ESSENTIALS OF PROBABILITY THEORY

Basic notions

Sample space $\mathcal{S}$

$\mathcal{S}$ is the set of all possible outcomes e of an experiment.

Example 1. In tossing of a die we have $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

Example 2. The life-time of a bulb $\mathcal{S} = \{x \in \mathcal{R} \mid x > 0\}$.

Event

An event is a subset of the sample space $\mathcal{S}$. An event is usually denoted by a capital letter $A,\ B, \ldots$

If the outcome of an experiment is a member of event A, we say that A has occurred.

Example 1. The outcome of tossing a die is an even number: $A = \{2, 4, 6\} \subset \mathcal{S}$.

Example 2. The life-time of a bulb is at least 3000 h: $A = \{x \in \mathcal{R} \mid x > 3000\} \subset \mathcal{S}$.

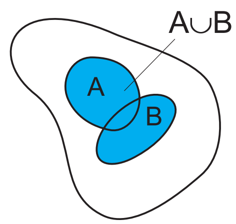
Certain event: The whole sample space $\mathcal{S}$.

Impossible event: Empty subset ϕ of $\mathcal{S}$.

Combining events

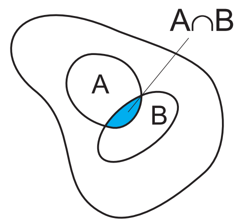
Union "A or B".

$$A \cup B = \{e \in \mathcal{S} \mid e \in A \text{ or } e \in B\}$$



Intersection (joint event) "A and B".

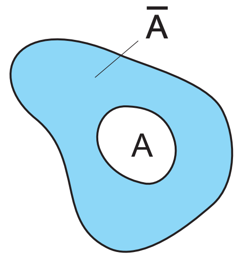
$$A \cap B = \{e \in \mathcal{S} \mid e \in A \text{ and } e \in B\}$$



Events A and B are mutually exclusive, if $A \cap B = \phi$.

Complement "not A".

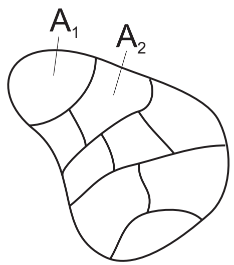
$$\bar{A} = \{e \in \mathcal{S} \mid e \notin A\}$$



Partition of the sample space

A set of events $A_1, A_2, \ldots$ is a partition of the sample space $\mathcal{S}$ if

1. The events are mutually exclusive, $A_i \cap A_j = \phi$, kun $i \neq j$.
2. Together they cover the whole sample space, $\cup_i A_i = \mathcal{S}$.



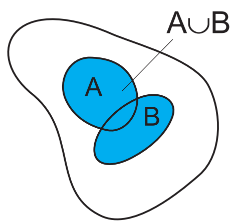
Probability

With each event A is associated the probability $\mathrm{P}\{A\}$.

Empirically, the probability $\mathrm{P}\{A\}$ means the limiting value of the relative frequency $N(A)/N$ with which A occurs in a repeated experiment

$$\mathrm{P}\{A\} = \lim_{N\to\infty} N(A)/N \qquad \begin{cases} N & = \text{ number of experiments} \\ N(A) & = \text{ number of occurrences of } A \end{cases}$$

Properties of probability

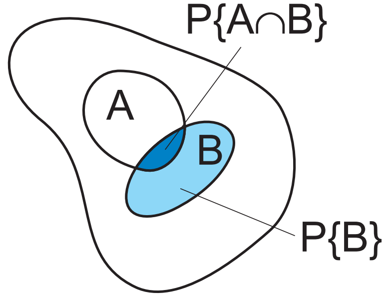


1. $0 \leq \mathrm{P}\{A\} \leq 1$
2. $\mathrm{P}\{\mathcal{S}\} = 1 \qquad \mathrm{P}\{\phi\} = 0$
3. $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\} - \mathrm{P}\{A \cap B\}$
4. If $A \cap B = 0$, then $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\}$
   If $A_i \cap A_j = 0$ for $i \neq j$, then $\mathrm{P}\{\cup_i A_i\} = \mathrm{P}\{A_1 \cup \ldots \cup A_n\} = \mathrm{P}\{A_1\} + \ldots \mathrm{P}\{A_n\}$
5. $\mathrm{P}\{\bar{A}\} = 1 - \mathrm{P}\{A\}$
6. If $A \subseteq B$, then $\mathrm{P}\{A\} \leq \mathrm{P}\{B\}$

Conditional probability

The probability of event A given that B has occurred.

