

Refinements Based on Midterm Feedback

More Detailed, Formal Proofs / Rigor

- ▶ I will incorporate more fully written, step-by-step proofs in lecture.
- ▶ For major theorems, I will prioritize presenting the formal proof structure first, followed by intuition.
- ▶ For longer, multi-part proofs, I will upload detailed written notes.
- ▶ Discussion sessions can be used to work through these longer technical proofs when appropriate.

Clearer Indication of Expected Level of Rigor

- ▶ At the beginning of major topics, I will clarify the level of rigor expected in exam answers.

More Examples

- ▶ I will include more concrete worked examples and applications immediately after major definitions and theorems.

Refinements Based on Midterm Feedback

Structure & Logistics

- ▶ **Earlier discussion problems:** I will post discussion problems one week in advance (immediately after the previous discussion session).
- ▶ **Homework feedback & solution discussion:**
 - ▶ I will frontload homework timing to allow earlier feedback before exams.
 - ▶ We will discuss homework solutions immediately after they are due, while the material is still fresh.
- ▶ **Big-picture connections (“the story”):**
 - ▶ I will periodically pause to review how definitions and theorems connect.
 - ▶ Discussion sessions may also be used for broader conceptual review when appropriate.
- ▶ **More detailed slides:** I will provide supplementary technical notes to support board proofs.
- ▶ **Midterm 2 timing:** March 13? But this is ICS Grad day!

Refinements Based on Midterm Feedback

Delivery & Board Clarity

- ▶ I will adopt a clearer sequence: **write** → **pause** → **explain**.
- ▶ I will allow time to finish copying before moving forward or asking conceptual questions.
- ▶ I will outline the logical roadmap of a proof before entering technical steps.
- ▶ Supplementary written notes will support complex board proofs when needed.

The overall pace of the course will remain the same.

Conceptual Integration & Abstraction

Feedback:

- ▶ Hard to see how definitions and theorems connect.
- ▶ Concepts feel abstract and difficult to implement.
- ▶ Challenging to relate measure-theoretic language back to statistics.
- ▶ Convergence theorems and generalization beyond the “stats lens” are difficult.

Response:

- ▶ I will periodically step back and map how definitions and results connect.
- ▶ I will explicitly distinguish closely related terms and clarify notation.
- ▶ I will incorporate short statistical interpretations when introducing major results.
- ▶ For convergence theorems, I will emphasize what each mode of convergence means conceptually and when it matters.

Independence

Independence of events

Definition

Two events $A, B \in \mathcal{F}$ are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Interpretation: learning whether A occurred does not change the probability of B .

Example: independence of events

Flip a fair coin twice.

$$A = \{\text{first flip is Heads}\}, \quad B = \{\text{second flip is Heads}\}.$$

Then

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B).$$

Independence of random variables

Definition

Random variables $X, Y : \Omega \rightarrow \mathbb{R}$ are *independent* if for all Borel sets $C, D \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D).$$

Equivalently, the events $\{X \in C\}$ and $\{Y \in D\}$ are independent for all Borel C, D .

Independence of σ -fields

Definition

σ -fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad \forall A \in \mathcal{G}, B \in \mathcal{H}.$$

Why this level? Independence of random variables is really independence of the *information* they generate.

How the notions connect

Hierarchy of independence

events \subset random variables \subset σ -fields

- ▶ Event independence is the base definition.
- ▶ $X \perp Y$ means $\{X \in C\}$ is independent of $\{Y \in D\}$ for all Borel C, D .
- ▶ σ -field independence is the most general formulation.

Payoff: independence \Rightarrow product laws \Rightarrow Fubini \Rightarrow factorization of expectations.

From random variables to σ -fields (and back)

Theorem

1. If X and Y are independent, then $\sigma(X)$ and $\sigma(Y)$ are independent.
2. Conversely, if \mathcal{G}, \mathcal{H} are independent, X is \mathcal{G} -measurable, and Y is \mathcal{H} -measurable, then X and Y are independent.

Proof (roadmap)

Roadmap: show independence on the generating events $\{X \in C\}$ and $\{Y \in D\}$.

Proof.

Assume X and Y are independent. Let $A \in \sigma(X)$ and $B \in \sigma(Y)$.

By definition of $\sigma(X)$, there exists $C \in \mathcal{B}(\mathbb{R})$ such that $A = \{X \in C\}$. Similarly, there exists $D \in \mathcal{B}(\mathbb{R})$ such that $B = \{Y \in D\}$. Then

$$\mathbb{P}(A \cap B) = \mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D) = \mathbb{P}(A)\mathbb{P}(B).$$

Thus $\sigma(X)$ and $\sigma(Y)$ are independent. □

Proof (converse)

Proof.

