Lecture 2: Random Variables and Expectation

Chapter 2.1

Charles A. Bouman

Purdue University

Wednesday, August 26, 2026

Outline

1. **Describing randomness**
   Probability spaces, random variables, the CDF, and the PDF.
2. **Properties of random variables**
   Expectation, joint distributions, and independence.
3. **Learning from observations**
   Conditional distributions and conditional expectation.

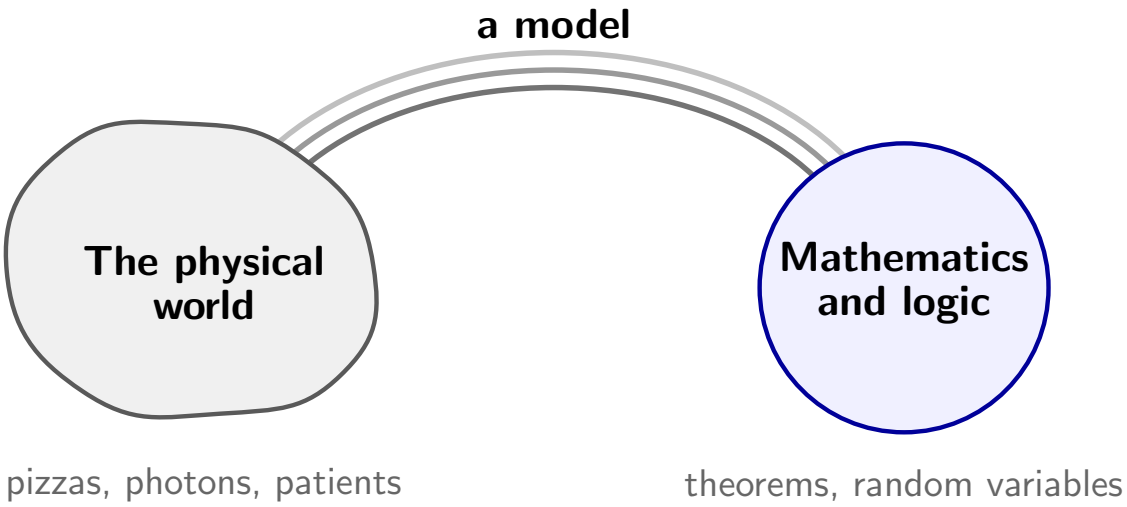
Reading: Chapter 2.1, *Foundations of Computational Imaging*.

Today's Big Ideas

- A random variable is a function on the sample space.
- The CDF and the PDF say how a random variable is distributed.
- Expectation is probability-weighted averaging, and it is always linear.
- Independence: the expectation of a product splits into a product.
- Observing one random variable gives a new distribution for another.

All of this depends on one object: conditional expectation, $\mathbb{E}[X|Y]$.

Two worlds, linked by models



- In mathematics a statement is true, false, or undecidable.
- In the physical world nothing is formally true; we gather evidence.
- A pizza is not a random variable; its diameter can be modeled as one.

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." (Albert Einstein, 1921)

Probability spaces

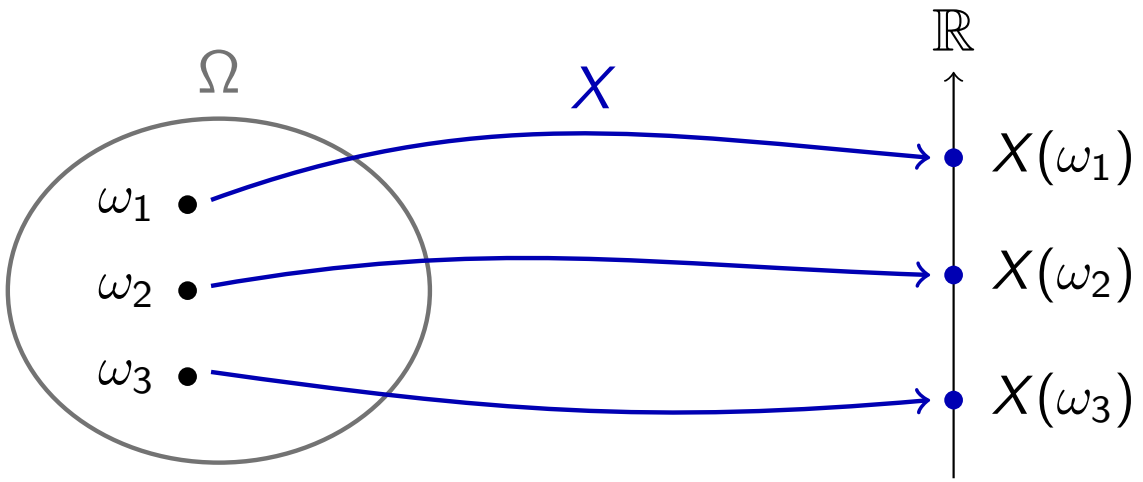
A **probability space** $(\Omega, \mathcal{B}, P)$ has three parts:

