Mathematics for Policy and Planning Science

SAMPLE Examination Questions and Answers

June 29, 2015

General instructions / 一般説明

If the English and Japanese versions of any text differ in meaning, the English text is more reliable. However, please ask for clarification if you have any doubt. Japanese versions of most questions are not available.

各文書の英文と和文の間に食い違いがあれば、英語の方に信頼をおいた方がよ い。ただし、不明な点については遠慮なく聞いてください。和文のない問題が 多いです。

Do not forget to write your name and student ID number on each page.

Several problems in basic mathematics are presented below. **You may answer in Japanese or English.** In Japanese, please take great care in writing kanji. Avoid abbreviated kanji; the only one I know is the 3-stroke mongamae.

Use of notes, textbooks, and so on is prohibited. All calculations are simple, and allowing use of electronic devices is unfortunately far to great a temptation to some students, so the use of calculators and electronic dictionaries is also prohibited. Some dictionaries will be provided.

On the desk you may place pens, pencils, erasers, pencil sharpeners, and tissues or handkerchief of reasonable size. Take out your cellphone, turn it off, and place it on the desk in front of you. All other items must be put in your bag and placed at your feet.

Except for calculations, most of the problems can be completely answered within 3 lines. Many questions can be answered within 2 or 3 words. Below each problem ample space is provided. Please write your answers there. Graph paper is provided for graph problems. Please use it. In calculations, in addition to the result itself, please also write any equations used.

Each question is worth 10 points, unless otherwise specified.

名前と学籍番号を忘れずに各ページに記入してください。

以下の設問のすべてに解答せよ。**解答の言語は日 本語でも英語でも構わない。**もし日本語で書けば漢字などの書き方に十分注意 してください。たとえ、省略した漢字などを使わないで。（三角門構えの他は 分からなく、そして私が読めない場合には省略した文字を「間違え」と採点します。）

ノート・教科書・電子辞書・電卓・携帯電話・その他のメモリ付き電子製品が 誘惑ものとなるので使用は禁止である。全ての計算は簡単で電卓などは不要。

机の上にペン・鉛筆・消しゴム・鉛筆削り・時計・この試験用紙の他の物を 置かないこと。携帯をお持ちなら電源を切って机の上に置くこと。その他のも のを鞄などに締めて足元に置くこと。

後ろの面空白を使ってもよい。ただ、その場合に表にメモを書くこと。メモが ない場合、別所で書いた文はカウントされない場合があります。

それぞれの問題は基本的点数が１０点。例外に５点、２０点などの問題もあるし、 注意してください。

Problems / 問題

Note: this is not a representative sample. These questions are all possible, but there are many others. You should not conclude anything about the relative likelihood of questions from this sample.

＊＊注意：代表的なサンプルでなく、試験に問われる可能性だけである。他の種類も可能である。サンプルでの頻度で確立を推計しては行けないこと。＊＊

Note: *This testbank is a* ***preliminary draft****. Among other things, I haven't checked that the problems really earn 10 points each.*

In this testbank, some questions are marked "This is a trick question." That means that the question seems to be asking for a specific answer, but in fact the answer is "there is insufficient information to answer the question as stated." "Insufficient information" plus a refusal to guess may be a sufficient answer, but in many cases a better answer is to state an obvious assumption, and use that to answer the question. Where a better answer is appropriate, it will get more points. An example of an "obvious" assumption is that a gambling device (such as dice, cards, or a wheel) is "fair", that is, each possible outcome has the same probability. ***Note:*** *"Trick question"* ***markers*** *will* ***not*** *be present in the examination!*

注意：本問題集は**草稿**です。各問題の点数はまだ確 認していません。

1. `[Problem ID #1] probability distribution computation`
   In class we showed that the distribution of sums of dots on a pair of dice is:

sum	2	3	4	5	6	7	8	9	10	11	12
frequency	1	2	3	4	5	6	5	4	3	2	1

Recall that we represented the state space of sums by a 6×6 *table:*

red/blue	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Count the number of combinations of the number of dots on a pair of dice that achieves a particular:

(a) product

The state space of (red, blue) pairs and products:

red/blue	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

The frequency of products:

sum	1	2	3	4	5	6	7	8	9	10	11	12
frequency	1	2	2	3	2	4	0	2	1	2	0	4

sum	13	14	15	16	17	18	19	20	21	22	23	24
frequency	0	0	2	1	0	2	0	2	0	0	0	2

sum	25	26	27	28	29	30	31	32	33	34	35	36
frequency	1	0	0	0	0	2	0	0	0	0	0	1

(b) difference

The state space of (red, blue) pairs and differences:

red/blue	1	2	3	4	5	6
1	0	-1	-2	-3	-4	-5
2	1	0	-1	-2	-3	-4
3	2	1	0	-1	-2	-3
4	3	2	1	0	-1	-2
5	4	3	2	1	0	-1
6	5	4	3	2	1	0

