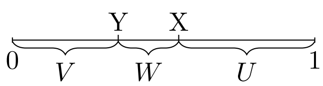
Probability and Statistics

Exercise sheet 11

Exercise 11.1 (Breaking a stick) Suppose $X \sim \mathrm{U}([0,1])$ and $Y \mid X \sim \mathrm{U}([0,X])$. Consider now $U = 1 - X, V = Y, W = X - Y$ (this represents breaking a stick into parts with length X and U, and then breaking the left piece again into V and W). Find $E[\max(U,V,W)]$.



Solution 11.1

Solution 1

Note that $(Y, X-Y, 1-X)$ has the same distribution as $(XZ, X(1-Z), 1-X)$ where $X \perp\!\!\!\perp Z$ and $Z \sim \mathrm{U}([0,1])$. To see this, we define $Z := \frac{Y}{X}$ and show that $Z \sim \mathrm{U}([0,1])$ independently of X.

Consider the map $g(x,y) = \left(x, \frac{y}{x}\right)^T$ defined on $\mathbb{R} \setminus \{0\} \times \mathbb{R}$.

Solving for $g(x,y) = (u,v)^T$ we have $x = u$ and $y = uv$. Hence, $g : \mathbb{R} \setminus \{0\} \times \mathbb{R} \to \mathbb{R} \setminus \{0\} \times \mathbb{R}$ is bijective. Also, g is $C^1(\mathbb{R} \setminus \{0\} \times \mathbb{R})$ with gradient at $(x,y)^T \in \mathbb{R} \setminus \{0\} \times \mathbb{R}$ given by

$$\nabla g(x,y) = \begin{pmatrix} \frac{\partial g_1}{\partial x}(x,y) & \frac{\partial g_1}{\partial y}(x,y) \\ \frac{\partial g_2}{\partial x}(x,y) & \frac{\partial g_2}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{y}{x^2} & \frac{1}{x} \end{pmatrix}.$$

Thus,

$$J_g(x,y) = \det(\nabla g(x,y)) = \frac{1}{x} \neq 0.$$

Furthermore, by construction of Y, the random vector (X,Y) satisfies

$$P((X,Y)^T \in \mathcal{O} := (0,1) \times (0,1)) = 1.$$

If f denotes the joint density of $(X,Y)^T$, then it follows from the Jacobian formula that

$$\begin{aligned} f_{(X,Z)^T}(x,z) &= f(g^{-1}(x,z)) \frac{1}{\frac{1}{x}} \mathbb{1}_{(x,z)^T \in g(\mathcal{O})} \\ &= f(g^{-1}(x,z)) x \mathbb{1}_{(x,z)^T \in (0,1) \times (0,+\infty)} \\ &= f(x,xz) x \mathbb{1}_{x \in (0,1)} \mathbb{1}_{z \in (0,+\infty)}. \end{aligned}$$

The joint density f is given by:

$$\begin{aligned} f(x,y) &\overset{\text{a.e.}}{=} f(y \mid x) f_X(x) \\ &= \frac{1}{x} \mathbb{1}_{y \in (0,x)} \mathbb{1}_{x \in (0,1)}. \end{aligned}$$

Thus,

$$\begin{aligned} f_{(X,Z)^T}(x,z) &\overset{\text{a.e.}}{=} \mathbb{1}_{xz \in (0,x)} \mathbb{1}_{x \in (0,1)} \mathbb{1}_{x \in (0,1)} \mathbb{1}_{z \in (0,+\infty)} \\ &= \mathbb{1}_{z \in (0,1)} \mathbb{1}_{x \in (0,1)} \mathbb{1}_{x \in (0,1)} \mathbb{1}_{z \in (0,+\infty)} \\ &= \mathbb{1}_{x \in (0,1)} \mathbb{1}_{z \in (0,1)}. \end{aligned}$$

We then conclude that X, Z are indeed i.i.d with $\mathrm{U}([0,1])$ distribution.