- Ω — the **sample space**: the set of all possible outcomes ω;
- $\mathcal{B}$ — the **events**: subsets of Ω with a probability;
- P — the **probability measure**: $P(A) \in [0, 1]$ for each event A.

A random variable is a function

A **random variable** is a function $X : \Omega \to \mathbb{R}$.

- $X(\omega)$ is the value X takes when the outcome is ω.
- The randomness is in which outcome occurs; X itself is fixed.
- Any function of a random variable is again a random variable.



Events written in terms of X

A probability statement about X is shorthand for a statement about outcomes.

$$\{X \leq t\} \triangleq \{\omega \in \Omega : X(\omega) \leq t\}$$

$$P\{X \leq 1\} \triangleq P\left(\{\omega \in \Omega : X(\omega) \leq 1\}\right)$$

- $\{X \leq t\}$ is a set of outcomes, so it is an event.
- $P\{X \leq 1\}$ is the probability measure applied to that event.

Random variables and the CDF

The **cumulative distribution function (CDF)** of a real-valued random variable X is given by

$$F(t) \triangleq P\{X \leq t\} \ ,$$

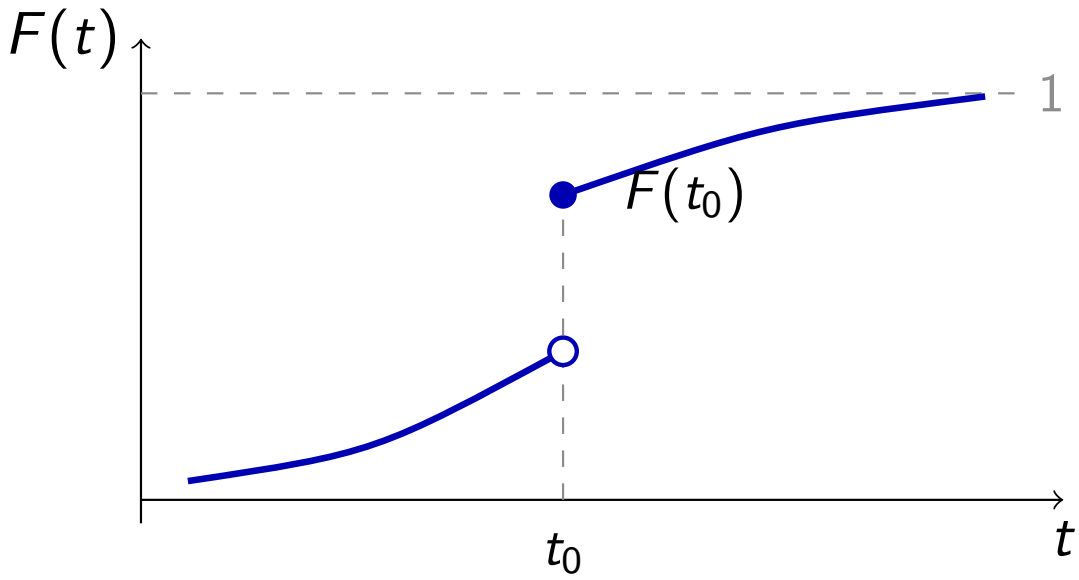
where $P\{X \leq t\}$ is the probability of the event that X is less than or equal to t.

$F(t)$ is a valid CDF if and only if

1. F is **monotone increasing**: $F(t_1) \leq F(t_2)$ whenever $t_1 < t_2$;
2. F is **right-hand continuous**: $\lim_{s \to t^+} F(s) = F(t)$ for every t;
3. $\lim_{t \to \infty} F(t) = 1$ and $\lim_{t \to -\infty} F(t) = 0$.

Right-hand continuity at a jump

A CDF is allowed to jump.



- Right-hand continuity: $F(t_0)$ is the top of the jump (filled dot).
- A jump means that $P\{X = t_0\} > 0$, and the height of the jump is that probability. Here X takes a discrete value.
- A CDF can have no jumps and no density: the Cantor function.

