Jae Hong Lee
Prof. C.E. Tsourakakis
Boston University

March 1, 2024

— **CAS CS 365 - Midterm Practice 1**

1 Practice Problems

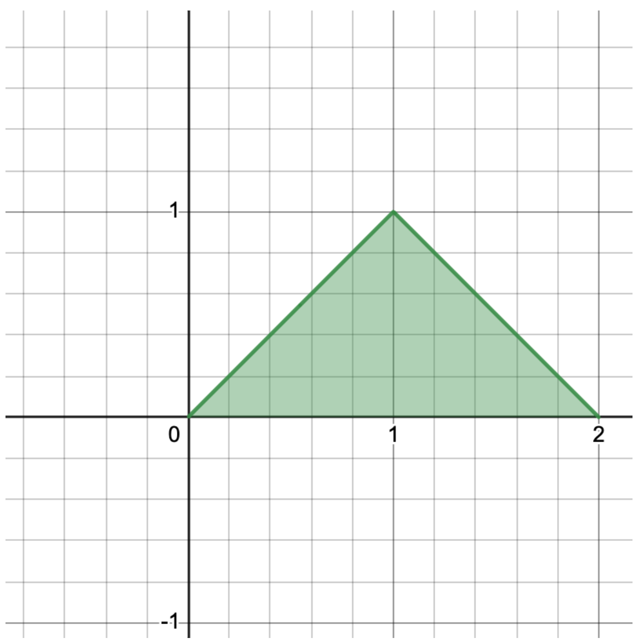
1.1 Problem 1.1

What is the distribution of the sum of two uniform variables in (0,1)?

Let's say two uniform variable X and Y. X and Y are uniform distribution $[X, Y \sim uniform(0,1)]$. Let's say there is another random variable $Z = X + Y$.

1.1.1 (i) Ideal 1

The PDF and CDF of sum of two uniform variables' distribution is as follows:



(a) The CDF graph of $Z = X + Y$

$Z = X + Y$, $Y = -X + Z$

1.1.2 (ii) Ideal 2

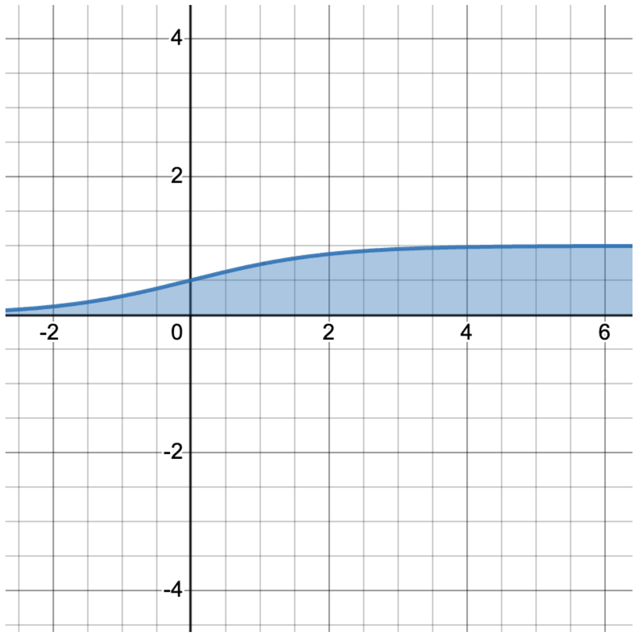
For Z smaller or equal to 1, $F_Z(z) = \int_0^z \int_0^{z-y} 1 \cdot dx \cdot dy = \int_0^z (z-y) \cdot dy = \left[zy - \frac{1}{2}y^2\right]_0^z = z^2 - \frac{1}{2}z^2 = \frac{1}{2}z^2$ The derivate of CDF of Z less or equal to 1 is $(\frac{1}{2}z^2)' = z$

