Probability Theory for EOR 2021/2022

—

Solutions to Exercises

—

Week 6

December 22, 2021

Means, medians, modes and moments

1. Median: $(a+b)/2$. Mode: all points in the interval (a,b).

   Median: U are equally likely to be larger or smaller than the midpoint of (a, b), which is (a + b) / 2. This point satisfies the definition of Median. Therefore, (a + b) / 2 is the median of U. Notice for a value, u_1, smaller that (a + b) / 2, the probability of U being smaller than u_1 would be smaller than 1/2; while for a value, u_2, larger that (a + b) / 2, the probability of U being larger than u_1 would be smaller than 1/2. Therefore, $(a+b)/2$ is the unique median.

   Unif(a,b)'s PDF takes positive constant on (a, b), the rest parts zero. Therefore, Every point in (a, b) is a mode of U.

   It might be helpful to draw the graph of PDF, and all these values can be found directly from the PDF figure. For example, here we have a symmtric PDF about the $\frac{a+b}{2}$, then the median and the mean coincides.

2. Median: $(\log 2)/\lambda$. The mode is 0 if the PDF has support for all $x \geq 0$ and value of the PDF at zero is larger than $\lambda e^{-\lambda \times 0}$ otherwise if we use the specification for Expo in the textbook (Definition 5.5.1) there is no mode. Whether mode exists would depend on how we specify the PDF on $x = 0$.

   Notice we can change finitely many points of a function, which does not affect the integral (if exists) of the function. This is also one reason why the textbook defines all continuous r.v.s by their PDFs, even though we only need to know CDFs (or MGFs if they exist) to know the corresponding

continuous r.v.s. By specifying the exact PDFs, the textbook avoid the confusions that different PDFs (PDFs that only differ at countably many points) can lead to the same CDFs and thus the same continuous r.v.s. Another reason could be looking at PDFs graphs gives us a better view of the probability allocation.

Median: Denote F_X the CDF of X, then we find the only solution of $F_X(x) = 1/2$ would be a unique solution $F^{-1}(1/2) = \log(2)/\lambda$, and thus the median is $\log(2)/\lambda$.

This PDF function is strictly decreasing in x. The mode here is a bit tricky, its existence depends on how we define the PDF. Notice changing values of the PDF over countably many points does not affect its integration values and thus does not affect the associated probabilities (distributions, CDFs).

Therefore, there is no mode if the PDF is defined to be 0 at x = 0, and the mode is 0 if the PDF is defined to be any value larger than $\lambda e^{-\lambda \times 0}$ at 0. **If we use the specification for Expo in the textbook (Definition 5.5.1), then there is no mode.** This is also one of the reasons that we use PDF instead of CDF to define different continuous real-valued random varaible.

Further comment: note another difference between the modes from discrete r.v.s and continuous r.v.s. If we observe data drawn from a discrete distribution, we would observe the mode many times (as its probability is large); while if data is drawn from continuous one, we are not likely to observe the exact mode even for a single time as the probability of a continuous r.v. being equal to an exact number is always zero. However, for one continuous r.v. with uni-mode continuous PDF, we can expect to observe values from a neighborhood of the mode many many times as the integration of the PDF over an interval containing the mode is relatively larger, which means the probability of observing values from that neighborhood (an interval containing the mode) is relatively larger.

3. Median: $2^{1/\alpha}$ (CDF: $1 - x^{-1\alpha}$). Mode: 1.

Denote F_X the CDF of continuous r.v. X, and if we check the definition of median, we know we need to find one value x such that

$$\mathbb{P}(X \leq x) = F_X(x) \geq 1/2; \;\; \mathbb{P}(X \geq x) = \mathbb{P}(X > x) = 1 - F_X(x) \geq 1/2 \qquad (1)$$

which implies that $F_X(x) = 1/2$ *(note this only holds for continuous r.v.s, as for discrete r.v.s X', $\mathbb{P}(X' \geq x) = \mathbb{P}(X' > x)$ may not hold true if there is non-zero probability mass at x. See exercise 6.05 for example.).* For one continuous r.v., to find its median, we only need to locate possible values of x such that

$$F_X(x) = 1/2.$$

Now using PDF, we can derive F_X by integration (for $x > 1$):

$$F_X(x) = \mathbb{P}(X \leq x) = \int_1^x a/x^{a+1}dx = 1 - x^{-a}$$

Solve $F_X(x) = 1 - x^{-a} = 1/2$ we have a unique solution (*in some cases, if CDF is flat at 1/2 over some intervals, solutions are not unique*) $2^{1/\alpha}$ which is the median.

