CS 6140 Spring 2019 - Homework 1

Tushar Sharma

Submitted on 02/01/2019

Problem 1. (5 points) Let (Ω, A, P) be a probability space. Using only axioms of probability and basic set operations, show that for any two sets A $\subseteq \Omega$ and B $\subseteq \Omega$, it holds that P(A $\bigcup$ B) = P(A) + P(B) − P(A $\bigcap$ B). Solution:

Rewriting the set (A $\bigcup$ B) by splitting in 2 disjoint sets:

A $\bigcup$ B = B$\bigcup$(A $\bigcap$ B^c)

Now B $\bigcap$ (A $\bigcap$ B^c) = ϕ,

So, using Axiom 2 of probabilities [Reference1], we can write above equation as:

P(A $\bigcup$ B) = P(B) + P(A $\bigcap$ B^c) [say result-1]

We can write set A as , A = (A$\bigcap\Omega$)

$\Rightarrow$A = A$\bigcap$(B $\bigcup$ B^c) [Since Ω = B$\bigcup B^c$, by the definition of set complementation, Reference 3]

$\Rightarrow$A = (A$\bigcap$B) $\bigcup$ (A $\bigcap$ B^c) [Using Distributivite property of sets, Reference 3]

$\Rightarrow$P(A) = P(A$\bigcap$B) + P(A $\bigcap$ B^c)

$\Rightarrow$P(A$\bigcap B^c$) = P(A) - P(A $\bigcap$ B) [say result-2]

Substituting result-2 in result-1 we get the following:

P(A $\bigcup$ B) = P(B) + P(A) - P(A $\bigcap$ B)

Problem 2. (5 points) Let (Ω, A, P) be a probability space and A$\subseteq\Omega$ and B$\subseteq\Omega$ any two subsets of . Prove the following expression or provide a counterexample if it does not hold P(A) = P(A|B) + P(A|B^c), where A^c is the complement of A. Solution:

This expression does not hold true. A counterexample is as follows:

Ω={1,2,3,4}, A={2,3}, B={1,2}$\Rightarrow B^c$ ={3,4}

A $\bigcap$ B = {2}, A $\bigcap$ B^c = {3}

Given above information, P(A|B) = P(A $\bigcap$ B) / P(B) = $(\frac{1}{4})/(\frac{2}{4}) = \frac{1}{2}$

and P(A|B^c) = P(A $\bigcap$ B^c) / P(B^c) = $\frac{1}{2}$

therefore, P(A|B) + P(A|B^c) = $\frac{1}{2} + \frac{1}{2} = 1$

but P(A) = $\frac{2}{4} = \frac{1}{2}$

Since the equality of LHS and RHS of the expression does not hold true for considered (Ω, A, P), hence this expression does not holds.

Problem 3. (15 points) Let X be a random variable on X = {a, b, c} with the probability mass function p(x). Let p(a) = 0.1, p(b) = 0.2, and p(c) = 0.7 and some function f(x) be

$$\text{f(x)}\begin{cases} 10 & x = a \\ 5 & x = b \\ 10/7 & x = c \end{cases}$$

a) (5 points) What is E[f(X)]?

b) (5 points) What is E[1/p(X)]?

c) (5 points) For an arbitrary finite set X with n elements and arbitrary p(x) on X, what is E[1=p(X)] Solution:

a) E[f(x)] = $\sum f(x).p(x)$ ∀x∈X [Reference 1]
=a.p(a)+b.p(b)+c.p(c)
=0.1*10+0.2*5+0.7*(10/7)
=1+1+1 = 3
b) E[1/p(x)]
Here we will replace f(x) by 1/p(x) and get the following equation:
$\sum 1/p(x).p(x) \forall \mathbf{x} \in \mathbf{X}$
=(1/0.1*0.1)+(1/0.2*0.2)+(1/0.7*0.7)
=3
c) E[1/p(x)] for n elements:
$\sum 1/p(x).p(x) \forall \mathbf{x} \in \mathbf{X}$
=(1/$\mathrm{p}_1(x)$*$\mathrm{p}_1(x)$)+(1/$\mathrm{p}_2(x)$*$\mathrm{p}_2(x)$) + ...+(1/$\mathrm{p}_n(x)$*$\mathrm{p}_n(x)$)
=1+1+1+...................n times = n