Now note that

$$\max(XZ, X(1-Z), 1-X) = \max(\max(XZ, X(1-Z)), 1-X) = \begin{cases} \max(XZ, 1-X), & \text{if } Z \geq \frac{1}{2} \\ \max(X(1-Z), 1-X), & \text{otherwise.} \end{cases}$$

Then, exploiting also the symmetry $(X,Z) \stackrel{\mathrm{d}}{=} (X, 1-Z)$,

$$\begin{aligned} E[\max(XZ, X(1-Z), 1-X)] &= E[\max(XZ, 1-X)\mathbb{1}_{Z\geq\frac{1}{2}}] + E[\max(X(1-Z), 1-X)\mathbb{1}_{Z<\frac{1}{2}}] \\ &= 2E[\max(XZ, 1-X)\mathbb{1}_{Z\geq\frac{1}{2}}]. \end{aligned}$$

We can then use the joint density $f(x,z) = \mathbb{1}_{x\in[0,1]} \times \mathbb{1}_{z\in[0,1]}$ to compute this expectation. Note that we will need to split the integral in order to compute the maximum, and the cutoff point can be calculated by

$$xz \leq 1-x \Leftrightarrow x(1+z) \leq 1 \Leftrightarrow x \leq \frac{1}{1+z}.$$

Thus we get:

$$\begin{aligned} E[\max(XZ, 1-X)\mathbb{1}_{Z\geq\frac{1}{2}}] &= \iint \max(xz, 1-x) f(x,z) \mathbb{1}_{z\geq\frac{1}{2}} dxdz \\ &= \int_{\frac{1}{2}}^{1} \int_{0}^{\frac{1}{1+z}} (1-x) dxdz + \int_{\frac{1}{2}}^{1} \int_{\frac{1}{1+z}}^{1} xzdxdz \\ &= I_1 + I_2. \end{aligned}$$

We compute each of these integrals:

$$\begin{aligned} I_1 &= \int_{\frac{1}{2}}^{1} \left[-\frac{(1-x)^2}{2}\right]_0^{\frac{1}{1+z}} dz \\ &= \frac{1}{2}\int_{\frac{1}{2}}^{1} \left(1 - \left(1 - \frac{1}{1+z}\right)^2\right) dz \\ &= \frac{1}{2}\int_{\frac{1}{2}}^{1} \left(\frac{2}{1+z} - \frac{1}{(1+z)^2}\right) dz \\ &= \frac{1}{2}\left(2[\log(1+z)]_{\frac{1}{2}}^{1} + \left[\frac{1}{1+z}\right]_{\frac{1}{2}}^{1}\right) \\ &= \log(2) - \log\left(\frac{3}{2}\right) + \frac{1}{2}\left(\frac{1}{2} - \frac{2}{3}\right) \\ &= \log\left(\frac{4}{3}\right) - \frac{1}{12}. \end{aligned}$$

$$\begin{aligned}
I_2 &= \int_{\frac{1}{2}}^{1} \left(\int_{\frac{1}{1+z}}^{1} x dx \right) z dz \\
&= \int_{\frac{1}{2}}^{1} \left[\frac{x^2}{2} \right]_{\frac{1}{1+z}}^{1} z dz \\
&= \frac{1}{2} \int_{\frac{1}{2}}^{1} \left(1 - \frac{1}{(1+z)^2} \right) z dz \\
&= \frac{1}{2} \left(\int_{\frac{1}{2}}^{1} z dz - \int_{\frac{1}{2}}^{1} \frac{z+1-1}{(1+z)^2} dz \right) \\
&= \frac{1}{2} \left(\int_{\frac{1}{2}}^{1} z dz - \int_{\frac{1}{2}}^{1} \frac{1}{1+z} dz + \int_{\frac{1}{2}}^{1} \frac{1}{(1+z)^2} dz \right) \\
&= \frac{1}{2} \left(\left[\frac{z^2}{2} \right]_{\frac{1}{2}}^{1} - [\log(1+z)]_{\frac{1}{2}}^{1} - \left[\frac{1}{1+z} \right]_{\frac{1}{2}}^{1} \right) \\
&= \frac{1}{2} \left(\frac{3}{8} - \log(2) + \log\left(\frac{3}{2}\right) + \frac{2}{3} - \frac{1}{2} \right) \\
&= \frac{13}{48} - \frac{1}{2} \log\left(\frac{4}{3}\right).
\end{aligned}$$

