STAT 512: Statistical Inference **Autumn 2022**

Quiz Session 2: Independence and Conditional Independence

Yikun Zhang *October 12, 2022*

Some problems in this notes are selected from the STAT 512 lecture notes [Perlman, 2020] written by *Prof. Michael Perlman* (MDP).

Problem 1 (Simpson's Paradox; Example 3.3 in MDP)**.** *Given three events A, B, C, prove or disprove the following implication:*

$$\begin{Bmatrix} P(A|B,C) > P(A|B^c,C), \\ P(A|B,C^c) > P(A|B^c,C^c) \end{Bmatrix} \overset{?}{\Longrightarrow} P(A|B) > P(A|B^c). \tag{1}$$

If the above implication does not hold, could we add some assumptions to make it hold?

Solution.

The implication (1) does not hold in general. Here is a counterexample.

Physics	Accept	Reject	
Female	60	40	P(A|F,Ph)=0.6
Male	50	50	P(A|M,Ph)=0.5
English	Accept	Reject	
Female	250	750	P(A|F,En)=0.25
Male	20	80	P(A|M,En)=0.2
Total	Accept	Reject	
Female	310	790	P(A|F)=0.28
Male	70	130	P(A|M)=0.35

One could verify that

$$\begin{aligned} P(A|F,Ph) &> P(A|M,Ph), \\ P(A|F,En) &> P(A|M,En), \end{aligned}$$

but

$$P(A|F) < P(A|M).$$

A key reasoning of the Simpson's paradox is that most female students applied to English, where the acceptance rate is considerably low; see also Figure 1.

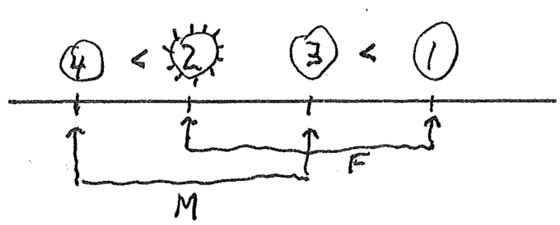


Figure 1: Explanation of the Simpson's paradox.

If B is independent with C, then the implication (1) does hold.

Proof. Note that

$$\begin{aligned} P(A|B) &= P(A|B,C)P(C|B) + P(A|B,C^c)P(C^c|B), \\ P(A|B^c) &= P(A|B^c,C)P(C|B^c) + P(A|B^c,C^c)P(C^c|B^c). \end{aligned}$$

Given the independence between B and C, it follows that

$$\begin{aligned} P(A|B) &= P(A|B,C)P(C) + P(A|B,C^c)P(C^c), \\ P(A|B^c) &= P(A|B^c,C)P(C) + P(A|B^c,C^c)P(C^c) \end{aligned}$$

and hence

$$\begin{aligned} & P(A|B) - P(A|B^c) \\ = \; & P(C)\underbrace{[P(A|B,C) - P(A|B^c,C))]}_{>0} + P(C^c)\underbrace{[P(A|B,C^c) - P(A|B^c,C^c)]}_{>0} \\ > \; & 0. \end{aligned}$$

□

Problem 2 (Uniform on the Unit Disk; MDP Example 1.12)**.** *Let (X,Y) follow a uniform distribution on the unit disk $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$.*

(a) Show that X and Y are not independent. (Notes: Try not to calculate the marginal distributions of X and Y at this point.)

