be improved by explicitly writing down a bijection to convince the listener; should not be hard.*

This is left as an exercise.

Action item: none

*(36:24) "countable INDEX set" has not been formally defined. (in general) some audio level fluctuations were observed at portions of the lecture*

Action item: Fix audio problems, if possible

Lec 3:

*(in general) The phrase "being able to find a LISTING" as a proxy for "countable set" can be misleading. It would be better to just stick to the formal definition of "there exists a bijection to the set of natural numbers".*

The reviewer is right, but what I mean is clear from the context, and the formal definition has been mentioned earlier anyway.

Action item: None

*(32:18) The function  $\tan(\pi x - \pi/2)$  is not defined for  $x=0$  and  $x=1$ , so the formula is incorrect*

Action item: Please put a comment at 32:18 as follows:

"There's a minor slip here:  $\tan(\pi x - \pi/2)$  is actually a bijection from the open interval  $(0,1)$  to the reals. The uncountability of  $(0,1)$  then implies the uncountability of the reals."

Lec 5:

Lec 6:

Lec 7:

*(general comment for all lecture videos): the volume level of the start/end credits is much higher than that of the lecture speaker, hence a sudden and unpleasant explosion of sound when lecture ends*

Action item: balance the sound if possible

Lec 8:

*(18:00) the theorem and proof is well-stated. however, there seem to be a lot of clarification questions about the process of generating a sigma-algebra; perhaps a simple example on a finite sample space would help greatly in building the concept*

I think the questions are mostly of a clarifying nature. I would prefer not to particularize to a finite sample space to preserve conceptual generality, especially since we're building up towards defining the Borel sigma-algebra.

Action item: none

*Lec 9:*

*(20:30) Would be better to say "non-Lebesgue measurable sets" instead of "non-measurable sets", since it is otherwise not clear what the underlying sigma algebra is*

Action item: At 20:30, we mean non-Lebesgue measurable set

*(32:45) Caratheodory is supposed to be a Greek mathematician (according to Wikipedia).*

Action item: ERRATUM: Caratheodory is a Greek mathematician.

*Lec 10:*

*(12:45) perhaps make an oral comment on why the definition of  $P_0$  for a member of the algebra is well-defined (w.r.t decomposing the same set as a different union of half-open intervals)*

Action item: None (this is a very nitty-gritty point)

*Lec 11:*

*(12:50) an alternative way of writing  $A$  is  $A^{(n)} \times [0,1]^{\infty}$ , if it helps; also, a quick example of what is NOT in  $F_n$  may also be instructive*

Action item: (12:50) Note that an alternative way of writing  $A$  is  $A = A^{(n)} \times \{0,1\}^{\infty}$

*(15:25) "measure space" instead of "measurable space"?*

Action item: none (it is correct as it is)

*Lec 12:*

*(37:30) mention that the property inside the limit is proven by induction, since it is a property of the natural number  $n$*

Action item: Note that the property inside the limit is proven by Induction on  $n$ .

*Lec 14:*

*(10:15) It would be helpful to also include the following definition for  $\{A_n \text{ infinitely often}\}$ :  $\{\omega \in \Omega: \omega \text{ in infinitely many } A_n\}$ . I think this 'pointwise' view is in fact essential to grasp the concept of a tail event, and that the phrase 'will occur' can be misleading in the sense of introducing a false notion of 'time'.*

Action item: (10:15) Note that the event  $A_n$  occurs infinitely often is simply the set of all  $\omega$  contained in infinitely many of the  $A_n$ s.

*(41:40) Useful to just state that a sigma algebra with the sought property indeed exists (e.g., independent Bernoulli coin tosses), and leave the construction to homework.*

Action item: (41:40) It can be shown that the sigma-algebra and probability measure with the sought properties can be constructed – e.g., independent Bernoulli coin tosses with “progressively more biased” coins.

*Lec 15:*

*(28:00) It would be instructive to present a concrete example of a r.v. and the induced law on  $(R, B(R))$ .*

Examples of concrete random variables and the induced measures are given subsequently.

Action item: None

*Lec 16:*

*(16:40) The proof needs to be done with some more precision (unless it is stated that it is an outline). Existence of any limit must be shown before calculating it. It is incorrect to claim a priori that the limit is equal to  $\lim_{n \rightarrow \infty} P(X \leq x_n)$  for any chosen seq.  $x_n \rightarrow -\infty$ .*

Putting the following remark should suffice.

Action item: (16:40) The existence of the said limit follows from the monotonicity and boundedness of the CDF.

