Week 1 : Probability essentials

Sébastien Gadat - APM 3F004 EP

5th of September 2025

Notations

- If A and B are subsets of some set Ω, $A \cup B$ denotes their union and $A \cap B$ their intersection. In this book, $\bar{A}$ (rather than CA or A^c) denotes the complement of A in Ω.
- The notation $A + B$ (the sum of A and B) implies by convention that A and B are disjoint, in which case it represents the union $A \cup B$.
- Similarly, the notation $\sum_{k=1}^{\infty} A_k$ is used for $\cup_{k=1}^{\infty} A_k$ only when the A_k 's are pairwise disjoint.
- The notation $A - B$ is used only if $B \subseteq A$, and it stands for $A \cap \bar{B}$. In particular, if $B \subseteq A$, then $A = B + (A - B)$. The symmetric difference of A and B, that is, the set $(A \cup B) - (A \cap B)$, is denoted by $A \triangle B$.

- Random phenomena are observed by means of experiments (performed either by man or nature). Each experiment results in an outcome. The collection of all possible outcomes ω is called the sample space Ω.
- A subset A of the sample space Ω will be regarded as a representation of some event.
  - Examples
    Tossing a die, take 1. The experiment consists in tossing a die once. The possible outcomes are $\omega = 1, 2, \dots, 6$ and the sample space is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$. The subset $A = \{1, 3, 5\}$ is the event "the result is odd."
  - Throwing a dart. The experiment consists in throwing a dart at a wall. The sample space can be chosen to be the plane $\mathbb{R}^2$. An outcome is the position $\omega = (x, y) \in \mathbb{R}^2$ hit by the dart. The subset $A = \left\{(x, y); x^2 + y^2 > 1\right\}$ is an event that could be named "you missed the dartboard" if the dartboard is a closed disk of radius 1 centered at 0 .

Probability theory assigns to each event a number, the probability of the said event. The collection $\mathcal{F}$ of events to which a probability is assigned is not always identical to the collection of all subsets of Ω. The requirement on $\mathcal{F}$ is that it should be a σ-field:

Definition

Let $\mathcal{F}$ be a collection of subsets of Ω, such that

(i) Ω is in $\mathcal{F}$,

(ii) if A belongs to $\mathcal{F}$, then so does its complement $\bar{A}$,

(iii) if $A_1, A_2, \ldots$ belong to $\mathcal{F}$, then so does their union $\cup_{k=1}^{\infty} A_k$.

One then calls $\mathcal{F}$ a σ-field on Ω (here the σ-field of events).

Note that the impossible event $\varnothing$, being the complement of the certain event Ω, is in $\mathcal{F}$. Note also that if $A_1, A_2, \ldots$ belong to $\mathcal{F}$, then so does their intersection $\cap_{k=1}^{\infty} A_k$.

Trivial σ-Field, gross σ-Field

- These are respectively the collection $\mathcal{P}(\Omega)$ of all subsets of Ω and the σ-field with only two sets: $\{\Omega, \varnothing\}$.
- If the sample space Ω is finite or countable, one usually (but not always and not necessarily) considers any subset of Ω to be an event, that is, $\mathcal{F} = \mathcal{P}(\Omega)$.

Borel σ-field

- The Borel σ-field on the n-dimensional euclidean space $\mathbb{R}^n$, denoted $\mathcal{B}(\mathbb{R}^n)$ and called the Borel σ-field on $\mathbb{R}^n$, is, by definition, the smallest σ-field on $\mathbb{R}^n$ that contains all rectangles, that is, all sets of the form $\prod_{j=1}^n I_j$, where the I_j 's are intervals of $\mathbb{R}$.
- This definition is not constructive and therefore one may wonder if there exists sets that are not Borel sets (that is, not sets in $\mathcal{B}(\mathbb{R}^n)$).
- The theory tells us that there are indeed such sets, but they are in a sense "pathological": the proof of existence of non-Borel sets is not constructive, in the sense that it involves the axiom of choice.

