Lecture 3: Commonly Used Distributions

Chapter 2.2

Charles A. Bouman

Purdue University

Friday, August 28, 2026

Outline

1. **Constructing random variables**
   A uniform random variable, then any CDF built from it.
2. **Gaussian models**
   Gaussians, joint Gaussianity, decorrelation, conditional expectation.
3. **Discrete distributions**
   The Bernoulli distribution and the probability mass function.
4. **Data and sufficient statistics**
   i.i.d. data, and compressing it without losing information.

Reading: Chapter 2.2, *Foundations of Computational Imaging*.

Today's Big Ideas

- Any random variable can be built from a uniform one.
- Covariance sets a Gaussian's shape: eigenvectors and eigenvalues.
- Jointly Gaussian is stronger than each component being Gaussian.
- Jointly Gaussian: uncorrelated means independent; $\mathbb{E}[X|Y]$ is linear.
- Sufficient statistics keep all information about the parameters.

Watch for the multivariate Gaussian $N(\mu, R)$. It appears in nearly every lecture from here on.

Example: Your first random variable!

The probability space

Let $\Omega = [0, 1]$, let $\mathcal{B}_{[0,1]}$ be the Borel sets on $[0, 1]$, and let

$$P(A) = \int_0^1 I_A(t)\, dt \qquad \text{for } A \in \mathcal{B}_{[0,1]} \ .$$

So $P(A)$ is the length of the set A.

The random variable

Define $U : \Omega \to \mathbb{R}$ by

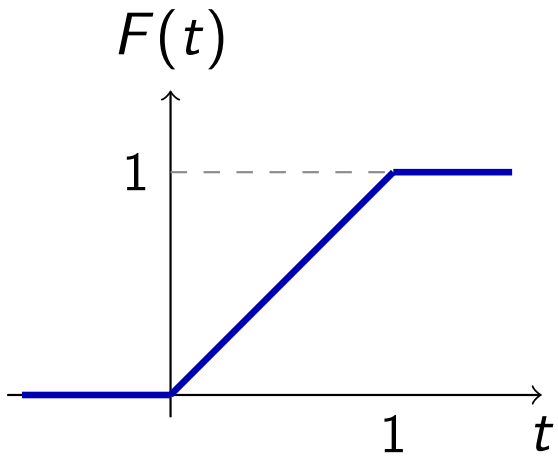
$$U(\omega) = \omega \ .$$

So U is the identity function, and all the randomness is in which outcome ω occurs.

Example: U is uniformly distributed

Compute the CDF of U from the definition. For $0 \leq t \leq 1$,

$$F(t) = P\{U \leq t\} = P\left(\{\omega \in \Omega : \omega \leq t\}\right) = P\left([0,t]\right)$$
$$= \int_0^1 I_{[0,t]}(s)\, ds = t\ .$$

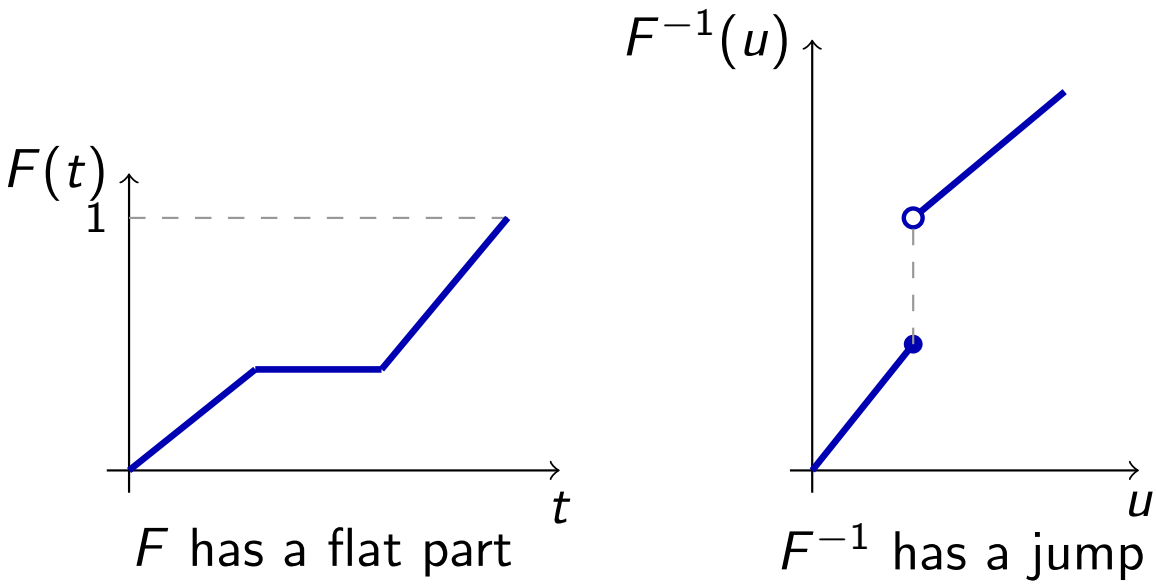


- Outside the interval, $F(t) = 0$ for $t < 0$ and $F(t) = 1$ for $t > 1$.
- This is the CDF of a uniform random variable on $[0, 1]$.

