Probability and Statistics

Exercise sheet 8

Exercise 8.1

(a) Let $X \sim \mathrm{U}([0,1])$. Compute $E(X^n)$, $E(X^{\frac{1}{n}})$ $(n \geq 1)$, and $\Psi_X(t) = E[e^{tX}]$ whenever it is defined.

(b) Let $X \sim \mathrm{Beta}(\alpha, \beta)$, $\alpha > 0$ and $\beta > 0$. Compute $E(X)$ and $\mathrm{var}(X)$.

Hint: Use the "trick" that any density function f has to integrate to 1.

(c) Let $X \sim \mathrm{Exp}(\lambda)$, for $\lambda > 0$. Compute the cdf of X and $E(X^n)$ for $n \geq 1$.

Remark: Watch out for the parametrisation - different sources may use different parametrisations. In the lectures we consider the density function of an $\mathrm{Exp}(\lambda)$-distributed random variable to be $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

Solution 8.1

(a) For $\alpha > -1$,

$$E(X^\alpha) = \int_0^1 x^\alpha dx = \frac{1}{\alpha + 1}.$$

In particular,

$$E(X^n) = \frac{1}{n+1}$$

and

$$E(X^{\frac{1}{n}}) = \frac{n}{n+1}.$$

Moreover,

$$\begin{aligned} \Psi_X(t) &= E[e^{tX}] \\ &= \int_0^1 e^{tx} dx \\ &= \frac{e^t - 1}{t} \end{aligned}$$

for any $t \neq 0$, while $\Psi_X(0) = 1$.

(b) The density of X is

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbb{1}_{x \in (0,1)}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function.

Thus,

$$\begin{aligned}
E(X) &= \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \int_0^1 \frac{x^{\alpha}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \int_0^1 \frac{x^{\alpha}(1-x)^{\beta-1}}{B(\alpha+1,\beta)} dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

(using the density "trick" and the formula for the beta function).

$$\begin{aligned}
E(X^2) &= \int_0^1 x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \int_0^1 \frac{x^{\alpha+1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \int_0^1 \frac{x^{\alpha+1}(1-x)^{\beta-1}}{B(\alpha+2,\beta)} dx \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+2)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}$$

Therefore we obtain

$$\begin{aligned}
\mathrm{var}(X) &= E(X^2) - E(X)^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha^3+\alpha^2+\alpha^2\beta+\alpha\beta-\alpha^3-\alpha^2\beta-\alpha^2}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

(c) The density of X is

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}.$$

For $x > 0$, we calculate

$$\int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$$

and therefore

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

We also compute

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n \lambda e^{-\lambda x} dx \\ &= \int_0^\infty \frac{y^n}{\lambda^n} \lambda e^{-y} \frac{dy}{\lambda} \quad (y = \lambda x) \\ &= \frac{1}{\lambda^n} \int_0^\infty y^n e^{-y} dy \\ &= \frac{\Gamma(n+1)}{\lambda^n} \\ &= \frac{n!}{\lambda^n} \end{aligned}$$

(using properties of the gamma function; alternatively one could integrate by parts).

Exercise 8.2 (Waiting time.)
An auto towing company services a 50 mile stretch of a highway. The company is located 20 miles from one end of the stretch. Breakdowns occur uniformly along the highway and the towing trucks travel at a constant speed of 50mph. Find the mean and variance of the time elapsed between the instant the company is called and a towing truck arrives.