The PDF given in the exercise is strictly decreasing over $[1, \infty)$, we know the mode is 1.

5. For n odd, the unique median is $(n+1)/2$. For n even, any x, $x \in [n/2, (n+2)/2]$ is a median. (How to prove the above claim? Use the definition of Median.) Mode: $\{1, 2, \cdots, n\}$.

   The PMF is equal for all points in $\{1, 2, \cdots, n\}$ and thus they are all modes.

   For n is odd, we have

   $\mathbb{P}(X \leq (n+1)/2) = 1/2 + 1/(2n) \geq 1/2$; $\mathbb{P}(X \geq (n+1)/2) = 1/2 + 1/(2n) \geq 1/2$;

   For $m < (n+1)/2$, $\mathbb{P}(X \leq m) \leq 1/2 + 1/(2n) - 1/n < 1/2$; for $mm(n+1)/2$, $\mathbb{P}(X \geq m) \leq 1/2 + 1/(2n) - 1/n < 1/2$. Therefore, $(n+1)/2$ is the unique median.

   For n is even, we have

   $$\mathbb{P}(X \leq n/2) = 1/2 \geq 1/2;\ \mathbb{P}(X \geq (n+2)/2) = 1/2 + 1/(2n) \geq 1/2;$$

   which shows that any value between n/ 2 and (n + 2) / 2 (inclusive) is also a median.

   For $m < n/2$, $\mathbb{P}(X \leq m) < 1/2$; for $m > (n+2)/2$, $\mathbb{P}(X \geq m) < 1/2$. Therefore, there are no other medians.

6.12.a. *Discussions for how we tackle the problem:*

> *To prove 6.12.a (you may link this question with the relation between sample moments (random varaibles) and moments (constants, degenertaed random varaibles) such that in large samples, sample moments would be good approximations for moments) (this question links certainty (deterministic) with uncertanty (random) all together, and hope can enahnce your understanding of r.v. which is nothing random but a special fucntion from the outcomes in the sample space to numbers), we consider the following two steps:*

(I). *prove the deterministic inequality (nothing rondom there, it just holds true for arbitrary fixed numbers):*

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 \geq 0.$$

The above deterministic inequality, suggested by the hint, can also be proved by properties of r.v.: variance of the Dunif random varaible, X, with suppport $\{x_1, \cdots, x_n\}$ and PMF $P(X = x_i) = 1/n$ is always non-negative:

$$V(X) = EX^2 - (EX)^2 = \sum_{i=1}^{n} x_i^2 \frac{1}{n} - (\sum_{i=1}^{n} x_i \frac{1}{n})^2 \geq 0$$

(here VX is only zero if $n = 1$ such that X is a degenerated random variable which equals one value. For arbitrary r.v., Y, the variance is always non-negative since by definition it is the expected values of the non-negative random variable $(Y - EY)^2$ which always takes non-negative values).
This appraoch shows we can use "uncertainty" to prove some "certainty" results.
We also prove the above inequality using traditional approach in this note (see equations (1) and (2) in the later discussions).

(II). *If an inequality holds true for arbitrary numbers, then if we replace these numbers (constants) by r.v.'s, the resulted stochastic inequality also holds true with probability equal to one. This is not surprising as r.v. is nothing but simply some functions (mapping from a sample space) into numbers. If an inequality holds true for arbitrary numbers, then it always hold true for whatever values taken by these r.v.'s and thus if we replace these numbers (constants) by r.v.'s, the resulted stochastic inequality also holds true with probability equal to one. (In steo II, we use "certainty" inequality to prove some "uncertanty" result).*

We explain the step (II): Suppose $\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 \geq 0$ holds true for any arbitrary constants x_i 's, then if now we replace constant x_i with r.v. X_i, we would have $P\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2 \geq 0\right) = 1$. Because for any arbitrary random events: $\cap\{X_i = s_i\}$ (suppose X_i takes arbitrary real value s_i), we know $\frac{1}{n}\sum_{i=1}^{n} s_i^2 - (\frac{1}{n}\sum_{i=1}^{n} s_i)^2 \geq 0$. Notice we choose s_i arbitrarily, so no matter what random events happen (or in other words, whatever elements from the underlying sample space and thus whatever numbers X_i's map these elements into) we always have $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2 \geq 0$ conditional on these events. Therefore, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2 \geq 0$ holds true (no matter what random events occurr) with probability one (you may also think about LOTP + Bayes'rule).

