Define a field and a $\sigma-$ field

Billingsley

Define a probability measure and give some of its properties

Billingsley

List some properties of probability measures

Billingsley

Definition. *A field on a set Ω is a collection of subsets $\mathcal{F}$ such that:*

1. *(at least contains two sets)* $\emptyset, \Omega \in \mathcal{F}$,
2. *(closer under complement) if* $A \in \mathcal{F}$, *then* $A^c \in \mathcal{F}$,
3. *(closure under finite union) if* $A, B \in \mathcal{F}$, *then* $A \cup B \in \mathcal{F}$

Definition. *A $\sigma -$ field $\mathcal{F}$ on the set Ω also has:*

1. *(closure under countable union) if* $A_1, \cdots \in \mathcal{F}$, *then* $\cup_i A_i \in \mathcal{F}$.

Definition. *A probability measure on a set Ω with field $\mathcal{F}$ is a function $P : \mathcal{F} \to [0, \infty)$ with:*

1. $0 \leq P(A) \leq 1, \forall A \in \mathcal{F}$,
2. $P(\emptyset) = 0$ *and* $P(\Omega) = 1$,
3. *if* $A_1, \ldots$ *are disjoint and* $\cup_i A_i \in \mathcal{F}$, *then*

$$P(\cup_i A_i) = \sum_i P(A_i)$$

Properties. *A probability measure P on Ω with field $\mathcal{F}$ has*

1. *(monotonicity), if* $A \subset B$, *then* $P(A) \leq P(B)$,
2. *(inclusion-exclusion)* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *and more generally*

$$P(\cup A_n) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$$

$$+ \cdots + (-1)^{n+1} P(A_1 \cap \ldots A_n),$$

3. *(countably subadditive) if* $A_1, \cdots \in \mathcal{F}$ *and* $\cup_i A_i \in \mathcal{F}$, *then* $P(\cup_i A_i) \leq \sum_i P(A_i)$,
4. *(continuous from below) if* $A_1 \subset A_2 \cdots \subset A$, *then* $P(A_n) \uparrow P(A)$
5. *(continuous from above) if* $A_1 \supset A_2 \cdots \supset A$, *then* $P(A_n) \downarrow P(A)$

Briefly describe the process of showing the Lebesgue measure exists

Billingsley

Define $\limsup_n A_n$ *and* $\liminf_n A_n$ *of sets and give an interpretation and an inequality*

Billingsley

Define measure-theoretic independence (events, collections of events, $\sigma-$ *field)*

Billingsley

Proof.

- The main work horse is the **Cartheodory extension theorem (theorem 3.1)**: a probability measure on a field can be uniquely extended to the generated $\sigma-$field if the measure is σ-finite.
- So to construct the Lebesgue measure: first we define the Lebesgue measure that assigns to half-open intervals the interval length.
- Second we verify that this is a well-defined measure on the Borel field, and then we apply theorem 3.1.
- Proving theorem 3.1 is involved. Also, studying $\sigma(\mathcal{B}_0)$ is necessary, where $\mathcal{B}_0$ is the field of finite unions and intersections of intervals.

$\square$

Definition. $\limsup_n A_n = \cup_n \cap_{k\geq n} A_k$. *If w is in LHS, then for every n, there exists some $k \geq n$ so that $w \in A_k$, hence w is in infinitely many of the A_n. "Infinitely often".*

Definition. $\liminf_n A_n = \cap_n \cup_{k\geq n} A_k$. *If w is in LHS, then there exists n such that for all $k \geq n$, $w \in A_k$ for all k. Hence, w is in all but finitely many A_n. "Eventually".*

Properties.

$$P(\liminf_n A_n) \leq \liminf_n P(A_n)$$
$$\leq \limsup_n P(A_n) \leq P(\limsup_n A_n)$$

Definition. *Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.*

Definition. *A collection of events $\{A_1, \dots, A_n\}$ are independent if*

$$P(A_{k_1} \cap \dots A_{k_j}) = P(A_{k_1}) \dots P(A_{k_j})$$

for all $2 \leq j \leq n$ and $1 \leq k_1 < \cdots < k_n \leq n$.

Definition. *A collection of classes $\mathcal{A}_1, \dots, \mathcal{A}_n$ in a $\sigma-$field $\mathcal{F}$ are independent if for each choice of $A_i \in \mathcal{A}_i$, the collection $\{A_n\}$ is independent.*

Definition. *Two $\sigma-$fields $\mathcal{A}$ and $\mathcal{B}$ are independent if for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we have $\mu(A \cap B) = \mu(A)\mu(B)$.*

State the Borel-Cantelli Lemma 1 and proof

BILLINGSLEY

State the Borel-Cantelli Lemma 2 and proof

BILLINGSLEY

What's a use case of the Borel-Cantelli lemmas (both)?

BILLINGSLEY

Theorem. *If* $\sum P(A_n) < \infty$, *then* $P(\limsup_n A_n) = 0$.