The inverse CDF

The **inverse CDF** of a valid CDF F is

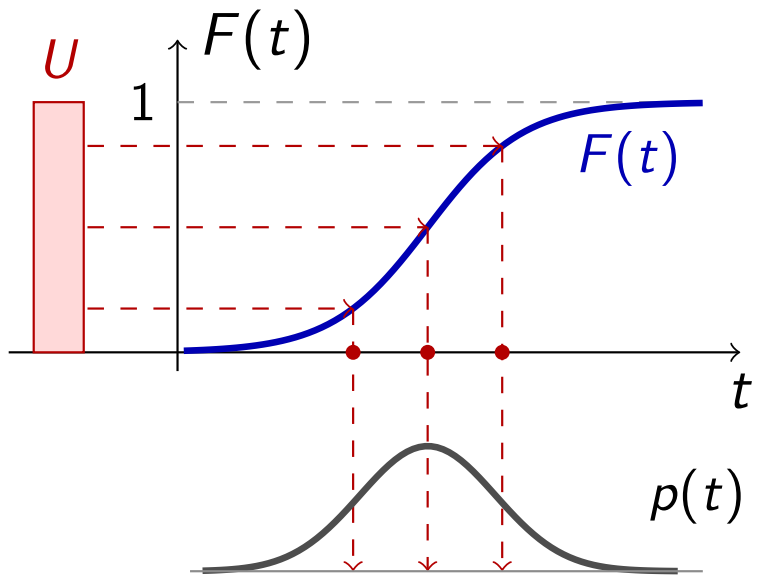
$$F^{-1}(u) \triangleq \inf\{t \in \mathbb{R} : F(t) \geq u\}, \qquad 0 < u < 1.$$



- If F is continuous and strictly increasing, F^{-1} is ordinary.
- Flat parts or jumps: no ordinary inverse. Infimum = smallest t.

Problem 2 of Chapter 2: building any random variable

Let U be uniform on $[0, 1]$ and let F be any valid CDF. Then $X = F^{-1}(U)$ is a random variable with CDF F.



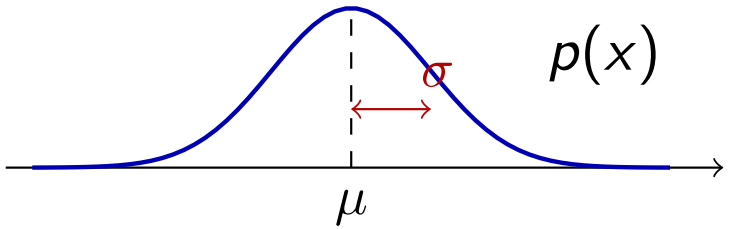
- We can now create a random variable with any distribution.
- Computers draw samples this way: generate U, then apply F^{-1}.

The Gaussian distribution

The **Gaussian distribution** of a random variable X is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \ ,$$

where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.



- $\mu = \mathbb{E}[X]$ is the **mean**; σ^2 the **variance**.
- Notation: $X \sim N(\mu, \sigma^2)$.
- This is our standard model for measurement noise.

Gaussian random vectors

A random column vector $X \in \mathbb{R}^p$ is a **Gaussian random vector** if its density has the form

$$p(x) = \frac{1}{(2\pi)^{p/2}} |R|^{-1/2} \exp\left\{ -\frac{1}{2}(x-\mu)^t R^{-1}(x-\mu) \right\}$$

for some $\mu \in \mathbb{R}^p$ and some symmetric positive definite matrix $R \in \mathbb{R}^{p \times p}$.

- This is the **multivariate Gaussian**; we write $X \sim N(\mu, R)$.
- Also **jointly Gaussian**: a set, when stacked, is Gaussian.
- R must be positive definite, or $p(x)$ does not integrate to 1.

The parameters are the mean and the covariance

Definition

A symmetric matrix R is **positive semidefinite** if for all $y \in \mathbb{R}^p$ such that $y \neq \mathbf{0}$,

$$\|y\|_R^2 \triangleq y^t R y \;\geq\; 0 \;.$$

If the inequality is strict, then R is **positive definite**.