The probability $P(A)$ of an event $A \in \mathcal{F}$ measures the likeliness of its occurrence. As a function defined on $\mathcal{F}$, the probability P is required to satisfy a few properties, the axioms of probability.

Definition

A probability on $(\Omega, \mathcal{F})$ is a mapping $P : \mathcal{F} \to \mathbb{R}$ such that

(i) $0 \leq P(A) \leq 1$ for all $A \in \mathcal{F}$,

(ii) $P(\Omega) = 1$, and

(iii) $P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P\left(A_k\right)$ for all sequences $\{A_k\}_{k \geq 1}$ of pairwise disjoint events in $\mathcal{F}$.

Property (iii) is called σ-additivity. The triple $(\Omega, \mathcal{F}, P)$ is called a probability space, or probability model.

Tossing a die

.

- An event A is a subset of $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- The formula

$$P(A) = \frac{|A|}{6}$$

  where $|A|$ is the cardinality of A (the number of elements in A), defines a probability P.

- Choose a probability P such that for any event of the form $A = \{x_1 = a_1, \ldots, x_n = a_n\}$, where $a_1, \ldots, a_n$ are in $\{0, 1\}$,

$$P(A) = \frac{1}{2^n}$$

- Note that this does not define the probability of all events of $\mathcal{F}$.
- But the theory tells us that there exists such a probability satisfying the above requirement and that this probability is unique.

Basic formula 1

Theorem

For any event A

$$P(\bar{A}) = 1 - P(A)$$

and

$$P(\varnothing) = 0.$$

Probability is monotone

Theorem

For any events A and B,

$$A \subseteq B \Longrightarrow P(A) \leq P(B)$$

Probability is subadditive

Theorem

For any sequence $A_1, A_2, \ldots$ *of events,*

$$P\left(\cup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P\left(A_k\right) \tag{1.4}$$

Theorem

Let $\{A_n\}_{n\geq 1}$ *be a non-decreasing sequence of events, that is,* $A_{n+1} \supseteq A_n$ *for all* $n \geq 1$*. Then*

$$P\left(\cup_{n=1}^{\infty} A_n\right) = \lim_{n\uparrow\infty} P\left(A_n\right) \tag{1.5}$$

Independent events

Definition

A family $\{A_n\}_{n\in\mathbb{N}}$ of events is called independent if for any finite set of indices $i_1 < \ldots < i_r$ where $i_j \in \mathbb{N}(1 \leq j \leq r)$,

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}\right) = P\left(A_{i_1}\right) \times P\left(A_{i_2}\right) \times \cdots \times P\left(A_{i_r}\right).$$

Probability distribution

Suppose that X is a random variable defined on a sample space Ω. If we define the event

$$[a \leq X \leq b] = \{\omega \in \Omega : a \leq X(\omega) \leq b\} = A$$

then $P(a \leq X \leq b) = P(X \in A) = P^X(A)$.

Definition

Let X be a random variable (defined on some sample space Ω). The distribution function of X is defined by

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$$

(The distribution function is often referred to as the cumulative distribution function.)

Properties of the cumulative distribution function

- If $x \leq y$ then $F(x) \leq F(y)$. (F is a non-decreasing function.)
- If $y \downarrow x$ then $F(y) \downarrow F(x)$. (F is a right-continuous function although it is not necessarily a continuous function.)
- $\lim_{x\to-\infty} F(x) = 0; \lim_{x\to\infty} F(x) = 1.$

Discrete random variables

Definition

A random variable X is discrete if its range is a finite or countably infinite set. That is, there exists a set $S = \{s_1, s_2, \cdots\}$ such that $P(X \in S) = 1$.

From the definition above, we can deduce that the probability distribution of a discrete random variable is completely determined by specifying $P(X = x)$ for all x.