Problem 4. (15 points) A biased four-sided die is rolled and the down face is a random variable X described by the following pmf

$$p(x) = \begin{cases} x/10 & x = 1, 2, 3, 4 \\ 0 & otherwise \end{cases}$$

Given the random variable X a biased coin is flipped and the random variable Y is 1 or zero according to whether the coin shows heads or tails. The conditional pmf is

$p(y|x) = (\frac{x+1}{2x})^y \ (1 - \frac{x+1}{2x})^{1-y}$**, where y ∈ {0, 1}.**

a) (5 points) Find the expectation E[X] and variance V [X].

b) (5 points) Find the conditional pmf p(x|y).

c) (5 points) Find the conditional expectation E[X|Y = 1]; i.e., the expectation with respect to the conditional pmf $\mathbf{p}_{X|Y}$ (X|1): Solution:

a) We will use the expectation formula E[f(x)] = $\sum f(x).p(x)$∀x∈**{1,2,3,4}** [Reference 1]

E[X] = $\sum x.\frac{x}{10}$

= 1.(1/10) + 2.(2/10) + 3.(3/10) + 4.(4/10)

= 0.1+0.4+0.9+1.6

= 3

b) $p(x|y) = \frac{p(y,x)}{p(y)}$ [Conditional Probability from Reference 1]

We are given, $p(y|x) = (\frac{x+1}{2x})^y\ (1-\frac{x+1}{2x})^{1-y}$ and $p(x)$ = x/10 ∀ x∈{1,2,3,4}

also, $p(y|x) = \frac{p(y,x)}{p(x)}$

⇒ p(y,x) = p(y|x).p(x) = $(\frac{x}{10})(\frac{x+1}{2x})^y\ (1-\frac{x+1}{2x})^{1-y}$ [Multiplication rule, Reference 1]

Using, the above equation to form the joint probability mass function as below:

$p(y,x)$	$x=1$	$x=2$	$x=3$	$x=4$
$y=0$	0	1/20	2/20	3/20
$y=1$	1	3/20	4/20	5/20

Marginalizing JPD of (y,x) over x we will get p(y):

$p(y,x)$	$x=1$	$x=2$	$x=3$	$x=4$	$p(y)$	
$y=0$	0	1/20	2/20	3/20	6/20	∀ x∈{1,2,3,4} and y∈{0,1}
$y=1$	1	3/20	4/20	5/20	14/20	

therefore, $p(y) = \begin{cases} 3/10 & y=0 \\ 7/10 & y=1 \end{cases}$

Now since we have p(y), using $p(x|y) = \frac{p(y,x)}{p(y)}$ we can get the conditional distribution for p(x|y) as follows:

$p(x\|y)$	$x=1$	$x=2$	$x=3$	$x=4$	
$y=0$	0	1/6	1/3	1/2	
$y=1$	1/7	3/14	2/7	5/14	∀ x∈{1,2,3,4} and y∈{0,1}

c) E[X|Y = 1] = $\sum x.p(x|y=1)$∀ x∈{1,2,3,4} [Conditional Expectation, Reference 1]

E[X|Y = 1] = 1.(1/7) + 2.(3/14) + 3. (2/7) + 4.(5/14)

= 40/14 ≈ 2.8571

Problem 5. (10 points) Let X, Y and Z be discrete random variables defined as functions on the same probability space (Ω, A, P). Prove or disprove the following expression P(X = x | Y = y) = ∑P(X = x | Y = y, Z = z). P(Z = z | Y = y) , where Z is the sample space defined by the random variable Z. Solution:

