Assignment #0 Solutions

Due: never

Problem 1

Given the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{1}$$

(a) Expand $\mathbf{Xy} + \mathbf{z}$

(b) Expand $\mathbf{y^T X y}$

Solution:

(a)

$$\mathbf{Xy} + \mathbf{z} = \begin{pmatrix} x_{11}y_1 + x_{12}y_2 \\ x_{21}y_1 + x_{22}y_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_{11}y_1 + x_{12}y_2 + z_1 \\ x_{21}y_1 + x_{22}y_2 + z_2 \end{pmatrix}$$

(b)

$$\begin{aligned} \mathbf{y^T X y} &= \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} x_{11}y_1 + x_{21}y_2 & x_{12}y_1 + x_{22}y_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= x_{11}y_1^2 + x_{21}y_1y_2 + x_{12}y_1y_2 + x_{22}y_2^2 \end{aligned}$$

Problem 2

Assume matrix $\mathbf{X}$ has dimensionality (or shape) $(n \times d)$, and vector $\mathbf{w}$ has shape $(d \times 1)$.

(a) What shape is $\mathbf{y} = \mathbf{Xw}$?

(b) What shape is $(\mathbf{X}^T\mathbf{X})^{-1}$?

(c) Using y from part (a), what shape is $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$?

(d) Assume vector $\mathbf{w'} = \mathbf{w}^T$. What shape is $\mathbf{y'} = \mathbf{Xw'}^T$?

Solution:

(a) $(n \times 1)$

(b) $\mathbf{X}^T\mathbf{X}$ has shape $(d \times d)$ and so $(\mathbf{X}^T\mathbf{X})^{-1}$ has shape $(d \times d)$

(c) $\mathbf{X}^T\mathbf{y}$ has shape $(d \times 1)$ and so $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ has shape $(d \times 1)$

(d) Transposing a matrix twice returns the original matrix so we have $\mathbf{y'} = \mathbf{Xw}$, which has shape $(n \times 1)$

Problem 3
Write $\mathbf{u} = \mathbf{u}^{\parallel} + \mathbf{u}^{\perp}$ where $\mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ is the projection of $\mathbf{u}$ onto $\mathbf{v}$. Verify that $\langle \mathbf{u}^{\parallel}, \mathbf{u}^{\perp}\rangle = 0$ and that $\mathbf{u} = \mathbf{u}^{\parallel}$ if and only if $\mathbf{u}$ is a scaled multiple of $\mathbf{v}$.

Solution: We have $\mathbf{u}^{\perp} = \mathbf{u} - \mathbf{u}^{\parallel} = \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$. Then

$$\langle \mathbf{u}^{\parallel}, \mathbf{u}^{\perp}\rangle = \left\langle \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}, \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} \right\rangle = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle} \left\langle \mathbf{v}, \mathbf{u} - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} \right\rangle =$$

$$\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle} \left(\langle \mathbf{v},\mathbf{u}\rangle - \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\langle \mathbf{v},\mathbf{v}\rangle \right) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}(\langle \mathbf{v},\mathbf{u}\rangle - \langle \mathbf{u},\mathbf{v}\rangle) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}(\langle \mathbf{v},\mathbf{u}\rangle - \langle \mathbf{v},\mathbf{u}\rangle) = 0,$$

where we note that $\langle \mathbf{v},\mathbf{u}\rangle = \langle \mathbf{u},\mathbf{v}\rangle$ since $\mathbf{u}$ and $\mathbf{v}$ are real vectors.

If $\mathbf{u} = \mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ then $\mathbf{u}$ is a scaled multiple of $\mathbf{v}$. For the other direction suppose $\mathbf{u} = c\mathbf{v}$ for some $c \in \mathbb{R}$. Then $\langle \mathbf{u},\mathbf{v}\rangle = \langle c\mathbf{v},\mathbf{v}\rangle = c\langle \mathbf{v},\mathbf{v}\rangle \implies \mathbf{u}^{\parallel} = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} = \frac{c\langle \mathbf{v},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v} = c\mathbf{v} = \mathbf{u}$.

Problem 4
For an invertible matrix $\mathbf{A}$ show that $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$ where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$.