Finally

$$\begin{aligned}
E[\max(XZ, X(1-Z), 1-X)] &= 2(I_1 + I_2) \\
&= 2\left(\log\left(\frac{4}{3}\right) - \frac{1}{12} + \frac{13}{48} - \frac{1}{2}\log\left(\frac{4}{3}\right) \right) \\
&= \frac{3}{8} + \log\left(\frac{4}{3}\right).
\end{aligned}$$

Solution 2
Note that

$$\max(u, v, w) = \max(u, \max(v, w)).$$

As a first step, consider $Z := \max(V, W)$. Conditionally on $X = x$,

$$\begin{aligned}
F_{Z|X}(z \mid x) &= P(Z \leq z \mid X = x) \\
&= P(\max(Y, x - Y) \leq z \mid X = x) \\
&= P(x - z \leq Y \leq z \mid X = x).
\end{aligned}$$

Since $Y \mid X = x \sim \mathrm{U}(0, x)$, a case-by-case calculation yields that

$$F_{Z|X}(z \mid x) = \begin{cases} 0, & z \leq \frac{x}{2} \\ \frac{2}{x}\left(z - \frac{x}{2}\right), & \frac{x}{2} \leq z \leq x \\ 1, & x \leq z. \end{cases}$$

For example, in the middle case, we can check that $0 \leq x - z \leq z \leq x$ and therefore

$$P(Y \in [x - z, z] \mid X = x) = \frac{1}{x}(z - (x - z)) = \frac{2}{x}\left(z - \frac{x}{2}\right).$$

The point of this is that, from looking at the conditional cdf above, and differentiating:

$$f_{Z|X}(z \mid x) = \frac{2}{x} \mathbb{1}_{\frac{x}{2} \leq z \leq x},$$

we conclude that $Z \mid X = x \sim \mathrm{U}(\frac{x}{2}, x)$.

To conclude, we need to calculate $E[\max(Z, U)]$. We will exploit the conditional distribution we found for Z by working with the law of iterated expectation:

$$E[\max(Z,U)] = E[\max(Z, 1-X)] = E[E[\max(Z, 1-X) \mid X]].$$

We calculate the conditional expectation as follows:

$$\begin{aligned}
E[\max(Z, 1-X) \mid X = x] &= \begin{cases} E(1 - X \mid X = x), & x \leq \frac{1}{2} \\ E[\max(Z, 1-X) \mid X = x], & \frac{1}{2} \leq x \leq \frac{2}{3} \\ E(Z \mid X = x), & x \geq \frac{2}{3} \end{cases} \\
&= \begin{cases} 1 - x, & x \leq \frac{1}{2} \\ \int_{\frac{x}{2}}^{1-x} \frac{2(1-x)}{x} dz + \int_{1-x}^{x} \frac{2z}{x} dz, & \frac{1}{2} \leq x \leq \frac{2}{3} \\ \frac{3x}{4}, & x \geq \frac{2}{3} \end{cases} \\
&= \begin{cases} 1 - x, & x \leq \frac{1}{2} \\ \left(1 - \frac{3x}{2}\right) \frac{2(1-x)}{x} + \frac{x^2 - (1-x)^2}{x}, & \frac{1}{2} \leq x \leq \frac{2}{3} \\ \frac{3x}{4}, & x \geq \frac{2}{3} \end{cases} \\
&= \begin{cases} 1 - x, & x \leq \frac{1}{2} \\ \left(1 - \frac{3x}{2}\right) \frac{2(1-x)}{x} + \frac{x^2 - (1-x)^2}{x}, & \frac{1}{2} \leq x \leq \frac{2}{3} \\ \frac{3x}{4}, & x \geq \frac{2}{3} \end{cases} \\
&= \left\{ \begin{matrix} 1 - x, & x \leq \frac{1}{2} \\ 3x - 3 + \frac{1}{x}, & \frac{1}{2} \leq x \leq \frac{2}{3} \\ \frac{3x}{4}, & x \geq \frac{2}{3}. \end{matrix} \right\} =: g(x)
\end{aligned}$$

