Independence

Last modified: 2026-09-28 23:45:35 (PDT)

Definition 0.1 (Statistical independence)**.** Random variables $X_1, \dots, X_n$ are **statistically independent** if, for all sets of real numbers $A_1, \dots, A_n$, the probability that every X_i falls in its set is the product of the individual probabilities:

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

Remark. We write $X \perp\!\!\!\perp Y$ for "X and Y are independent". The symbol $\perp\!\!\!\perp$ is essentially $\prod$ upside-down, which can remind you of the definition.

Theorem 0.1 (Independence of discrete random variables: the joint PMF factors)**.** *Discrete random variables $X_1, \dots, X_n$ are statistically independent if and only if their joint PMF factors into the product of their PMFs: for all real numbers $x_1, \dots, x_n$,*

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

Proof

Proof. **Only if.** Take $A_i = \{x_i\}$ for each i in Definition 0.1.
If. Let $A_1, \dots, A_n$ be sets of real numbers, and let C be the set of tuples $(x_1, \dots, x_n)$ with each $x_i \in A_i \cap \mathcal{R}(X_i)$. Each range $\mathcal{R}(X_i)$ is countable, so C is countable. Each X_i takes its values in $\mathcal{R}(X_i)$, so the event $\{X_1 \in A_1, \dots, X_n \in A_n\}$ is the disjoint union of the events $\{X_1 = x_1, \dots, X_n = x_n\}$ over $(x_1, \dots, x_n) \in C$:

$$
\begin{aligned}
\Pr(X_1 \in A_1, \dots, X_n \in A_n) &= \sum_{(x_1,\dots,x_n)\in C} \mathrm{P}(X_1 = x_1, \dots, X_n = x_n) && \text{(countable additivity over the disjoint eve} \\
&= \sum_{(x_1,\dots,x_n)\in C} \prod_{i=1}^{n} \mathrm{P}(X_i = x_i) && \text{(the joint PMF factors)} \\
&= \prod_{i=1}^{n} \sum_{x_i \in A_i \cap \mathcal{R}(X_i)} \mathrm{P}(X_i = x_i) && \text{(a sum over the product set } C \text{ of non-nega} \\
&= \prod_{i=1}^{n} \Pr(X_i \in A_i) && \text{(countable additivity over the disjoint eve}
\end{aligned}
$$

□

Example 0.1 (Two fair coin flips)**.** Flip two fair coins, and let X_1 and X_2 indicate heads on the first and second flip. Each of the four outcomes has probability 1/4, so for example:

$$\mathrm{P}(X_1 = 1, X_2 = 1) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = \mathrm{P}(X_1 = 1)\ \mathrm{P}(X_2 = 1)$$

and the same factorization holds for the other three pairs of values, so $X_1 \perp\!\!\!\perp X_2$ by Theorem 0.1. In contrast, X_1 and the total number of heads, $X_1 + X_2$, are not independent: $\mathrm{P}(X_1 = 0, X_1 + X_2 = 2) = 0$, but $\mathrm{P}(X_1 = 0)\ \mathrm{P}(X_1 + X_2 = 2) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$.

Theorem 0.2 (Independence with densities: the joint density factors)**.** ***Continuous case.*** *Continuous random variables X and Y with densities $\mathrm{p}(X = x)$ and $\mathrm{p}(Y = y)$ are statistically independent if and only if the product $\mathrm{p}(X = x)\ \mathrm{p}(Y = y)$ is a joint density of X and Y.*
Mixed case. *A discrete random variable X and a continuous random variable Y with density $\mathrm{p}(Y = y)$ are statistically independent if and only if the product $\mathrm{P}(X = x)\ \mathrm{p}(Y = y)$ is a joint density-mass function of X and Y.*

Remark. The proof is beyond these notes' scope: it needs measure theory that the notes do not develop, to extend a density's integrals from intervals to general sets and to exchange the order of integration (Casella and Berger 2002; Billingsley 1995). The continuous case extends to n random variables with a joint density of all n. The statement says "a joint density" rather than "the joint density" because a density is not unique: changing it on a set of zero area (zero length, in the mixed case) changes none of its integrals.