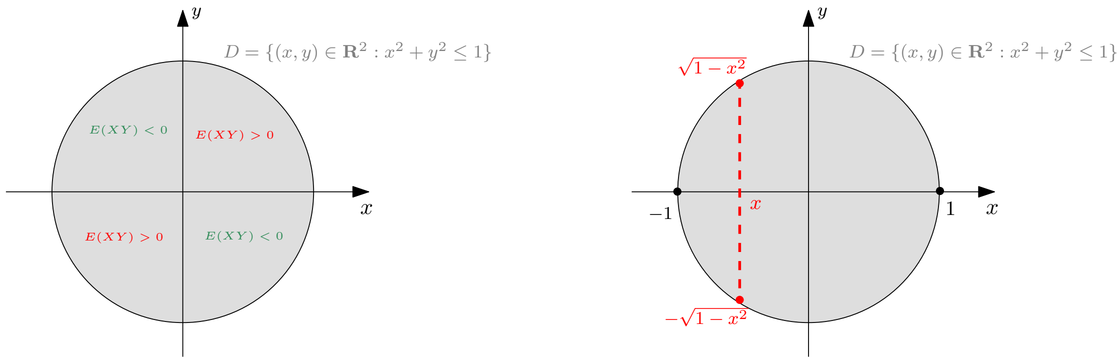
(b) Verify that X and Y are uncorrelated.

(c) Calculate the marginal probability density functions (PDF) of X and Y.

(d) Find the conditional PDF of $Y|X$.

(e) Show that $\frac{Y}{\sqrt{1-X^2}}$ is independent of X.

(f) Consider representing the random vector (X,Y) in polar coordinates as (R,Θ), where $R = \sqrt{X^2+Y^2}$ and $\Theta = \arctan(Y/X)$. Verify that R and Θ are independent.



(a) The sign information of $\mathbb{E}(XY)$.

(b) Conditional distribution of $Y|X$.

Figure 2: Unit disk D and other graphical illustrations.

(g) Let* $S = \frac{Y}{\sqrt{1-X^2}}$ *and* $T = \frac{X}{\sqrt{1-Y^2}}$. *Prove or disprove: S and T are independent.*

*(h**) Find the joint PDF $f(s,t)$ and cumulative distribution function $F(s,t) = \mathbb{P}(S \leq s, T \leq t)$ of (S,T).*

Solution.

(a) Notice that the range of (X, Y) is the unit disk D, while the marginal ranges of X and Y are both $[-1, 1]$. Given that $D \neq [-1, 1] \times [-1, 1]$, we conclude that X and Y are not independent.

(b) By the symmetry of D, we know that $\mathbb{E}(XY) = \mathbb{E}(X) = \mathbb{E}(Y) = 0$; see also Figure 2a. Hence, X and Y are uncorrelated, i.e., has no linear trend.

(c) Notice that the joint PDF of (X, Y) is

$$f_{X,Y}(x, y) = \frac{1}{\pi} \cdot \mathbb{1}_D(x, y), \tag{2}$$

where

$$\mathbb{1}_D(x, y) = \begin{cases} 1 & \text{if } (x, y) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

By definition, the marginal PDF of X is

$$f_X(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1 - x^2} \cdot \mathbb{1}_{[-1,1]}(x).$$

Likewise, the marginal PDF of Y is $f_Y(y) = \frac{2}{\pi} \sqrt{1 - y^2} \cdot \mathbb{1}_{[-1,1]}(y)$.

(d) By definition (see Section 1.5 in Lecture Note 1), the conditional PDF of $Y|X$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\mathbb{1}_D(x, y)/\pi}{2\sqrt{1 - x^2} \cdot \mathbb{1}_{[-1,1]}(x)/\pi} = \frac{1}{2\sqrt{1 - x^2}} \cdot \mathbb{1}_{\left[-\sqrt{1-x^2}, \sqrt{1-x^2}\right]}(y).$$

Remark 1. *It is worth mentioning that the domains/supports of the conditional PDF $f_{Y|X}(y|x)$ and marginal PDF $f_Y(y)$ are different! The former one, $\left[-\sqrt{1 - x^2}, \sqrt{1 - x^2}\right]$ depends on the choice of $X = x$, while the latter one, $[-1, 1]$, is independent of X. Furthermore, given that $f_{Y|X}(y|x) \neq f_Y(y)$, it provides another way to verify the dependency between X and Y.*

(e) By (d), we know that the conditional distribution $Y|X = x$ is Uniform $\left[-\sqrt{1-x^2}, \sqrt{1-x^2}\right]$; see also Figure 2b. Thus, $\frac{Y}{\sqrt{1-X^2}}|X$ follows the distribution Uniform$[-1,1]$, which in turn shows that $\frac{Y}{\sqrt{1-X^2}}$ is independent of X.