The frequency of differences:

difference	-5	-4	-3	-2	-1	0	1	2	3	4	5
frequency	1	2	3	4	5	6	5	4	3	2	1

(c) quotient

The state space of (red, blue) pairs and quotients:

red/blue	1	2	3	4	5	6
1	1	1/2	1/3	1/4	1/5	1/6
2	2	1	2/3	1/2	2/5	1/3
3	3	3/2	1	3/4	3/5	1/2
4	4	2	4/3	1	4/5	2/3
5	5	5/2	5/3	5/4	1	5/6
6	6	3	2	3/2	6/5	1

The frequency of quotients:

quotient	1/6	1/5	1/4	1/3	2/5	1/2	3/5	2/3	3/4	4/5	5/6	1
frequency	1	1	1	2	1	3	1	2	1	1	1	6

quotient	6/5	5/4	4/3	3/2	5/3	2	5/2	3	4	5	6	
frequency	1	1	1	2	1	3	1	2	1	1	1	

Discussion:

The distribution of ordered pairs is said to be uniform*. For each pair, the frequency seen in a series of throws of the dice should be about the same. Theoretically, this occurs because each (ordered) pair occurs once in a list of pairs (as seen in the* state space *tables, square tables with six rows and six columns). The frequency of sums is* nonuniform *(different frequencies for different values). Theoretically this difference occurs because there are often multiple pairs of dice that result in the same sum and the number of appropriate pairs varies according to the sum specified.*

Note that this is the prototypical example of

- *a* random variable *(a function from a set whose probabilities are known, to a set of more interesting outcomes, usually but not always numerical), and*
- *the difference between* observed *variables (the sum of the dice is what the croupier announces) and underlying* latent *variables (the pair of numbers of dots on each die).*

Random variables are often used to model *latent-observed systems, both in social science theory (especially economics) and in statistical analysis. It is very common, though not theoretically necessary, to use a uniform distribution on the underlying set.*

2. `[Problem ID #2] probability distribution computation`
   Did you notice any interesting similarities or differences between the frequency distributions? Did you notice any interesting similarities or differences between the random variables (*i.e.*, the function from the pairs of dice to the mathematical result)?

   The obvious interesting fact is that the distributions of sums and differences are identical (with a shift of -7 from the sums to the differences). This is not an accident. *Look at the state space to value mappings (they are not quite a random variables, despite what I wrote before, because there is no probability here yet). Notice that the patterns are the same, with a tall ridge along the diagonal, except that the diagonals are perpendicular to each other.*

   The mathematical reason why this happens is that the operation of subtraction is "the same" as addition in the sense that we can transform the subtracted numbers y (in the distribution of differences $x - y$) into the added numbers z (in the distribution of sums $x + z$) by the formula $z = 7 - y$. However, although the set *of numbers is the same, the* order *of numbers is reversed (because z and y have opposite sign in the equation relating them). This is why the diagonals with constant frequency have different orientation. If you're not a mathematician, that may not have much flavor.*

   But think about this: reversing the order of rows in the difference table doesn't change the distribution. Now add 7 to each number in the table, and compare to the sum table. It's the same.

3. `[Problem ID #3] probability model`
   Describe a model of "what is the gender of the first person to arrive in the classroom of 'Mathematics for Policy and Planning Science' ". What "model" means here is "What things do you count to determine the probability that the first person to arrive is female?"

   (a) Use an underlying set whose probabilities are *uniform*.

   One convenient underlying set is the set of students in the class. If we assume that the probability of each person arriving first is the same (symmetry of people in terms of class arrival), then we can just count the number of women in the class and divide by the number of students to get the probability that the first person to arrive is female.

   (b) Do you think this model accurately describes the probability that the first person to arrive is female? If so, why? If not, why not?

   No, it's not very accurate. First, to be precise, I *might arrive at the class first. But ignoring that was just a convenience. More important, the first person to arrive might be someone with no connection to the class itself, but who needs to talk to someone in the class for a few minutes.*

That would leave us with no good way to estimate probabilities. Second, even if we ignore non-students (that is, redefine the problem to be the gender of the first student *to arrive), the symmetry assumption seems unlikely to hold. Students who had a class in the same room in 4th period are far more likely to arrive first (or simply remain in the room). Students who smoke who had a class in 4th period are likely to want to go somewhere to smoke between classes, so they are relatively unlikely to be first. And so on—there are many personal characteristics that affect the likely timing of arrival.*

4. `[Problem ID #4] random variable model`
   Think of a practical or daily life application where you can explain observed rates of occurance with a model with an underlying uniform distribution, but a nonuniform distribution for the observed (or practically relevant) outcomes. What is the underlying set composed of? Why should its elements be uniformly distributed? What is the function relating the underlying things to observed outcomes?

   It's actually quite difficult to come up with a practical application that matters where the underlying set is symmetrical unless it's either very abstract or deliberately constructed to generate equal probabilities (like a gambling game using dice, cards, or a wheel with equally-sized slots).

   The best I could come up with is a pencil. It has six equally wide sides. So if you roll it, each side should come up with equal probability. But usually only two sides have imprinting on them, so the probability of "printed side up" is 1/3, while the probability of "blank side up" is 2/3.

   The difficulty of coming up with realistic underlying sets that have uniform probability, or even easily defined probability, is why I say that state spaces are abstract.