Definition

The frequency function of a discrete random variable X is defined by

$$f(x) = P(X = x)$$

The frequency function of a discrete random variable is known by many other names: some examples are probability mass function, probability function and density function. We will reserve the term "density function" for continuous random variables.

Given the frequency function $f(x)$, we can determine the distribution function:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

Thus $F(x)$ is a step function with jumps of height $f(x_1), f(x_2), \cdots$ occurring at the points $x_1, x_2, \cdots$. Likewise, we have

$$P(X \in A) = \sum_{x \in A} f(x)$$

in the special case where $A = (-\infty, \infty)$, we obtain

$$1 = P(-\infty < X < \infty) = \sum_{-\infty < x < \infty} f(x)$$

Definition

A random variable X is said to be continuous if its distribution function $F(x)$ is continuous at every real number x.

Equivalently, we can say that X is a continuous random variable if $P(X = x) = 0$ for all real numbers x. Thus we have

$$P(a \leq X \leq b) = P(a < X < b)$$

for $a < b$. The fact that $P(X = x) = 0$ for any x means that we cannot usefully define a frequency function as we did for discrete random variables; however, for many continuous distributions, we can define an analogous function that is useful for probability calculations.

Probability density function

Definition

A continuous random variable X has a probability density function $f(x) \geq 0$ if for $-\infty < a < b < \infty$, we have

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

It is important to note that density functions are not uniquely determined; that is, given a density function f, it is always possible to find another density function f^* such that

$$\int_a^b f(x)dx = \int_a^b f^*(x)dx$$

for all a and b but for which $f^*(x)$ and $f(x)$ differ on a set of Lebesgue measure 0. This non-uniqueness does not pose problems, in general, for probability calculations but can pose subtle problems in certain statistical applications. If the distribution function F is differentiable at x, we can take the density function f to be the derivative of F at $x : f(x) = F'(x)$.

(Normal distribution)

- A random variable X is said to have a Normal distribution with parameters μ and σ^2 $(X \sim N(\mu, \sigma^2))$ if its density is

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

  It is easy to see that $f(x)$ has its maximum value at $x = \mu$ (which is called the mode of the distribution).
- When $\mu = 0$ and $\sigma^2 = 1$, X is said to have a standard Normal distribution; we will denote the distribution function of $X \sim N(0,1)$ by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-t^2/2\right) dt$$

- If $X \sim N(\mu, \sigma^2)$ then its distribution function is

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

A useful result

Theorem

Suppose that X is a continuous random variable with distribution function $F(x)$ and let $U = F(X)$. Then $U \sim \mathrm{Unif}(0,1)$.

Quantile function

The quantile function is given by

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad \forall p \in [0, 1]$$

1. F^{-1} is nondecreasing
2. $F^{-1}(F(x)) \leq x$
3. $F\left(F^{-1}(p)\right) \geq p$
4. $F^{-1}(p) \leq x$ if and only if $p \leq F(x)$

Theorem

Let F be any distribution function and define $F^{-1}(t) = \inf\{x : F(x) \geq t\}$ to be its inverse function for $0 < t < 1$. If $U \sim \mathrm{Unif}(0,1)$ and $X = F^{-1}(U)$ then the distribution function of X is F.

Sampling an exponential distribution

Suppose that X is a continuous random variable with density

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

X is said to have an Exponential distribution with parameter λ ($X \sim \mathrm{Exp}(\lambda)$). The distribution function of X is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \exp(-\lambda x) & \text{for } x \geq 0 \end{cases}$$

Since $F(x)$ is strictly increasing over the set where $F(x) > 0$, we can determine the inverse $F^{-1}(t)$ by solving the equation

$$1 - \exp\left(-\lambda F^{-1}(t)\right) = t$$

which yields

$$F^{-1}(t) = -\frac{1}{\lambda} \ln(1 - t)$$

Thus if $U \sim \mathrm{Unif}(0, 1)$ then $Y = -\lambda^{-1} \ln(1 - U) \sim \mathrm{Exp}(\lambda)$. Note that since $1 - U$ has the same distribution as U, we also have $-\lambda^{-1} \ln(U) \sim \mathrm{Exp}(\lambda)$.