Solution: We have $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ so $|\mathbf{A}\mathbf{A}^{-1}| = |\mathbf{A}| \cdot |\mathbf{A}^{-1}| = |\mathbf{I}| = 1 \implies |\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$, where we use the fact that the determinant factors over products and that $|\mathbf{A}| \neq 0$ since $\mathbf{A}$ is invertible.

Problem 5

Solve the following vector/matrix calculus problems. In all of the below, $\mathbf{x}$ and $\mathbf{w}$ are column vectors (i.e. $n \times 1$ vectors). It may be helpful to refer to *The Matrix Cookbook* by Petersen and Pedersen, specifically sections 2.4, 2.6, and 2.7.

(a) Let $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$. Find $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{x})$.

Hint: As a first step, you can expand $\mathbf{x}^T\mathbf{x} = (x_1^2 + x_2^2 + ... + x_n^2)$, where $\mathbf{x} = (x_1, ..., x_n)$.

(b) Let $f(\mathbf{w}) = (1 - \mathbf{w}^T\mathbf{x})^2$. Find $\nabla_{\mathbf{w}} f(\mathbf{w}) = \frac{\delta}{\delta \mathbf{w}} f(\mathbf{w})$.

(c) Let $\mathbf{A}$ be a symmetric n-by-n matrix. If $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{w}^T\mathbf{x}$, find $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{x})$.

Solution:

(a)

$$\begin{aligned}
\nabla(\mathbf{x}^T\mathbf{x}) &= \nabla(x_1^2 + x_2^2 + \ldots + x_n^2) \\
&= \nabla(x_1^2 + x_2^2 + \ldots + x_n^2) \\
&= \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} \\
&= 2\mathbf{x}
\end{aligned}$$

(b) By the chain rule:

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{w}} &= 2(1 - \mathbf{w}^T\mathbf{x})\frac{\partial}{\partial \mathbf{w}}(1 - \mathbf{w}^T\mathbf{x}) \\
&= -2(1 - \mathbf{w}^T\mathbf{x})\mathbf{x}
\end{aligned}$$

(c) The partial of $\mathbf{x^T A x}$ with respect to x_i is:

$$\begin{aligned}
\frac{\partial}{\partial x_i}\mathbf{x^T A x} &= \frac{\partial}{\partial x_i}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_i x_j \\
&= \sum_{k \neq i} a_{ik}x_k + \sum_{j \neq i} a_{ji}x_j + 2a_{ii}x_i \\
&= \sum_{k=1}^{n} a_{ik}x_k + \sum_{j=1}^{n} a_{ji}x_j \\
&= \sum_{k=1}^{n} a_{ik}x_k + \sum_{k=1}^{n} a_{ik}x_k \ \text{ since } A \text{ is symmetric} \\
&= 2\sum_{k=1}^{n} a_{ik}x_k
\end{aligned}$$

This is the ith row that results from multiplying $2Ax$. Thus $\frac{\partial}{\partial \mathbf{x}}\mathbf{x^T A x}$ is $2Ax$.

Since $\frac{\partial}{\partial \mathbf{x}}\mathbf{w^T x}$ is $\mathbf{w}$, the total answer is:

$$\mathbf{Ax} + \mathbf{w} \tag{2}$$

Problem 6

In her most recent work-from-home shopping spree, Nari decided to buy several house plants. She would like for them to grow as tall as possible, but needs your calculus help to understand how to best take care of them.

(a) After perusing the internet, Nari learns that the height y in mm of her Weeping Fig plant can be directly modeled as a function of the oz of water x she gives it each week:

$$y = -3x^2 + 72x + 70$$

Is this function concave, convex, or neither? Explain why or why not.

(b) Solve analytically for the critical points of this expression. For each critical point, use the second-derivative test to identify if each point is a local max, global max, local min, or global min.

(c) How many oz per week should Nari water her plant to maximize its height? With this much water how tall will her plant grow?

(d) Nari also has a Money Tree plant. The height y in mm of her Money Tree can be directly modeled as a function of the oz of water x she gives it per week:

$$y = -x^4 + 16x^3 - 93x^2 + 230x - 190$$

Is this function concave, convex, or neither? Explain why or why not.

Solution:

(a) It is concave since the 2nd derivative is $y'' = -6 < 0$.