5. `[Problem ID #5] random variable model`
   (Optional) Are there other ways to combine the two numbers on a pair of dice into an interesting single result with non-uniform probability? (Note: I don't have a good answer in mind as of writing this question, so don't lose sleep trying to answer it!)

   The best I could come up with is "Do the numbers match?" The probability of such a "doublet" is 6/36 = 1/6, and the probability of non-match is 5/6.

6. `[Problem ID #6] mathematics`
   (Optional) Here's a brain-breaker for those of you who think you're good at math. The numbers on dice are all positive, and therefore it makes sense to take logarithms. There's a one-to-one relationship between numbers and their logarithms, so given a number we can find its logarithm, and it is the

unique number with that logarithm. And given a number that is a logarithm there's a unique number it is the logarithm for.

Now, after taking logarithms, multiplication becomes addition. Therefore we might expect that there should be a one-to-one relationship between the distribution of *sums* of two dice and the distribution of *products* of two dice. That is wrong.

Explain why.

Basically the problem is that we haven't taken exponentials of the addends to get factors. The equivalence of multiplication to addition under the logarithmic transformation requires transforming both sides.

Thus, if instead of taking products of 1, 2, 3, 4, 5, and 6, where the equivalence doesn't work, we take products of 1, 2, 4, 8, 16, and 32, we do *get an equivalent distribution. Check it!*

7. `[Problem ID #7] set computation`
   For any set A:

   (a) What is $A \cup \{\}$? Explain why in terms of the membership relationship.

   $A \cup \{\} = A$, because all of the members of A are members of $A \cup \{\}$, and there is nothing else because $\{\}$ has no members.

   (b) What is $A \cap \{\}$? Explain why in terms of the membership relationship.

   $A \cap \{\} = \{\}$, because anything that is a member of $A \cap \{\}$ must be a member of both A and $\{\}$, and $\{\}$ has no members.

8. `[Problem ID #8] set logic`
   Consider the following two descriptions of collections:

   - The collection of students in this class who understand mathematics well.
   - The collection of students in this class who get the grade of A.

   Which of these describes a well-defined *set*? Explain why you chose the one you did, and why you rejected the other one.

   Students who get an A is a well-defined set. The definition of "understand mathematics well" varies both by the person who is evaluating the students, and for the same person by situation. (That is, I might say that a student "understands well" in grading this class, but not if that same student proposed to do a thesis in game theory.)

9. `[Problem ID #9] set logic`
   Consider the two sets containing one member: A = {a professor named

Stephen Turnbull} and B = {a professor named Stephen Turnbull}. Explain why you *cannot* say $A = B$.

Consider `https://en.wikipedia.org/wiki/Stephen_Turnbull_(historian)`. *A set member must be unambiguously identifiable, and "professor named Stephen Turnbull" is ambiguous.*

10. `[Problem ID #10] set computation`
Let $A = \{1, 2, 3\}$ and $B = \{2, 4\}$. Compute the following sets:

   (a) $A \cup B$

   $A \cup B = \{1, 2, 3, 4\}$.

   (b) $A \cap B$

   $A \cap B = \{2\}$.

   (c) $A \backslash B$

   $A \backslash B = \{1, 3\}$.

   (d) $B \backslash A$

   $B \backslash A = \{4\}$.

11. `[Problem ID #11] set algebra`
Consider two sets C and D. What can you say about $(C \cap D) \cup (C \backslash D)$? Explain.

$(C \cap D) \cup (C \backslash D) = C$ because $(C \cap D)$ is all the members of C that are in D, while $(C \backslash D)$ is all the members of C that are not in D. So any member of C that is not in the former is in the latter, and so the union contains all members of C. On the other hand, to be a member of either $(C \cap D)$ or $(C \backslash D)$, an element must be a member of C, so everything in $(C \cap D) \cup (C \backslash D)$ is in C. Thus they must be equal to each other for any C and D.

12. `[Problem ID #12] set logic`
Let E = {stocks listed on the first section of the Tokyo Stock Exchange whose price was at least as high on June 30, 2015 as on December 31, 2014} and F = {stocks listed on the first section of the Tokyo Stock Exchange which traded more shares January 1–June 30, 2015 than July 1–December 31, 2014}. What are the following sets (described in the same style as E and F)?

Let "stocks" be restricted to those stocks listed on the Tokyo Stock Exchange first section.

(a) $E \cap F$

*$E \cap F$ = {stocks whose price was higher on June 30 than on December 31, **and** which traded more shares in the first half of 2015 than they did in the second half of 2014.*

(b) $E \backslash F$

*$E \cap F$ = {stocks whose price was higher on June 30 than on December 31, **but** which did not trade more shares in the first half of 2015 than they did in the second half of 2014.*

13. `[Problem ID #13] set logic computation`
Let G = {ordered pairs of 6-sided dice whose sum is 4} and H = {ordered pairs of 6-sided dice whose product is 4}.

(a) Describe $G \cap H$ in words, in the same style as G and H were defined.

$G \cap H$ = {ordered pairs of 6-sided dice whose sum and product are both 4}.

(b) Compute $G \cup H$ as an explicit set of ordered pairs (x, y) of numbers between 1 and 6 (inclusive).