Example 0.2 (The PMF form fails for continuous random variables)**.** The factorization in Theorem 0.1 does not work as a definition of independence for continuous random variables: there, both sides are 0 at every point, so it would call every pair of continuous random variables independent. For instance, let $X \sim \text{Uniform}(0,1)$ (uniform distribution), whose density is 1 on $[0,1]$, and let $Y = X$. For all real numbers x and y, the event $\{X = x, Y = y\}$ is $\{X = x\}$ if $y = x$ and empty otherwise, so it has probability 0, because $\Pr(X = x) = 0$ for the continuous X. Likewise $\Pr(X = x)\Pr(Y = y) = 0 \cdot 0 = 0$, so the PMF form holds. But X and Y are not independent:

$$\begin{aligned}
\Pr(X \in [0, \tfrac{1}{2}], Y \in [0, \tfrac{1}{2}]) &= \Pr(X \in [0, \tfrac{1}{2}]) && (Y = X) \\
&= \int_0^{1/2} 1\,dx && (\text{the density of } X \text{ is 1 on } [0,1]) \\
&= \tfrac{1}{2} && (\text{integrate})
\end{aligned}$$

while $\Pr(X \in [0, \tfrac{1}{2}])\Pr(Y \in [0, \tfrac{1}{2}]) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Definition 0.2 (Conditional independence)**.** Random variables $Y_1, \dots, Y_n$ are **conditionally independent** given a discrete random variable (or vector) $\tilde{X}$ if, for every value $\tilde{x}$ with $\Pr(\tilde{X} = \tilde{x}) > 0$ and all sets of real numbers $A_1, \dots, A_n$, the conditional probability that every Y_i falls in its set is the product of the individual conditional probabilities:

$$\Pr(Y_1 \in A_1, \dots, Y_n \in A_n \mid \tilde{X} = \tilde{x}) = \prod_{i=1}^{n} \Pr(Y_i \in A_i \mid \tilde{X} = \tilde{x})$$

When $\tilde{X}$ is continuous, every event $\{\tilde{X} = \tilde{x}\}$ has probability 0, so the condition is stated with densities instead: for every $\tilde{x}$ with $\mathrm{p}(\tilde{X} = \tilde{x}) > 0$ and all $y_1, \dots, y_n$,

$$\frac{\mathrm{p}(\tilde{X} = \tilde{x}, Y_1 = y_1, \dots, Y_n = y_n)}{\mathrm{p}(\tilde{X} = \tilde{x})} = \prod_{i=1}^{n} \frac{\mathrm{p}(\tilde{X} = \tilde{x}, Y_i = y_i)}{\mathrm{p}(\tilde{X} = \tilde{x})}$$

where each $\mathrm{p}(\cdot)$ is a joint density (with probability masses in place of densities for any discrete Y_i, as in a joint density-mass function).

Remark. For continuous $\tilde{X}$, each ratio in the density form is the density of the Y's under their conditional distribution given $\tilde{X} = \tilde{x}$, so both forms say the same thing: given $\tilde{X} = \tilde{x}$, the Y_i's joint distribution factors.

Example 0.3 (Two tests of the same patient)**.** Let X indicate whether a patient has a disease, and let Y_1 and Y_2 indicate positive results on two tests whose errors are unrelated, so that Y_1 and Y_2 are conditionally independent given X. Suppose $\Pr(X = 1) = 0.1$, each test is positive with probability 0.9 if $X = 1$ and with probability 0.1 if $X = 0$. Then, by the law of total probability:

$$
\begin{aligned}
\Pr(Y_1 = 1, Y_2 = 1) &= (0.9)(0.9)(0.1) + (0.1)(0.1)(0.9) && \text{(condition on } X\text{; factor given } X\text{)} \\
&= 0.081 + 0.009 && \text{(multiply)} \\
&= 0.09 && \text{(add)}
\end{aligned}
$$

but $\Pr(Y_1 = 1) = (0.9)(0.1) + (0.1)(0.9) = 0.18$, so $\Pr(Y_1 = 1)\Pr(Y_2 = 1) = 0.0324 \neq 0.09$: the tests are conditionally independent given X, but not independent.