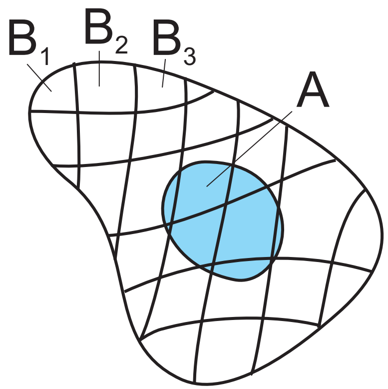
$$\boxed{\mathrm{P}\{A\,|\,B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}}} \qquad \Rightarrow \mathrm{P}\{A \cap B\} = \mathrm{P}\{A\,|\,B\}\mathrm{P}\{B\}$$



Law of total probability

Let $\{B_1, \ldots, B_n\}$ be a complete set of mutually exclusive events, i.e. a <u>partition</u> of the sample space $\mathcal{S}$,

$$
\begin{array}{llll}
1. & \cup_i B_i = \mathcal{S} & \text{certain event} & \mathrm{P}\{\cup_i B_i\} = 1 \\
2. & B_i \cap B_j = \phi & \text{for } i \neq j & \mathrm{P}\{B_i \cap B_j\} = 0
\end{array}
$$



Then $A = A \cap \mathcal{S} = A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$ and

$$\boxed{\mathrm{P}\{A\} = \sum_{i=1}^{n} \mathrm{P}\{A \cap B_i\} = \sum_{i=1}^{n} \mathrm{P}\{A \,|\, B_i\}\mathrm{P}\{B_i\}}$$

Calculation of the probability of event A by conditioning on the events B_i. Typically the events B_i represent all the possible outcomes of an experiment.

Bayes' formula

Let again $\{B_1, \ldots, B_n\}$ be a partition of the sample space.

The problem is to calculate the probability of event B_i given that A has occurred.

$$\boxed{\mathrm{P}\{B_i \,|\, A\} = \frac{\mathrm{P}\{A \cap B_i\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{A \,|\, B_i\}\mathrm{P}\{B_i\}}{\sum_j \mathrm{P}\{A \,|\, B_j\}\mathrm{P}\{B_j\}}}$$

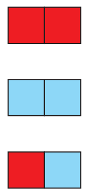
Bayes' formula enables us to calculate a conditional probability when we know the reverse conditional probabilities.

Example: three cards with different colours on different sides.

rr: both sides are red

bb: both sides are blue

rb: one side red, the other one blue



The upper side of a randomly drawn card is red. What is the probability that the other side is blue?

$$\begin{aligned}
\mathrm{P}\{\mathrm{rb} \,|\, \mathrm{red}\} &= \frac{\mathrm{P}\{\mathrm{red} \,|\, \mathrm{rb}\}\mathrm{P}\{\mathrm{rb}\}}{\mathrm{P}\{\mathrm{red} \,|\, \mathrm{rr}\}\mathrm{P}\{\mathrm{rr}\} + \mathrm{P}\{\mathrm{red} \,|\, \mathrm{bb}\}\mathrm{P}\{\mathrm{bb}\} + \mathrm{P}\{\mathrm{red} \,|\, \mathrm{rb}\}\mathrm{P}\{\mathrm{rb}\}} \\
&= \frac{\frac{1}{2}\times\frac{1}{3}}{1\times\frac{1}{3}+0\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}} = \frac{1}{3}
\end{aligned}$$