The probability density function

The **probability density function (PDF)** $p(t)$ of a random variable X is the function that satisfies

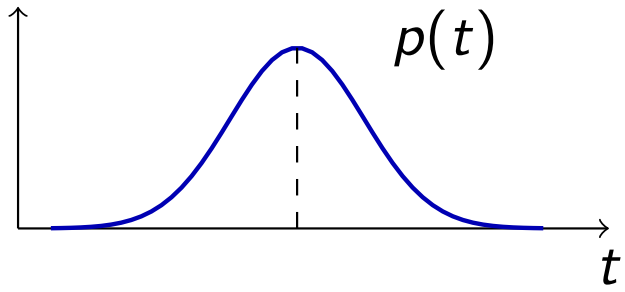
$$F(t) = \int_{-\infty}^{t} p(\tau)\, d\tau \ .$$

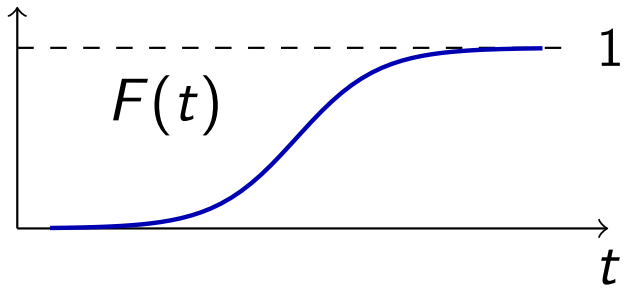
It exists when $F(t)$ is absolutely continuous, which is a reasonable assumption for most physical problems. Then

$$p(t) = \frac{dF(t)}{dt} \ .$$

A PDF and its CDF

Let X be the true brightness of a single pixel in a noisy image.





- The PDF says which brightness values are probable. The CDF accumulates that probability from the left.

Expectation

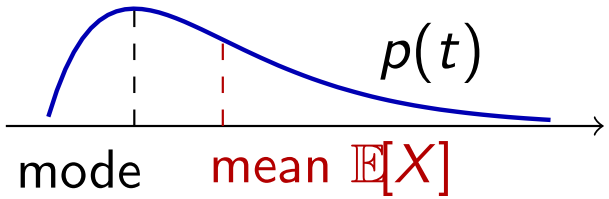
Expectation is probability-weighted averaging:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \tau \, p(\tau) \, d\tau \ .$$

More generally,

$$\mathbb{E}[X] \triangleq \int_{-\infty}^{\infty} \tau \, dF(\tau) \ ,$$

a **Riemann–Stieltjes** integral, which is defined even when the density does not exist.



- $\mathbb{E}[X]$ is the balance point, not the peak; mean and mode can differ.
- We assume $\mathbb{E}[|X|] < \infty$, so $\mathbb{E}[X]$ is well defined.

Property: linearity of expectation

Linearity of expectation

For all random variables X and Y,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \ .$$

- This follows directly from the integral definition of expectation.
- No independence is required. X and Y can be related in any way.

Joint distributions

One pixel is not an image. Let $X_1, \cdots, X_n$ be n random variables — say, n pixels. Their joint distribution is specified by the n-dimensional CDF

$$F(t_1, \cdots, t_n) = P\{X_1 \leq t_1, \cdots, X_n \leq t_n\} \ .$$

There is typically an associated n-dimensional PDF, $p(t_1, \cdots, t_n)$, so that

$$F(t_1, \cdots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} p(\tau_1, \cdots, \tau_n) \, d\tau_1 \cdots d\tau_n \ .$$

Functions of a random vector

Again, any function of the vector $X = (X_1, \cdots, X_n)$ is a new random variable.

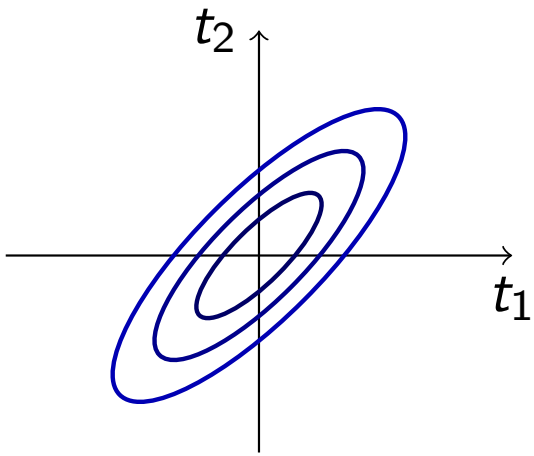
If $Y = g(X)$ where $g : \mathbb{R}^n \to \mathbb{R}$, then Y is a random variable, and we may compute its expectation as

$$\mathbb{E}[Y] = \int_{\mathbb{R}^n} g(\tau_1, \cdots, \tau_n)\, p(\tau_1, \cdots, \tau_n)\, d\tau_1 \cdots d\tau_n \ .$$