Proof. Observe that $\limsup_n A_n \subset \cup_{k\geq m} A_k$ for all m. This implies that

$$P(\limsup_n A_n) \leq P(\cup_{k\geq m} A_k) \leq \sum_{k\geq m} P(A_k).$$

Since this holds for arbitrary m and the right hand side sum goes to 0 if the infinite sum converges, the lemma follows. □

Theorem. *If* $\{A_n\}$ *are independent and* $\sum P(A_n) = \infty$ *then* $P(\limsup_n A_n) = 1$.

Proof. It is enough to show that $P(\cup_n \cap_{k\geq n} A_k^c) = 0$ for which it is enough to show that $P(\cap_{k\geq n} A_k^c) = 0$ for all k. Note that $1 - x \leq e^{-x}$, then (by independence)

$$P(\cap_{k=n}^{n+j} A_k^c) = \prod_{k=n}^{n+j} 1 - P(A_k) \leq \exp\{-\sum_{k=n}^{n+j} P(A_k)\}.$$

But since the sum diverges, as $j \to \infty$, the RHS goes to 0, hence

$$P(\cap_{k=n}^{\infty} A_k^c) = \lim_j P(\cap_{k=n}^{n+j} A_k^c) = 0$$

□

Use-cases:

- The first one can be used to show the strong law of large numbers. The process there is to show that the deviations of the sum have a decaying (summable) probability, hence the tail event is probability zero.
- The second one can be used like this. Consider a sequence of fair coin flips $X_1, X_2, \ldots$ and define the event $A_n = [X_n = 1]$. Then, $\Pr(A_n) = \frac{1}{2}$, are all independent, and hence $\Pr([X_n = 1\text{i.o.}]) = 1$.
- In general, it's a way to get information about the limit event when you have information about the individual events.

Define the tail $\sigma-$field and state the Kolmogorov 0-1 law

Billingsley

What's a use case of the 0-1 law?

Billingsley

What are simple random variables?

Billingsley

Definition. *Given a sequence of events* $A_1, A_2, \dots$ *in a probability space* $(\Omega, \mathcal{F}, P)$*, the tail* $\sigma-$*field associated with the sequence* $\{A_n\}$ *is*

$$\mathcal{T} = \cap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \dots).$$

The $\limsup_n A_n$ *and* $\liminf_n A_n$ *are events in the tail* $\sigma-$*field.*

Theorem. *If* $A_1, A_2, \dots$ *are independent, then for each event* A *in the tail* $\sigma-$*field,* $P(A)$ *is either 0 or 1.*

Use cases:

- Given a sequence of independent events A_n, we know that $P(\limsup_n A_n)$ is either 0 or 1. The Borel-Cantelli lemmas go further and give us a summability condition to decide whether the probability is 0 or 1.
- Any random variable measurable with respect to the tail $\sigma-$field is equal to a constant almost surely. Consider a random variable X on $\mathbb{R}$ that is measurable with respect to a tail $\sigma-$field $\mathcal{T}$. By Kolmogorov's 0-1 law, we have $P(A) \in \{0, 1\}$ for all $A \in \mathcal{T}$. Hence, since $A_n = [X(w) \in (n, n+1]]$ is measurable and is either 0 or 1. Since $\cup_n A_n$ is the whole measure space, whose measure is 1, we know that $P(A_n) = 1$ only for a single A_n. We can then continue subdividing intervals in a similar way and using Cantor's intersection theorem (that the intersection of a nested sequence of compact sets is non-empty) we know that $P(X(w) = c) = 1$ and hence almost surely.

Definition. *A random variables* X *on* $(\Omega, \mathcal{F})$ *is* ***simple*** *iff it can be written as*

$$X(w) = \sum_i x_i I_{A_i}$$

for some finite set of x_i *and* $A_i \in \mathcal{F}$*.*

Remark. *Simple rand om variables* X_n *converge to* X *with probability 1 (*$\lim_n X_n = X$*) iff* $\forall \epsilon > 0$,

$$P(|X_n - X| > \epsilon \text{ } i.o.) = 0$$

which, if the above holds, implies that

$$\lim_n P(|X_n - X| > \epsilon) = 0.$$

Note that

$$\{\lim_n X_n = X\}^c = \cup_\epsilon \{|X_n - X| \geq \epsilon \text{ } i.o.\} = \cup_\epsilon \cup_n \cap_{k \geq n} \{|X_n - X| \geq \epsilon\}.$$

State and prove the Markov's inequality

Billingsley

What's a use case of Markov's inequality?