Independence

Two events A and B are <u>independent</u> if and only if

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\} \cdot \mathrm{P}\{B\}$$

For independent events holds

$$\mathrm{P}\{A\,|\,B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}} = \frac{\mathrm{P}\{A\}\mathrm{P}\{B\}}{\mathrm{P}\{B\}} = \mathrm{P}\{A\} \quad \text{“}B \text{ does not influence occurrence of } A\text{”}.$$

<u>Example 1:</u> Tossing two dice, $A = \{n_1 = 6\},\ B = \{n_2 = 1\}$

$$A \cap B = \{(6,1)\}, \quad \mathrm{P}\{A \cap B\} = \tfrac{1}{36}, \qquad \text{all combinations equally probable}$$
$$\mathrm{P}\{A\} = \mathrm{P}\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} = \tfrac{6}{36} = \tfrac{1}{6}; \quad \text{similarly } \mathrm{P}\{B\} = \tfrac{1}{6}$$
$$\mathrm{P}\{A\}\mathrm{P}\{B\} = \tfrac{1}{36} = \mathrm{P}\{A \cap B\} \quad \Rightarrow \quad \text{independent}$$

<u>Example 2:</u> $A = \{n_1 = 6\},\ B = \{n_1 + n_2 = 9\} = \{(3,6),(4,5),(5,4),(6,3)\}$

$$A \cap B = \{(6,2)\}$$
$$\mathrm{P}\{A\} = \tfrac{1}{6}, \quad \mathrm{P}\{B\} = \tfrac{4}{36}, \quad \mathrm{P}\{A \cap B\} = \tfrac{1}{36}$$
$$\mathrm{P}\{A\} \cdot \mathrm{P}\{B\} \neq \mathrm{P}\{A \cap B\} \quad \Rightarrow \quad A \text{ and } B \text{ dependent}$$

Probability theory: summary

- Important in modelling phenomena in real world
  - e.g. telecommunication systems
- Probability theory has a natural, intuitive interpretation and simple mathematical axioms
- Law of total probability enables one to decompose the problem into subproblems
  - analytical approach
  - a central tool in stochastic modelling
- The probability of the joint event of independent events is the product of the probabilities of the individual events

Random variables and distributions

Random variable

We are often more interested in a some number associated with the experiment rather than the outcome itself.

Example 1. The number of heads in tossing coin rather than the sequence of heads/tails

> A real-valued random variable X is a mapping
> $X: \quad \mathcal{S} \mapsto \mathcal{R}$
> which associates the real number $X(e)$ to each outcome $e \in \mathcal{S}$.

Example 2. The number of heads in three consecutive tossings of a coin (head = `h`, tail=`t` (tail))

e	$X(e)$
hhh	3
hht	2
hth	2
htt	1
thh	2
tht	1
tth	1
ttt	0

- The values of X are "drawn" by "drawing" e
- e represents a "lottery ticket", on which the value of X is written

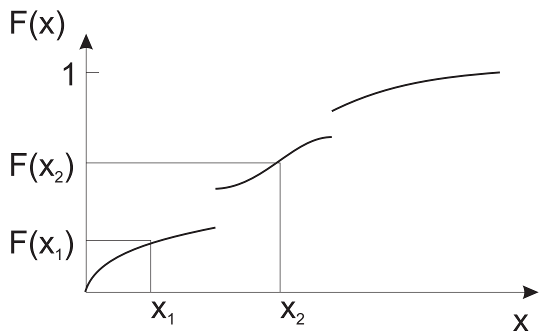
The image of a random variable X

$$\boxed{\mathcal{S}_X = \{x \in \mathcal{R} \mid X(e) = x,\ e \in \mathcal{S}\}}$$ (complete set of values X can take)

- may be finite or countably infinite: <u>discrete random variable</u>
- uncountably infinite: <u>continuous random variable</u>