Finally,

$$\begin{aligned}
E(\max(U, V, W)) &= E[E[\max(Z, 1-X) \mid X]] \\
&= E[g(X)] \\
&= \int_0^1 g(x) dx \\
&= \int_0^{\frac{1}{2}} (1 - x) dx + \int_{\frac{1}{2}}^{\frac{2}{3}} \left(3x - 3 + \frac{1}{x}\right) dx + \int_{\frac{2}{3}}^{1} \frac{3x}{4} dx \\
&= \frac{1}{2} - \frac{1}{8} + \frac{7}{24} - \frac{1}{2} + \log\left(\frac{4}{3}\right) + \frac{5}{24} \\
&= \log\left(\frac{4}{3}\right) + \frac{3}{8}.
\end{aligned}$$

Exercise 11.2 (Uniforms, uniforms...) Suppose $X \sim U([0,1])$ and consider $Y = 2X$.

(a) What is the joint distribution of (X, Y)?

(b) Does this joint distribution have a density with respect to the Lebesgue measure on $\mathbb{R}^2$?

(c) (Probability of a diamond) Let X, Y and Z be $\stackrel{iid}{\sim} \mathrm{U}([-1, 1])$. Find $P(|X| + |Y| + |Z| \leq 1)$.

Solution 11.2
Remark: We use here the common shorthand notation

$$x \wedge y := \min(x, y).$$

(a) The (joint) cdf of (X, Y) is given by

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= P(X \leq x, 2X \leq y) \\ &= P\left(X \leq x \wedge \frac{y}{2}\right) \\ &= \begin{cases} 0, & \text{if } x \wedge \frac{y}{2} < 0 \\ x \wedge \frac{y}{2}, & \text{if } 0 \leq x \wedge \frac{y}{2} < 1 \\ 1, & \text{if } x \wedge \frac{y}{2} \geq 1. \end{cases} \end{aligned}$$

(b) We have that $P((X, Y) \in B) = 1$, where $B = \{(x, y) \in \mathbb{R}^2 : y = 2x\}$. Since $\lambda_2(B) = 0$, where λ_2 denotes the Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$, it follows that $P_{(X,Y)}$ is not absolutely continuous with respect to λ_2, and hence cannot, by the Radon-Nikodym theorem, admit a density with respect to λ_2.

(c) We want to calculate

$$p := P(|X| + |Y| + |Z| \leq 1) = \frac{1}{8} \iiint_D \mathbb{1}_{x \in [-1,1]} \mathbb{1}_{y \in [-1,1]} \mathbb{1}_{z \in [-1,1]} dx dy dz$$

where $D := \left\{(x, y, z) \in \mathbb{R}^3 : |x| + |y| + |z| \leq 1\right\}$.

By symmetry, the integral on one octant is equal to the integral on any of the other octants:

$$p = \frac{8}{8} \iiint_{D^+} \mathbb{1}_{x \in [-1,1]} \mathbb{1}_{y \in [-1,1]} \mathbb{1}_{z \in [-1,1]} dx dy dz$$

where $D^+ := \left\{(x, y, z) \in \mathbb{R}^3 : |x| + |y| + |z| \leq 1, 0 \leq x, y, z \leq 1\right\}$.

After these considerations, we compute (using Fubini's Theorem):

$$\begin{aligned} p &= \int_0^1 \int_0^{1-z} \left(\int_0^{1-z-y} dx\right) dy dz \\ &= \int_0^1 \left(\int_0^{1-z} (1 - z - y) dy\right) dz \\ &= \int_0^1 \frac{(1 - z)^2}{2} dz \\ &= -\frac{1}{6} \left[(1 - z)^3\right]_{z=0}^{z=3} \\ &= \frac{1}{6}. \end{aligned}$$

Exercise 11.3 Recall that $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$ (a Gaussian vector with expectation $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$) if for any $v \in \mathbb{R}^n$,

$$v^T X = \sum_{i=1}^n v_i X_i \sim \mathcal{N}(v^T\mu, v^T\Sigma v).$$

The goal of this exercise is to show the following remarkable property:
$(*)$ $X_1, ..., X_n$ are independent if and only if for all $i \neq j$, $\Sigma_{ij} = \text{cov}(X_i, X_j) = 0$.