Example 0.4 (Two coins, given their total)**.** Let Y_1 and Y_2 indicate heads on two fair coin flips, which are independent (Example 0.1), and let $X = Y_1 + Y_2$ be the number of heads. Given $X = 1$, which has probability $1/2$, exactly one coin is heads, so:

$$
\begin{aligned}
\Pr(Y_1 = 1, Y_2 = 1 \mid X = 1) &= \frac{\Pr(Y_1 = 1, Y_2 = 1, X = 1)}{\Pr(X = 1)} && \text{(definition of conditional probability)} \\
&= \frac{0}{1/2} && \text{(two heads make } X = 2\text{, not 1)} \\
&= 0 && \text{(divide)}
\end{aligned}
$$

but

$$
\begin{aligned}
\Pr(Y_1 = 1 \mid X = 1) &= \frac{\Pr(Y_1 = 1, X = 1)}{\Pr(X = 1)} && \text{(definition of conditional probability)} \\
&= \frac{\Pr(Y_1 = 1, Y_2 = 0)}{\Pr(X = 1)} && (Y_1 = 1 \text{ and } X = 1 \text{ means } Y_2 = 0) \\
&= \frac{1/4}{1/2} && \text{(substitute)} \\
&= \tfrac{1}{2} && \text{(divide)}
\end{aligned}
$$

and likewise $\Pr(Y_2 = 1 \mid X = 1) = 1/2$, so the product of the conditional probabilities is $1/4 \neq 0$: Y_1 and Y_2 are independent, but not conditionally independent given X.

Proposition 0.1 (Neither independence nor conditional independence implies the other)**.** *There are random variables Y_1, Y_2, and X such that Y_1 and Y_2 are conditionally independent given X but not independent, and there are random variables Y_1, Y_2, and X such that Y_1 and Y_2 are independent but not conditionally independent given X.*

Proof

Proof. Example 0.3 gives the first case, and Example 0.4 gives the second. □

Proposition 0.2 (Conditional independence when each Y_i depends only on its own X_i)**.** *Let $\tilde{X} = (X_1, \dots, X_n)$ be a discrete random vector, and let $Y_1, \dots, Y_n$ be conditionally independent given $\tilde{X}$. Suppose also that each Y_i depends on $\tilde{X}$ only through X_i: for every $\tilde{x} = (x_1, \dots, x_n)$ with $\Pr(\tilde{X} = \tilde{x}) > 0$ and every set of real numbers A_i,*

$$\Pr(Y_i \in A_i \mid \tilde{X} = \tilde{x}) = \Pr(Y_i \in A_i \mid X_i = x_i)$$

Then, for every such $\tilde{x}$ and all sets of real numbers $A_1, \dots, A_n$:

$$\Pr(Y_1 \in A_1, \dots, Y_n \in A_n \mid \tilde{X} = \tilde{x}) = \prod_{i=1}^{n} \Pr(Y_i \in A_i \mid X_i = x_i)$$

Proof

Proof. Each conditional probability given $X_i = x_i$ is defined, because $\Pr(X_i = x_i) \geq \Pr(\tilde{X} = \tilde{x}) > 0$. So:

$$\begin{aligned} \Pr(Y_1 \in A_1, \dots, Y_n \in A_n \mid \tilde{X} = \tilde{x}) &= \prod_{i=1}^{n} \Pr(Y_i \in A_i \mid \tilde{X} = \tilde{x}) && \text{(conditional independence given } \tilde{X}\text{)} \\ &= \prod_{i=1}^{n} \Pr(Y_i \in A_i \mid X_i = x_i) && \text{(each } Y_i \text{ depends on } \tilde{X} \text{ only through } X_i\text{)} \end{aligned}$$

□

Remark. In regression, $\tilde{X}$ is usually the full set of covariates $(X_1, \dots, X_n)$, and a model usually also assumes, as Proposition 0.2 does, that each Y_i depends on the covariates only through its own X_i. That second assumption is a separate one, not part of conditional independence.

