

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 $A \cup B \in \mathcal{F}$
- A σ field \mathcal{F} on the set Ω also has:
 - 4. (closure under countable union) if $A_1, \dots \in \mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$.

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

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

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

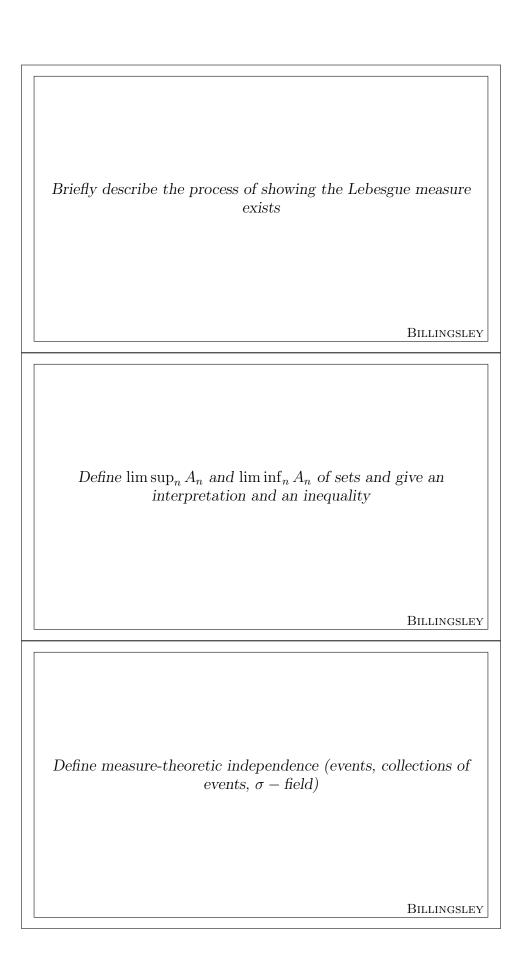
A probability measure P on Ω with field \mathcal{F} has

- 1. (monotonicity), if $A \subset B$, then P(A) < 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 $A_1, \dots \in \mathcal{F}$ and $\bigcup_i A_i \in \mathcal{F}$, then $P(\bigcup_i A_i) \leq \sum_i P(A_i)$,
- 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 $P(A_n) \downarrow P(A)$



Theorem 3.1 (Cartheodory extension theorem): a probability measure on a field can be uniquely extended to the generated σ – field if the measure is σ -finite.

Hence, 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.

 $\limsup_n A_n = \bigcup_n \cap_{k \geq n} A_k$. 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".

 $\liminf_n A_n = \cap_n \cup_{k \geq n} A_k$. 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".

$$P(\liminf_{n} A_n) \le \liminf_{n} P(A_n)$$

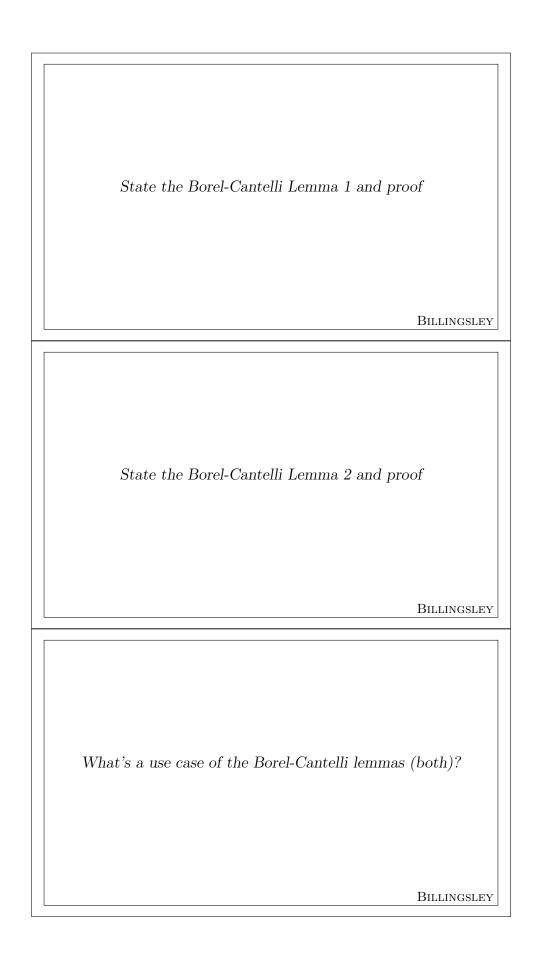
$$\le \lim \sup_{n} P(A_n) \le P(\limsup_{n} A_n)$$

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- A collection of events $\{A_1, \ldots, A_n\}$ are independent if

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

for all $2 \le j \le n$ and $1 \le k_1 < \dots < k_n \le n$.