(a) Show that $(*)$ is necessary.

(b) To show it is sufficient, we shall use the following result:
$X_1, ..., X_n$ are independent if and only if

$$\Psi_X(t) := E(e^{t^T X}) = \prod_{i=1}^n \Psi_{X_i}(t_i) \left(= \prod_{i=1}^n E(e^{t_i X_i}) \right)$$

for all $t \in \mathbb{R}^n$.

Compute $\Psi_X(t)$ when $\Sigma_{ij} = 0$ for all $i \neq j$ and conclude.

Hint: $t^T X$ is a normal random variable, for which we know the expression of the moment generating function.

(c) Taking $n \geq 3$, let $Y \in \mathbb{R}^p$ (for $2 \leq p \leq n-1$) be a subset of the original vector X. Using a simple argument, explain why Y is also a Gaussian vector. When are the components of Y independent?

(d) Let $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} \in \mathbb{R}^5$ have a $\mathcal{N}(\mu, \Sigma)$ distribution, where

$$\mu = \begin{pmatrix} -1 \\ 2 \\ 0 \\ \frac{1}{2} \\ 3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 9 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & -1 & 6 \\ 0 & 0 & 16 & 0 & 3 \\ 1 & -1 & 0 & 4 & 3 \\ 2 & 6 & 3 & 3 & 49 \end{pmatrix}.$$

Which subsets of $X_1, ..., X_5$ can you say are independent?

(e) Consider the case $n = 2$, and $(X_1, X_2)^T$ a Gaussian pair with expectation $\mu = (\mu_1, \mu_2)^T$ and a 2×2 covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_1^2, \sigma_2^2 > 0$ are the (marginal) variances and ρ is the correlation.

Find a and b such that $X_1 + X_2$ and $aX_1 + bX_2$ are independent.

Solution 11.3

(a) If $X_1, ..., X_n$ are independent, then for $i \neq j$,

$$\begin{aligned}\Sigma_{ij} &= \text{cov}(X_i, X_j)\\ &= E[(X_i - E(X_i))(X_j - E(X_j))]\\ &= E[(X_i - E(X_i))]E[(X_j - E(X_j))] = 0.\end{aligned}$$

(b) For $t \in \mathbb{R}^n$,

$$\Psi_X(t) = E[e^{t^T X}] = \exp\left(t^T\mu + \frac{t^T\Sigma t}{2}\right)$$

since $t^T X \sim \mathcal{N}(t^T\mu, t^T\Sigma t)$, and we know the mgf of the normal distribution. If $\Sigma_{ij} = 0$ for all $i \neq j$, it follows immediately that

$$t^T\Sigma t = \sum_{i=1}^{n} t_i^2 \Sigma_{ii}$$

implying that

$$\begin{aligned}\Psi_X(t) &= \exp\left(\sum_{i=1}^{n} t_i\mu_i + \frac{1}{2}\sum_{i=1}^{n} t_i^2\Sigma_{ii}\right)\\ &= \prod_{i=1}^{n} \exp\left(t_i\mu_i + \frac{1}{2}t_i^2\Sigma_{ii}\right)\\ &= \prod_{i=1}^{n} \Psi_{X_i}(t_i).\end{aligned}$$

Therefore, $X_1, ..., X_n$ are independent, using the hint.

(c) Without loss of generality, we can focus on the case $Y = (X_1, ..., X_p)^T$. For any $a \in \mathbb{R}^p$,

$$a^T Y = \sum_{i=1}^{p} a_i X_i = \sum_{i=1}^{n} v_i X_i$$

where $v_i = a_i$ for $1 \leq i \leq p$ and $v_i = 0$ otherwise. Thus $a^T Y$ is a linear combination of the X_i, and so it is normally distributed. As $a \in \mathbb{R}^p$ is arbitrary, it follows that Y is Gaussian in $\mathbb{R}^p$.