(b) The first derivative is $y' = -6x + 72 = -6(x - 12)$. We have $y' = 0$ if and only if $x = 12$, so $x = 12$ is the only critical point. Since $y' > 0$ for $x < 12$ and $y' < 0$ for $x > 12$ we know that $x = 12$ is a local maximum. Since y is concave ($y'' < 0$) the point $x = 12$ is the global maximum.

(c) She should give her plant 12 oz of water a week for it to achieve the maximum height of 502 mm.

(d) Neither, the 2nd derivative is $y'' = -12x^2 + 96x - 186$, which is negative and positive depending on x.

Problem 7

Solve the following:

(a) Verify that $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$.

(b) Verify that $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

(c) Verify that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(d) Verify that $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y)$

(e) Suppose that $X_1, ..., X_n$ are i.i.d., scalar random variables with mean μ and variance σ^2. Let $\bar{X}$ be the mean $\frac{1}{n}\sum_i^n X_i$. Find $\mathbb{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

Solution:

(a) We have

$$\begin{aligned}\mathbb{E}[aX+b] &= \sum_{x\in X}(ax+b)\Pr(x)\\ &= a\sum_x x\Pr(x) + b\sum_x \Pr(x)\\ &= a\mathbb{E}[x]+b\end{aligned}$$

(b) Let $\mathbb{E}[X]=\mu$. Then applying the result from part a, we have

$$\begin{aligned}\mathrm{Var}(aX+b) &= \sum_{x\in X}\left(ax+b-\mathbb{E}[aX+b]\right)^2\Pr(X)\\ &= \sum_{x\in X}\left(ax+b-a\mu-b\right)^2\Pr(X)\\ &= \sum_{x\in X}a^2(x-\mu)^2\Pr(X)\\ &= a^2\mathrm{Var}(X)\end{aligned}$$

(c) Let $\mathbb{E}[X]=\mu$. Then we have

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x\in X}(x-\mu)^2\Pr(x)\\ &= \sum_x\left(x^2-2x\mu+\mu^2\right)\Pr(x)\\ &= \sum_x x^2\Pr(x) - 2\mu\sum_x x\Pr(x) + \mu^2\sum_x\Pr(x)\\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2\\ &= \mathbb{E}(X^2) - \mu^2\end{aligned}$$

(d) Using the result from part c, we have that

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2\\ &= \mathbb{E}[X^2+2XY+Y^2] - (\mathbb{E}[X]+\mathbb{E}[Y])^2\\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 + 2(\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y])\\ &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)\end{aligned}$$

(e) We have

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_i^n X_i\right) = \frac{1}{n}\sum_i^n\mathbb{E}(X_i) = \frac{1}{n}\sum_i^n\mu = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_i^n X_i\right) = \frac{1}{n^2}\sum_i^n\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_i^n\sigma^2 = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

Problem 8

Suppose $X_1, X_2, X_3, \dots X_n \overset{\text{iid}}{\sim} \text{Unif}[0,1]$. What is the distribution of

(a) (X_1, X_2)

(b) $X_1 + X_2$

(c) $\sum_{i=1}^n X_i$

(d) $\sum_{i=1}^n X_n$

Feel free to give just a sketch or a qualitative description of each case (no need for formal derivations). What do you notice about the sum in part (c) as $n \to \infty$? How does it differ from the sum in part (d)?

Solution

(a) Because X_1 and X_2 are independent, their joint distribution is simply the product of their individual distributions — which is 1, over the unit square.

(b) The sum of two independent uniform distributions forms a triangular distribution with a maximum density of 1 at $x_1 + x_2 = 1$, and tapers to zero symmetrically at either end $x_1 + x_2 = 0$ and $x_1 + x_2 = 2$. Note, the density of $X_1 + X_2$ is not simply 2. One way to intuit this is to think about the sum of two independent dice rolls, which will tend to result closer to 7 rather than uniformly covering the whole possible range from 2 through 12. *The Settlers of Catan* depends on precisely this property.

(c) As you add more and more uniform random variables, one can intuitively imagine that the sum will concentrate around $n/2$, and taper off the ends at n and 0. Indeed, the distribution of the sum will look more Gaussian as $n \to \infty$, a consequence of the Central Limit Theorem.

(d) This quantity is equivalent to nX_n, which is distributed uniformly on the interval $[0, N]$. Note, this differs crucially from the quantity in part (c) because we sum a single random variable, rather than sum over a collection of independent random variables.