*Lec 17:*

*Lec 18:*

*(38:23) This is perhaps not the most appropriate module to explain the memorylessness of the exponential r.v.*

Action item: None

*Lec 19:*

*(07:45) The Gaussian pdf is better drawn as the familiar 'bell curve'*

*with a smooth portion at the top, and inflection points to the left and right sides of the curve. Here it looks more like a 'pointed hat'.*

Action item: None

(Could have drawn it better, but it looks reasonably like a bell-curve)

*(30:45) '... but it is not continuous either because all the measure is sitting on the Cantor set' -- this line is vague; perhaps you meant \*absolutely\* continuous?*

Action item: (30:45) In the sentence "it is not continuous either," we mean the measure is not absolutely continuous.

*Lec 20:*

*(07:26) 'If the random variable  $X$  is constant with probability 1, then the sigma algebra generated by it is the trivial sigma algebra' -- incorrect statement!*

Both reviewers have pointed out this minor error, which can be corrected with the following remark:

Action item: (7:26) The statement '*If the random variable  $X$  is constant with probability 1, the sigma algebra generated by it is the trivial sigma algebra*' is not quite correct; it is correct if the random variable always maps to a constant value for *all* points in the sample space.

*Lec 21:*

*(42:58) Formally, one must include the condition  $P_Y(Y = y) > 0$  in the beginning of the definition (since the definition only holds for  $y$  satisfying the condition).*

Action item: (42:58) ) Formally, one must include the condition  $P_Y(Y = y) > 0$  in the beginning of the definition (since the definition only holds for ' $y$ ' satisfying the condition).

*Lec 22:*

*(13:20) The summation  $\sum_{x \in B_1}$  for an arbitrary Borel set  $B_1$  does not make formal sense. Would be better to write, say,  $\sum_{x \in B_1 \cap S_X}$  where  $S_X$  is known to be a countable set, so that the summation makes sense. Or alternatively, use something like  $\sum_{x \in S_X} P[X = x] * \text{Indicator}[x \in B_1]$ .*

Both reviewers have pointed out a similar slip. This can be rectified by adding the following remark.

Action item: (13:20) The summation over all  $(x,y)$  belonging to the Borel set  $B$  is not well-defined as stated in the lecture. What I really mean is the following: Let  $E$  be a countable subset of  $\mathbb{R}^2$  over which the discrete random variables take

values with positive probability. Then, the sum should run over  $E \cap B$  (i.e., the intersection of the countable set  $E$  with the Borel set  $B$ ).

*Lec 23:*

*(11:50) You can perhaps make the statement "... except possibly on a set of Lebesgue measure zero" clearer by writing "..., i.e.,  $\text{Lebesgue}(\{(a,b) \in \mathbb{R}^2: \text{relation does not hold}\}) = 0$ ". (Students not exposed to measure theory may find the former statement somewhat vague and uncomfortable. Or at least mention that the integrals being talked about here are more general than the (undergraduate) Riemann integral, etc.*

Action item: (11:50) By the statement "... except possibly on a set of Lebesgue measure zero," we mean that the Lebesgue measure (on  $\mathbb{R}^2$ ) of the set on which the statement fails to hold, is zero.

*(24:20) You must explicitly assume that  $f_Y(y) > 0$  before you make the definition of  $F_{\{X|Y\}}(x|y)$ .*

Action item: (24:20) Note: we must explicitly assume that  $f_Y(y) > 0$  before we define the conditional CDF.

*(44:40) I think too much time has been spent discussing the single example of jointly continuous random variables. This could have been assigned as a detailed homework exercise.*

Action item: None

(There are not many worked out examples in these lectures – this in one of the few).

*Lec 24:*

*(21:30) The example for max of r.v.s is perhaps redundant and can be deferred to homework.*

Action item: None

(There are not many worked out examples in these lectures – this in one of the few).

*Lec 25:*

*Lec 26:*

*Lec 27:*

*(27:30) Perhaps spending too much time on computations involving the Jacobian formula is not very useful, since this may have been taught/worked out in previous (undergraduate-level) probability courses.*

Action item: None



(This was done for completeness).

Lec 28:

*(07:50) By "n going to infinity" here, you mean a particular sequence of successively finer partitions, and so generally speaking, the limits can depend on the partitioning scheme. The Riemann integral, however, is really defined as the (common value of the) supremum/infimum of the upper/lower sums over \*all\* finite partitions of the domain.*

Action item: (07:50) Some clarification on the statement "*n going to infinity*" is in order here: To be precise, the Riemann integral is defined as the (common value of the) supremum/infimum of the upper/lower sums over \*all\* finite partitions of the domain.