- A collection of classes A_1, \ldots, A_n in a σ field \mathcal{F} are independent if for each choice of $A_i \in \mathcal{A}_i$, the collection $\{A_n\}$ is independent.
- Two σ fields \mathcal{A} and \mathcal{B} are independent if for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we have $\mu(A \cap B) = \mu(A)\mu(B)$.



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

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

$$P(\limsup_{n} A_n) \le P(\bigcup_{k \ge m} A_k) \le \sum_{k \ge m} P(A_k).$$

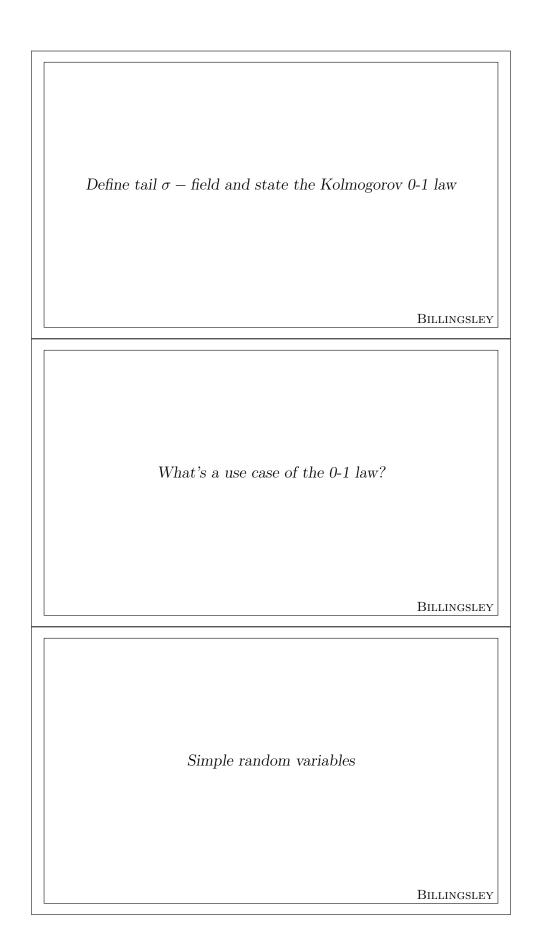
Since this holds for arbitrary m and the right hand side sum goes to 0 if the infinite sum converges, the lemma follows.

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

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

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

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



Given a sequence of events A_1, A_2, \ldots in a probability space (Ω, \mathcal{F}, P) , the tail σ – field associated with the sequence $\{A_n\}$ is

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

The $\limsup_n A_n$ and $\liminf_n A_n$ are events in the tail σ – field.

The Kolmogorov zero-one law: if A_1, A_2, \ldots are independent, then for each event A in the tail σ – field, P(A) is either 0 or 1.

A random variables X on (Ω, \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 $A_i \in \mathcal{F}$.

Simple random 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

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

State and prove the Markov's inequality What's a use case of Markov's inequality? Billingsley State and prove Chebyshev's inequality	_		
What's a use case of Markov's inequality? BILLINGSLEY		State and prove the Markov's inequality	
BILLINGSLEY			BILLINGSLEY
BILLINGSLEY	L		
		What's a use case of Markov's inequality?	,
			BILLINGSLEY
State and prove Chebyshev's inequality			DILLINGSLE I
Billingsley		State and prove Chebyshev's inequality	Billingsi ev

Markov's inequality: For a random variable X, nonnegative, then for positive α , we have

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

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

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

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 :