Solution 8.2 Call the left endpoint of the 50 mile stretch zero, and let X be the number of miles from the left endpoint that a breakdown occurs. Then $X \sim \mathrm{U}[0, 50]$. Assume that the towing company is located 20 miles from the left endpoint, so that the distance Y of the breakdown from the location of the towing company is $Y = |X - 20|$. It will take the truck $Z = \frac{Y}{50} = \frac{|X-20|}{50}$ hours to reach the location of the breakdown. We want the mean and variance of Z. First,

$$\begin{aligned} E(Z) = E\left(\frac{|X - 20|}{50}\right) &= \frac{1}{50} \int_0^{50} |x - 20| f(x) dx = \frac{1}{50^2} \int_0^{50} |x - 20| dx \\ &= \frac{1}{2500} \int_0^{20} (20 - x) dx + \int_{20}^{50} (x - 20) dx = \frac{1}{2500} [200 + 450] = 0.26h. \end{aligned}$$

Next,

$$\begin{aligned} E(Z^2) &= \frac{1}{2500} E[(X - 20)^2] = \frac{1}{2500} E[X^2 - 40X + 400] \\ &= \frac{1}{2500} \frac{1}{50} \int_0^{50} (x^2 - 40x + 400) dx \approx 0.0933, \end{aligned}$$

and therefore

$$\mathrm{var}(Z) = E(Z^2) - E(Z)^2 = 0.0933 - 0.26^2 \approx 0.0257.$$

(if one wanted the units, one could write $\mathrm{var}(Z) \approx 0.0257h^2$).

Exercise 8.3 (Uniforms, uniforms...)

(a) Consider a random variable $X \sim \mathrm{U}\left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)$. Find $E(\sin(X))$ and $\mathrm{var}(\sin(X))$.

(b) The lengths of the sides of a triangle are X, $2X$ and $2.5X$ with $X \sim \mathrm{U}([0, \alpha])$ for some $\alpha > 0$.

- Find the mean and variance of its area.
  Hint: Recall that if
  $$s = \frac{a+b+c}{2}$$
  with a, b, c the lengths, then the area of the triangle is
  $$|\Delta| = \sqrt{s(s-a)(s-b)(s-c)}$$
  (Heron's formula).
- How should we choose α so that the mean area is ≥ 1?

(c) Take $X_1, ..., X_n$ to be $\overset{\text{iid}}{\sim} \mathrm{U}([0,1])$. Let $M_n = \max(X_1, ..., X_n)$. Find the cdf and pdf of M_n. Can you recognise this distribution? What are $E(M_n)$ and $\mathrm{var}(M_n)$?

Solution 8.3

(a) We have

$$\begin{aligned} E(\sin(X)) &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin(x)}{\pi} dx \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}(\sin(X)) = E(\sin(X)^2) &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin(x)^2}{\pi} dx \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1-\cos(2x)}{2\pi} dx \\ &= \frac{1}{2}. \end{aligned}$$

(b) The area of the triangle is given by Heron's formula:

$$|\Delta| = X^2\sqrt{2.75 \times 0.25 \times 0.75 \times 1.75} = aX^2$$

for a constant $a \approx 0.95$.

- Note that if $X \sim \mathrm{U}([0, \alpha])$, then $Y = \frac{X}{\alpha} \sim \mathrm{U}([0, 1])$. Therefore,
  $$\begin{aligned} E[|\Delta|] &= E[aX^2] \\ &= a\alpha^2 E[Y^2] \\ &= \frac{a\alpha^2}{3} \end{aligned}$$
  as we found in question 1(a).
  Moreover
  $$\begin{aligned} E[|\Delta|^2] &= a^2\alpha^4 E[Y^4] \\ &= \frac{a^2\alpha^4}{5} \end{aligned}$$

also by question 1(a). Finally

$$\operatorname{var}(|\Delta|) = E[|\Delta|^2] - E[|\Delta|]^2 = \frac{4a^2\alpha^4}{45}.$$

- We need to ensure that

$$\frac{a\alpha^2}{3} \geq 1 \Leftrightarrow \alpha \geq \sqrt{\frac{3}{a}} \approx 1.78,$$

so this is the suitable value of α.

(c) We can easily compute the cdf: for $x \in (0,1)$,

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = x^n.$$

For general x,

$$F_{M_n}(x) = \begin{cases} 0, & x < 0, \\ x^n, & 0 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$

Therefore, its density is

$$f_{M_n}(x) = nx^{n-1}$$

on $(0,1)$ (0 otherwise).