Assume \mathcal{G}, \mathcal{H} are independent, X is \mathcal{G} -measurable, and Y is \mathcal{H} -measurable. For any $C, D \in \mathcal{B}(\mathbb{R})$, measurability implies $\{X \in C\} \in \mathcal{G}$ and $\{Y \in D\} \in \mathcal{H}$. By independence of \mathcal{G}, \mathcal{H} ,

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(\{X \in C\} \cap \{Y \in D\}) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D).$$

So X and Y are independent. □

Complements preserve independence

Theorem

If A and B are independent events, then so are A^c and B , A and B^c , and A^c and B^c .

Proof.

Start from $\mathbb{P}(B) = \mathbb{P}((A \cap B) \dot{\cup} (A^c \cap B))$ to get

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Using independence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, we have

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B).$$

Thus A^c and B are independent. The other cases follow by symmetry and applying the same argument to (A, B^c) . □

Indicators encode event independence

Theorem

Events A, B are independent if and only if their indicator variables $\mathbf{1}_A$ and $\mathbf{1}_B$ are independent.

Proof.

(\Rightarrow) Assume A, B are independent. For any Borel set $C \subseteq \mathbb{R}$, $\{\mathbf{1}_A \in C\}$ is one of $\emptyset, A, A^c, \Omega$ (since $\mathbf{1}_A$ only takes values 0, 1). Likewise, $\{\mathbf{1}_B \in D\}$ is one of $\emptyset, B, B^c, \Omega$. Independence of $\mathbf{1}_A, \mathbf{1}_B$ requires checking

$$\mathbb{P}(\{\mathbf{1}_A \in C\} \cap \{\mathbf{1}_B \in D\}) = \mathbb{P}(\{\mathbf{1}_A \in C\})\mathbb{P}(\{\mathbf{1}_B \in D\})$$

for all such C, D , which reduces to the four nontrivial cases

$(A, B), (A^c, B), (A, B^c), (A^c, B^c)$, all true by Theorem 5.

(\Leftarrow) Conversely, if $\mathbf{1}_A$ and $\mathbf{1}_B$ are independent, take $C = \{1\}$ and $D = \{1\}$: then $\{\mathbf{1}_A \in C\} = A$ and $\{\mathbf{1}_B \in D\} = B$, giving $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

□

Independence of many objects

Definition

Events A_1, \dots, A_n are independent if for every index set $I \subseteq \{1, \dots, n\}$,

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

An infinite collection is independent if every finite subcollection is independent.

Pairwise independence is not enough

Definition

A_1, \dots, A_n are *pairwise independent* if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \text{for all } i \neq j.$$

Pairwise independence $\not\Rightarrow$ joint independence.

Counterexample: pairwise but not jointly independent

Let X_1, X_2, X_3 be i.i.d. Bernoulli(1/2).

Define

$$A_1 = \{X_2 = X_3\}, \quad A_2 = \{X_3 = X_1\}, \quad A_3 = \{X_1 = X_2\}.$$

- ▶ For $i \neq j$,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(X_1 = X_2 = X_3) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

So the events are pairwise independent.

- ▶ But

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(X_1 = X_2 = X_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

So they are not jointly independent.

How to check independence in practice

To verify independence of X_1, \dots, X_n , it suffices to check a generating class.

- ▶ **CDF factorization:** for all $x_1, \dots, x_n \in (-\infty, \infty]$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i).$$

- ▶ **(If densities exist)** joint density factorizes into the product of marginals.

Takeaway: independence reduces a joint object to marginal objects.

Practical checking: CDF condition

Theorem (Sufficient condition for independence)

It is sufficient for X_1, \dots, X_n to be independent that for all $x_1, \dots, x_n \in (-\infty, \infty]$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i).$$

Product distributions (rectangle computation)

Theorem

If X_1, \dots, X_n are independent and μ_i is the law of X_i , then for all Borel sets A_1, \dots, A_n ,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mu_i(A_i).$$

This shows the joint distribution agrees with $\mu_1 \times \dots \times \mu_n$ on rectangles.

Proof.

By independence,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

Since μ_i is the distribution of X_i , $\mathbb{P}(X_i \in A_i) = \mu_i(A_i)$. Substituting yields the claimed identity. □

Why product measures matter

Independence implies a structural identity:

joint law = product of marginal laws.