(f) By the uniformity of (X,Y) on D, the joint range of (R,Θ) is $[0,1]\times[0,2\pi]$, which is the cross product of the marginal ranges of R and Θ. Moreover, the joint cumulative distribution function (CDF) of (R,Θ) is

$$\begin{aligned}
F_{R,\Theta}(r,\theta) &\equiv \mathbb{P}\left(0 \le R \le r, 0 \le \Theta \le \theta\right) \\
&= \frac{\theta r^2/2}{\pi} \\
&= r^2 \cdot \frac{\theta}{2\pi} \\
&= \mathbb{P}(0 \le R \le r) \cdot \mathbb{P}\left(0 \le \Theta \le \theta\right),
\end{aligned}$$

yielding the product of marginal CDFs of R and Θ. Hence, R and Θ are independent.

Remark 2. *One can also leverage the Jacobian method (see Section 8.2 in Lecture Note 8) to directly compute the joint PDF of (R,Θ) as:*

$$\begin{aligned}
f_{R,\Theta}(r,\theta) &= f_{X,Y}(r\cos\theta, r\sin\theta) \\
&= \frac{1}{\pi} \cdot \mathbb{1}_D(r\cos\theta, r\sin\theta) \cdot \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| \\
&= \frac{1}{\pi} \cdot \mathbb{1}_{\{[0,1]\times[0,2\pi]\}}(r,\theta) \cdot \left|\det\begin{pmatrix}\cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{pmatrix}\right| \\
&= \frac{r}{\pi} \cdot \mathbb{1}_{\{[0,1]\times[0,2\pi]\}}(r,\theta) \\
&= 2r\mathbb{1}_{[0,1]}(r) \cdot \frac{1}{2\pi}\mathbb{1}_{[0,2\pi]}(\theta),
\end{aligned}$$

which is the product of two marginal PDFs $f_R(r)$ and $f_\Theta(\theta)$.

(g) We claim that S and T are not independent. To prove this, it suffices to show that

$$\mathbb{P}(S \le 1/2, T \le 1/2) \ne \mathbb{P}(S \le 1/2) \cdot \mathbb{P}(T \le 1/2). \tag{3}$$

By (e), we know that S and T follow Uniform$[-1,1]$. Hence,

$$\mathbb{P}(S \le 1/2) = \mathbb{P}(T \le 1/2) = \int_{-1}^{1/2} \frac{1}{2} ds = \frac{3}{4}.$$

On the other hand,

$$\begin{aligned}
\mathbb{P}(S \le 1/2, T \le 1/2) &= \mathbb{P}\left(-1 \le \frac{Y}{\sqrt{1-X^2}} \le \frac{1}{2}, -1 \le \frac{X}{\sqrt{1-Y^2}} \le \frac{1}{2}\right) \\
&= \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4),
\end{aligned}$$

where, referring to Figure 3,

$$\begin{aligned}
A_1 &= \{X \le 0, Y \le 0\}, \\
A_2 &= \left\{X < 0, Y > 0, X^2 + 4Y^2 \le 1\right\}, \\
A_3 &= \left\{X > 0, Y < 0, 4X^2 + Y^2 \le 1\right\}, \\
A_4 &= \left\{X \ge 0, Y \ge 0, X^2 + 4Y^2 \le 1, 4X^2 + Y^2 \le 1\right\},
\end{aligned}$$

and they are mutually exclusive.