Billingsley

State and prove Chebyshev's inequality

Billingsley

Theorem. *For a random variable X, nonnegative, then for positive α, we have*

$$P(X \geq \alpha) \leq \frac{1}{\alpha}\mathbb{E}[X].$$

Proof. Note that for any convex f and any set A, we have that

$$\min_{x \in A} f(x)\mathbf{1}_A \leq E[X\mathbf{1}_A] \leq E[X]$$

Hence, with $f(x) = x$ and $A = [\alpha, \infty)$, the result follows. If we use $f(x) = |x|^k$, then we have for positive α:

$$\Pr(|X| \geq \alpha) \leq \frac{1}{\alpha}^k \mathbb{E}[|X|^k]$$

□

Use cases:

- Well, say we flip n fair coins and count the number of heads. What's a bound on the probability of getting $.9n$ heads?

$$\Pr(X \geq 0.9n) \leq \frac{0.5n}{0.9n} = \frac{5}{9}$$

- The main limitation on Markov's inequality seems to be that it works on positive random variables. Hence, given X we can do:

$$\Pr(|X - \mu| \geq \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}$$

Theorem. *For a random variable X, we have*

$$\Pr(|X - m| \geq \alpha) \leq \frac{1}{\alpha}^2 \mathit{Var}(X)$$

Proof. Applying Markov's inequality with the absolute value function, exponent $k = 2$, and subtracting $m = \mathbb{E}[X]$, we obtain the desired result. □

What's a use case of Chebyshev's inequality?

Billingsley

State and prove Jensen's inequality (finite case).

Billingsley

What's a use case of Jensen's inequality?

Billingsley

When proving the strong law of large numbers by appealing to moments, we can bound the deviations of S_n by its variance times a factor in n.

Theorem. *Jensen's inequality says that for a convex function $\phi(x)$ and a random variable X, we have*

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$$

Proof. The proof follows by induction (base case follows form convexity; induction step follows from grouping n of the sum terms together).
More details here. □

Uses:

- Markov's inequality.
- Non-negativity of the Kullback-Leibler divergence and hence mutual information.
- The fact that $H(X) \leq \log |\mathcal{X}|$.
- The fact that $H(X|Y) \leq H(X)$.
- Independence bound on entropy (entropy of a collection of random variables is bounded above by the sum of their individual entropies).

State and prove Holder's inequality.

Billingsley

What's a use case of Holder's inequality?

Billingsley

State and prove the strong law of large numbers

Billingsley

Theorem. *Suppose that* $\frac{1}{p} + \frac{1}{q} = 1$ *for* $p, q > 1$. *Then:*

$$\mathbb{E}[|XY|] \le \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^q]^{\frac{1}{q}}$$

Proof. Young's. Here. □

We can show that $\|f\|_r \le \|f\|_p \mu(X)^s$ for $0 \le r \le p$ and $\frac{1}{s} + \frac{r}{p} = 1$. Using Holder's:

$$\|f\|_r^r = \int_X |f|^r d\mu \le \left(\int_X (|f|^r)^{p/r} d\mu\right)^{r/p} \left(\int_X 1 d\mu\right)^{1/s} \le \|f\|_p^r \mu(X)^{1/s}$$

Theorem. *If* X_n *are iid and* $\mathbb{E}[X_n] = m$, *then*

$$\Pr\Big(\lim_n n^{-1} S_n = m\Big) = 1$$

Proof. WLOG $m = 0$. It is enough to show that $\Pr\big(|n^{-1}S_n| \ge \epsilon \text{ i.o}\big) = 0$ for each ϵ.
Let $\mathbb{E}[X_i^2] = \sigma^2$ and $\mathbb{E}[X_i^4] = \xi^4$. By independence, we have

$$\mathbb{E}[S_n^4] = n\xi^4 + 3n(n-1)\sigma^4 \le Kn^2$$

where K does not depend on n. By Markov's inequality for $k = 4$,

$$\Pr(|S_n| \ge n\epsilon) \le Kn^{-2}\epsilon^{-4},$$

so the result follows by the first Borel-Cantelli lemma (the event probs are summable, hence the lim sup is 0). □

State and prove the weak law of large numbers

BILLINGSLEY

Demonstrate a use of the strong law

BILLINGSLEY

Demonstrate a use of the weak law

BILLINGSLEY

Theorem. *If X_n are iid and $\mathbb{E}[X_n] = m$, then for all ϵ*

$$\lim_n \Pr\big(|n^{-1}S_n - m| \geq \epsilon\big) = 0.$$

Proof. By appealing to the strong law, we have

$$\Pr\big(|n^{-1}S_n - m| \geq \epsilon\big) \leq \frac{\mathrm{Var}(S_n)}{n^2\epsilon^2} = \frac{n\mathrm{Var}(X_1)}{n^2\epsilon^2} \to 0.$$

□

Obviously, the probability that the sample mean is $\epsilon > 0$ away infinitely often is zero.

The probability that the sample mean is $\epsilon > 0$ away decays to zero with n (can still be infinitely often).

What is a measurable mapping? What are some properties?

Billingsley

What is the pushforward measure?

Billingsley

What is the change-of-variables formula?