For Z bigger than 1, $1 < Z \leq 2$ $F_Z(z) = \int_{z-1}^1 \int_{z-1}^{z-y} 1 \cdot dx \cdot dy = \int_{z-1}^1 (-y+1) \cdot dy = \left[-\frac{1}{2}y^2 + y\right]_{z-1}^1 = \frac{1}{2} - (-\frac{1}{2}(z-1)^2 + (z-1)) = -\frac{1}{2}z^2 + 2z - 2$ The derivative of CDF of Z bigger than 1 is $(-\frac{1}{2}z^2 - +2z - 2)' = 2 - z$

1.2 Problem 1.2

Verify that $F_X(x) = \frac{1}{1+e^{-x}}, x \in R$ is a valid distribution function(cdf). What is the corresponding pdf?

In order to verify that cdf is valid, we need to prove that cdf is monotonically increasing and showing non negativity. $F_x(x) = \frac{1}{1+e^{-x}}$, $x \in R$. From the cdf we can see the function as below



(a) The CDF graph of $F_x(x) = \frac{1}{1+e^{-x}}$

The graph is monotonically increase and it has non - negativity, so it is valid. PDF is derivative of CDF so

$$f_X(x) = F_X(x)' = (\frac{1}{1+e^{-x}})' = ((1+e^{-x})^{-1})'$$

$$= -1 \cdot (1+e^{-x})^{-2} \cdot (1+e^{-x})' = -1 \cdot (1+e^{-x})^{-2} \cdot (-1) \cdot (e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^2}$$

1.3 Problem 1.3

Two people each toss a coin n times. What is the probability that they will toss the same number of heads?

1.3.1 (i) Using binomial distribution

N coins toss and getting the same number of k heads with the probability of a fair coin, $\frac{1}{2}$. We know this is the binomial distribution. The first person getting k heads out of n coins is $\binom{n}{h} \cdot p^k \cdot (1-p)^{n-k}$. The second person getting k heads out of n coins' probability is also the same and these two people are independent, so the probability of these two people getting the same number k heads out of n coins is:

$$\binom{n}{h} \cdot p^k \cdot (1-p)^{n-k} \times \binom{n}{h} \cdot p^k \cdot (1-p)^{n-k} = (\binom{n}{h} \cdot (\frac{1}{2})^n)^2$$

The number of head can be 0 up to n heads, so the probability of these two people getting the same number of heads out of n coins is :

$$\sum_{i=0}^{n}(\binom{n}{i} \cdot (\frac{1}{2})^n)^2$$

1.3.2 (ii) the specialty of fair coin

1.4 Problem 1.4

Toss two dice. What is the probability of having 2 or 5 on at least one die?

The probability of having 2 or 5 on one die is $\frac{2}{6} = \frac{1}{3}$. There are two independent dice, so let's say the event of the first dice to have 2 or 5 A_1 and second die A_2. We are looking for the probability of having 2 or 5 on **at least** one die, so we are looking for $Pr(A_1 \cup A_2)$. $Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2) - Pr(A_1 \cap A_2) = \frac{1}{3} + \frac{1}{3} - (\frac{1}{3} \times \frac{1}{3}) = \frac{2}{3} - \frac{1}{9} = \frac{5}{9}$.