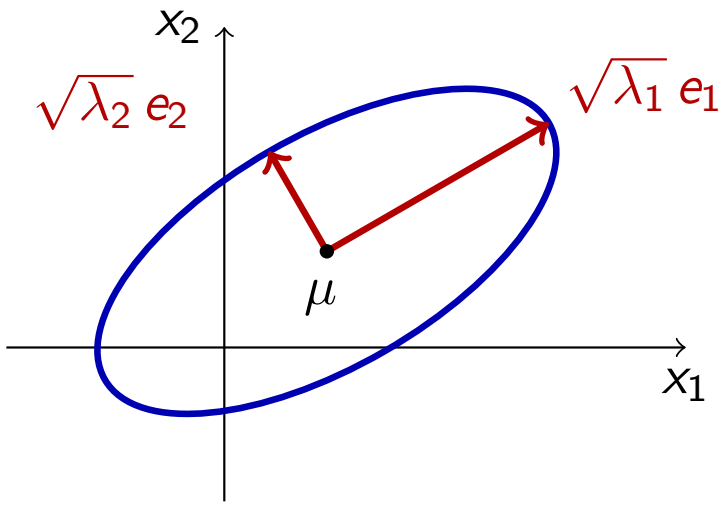
Let $X \sim N(\mu, R)$. It can be shown that

$$\mathbb{E}[X] = \mu \;,$$
$$\mathbb{E}\left[(X-\mu)(X-\mu)^t\right] = R \;.$$

- So μ is the **mean vector** of X, and R is the **covariance matrix** of X.

The covariance matrix sets the shape

A symmetric positive definite matrix can be diagonalized as $R = E\Lambda E^t$. The columns of the orthonormal matrix E are the eigenvectors $e_1, \cdots, e_p$ of R, and $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p)$ holds its strictly positive eigenvalues.



- Contour: ellipse at μ; axes along eigenvectors, half-widths $\sqrt{\lambda_i}$.

Gaussian marginals are not enough

A vector of Gaussian random variables need not be jointly Gaussian. Every component can be Gaussian while the joint density does not have the multivariate Gaussian form.

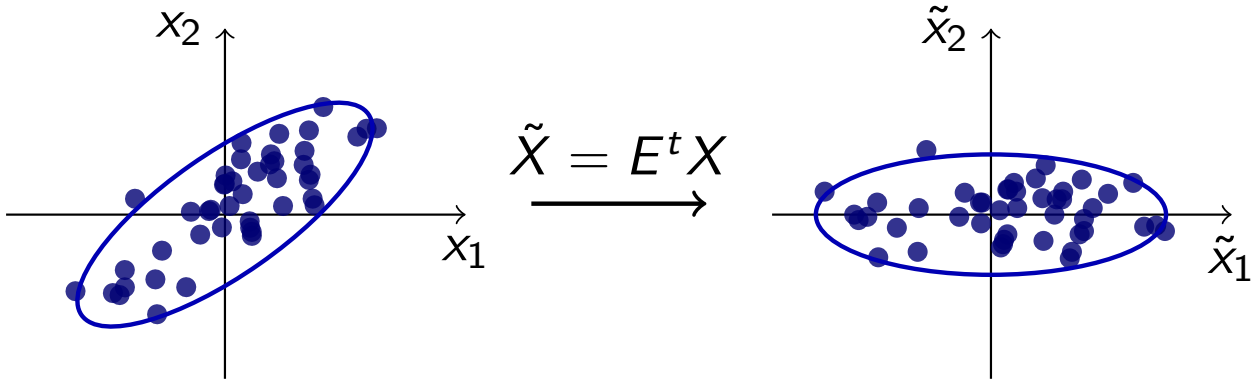
Example

Let $X_1 \sim N(0,1)$, and let S be $+1$ or -1 with equal probability and independent of X_1. Set $X_2 = SX_1$. Then $X_2 \sim N(0,1)$, but $Y = X_1 + X_2$ is zero with probability $1/2$, so Y is not Gaussian.

- So X_1 and X_2 are each Gaussian, and they are **not** jointly Gaussian.
- Uncorrelated ($\mathbb{E}[X_1 X_2] = 0$), but clearly not independent.
- Jointly Gaussian iff every linear transform of it is Gaussian.

Decorrelation gives independence

Let X be zero-mean Gaussian with covariance R, and set $\tilde{X} = E^t X$. Then $\mathbb{E}[\tilde{X}\tilde{X}^t] = \Lambda$.