Definition 0.3 (Identically distributed)**.** Random variables $X_1, \dots, X_n$ are **identically distributed** if they all have the same CDF:

$$\forall i \in \{1, \dots, n\}, \forall x \in \mathbb{R} : \Pr(X_i \leq x) = \Pr(X_1 \leq x)$$

Theorem 0.3 (Identically distributed discrete random variables have the same PMF). *Discrete random variables $X_1, \dots, X_n$ are identically distributed if and only if they all have the same PMF: for every $i \in \{1, \dots, n\}$ and every real number x,*

$$\mathrm{P}(X_i = x) = \mathrm{P}(X_1 = x)$$

Proof

Proof. Write $F_i(t) \stackrel{\text{def}}{=} \Pr(X_i \le t)$ for the CDF of X_i.

Only if: the CDF determines the PMF through its jumps. Fix i and x. Every number below x lies in exactly one of the intervals $(-\infty, x-1]$, $(x-1, x-\frac{1}{2}]$, $(x-\frac{1}{2}, x-\frac{1}{3}]$, ..., so the event $\{X_i < x\}$ is the disjoint union of the events $B_1 \stackrel{\text{def}}{=} \{X_i \le x-1\}$ and $B_k \stackrel{\text{def}}{=} \{x - \frac{1}{k-1} < X_i \le x - \frac{1}{k}\}$ for $k \ge 2$, and for each K, the events $B_1, \dots, B_K$ have union $\{X_i \le x - \frac{1}{K}\}$:

$$\begin{aligned}
\Pr(X_i < x) &= \sum_{k=1}^{\infty} \Pr(B_k) && \text{(countable additivity)} \\
&= \lim_{K\to\infty} \sum_{k=1}^{K} \Pr(B_k) && \text{(definition of an infinite series)} \\
&= \lim_{K\to\infty} \Pr(X_i \le x - \tfrac{1}{K}) && \text{(finite additivity)} \\
&= \lim_{K\to\infty} F_i(x - \tfrac{1}{K}) && \text{(definition of the CDF)}
\end{aligned}$$

The event $\{X_i \le x\}$ is the disjoint union of $\{X_i < x\}$ and $\{X_i = x\}$, so:

$$\begin{aligned}
\mathrm{P}(X_i = x) &= \Pr(X_i \le x) - \Pr(X_i < x) && \text{(finite additivity)} \\
&= F_i(x) - \Pr(X_i < x) && \text{(definition of the CDF)} \\
&= F_i(x) - \lim_{K\to\infty} F_i(x - \tfrac{1}{K}) && \text{(the display above)} \\
&= F_1(x) - \lim_{K\to\infty} F_1(x - \tfrac{1}{K}) && (F_i = F_1) \\
&= \mathrm{P}(X_1 = x) && \text{(the same two steps, for } X_1\text{)}
\end{aligned}$$

If: the PMF determines the CDF through sums. Let S be the set of values u with $\mathrm{P}(X_1 = u) > 0$. Since the PMFs are equal, S is also the set of u with $\mathrm{P}(X_i = u) > 0$, so $S \subseteq \mathcal{R}(X_i)$ and $S \subseteq \mathcal{R}(X_1)$. For every real t, the event $\{X_i \le t\}$ is the disjoint union of the events $\{X_i = u\}$ over the countably many $u \in \mathcal{R}(X_i)$ with $u \le t$, so:

$$
\begin{aligned}
F_i(t) &= \sum_{u \in \mathcal{R}(X_i),\, u \le t} \mathrm{P}(X_i = u) && \text{(countable additivity)} \\
&= \sum_{u \in S,\, u \le t} \mathrm{P}(X_i = u) && \text{(drop the terms that are 0)} \\
&= \sum_{u \in S,\, u \le t} \mathrm{P}(X_1 = u) && \text{(the PMFs are equal)} \\
&= \sum_{u \in \mathcal{R}(X_1),\, u \le t} \mathrm{P}(X_1 = u) && \text{(restore the terms that are 0)} \\
&= F_1(t) && \text{(countable additivity)}
\end{aligned}
$$

□

Remark. As with independence (Example 0.2), the PMF form would not work as a definition in general: every continuous random variable has $\Pr(X_i = x) = 0$ at every x.