Now let's proceed to the proof.

For arbitrary real **numbers** x_i's,

we first prove two deterministic equalities using classic approaches (the probability approach has been shown in the above discussions) (nothing rondom in these equalities, they just hold true for all different numbers) (denote $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i/n$):
(1)

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

((1) is quie commonly used in your future studies.)
(2)

$$\frac{1}{n}\sum x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2 = \frac{1}{n^2}\sum_{i<j}(x_i - x_j)^2$$

We first prove 6.12.a.(1)

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + (\bar{x})^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}(\bar{x})^2 \\
=&\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2(\bar{x})^2 + (\bar{x})^2 \\
=&\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 \\
=&\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2
\end{aligned}$$

As for 6.12.a.(2)

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \\
=&\frac{1}{n^2}\sum_{i=1}^{n} n^2 x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2 \\
=&\frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) \\
=&\frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j\right) \\
=&\frac{1}{n^2}\left(\sum_{i<j}\left(x_i^2 + x_j^2\right) + \sum_{i=j}\left(x_i^2\right) - \sum_{i<j} 2x_i x_j - \sum_{i=j}\left(x_i^2\right)\right) \\
=&\frac{1}{n^2}\sum_{i<j}\left(\left(x_i^2 + x_j^2\right) - 2x_i x_j\right) \\
=&\frac{1}{n^2}\sum_{i<j}\left(x_i - x_j\right)^2
\end{aligned}
$$

Now we prove using 6.12.a.(2) (the arguments using 6.12.a.(1) are similar where $x_i = \bar{x}, \forall i$ implies again they are equal to each other): note that for arbitrary numbers

$$\frac{1}{n^2}\sum_{i<j}(x_i - x_j)^2 \geq 0$$

where the equality holds only when $x_i = x_j, \forall i, j$ and thus (we replace x_i with X_i) we know

$$\frac{1}{n^2}\sum_{i<j}(X_i - X_j)^2 \geq 0;$$

and

$$T_2 \geq T_1;$$

where the equalities hold only when $X_i = X_j, \forall i, j,$

$$\mathbb{P}\left(T_1 \leq T_2\right) = 1$$

Note that

$$\mathbb{P}(T_1 = T_2) = \mathbb{P}(X_1 = \cdots = X_n) = 0$$

The last equality holds true because $X_1 - X_2$ is still a continuous r.v. and for continuous r.v. the probability of it being equal to one exact number is zero, therefore, $0 \leq \mathbb{P}(X_1 = \cdots = X_n) \leq \mathbb{P}(X_1 = X_n) = \mathbb{P}(X_1 - X_n = 0) = 0$ (**from a real line to one point, one dimension to zero dimension**). Interestingly, if we consider (X_1, X_2) as a two dimensional r.v. with independent entries in a plane, then $X_1 = X_2$ represents the event that the **two dimensional r.v. in a plane reduces to one dimensional line**, see the probability is always zero when high dimensional r.v. takes low dimensional values.

and thus by the second axiom of probability (recall the general definition of probability)

$$\mathbb{P}(T_1 < T_2) = \mathbb{P}(T_1 \leq T_2) - \mathbb{P}(T_1 = T_2) = 1$$

6.12.b. First use the sum of indepedent normal, $N(\mu_i, \sigma_i^2)$, is still normal $N(\sum_i \mu_i, \sum_i \sigma_i^2)$: $\sum_i X_i \sim N(nc, n\sigma^2)$.

Then the random varaible $\bar{X}_n$ is just a location-scale transformation of $\frac{1}{n}\sum_i X_i$, so it is still a normal distribution (a normal distribution is determined by two parameters: mean and variance, once you derive the mean, $E\frac{1}{n}\sum_i X_i = c$, and the variance, $V\frac{1}{n}\sum_i X_i = \sigma^2/n$, of $\bar{X}_n$ you will find its distribution) and thus $\bar{X}_n \sim N(c, \sigma^2/n)$.

$$\mathbb{E}\bar{X}_n^2 = \mathbb{V}\bar{X}_n + (\mathbb{E}\bar{X}_n)^2 = c^2 + \frac{\sigma^2}{n}$$

The bias of T_1: $ET_1 - c^2$ (here we want to use the value taken by the random varaible to approximate/estimate the unknown number c^2) is $\frac{\sigma^2}{n}$ (getting smaller when n is large).

$$\mathbb{E}T_2 = \frac{1}{n}\sum \mathbb{E}X_i^2 = \frac{1}{n}\sum(c^2 + \sigma^2) = c^2 + \sigma^2$$

The bias of T_2 is σ^2.