Billingsley

Definition. *For two measure spaces* $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$*, a transformation* $T : \Omega \to \Omega'$ *is measurable* $\mathcal{F}/\mathcal{F}'$ *iff for all* $A \in \mathcal{F}'$*,* $T^{-1}(A) \in \mathcal{F}$*.*

Properties.

- *If* $T^{-1}(A) \in \mathcal{F}$ *for each* $A \in \mathcal{A}$*, where* $\mathcal{A}$ *generates* $\mathcal{F}'$*, then* T *is* $\mathcal{F}/\mathcal{F}'$ *measurable.*
- *A random vector is measurable iff each component function is measurable.*
- *Continuous functions are measurable. If* $f_k : \Omega \to \mathbb{R}$ *are measurable* $\mathcal{F}$*, then* $g(f_1(w), \ldots, f_k(w))$ *is measurable* $\mathcal{F}$ *if* $g : \mathbb{R}^k \to \mathbb{R}$ *is measurable.*
- *Composition of measurable functions is measurable. Sum, sup, lim sup, product are measure-preserving. A limit of measurable functions is measurable if the limit exists everywhere. We can construct a sequence of simple measurable functions that increase to any given measurable function.*

Definition. *Given* $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$ *and a measurable transformation* $T : \Omega \to \Omega'$ *and a measure* μ *on* $\mathcal{F}$*, then* $\mu T^{-1}(A') = \mu(T^{-1}(A'))$ *is a pushforward measure on* $\mathcal{F}'$*.*

Definition. *A measurable function* g *on* Ω' *is integrable with respect to the pushforward measure* $\mu T^{-1} = T(\mu)$ *iff the composition* $g \circ T$ *is integrable with respect to the measure* μ*. In that case,*

$$\int_{\Omega'} g d(\mu T^{-1}) = \int_{\Omega} g \circ T d\mu$$

Say a few things about distribution functions.

Billingsley

How is the integral defined for a non-negative measurable function? Properties?

Billingsley

State some properties of the general integral (addition, limits)

Billingsley

Properties.

- *A distribution function for a random variable X on $\mathbb{R}$ is $F(x) = \Pr(X \le x)$. It is non-decreasing, right-continuous (by continuity from above). By continuity from below, $\lim_{y\uparrow x} F(y) = F(x^-) = \Pr(X < x)$.*
- *For every non-decreasing, right-continuous function with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$, there exists on some probability space a random variable X for F is the distribution function.*
- *If $\lim_n F_n(x) = F(x)$ for all x, then we write $F_n \Rightarrow F$ and say that the distributions converge weakly and their corresponding random variables converge weakly.*

Definition. $\int_\Omega f d\mu = \sup \sum_i (\inf_{w\in A_i} f(w))\mu(A_i)$ *where the sup is taken over all partitions of* Ω.

Properties.

- *If $f \le g$ then $\int f \le \int g$.*
- *If $f_n \uparrow f$ then $\int f_n \uparrow \int f$.*
- *The integral is linear.*
- *If $f = 0$ a.e., then $\int f = 0$. If the measure of the set where f is non-zero is positive, then the integral is positive. If the integral exists, then $f < \infty$ a.e.*

Properties.

- *Monotonicity: (i) if $f \le g$ and are integrable, then $\int f d\mu \le \int g d\mu$,*
- *Linearity: if f, g are integrable, then $\int(\alpha f + \beta g)d\mu \le \alpha \int f d\mu + \beta \int g d\mu$.*
- *Monotone convergence: if $0 \le f_n \uparrow f$ a.e., then $\int f_n d\mu \uparrow \int f d\mu$,*
- *Fatou's lemma: for non-negative f_n, $\int \liminf f_n d\mu \le \liminf \int f_n d\mu$,*
- *Dominated convergence: if $|f_n| \le g$ a.e., where g is integrable, and if $f_n \to f$ a.e., then f and f_n are integrable and $\int f_n d\mu \to \int f d\mu$.*

Give an application of MCT.

Billingsley

Give an application of DOM.

Billingsley

Give an application of Fatou's lemma.

Billingsley

Example. *Consider the space* $\{1, 2, \dots\}$ *with the counting measure. If for all* m*, we have* $0 \le x_{n,m} \uparrow x_m$ *as* $n \to \infty$*, then* $\lim_n \sum_m x_{n,m} = \sum_m x_m$.

Example. *For an infinite sequence of measures* μ_n *on* $\mathcal{F}$, $\mu(A) = \sum_n \mu_n(A)$ *defines another measure (countably additive because sums can be reversed in a nonnegative double series). You can show that* $\int f d\mu = \sum_n \int f d\mu_n$ *holds for all nonnegative* f.

Consider the sequence $f_n = X 1_{A_n}$ where $A_n \downarrow A$ with $\mu(A) = 0$ and $A_1 = \Omega$. Assuming that f_1 is absolutely integrable, note that $|f_1| \ge |f_n|$ hence the sequence is dominated and therefore $\lim_n \int f_n d\mu = \int X 1_A d\mu = 0$.