$\therefore \frac{5}{9}$

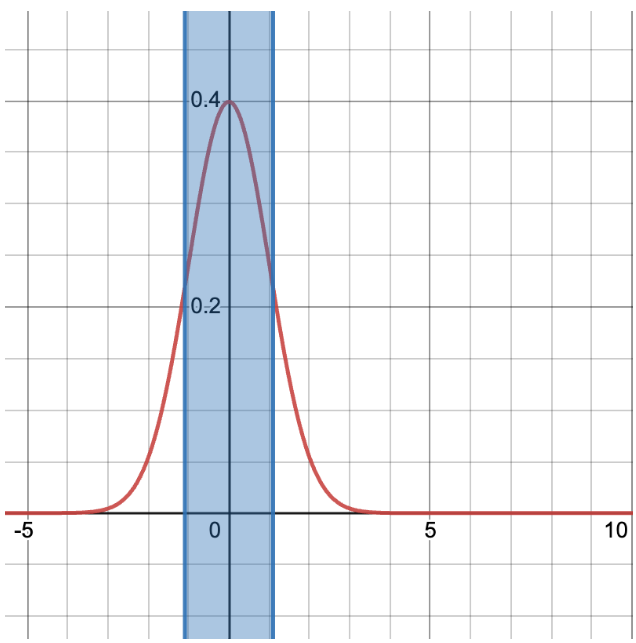
1.5 Problem 1.5

Suppose X is a continuous random variable with distribution F_X and pdf f_x. What is the pdf of X^2?

X is the continuous random variable with CDF of F_X and pdf of f_x. From the CDF of X^2, F_{x^2} let's say there is a random point u. The probability of u is $F_{X^2}(u) = Pr[X^2 \leq u] = Pr[-\sqrt{u} \leq X \leq \sqrt{u}]$.

X^2를 X의 형태로 바꾸어주는 게 키포인트!

$Pr[-\sqrt{u} \leq X \leq \sqrt{u}] = F_X(\sqrt{u}) - F_X(-\sqrt{u})$



(a) The CDF graph of $F_{X^2}(\sqrt{u}) - F_{X^2}(-\sqrt{u})$

The derivative of CDF is PDF and we figured out CDF of $F_{X^2} = F_X(\sqrt{u}) - F_X(-\sqrt{u})$.

$$(F_X(\sqrt{u}) - F_X(-\sqrt{u}))' = (F_X(u^{\frac{1}{2}}) - F_X(-u^{\frac{1}{2}}))' = f_x(u^{\frac{1}{2}}) \cdot (u^{\frac{1}{2}})' - f_x(-u^{\frac{1}{2}}) \cdot (-u^{\frac{1}{2}})'$$

$$=\therefore \frac{1}{2} \cdot (u^{-\frac{1}{2}}) \cdot f_x(u^{\frac{1}{2}}) - \frac{1}{2} \cdot (-u^{-\frac{1}{2}}) \cdot f_x(-u^{\frac{1}{2}}) = \sqrt{u} \cdot (f_x(\sqrt{u}) + f_x(-\sqrt{u}))$$

1.6 Problem 1.6

Compute E(X) and Var(X) if X is a random variable with the discrete uniform distribution on $N_0, N_0 + 1, ..., N_1$ (i.e., equal mass to each of those numbers) where $N_0 \leq N_1$.

The PDF of the discrete uniform distribution is $\frac{1}{(N_1 - N_0 + 1)}$

1.6.1 (i) E(X)

X is a discrete uniform random variable. The E(X) is the average of the range $N_0, N_0 + 1, ..., N_1$. The expectation E(X) $= \frac{N_0 + N_1}{2}$

1.6.2 (ii) Var(X)

Var(X) $= \frac{\sum_{i=N_0}^{N_1} (i - E(X))^2}{n}$

1.7 Problem 1.7

Suppose that 5% of men and .25% of women are color-blind. A person is chosen at random and that person is color-blind. What is the probability that the person is male? (Assume males and females to be in equal numbers).

With the given information, we can say that out of male, the probability of color-blind male is P(B|M) = 0.05 and out of female, the probability of color-blind women is $P(B|\bar{M}) = 0.0025$. The problem is asking the probability of the person is color-blind given that person is male $Pr(M|B)$.

From Bayes rule, $Pr(M|B) = \frac{Pr(B|M) \cdot Pr(M)}{Pr(B)}$. $Pr(B) = Pr(B|M) \cdot Pr(M) + Pr(B|\bar{M}) \cdot Pr(\bar{M}) = 0.05 \cdot 0.5 + 0.0025 \cdot 0.5 = 0.02625$.