{(1,3), (1,4), (2,2), (3,1), (4,1)}. Only (2,2) is in both G and H.

14. `[Problem ID #14] set algebra`
The event A_1 = {the sum of a pair of 6-sided dice is 1}, and A_2, …, A_{50} are defined in the same way (the number of the set is the sum). Similarly, B_1 = {the product of a pair of 6-sided dice is 1}, and B_2, …, B_{50} are defined in the same way.

(a) Describe the event C = {the sum of a pair of dice is equal to their product} in terms of the events A_i and B_j.

$C = (A_1 \cap B_1) \cup (A_2 \cap B_2) \cup \cdots \cup (A_{50} \cap B_{50})$. *$A_1 \cap B_1$ means that the sum is 1 (from A_1) and the product is 1 (from B_1), so $A_1 \cap B_1$ means product = sum = 1, and similarly for all the other A, B pairs. C is the union of all of these.*

(b) Compute the event C from Part a as a set of ordered pairs of numbers. Explain the method you used to compute this set.

{(2,2)}. The obvious way to "compute" this is to compare the state space tables for sum and product in Problem #1.
A trickier way to do the problem is to solve $a + b = ba$ to get $a = \frac{b}{b-1}$, and notice that a can be an integer only if $b = 0$ or $b - 1 = 1$. $b = 0$ is out of the range of dice, but $b = 2$ gives $a = 2$.

(c) What is the probability of the event C? What assumptions do you need to make to compute this probability?

The probability of C is 1/36, on the assumption that the dice are "fair" (uniform probability of each number = 1/6), and independent of each other.

(d) Are $A_1 \cup \cdots \cup A_{50}$ and $B_1 \cup \cdots \cup B_{50}$ *partitions* of U = {all the ways two 6-sided dice can fall}? Explain why or why not. If they are partitions, are they the *same* partition?

Neither is a partition, because many of the A_i and B_j are empty. E.g., $A_{50} = B_{50} = \{\}$.

15. `[Problem ID #15] probability estimation`
Mr. X is a student in this class. Describe *three* different ways of assessing the probability that Mr. X will *fail to attend* class on June 15 (a future date).

For each method, explain any assumptions you need to make to justify using the method.

There is an infinite number of ways to assess the probability. Here are three examples from the different categories.

(a) *Pure subjective estimate: I know Mr. X is a very serious student and would not miss class except for a very serious reason. I assess the probability that he would miss that class to be 0.005. No assumptions are needed.*

(b) *Pure frequentist estimate: In the first 6 weeks of class, Mr. X missed class once. I assess the probability that Mr. X will miss any class meetig to be 1/6, and therefore the probability that he will miss class on June 15 is 1/6. A typical assumption that leads to this estimate is that Mr. X's attendence is the same for each class meeting, and does not depend on how often (including zero) he has missed class before.*

(c) *Bayesian estimate: The program leader told me he is missing one class out of 5 overall recently. The probability is 1/3 that there will be a meeting during any given class meeting. 2/5 of students who miss class say they had a recruiting meeting. I know that there is a recruiting meeting on that day. Summarizing, let RM = there is a recruiting meeting, and MC = Mr. X misses class. Then P(MC) = 1/5, P(RM) = 1/3, and P(RM|MC) = 2/5. Since I know there is a recruiting meeting, what I want to know is P(MC|RM). Using Bayes' Rule, P(MC|RM) = P(RM|MC)P(MC)/P(RM) = (2/5)(1/5)/(1/3) = 6/25 = 24%.*

The assumptions that need to be made are that Mr. X's reasons for missing class are the same as other students' reasons, and that students are reporting their reasons for absence truthfully.

16. `[Problem ID #16] random variable`
Recall Problem 1, which defined the sum, difference, product, and quotient of the numbers produced by rolling a pair of dice. Which of these four are *random variables*? Explain why you chose the ones you did.

All are random variables, because they are functions of random variables. However, in order to be technically correct, one needs to assume that there is an state space, where the number of dots on the top face of each die is a function of the state.

17. `[Problem ID #17] random variable model`
Consider a different situation with dice. This time we have *one* die, but we roll it twice, and then add the two numbers. Carefully describe a model of this experiment which allows you to compute the probability of each value of the sum.

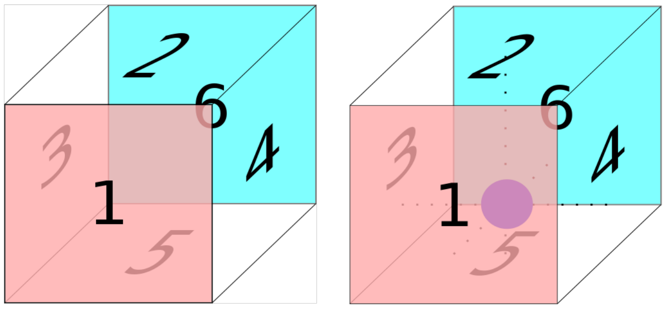
Compare this model to the model used in Problem 1.