Consider on $(\mathbb{R}, \mathcal{R}, \lambda)$ the functions $f_n = n^2 I_{(0,n^{-1})}$ and $f = 0$ satisfy $f_n \to 0$ for each x, but $\int f d\lambda = 0$ and $\int f_n d\lambda = n \to \infty$. Hence Fatou's lemma inequality can be strict. Note that DOM and MCT do not apply here, as f_n are unbounded.

What is the continuous mapping theorem?

BILLINGSLEY

When do the integral values determine the integrands?

BILLINGSLEY

What is a density?

BILLINGSLEY

Theorem. *A sequence of random variables X_n converging in distribution/probability/as to X implies that $g(X_n)$ converges to $g(X)$ if g is continuous.*

Remark. *If g is just a function on $\mathbb{R} \to \mathbb{R}$ and it's a composition, this is fine. However, if you want this to state something about the expectation, then $g(X) \coloneqq \mathbb{E}[X^k]$ (so g is now an operator), this is only continuous if the moments are bounded. However even in a space of bounded moment random variables (aka. $L^p(X)$), one can find a sequence of random variables whose integrals will blow up, so this g is not a bounded operator, hence not continuous.*

Remark. *If random variables converge in distribution, then do their moments converge? No: $Pr(X_n = n^2) = 1/n, Pr(X_n = 0) = 1 - 1/n$.*

Theorem. *If f and g are nonnegative and $\int_A f d\mu = \int_A g d\mu$ for all $A \in \mathcal{F}$ and if μ is $\sigma -$ field then $f = g$ a.e.*

Theorem. *If f and g are integrable and $\int_A f d\mu = \int_A g d\mu$ for all $A \in \mathcal{F}$ then $f = g$ a.e.*

Definition. *δ is a density if it is a nonnegative function so that for two measures ν and μ we have*

$$\nu(A) = \int_A \delta d\mu, \quad A \in \mathcal{F}$$

What is Scheffe's theorem?

Billingsley

What is uniform integrability?

Billingsley

What is an application of uniform integrablity?

Billingsley

Theorem. *Suppose* $\nu_n(A) = \int_A \delta_n d\mu$ *and* $\nu(A) = \int_A \delta d\mu$ *for densities* δ_n *and* δ*. If*

$$\nu_n(\Omega) = \nu(\Omega) < \infty, \quad n = 1, 2 \ldots,$$

and if $\delta_n \to \delta$ *except on a set of* μ*-measure 0, then*

$$\sup_{A \in \mathcal{F}} |\nu(A) - \nu_n(A)| \leq \int_\Omega |\delta - \delta_n| d\mu \to 0$$

Remark. *In other words, the total variation of two measures is bounded above by the* L^1 *difference of the densities, hence if the densities convergence, then the measures converge.*

Definition. *A sequence* f_n *is uniformly integrable if*

$$\lim_{\alpha \to \infty} \sup_n \int_{[|f_n| \geq \alpha]} |f_n| d\mu = 0$$

Theorem. *Suppose* $\mu(\Omega) < \infty$ *and* $f_n \to f$ *a.e.*
(i) if f_n *are uniformly integrable, then* f *is integrable and*

$$\int f_n d\mu \to \int f d\mu$$

(ii) if f *and* f_n *are nonnegative and integrable, then the conclusion of (i) implies that* f_n *are uniformly integrable.*

Theorem. *If* $\mu(\Omega) < \infty$*,* f *and* f_n *are integrable, and* $f_n \to f$ *a.e., then the following are equivalent:*

- f_n *are uniformly integrable,*
- $\int |f - f_n| d\mu \to 0$*,*
- $\int |f_n| d\mu \to \int |f| d\mu$*.*

What is Riemann integrability?

Billingsley

What is a product space? What is the product measure?

Billingsley

What is Fubini's theorem?

Billingsley

Definition. *A real function f on $(a, b]$ is Riemann integrable with integral r if: for all ϵ there exists a δ with*

$$|r - \sum_i f(x_i)\lambda(I_i)| < \epsilon$$

if $\{I_i\}$ is any finite partition of $(a, b]$ into subintervals satisfying $\lambda(I_i) < \delta$ and if $x_i \in I_i$ for each i.

Definition. *Given two measure spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$, the product space is $(X \times Y, \mathcal{X} \times \mathcal{Y})$ where $\mathcal{X} \times \mathcal{Y}$ is the set of sets $A \times B$ where $A \in \mathcal{X}$ and $B \in \mathcal{Y}$.*

Definition. *Given two measures μ, ν, the product measure is a measure π with $\pi(A \times B) = \mu(A)\nu(B)$. This measure is unique.*

Theorem. *If $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \hat{)}$ are $\sigma -$ field measure spaces, then*

$$\int_{X\times Y} f(x,y)\pi(d(x,y)) = \int_X \left[\int_Y f(x,y)\nu(dy)\right]\mu(dx)$$

$$\int_{X\times Y} f(x,y)\pi(d(x,y)) = \int_Y \left[\int_X f(x,y)\nu(dx)\right]\mu(dy)$$

What is convergence in probability?