We can recognise this as a particular case of a Beta distribution, $M_n \sim \text{Beta}(n,1)$. Thus from question 1(b),

$$E(M_n) = \frac{n}{n+1}$$

and

$$\operatorname{var}(M_n) = \frac{n}{(n+1)^2(n+2)}.$$

Exercise 8.4 (Quantile transformation.) Recall that for a given cdf F, the quantile t_α of order $\alpha \in (0,1)$ is defined as

$$t_\alpha = \inf\{t : F(t) \geq \alpha\} =: F^{-1}(\alpha).$$

F^{-1} denotes the generalised inverse of F. When the latter is bijective (at least in the neighbourhood of t_α), then F^{-1} is the inverse of F in the classical sense.

(a) Consider $U \sim \text{U}([0,1])$. Show that $1 - U \sim \text{U}([0,1])$.

Hint: Compute, for example, the cdf or the pdf of $1 - U$.

(b) Consider

$$X := -\frac{1}{\lambda}\log(U)$$

(for $\omega : U(\omega) = 0$, take $X(\omega) = 0$, say).

Find the cdf of X. Can you recognise this distribution?

(c) Now, consider the following problem: take a cdf F which is bijective when viewed as a map $F : (a,b) \to (0,1)$, for some $-\infty \leq a < b \leq +\infty$.

Define $X = F^{-1}(U)$, with $U \sim \text{U}([0,1])$.

- Compute the cdf of X.
  Hint: You may need to consider the cases $a = -\infty$, $a > -\infty$, $b = +\infty$, $b < +\infty$.
- Compute the pdf of X, assuming that F is C^1 on (a, b) with $F'(x) > 0 \ \forall x \in (a, b)$.
- Can you make the link to (b)?

(d) Suppose you are given a numerical algorithm which enables you to generate a random number from $[0, 1]$. You would like to generate a random number X which follows the Cauchy distribution, i.e.

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \ x \in \mathbb{R}.$$

Can you propose a way to do that, based on your previous findings?

Solution 8.4

(a) For $x \in (0, 1)$, it is clear that

$$\begin{aligned} P(1-U \leq x) &= P(U \geq 1-x) \\ &= P(U > 1-x) \\ &= 1 - P(U \leq 1-x) \\ &= 1-(1-x) = x \end{aligned}$$

(using absolute continuity of U in the second line). For general x, note that $1 - U$ cannot be smaller than 0 or greater than 1, so

$$F_{1-U}(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$

So we conclude that $1 - U \sim \mathrm{U}([0, 1])$.

(b) Similarly we compute for $x > 0$:

$$\begin{aligned} P(X \leq x) &= P\left(-\frac{1}{\lambda}\log(U) \leq x\right) \\ &= P(\log(U) \geq -\lambda x) \\ &= P(U \geq e^{-\lambda x}) \\ &= P(U > e^{-\lambda x}) \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

and

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & 0 \leq x. \end{cases}$$

Therefore X has an $\mathrm{Exp}(\lambda)$ distribution, by 1(c).

(c)
- We calculate for $x \in (a, b)$,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \qquad (*) \\ &= F(x). \qquad (**) \end{aligned}$$

$(**)$ holds since $F(x) \in [0,1]$. $(*)$ requires more justification. By definition, for $u \in (0,1)$,

$$F^{-1}(u) = t_u = \inf\{t : F(t) \geq u\}.$$

Since F is non-decreasing and right-continuous, we can see that the set on the right-hand side is a closed interval of the form $[x_0, +\infty]$ for some x_0, indeed we have

$$\{t : F(t) \geq u\} = [t_u, +\infty).$$

Therefore,

$$\begin{aligned} & x \geq t_u \\ \Leftrightarrow\ & x \in [t_u, +\infty) = \{t : F(t) \geq u\} \\ \Leftrightarrow\ & F(x) \geq u \end{aligned}$$

Therefore, since $U \in (0,1)$ almost surely, the step $(*)$ is justified. We observe that X has the cdf F, and moreover while we only wrote this for $x \in (a,b)$, we can see that $F(x) = F_X(x) = 0$ on $(-\infty, a]$ and $F(x) = F_X(x) = 1$ on $[b, +\infty)$. We conclude that X and $F^{-1}(U)$ have the same cdf, which is F.