Example 0.5 (Identically distributed but not independent)**.** In Example 0.1, X_1 and $1 - X_1$ (the indicator of tails on the first flip) both take the values 0 and 1 with probability 1/2 each, so they are identically distributed (Theorem 0.3). They are not independent: knowing one determines the other.

Definition 0.4 (Conditionally identically distributed)**.** Random variables $Y_1, \dots, Y_n$ are **conditionally identically distributed** given random variables $X_1, \dots, X_n$ if the conditional CDF of each Y_i given $X_i = x$ is one shared function $G(y \mid x)$ of y and x: for every $i \in \{1, \dots, n\}$, every y, and every x with $\Pr(X_i = x) > 0$,

$$\Pr(Y_i \le y \mid X_i = x) = G(y \mid x)$$

When X_i is continuous, the condition applies at every x with $\mathrm{p}(X_i = x) > 0$, with $\Pr(Y_i \le y \mid X_i = x)$ computed from densities: the integral of $\mathrm{p}(X_i = x, Y_i = u)/\,\mathrm{p}(X_i = x)$ over $u \le y$ (a sum over values $u \le y$ when Y_i is discrete).

Example 0.6 (A shared regression model)**.** Suppose each Y_i is binary, with $\Pr(Y_i = 1 \mid X_i = x) = \pi(x)$ for one function π shared by every i (for instance, $\pi(x) = x/(1+x)$ for a dose $x \ge 0$). Then $Y_1, \dots, Y_n$ are conditionally identically distributed given $X_1, \dots, X_n$, with $G(y \mid x) = 1 - \pi(x)$ for $0 \le y < 1$ (and 0 for $y < 0$, 1 for $y \ge 1$). Their marginal distributions can still differ: if participant 1 always receives dose 0 and participant 2 always receives dose 1, then $\Pr(Y_1 = 1) = 0$ but $\Pr(Y_2 = 1) = 1/2$.

Definition 0.5 (Independent and identically distributed)**.** Random variables $X_1, \dots, X_n$ are **independent and identically distributed** (shorthand: "X_i iid") if they are:

- statistically independent, and
- identically distributed.

Remark. The IID assumption is one of the most common assumptions in introductory statistics: it says a sample $X_1, \dots, X_n$ can be treated as n independent draws from a single shared distribution.

Example 0.7 (Repeated die rolls)**.** The results of n rolls of the same fair die are IID: the rolls are independent, and each is uniform on $\{1, \dots, 6\}$.

Definition 0.6 (Conditionally independent and identically distributed)**.** Random variables $Y_1, \dots, Y_n$ are **conditionally independent and identically distributed** given $X_1, \dots, X_n$ (shorthand: "$Y_i \mid X_i$ ciid" or just "$Y_i \mid X_i$ iid") if:

- $Y_1, \dots, Y_n$ are conditionally independent given $(X_1, \dots, X_n)$,
- each Y_i depends on $(X_1, \dots, X_n)$ only through X_i, and
- $Y_1, \dots, Y_n$ are conditionally identically distributed given $X_1, \dots, X_n$.

Example 0.8 (The usual regression assumption)**.** In Example 0.6, if the Y_i are also conditionally independent given all the doses, and each Y_i depends on the doses only through its own X_i, then $Y_i \mid X_i$ ciid, and the joint conditional PMF is $\prod_{i=1}^{n} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$: one shared function evaluated at each (x_i, y_i).

Video lecture

Hutchinson's Probability Refresher (27 min) covers independence (Hutchinson, n.d.). The login for the video site is posted on Canvas.

References

Billingsley, Patrick. 1995. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley.

Casella, George, and Roger Berger. 2002. *Statistical Inference*. 2nd ed. Cengage Learning. https://www.cengage.com/c/statistical-inference-2e-casella-berger/9780534243128/.

Hutchinson, Brian. n.d. *DATA 471/571 (Machine Learning) and CSCI 481/581 (Deep Learning) Video Lectures*. Western Washington University. Accessed September 28, 2026. https://facultyweb.cs.wwu.edu/~hutchib2/video_lectures/data371/.