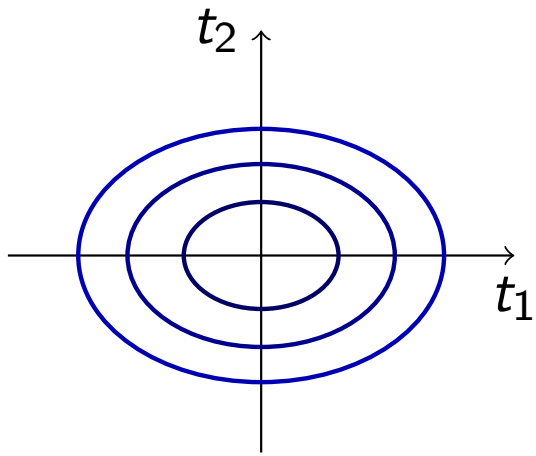
- Note that the expectation of Y is computed using the density of X.

Joint distribution of two pixels

What can the joint density of two pixels look like? Contours of $p(t_1, t_2)$:



neighboring pixels



distant pixels

- Left: knowing t_1 shifts t_2, so they are *dependent*.
- Right: knowing t_1 tells you nothing about t_2.

Joint independence

The random variables $X_1, \cdots, X_n$ are **jointly independent** if the CDF factors into a product of the form

$$F(t_1, \cdots, t_n) = F_{x_1}(t_1) \cdots F_{x_n}(t_n) \ ,$$

where $F_{x_k}(t_k)$ denotes the CDF for the random variable X_k.

When the PDF exists, it factors in the same way:

$$p(t_1, \cdots, t_n) = p_{x_1}(t_1) \cdots p_{x_n}(t_n) \ .$$

- Pairwise independence does *not* imply *joint* independence.

Property: expectation of independent random variables

Expectation of independent random variables

If $X_1, \cdots, X_n$ are a set of jointly independent random variables, then

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

- Under joint independence, expectation and multiplication interchange.
- Linearity: sums split *always*; products split only under independence.

Conditional distributions

The **conditional PDF** of X given $Y = y$ is

$$p(x|y) = \frac{p(x,y)}{p_y(y)} \ .$$

In terms of the joint and marginal CDFs, the densities are

$$\begin{aligned} p(x,y) &= \frac{d^2 F(x,y)}{dx\, dy} \ , \\ p_y(y) &= \frac{dF(\infty,y)}{dy} \ . \end{aligned}$$

This is how the description of X changes when we observe Y. We use it all semester.

Conditional CDF: the formal definition

The **conditional CDF** of X given $Y = y$ is any function $F(x|y)$ that solves

$$F(x, y) = \int_{-\infty}^{y} F(x|t)\, dF_y(t) \ ,$$

where $F(x, y)$ is the **joint CDF** and $F_y(y) = \lim_{x\to\infty} F(x, y)$ is the **marginal CDF**. In most practical cases,

$$F(x|y) = \frac{dF(x, y)}{dy} \left(\frac{dF(\infty, y)}{dy} \right)^{-1} \ .$$

- Radon–Nikodym: a solution exists, not unique. Not required here.

Conditional expectation

The **conditional expectation** of X given Y is

$$\mathbb{E}[X|Y] = \int_{-\infty}^{\infty} x \, dF(x|Y) \ ,$$

or alternatively, using the conditional PDF,

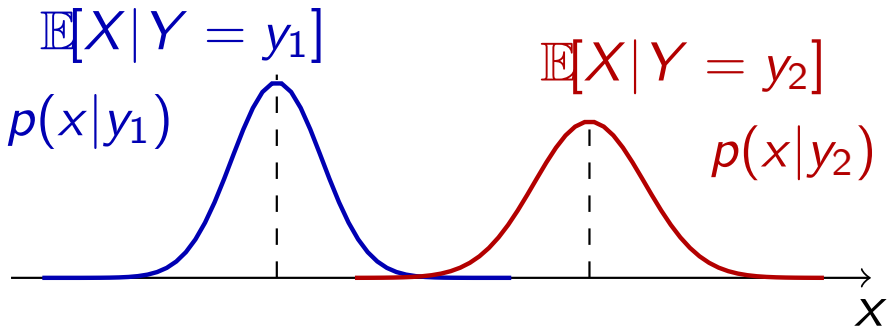
$$\mathbb{E}[X|Y] = \int_{-\infty}^{\infty} x \, p(x|Y) \, dx \ .$$