Distribution function (cdf, cumulative distribution function)

$$\boxed{F(x) = \mathrm{P}\{X \le x\}}$$

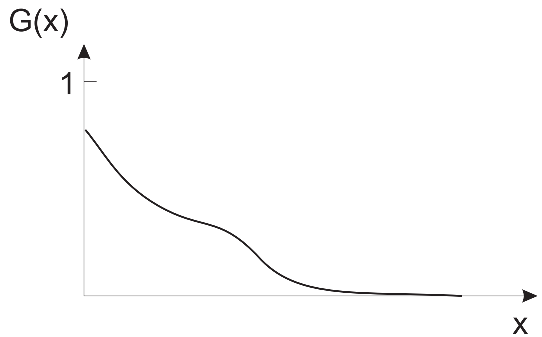


The probability of an interval

$$\boxed{\mathrm{P}\{x_1 \le X \le x_2\} = F(x_2) - F(x_1)}$$

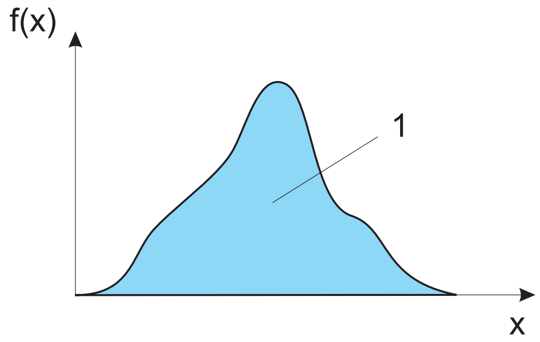
Complementary distribution function (tail distribution)

$$\boxed{G(x) = 1 - F(x) = \mathrm{P}\{X > x\}}$$



Continuous random variable: probability density function (pdf)

$$\boxed{f(x) = \frac{dF(x)}{dx} = \lim_{dx\to 0} \frac{\mathrm{P}\{x < X \le x + dx\}}{dx}}$$



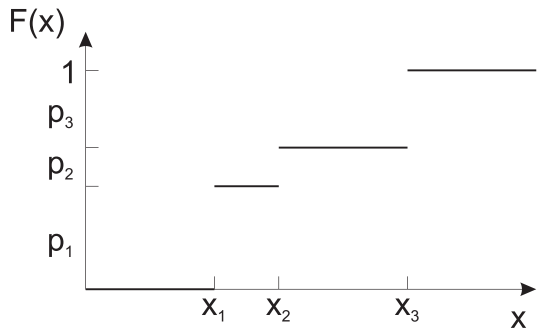
Discrete random variable

The set of values a discrete random variable X can take is either finite or countably infinite, $X \in \{x_1, x_2, \ldots\}$.

With these are associated the point probabilities

$$p_i = \mathrm{P}\{X = x_i\}$$

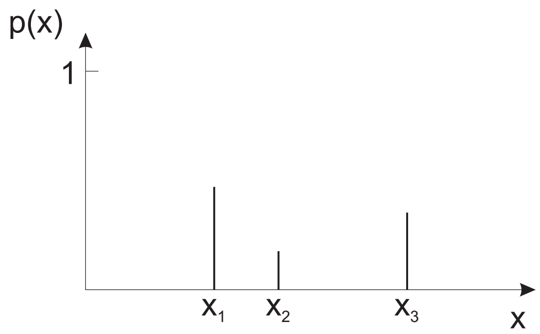
which define the discrete distribution



The distribution function is a step function, which has jumps of height p_i at points x_i.