LHS = P(X = x | Y = y) = $\frac{P(X,Y)}{P(Y)}$

RHS = ∑P(X = x | Y = y, Z = z). P(Z = z | Y = y)∀**z**∈**Z**

=$\sum \frac{P(X,Y,Z)}{P(Y,Z)}.\ \frac{P(Y,Z)}{P(Y)}$∀**z**∈**Z**

= $\sum \frac{P(X,Y,Z)}{P(Y)}$∀**z**∈**Z**

Now, $\sum$P(X,Y,Z) $\forall \mathbf{z} \in \mathbf{Z}$ is nothing but marginal distribution of joint distribution of X,Y,Z over all Z [Joint Probability Distributions, Reference 1]

therefore, $\sum$P(X,Y,Z) $\forall \mathbf{z} \in \mathbf{Z}$ = P(X,Y)

RHS becomes $\frac{P(X,Y)}{P(Y)}$

=P(X = x|Y = y) = LHS

Problem 6. (10 points) Let X and Y be random variables. Prove or disprove the following formula V [X + Y] = V [X] + V [Y] + 2Cov[X, Y], where V [X] is the variance of X and Cov[X, Y] is the covariance between X and Y . Solution:

By the definition of Variance of a random variable (say X) we have the following equation:

V[X] = E[(X-E[X])^2] [From Reference 1]

therefore,

V[X+Y] = E[((X+Y) - E[(X+Y)])^2] (Expanding square terms)

= E[(X+Y)^2+ E^2(X+Y) - 2(X+Y)E[X+Y]] (Expanding square terms again)

= E[X^2+ Y^2+ 2X.Y + E^2[X] + E^2[Y] + 2E[X].E[Y] - 2X.E[X] - 2X.E[Y] - 2Y.E[X] - 2Y.E[Y]]

Regrouping above equation,

= E[(X^2+ E^2[X] - 2.X.E[X]) + (Y^2 + E^2[Y] - 2Y.E[Y]) + 2X.Y - 2X.E[Y] - 2Y.E[X] + 2E[X].E[Y]]

= E[(X-E[X])^2 + (Y-E[Y])^2 + 2(X.Y - X.E[Y] - Y.E[X] + E[X].E[Y])]

= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[X(Y - E[Y]) - E[X](Y - E[Y])]

= V[X] + V[Y] + 2E[(X-E[X])(Y-E[Y])]

= V[X] + V[Y] + 2Cov[X,Y] [From Reference 1]

Problem 7. (15 points) Let Y0 and Y1 be two random variables and Z $\sim$ Bernoulli(α). Let X be a random variable defined as X = ZY1 + (1 $-$ Z)Y0. Assuming that the probability density functions of Y0 and Y1 exist, show that the density of X is a mixture of the densities of Y1 and Y0 with α and 1 $-$ α as the mixing proportions, respectively. Solution:

Given is Z ~ Bernoulli(α)

$\Rightarrow$ E[Z] = 0(1 - α) + 1(α)

$\Rightarrow$ E[Z] = α

Also,

X = Z.Y_1+ (1 - Z).Y_0

$\Rightarrow$ E[X] = E[Z.Y_1+ (1 - Z).Y_0]

$\Rightarrow$ E[X] = E[Z.Y_1] + E[(1 - Z).Y_0]

Making a safe assumption that Y_1 and Y_0 are independent of Z leads to following equation,

E[X] = E[Z].E[Y_1] + E[(1 - Z)].E[Y_0]

$\Rightarrow$ E[X] = α.E[Y_1] + (E[1]-E[Z])E[Y_0]

$\Rightarrow$ E[X] = α.E[Y_1] + (1 - α)E[Y_0] (substituting E[Z] = α everywhere)

Now assigning w_1 = α and w_0 =(1- α) (Note that $\sum w = 1$ which is a property for proportions of mixtures),

we rewrite last equation as follows:

E[X] = w_1.E[Y_1] + w_0 E[Y_0]

or E[X] =$\sum w_i$E[Y_i] ∀**i**∈**{0,1}**

This satisfies the expectation property of mixture distributions [Mixture Distributions from Reference 1]

Hence, we can say that p(x) = $\sum w_i$p(y)$_i$ ∀**i**∈**{0,1}** where $\sum w = 1$

with mixing proportion of p(y)$_1$ as α and of p(y)$_0$ as (1 - α).