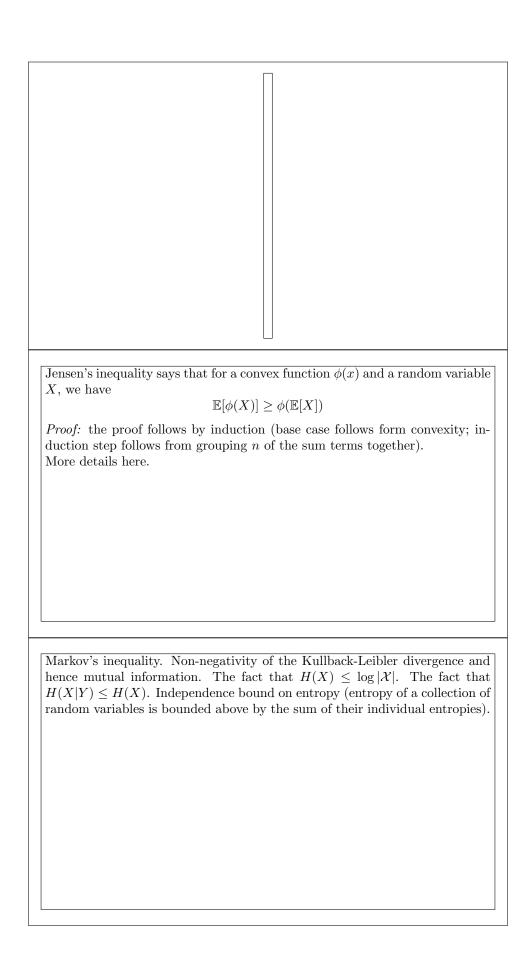
$$\Pr(|X| \ge \alpha) \le \frac{1}{\alpha}^k \mathbb{E}[|X|^k]$$

Chebyshev's inequality: for a random variable X, we have

$$\Pr(|X - m| \ge \alpha) \le \frac{1}{\alpha} \operatorname{Var}(X)$$

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

1	
	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?



State and prove Holder's inequality. Billingsley What's a use case of Holder's inequality? Billingsley
DILLINGSLEY
State and prove the strong law of large numbers. Billingsley

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.

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

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

Proof: WLOG m=0. It is enough to show that $\Pr(|n^{-1}S_n| \ge \epsilon \text{ i.o}) = 0$ 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).

$\overline{}$	
	State and prove the weak law of large numbers. Billingsley
	DILLINGSLEY
	Demonstrate a use of the strong law. Billingsley
	Demonstrate a use of the weak law. Billingsley

If X_n are iid and $\mathbb{E}[X_n] = m$, then for all ϵ $\lim_{n} \Pr(|n^{-1}S_n - m| \ge \epsilon) = 0.$ *Proof:* By appealing to the strong law, we have $\Pr(|n^{-1}S_n - m| \ge \epsilon) \le \frac{\operatorname{Var}(S_n)}{n^2 \epsilon^2} = \frac{n\operatorname{Var}(X_1)}{n^2 \epsilon^2} \to 0.$

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

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.

Given (Ω, \mathcal{F}) and (Ω', \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}' .

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,

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

A distribution function for a random variable X on \mathbb{R} is $F(x) = \Pr(X \leq x)$. 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 $F_n \implies F$ and say that the distributions converge weakly and their corresponding random variables converge weakly.

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

If $f \leq g$ then $\int f \leq \int g$.

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.

- Monotonicity: (i) if $f \leq g$ and are integrable, then $\int f d\mu \leq \int g d\mu$,
- Linearity: if f, g are integrable, then $\int (\alpha f + \beta g) d\mu \leq \alpha \int f d\mu + \beta \int g d\mu$.
- 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 , $\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$,
- Dominated convergence: if $|f_n| \leq g$ 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
	DITTINGSTE I

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 $\lim_n \sum_m x_{n,m} = \sum_m x_m$. 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 $f_n = X1_{A_n}$ where $A_n \downarrow A$ with $\mu(A) = 0$ and $A_1 = \Omega$. Assuming that f_1 is absolutely integrable, note that $|f_1| \geq |f_n|$ hence the sequence is dominated and therefore $\lim_n \int f_n d\mu = \int X1_A d\mu = 0$.

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?
When do the integral values determine the integrands? BILLINGSLEY
DIBBINGBET
What is a density?

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.

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 $g(X) := \mathbb{E}[X^k]$ (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.

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

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 σ – field then f = g a.e.

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.

 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?
	DILLINGSLET
	What is uniform integrability?
	BILLINGSLEY
L	
	What is an application of uniform integrablity?

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

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

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

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

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.

A sequence f_n is uniformly integrable if

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

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.

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?
What is Fubini's theorem?