Billingsley

What is the Glivenko-Cantelli theorem?

Billingsley

What is the definition of expected value? Same some things about it and limits.

Billingsley

Definition. *Random variables* X_n *converge in probability to* X*, written* $X_n \to_P X$ *if for each positive* ϵ

$$\lim_n P(|X_n - X| \geq \epsilon) = 0$$

Theorem. *A necessary and sufficient condition for convergence in probability is that each subsequence* X_{n_i} *contain a further subsequence* $X_{n_{k(i)}}$ *such that* $X_{n_{k(i)}} \to X$ *with probability 1 as* $i \to \infty$*.*

Remark. *In nonprobabilistic contexts, convergence in probability becomes convergence in measure.*

Theorem. *Suppose that* $X_1, X_2, \ldots$ *are independent and have a common distribution function* F*; put* $D_n(w) = \sup_x |F_n(x, w) - F(x)|$*, where*

$$F_n(x, w) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i(w))$$

Then $D_n \to 0$ *with probability 1.*

Definition.

- $\mathbb{E}[X] = \int X dP = \int_\Omega X(w) P(dw)$.
- $\mathbb{E}[X 1_A] = \int_A X dP$.
- $\mathbb{E}[g(X)] = \int_\Omega g(X(w)) P(dw)$.
- *Absolute moments:* $\mathbb{E}[|X|^k] = \int_{-\infty}^{\infty} |x|^k P(dx)$.

Remark. *By the way* $P(dx)$ *intuitively refers to the* P *measure of a small change in* x*.*

Remark. *If* X_n *are dominated by an integrable random variable (or uniformly integrable), then* $\mathbb{E}[X_n] \to \mathbb{E}[X]$ *follows if* $X_n \to_P X$*.*

What is the MGF? What is it for normal, exponential, Poisson, and Binomial dists?

Billingsley

Roughly sketch the proof of the strong law of large numbers from first principles.

Billingsley

The weak law of large numbers? State and prove it.

Billingsley

Definition. *The MGF is*

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{sx} \mu(dx)$$

Properties.

- *The standard normal distribution:* $M_X(t) = e^{t^2/2}$.
- *The exponential:* $M_X(t) = \frac{\alpha}{\alpha - s}$ *defined for* $s < \alpha$.
- *The Poisson:* $M_X(t) = e^{\lambda(e^t - 1)}$.
- *The Binomial:* $M_X(t) = pe^t + (1-p)$.

Proof.

- You only need to show it for nonnegative rvs, by $\mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X]$.
- Truncated random variables and two applications of Borel-Cantelli and some other things.

$\square$

Theorem. $n^{-1} S_n \to_P \mathbb{E}[X_1]$.

Proof. This follows in the same condition from the strong law, because a.s. convergence implies convergence in probability. $\square$

What is the connection between the MGF and the Laplace transform?

Billingsley

What is Kolmogorov's maximal inequality?

Billingsley

What is a condition for random series to converge?

Billingsley

Theorem. *Let μ and ν be probability measures on $[0,\infty)$. If*

$$\int_0^\infty e^{-sx}\mu(dx) = \int_0^\infty e^{-sx}\nu(dx), \quad s \geq s_0$$

where $s_0 \geq 0$, then $\mu = \nu$.

Remark. *A distribution concentrated on $[0,\infty)$ is determined by its MGF or its Laplace transform.*

Theorem. *Suppose that $X_1, X_2, \ldots$ are independent with mean 0 and finite variances. For $\alpha > 0$,*

$$P(\max_{1\leq k\leq n} |S_k| \geq \alpha) \leq \frac{1}{\alpha^2}\, Var(S_n)$$

Theorem. *Suppose that $\{X_n\}$ are independent and $\mathbb{E}[X_n] = 0$. If $\sum Var(X_n) < \infty$, then $\sum X_n$ converges with probability 1.*

Theorem. *For an independent sequence $\{X_n\}$, the S_n converge with probability if and only if they converge in probability.*

Theorem. *Suppose that $\{X_n\}$ are independent and consider the three series*

$$\sum P(|X_n| > c), \quad \sum \mathbb{E}[X_n^{(c)}], \quad \sum Var(X_n^{(c)}).$$

In order that $\sum X_n$ converge with probability 1, it is necessary that the three series converge for all positive c and sufficient that they converge for some positive c.

Define weak convergence

Billingsley

What's an example of weak convergence of random variables?

Billingsley

What's the relationship between convergence in probability and weak convergence?