Problem 9

Prove or come up with counterexamples for the following statements:

(a) Random variables A and B are conditionally independent given C. Does this imply that A and B are (unconditionally) independent?

(b) Random variables A and B are independent. Does this imply that A and B are conditionally independent given some random variable C?

Solution: (a) No! Suppose we have a fair coin C_1 and an unfair coin C_2 that has Heads on both sides. We will select 1 coin and then flip the coin twice. Let C be the event that we select C_1. Let A be be the event that first flip lands Heads and let B be the event that the second flip lands Heads. Given C we have that A and B are two separate flips of a fair coin, and so the flips are independent given C. However, suppose we do not know which coin has been selected. Then given A has occurred the probability of selecting coin C_1 is $1/3$ and that of selecting coin C_2 is $2/3$. But then $\mathbb{P}(B|A) = \mathbb{P}(B|C_1)\mathbb{P}(C_1|A) + \mathbb{P}(B|C_2)\mathbb{P}(C_2|A) = (1/2) \times 1/3 + (1) \times 2/3 = 5/6 \neq 3/4 = \mathbb{P}(B)$, so A and B are not independent.

(b) No! First, consider two fair, independent coin flips A, B. Let C be the event that $A = B$. On their own, A and B are independent but given C we can determine B from A or A from B (they are either perfectly correlated or perfectly anti-correlated), so A and B are not conditionally independent given C.

Problem 10

Consider the following:

(a) Your child has been randomly selected for Type I diabetes screening, using a highly accurate new test that boasts of a false positive rate of 1% and a false negative rate of 0%. The prevalence of of Type I diabetes in children is approximately 0.228%. Should your child test positive, what is the probability that they has Type I diabetes?

(b) Should you be concerned enough to ask for further testing or treatment for your child?

(c) Later, you read online that Type I diabetes is 6 times more prevalent in prematurely born children. If this statistic is true, what is the probability that your child, who is prematurely born, has Type I diabetes?

(d) Given the new information, should you be concerned enough to ask for further testing or treatment for your child?

Solution:

(a) Let D be the event that your child has diabetes, and $+$ be the event that your child tests positive. Then applying Bayes' Rule,

$$\begin{aligned}
\Pr(D|+) &= \frac{\Pr(+|D)\Pr(D)}{\Pr(+|D)\Pr(D) + \Pr(+|\overline{D})\Pr(\overline{D})} \\
&= \frac{(1)(0.00228)}{(1)(0.00228) + (0.01)(1 - 0.00228)} \\
&= 0.0223
\end{aligned}$$

Your child has a **2.23%** chance of having diabetes.

(b) This depends on your personal taste for risk. If we assume your child has a prior probability for diabetes that matches the global average, then the chance that they have diabetes is negligible, and further testing is not required. However, if you suspect your child has an elevated risk for diabetes as compared to the global average (and certainly, seeing a positive test result may retroactively bias your prior), then the chance that your child has diabetes may be considerably higher. (If it were *my* child, I would definitely request further testing, whatever the population statistics.) We explore an example of the latter case in the next section.

(c) If your child was born premature, then $\Pr(D) = 6 \cdot 0.00228 = 0.0137$. Substituting this new prior, applying the same Bayes' calculation as above shows us that the child's chance of diabetes is now **12.1%**.

(d) A small change in prior assumptions results in a significant change in the child's chance for diabetes. One goal of this exercise is to demonstrate the sensitivity of Bayesian analysis to the choice of priors. Before you apply any technique from a Bayesian toolbox, remember to always weigh your choice of priors carefully, and examine their impact on the final analysis.

Problem 11

During shopping week, you're trying to decide between two classes based on the criteria that the class must have a lenient grading system. You hear from your friends that one of these classes is rumored to award grades lower than the work merits 35% of the time while the other awards lower grades 15% of the time. However, the rumor doesn't specify which class has harsher grading. So, you decide to conduct an experiment: submit an assignment to be graded.

Fortunately, both classes offer an optional Homework 0 that is graded as extra credit. Unfortunately, you only have time to complete the problem set for just one of these classes.

Suppose you randomly pick the Homework 0 from Class A to complete and suppose that you received a grade that you believe is lower than the quality of your work warrents. Based on this evidence, what is the probability that Class A has the harsher grading system? Which class should you drop based on the results of your experiment (or do you not have sufficient evidence to decide)?