*(12:40) Suggestion: Before formally developing the abstract integral, it would be nice to draw a quick picture to compare the abstract and Riemann integrals -- the "layer cake" or "level set" depiction of the abstract integral (this is in fact presented much later, Lec 31, 45:50)*

This comes later, as the reviewer points out.

Action item: None

Lec 29:

Lec 30:

Lec 31:

Lec 32:

Lec 33:

Lec 34:

(28:10) Audio glitch

Lec 35:

Lec 36:

*(07:43) It would be nice to have a quick explanation of the statement "Remember how we constructed the approximation of a r.v.  $X$  from below, and monotonically increasing, by simple functions? We can show that that both the approximations of  $X$  and  $Y$  will continue to be independent."*

*(This may not be obvious to a student at first glance.)*

Action item: (07:43) The claim "*We can show that that both the approximations of  $X$  and  $Y$  will continue to be independent*" can be shown from the independence of  $X$  and  $Y$  – the details are left as homework.

Lec 37:

*(43:50) Suggestion: Having introduced fairly measure-theoretic concepts like integration, expectation and monotone convergence, one*

*could rather comfortably start out defining conditional expectation using this property itself (i.e.,  $E[Y|X]$  is a ' $\sigma(X)$ -measurable' r.v.  $Z$  such that  $E[Y \cdot 1_A] = E[Z \cdot 1_A]$  for all  $A$  in  $\sigma(X)$ ), and derive useful properties (iterated expectation, expectation under independent sigma algebras, etc.). This would eliminate the need to resort to discrete pmfs, etc. and build up (perhaps these could be done as examples a posteriori).*

*[NOTE: I notice that you make this comment in the next lecture, but the suggestion above still stands.]*

*(50:45) The Hilbert-space view of conditional expectation is probably too heavy to introduce in the same lecture as conditional expectation itself -- both are deep subjects to grasp!*

Response: As the reviewer points out, a fully rigorous treatment of conditional expectations and the Hilbert space interpretation is a rather advanced subject, and is beyond the scope of this course. However, I have given a brief vignette as to how these concepts are approached from the more elementary treatments.

Action item: None

Lec 38:

Lec 39:

*(23:50) Statement (ii) of the theorem should formally include the additional condition that  $M_x$  and  $M_y$  are finite in the region around 0.*

Action item: (23:50) In the statement (ii) of the theorem, we have to assume that the MGFs  $M_x$  and  $M_y$  are finite in an interval around the origin.

Lec 40:

Lec 41:

Lec 42:

Lec 43:

*(33:00) Of course, one must mention that all the r.v.s are on the same probability space to compare convergence in probability and in distribution.*

Action item: (33:00) Note that to compare convergence in probability with convergence in distribution, we have to assume that the RVs are in the same probability space.

*(42:13) Interesting counterexample to " $X_n \rightarrow X$  in probability but not almost surely". I wonder if there is a more explicit proof (without using Borel-Cantelli part 2) where you explicitly construct a sequence (with 1s recurring infinitely often in some pattern) and show that the probability of that single sequence is positive.*

Response: I am not sure if one can construct an explicit example of a particular sequence with positive probability. However, an elementary proof is possible,

which mirrors the proof of the Bore-Cantelli lemma, so I've preferred to just utilize the said lemma directly.

Action item: None

Lec 44:

*(12:00) If I am not mistaken, the proof seems incomplete and thus somewhat unconvincing. You also need to take epsilon to zero along a countable sequence to explicitly reach the definition of almost sure convergence.*

Response: The proof is indeed complete, since there are only finitely many excursions outside an epsilon band, with probability one, for any  $\epsilon > 0$ .

Action item: none

Lec 45:

*(39:41) Isn't this immediate by the fact that " $x$  maps to  $e^{(itx)}$ " is continuous and bounded?*

Response: Indeed. I have just derived it from fundamental results.

Action item: None

Lec 46:

*(15:40)  $p = 1$  or  $q = 1$  is also okay for Holder's inequality, and often important.*

Response: Thanks for pointing this out.

Action item: (15:40) In Hölder's inequality,  $p$  and  $q$  can be equal to one or infinity.

(26:16) Audio glitch

Action item: Fix the glitch if possible

Lec 47:

*(47:46) You mean "Define  $X_{\tilde{i}} = \min(B, X_i)$ "?*

Action item: (47:46) In the truncation argument, I mean  $X_{\tilde{i}} = \min(B, x_i)$ , not max.

Lec 48:

Lec 49: