ECON 7710 TA Session

Midterm Review

Jiarui(Jerry) Qian

University of Virginia, Department of Economics

arr3ra@virginia.edu

Sep 2023

Outline

General Suggestions

1 Relax. Grades doesn't matter. They just help you check how you are doing in this class. Spend some time for fun during the fall break [Look at your classmates who are not here today;)]

2 The exam date is OCT 04 at your class room Dell 2 101. I will be the proctor. If anything urgent happens, let me know.

2 Since it is an open book and open internet exam, you don't need to memory a lot but you want to know where they are quickly.

4 Review notes and C & B and figure out the unclear concepts. Redo homework and practice questions in my TA session if you cannot do it without the help of solutions.

5 Some questions can be tricky and that's what we will see today.

Midterm 2018 Q1

Powerball is an American lottery game offered by 44 states, the District of Columbia, Puerto Rico and the US Virgin Islands. Powerball tickets have 6 numbers on them from 1 to 99 each. While the numbers on each ticket are randomly generated, each combination of numbers is unique, i.e. no two tickets share the same combination. The winning combination is generated by randomly drawing 6 numbers from 1 to 99 with replacement. Find the probability that a ticket that you bought today wins powerball given the total number of tickets sold by today's drawing is 90,674,762.

Midterm 2018 Q1

After moving out of irrelevant information:
Powerball tickets have 6 numbers on them from 1 to 99 each. While the numbers on each ticket are randomly generated, each combination of numbers is unique, i.e. no two tickets share the same combination. The winning combination is generated by randomly drawing 6 numbers from 1 to 99 with replacement. Find the probability that a ticket that you bought today wins powerball.

- The total number of tickets sold out doesn't affect your probability of wining as long as you have only have one ticket in hand.
- If **order matters**, the probability is $P = \frac{1}{99^6}$

* If **order doesn't matter**, the probability is $P = \frac{1}{\binom{99+6-1}{6}} = \frac{1}{\binom{104}{6}}$

Midterm 2019 Q5

A gambler rolls two dice. Find the expectation of the sum of the outcomes, given that the outcomes of the dice are different.

Midterm 2019 Q5

A gambler rolls two dice. Find the expectation of the sum of the outcomes, given that the outcomes of the dice are different.

- For this easy case, it will be straightforward to list all the possible combinations from rolling two dice with different outcomes.

	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)		(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)		(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)		(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)		(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	

As you can see, there are 30 possible combinations. we can easily calculate the expectation of sum of outcomes from here.
Group up to avoid math.

$$E = \frac{1}{30}(3*2+4*2+5*4+6*4+7*6+8*4+9*4+10*2+11*2)$$

$$\Rightarrow E = 7$$

Midterm 2017 Q4

n-dimensional random vector $(X_1, ..., X_n)$ such that $0 \leq X_i \leq 1$ has density

$$f(x_1, ..., x_n) = \begin{cases} 1 + \prod_{i=1}^n (x_i - \frac{1}{2}), & \text{if } 0 \leq x_i \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

a Are $X_1, ..., X_n$ independent?

b Take a subset of k variables $(2 \leq k \leq n-1)$ from $X_1, ..., X_n$. Could the variables in this subset be independent?

Midterm 2017 Q4

n-dimensional random vector $(X_1, ..., X_n)$ such that $0 \leq X_i \leq 1$ has density

$$f(x_1, ..., x_n) = \begin{cases} 1 + \prod_{i=1}^n (x_i - \frac{1}{2}), & \text{if } 0 \leq x_i \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

a Are $X_1, ..., X_n$ independent?
Random variables are independent iff joint density is equal to the product of marginal densities

$$f(x_1, ..., x_n) = f_{X_1}(x_1) \times ... \times f_{X_n}(x_n)$$

For any X_i the marginal density of $f_{X_i}(x)$ when $x \in [0, 1]$

$$f_{X_i}(x) = \int_0^1 (1 + \prod_{i=1}^n (x_i - \frac{1}{2}))dx_i = \int_0^1 dx_i + \prod_{j \neq i}^n (x_j - \frac{1}{2}) \int_0^1 (x_i - \frac{1}{2})dx_i = 1$$

Midterm 2017 Q4

n-dimensional random vector $(X_1, ..., X_n)$ such that $0 \leq X_i \leq 1$ has density

$$f(x_1, ..., x_n) = \begin{cases} 1 + \prod_{i=1}^{n}(x_i - \frac{1}{2}), & \text{if } 0 \leq x_i \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

a Are $X_1, ..., X_n$ independent?