Note that the bias of T_1 is smaller and smaller (goes to zero) when the sample size is larger and larger (n increases), so if we can redraw values from the distribution of T_1 then the average would give a good approximation for the unknown number c^2.

Moment generating functions

13.

$$\left(\frac{1}{6}\sum_{i=1}^{6} e^{it}\right)^2$$

Note X and Y are i.i.d. (independent and identical distributed). By LOTUS, we calculate the MGF:

$$\begin{aligned}
M_{X+Y}(t) &=_{\textit{definition of MGF}} \mathbb{E}e^{t(X+Y)} \\
&=_{\textit{independence}} \mathbb{E}e^{t(X)}\mathbb{E}e^{t(Y)} \\
&=_{LOTUS} \left(\sum_{i=1}^{6} e^{it}\frac{1}{6}\right)\left(\sum_{i=1}^{6} e^{it}\frac{1}{6}\right) \\
&= \left(\frac{1}{6}\sum_{i=1}^{6} e^{it}\right)^2
\end{aligned}$$

which is well defined for all $t \in \mathbb{R}$.

14.

$$\frac{(e^t-1)^{60}}{t^{60}}$$

We need to specify the value of t.

Suppose $t \neq 0$ first, for U_i,

$$\begin{aligned}
M_{U_i}(t) &=_{\textit{MGF definition}} \mathbb{E}e^{tU_i} \\
&=_{LOTUS} \int_0^1 e^{ta}da \\
&= (e^t-1)/t
\end{aligned}$$

Then

$$M_{\sum_i U_i}(t) =_{\textit{independence}} \prod_i M_{U_i}(t) =_{\textit{identical distribution}} \prod_i (e^t-1)/t = \frac{(e^t-1)^{60}}{t^{60}}$$

When $t = 0$, follow the same step, you will find $M_{U_i}(0)$ and $M_{\sum_i U_i}(0)$ are both equal to one.

Note that, for all $t \in \mathbb{R}$, $M_{\sum_i U_i}(t)$ takes finite values and thus is well-defined.

Note that, the results are quite consistent as $\lim_{t\to 0}(e^t-1)/t = 1$ *and* $\lim_{t\to 0}\frac{\left(e^t-1\right)^{60}}{t^{60}} = 1$.

Why we need to check 0? Notice MGF is only defined if the function is defined in some interval (-a,a) which contains 0!! *Therefore, to show that one MGF is well defined, we need show that the values taken by the function within some interval (-a,a) for* $a > 0$.

16. Note that $Y = \lambda X \sim$Expo(1) (Intuition would be: if we normalize the weighting time by the arrival rate, it is 1 then. For example, the arrival (service) rate is 2 per day and thus the weighting time would be half day on average, 2 times half is one.). The skewness is the third standardized moment which in this case is the third moment of $\frac{X-1/\lambda}{1/\lambda} = Y-1$ (whose distribution now is free of λ, and thus of course Skenness has nothing to do with λ).

$$Skew(X) = \mathbb{E}(Y-1)^3 = \mathbb{E}(Y^3 - 3Y^2 + 3Y - 1) = 3! - 3 * 2! + 3 - 1 = 2 > 0$$

Right (positive) skewed. if you draw the graph, you will see that it has a long right tail, and thus could take some very large positive numbers, and the skewness is positive.

Now how do we calculate the ith moment of $E(Y^i)$, you can use the definition of expectation and LOTUT $= \int_0^\infty y^i e^{-y} dy$. Or you can make use of the moment generting function:

$$M_Y(t) = \int_0^\infty e^{ty} e^{-y} dy = \frac{1}{1-t},$$

which is well defined since $\frac{1}{1-t}$ is finite for, e.g., $t \in (-0.5, 0.5)$ and thus

$$E(Y^i) = (M_Y(t))^{(i)} = i!$$

18. Check Example 6.4.3, X has MGF $M_X(t) = p/(1-qe^t)$ for $qe^t < 1$ (which is equivalent to $t \in (-\infty, -\ln(q))$, and $t \in (-\infty, -\ln(q))$ apparently contains zero as $\ln(q) < 0$. $M_X(t)$ takes finite values for $t \in (-\infty, -\ln(q))$ and thus is well-defined.). Then,

$$\mathbb{E}X = M^{(1)}(0) = pqe^t/(1-qe^t)^2|_{t=0} = q/p$$
$$\mathbb{E}X^2 = M^{(2)}(0) = pq\left((1-qe^t)^2 e^t + 2qe^{2t}(1-qe^t)\right)/(1-qe^t)^4|_{t=0} = q(1+q)/p^2$$