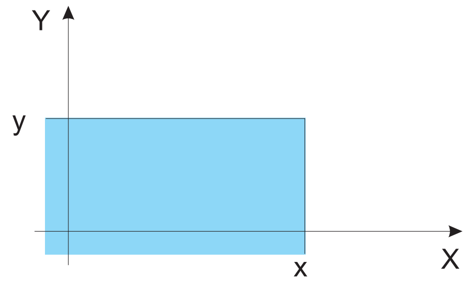
Probability mass function (pmf)

$$p(x) = \mathrm{P}\{X = x\} = \begin{cases} p_i & \text{when } x = x_i \\ 0, & \text{otherwise} \end{cases}$$



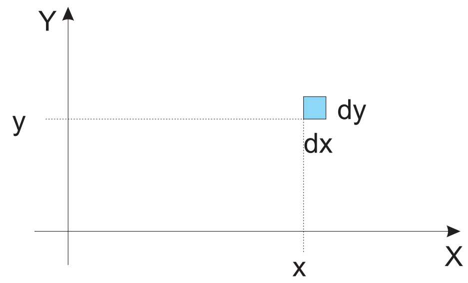
Joint random variables and their distributions

Joint distribution function



$$\boxed{F_{X,Y}(x,y) = \mathrm{P}\{X \le x, Y \le y\}}$$

Joint probability density function



$$\boxed{f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)}$$

The above definitions can be generalized in a natural way for several random variables.

Independence

The random variables X and Y are independent if and only if the events $\{X \le x\}$ and $\{Y \le y\}$ are independent, whence

$$\boxed{F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)} \qquad \boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$$

(the conditions are equivalent)

Function of a random variable

Let X be a (real-valued) random variable and $g(\cdot)$ a function (g: $\mathcal{R} \mapsto \mathcal{R}$). By applying the function g on the values of X we get another random variable $Y = g(X)$.

$$\boxed{F_Y(y) = F_X(g^{-1}(y))} \qquad \text{since} \qquad Y \le y \;\Leftrightarrow\; g(X) \le y \;\Leftrightarrow\; X \le g^{-1}(y)$$

Specifically, if we take $g(\cdot) = F_X(\cdot)$ (image [0,1]), then

$$F_Y(y) = F_X(F_X^{-1}(y)) = y$$

and the pdf of Y is $f_Y(y) = \frac{d}{dy}F_Y(y) = 1$, i.e. Y obeys the uniform distribution in the interval (0,1).

$$\boxed{F_X(X) \sim U} \qquad \boxed{X \sim F_X^{-1}(U)} \qquad \sim \text{ means "identically distributed"}$$

This enables one to draw values for an arbitrary random variable X (with distribution function $F_X(x)$), e.g. in simulations, if one has at disposal a random number generator which produces values of a random variable U uniformly distributed in (0,1).

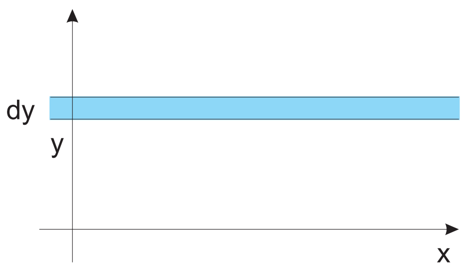
The pdf of a conditional distribution

Let X and Y be two random variables (in general, dependent). Consider the variable X conditioned on that Y has taken a given value y. Denote this conditioned random variable by $X_{|Y=y}$.

The conditional pdf is denoted by $f_{X_{|Y=y}} = f_{X|Y}(x, y)$ and defined by

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

where the marginal distribution of Y is $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$



the distribution is limited in the strip $Y \in (y, y + dy)$

$f_{X,Y}(x, y)dydx$ is the probability of the element $dxdy$ in the strip

$f_Y(y)dy$ is the total probability mass of the strip

If X and Y are independent, then $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ and $f_{X|Y}(x, y) = f_X(x)$, i.e. the conditioning does not affect the distribution.