We know $f_{X_1}(x) = f_{X_2}(x) = ... = f_{X_n}(x) = \begin{cases} 1 & \forall x \in [0, 1] \\ 0 & \text{Otherwise} \end{cases}$

$$f(x_1, ..., x_n) \neq f_{X_1}(x_1) \times ... \times f_{X_n}(x_n)$$

So they are not independent.

Midterm 2017 Q4

n-dimensional random vector $(X_1, ..., X_n)$ such that $0 \leq X_i \leq 1$ has density

$$f(x_1, ..., x_n) = \begin{cases} 1 + \prod_{i=1}^{n}(x_i - \frac{1}{2}), & \text{if } 0 \leq x_i \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

b Take a subset of k variables $(2 \leq k \leq n-1)$ from $X_1, ..., X_n$. Could the variables in this subset be independent?
WLOG, X_1, X_2 are picked, We already know
$f_{X_1}(x) = f_{X_2}(x) = \begin{cases} 1 & \forall x \in [0,1] \\ 0 & \text{Otherwise} \end{cases}$
You can also show that when $x \in [0,1]$,
$f_{X_1,X_2} = \int_0^1 \int_0^1 1 dx_1 dx_2 + \prod_{i=3}^{n}(x_i - \frac{1}{2}) \int_0^1 \int_0^1 (x_1 - \frac{1}{2})(x_2 - \frac{1}{2}) dx_1 dx_2$
$= 1 + \prod_{i=3}^{n}(x_i - \frac{1}{2}) \int_0^1 (x_1 - \frac{1}{2}) dx_1 \int_0^1 (x_2 - \frac{1}{2}) dx_2 = 1$

- Basically, this tells you $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) * f_{X_2}(x_2)$
- This holds when k grows from 2 to $n-1$. so you can see the variables in any subset are independent.

Midterm 2021 Q4

Random variables X and Y are independent and identically distributed such that

$$P(X = k) = P(Y = k) = 1/N \quad \text{for } k = 1, 2, ..., N$$

Find distribution of random variable:

$$Z = max\{X, Y\} - min\{X, Y\}$$

Random variables X and Y are independent and identically distributed such that

$$P(X = k) = P(Y = k) = 1/N \quad \text{for } k = 1, 2, ..., N$$

Find distribution of random variable:

$$Z = max\{X, Y\} - min\{X, Y\}$$

- After getting rid of max and min:

$$Z = \begin{cases} X - Y & \text{if } X > Y \\ Y - X & \text{if } Y > X \\ 0 \quad \text{if } X = Y \end{cases} \quad \Rightarrow Z = \begin{cases} 1 \\ 2 \\ 3 \\ ... \\ N - 1 \\ 0 \quad \text{if } X = Y \end{cases}$$

Midterm 2021 Q4

Random variables X and Y are independent and identically distributed such that

$$P(X = k) = P(Y = k) = 1/N \quad \text{for } k = 1, 2, ..., N$$

Find distribution of random variable:

$$Z = max\{X, Y\} - min\{X, Y\}$$

- $P(Z = 0) = P(X = Y) = \frac{1}{N} * \frac{1}{N} * N = \frac{1}{N}$
- $P(Z = i) =?, i = 1, 2, 3, ...N - 1$

Midterm 2021 Q4

- Again, start from the most extreme case:
  - $i = N - 1$

    $$1 \quad 2 \quad 3 \quad \ldots \quad N-2 \quad N-1 \quad N$$

    Clearly, only pair $(X, Y) = (1, N), (N, 1)$ work. So 2 cases.
    $P(Z = N - 1) = \frac{2}{N^2}$
  - $i = N - 2$

    $$1 \quad 2 \quad 3 \quad \ldots \quad N-2 \quad N-1 \quad N$$

    Clearly, only pair$(X, Y) = (1, N-1), (N-1, 1)$ or $(2, N), (N, 2)$ work.
    So $2 * 2 = 4$ cases. $P(Z = N - 2) = \frac{4}{N^2}$
  - $i = N - 3$

    $$1 \quad 2 \quad 3 \quad \ldots \quad N-2 \quad N-1 \quad N$$

    Clearly, only pair$(X, Y) = (1, N-2), (N-2, 1)$ or
    $(2, N-1), (N-1, 2)$ or $(3, N), (N, 3)$ work.
    So $2 * 3 = 6$ cases. $P(Z = N - 3) = \frac{6}{N^2}$
  - You shall observe the pattern that when i decreases by 1, we have two more possible pairs. $P(Z = i) = \frac{2(N-i)}{N^2}$. Sum of $\frac{\#\text{pairs}}{2}$ and i is N
- $Z = \begin{cases} i, & P(Z = i) = \frac{2(N-i)}{N^2} \quad (i = 1, 2, ..N-1) \\ 0, & P(Z = 0) = \frac{1}{N} \end{cases}$

Midterm 2022 Q2

Suppose that X and Y are absolutely continuous random variables. Can random variables X and $Z = X + Y$ be independent? If your answer is yes. prove a necessary and sufficient conditions under which this is possible. If your answer is "no", prove that formally.