Billingsley

Definition. *Distribution functions* F_n *converge weakly to* F *if*

$$\lim_n F_n(x) = F(x)$$

for every continuity point x *of* F. *We write* $F_n \Rightarrow F$.

Definition. *Measures* μ_n *converge weakly if*

$$\lim_n \mu_n((-\infty, x]) = \mu((-\infty, x])$$

for every x *for which* $\mu(\{x\}) = 0$.

Definition. *Random variables* X_n *converge weakly to* X *if their respective distribution functions converge weakly* $F_n \Rightarrow F$.

Remark. *Let* μ_n *be the binomial distribution for* $p = \lambda/n$ *and let* μ *be the Poisson distribution. For nonnegative integers* k,

$$\mu_n(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k (1 - \lambda/n)^n}{k!} \times \frac{1}{(1 - \lambda/n)^k} \prod_{i=0}^{k-1} (1 - \frac{i}{n})$$

$$\overset{n \to \infty}{\to} \frac{\lambda^k e^{-\lambda}}{k!} = \mu(k)$$

if $n \geq k$ *and* k *stays fixed as we take the limit.*

Remark.

Remark. *Convergence in probability is equivalent to convergence in distribution when limiting to constants.*

Give examples of convergence in distribution but not in probability and in probability but not almost surely

Billingsley

When does a parallel converging weak sequence converge to the same limit?

Billingsley

What are some equivalent conditions for weak convergence of measures?

Billingsley

Example. *Consider two independent Bernoulli random variables X and Y with $p = \frac{1}{2}$. If $X_n = Y$, then certainly $X_n \Rightarrow X$ because their distributions are the same. However,* $\Pr(|X_n - X|) = \frac{1}{2}$.

Example. *Consider the independent random variables* $X_n = \begin{cases} 1, & p = \frac{1}{n} \\ 0, & 1 - p = 1 - \frac{1}{n} \end{cases}$. *Clearly $X_n \to 0$ a.e. as functions and* $\Pr(|X_n| \geq \epsilon) = \frac{1}{n} \to 0$. *However,* $\sum_n \Pr(X_n = 1) = \infty$, *while the X_n are independent, hence by Borel-Cantelli 2, we have* $\Pr(\limsup_n X_n = 1) = 1$ *so X_n do not converge almost surely.*

Example. *An analysis version, aka the "typewriter sequence": partition the interval* $[0, 1]$ *into two sets and choose the indicator functions of those sets to be f_1, f_2. Then partition into three and set the indicators of those sets to be f_3, f_4, f_5. Repeat this and note that in measure these functions converge to the zero function, while not converging pointwise.*

Theorem. *If $X_n \Rightarrow X$ and $X_n - Y_n \Rightarrow 0$, then $Y_n \Rightarrow X$.*

Theorem. *The following are equivalent:*

- $\mu_n \Rightarrow \mu$,
- $\int f d\mu_n \to \int f d\mu$ *for every bounded, continuous real f,*
- $\mu_n(A) \to \mu(A)$ *for every μ-continuity set A*

What is a theorem on weak convergence and compactness?

Billingsley

What is tightness?

Billingsley

Does weak convergence imply convergence of means?

Billingsley

Theorem. *(Helly selection theorem): For every sequence* $\{F_n\}$ *of distribution functions there exists a subsequence* $\{F_{n_k}\}$ *and a non-decreasing, right-continuous function* F *such that* $\lim_k F_{n_k}(x) = F(x)$ *at continuity points* x *of* F.

Theorem. *Tightness is necessary and sufficient so that for every subsequence* $\{\mu_{n_k}\}$ *there exists a further subsequence* $\{\mu_{n_{k(i)}}\}$ *and a probabilistic measure* μ *such that* $\mu_{n_{k(i)}} \Rightarrow \mu$.

Remark. *The first theorem guarantees compactness of distribution functions, but in the larger space of functions – that is* F, *may not be a distribution. The second theorem gives the necessary and sufficient condition for that compactness be in the space of probability measures.*

Remark. *Tightness for sequences of unit mass probability measures* μ_n *at* x_n *corresponds to boundedness of the sequence. A sequence of normal distributions is tight iff the means and variances are bounded.*

Definition. *A sequence of measures is tight if* $\forall \epsilon > 0, \exists (a, b]$ *s.t.* $\forall n, \mu_n(a, b] > 1 - \epsilon$. *The equivalent condition on the distributions is* $\forall \epsilon > 0, \exists (a, b]$ *such that* $F_n(a) < \epsilon, F_n(b) > 1 - \epsilon$.

Theorem. *If* $X_n \Rightarrow X$ *and* X_n *are uniformly integrable, then* X *is integrable and* $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Theorem. *Let* r *be a positive integer. If* $X_n \Rightarrow X$ *and* $\sup_n \mathbb{E}[|X_n|^{r+\epsilon}] < \infty$, *where* $\epsilon > 0$, *then* $\mathbb{E}[|X|^r] < \infty$ *and* $\mathbb{E}[X^r] \to \mathbb{E}[X^r]$.