Sampling a binomial distribution

Suppose that X has a Binomial distribution with parameters $n = 3$ and $\theta = 1/2$. We then have $F(x) = 1/8$ for $0 \leq x < 1, F(x) = 1/2$ for $1 \leq x < 2$ and $F(x) = 7/8$ for $2 \leq x < 3$ ($F(x) = 0$ and $F(x) = 1$ for $x < 0$ and $x \geq 3$ respectively). The inverse of F is given by

$$F^{-1}(t) = \begin{cases} 0 & \text{for } 0 < t \leq 1/8 \\ 1 & \text{for } 1/8 < t \leq 1/2 \\ 2 & \text{for } 1/2 < t \leq 7/8 \\ 3 & \text{for } 7/8 < t < 1 \end{cases}$$

If $U \sim \mathrm{Unif}(0,1)$ then it is simple to see that $F^{-1}(U) \sim \mathrm{Bin}(3, 0.5)$.

Change of variable formula - monotonic case

- Let X be a continuous random variable with p.d.f f_X.
- Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonic function.
- Let $Y = g(X)$ is a continuous random variable with p.d.f f_Y

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}\left(g^{-1}(y)\right)\right|$$

where g^{-1} denotes the inverse function.

Change of variable formula - nonmonotonic case

Consider the quadratic transformation $Y = X^2$ for $-\infty < x < \infty$:
We can begin with an arbitrary cdf of Y:

$$\begin{aligned} F_{\mathrm{Y}}(y) = P(Y \leq y) = P\left(X^2 \leq y\right) &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_{\mathbf{X}}(\sqrt{y}) - F_{\mathbf{X}}(-\sqrt{y}) \end{aligned}$$

So if we wanted the pdf of the random variable Y then naturally we take the derivative of the cdf :

$$\begin{aligned} f_{\mathrm{Y}}(y) &= F'_{\mathrm{X}}(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - F'_{\mathrm{X}}(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \frac{1}{2\sqrt{y}}\left[f_{\mathrm{X}}(\sqrt{y}) + f_{\mathrm{X}}(-\sqrt{y})\right], \quad y > 0 \end{aligned}$$

Theorem

Let X *have pdf* $f_X(x)$*, let* $Y = g(X)$*. Suppose there exists a partition,* $A_0, A_1, \ldots, A_K$*, of X such that* $P(X \in A_0) = 0$*. Further, suppose there exist functions* $g_1(x), g_2(x), \ldots, g_K(x)$*, defined on* $A_0, A_1, \ldots, A_K$*, respectively, satisfying:*

1. $g(x) = \mathrm{g_i(x)}$*, for* $x \in A_\mathrm{i}$*,*
2. $\mathrm{g_i(x)}$ *is monotone on* $\mathrm{A_i}$*,*
3. *The set* $Y = \{y : y = g_\mathrm{i}(x) \exists x \in A_\mathrm{i}\}$ *is the same for each* $\mathrm{i} = 1, \ldots, \mathrm{k}$*, and*
4. $\mathrm{g_i}^{-1}(\mathrm{y})$ *has a continuous derivative on Y , for each* $\mathrm{i} = 1, \ldots, \mathrm{k}$*.*

Then,

$$f_\mathrm{Y}(y) = \sum_{i=1}^{k} f_\mathrm{X}\left(g_\mathrm{i}^{-1}(y)\right) \left|\frac{d}{dy} g_\mathrm{i}^{-1}(y)\right| y \in Y$$
$$= 0 \textit{ otherwise}$$