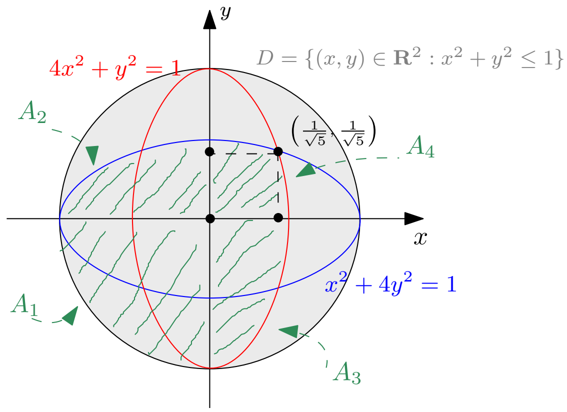


Figure 3: Graphical display of the probability $\mathbb{P}(S \leq 1/2, T \leq 1/2)$.

Given the uniformity of (X, Y) on D, we know that $\mathbb{P}(A_1) = \frac{1}{4}$. In addition, using the property that the area of an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is πab (which can be proved by integrating $\int_{-a}^{a} 2b\sqrt{1 - x^2/a^2}dx$), we have that

$$\mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{\frac{1}{4} \cdot \pi \cdot \frac{1}{2}}{\pi} = \frac{1}{8}.$$

(Or, one can resort to Cavalieri's principle to argue that the area of $A_2 + A_3$ is equal to $1/4$ of the area of D.) Finally, we are not going to directly compute $\mathbb{P}(A_4)$. Instead, according to (3), we intend to show that

$$\mathbb{P}(S \leq 1/2, T \leq 1/2) = \sum_{i=1}^{4} \mathbb{P}(A_i) = \frac{1}{4} + 2 \times \frac{1}{8} + \mathbb{P}(A_4) = \frac{1}{2} + \mathbb{P}(A_4)$$
$$> \mathbb{P}(S \leq 1/2) \cdot \mathbb{P}(T \leq 1/2) = \frac{9}{16}.$$

Notice that two ellipses $4x^2 + y^2 = 1$ and $x^2 + 4y^2 = 1$ intersects at $\left(\frac{1}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)$ in the first quadrant. By the shape of the circular boundaries of A_4, it contains the square with lower left vertex $(0, 0)$ and upper right vertex $\left(\frac{1}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)$, whose probability is $\frac{1}{5\pi}$. Therefore,

$$\mathbb{P}(S \leq 1/2, T \leq 1/2) = \frac{1}{2} + \mathbb{P}(A_4) \geq \frac{1}{2} + \frac{1}{5\pi} > \frac{9}{16}.$$

It completes our proof for the claim that S and T are not independent.

(h) Based on the Jacobian method (Theorem 8.1 in Lecture Note 8), we take $s = \frac{y}{\sqrt{1-x^2}}$, $t = \frac{x}{\sqrt{1-y^2}}$, and calculate the determinant

$$\left|\frac{\partial(s, t)}{\partial(x, y)}\right| = \begin{vmatrix} \frac{\partial s}{\partial x} & \frac{\partial s}{\partial y} \\ \frac{\partial t}{\partial x} & \frac{\partial t}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{xy}{(1-x^2)^{\frac{3}{2}}} & \frac{1}{\sqrt{1-x^2}} \\ \frac{1}{\sqrt{1-y^2}} & \frac{xy}{(1-y^2)^{\frac{3}{2}}} \end{vmatrix} = \frac{1 - x^2 - y^2}{(1 - x^2)^{\frac{3}{2}}(1 - y^2)^{\frac{3}{2}}}.$$

Solving for x^2, y^2, we obtain that $x^2 = \frac{t^2(1-s^2)}{1-s^2t^2}$ and $y^2 = \frac{s^2(1-t^2)}{1-s^2t^2}$. Hence, the joint PDF of (S,T) is