What is the characteristic function?

Billingsley

What are important properties of the characteristic function?

Billingsley

State the CLT and sketch the proof

Billingsley

Definition. *The characteristic function of a probability measure μ is defined for real t by*

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \mu(dx).$$

A random variable with distribution μ has characteristic function

$$\phi(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \mu(dx).$$

Properties.

- *It is the Fourier transform of the random variable's density. It is the MGF with t replaced by it. It always exists and is bounded.*
- *if μ_1 and μ_2 have characteristic functions $\phi_1(t)$ and $\phi_2(t)$ then $\mu_1 * \mu_2$ has characteristic function $\phi_1(t)\phi_2(t)$*
- *The characeristic function always determines the distribution.*
- *From the pointwise convergence of characteristic functions follows weak convergence of the corresponding distributions.*

Theorem. *Suppose that X_n are iid random variables with mean c and finite positive variance σ^2. If $S_n = X_1 + \cdots + X_n$, then*

$$\frac{S_n - nc}{\sigma\sqrt{n}} \Rightarrow N(0,1)$$

Proof. Consider the special case where X_n takes the values ± 1 with probability $\frac{1}{2}$ each. Then the characteristic function is $\phi(t) = \cos t$ and the characteristic function of $S_n/\sqrt{n}$ is $\phi(t/\sqrt{n})^n = \cos(t/\sqrt{n})^n$. If we show that this converges to $e^{-t^2/2}$, then by the continuity mapping theorem we get convergence in distribution. The rest requires analysis. $\square$

When is a measure determined by its moments?

BILLINGSLEY

What is an additive set function?

BILLINGSLEY

What is the Hahn decomposition theorem?

BILLINGSLEY

Theorem. *A measure μ is determined by its moments if it has finite moments α_k of all orders and the power series $\sum_k \alpha_k \frac{r^k}{k!}$ has a positive radius of convergence.*

Theorem. *If a random variable X has finite moments of all orders and a sequence of random variables X_n has $\lim_n \mathbb{E}[X_n^r] = \mathbb{E}[X^r]$ for all r, then $X_n \Rightarrow X$.*

Remark. *One can also show the CLT using the method of moments.*

Definition. *A functino ϕ on $\mathcal{F}$ is an additive set function if*

$$\phi(\cup A_i) = \sum \phi(A_i)$$

whenever $\{A_i\}$ are a countable sequence of disjoint sets.

Theorem. *For any additive set function, there exist two sets A^+ and A^- with $A^+ \cup A^- = \Omega$, and $\phi(E) \geq 0$ for all E in A^+, and $\phi(E) \leq 0$ for all E in A^-.*

What is the Radon-Nikodym theorem?

Billingsley

How is conditional probability defined measure theoretically?

Billingsley

What is conditional expectation?

Billingsley

Theorem. *If μ and ν are $\sigma -$field measures and $\nu \ll \mu$, then there exists an almost everywhere unique non-negative function f such that*

$$\nu(A) = \int_A f d\mu$$

for all $A \in \mathcal{F}$.

Remark. *This is the converse of the statement that if $\nu(A) = \int_A f d\mu$, then $\nu \ll \mu$.*

Theorem. *Let X be a random variable $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a $\sigma -$field in $\mathcal{F}$. Then there exists a function $\mu(H, w)$ defined for $H \in \mathcal{R}$ and $w \in \Omega$ with:*

- *for each w in Ω, $\mu(\cdot, w)$ is a probability measure on $\mathcal{R}$,*
- *for each H in $\mathcal{F}$, $\mu(H, \cdot)$ is a version of $P(X \in H|\mathcal{G})$*

Definition. *A version is a function $f = P(A|\mathcal{G})$ specifying a random variable such that $P(A|\mathcal{G})$ is measurable $\mathcal{G}$ and integrable and*

$$\int_G P(A|\mathcal{G}) dP = P(A \cap G), \quad G \in \mathcal{G}$$

Definition. *A conditional expectation of X given $\mathcal{G}$ is a random variable $\mathbb{E}[X|\mathcal{G}]$ satisfying (i) that $\mathbb{E}[X|\mathcal{G}]$ is measurable $\mathcal{G}$ and integrable, (ii) that*

$$\int_G E[X|\mathcal{G}] dP = \int_G X dP$$

What is the tower property of expectation?

Billingsley

Theorem. *If X is a random variable $\mathcal{G}_1$ and $\mathcal{G}_2$ are two $\sigma-$field with $\mathcal{G}_1 \subset \mathcal{G}_2$, then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1]$$

Remark. *The average is a smoothing operation, in which we loose information about X up to the average of X in $\mathcal{G}$. Hence, averaging over a fine and then a coarser $\sigma-$field leaves the coarser result.*