Mathematically it's the same model as in Problem 1. We just change "face of red die after roll" to "face of (only) die after first roll" and "face of blue die after roll" to "face of (only) die after second roll". We define a state space by taking pairs (first face, second face), and assume each face has equal probability of landing on top in the first roll and in the second roll, and assume independence of the two roll. That gives each pair of faces equal probability of 1/36. The set of first and second face pairs is Ω. *Then we map each pair to the numbers 1 to 6 by taking the number on the first face to get a random variable* $X : \Omega \to \{1, \ldots, 6\}$ *and the number on the second face to define a second random variable* $Y : \Omega \to \{1, \ldots, 6\}$. *Then the sum of the dice is defined by applying the function* $f(x, y) = x + y$ *to the two random variables, giving a random variable* $Z = f(X, Y)$ *defined by* $Z(\omega) = f(X(\omega), Y(\omega)) = X(\omega) + Y(\omega)$. *Z has the same distribution as given in Problem 1:*

sum	2	3	4	5	6	7	8	9	10	11	12
frequency	1	2	3	4	5	6	5	4	3	2	1

18. `[Problem ID #18] random variable model computation probability estimation`
Now consider a new kind of problem. Suppose you have a die that you think is used to cheat. That is, inside of the die is a weight placed near one face of the die at the center of that face. That means that the face on the opposite side of the die is more likely to come up. However, you don't know which side is "loaded".

The image below shows an "unloaded" die on the left, and a die loaded near "5" (so that "2" is most likely to come up) on the right:



Describe a probabilistic model in which you can compute the probability of each side of the die coming up if rolled *once* based on the information above. Note: Even if you know the physics, you can't calculate a number without more information than is given here, so you will need one or more parameters in your model. (Use as many as is convenient, but make sure you define all of them.)

Discussion:
This problem is a lot more work than I thought when I made it up. Nevertheless, except that the number of cases could be reduced by discussing coin tosses rather than dice rolls, this is about as simple a model of latent variables (the "load" in the die) as you will find. Although you need not study the calculations carefully, there are several important points that you should pay attention to.

***Latent variables** must be included in the probability model. They do affect the behavior of the system in observable ways with enough data. Compare the unobservable differences in the distribution of one die roll (Parts (a) and (b), where if the position of the "load" is unknown, the probabilities are uniform) with the observable differences in the joint distribution of two rolls (Parts (c), (d), (e), and (f), where different sequences have different probabilities, i.e., $A > B > C$.)*

***Specifying distributions of variables** is typically done by assuming a uniform distribution, as in part (b) for the latent variable of "load position," or a parametric model based on a domain model as for the distribution of faces for a loaded die (α, β, and γ). In an empirical analysis a means for estimating them would be used.*

***State spaces** used in theoretical analyses must be sufficiently complex to handle latent variables as in Part (c).*

***Composition of simple functions** can be used to simplify complex random variables as in Parts (d) and (f).*

***Independence** is also useful to simplify calculations, although it may not* *apply to real situations.*

This question will not appear on the exam. *However, calculation and modeling problams similar to parts of the problem may.*

(a) What is your state space? y

Since the problem is fundamentally asymmetric, we may as well choose a state space consisting of the pairs of sets of faces of the die. The first element of the s pair is the face which is weighted, the second is the face that comes up on the roll. We need the first face of the pair because the problem says we don't know which face is loaded. The second face of the pair is the one which lands on the top.

(b) What is the probability of each number? e You probably need to use two different methods to assess the probability in your model.

First, we suppose that since we don't know anything about which face is "loaded," each face has equal probability of being loaded. (This is partly an assumption of symmetry, and partly just plain guessing.) But "face 1 is loaded" is not a state, it is an event, because it corresponds to the set of pairs {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)}. Corresponding statements about "is loaded" being an event will be true for the other faces 2, ..., 6. This is different from the way we defined the probabilities in previous problems.

Now, if face 1 is loaded, it has the lowest probability of landing on top. On the contrary, 6 has the highest probability (6 is always on the opposite side from 1). The side faces are physically symmetric. But we started with the condition "1 is loaded," so these are conditional *probabilities. For convenience, let's read* $P(i \mid \ell)$ *as "the probability that face* i *lands on top given that face* ℓ *is loaded". For our example case, we can say* $P(1 \mid 1) = \alpha$, $P(2 \mid 1) = P(3 \mid 1) = P(4 \mid 1) = P(5 \mid 1) = \beta$, *and* $P(6 \mid 1) = \gamma$, *where* $0 \leq \alpha < \beta < \gamma$ *and* $\alpha + 4\beta + \gamma = 1$. *Corresponding statements will be true for each face that could be loaded.*

Note that the dots on a face and its opposite always sum to one, so each face has one way to be the loaded, one way to be the opposite, and four cases where it is a side face.

We use the definition of conditional probability to show that $P(\ell, i) = P(i \mid \ell)P(\ell)$. *Working out all cases gives the following table.*

Now the probability that the roll is one is $P(\{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1)\})$. *I.e., the sum of column 1. Using the fact that* $\alpha + 4\beta + \gamma = 1$ *we have that* $P(\mathit{roll} = 1) = 1/6$. *Corresponding statements hold for all faces.*

Most likely you expected this result. Now you know how to prove it.