$$\begin{aligned}
f(s,t) &= \frac{1}{\pi}\left|\frac{\partial(x,y)}{\partial(s,t)}\right| \cdot \mathbb{1}_{\{x^2+y^2\leq 1\}} \\
&= \frac{1}{\pi}\cdot\frac{(1-x^2)^{\frac{3}{2}}(1-y^2)^{\frac{3}{2}}}{1-x^2-y^2}\cdot \mathbb{1}_{\left\{\frac{t^2(1-s^2)}{1-s^2t^2}+\frac{s^2(1-t^2)}{1-s^2t^2}\leq 1\right\}} \\
&= \frac{\sqrt{(1-t^2)(1-s^2)}}{\pi(1-s^2t^2)}\cdot \mathbb{1}_{\{(1-s^2)(1-t^2)\geq 0\}} \\
&= \frac{\sqrt{(1-t^2)(1-s^2)}}{\pi(1-s^2t^2)}\cdot \mathbb{1}_{\{(s,t)\in[-1,1]\times[-1,1]\}}.
\end{aligned}$$

Notice that $f(s,t)$ cannot be factored into the product of two marginal densities, so it provides another way to justify the dependence between S and T.

To compute the CDF $F(s,t) = \int_{-1}^{s}\int_{-1}^{t} f(u,v)\,dudv$, we need to consider 4 different cases regarding the signs of s and t.

Case 1: $-1 \leq s < 0, -1 \leq t < 0$. Then, from $-1 < \frac{y}{\sqrt{1-x^2}} \leq s, -1 < \frac{x}{\sqrt{1-y^2}} \leq t$, we know that the integral range of $f_{X,Y}(x,y)$ on D satisfies

$$x^2+y^2 \leq 1 \quad \frac{y^2}{s^2}+x^2 \geq 1, \quad \frac{x^2}{t^2}+y^2 \geq 1.$$

See also Figure 4 for a graphical illustration. Thus,

$$\begin{aligned}
F(s,t) &= \int\int_{D_1} \frac{1}{\pi}\,dxdy \\
&= \frac{4}{\pi}\left[\frac{\pi}{4} - \left(\frac{\pi t}{4} + \frac{\pi s}{4} - \int_0^{\theta_1}\int_0^1 tr\,drd\theta - \int_{\theta_1}^{\frac{\pi}{2}}\int_0^1 sr\,drd\theta\right)\right] \\
&= (1-t) + \frac{2(1-s)\theta_1}{\pi},
\end{aligned}$$

where $\theta_1 = \arctan\left(\frac{s\sqrt{1-t^2}}{t\sqrt{1-s^2}}\right)$.

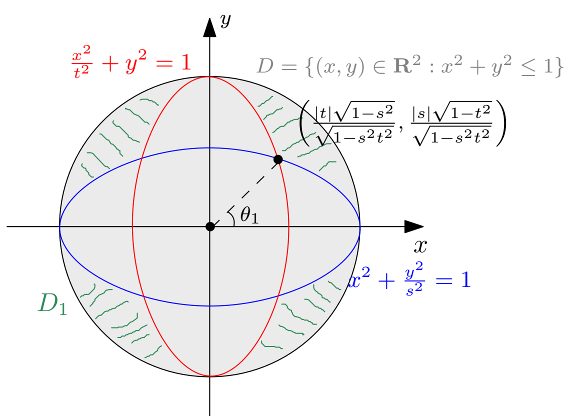


Figure 4: Integrating range of $f_{X,Y}(x,y)$ when $-1 \leq s < 0, -1 \leq t < 0$.

Case 2: $0 \leq s \leq 1, -1 \leq t < 0$. Then, from $-1 < \frac{y}{\sqrt{1-x^2}} \leq s, -1 < \frac{x}{\sqrt{1-y^2}} \leq t$, we know that the integral range of $f_{X,Y}(x,y)$ on D satisfies

$$\frac{x^2}{t^2} + y^2 \geq 1$$

and

$$\frac{y^2}{s^2} + x^2 \leq 1, y \geq 0 \quad \text{or} \quad \frac{x^2}{t^2} + y^2 \leq 1, x > 0.$$