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

$$|r - \sum_{i} f(x_i)\lambda(I_i)| < \epsilon$$

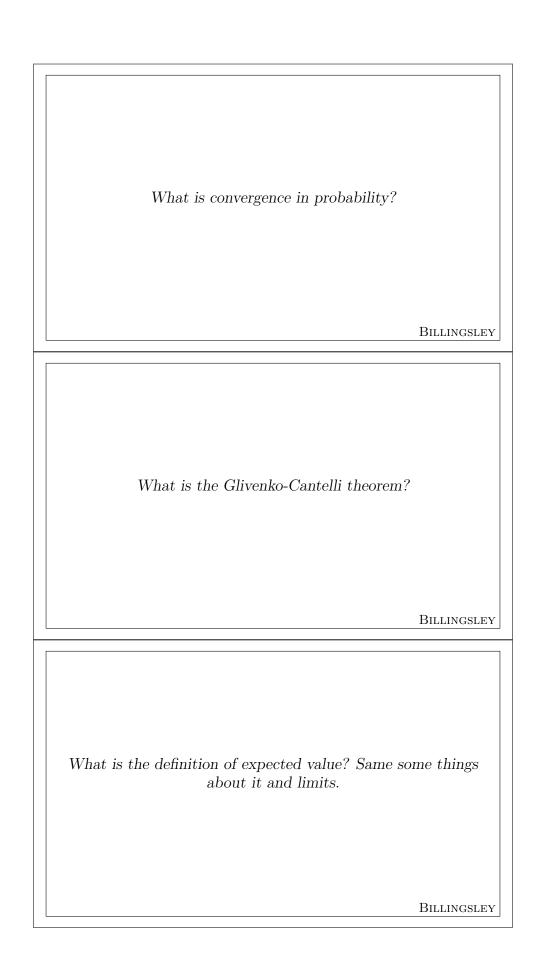
if $\{I_i\}$ is any finite partition of (a,b] into subintervals satisfying $\lambda(I_i) < \delta$ and if $x_i \in I_i$ for each i.

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 $B \in \mathcal{Y}$. 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.

If (X, \mathcal{X}, μ) and $(Y, \mathcal{Y}, \hat{)}$ are σ – field measure spaces, then

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

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



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

$$\lim_{n} P(|X_n - X| \ge \epsilon) = 0$$

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 $i \to \infty$.

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

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.

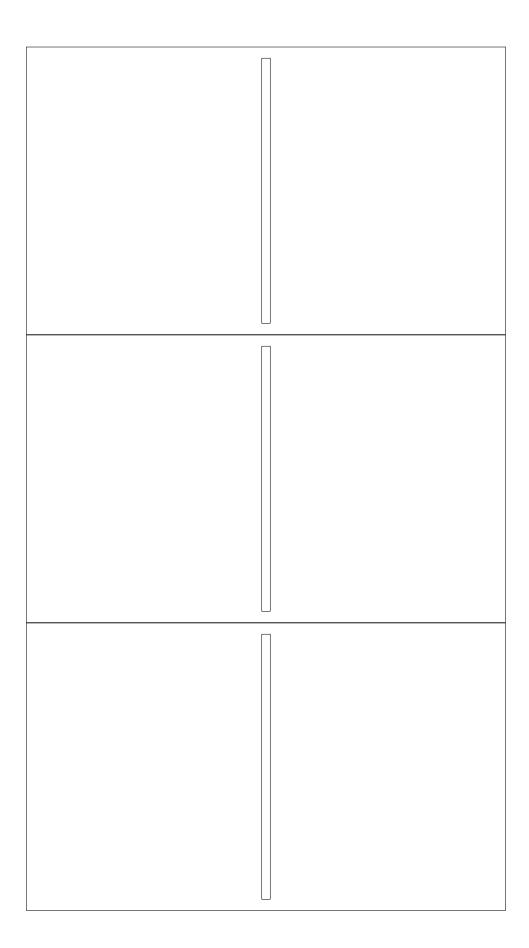
 $\mathbb{E}[X] = \int X dP = \int_{\Omega} X(w) P(dw).$ $\mathbb{E}[X1_A] = \int_{A} X dP.$ $\mathbb{E}[g(X)] = \int_{\Omega} g(X(w)) P(dw).$

Absolute moments $\mathbb{E}[|X|^k] = \int_{-\infty}^{\infty} |x|^k P(dx)$. By the way P(dx) intuitively refers to the P measure of a small change in x. 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
Billingsley
Billingsley

The MGF is $M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{sx} \mu(dx)$
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: $M_X(t) = pe^t + (1 - p)$.

BILLINGSLEY	
Billingsley	
Billingsley	BILLINGSLEY
	BILLINGSLEV
	PILLINGSHE I
	D
Billingsley	RILLINGGLEV



BILLINGSLEY	
Billingsley	
Billingsley	BILLINGSLEY
	BILLINGSLEV
	PILLINGSHE I
	D
Billingsley	RILLINGGLEV