(c) Now consider rolling the same (possibly "loaded") die twice. Think of this as two random variables with a common state space. What further

roll/load	1	2	3	4	5	6
1	$\alpha/6$	$\beta/6$	$\beta/6$	$\beta/6$	$\beta/6$	$\gamma/6$
2	$\beta/6$	$\alpha/6$	$\beta/6$	$\beta/6$	$\gamma/6$	$\beta/6$
3	$\beta/6$	$\beta/6$	$\alpha/6$	$\gamma/6$	$\beta/6$	$\beta/6$
4	$\beta/6$	$\beta/6$	$\gamma/6$	$\alpha/6$	$\beta/6$	$\beta/6$
5	$\beta/6$	$\gamma/6$	$\beta/6$	$\beta/6$	$\alpha/6$	$\beta/6$
6	$\gamma/6$	$\beta/6$	$\beta/6$	$\beta/6$	$\beta/6$	$\alpha/6$

assumption(s) do you need to make to compute the probability of each possible sequence of two rolls of one die?

We need a state space, which is the set of all possible triples of faces. The first component is the loaded face, and the second and third components correspond to the first and second rolls, respectively. There are $6 \times 6 \times 6 = 216$ *such triples. Then we define the random variables* $L(\omega) = \omega_1$, $I(\omega) = \omega_2$, *and* $J(\omega) = \omega_3$.

As before, the presence of the "load" means that we cannot assume that the probabilities of faces in rolls are independent of the loaded face. But we can assume that the rolls are independent of each other.

(d) Compute the probability of each sequence of two rolls.

An important aspect of conditional probability is that conditional probabilities are probabilities, and all of the same operations apply. So for the r.v.s defined above, independence of rolls means $P(\{\omega : I(\omega) = i, J(\omega) = j \mid L(\omega) = \ell\})o = P(\{\omega : I(\omega) = i \mid L(\omega) = \ell\}P(\{\omega : J(\omega) = j \mid L(\omega) = \ell\}$. *Let's abbreviate this equation as* $P(i, j \mid \ell) = P(i \mid \ell)P(j \mid \ell)$. *We can also "move a r.v. across the bar" because* $P(i \mid j, \ell) = P(i, j \mid \ell)/P(j|mid\ell)$. *That is, we don't need notation like* $P(i|midj \mid \ell)$ *to express multiple conditions.*

All this means that we can express probabilities in the state space as $P(\omega) = P(\omega_2, \omega_3 \mid \omega_1)P(\omega_1) = P(\omega_3 \mid \omega_1, \omega_2)P(\omega_2 \mid \omega_1)P(\omega_1) = P(\omega_3 \mid \omega_1)P(\omega_2 \mid \omega_1)P(\omega_1)$, *where the first two equalities are the definition of conditional probability, while the third comes from independence of rolls. The values for the conditional probabilities in the far right-hand side are as given in the first part. Thus we have the table:*

		Load						
R1	R2	1	2	3	4	5	6	
1	1	$\alpha^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\gamma^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$
	2	$\alpha\beta/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta^2/6$	$\beta\gamma/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	3	$\alpha\beta/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	4	$\alpha\beta/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	5	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
u	6	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$2\alpha\gamma+4\beta^2$
2	1	$\alpha\beta/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta^2/6$	$\beta\gamma/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	2	$\beta^2/6$	$\alpha^2/6$	$\beta^2/6$	$\beta^2/6$	$\gamma^2/6$	$\beta^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$
	3	$\beta^2/6$	$\alpha\beta/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta\gamma/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	4	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	5	$\beta^2/6$	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$\beta^2/6$	$2\alpha\gamma+4\beta^2$
	6	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
3	1	$\alpha\beta/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	2	$\beta^2/6$	$\alpha\beta/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta\gamma/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	3	$\beta^2/6$	$\beta^2/6$	$\alpha^2/6$	$\gamma^2/6$	$\beta^2/6$	$\beta^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$
	4	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$2\alpha\gamma+4\beta^2$
	5	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	6	$\beta\gamma/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
4	1	$\alpha\beta/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	2	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	3	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$2\alpha\gamma+4\beta^2$
	4	$\beta^2/6$	$\beta^2/6$	$\gamma^2/6$	$\alpha^2/6$	$\beta^2/6$	$\beta^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$
	5	$\beta^2/6$	$\beta\gamma/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\alpha\beta/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	6	$\beta\gamma/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
5	1	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	2	$\beta^2/6$	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$\beta^2/6$	$2\alpha\gamma+4\beta^2$
	3	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	4	$\beta^2/6$	$\beta\gamma/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\alpha\beta/6$	$\beta^2/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	5	$\beta^2/6$	$\gamma^2/6$	$\beta^2/6$	$\beta^2/6$	$\alpha^2/6$	$\beta^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$
	6	$\beta\gamma/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\beta/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$

		Load						
R1	R2	1	2	3	4	5	6	
6	1	$\alpha\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\gamma/6$	$2\alpha\gamma+4\beta^2$
	2	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	3	$\beta\gamma/6$	$\beta^2/6$	$\alpha\beta/6$	$\beta\gamma/6$	$\beta^2/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	4	$\beta\gamma/6$	$\beta^2/6$	$\beta\gamma/6$	$\alpha\beta/6$	$\beta^2/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	5	$\beta\gamma/6$	$\beta\gamma/6$	$\beta^2/6$	$\beta^2/6$	$\alpha\beta/6$	$\alpha\beta/6$	$2\alpha\beta+2\beta\gamma+2\beta^2$
	6	$\gamma^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\beta^2/6$	$\alpha^2/6$	$2\alpha^2+2\beta^2+2\gamma^2$

(e) Define the random variable X which gives the sum of two rolls carefully.