- If F is piecewise C^1, then by the fundamental theorem of calculus,

$$F(x) = \int_a^x F'(y)dy$$

so the density of F is $f(x) = F'(x)$.

- In (b) we defined

$$X = -\frac{1}{\lambda}\log(U).$$

Note that an $\mathrm{Exp}(\lambda)$ random variable has a cdf $F(x) = 1 - e^{-\lambda x}$ on $(0, +\infty)$. Since F is smooth and strictly increasing, we find the inverse by calculating

$$y = 1 - e^{-\lambda x} \Leftrightarrow x = -\frac{1}{\lambda}\log(1-y) = F^{-1}(y).$$

Therefore, $X = -\frac{1}{\lambda}\log(1-U)$ has an $\mathrm{Exp}(\lambda)$-distribution. This is not exactly the same as in (b), but since by (a), $1-U$ also has a $\mathrm{U}([0,1])$ distribution, we can replace U by $1-U$ in the expression above.

(d) Let U be the random number generated by the algorithm, with distribution $U \sim \mathrm{U}([0,1])$. For the Cauchy distribution, we can compute

$$F(x) = \int_{-\infty}^x \frac{1}{\pi(1+y^2)}dy = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan(x)\right)$$

and therefore by the previous parts, we can use the inverse

$$F^{-1}(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$$

to simulate a random variable X with a Cauchy distribution, namely by setting

$$X = F^{-1}(U) = \tan\left(\pi\left(U - \frac{1}{2}\right)\right).$$

Exercise 8.5 (Optional.) Let $X_1, ..., X_n$ (for some $n \geq 2$) be random variables defined on the same probability space $(\Omega, \mathcal{A}, P)$. A necessary and sufficient condition for (mutual) independence of $X_1, ..., X_n$ is that

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

for any $(x_1, ..., x_n) \in \mathbb{R}^n$.

The goal of this exercise is to show that when $X_1, ..., X_n$ are discrete, (1) is equivalent to

$$P(X_{i_1} = x_1, ..., X_{i_m} = x_m) = \prod_{j=1}^{m} P(X_{i_j} = x_j) \tag{2}$$

for any $1 \leq m \leq n$ and $i_1 < ... < i_m$, $(x_1, ..., x_m) \in \mathbb{R}^m$.

(a) Take $m = n$ (in which case $i_1 = 1, i_2 = 2, ..., i_m = n$).

- Focus on X_1 and show that (2) implies that for any $(x_1, ..., x_n) \in \mathbb{R}^n$,

$$P(X_1 \leq x_1, X_2 = x_2, ..., X_n = x_n) = P(X_1 \leq x_1)P(X_2 = x_2)...P(X_n = x_n).$$

- Repeating this argument inductively, show that (1) holds.

(b) Now, we want to show that (1) implies (2). Fix $\{i_1, ..., i_m\} = \mathcal{J} \subset \{1, ..., n\}$, with strict inclusion.

- Show that (1) implies that

$$P(X_{i_1} \leq x_1, ..., X_{i_m} \leq x_m) = \prod_{j=1}^{m} P(X_{i_j} \leq x_j).$$

Hint: You may want to take limits in (1) as $x_i \to +\infty$ for $i \notin \mathcal{J}$.

- Focus on i_1 (hold the other events depending on $i_2, ..., i_m$ fixed). Show that we have

$$P(X_{i_1} < x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) = P(X_{i_1} < x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m)$$

for any $(x_1, ..., x_m) \in \mathbb{R}^m$.
Hint: Consider $A_k := \{X_{i_1} \leq x_1 - \frac{1}{k}, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m\}$ and use the monotone convergence theorem.