Problem 8. (15 points) Consider a medical diagnosis problem in which there are two alternative hypotheses: (1) the patient has a particular form of disease, and (2) the patient does not have that disease. The available data is from a particular laboratory test with two possible outcomes: + (positive) and - (negative). We have prior knowledge that over the entire population of people, only 0.8% have this disease. Also, the lab test is imperfect. It returns a correct positive result in 97% of cases in which the disease is actually present and a correct negative result in 98% of cases in which the disease is not present. In other cases, the test returns the opposite result.

a) (5 points) Suppose we pick a random person from the street and perform the lab test. The lab test then returns a positive result. What is the probability that the patient has disease given the positive test.

b) (10 points) Suppose the doctor is concerned after the first positive result and decides to perform another test (tests are independent of one another). The second test comes as positive. What is the probability that the patient has disease given two independent positive tests? Solution:

Let D be the set of diseased population and ND be the set of non diseased population.

we are given the following information:

P(+|D) = 0.97, P(− |ND) = 0.98, P(− |D) = 1- P(+|D) = 0.03, P(+|ND) = 1- P(− |ND) = 0.02

P(D) = 0.008, P(ND) = 1-0.008= 0.992

Also, P(+,D) = P(+|D).P(D) = 0.97 * 0.008 = 0.00776

P(+) = P(+|D).P(D) + P(+|ND).P(ND) [Total probability rule]

= 0.97*0.008 + 0.02 * 0.992 = 0.0276

a) Since we are given the test reult as + and probability of disease is to be determined given test evidence,

We need to find P(D|+) = $\frac{P(+,D)}{P(+)}$= 0.00776/0.0276 = 0.2811

b) Now we need determine the chances of disease given 2 independent tests resulting in +, therefore, P(D|+,+) needs to be determined.

P(D|+,+) = $\frac{P(+,+|D)P(D)}{P(+,+)}$

Here we will consider conditional indpependence of two tests since the doctor is worries the person has diseas and then goes for the second test.

P(+,+|D) = P(+|D).P(+|D) [Using conditional independence, Reference 1]

also, P(+,+) = P(+,+|D).P(D) + P(+,+|ND)P(ND) (Marginalizing P(+,+) over disease/no disease)

Therefore P(D|+,+)= $\frac{P(+|D).P(+|D).P(D)}{P(+|D)P(+|D)P(D)+P(+|ND)P(+|ND)P(ND)}$ [Here we used conditional independence for P(+,+|ND)]

= $\frac{0.97*0.97*0.008}{0.97*0.97*0.008+0.02*0.02*0.992}$ = $\frac{0.0075}{0.0075+0.0004}$= 0.949

Problem 9. (10 points) Let D = {x$_i$} ∀i∈{1,...,n} be an i.i.d. sample from

p(x) = $\begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & otherwise \end{cases}$**. Determine θ_{ML} - the maximum likelihood estimate of θ_0.** Solution:

Note that at x = θ the function is not continuous and hence not differentiable,

So we will assume x>θ for all our calculations.

We can write likelihood function as,

p(D|θ) = $\prod e^{-(x_i-\theta)}$ [Reference 2]

and loglikelihood function as,

ln(p(D|θ)) = $\sum -(x_i - \theta)$ [Reference 2]

or ll = $\sum \theta - \sum x_i$

⇒ ll = nθ

therefore,

$\frac{d(ll)}{d\theta} = n$ which is always greater than 0.

Since derivative of log-likelihood function did not give an important result, let us try working with p(x)

Taking log of p(x),

ln(p(x)) = -(x - θ)

= θ - x

$\frac{dln(p(x))}{dx}$= -1

Now we get first derivative of p(x) = -1 which is less than 0 and second derivative is 0.

This suggests the function of ln(p(x)) is a strictly decreasing function.

p(x) will also follow the similar decreasing trend (since function follows same trend as log of the function as discussed in class on 02/01).

Hence we can conclude that starting point of this distribution will be the highest point.

This means for the minimum value of x we will get Maximum p(x|θ), i.e. Likelihood of θ is maximum at minimum value of x.