With state state space Ω defined above, we can write $X : \Omega \to R$ where $X(\omega) = \omega_2 + \omega_3$.

Note that although ω_1 doesn't enter the formula for $X(\omega)$, X still depends on it in the sense that the conditional distribution of X will change if ω_1 changes.

(f) Recall that the distribution of a random variable $X : \Omega \to R$ is the function $F : R \to [0, 1]$ defined by $F(c) = P(\omega \in \Omega | X(\omega) \leq c)$. Compute the distribution of the sum of two rolls.

Remember that the composition of a function $f : S \to T$ with a random variable $X : \Omega \to S$ is another random variable $Y : \Omega \to S$, where $Y(\omega) = f(X(\omega))$. We split the sum random variable X defined above into two parts, $X_1 : \Omega \to D^2$ and $X_2 : D^2 \to R$, where $D = \{1, \ldots, 6\}$ and thus D^2 is the set of ordered pairs of dice rolls. Then $X(\omega) = X_2(X_1(\omega))$.

We saw that state space *was an abstract concept. Being abstract, if it satisfies a few conditions (being a partition, and having enough elements to define all events of interest, including all events that can be generated by the set operations on those events), it can be whatever we want. If we want, it can be what we need—and what we need to make things easy to understand here is for D^2 to be the state space for X_2. It's a partition, and it allows us to define all events with two rolls of the die—so we can do that. The function for the random variable is easy —use the identity (the value is the argument, $f(i, j) = (i, j)$). All that is left is to define the probabilities.*

From the point of view of the random variable X_1, the event $(1, 1)$ is $\{\omega : \omega_2 = 1, \omega_3 = 1\}$, which is

$$\{(1, 1, 1), (2, 1, 1), (3, 1, 1), (4, 1, 1), (5, 1, 1), (6, 1, 1)\}$$

. The probability of this event is

$$P(1, 1, 1)+P(2, 1, 1)+P(3, 1, 1)+P(4, 1, 1)+P(5, 1, 1)+P(6, 1, 1)$$

, and we can get the values for $P(\ell, i, j)$ from the previous table. The results are in the last column, and formatted in the usual 6×6 format below:

red/blue	1	2	3	4	5	6
1	A	B	B	B	B	C
2	B	A	B	B	C	B
3	B	B	A	C	B	B
4	B	B	C	A	B	B
5	B	C	B	B	A	B
6	C	B	B	B	B	A

where $A = 2\alpha^2 + 2\beta^2 + 2\gamma^2$, $B = 2\alpha\beta + 2\beta\gamma + 2\beta^2$, and $C = 2\alpha\gamma + 4\beta^2$. We can show $A > B > C$ under some conditions. One easy case is where $\gamma - \beta = \beta - \alpha$. Let the common difference be δ. Then we have

$$2\beta^2 + 2\delta > 2\beta^2 > 2\beta^2 - 2\delta^2$$

which is a simplification of

$$(\beta - \delta)^2 + (\beta + \delta)^2 > (\beta - \delta)\beta + \beta(\beta + \delta) > (\beta - \delta)(\beta + \delta) + \ \ \beta^2$$

which may be rewritten by the definition of δ as

$$\alpha^2 + \gamma^2 > \alpha\beta + \beta\gamma > \alpha\gamma + \beta^2.$$

Multiplying by 2 and adding $2\beta^2$ gives

$$2\alpha^2 + 2\beta^2 + \gamma^2 > 2\alpha\beta + 2\beta^2 + 2\beta\gamma > 2\alpha\gamma + 4\beta^2.$$

The definitions of A, B, and C may be substituted, giving

$$A > B > C.$$

(g) Suppose you are told that the die on the right in the image above is the actual case. What do you expect the probability of each number 1, ..., 6 that can appear in a *single* roll of the die is?

Using the notation from the previous parts, $P(1) = P(3) = P(4) = P(6) = \beta$, $P(2) = \gamma$, *and* $P(5) = \alpha$.

(h) For each value $c = 1, ..., 6$, compute $P(X = c \mid E)$, where E is the event "the die is loaded on the 5 side".

Using the notation of Part 18b,

$$\begin{aligned} P(X = 1 \mid E) &= \frac{P(\{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1)\} \cap \{(5,1),(5,2),(5,3),(5,4),(}{P(\{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\})} \\ &= \frac{P(\{(5,1)\})}{P((5,1)) + P((5,2)) + P((5,3)) + P((5,4)) + P((5,5)) + P((5,6))} \\ &= \frac{\beta}{6} / \frac{1}{6} = \beta \end{aligned}$$

as assumed in that part.