Midterm 2022 Q2

We want to see if X and $Z = X + Y$ can be independent. By definition:

$$X \perp Z \Rightarrow P(X \in B_1, Z \in B_2) = P(X \in B_1) \times P(Z \in B_2)$$

- We can always construct $Y = -X$, then we get $Z = 0$, a degenerate random variable, with $P(Z \in B_2) = \{0, 1\}$. Then for any value of X:

$$P(X \in B_1, Z \in B_2) = \begin{cases} P(X \in B_1) & \text{if } Z = 0 \\ 0 & \text{otherwise} \end{cases}$$

  Clearly it gives us $P(X \in B_1, Z \in B_2) = P(X \in B_1) \times P(Z \in B_2)$. So the answer should be yes they can be independent.
- A formal condition is constructed on characteristic functions.
  - X and Y are independent if and only if $\phi_{X,Y}(t,s) = \phi_X(t) * \phi_Y(s)$
  - $\phi_{X,Z}(t,s) = E[e^{itX} e^{is(X+Y)}] = E[e^{itX}]E[e^{is(X+Y)}] = \phi_X(t)\phi_Z(s)$

Core Jun 2018 Q2

Consider the non-negative function of two arguments:

$$G(x, y) = \begin{cases} 0, & \text{if } x \leq 0 \text{ or } y \leq 0 \\ min\{1, max\{x, y\}\}, & \text{otherwise.} \end{cases}$$

Formally show whether this function is a joint cumulative distribution function.

Core Jun 2018 Q2

Consider the non-negative function of two arguments:

$$G(x, y) = \begin{cases} 0, & \text{if } x \leq 0 \text{ or } y \leq 0 \\ min\{1, max\{x, y\}\}, & \text{otherwise.} \end{cases}$$

Always try to test counterexamples before you attempt to come up with a formal proof.
$Pr(a \leq x \leq b, c \leq y \leq d) = G(b, d) - G(a, d) - G(b, c) + G(a, c)$ Why?
Because $P(a \leq x \leq b, y \leq d) = G(b, d) - G(a, d)$
and $P(a \leq x \leq b, y \leq c) = G(b, c) - G(a, c)$.
Then if we test $a = 0, c = 0, b = 1, d = 1$ We have:

$$Pr(0 \leq x \leq 1, 0 \leq y \leq 1) = G(1, 1) - G(0, 1) - G(1, 0) + G(0, 0)$$
$$= 1 - 1 - 1 + 0 = -1$$

Clearly it is not a joint CDF.

Midterm 2020 Q3

Suppose that random variable Z is such that $a \leq Z \leq b$ and $E[Z] = 0$. Show that:

$$E[e^{tZ}] \leq e^{t^2(b-a)^2/8}$$

Hint: It may be useful to explore convexity of function e^{tZ} for each Z to produce deterministic bound $e^{tZ} \leq \frac{Z-a}{b-a}e^{tb} + \frac{b-Z}{b-a}e^{ta}$

Suppose that random variable Z is such that $a \le Z \le b$ and $E[Z] = 0$. Show that:

$$E[e^{tZ}] \le e^{t^2(b-a)^2/8}$$

Let $Z = \alpha a + (1-\alpha)b. 0 \le \alpha \le 1$ Then by convexity of e^{tZ} we know

$$e^{tZ} = e^{t(\alpha a+(1-\alpha)b)} \le \alpha e^{ta} + (1-\alpha)e^{tb}$$

Let $\alpha = \frac{b-Z}{b-a}$ and $1-\alpha = \frac{Z-a}{b-a}$
Then we know: $e^{tZ} \le \frac{Z-a}{b-a}e^{tb} + \frac{b-Z}{b-a}e^{ta} \Rightarrow E[e^{tZ}] \le E[\frac{Z-a}{b-a}e^{tb} + \frac{b-Z}{b-a}e^{ta}]$
Since $E[Z] = 0$, our inequality will be $E[e^{tZ}] \le \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}$
Can we transform the RHS to some function and carry out Tylor Expansion?