So, $Pr(M|B) = \frac{0.05 \cdot 0.5}{0.02625} = \frac{20}{21} \approx 0.95238$.

The probability that the person is make is $\frac{20}{21} \approx 0.95238$

1.8 Problem 1.8

Let $x_1, ..., x_n$ be iid samples from uniform (0, θ) where θ is an unknown parameter. Estimate θ using (i) MoM (ii) MLE. What do you observe?

x_n is the uniform distribution of (0, θ). The pdf of $x_n = \frac{1}{\theta}$. And the expectation of X is $\frac{\theta}{2}$

1.8.1 (i) MoM

The method of moment takes the average of the observed samples as the expectation number of the model. so the E(X) of MoM is $\frac{\sum_1^n x_i}{n}$, which is also $\frac{\theta}{2}$. so θ is $2 \cdot \frac{\sum_1^n x_i}{n}$

1.8.2 (ii) MLE

We have the PDF of Random variable X is $\frac{1}{\theta}$. So the likelihood function is $\prod_1^n \frac{1}{\theta} = (\frac{1}{\theta})^n$. The log likelihood function is $log((\frac{1}{\theta})^n) = n \cdot log(\frac{1}{\theta})$. The derivative of the log likelihood function is $(n \cdot log(\frac{1}{\theta}))' = n \cdot \theta \cdot -\frac{1}{\theta^2} = -\frac{n}{\theta}$. The log MLE has to be 0 to maximize. So $-\frac{n}{\theta} = 0$. Lower the θ, $-\frac{n}{\theta}$ closer to 0. However, θ has to be bigger than the max value of x_i.

$\therefore \theta \geq max(x_i)$.

1.9 Problem 1.9

Let X, Y be independent random variable with common density function f. Prove that the density function of $Z = max(X, Y)$ is given by $f_Z(x) = 2f(x)P(X \leq x)$.

X and Y are independent random variable with common density function, so they have identical pdf $f_X(x) = f_Y(y)$ Let's say there is a random variable T and the probability to get T from Z distribution is $f_Z(T) = max(X,Y)(T)$. So the T on X and Y distribution cannot exceed Z distribution. so

$$f_Z(T) = Pr[z \leq T]$$

$$= Pr[x \leq T \; and \; y \leq T]$$

$$= Pr[x \leq T] \cdot Pr[y \leq T]$$

$$Pr[x \leq T]^2$$

This is the CDF of F_Z, so we need to derivative to get the pdf.

$$(Pr[x \leq T]^2)' = 2 \cdot f(x) \cdot P(X \leq T)$$

This is the same form as the problem asked to prove so the prove is done.

1.10 Problem 1.10

If U is a uniform random variable in [0, 1], what is the distribution of $\lfloor 100U \rfloor + 1$

Since U is the continuous uniform random variable. 100U is the a continuous random variable [0, 100]. when you floor it, then the continuous random variable change to the discrete random variable range of [0, 99]. adding 1 gives us the discrete uniform random variable [1, 100].

1.11 Problem 1.11

If U is a uniform random variable in $[0,1]$ and $0 < q < 1$, prove that $X = 1 + \left\lfloor \frac{logU}{logq} \right\rfloor$ has a geometric distribution. What is the parameter of the geometric distribution?

Since q is in the range of $[0,1]$, $log(q)$ is less than 0. $(log(q) < 0)$ $log(U)$ is also $-\infty < log(U) < 0$, $0 < \frac{log(U)}{log(q)} < 1$. Let's say $\frac{log(U)}{log(q)} = k$ (k is an integer), then $k < \left\lfloor \frac{log(U)}{log(q)} \right\rfloor < k+1$.