- Conclude that

$$P(X_{i_1} = x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) = P(X_{i_1} = x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m).$$

- Repeat your argument inductively to conclude.

Solution 8.5

(a)
- Since $X_1, ..., X_n$ are discrete, assume that X_1 is supported on $\{x_1^1, x_2^1, ...\}$. Then,

$$\begin{aligned}
P(X_1 \le x_1, X_2 = x_2, ..., X_n = x_n) &= \sum_{i=1}^{\infty} P(X_1 = x_i^1, X_1 \le x_1, X_2 = x_2, ..., X_n = x_n) \\
&= \sum_{i=1}^{\infty} P(X_1 = x_i^1, X_2 = x_2, ..., X_n = x_n) \mathbb{1}_{x_i^1 \le x_1} \\
&= \sum_{i=1}^{\infty} \left(\prod_{j=2}^{m} P(X_j = x_j) \right) P(X_1 = x_i^1) \mathbb{1}_{x_i^1 \le x_1} \\
&= \prod_{j=2}^{m} P(X_j = x_j) \sum_{i=1}^{\infty} P(X_1 = x_i^1) \mathbb{1}_{x_i^1 \le x_1} \\
&= \prod_{j=2}^{m} P(X_j = x_j) P(X_1 \le x_1) \\
&= P(X_1 \le x_1) P(X_2 = x_2) ... P(X_n = x_n).
\end{aligned}$$

- Similarly if we assume that

  $$\begin{aligned}
  &P(X_1 \le x_1, ..., X_{k-1} \le x_{k-1}, X_k = x_k, ..., X_n = x_n) = \\
  &P(X_1 \le x_1) ... P(X_{k-1} \le x_{k-1}) P(X_k = x_k) ... P(X_n = x_n),
  \end{aligned}$$

  we assume that X_k is supported on $\{x_1^k, x_2^k, ...\}$ and redo the computations:

$$\begin{aligned}
&P(X_1 \le x_1, X_2 \le x_2, ..., X_k \le x_k, X_{k+1} = x_{k+1}, ..., X_n = x_n) \\
&= \sum_{i=1}^{\infty} P(X_1 \le x_1, X_2 \le x_2, ..., X_k \le x_k, X_k = x_i^k, X_{k+1} = x_{k+1}, ..., X_n = x_n) \\
&= \sum_{i=1}^{\infty} P(X_1 \le x_1, X_2 \le x_2, ..., X_k = x_i^k, X_{k+1} = x_{k+1}, ..., X_n = x_n) \mathbb{1}_{x_i^k \le x_k} \\
&= \sum_{i=1}^{\infty} \left(\prod_{j=1}^{k-1} P(X_j \le x_j) \right) \left(\prod_{j=k+1}^{m} P(X_j = x_j) \right) P(X_k = x_i^k) \mathbb{1}_{x_i^k \le x_k} \\
&= \prod_{j=1}^{k-1} P(X_j \le x_j) \prod_{j=k+1}^{m} P(X_j = x_j) \sum_{i=1}^{\infty} P(X_k = x_i^k) \mathbb{1}_{x_i^k \le x_k} \\
&= \prod_{j=1}^{k-1} P(X_j \le x_j) \prod_{j=k+1}^{m} P(X_j = x_j) P(X_k \le x_k) \\
&= P(X_1 \le x_1) P(X_2 \le x_2) ... P(X_k \le x_k) P(X_{k+1} = x_{k+1}) ... P(X_n = x_n).
\end{aligned}$$

  Thus, by induction we obtain the desired result.

(b) Now, we want to show that (1) implies (2). Fix $\{i_1, ..., i_m\} = \mathcal{J} \subset \{1, ..., n\}$, with strict inclusion.