Conditional expectation is a random variable

The integral on the right is a function of the random variable Y. So the conditional expectation is itself a random variable.

Conditional expectation is a function of Y

Different observations y give different conditional densities, and so different best guesses.



- So $\mathbb{E}[X|Y]$ is a function of Y: observe a value, get a guess.
- Section 2.3: the MMSE estimator is $\mathbb{E}[X|Y]$, the core of Bayes.

Indicator functions and conditional probability

The **indicator function** of a set A is

$$I_A(X) \triangleq \begin{cases} 1 & \text{if } X \in A \ , \\ 0 & \text{otherwise} \end{cases} \ .$$

Using the indicator function, the conditional probability of the event $\{X \in A\}$ given Y is

$$P\{X \in A|Y\} = \mathbb{E}[I_A(X)|Y] \ .$$

- For $A =$ "pixel is in the object," $P\{X \in A|Y\}$ is our belief.
- Conditional probabilities are a special case of conditional expectations.

Property: filtration

In words: if I temporarily give you extra information Z, and you then average over it, you are back where you started.

Filtration property of conditional expectation

For all random variables X, Y, and Z,

$$\mathbb{E}\Big[\,\mathbb{E}[X|Y,Z]\,\Big|Y\Big] = \mathbb{E}[X|Y] \ .$$

- Special case: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$: condition first, then average.
- Another special case, using $P\{X \in A|Y\} = \mathbb{E}[I_A(X)|Y]$:

$$\mathbb{E}\Big[\,P\{X \in A|Y,Z\}\Big|Y\Big] = P\{X \in A|Y\} \ .$$

Property: conditional expectation of known quantities

Conditional expectation of known quantities

For all random variables X, Y, and Z and all functions $f(\cdot)$,

$$\mathbb{E}[f(X)Z|X,Y] = f(X)\,\mathbb{E}[Z|X,Y] \ ,$$

which implies that $\mathbb{E}[X|X,Y] = X$.

- Given X, the value of $f(X)$ is known.
- If the outside temperature is exactly 100 F, its expectation is 100 F.

A corollary

Since $\mathbb{E}[X|Y] = f(Y)$ for some function $f(\cdot)$, the previous property gives

$$\begin{aligned} \mathbb{E}\big[\,\mathbb{E}[X|Y]\,\big|\,Y,Z\big] &= \mathbb{E}[f(Y)|Y,Z] \\ &= f(Y) \\ &= \mathbb{E}[X|Y] \ . \end{aligned}$$

Therefore, for all random variables X, Y, and Z,

$$\mathbb{E}\big[\,\mathbb{E}[X|Y,Z]\,\big|\,Y\big] = \mathbb{E}[X|Y] = \mathbb{E}\big[\,\mathbb{E}[X|Y]\,\big|\,Y,Z\big] \ .$$

- Conditioning on more, then less, equals less, then more.

Summary

- A random variable is a function $X : \Omega \to \mathbb{R}$; $\{X \leq t\}$ is an event.
- The CDF $F(t) = P\{X \leq t\}$ is monotone increasing and right-hand continuous. When a density exists, $p(t) = dF(t)/dt$.
- Expectation is weighted averaging. Sums always split: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- Products split only under joint independence: $\mathbb{E}[\prod_k X_k] = \prod_k \mathbb{E}[X_k]$.
- Observing one random variable gives a new distribution for another: the conditional PDF is $p(x|y) = p(x, y)/p_y(y)$, and $\mathbb{E}[X|Y]$ is a *random variable*, a function of Y. Its two properties are $\mathbb{E}\big[\mathbb{E}[X|Y, Z]\big|Y\big] = \mathbb{E}[X|Y]$ and $\mathbb{E}[f(X)Z|X, Y] = f(X)\mathbb{E}[Z|X, Y]$.

Reading: Chapter 2.1. **Next lecture:** Chapter 2.2, Commonly Used Distributions.