X: correlated

$\tilde{X}$: uncorrelated

- Λ is diagonal, so the components of $\tilde{X}$ are **uncorrelated**.
- $\tilde{X}$ is jointly Gaussian, so its components are jointly independent.

Property: linearity of conditional expectation

Linearity of conditional expectation for Gaussian random variables

If $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^p$ are jointly Gaussian random vectors, then

$$\mathbb{E}[X|Y] = AY + b \ ,$$
$$\mathbb{E}\left[(X - \mathbb{E}[X|Y])(X - \mathbb{E}[X|Y])^t \,|\, Y\right] = C \ ,$$

where $A \in \mathbb{R}^{n \times p}$ is a matrix, $b \in \mathbb{R}^n$ is a vector, and $C \in \mathbb{R}^{n \times n}$ is a positive-definite matrix.

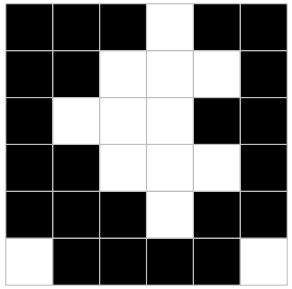
- If X and Y are zero mean, then $b = \mathbf{0}$ and $\mathbb{E}[X|Y]$ is linear in Y.
- $\mathbb{E}[X|Y]$ is the best guess of X; Gaussian models give linear ones.

The Bernoulli distribution

The **Bernoulli distribution** models a binary quantity. If X is Bernoulli, then

$$P\{X = 1\} = \theta \ ,$$
$$P\{X = 0\} = 1 - \theta \ ,$$

where $\theta \in [0, 1]$.



- Example: each pixel of a binary image is object or background.

The probability mass function

A **probability mass function (PMF)** gives the probability of each value of a discrete random variable. A Bernoulli random variable is discrete, so its CDF has jumps, and we use a PMF in place of a PDF.

- For a Bernoulli random variable,

$$p(x) = (1-\theta)^{1-x}\theta^x \ , \qquad x \in \{0,1\} \ .$$

- An equivalent form is

$$p(x) = (1-\theta)\,\delta(x-0) + \theta\,\delta(x-1) \ ,$$

where $\delta(x)$ is one when its argument is zero, and zero otherwise.

Independent and identically distributed data

Random variables $X_1, \cdots, X_n$ are **independent and identically distributed (i.i.d.)** when each one is an independent measurement from the same distribution. Write x for the whole vector of measurements.

Independence alone makes the joint density the product of the densities of the individual measurements,

$$p(x) = p(x_1, \cdots, x_n) = \prod_{k=1}^{n} p_k(x_k) \ .$$

Identical distribution makes every p_k the same function p, so

$$p(x) = \prod_{k=1}^{n} p(x_k) \ .$$

i.i.d. Gaussian measurements

Specialize to n i.i.d. measurements with distribution $N(\mu, \sigma^2)$. Then $p(x_k)$ is the Gaussian density, and

$$\begin{aligned} p(x) &= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_k-\mu)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k-\mu)^2\right\} \quad . \end{aligned}$$

- Products of exponentials became a sum, so squared differences recur.

Example 2.1: n i.i.d. multivariate Gaussian vectors

Let $X_1, \cdots, X_n$ be i.i.d. with distribution $N(\mu, R)$, where $\mu \in \mathbb{R}^p$ and $R \in \mathbb{R}^{p \times p}$. What is the density of the whole observation?

Collect the data in a matrix, $X = [X_1, \cdots, X_n]$. Since the observations are i.i.d.,

$$\begin{aligned} p(x) &= \prod_{k=1}^{n} \frac{1}{(2\pi)^{p/2}} |R|^{-1/2} \exp\left\{ -\frac{1}{2}(x_k - \mu)^t R^{-1}(x_k - \mu) \right\} \\ &= \frac{1}{(2\pi)^{np/2}} |R|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{n} (x_k - \mu)^t R^{-1}(x_k - \mu) \right\} . \end{aligned}$$

Example 2.1: sample statistics

Define two sample statistics,

$$b = \sum_{k=1}^{n} x_k = x\mathbf{1} \ ,$$

$$S = \sum_{k=1}^{n} x_k x_k^t = xx^t \ ,$$

where $\mathbf{1}$ is a column vector of n ones.

- A **statistic** is any function of the data. Useful statistics summarize it.
- b sums the data; S sums their products.