Midterm 2020 Q3

- $\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{ta} + \frac{a}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{t}a + \frac{a}{b-a}e^{ta}[1 - e^{t(b-a)}]$ $e^{ta}[1 + \frac{a-ae^{t(b-a)}}{b-a}]$
- Let $h = t(b-a)$ and $e^{ta}[1 + \frac{a-ae^{t(b-a)}}{b-a}]$ can be written as $e^{L(h)}$, where $L(h) = \frac{ha}{b-a} + ln(1 + \frac{a-e^{h}a}{b-a})$
  - $L(0) = 0,\ L'(0) = 0$
  - $L''(h) = -\frac{abe^{h}}{(b-ae^{h})^2}$ Let $b = M$ and $-ae^{h} = N$ We have $L''(h) = -\frac{abe^{h}}{(b-ae^{h})^2} = \frac{MN}{(M+N)^2} \leq \frac{1}{4}$
    Why? $(M-N)^2 \geq 0 \Rightarrow (M+N)^2 \geq 4MN \Rightarrow \frac{MN}{(M+N)^2} \leq \frac{1}{4}$
- Then by Tylor's theorem there is some $0 \leq \theta \leq 1$ s.t.

$$L(h) = L(0) + hL'(0) + \frac{1}{2}h^2L''(h\theta) \leq \frac{1}{8}h^2$$

Wait, what is $\frac{1}{8}h^2$? It is $\frac{t^2(b-a)^2}{8}$! So we proved
$E[e^{tZ}] \leq \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} \leq e^{t^2(b-a)^2/8}$

Midterm 2020 Q4

Prove that for any random variable Z for which the moment generating function is well-defined

$$P(Z > \epsilon) \leq e^{-t\epsilon} E[e^{tZ}]. \quad (**)$$

Using inequality$(**)$ in combination with the result in Question 3,

$$E[e^{tZ}] \leq e^{t^2(b-a)^2/8}$$

evaluate the bound for deviation probability for the sample mean $\bar{Z}$ constructed from the sample of i.i.d. random variables $Z_1, ..., Z_n$ such that $a \leq Z_i \leq b$ and $E[Z_i] = 0$:

$$P(\bar{Z} > \epsilon)$$

i.e. bound this probability from above as a function of ϵ, n, a, b. Discuss the difference between this bound and the bound that can be obtained using the Chbychev's inequality.

Hint: Note that the property of characteristic function of the sum of independent random variables also applies to the moment generating function. The bound can then be computed by minimizing the expression you obtain over t.

Actually, this is easier than Q3. So if you cannot prove sth, just don't stuck there. Move on.
We are given the inequality (**) to carry out research:

$$P(\bar{Z} > \epsilon) \leq e^{-t\epsilon} E[e^{t\bar{Z}}]$$

$$\Rightarrow P(\bar{Z} > \epsilon) \leq e^{-t\epsilon} E[e^{\frac{t}{n}\sum_{i=1}^{n} Z_i}] \Rightarrow P(\bar{Z} > \epsilon) \leq e^{-t\epsilon} E[e^{\frac{t}{n}Z_1} e^{\frac{t}{n}Z_2} ...]$$

$$\Rightarrow P(\bar{Z} > \epsilon) \leq e^{-t\epsilon} E[e^{\frac{t}{n}Z_1}] ... E[e^{\frac{t}{n}Z_n}] \quad (\text{ Independence of } Z_i)$$

Use the inequality we proved in Q3, we know:

$$e^{-t\epsilon} E[e^{\frac{t}{n}Z_1}] ... E[e^{\frac{t}{n}Z_n}] \leq e^{-t\epsilon} e^{\frac{t^2(b-a)^2}{8n^2}} ... e^{\frac{t^2(b-a)^2}{8n^2}}$$

$$\Rightarrow P(\bar{Z} > \epsilon) \leq e^{-t\epsilon} E[e^{\frac{t}{n}Z_1}] ... E[e^{\frac{t}{n}Z_n}] \leq e^{\frac{t^2(b-a)^2}{8n} - t\epsilon}$$

Midterm 2020 Q4

We know

$$e^{\frac{t^2(b-a)^2}{8n}-t\epsilon} = e^{\frac{(b-a)^2}{8n}[t-\frac{4n\epsilon}{(b-a)^2}]^2-\frac{2n\epsilon^2}{(b-a)^2}}, \forall t$$

and $\frac{(b-a)^2}{8n}[t-\frac{4n\epsilon}{(b-a)^2}]^2 \geq 0$ Therefore:

$$P(\bar{Z} > \epsilon) \leq e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

Chebyshev's inequality tells us suppose we have a random variable X with expected value μ and variance σ^2, for any real number $n > 0$

$$P(|X - \mu| \geq n\sigma) \leq \frac{1}{n^2}.$$

The bound in Chebyshev's inequality is polynomial in n. But our inequality is exponential in n. Basically, this new inequality shows much faster convergence of the sample mean.