See also Figure 5 for a graphical illustration. Thus,

$$\begin{aligned} F(s,t) &= \int\int_{D_2} \frac{1}{\pi}\, dxdy \\ &= \frac{1}{\pi}\left[\frac{\pi s}{2} - 2\left(\int_0^{\theta_1}\int_0^1 tr\, drd\theta + \int_{\theta_1}^{\frac{\pi}{2}}\int_0^1 sr\, drd\theta\right) + \frac{\pi}{2} - \frac{\pi t}{2}\right] \\ &= \frac{1}{\pi}\left[\frac{\pi s}{2} - 2\left(\frac{t\theta_1}{2} + \frac{s}{2}\left(\frac{\pi}{2} - \theta_1\right)\right) + \frac{\pi}{2} - \frac{\pi t}{2}\right] \\ &= \frac{1-t}{2} + \frac{(1-s)\theta_1}{2\pi}, \end{aligned}$$

where $\theta_1 = \arctan\left(-\frac{s\sqrt{1-t^2}}{t\sqrt{1-s^2}}\right)$.

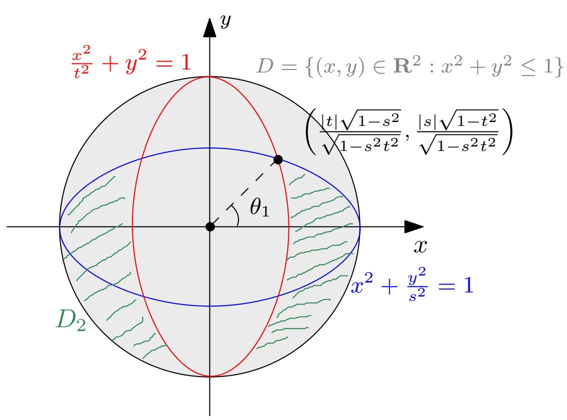


Figure 5: Integrating range of $f_{X,Y}(x,y)$ when $0 \leq s \leq 1, -1 \leq t < 0$.

Case 3: $-1 \leq s < 0, 0 \leq t \leq 1$. Then, from $-1 < \frac{y}{\sqrt{1-x^2}} \leq s, -1 < \frac{x}{\sqrt{1-y^2}} \leq t$, we know that the integral range of $f_{X,Y}(x,y)$ on D satisfies

$$x^2 + \frac{y^2}{s^2} > 1$$

and

$$x^2 + y^2 \leq 1, x \leq 0 \quad \text{or} \quad \frac{x^2}{t^2} + y^2 \leq 1, x > 0.$$

See also Figure 6 for a graphical illustration. Thus,

$$\begin{aligned} F(s,t) &= \int\int_{D_3} \frac{1}{\pi}\, dxdy \\ &= \frac{1}{\pi}\left[\frac{\pi}{2} - \frac{\pi s}{2} + \frac{\pi t}{2} - 2\left(\frac{t\theta_1}{2} + \frac{s}{2}\left(\frac{\pi}{2} - \theta_1\right)\right)\right] \\ &= \frac{1+t}{2} - s + \frac{(s-t)\theta_1}{\pi}, \end{aligned}$$

where $\theta_1 = \arctan\left(-\frac{s\sqrt{1-t^2}}{t\sqrt{1-s^2}}\right)$.

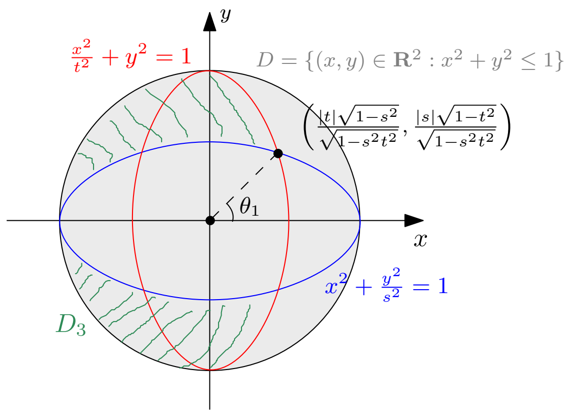


Figure 6: Integrating range of $f_{X,Y}(x,y)$ when $-1 \le s < 0, 0 \le t \le 1$.