Therefore,

$$\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = q/p^2$$

19. Not Poisson. $M_{X+2Y}(t) = e^{\lambda(e^{\lambda}+e^{2\lambda}-2)}$.

    First, derive MGF. For any t from a real line (certainly the whole real line would contain one open interval around zero), we have:

$$\begin{aligned} M_{X+2Y}(t) &=_{\text{definition MGF}} \mathbb{E}e^{t(X+2Y)} \\ &=_{\text{independence}} \mathbb{E}e^{tX}\mathbb{E}e^{2tY} \\ &=_{\text{definition MGF}} M_X(t)M_Y(2t) \\ &=_{\text{Poisson MGF}} e^{\lambda(e^t-1)}e^{\lambda(e^{2t}-1)} \\ &= e^{\lambda(e^{\lambda}+e^{2\lambda}-2)} \end{aligned}$$

    Now we know if X+2Y follows Poisson it should follow Pois($\mathbb{E}(X+2Y)$) as the arrival rate should be its expectation by Poisson property. However, this contradicts the above MGF we derive as it is not equal to $M_{Pois(\mathbb{E}(X+2Y))}(t) = M_{Pois(3\lambda)}(t) = e^{3\lambda(e^t-1)}$.

20. (you may find the consistency among different concepts is delightful *Taylor expansion*.)

    You can either use Taylor series for e^t or you can simply use the fact that we have used many times in this course such that $\sum_{i=0}^{\infty} t^i/i! = e^t$ (think about the sum of the PMF funcrion of Pois(t)). Both ways would lead to

$$g(t) = \lambda(e^t - 1) = \lambda(\sum_{i=0}^{\infty} \lambda^i/i! - 1) = \lambda(1 + \sum_{i=1}^{\infty} t^i/i! - 1) = \sum_{i=1}^{\infty} \lambda \frac{t^i}{i!}$$

    Then if we match the coefficients of the corresponding exponentiations we have $c_j = \lambda, \forall j \geq 1$.

22.a. MGFs are useful when dealing with the sum of independent r.v.'s since

$$M_{\sum_i Y_i} = \prod_i M(Y_i)$$

if Y_i's are independent from each other. Therefore,

$$M_{I(A_i)}(t) = \mathbb{E}e^{tI(A_i)} = (1-p_i)e^0 + p_ie^t$$

$$M_X(t) = M_{\sum_i I(A_i)}(t) = \prod_i M_{I(A_i)}(t) = \prod_i \left((1-p_i) + p_i e^t\right)$$

22.b. In the exam, we will try to avoid using the knowledge relating to some approximations in the limit (such as the following formula in color blue), we will give hints directly when these approximations are needed:

$$\begin{aligned}
&\prod_i \left((1-p_i) + p_i e^t\right) \\
=&\prod_i \left(1 + p_i(e^t - 1)\right) \\
\approx&\prod_i e^{p_i(e^t-1)} \\
=&e^{(e^t-1)\sum_i p_i} \\
=&e^{(e^t-1)\lambda}
\end{aligned}$$

$e^{(e^t-1)\lambda}$ is the MGF of Pois(λ). Poisson paradigm (Poisson distribution is a good approximation for the distribution of a sum of many weakly dependent indicator r.v.'s (which equal 1 with very small probability values), you may regard each indicator r.v. is the occurance of one specific rare event): Poisson approximates the distribution of the arrivals of a collection rare events.

Add. Ex. Prove for independent discrete r.v.'s $X, Y : E(XY) = E(X)E(Y)$.

$$\begin{aligned}
E(XY) = &\sum_{x \in Supp(X), y \in Supp(Y)} xyP(X = x, Y = Y) \\
&=_{independence} \sum_{x \in Supp(X), y \in Supp(Y)} xyP(X = x)P(Y = Y) \\
&= \sum_{x \in Supp(X)} xP(X = x) \sum_{y \in Supp(Y)} yP(Y = Y) \\
&= EXEY.
\end{aligned}$$