This is the bridge from probability to integration:

product law \Rightarrow Fubini/Tonelli \Rightarrow factorization of expectations.

Everything that follows comes from this structural identity.

Joint law as a product measure

Theorem

If X_1, \dots, X_n are independent with laws μ_1, \dots, μ_n , then

$$(X_1, \dots, X_n) \sim \mu_1 \times \dots \times \mu_n \quad \text{on } (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)).$$

Remark. We already proved equality on rectangles. One can show (from standard uniqueness of measures on generating classes) that this determines the whole law on $\mathcal{B}(\mathbb{R}^n)$. We will treat this as a known measure-theoretic fact.

Fubini payoff: expectation of $h(X, Y)$

Theorem

Let X, Y be independent with laws μ, ν . If $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ is measurable and either $h \geq 0$ or $\mathbb{E}|h(X, Y)| < \infty$, then

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) \mu(dx) \nu(dy).$$

Proof.

By Theorem 11, (X, Y) has distribution $\mu \times \nu$ on \mathbb{R}^2 . By the definition of expectation as an integral with respect to the law,

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) (\mu \times \nu)(dx, dy).$$

When $h \geq 0$ (or integrable), Fubini/Tonelli applies to give

$$\int_{\mathbb{R}^2} h d(\mu \times \nu) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) \mu(dx) \nu(dy).$$



Factorization of expectations

Corollary

If X, Y are independent and $f, g : \mathbb{R} \rightarrow \mathbb{R}$ are measurable with $\mathbb{E}|f(X)| < \infty$ and $\mathbb{E}|g(Y)| < \infty$, then

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)].$$

Proof.

Apply Theorem 12 with $h(x, y) = f(x)g(y)$:

$$\mathbb{E}[f(X)g(Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)g(y) \mu(dx) \nu(dy).$$

Since $g(y)$ does not depend on x ,

$$\int_{\mathbb{R}} f(x)g(y) \mu(dx) = g(y) \int_{\mathbb{R}} f(x) \mu(dx) = g(y)\mathbb{E}[f(X)].$$

Therefore

$$\mathbb{E}[f(X)g(Y)] = \int_{\mathbb{R}} g(y)\mathbb{E}[f(X)] \nu(dy) = \mathbb{E}[f(X)] \int_{\mathbb{R}} g(y) \nu(dy) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$



Expectation of products: n variables

Theorem

If X_1, \dots, X_n are independent and $\mathbb{E}|X_i| < \infty$ for all i , then

$$\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mathbb{E}[X_i].$$

Proof.

Let $Y = \prod_{i=2}^n X_i$. By independence, X_1 is independent of (X_2, \dots, X_n) , and hence independent of Y (as Y is a measurable function of X_2, \dots, X_n).

By factorization Corollary,

$$\mathbb{E}[X_1 Y] = \mathbb{E}[X_1] \mathbb{E}[Y].$$

Apply the same argument inductively to $\mathbb{E}[Y] = \mathbb{E}[\prod_{i=2}^n X_i]$ to obtain

$$\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mathbb{E}[X_i].$$



Independence vs uncorrelated (quick warning)

Key idea: uncorrelatedness checks one moment identity; independence requires factorization for *all* measurable events.

If $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$ and

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y],$$

then X and Y are *uncorrelated*. Independence implies uncorrelatedness, but not conversely.

Takeaway: moment factorization at one order does not imply full joint factorization.

Uncorrelated does not imply independent (setup)

Example

Let

$$X \sim \text{Uniform}(-1, 1), \quad Y = X^2.$$

Compute expectations:

$$\mathbb{E}[X] = 0, \quad \mathbb{E}[Y] = \mathbb{E}[X^2] = \frac{1}{3}.$$

By symmetry,

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0.$$

Therefore,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y],$$

so X and Y are *uncorrelated*.

Why independence fails

If X and Y were independent, then

$$\mathbb{P}(Y \leq 1/4 | X = x) = \mathbb{P}(Y \leq 1/4) \quad \text{for all } x.$$

But since $Y = X^2$,

$$\mathbb{P}(Y \leq 1/4 | X = x) = \begin{cases} 1, & |x| \leq 1/2, \\ 0, & |x| > 1/2. \end{cases}$$

This conditional probability depends on x . Unconditionally,

$$\mathbb{P}(Y \leq 1/4) = \mathbb{P}(|X| \leq 1/2) = \frac{1}{2}.$$

Thus,

$$\mathbb{P}(Y \leq 1/4 | X = x) \neq \mathbb{P}(Y \leq 1/4),$$

so X and Y are *not independent*.