Thus, by (c), the components of Y are independent if and only if all their covariances are 0, i.e. if all the entries $\Sigma_{ij} = 0$, where $i \neq j$ and X_i, X_j are components of Y.

(d) Note that $\Sigma_{12} = \Sigma_{13} = \Sigma_{23} = \Sigma_{34} = 0$. Therefore, $\{X_1, X_2, X_3\}$ (and subsets) are independent, and $\{X_3, X_4\}$ are independent.

(e)

$$\begin{pmatrix} X_1 + X_2 \\ aX_1 + bX_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

hence it is a Gaussian vector, as a linear transformation of the Gaussian vector $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

Thus, $X_1 + X_2$ and $aX_1 + bX_2$ are independent if and only if

$$\begin{aligned}
&\operatorname{cov}(X_1 + X_2, aX_1 + bX_2) = 0 \\
\Leftrightarrow\ & a \operatorname{var}(X_1) + b \operatorname{cov}(X_1, X_2) + a \operatorname{cov}(X_1, X_2) + b \operatorname{var}(X_2) = 0 \\
\Leftrightarrow\ & a\sigma_1^2 + b\sigma_1\sigma_2\rho + a\sigma_1\sigma_2\rho + b\sigma_2^2 = 0 \\
\Leftrightarrow\ & \sigma_1(\sigma_1 + \sigma_2\rho)a = -\sigma_2(\sigma_2 + \sigma_1\rho)b
\end{aligned}$$

Therefore, $X_1 + X_2$ and $aX_1 + bX_2$ are independent if and only if a and b are in the above ratio.

We should look at the special case where the coefficients are both 0. Since $\sigma_1 > 0$, $\sigma_2 > 0$, we must have $\sigma_1 + \sigma_2\rho = \sigma_2 + \sigma_1\rho = 0$ and then

$$-\rho = \frac{\sigma_1}{\sigma_2} = -\frac{1}{\rho}.$$

Since $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in [-1, 1]$, this is only possible if and only if $\sigma_1 = \sigma_2$ and $\rho = -1$. In that special case (and only then), indeed, $X_1 + X_2 = 0$ and for any $(a, b) \in \mathbb{R}^2$, $X_1 + X_2$ is independent from $aX_1 + bX_2$.

Exercise 11.4 (I hope you're arriving soon)

Alice and Viera plan to meet at a café, and each will arrive at a random time between 15:00 and 15:30, independently of each other. Find the probability that the first to arrive has to wait between 5 and 10 minutes for the other to arrive.

Solution 11.4

Remark: Similarly to question 2(c), we use both notations

$$x \wedge y := \min(x, y), \quad x \vee y := \max(x, y).$$

Let Alice arrive X minutes after 15h, and similarly let Viera arrive Y minutes after 15h. It is assumed that $X, Y \overset{\text{iid}}{\sim} \mathrm{U}([0, 30])$.

We have

$$p = P(X \vee Y - X \wedge Y \in [5, 10]) = \iint_D f(x, y)dxdy$$

where $D = \{(x, y) \in \mathbb{R}^2 : x \vee y - x \wedge y \in [5, 10]\}$ and $f(x, y) = \frac{1}{30}\mathbb{1}_{x\in[0,30]}\frac{1}{30}\mathbb{1}_{y\in[0,30]}$ is the joint density.

We can decompose $D = D_1 \cup D_2$ where $D_1 = \{(x, y) \in \mathbb{R}^2 : x \geq y, x - y \in [5, 10]\}$ and $D_2 = \{(x, y) \in \mathbb{R}^2 : x < y, y - x \in [5, 10]\}$, and by symmetry it is clear that the integral on each is the same. Therefore,

$$\begin{aligned}
p &= 2\iint_{D_1} f(x, y)dxdy \\
&= \frac{2}{900}\left(\int_0^{20}\left(\int_{y+5}^{y+10} dx\right)dy + \int_{20}^{25}\left(\int_{y+5}^{30} dx\right)dy\right) \\
&= \frac{2}{900}\left(\int_0^{20} 5dy + \int_{20}^{25}(25 - y)dy\right) \\
&= \frac{2}{9} + \frac{1}{36} \\
&= \frac{1}{4}
\end{aligned}$$

Exercise 11.5 (Cauchy-Schwarz) (optional)

The goal of this exercise is to show the Cauchy-Schwarz inequality, which is stated as follows.