Parameters of distributions

Expectation

Denoted by $\mathrm{E}[X] = \bar{X}$

Continuous distribution: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx}$

Discrete distribution: $\boxed{\mathrm{E}[X] = \sum_i x_i p_i}$

In general: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, dF(x)}$ $dF(x)$ is the probability of the interval dx

Properties of expectation

$\mathrm{E}[cX] = c\mathrm{E}[X]$	c constant
$\mathrm{E}[X_1 + \cdots X_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]$	always
$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$	only when X and Y are independent

Variance

Denoted by V[X] (also Var[X])

$$\boxed{\mathrm{V}[X] = \mathrm{E}[(X - \bar{X})^2]} = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Covariance

Denoted by Cov[X, Y]

$$\boxed{\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \bar{X})(Y - \bar{Y})]} = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

$\mathrm{Cov}[X, X] = \mathrm{V}[X]$

If X are Y independent then $\mathrm{Cov}[X, Y] = 0$

Properties of variance

$\mathrm{V}[cX] = c^2\mathrm{V}[X]$	c constant; observe square
$\mathrm{V}[X_1 + \cdots X_n] = \sum_{i,j=1}^{n} \mathrm{Cov}[X_i, X_j]$	always
$\mathrm{V}[X_1 + \cdots X_n] = \mathrm{V}[X_1] + \cdots + \mathrm{V}[X_n]$	<u>only when the X_i are independent</u>

Properties of covariance

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$$

$$\mathrm{Cov}[X + Y, Z] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

Conditional expectation

The expectation of the random variable X given that another random variable Y takes the value $Y = y$ is

$$\boxed{\mathrm{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx}$$ obtained by using the conditional distribution of X.

$\mathrm{E}[X \mid Y = y]$ is a function of y. By applying this function on the value of the random variable Y one obtains a random variable $\mathrm{E}[X \mid Y]$ (a function of the random variable Y).

Properties of conditional expectation

$\mathrm{E}[X \mid Y] = \mathrm{E}[X]$	if X and Y are independent
$\mathrm{E}[cX \mid Y] = c\,\mathrm{E}[X \mid Y]$	c is constant
$\mathrm{E}[X + Y \mid Z] = \mathrm{E}[X \mid Z] + \mathrm{E}[Y \mid Z]$	

Conditional variance

$$\boxed{\mathrm{V}[X\,|\,Y] = \mathrm{E}\left[(X - \mathrm{E}\,[X\,|\,Y])^2\,|\,Y\right]}$$

Deviation with respect to the conditional expectation

Conditional covariance

$$\boxed{\mathrm{Cov}[\,X, Y\,|\,Z] = \mathrm{E}\,[(X - \mathrm{E}\,[X\,|\,Z])(Y - \mathrm{E}\,[Y\,|\,Z])\,|\,Z]}$$

Conditioning rules

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}\,[X\,|\,Y]] \quad \text{(inner conditional expectation is a function of } Y\text{)}$$

$$\mathrm{V}[X] = \mathrm{E}[\mathrm{V}[X\,|\,Y]] + \mathrm{V}[\mathrm{E}\,[X\,|\,Y]]$$

$$\mathrm{Cov}[X, Y] = \mathrm{E}[\mathrm{Cov}[\,X, Y\,|\,Z]] + \mathrm{Cov}[\mathrm{E}\,[X\,|\,Z]\,, \mathrm{E}\,[Y\,|\,Z]]$$

The distribution of max and min of independent random variables

Let $X_1, \ldots, X_n$ be independent random variables
(distribution functions $F_i(x)$ and tail distributions $G_i(x),\ i = 1, \ldots, n$)

Distribution of the maximum

$$\begin{aligned}
\mathrm{P}\{\max(X_1, \ldots, X_n) \le x\} &= \mathrm{P}\{X_1 \le x, \ldots, X_n \le x\} \\
&= \mathrm{P}\{X_1 \le x\} \cdots \mathrm{P}\{X_n \le x\} \qquad \text{(independence!)} \\
&= F_1(x) \cdots F_n(x)
\end{aligned}$$

Distribution of the minimum

$$\begin{aligned}
\mathrm{P}\{\min(X_1, \ldots, X_n) > x\} &= \mathrm{P}\{X_1 > x, \ldots, X_n > x\} \\
&= \mathrm{P}\{X_1 > x\} \cdots \mathrm{P}\{X_n > x\} \qquad \text{(independence!)} \\
&= G_1(x) \cdots G_n(x)
\end{aligned}$$