$k \cdot log(q) > log(U) > (k+1) \cdot log(q)$. I don't know why but we need to put this in exponential eqaution.

$$k \cdot log(q) > log(U) > (k+1) \cdot log(q)$$

$$log(q^k) > log(U) \; and \; log(U) > log(q^{(k+1)})$$

$$e^{log(q^k)} > e^{log(U)} \; and \; e^{log(U)} > e^{log(q^{(k+1)})}$$

$$= q^k > U \; and \; U > q^{(k+1)}$$

$$= Pr[U < q^k] \; and \; Pr[q^{(k+1)} < U]$$

$$= Pr[q^k - q^{(k+1)}]$$

$$= Pr[q^k(1-q)]$$

which is a geometric distribution.

1.12 Problem 1.12

Let N be a discrete random variable that takes values from the set 1, n with equal probability, i.e., Pr[N = 1] = Pr[N = n] = $\frac{1}{2}$. Consider the following process.

First we draw a value for N.

Then, we draw N iid uniform RV $\{X_i\}_{i=1,...,N}$ in [0, 1]

Someone tells you the value Z = $min_{i=1,...,N}\ X_i = 0.05$, namely the smallest value among the N uniform RVs drawn. However you do not know he value of N

What is the probability Pr[$N = 1|Z = 0.05$] when: (a) n = 2, and (b) n = 10.

pdf of N discrete random variable is $N = \begin{cases} 1 & p = \frac{1}{2} \\ n & p = \frac{1}{2} \end{cases}$

We need to calculate Pr[$N = 1|Z = 0.05$]. Using Bayes theorem, we get

$$Pr[N = 1|Z = 0.05] \ = \ \frac{Pr[Z = 0.05|N = 1] \cdot Pr[N = 1]}{Pr[Z = 0.05]}$$

The probability to get Pr[N = 1] is $\frac{1}{2}$ since we know from the discrete random variable PDF. When N = 1, pdf of RV X is the uniform distribution in [0, 1], so the PDF of x is 1. So we get that Pr[Z = 0.05 | N = 1] = $1 \cdot \frac{1}{2} = \frac{1}{2}$.

Pr[Z = 0.05] is $Pr[Z = 0.05|N = 1] \cdot Pr[N = 1] + Pr[Z = 0.05|N \neq 1] \cdot Pr[N \neq 1]$.

$$Pr[Z = 0.05|N = 1] \cdot Pr[N = 1] + Pr[Z = 0.05|N \neq 1] \cdot Pr[N \neq 1] = 1 \cdot \frac{1}{2} + Pr[Z = 0.05|N \neq 1] \cdot Pr[N \neq 1]$$

$Pr[Z = 0.05|N \neq 1]$ is the probability that the min X_i value is 0.05 when there are n iid uniform RV. We can write this as Pr[min(z) = 0.05] = 1 - Pr[all x_i bigger than 0.05]. This is the same as 1 - Pr[all x_i to x_n > x] = 1 - $(1 - x)^n$. This is the form of CDF, in order to get PDF we need to derivative to get pdf.

$$(1 - (1 - x)^n)' = -1 \cdot n \cdot (1 - x)^{n-1} \cdot (1 - x)' = n \cdot (1 - x)^{n-1}$$

we put this equation in substitute of $Pr[Z = 0.05|N \neq 1]$. so the Bayes equation is

$$Pr[N = 1|Z = 0.05] \ = \ \frac{\frac{1}{2}}{1 \cdot \frac{1}{2} + n \cdot (1 - x)^{n-1} \cdot \frac{1}{2}}$$

$$= \frac{1}{1 + (n \cdot (1 - x)^{n-1})} = \frac{1}{1 + (n \cdot (1 - 0.05)^{n-1})} = \frac{1}{1 + (n \cdot (0.95)^{n-1})}$$

When n = 2, $\frac{1}{1+(2\cdot(0.95)^{2-1})} = \frac{1}{2.9} = \frac{10}{29}$

When n = 10, $\frac{1}{1+(10\cdot(0.95)^{10-1})} = \frac{1}{7.302} = 0.136939515$

Submitted by Jae Hong Lee on March 1, 2024.