Quadratric transform of a Gaussian random variable

Consider $\mathrm{Y} = \mathrm{X}^2$. The function $\mathrm{g(x)} = \mathrm{x}^2$ is monotone on $(-\infty, 0)$ and $(0, \infty)$. We should also note that the set $Y = (0, \infty)$.
We can partition X such that:

$$
\begin{aligned}
&A_0 = \{0\} \\
&A_1 = (-\infty, 0), \quad g_1(x) = x^2, \quad g_1^{-1}(y) = -\sqrt{y} \\
&A_2 = (0, \infty), \quad g_2(x) = x^2, \quad g_2^{-1}(y) = \sqrt{y}
\end{aligned}
$$

Thus if we employ monotone transformations on the A_1 and A_2 intervals then we can solve for the pdf of the random variable Y.

$$
\begin{aligned}
f_{\mathrm{Y}}(y) &= \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| \\
&= \frac{1}{\sqrt{2\pi y}} e^{-\mathrm{y}/2}, \quad 0 < y < \infty
\end{aligned}
$$

Definition

Suppose that X is a nonnegative random variable with distribution function F. The expected value or mean of X (denoted by $E(X)$) is defined to be

$$E(X)=\int_0^{\infty}(1-F(x))dx$$

which may be infinite. In general, if $X=X^{+}-X^{-}$, we define $E(X)=E\left(X^{+}\right)-E\left(X^{-}\right)$provided that at least one of $E\left(X^{+}\right)$ and $E\left(X^{-}\right)$is finite; if both are infinite then $E(X)$ is undefined. If $E(X)$ is well-defined then

$$E(X)=\int_0^{\infty}(1-F(x))dx-\int_{-\infty}^{0}F(x)dx$$

If X is a continuous random variable with density function $f(x)$, we have

$$\begin{aligned}
E(X) &= \int_0^{\infty}(1-F(x))dx - \int_{-\infty}^{0} F(x)dx \\
&= \int_0^{\infty}\int_x^{\infty} f(t)dtdx - \int_{-\infty}^{0}\int_{-\infty}^{x} f(t)dtdx \\
&= \int_0^{\infty}\int_0^{t} f(t)dxdt - \int_{-\infty}^{0}\int_{t}^{0} f(t)dxdt \\
&= \int_0^{\infty} tf(t)dt + \int_{-\infty}^{0} tf(t)dt \\
&= \int_{-\infty}^{\infty} tf(t)dt
\end{aligned}$$

Expectation for "mixed" random variable

Theorem

Suppose that X has a distribution function F with $F(x) = pF_c(x) + (1-p)F_d(x)$ where F_c is a continuous distribution function with density function f_c and F_d is a discrete distribution function with frequency function f_d. If $E(X)$ is welldefined then

$$E(X) = p\int_{-\infty}^{\infty} x f_c(x)dx + (1-p)\sum_x x f_d(x)$$

Expectation for the Gamma distribution

Suppose that X has the density function

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\lambda x) \quad \text{for } x > 0$$

where $\lambda > 0, \alpha > 0$ and $\Gamma(\alpha)$ is the Gamma function defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt$$

X is said to a Gamma distribution with shape parameter α and scale parameter $\lambda (X \sim \mathrm{Gamma}(\alpha, \lambda))$. Using properties of the function $\Gamma(\alpha)$, we have

$$\begin{aligned} E(X) &= \int_{-\infty}^\infty x f(x) dx \\ &= \int_0^\infty \frac{\lambda^\alpha x^\alpha}{\Gamma(\alpha)} \exp(-\lambda x) dx \\ &= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+1} x^{\alpha+1-1}}{\Gamma(\alpha+1)} \exp(-\lambda x) dx \\ &= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \\ &= \frac{\alpha}{\lambda} \end{aligned}$$

Joint Distribution of a random vector

Definition

The joint distribution function of a random vector $(X_1, \cdots, X_k)$ is

$$F(x_1, \cdots, x_k) = P(X_1 \leq x_1, \cdots, X_k \leq x_k)$$

where the event $[X_1 \leq x_1, \cdots, X_k \leq x_k]$ is the intersection of the events $[X_1 \leq x_1], \cdots, [X_k \leq x_k]$.
Given the joint distribution function of random vector $\boldsymbol{X}$, we can determine $P(\boldsymbol{X} \in A)$ for any (Borel) set $A \subset \mathbb{R}^k$.