Case 4: $0 \le s \le 1, 0 \le t \le 1$. Then, from $-1 < \frac{y}{\sqrt{1-x^2}} \le s, -1 < \frac{x}{\sqrt{1-y^2}} \le t$, we know that the integral range of $f_{X,Y}(x,y)$ on D satisfies

$$x^2+y^2 \le 1, y<0 \quad \text{or} \quad x^2+\frac{y^2}{s^2} \le 1, y \ge 0$$

and

$$x^2+y^2 \le 1, x<0 \quad \text{or} \quad \frac{x^2}{t^2}+y^2 \le 1, x \ge 0.$$

See also Figure 7 for a graphical illustration. Thus,

$$\begin{aligned} F(s,t) &= \int\int_{D_4} \frac{1}{\pi}\,dxdy \\ &= \frac{1}{\pi}\left[\frac{\pi}{4}+\frac{\pi s}{4}+\frac{\pi t}{4}+2\left(\frac{t\theta_1}{2}+\frac{s}{2}\left(\frac{\pi}{2}-\theta_1\right)\right)\right] \\ &= \frac{1+t+2s}{4}+\frac{(s-t)\theta_1}{2\pi}, \end{aligned}$$

where $\theta_1 = \arctan\left(\frac{s\sqrt{1-t^2}}{t\sqrt{1-s^2}}\right)$. □

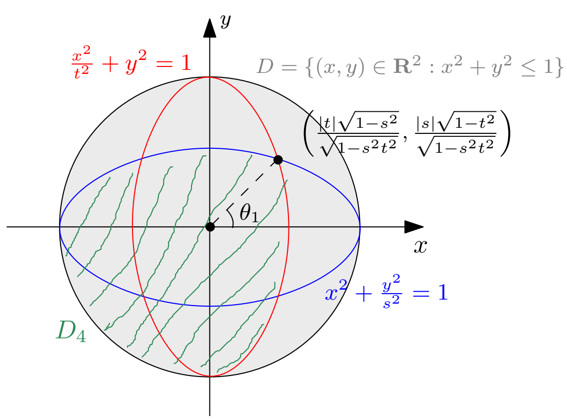


Figure 7: Integrating range of $f_{X,Y}(x,y)$ when $0 \le s \le 1, 0 \le t \le 1$.

Problem 3 (Homework 1; Exercise 1.32 in Casella and Berger 2002)**.** *An employer is about to hire one new employee from a group of N candidates, whose future potential can be rated on a scale from 1 to N. The employer proceeds according to the following rules:*

(a) Each candidate is seen in succession (in random order) and a decision is made whether to hire the candidate.
(b) Having rejected $m-1$ candidates ($m>1$), the employer can hire the mth candidate only if the mth candidate is better than the previous $m-1$.

Suppose a candidate is hired on the ith trial. What is the probability that the best candidate was hired?

Solution. Let E be the event that i^{th} candidate is the best, and F be the event that i^{th} candidate is better than the previous $i-1$ candidates. The probability that the best candidate was hired is

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\cdot\mathbb{P}(E)}{\mathbb{P}(F)}.$$

Here, $\mathbb{P}(F|E) = 1$ because if the i^{th} candidate is the best among all, it must be better than the previous $i-1$ candidates. Moreover, $\mathbb{P}(E) = \frac{1}{N}$ and $\mathbb{P}(F) = \frac{1}{i}$. Thus,

$$\mathbb{P}(E|F) = \frac{i}{N}.$$

References

G. Casella and R. Berger. *Statistical Inference*. Duxbury advanced series. Thomson Learning, 2nd ed. edition, 2002.

M. Perlman. Probability and mathematical statistics i (stat 512 lecture notes), 2020. URL https://sites.stat.washington.edu/people/mdperlma/STAT%20512%20MDP%20Notes.pdf.