Solution: There are many ways to approach this problem. Your answer may vary.

We take a vanilla Bayesian approach. Let A be the event that class A is harsher, and let B be the event that class B is harsher. Let E be the event that your HW0 was graded lower than expected. Suppose we have a uniform prior where $\Pr(A) = \Pr(B) = 0.5$. Then applying Bayes' Rule,

$$\begin{aligned}\Pr(A|E) &= \frac{\Pr(E|A)\Pr(A)}{\Pr(E|A)\Pr(A) + \Pr(E|B)\Pr(B)} \\ &= \frac{(0.35)(0.5)}{(0.35)(0.5) + (0.15)(0.5)} \\ &= 0.7\end{aligned}$$

According to this analysis, there is a 70% chance that course A is more difficult.

Is this enough to drop A and take B? That depends on your personal taste. 70% is not that wide a margin, and more evidence would be helpful in making a confident judgement. Additional factors like your impression of the instructor, syllabus, course reviews, and so forth, should all influence your priors, so a 50-50 split may not be realistic. All this is to say, the final judgement is subjective, and will vary based on your priorities and assumptions.

Problem 12

A random point (X, Y, Z) is chosen uniformly in the ball

$$B = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$$

(a) Find the joint PDF of (X, Y, Z).

(b) Find the joint PDF of (X, Y).

(c) Write an expression for the maginal PDF of X, as an integral.

Solution:

(a) Let $B = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$ be the closed unit ball. The volume of B is $\mathrm{vol}(B) = \int_B 1dx\ dy\ dz = \frac{4}{3}\pi$. Since the distribution of (X, Y, Z) is uniform over B the PDF is then

$$f(x, y, z) = \frac{3}{4\pi} \cdot \chi((x, y, z) \in B)$$

where $\chi((x, y, z) \in B) = \begin{cases} 1 & (x, y, z) \in B \\ 0 & (x, y, z) \notin B \end{cases}$

(b) Let $C = \{(x, y) \mid x^2 + y^2 \leq 1\}$ be the unit circle. We have

$$f(x, y) = \int_{\mathbb{R}} f(x, y, z)\ dz = \frac{3}{4\pi} \int_{\mathbb{R}} \chi((x, y, z) \in B)\ dz = \frac{3}{4\pi} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} \chi((x, y) \in C)\ dz =$$

$$\frac{3}{2\pi} \sqrt{1 - x^2 - y^2} \cdot \chi((x, y) \in C).$$

(c) We have

$$f(x) = \int_{\mathbb{R}} f(x, y)\ dy = \frac{3}{2\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1 - x^2 - y^2} \cdot \chi(x \in [-1, 1])\ dy.$$

Problem 13

Suppose we randomly sample a Harvard College student from the undergraduate population. Let X be the indicator of the sampled individual concentrating in computer science, and let Y be the indicator of their working in the tech industry after graduation.

Suppose that the below table represented the joint PMF of X and Y:

	$Y = 1$	$Y = 0$
$X = 1$	$\frac{10}{100}$	$\frac{5}{100}$
$X = 0$	$\frac{15}{100}$	$\frac{70}{100}$

(a) Calculate marginal probability $P(Y = 1)$. In the context of this problem, what does this probability represent?

(b) Calculate conditional probability $P(Y = 1|X = 1)$. In the context of this problem, what does this probability represent?

(c) Are X and Y independent? Why or why not?

Solution:

(a) We have $P(Y = 1) = P(Y = 1, X = 1) + P(Y = 1, X = 0) = \frac{10}{100} + \frac{15}{100} = \frac{1}{4}$. This represents the probability that a Harvard student works in the tech industry after graduation.

(b) Similarly, we compute $P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 0) = \frac{10}{100} + \frac{5}{100} = \frac{3}{20}$. Then we compute $P(Y = 1|X = 1) = \frac{P(Y=1,X=1)}{P(X=1)} = \frac{10/100}{3/20} = \frac{2}{3}$. This represents the probability of a CS concentrator at Harvard working in the tech industry after graduation.

(c) X and Y are not independent since $P(Y = 1|X = 1) \neq P(Y = 1)$.

Credits: Problems 12 and 13 were inspired by Exercise 7.19 and Example 7.1.5 in Blitzstein & Hwang's "Introduction to Probability".