Joint Frequency Function of a discrete random vector

Definition

Suppose that $X_1, \cdots, X_k$ are discrete random variables defined on the same sample space. Then the joint frequency function of $X = (X_1, \cdots, X_k)$ is defined to be

$$f(x_1, \cdots, x_k) = P(X_1 = x_1, \cdots, X_k = x_k)$$

Joint Density function of a random vector

Definition

Suppose that $X_1, \cdots, X_n$ are continuous random variables defined on the same sample space and that

$$P\left[X_1 \leq x_1, \cdots, X_k \leq x_k\right] = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f\left(t_1, \cdots, t_k\right) dt_1 \cdots dt_k$$

for all $x_1, \cdots, x_k$. Then $f\left(x_1, \cdots, x_k\right)$ is the joint density function of $\left(X_1, \cdots, X_k\right)$ (provided that $f\left(x_1, \cdots, x_k\right) \geq 0$).

Marginal distributions

Theorem

(a) *Suppose that* $\boldsymbol{X} = (X_1, \cdots, X_k)$ *has joint frequency function* $f(\boldsymbol{x})$. *For* $\ell < k$, *the joint frequency function of* $(X_1, \cdots, X_\ell)$ *is*

$$g(x_1, \cdots, x_\ell) = \sum_{x_{\ell+1}, \cdots, x_k} f(x_1, \cdots, x_k)$$

(b) *Suppose that* $\boldsymbol{X} = (X_1, \cdots, X_k)$ *has joint density function* $f(\boldsymbol{x})$. *For* $\ell < k$, *the joint density function of* $(X_1, \cdots, X_\ell)$ *is*

$$g(x_1, \cdots, x_\ell) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_k)\, dx_{\ell+1} \cdots dx_k$$

Uniform distribution on a disk

Suppose that X and Y are continuous random variables with joint density function

$$f(x,y) = \frac{1}{\pi} \quad \text{for } x^2 + y^2 \leq 1$$

X and Y thus have a Uniform distribution on a disk of radius 1 centered at the origin.

- Determine $P(X \leq u)$ for $-1 \leq u \leq 1$.
- Determine the probability density function (pdf) of X

Independent random variables

Definition

Let $X_1, \cdots, X_k$ be random variables defined on the same sample space. $X_1, \cdots, X_k$ are said to be independent if the events
$[a_1 < X_1 \leq b_1], [a_2 < X_2 \leq b_2], \cdots, [a_k < X_k \leq b_k]$ are independent for all $a_i < b_i, i = 1, \cdots, k$.
An infinite collection $X_1, X_2, \cdots$ of random variables are independent if every finite collection of random variables is independent.

Joint density of independent random variables

Theorem

If $X_1, \cdots, X_k$ are independent and have joint density (or frequency) function $f(x_1, \cdots, x_k)$ then

$$f(x_1, \cdots, x_k) = \prod_{i=1}^{k} f_i(x_i)$$

where $f_i(x_i)$ is the marginal density (frequency) function of X_i. Conversely, if the joint density (frequency) function is the product of marginal density (frequency) functions then $X_1, \cdots, X_k$ are independent.

Minimum and Maximum of Uniform random variables

Suppose that $X_1, \cdots, X_n$ are i.i.d. continuous random variables with common (marginal) density $f(x)$ and distribution function $F(x)$. Given $X_1, \cdots, X_n$, we can define two new random variables

$$U = \min(X_1, \cdots, X_n) \quad \text{and} \quad V = \max(X_1, \cdots, X_n)$$

(a) Determine the marginal densities of U and V.

(b) Determine the joint density of (U, V)