- For each $k \in \mathbb{N}_0$ and $(x_1, ..., x_m) \in \mathbb{R}^m$, consider the measurable set

  $$A_k := \left(\bigcap_{l \neq \mathcal{J}} \{X_l \le k\} \right) \cap \left(\bigcap_{j=1}^{m} \{X_{i_j} \le x_j\} \right).$$

The sequence $(A_k)_{k\geq 0}$ is increasing with

$$\bigcup_{k=0}^{\infty} A_k = A = \bigcap_{j=1}^{m}\{X_{i_j} \leq x_j\}.$$

By the monotone convergence theorem,

$$P(A) = \lim_{k\to\infty} P(A_k)$$

where

$$P(A_k) = \prod_{l\neq\mathcal{J}} P(X_l \leq k) \times \prod_{j=1}^{m} P(X_{i_j} \leq x_j)$$

and $\lim_{k\to\infty} P(X_l \leq k) = 1$ using the same theorem. This implies that

$$\lim_{k\to\infty} P(A_k) = \prod_{j=1}^{m} P(X_{i_j} \leq x_j).$$

It follows that

$$P\left(\bigcap_{i=1}^{m}\{X_{i_j} \leq x_j\}\right) = \prod_{j=1}^{m} P(X_{i_j} \leq x_j),$$

or equivalently,

$$P(X_{i_1} \leq x_1, ..., x_{i_m} \leq x_m) = \prod_{i=1}^{m} P(X_{i_j} \leq x_j).$$

- Now, let us focus on X_{i_1}. As in the hint, we consider the measurable set $B_k := \{X_{i_1} \leq x_1 - \frac{1}{k}, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m\}$ for $k \geq 1$.
  The sequence $(B_k)_{k\geq 1}$ is increasing with

$$\bigcup_{k=1}^{\infty} B_k = B := \{X_{i_1} < x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m\}.$$

By the monotone convergence theorem, $\lim_{k\to\infty} P(B_k) = P(B)$. But

$$\begin{aligned} P(B_k) &= P\left(X_{i_1} \leq x_1 - \frac{1}{k}, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m\right) \\ &= P\left(X_{i_1} \leq x_1 - \frac{1}{k}\right) P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m) \end{aligned}$$

and using the same theorem, $\lim_{k\to\infty} P\left(X_{i_1} \leq x_1 - \frac{1}{k}\right) = P(X_{i_1} < x_1)$. This implies that

$$P(X_{i_1} < x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) = P(X_{i_1} < x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m)$$

as we wanted.

- Now,

$$
\begin{aligned}
&P(X_{i_1} = x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) \\
&= P(X_{i_1} \leq x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) - P(X_{i_1} < x_1, X_{i_2} \leq x_2, ..., X_{i_m} \leq x_m) \\
&= P(X_{i_1} \leq x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m) - P(X_{i_1} < x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m) \\
&= [P(X_{i_1} \leq x_1) - P(X_{i_1} < x_1)]P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m) \\
&= P(X_{i_1} = x_1)P(X_{i_2} \leq x_2)...P(X_{i_m} \leq x_m).
\end{aligned}
$$

- Similarly, given the previous result, one can obtain by the monotone convergence argument that

$$P(X_{i_1} < x_1, X_{i_2} < x_2, X_{i_3} \leq x_3..., X_{i_m} \leq x_m) = P(X_{i_1} < x_1)P(X_{i_2} < x_2)P(X_{i_3} \leq x_3)...P(X_{i_m} \leq x_m),$$

  and therefore by taking a difference we get

$$P(X_{i_1} = x_1, X_{i_2} = x_2, X_{i_3} \leq x_3, ..., X_{i_m} \leq x_m) = P(X_{i_1} = x_1)P(X_{i_2} = x_2)P(X_{i_3} \leq x_3)...P(X_{i_m} \leq x_m).$$

  We can keep on going, and by an inductive argument we get eventually

$$P(X_{i_1} = x_1, ..., X_{i_m} = x_m) = \prod_{j=1}^{m} P(X_{i_j} = x_j)$$

  as we wanted.