Let X and Y be two random variables defined on the same probability space, such that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then,

$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)} \tag{1}$$

with equality if and only if $P(X = 0) = 1$ or $P(Y = aX) = 1$ for some $a \in \mathbb{R}$.

1. For $t \in \mathbb{R}$, write $E[(Y - tX)^2]$ as a function of t.
2. Using the fact that $E[(Y - tX)^2] \geq 0$ for any $t \in \mathbb{R}$, conclude that either $P(X = 0) = 1$ or
$$E(X^2)\left(t - \frac{E(XY)}{E(X^2)}\right)^2 - \frac{E(XY)^2}{E(X^2)} + E(Y^2) \geq 0$$
for $t \in \mathbb{R}$.
3. Prove the inequality (1).
4. If $P(X = 0) < 1$, show that equality case in (1) holds if and only if $P(Y = t^*X) = 1$, and give the expression for t^*.

   Hint: Remember that for Z a non-negative random variable, $E(Z) = 0$ if and only if $P(Z = 0) = 1$.
5. Conclude on the condition for the equality case to hold.

Solution 11.5

1. $$E((Y - tX)^2) = E(Y^2 - 2tXY + t^2X^2) = t^2E(X^2) - 2tE(XY) + E(Y^2).$$
2. If $E(X^2) = 0$, then $P(X = 0) = 1$. Otherwise, if $E(X^2) > 0$,
$$\begin{aligned} E((Y - tX)^2) &= E(X^2)\left(t^2 - 2t\frac{E(XY)}{E(X^2)} + \frac{E(XY)^2}{E(X^2)^2}\right) - \frac{E(XY)^2}{E(X^2)} + E(Y^2) \\ &= E(X^2)\left(t - \frac{E(XY)}{E(X^2)}\right)^2 + E(Y^2) - \frac{E(XY)^2}{E(X^2)} \geq 0 \end{aligned}$$
since the original expression $E((Y - tX)^2) \geq 0$ for any $t \in \mathbb{R}$.
3. In the case that $P(X = 0) = 1$, $E(XY) = 0$ and the inequality (1) holds. Otherwise, we have the inequality above for any $t \in \mathbb{R}$. That applies in particular to the choice
$$t^* = \frac{E(XY)}{E(X^2)},$$
and therefore
$$E(Y^2) \geq \frac{E(XY)^2}{E(X^2)}$$
or equivalently,
$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$$
as we wanted.

4. Suppose that equality holds and $P(X=0)<1$. As above, we must then have $E(Y^2)=\frac{E(XY)^2}{E(X^2)}$. Then, with the same t^* as before, we get that

$$E[(Y-t^*X)^2]=E(X^2)\left(t^*-\frac{E(XY)}{E(X^2)}\right)^2=0.$$

   Following the hint, this means that $1=P((Y-t^*X)^2=0)=P(Y=t^*X)$.

   The converse is obvious: if $P(Y=t^*X)=1$, then

$$E(XY)=E(Xt^*X)=t^*E(X^2)=\sqrt{E(X^2)}\sqrt{E((t^*X)^2)}=\sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

5. From the previous part, if equality holds then either $P(X=0)=1$ or $P(Y=t^*X)=1$ for some $t^*\in\mathbb{R}$. Conversely, we already saw that $P(Y=t^*X)=1$ implies that equality holds, and similarly if $P(X=0)=1$ then

$$0=E(XY)=\sqrt{0}\sqrt{E(Y^2)}.$$

   Therefore equality holds if and only if either $P(X=0)=1$ or $P(Y=t^*X)=1$ for some $t^*\in\mathbb{R}$.