Thus, θ_{ML}= min{x$_i$}

Problem 10. (20 points) High dimensional spaces.

a) (10 points) Show that in a high dimensional space, most of the volume of a cube is concentrated in corners, which themselves become very long "spikes". Hints: compute the ratio of the volume of a hypersphere of radius a to the volume of a hypercube of side 2a and also the ratio of the distance from the center of the hypercube to one of the corners divided by the perpendicular distance to one of the edges. Solution:

a) From reference we get the following formula for n dimensional and R radius hypersphere:

$V_n(R) = (\frac{\pi^{\frac{n}{2}}}{\Gamma\left((\frac{n}{2})+1\right)})R^n$

Also, volume for n-dimensional hypercube with dimensions 2R is (2R)^n

taking ratio of $V_{hypersphere}/V_{hypercube}$,

$\frac{V_{hypersphere}}{V_{hypercube}} = \frac{\pi^{\frac{n}{2}}}{\Gamma\left((\frac{n}{2})+1\right)} R^n/2^n R^n$

V$_{ratio}$ $= \frac{\pi^{\frac{n}{2}} 2^{-n}}{\Gamma\left((\frac{n}{2})+1\right)}$

For higher dimensional space where n will become high,

lim (V$_{ratio}$) when n $->\infty = 0$ [Reference 4 (Wolfram calculator)]

We can say here the volume inside the hypersphere is infinitesimally small compared to volume of hypercube.

Or the most of the volume of hypercube is outside the hypersphere (or concentrated in the corners).

Let D be the distance from centre of hypercube of side 2R to one of its corners.

then, D $= \sqrt{(2a)^2 + (2a)^2 + (2a)^2 + ...n - times} = \sqrt{n(2a)^2} = 2a\sqrt{n}$ [Pythagoras' Theorem]

Let d be the perpendicular distance from centre of hypercube to an edge.

d=2a

ratio (D/d)= $\frac{2a\sqrt{n}}{2a} = \sqrt{n}$

therefore, lim D/d as n $->\infty = \infty$

This means the growth of centre to corner distance becomes very high compared to growth of centre to edge distance.

Higher growth rate will result in making very long spikes on the corners of the hypercube where most of the volume will get accumulated.

b) (10 points) Show that for points which are uniformly distributed inside a sphere in d dimensions where d is large, almost all of the points are concentrated in a thin shell close to the surface. Hints: compute the fraction of the volume of the sphere which lies at values of the radius between a–ε and 0 <ε < a; Evaluate this fraction for ε= 0.01a and also for ε= 0.5a for d $\in$ {2, 3, 10, 100}. Solution:

b) Volume of a hypersphere or radius R and d-dimension is:

$V_d(R) = (\frac{\pi^{\frac{d}{2}}}{\Gamma\left((\frac{d}{2})+1\right)})R^d$

$=C_dR^d$ [where $C_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left((\frac{d}{2})+1\right)}$]

Calculating the ratio of volume between ε to a, and 0 to a

$\frac{V_d(a)-V_d(a-\varepsilon)}{V_d(a)} = \frac{C_da^d-C_d(a-\varepsilon)^d}{C_da^d} = \frac{a^d-(a-\varepsilon)^d}{a^d} = 1 - \frac{(a-\varepsilon)^d}{a^d}$

Here is the table for this fraction for ε = 0.01a and ε = 0.5a and d ∈**{2,3,10,100}**

d	2	3	10	100
$\in = 0.01a$	0.0199	0.029701	0.0956179	0.633968
$\in = 0.5a$	0.75	0.875	0.999023	1

It is evident from value pair (ε=0.5a and d=100) that as dimension gets bigger the volume that lies closer to the surface increase significantly.

This proves that most of the points in the hypersphere lies closer to the shell.

References

[1] Lecture Notes on Probability Theory

[2] Lecture Notes on Parameter Estimation

[3] Wikipedia page on Set Operations - https://en.wikipedia.org/wiki/Set_(mathematics)#Basic_operatio

[4] Limit Calculator on Wolfram - https://www.wolframalpha.com/input/?i=lim+(pi%5E(n%2F2)2%5E(-n)%2Fgamma((n%2F2%2B1)))+as+n+-%3E+inf