Example 2.1: multivariate Gaussian trace form

With these statistics, the PDF of X becomes

$$p(x) = \frac{1}{(2\pi)^{np/2}} |R|^{-n/2} \exp\left\{ -\frac{1}{2}\text{tr}\left\{SR^{-1}\right\} + b^t R^{-1}\mu - \frac{n}{2}\mu^t R^{-1}\mu \right\} ,$$

where $\text{tr}\{A\}$ is the trace of A.

- This follows from the identity

$$\text{tr}\{AB\} = \text{tr}\{BA\}$$

  for any $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$.
- The data now enter the density *only* through b and S.

Sufficient statistics

Let $T = (b, S)$. Then T is a **sufficient statistic** for the multivariate Gaussian family with parameters $\theta = (\mu, R)$.

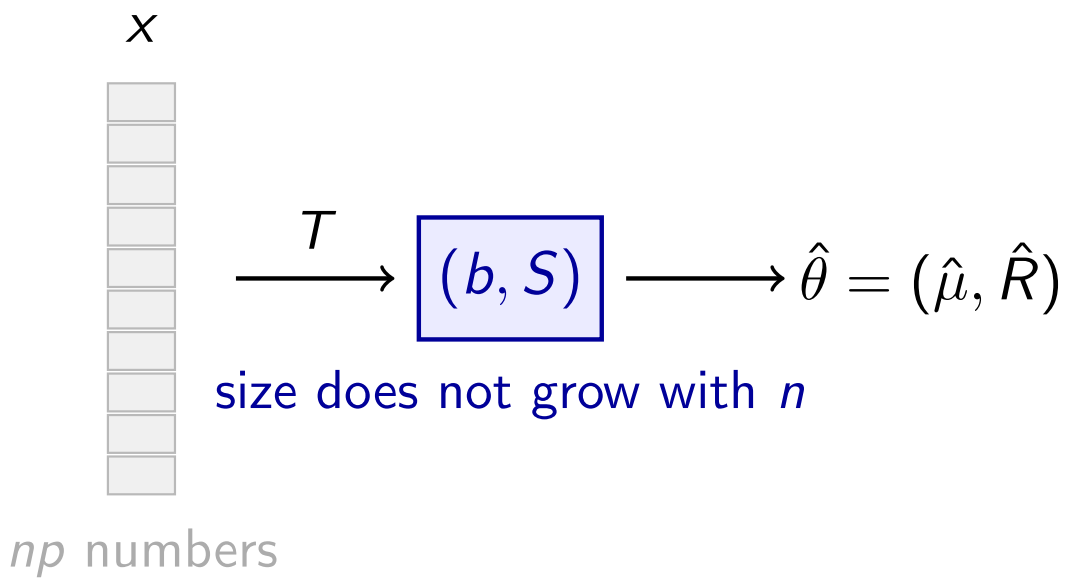
Definition

A statistic T is **sufficient** if the distribution can be written as

$$p(x|\theta) = h(x)\, g\big(T(x), \theta\big)$$

for some functions $h(\cdot)$ and $g(\cdot, \cdot)$.

Why sufficient statistics matter



- Given (b, S), the raw data can go; nothing about θ is lost.
- Compression: np numbers become (b, S), whatever n is.
- Next lecture: ML estimates of (μ, R) from b and S; EM reuses them.

Example 2.2: n i.i.d. Bernoulli random variables

Let $X_1, \cdots, X_n$ be i.i.d. Bernoulli with parameter θ. What is the PMF of the whole observation?

Since the observations are i.i.d.,

$$\begin{aligned} p(x) &= \prod_{k=1}^{n} \left\{ (1-\theta)^{1-x_k} \theta^{x_k} \right\} \\ &= (1-\theta)^{n-N_1} \theta^{N_1} \ , \end{aligned}$$

where

$$N_1 = \sum_{k=1}^{n} x_k$$

counts how many of the X_k equal one.

- N_1 is sufficient: ten bits become one count, $N_1 = 6$, losing nothing.

Summary

- Any random variable: $X = F^{-1}(U)$, with U uniform on $[0, 1]$.
- X is a Gaussian random vector when $p(x) = (2\pi)^{-p/2}|R|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t R^{-1}(x-\mu)\right\}$ for a symmetric positive definite R. Then $\mathbb{E}[X] = \mu$ and $\mathbb{E}[(X-\mu)(X-\mu)^t] = R$.
- $R = E\Lambda E^t$: eigenvectors are the ellipse axes; $\tilde{X} = E^t X$ decorrelates.
- Marginals are not enough; jointly, uncorrelated implies independent.
- Sufficient: $T = (b, S)$ for the Gaussian, $T = N_1$ for the Bernoulli.

Reading: Chapter 2.2 Bayesian estimators.

Next lecture: Chapter 2.3, Frequentist and