19. `[Problem ID #19] distribution computation`
Consider the state space which is the set of formations used by my daughter's dance team (each dot indicates the position of a dancer): $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$, and two random variables $X : \Omega \to R$ and $Y : \Omega \to R$ defined by

$X($⚀$) = 1, \ldots, X($⚅$) = 6$ and $Y($⚀$) = 6, \ldots, Y($⚅$) = 1$. Assume the formations are used with equal probability.

Compute the cumulative distributions of X and Y themselves, and the cumulative joint distribution of X and Y together. Conclude that X and Y are not independent. Explain why not.

Hint: in expressing the distributions, you may use any convenient table format, and you may abbreviate expressions like "$X \leq b$" to just "b" in the table.

Actually it turns out to more convenient to use b to mean $b \leq X < b + 1$, except that for $b = -\infty$ it means $-\infty < X < 1$, and for $b = 6$ it means $6 \leq X < \infty$.

Here is the table. The last row and last column are the marginal distributions of X and Y respectively.

	$-\infty$	1	2	3	4	5	6	
$-\infty$	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1/6	1/6
2	0	0	0	0	0	1/6	1/3	1/3
3	0	0	0	0	1/6	1/3	1/2	1/2
4	0	0	0	1/6	1/3	1/2	2/3	2/3
5	0	0	1/6	1/3	1/2	2/3	5/6	5/6
6	0	1/6	1/3	1/2	2/3	5/6	1	1
	0	1/6	1/3	1/2	2/3	5/6	1	

$F_{xy}(1, 1) \neq F_x(1)F_y(1)$, so X and Y are not independent. You only have to check one case if it's not equal!

20. `[Problem ID #20] random variable`
Construct two r.v.s X and Y such that $\text{Cov}[X, Y] = 0$ but X and Y are not independent. Hint: a simple way involves three values for each of -1, 0, and 1, and a condition on possible combinations so that $XY = 0$ for all ω. You must define the state space Ω, and prove that $\text{Cov}[X, Y] = 0$, and the r.v.s are not independent.

Let $\Omega = \{A, B, C, D, E\}$, and define $P(\omega)$, X, and Y as in the following table.

ω	A	B	C	D	E
$P(\omega)$	1/5	1/5	1/5	1/5	1/5
$X(\omega)$	-1	0	0	0	1
$Y(\omega)$	0	-1	0	1	0
$X(\omega)Y(\omega)$	0	0	0	0	0

The covariance is 0 because the mean of each variable is 0, and all cross-products for each ω are zero. X and Y are not independent because $F_x(-1) = F_y(-1) = 1/5$, but $F_{xy}(-1,-1) = 0 \neq F_x(-1)F_y(-1)$.

21. `[Problem ID #21] statistics model`
An airplane regularly flies from Tokyo to Sapporo. It takes off at the same time every day, but it faces varying weather conditions. The airline checks its records and computes that the equation for the time to arrive is $t = 92 + \epsilon$, where ϵ is distributed approximately normally with mean 0 and standard deviation 4.5. Time units are minutes.

    Explain each model in some detail.

    (a) What is the *domain model* being used?

    The domain model *is that it takes a certain amount of time to fly the route. This time is 92 minutes. It doesn't say why a constant time is an appropriate model. The model I had in mind is that the plane needs a certain time to take off and land, and then travels using a specific amount of power to the engines over the same route every time. There are other models that would give the same equation.*

    (b) What is the *statistical model* being used?

    The statistical model is that there is a random variation in flight time. The distribution was measured empirically, and the use of a statistical model is based on the weather (presumably wind direction and speed). This last statement matters because it means that

22. `[Problem ID #22] descriptive distribution computation`
Here is some data about 10 college students:

Person	1	2	3	4	5	6	7	8	9	10
Gender	F	F	F	F	F	M	F	F	F	F
Age	23	20	21	20	19	18	20	22	18	22
Height	161.7	169.5	159.2	159.4	165.1	166.4	153.8	163.3	164.0	168.3
Weight	62.0	65.1	57.2	62.9	66.1	64.9	57.4	62.7	63.6	64.0

    (a) Compute the joint frequency distribution of Gender and Age (not cumulative).

	18	19	20	21	22	23
M	1	0	0	0	0	0
F	1	1	3	1	2	1

(b) Compute the cumulative joint distribution of Height and Weight.

Weight	Height									
	153.8	159.2	159.4	161.7	163.3	164.0	165.1	166.4	168.3	169.5
57.2	0	0	1	1	1	1	1	1	1	1
57.4	1	1	2	2	2	2	2	2	2	2
62.0	1	1	2	3	3	3	3	3	3	3
62.7	1	1	2	3	4	4	4	4	4	4
62.9	1	1	3	4	5	5	5	5	5	5
63.6	1	1	3	4	5	6	6	6	6	6
64.0	1	1	3	4	5	6	6	6	7	7
64.9	1	1	3	4	5	6	6	7	8	8
65.1	1	1	3	4	5	6	6	7	8	9
66.1	1	1	3	4	5	6	7	8	9	10

(c) Compute a histogram of heights for 5cm ranges from 150 to 170 cm. Draw it as a bar graph.

Range	150-155	155-160	160-165	165-170
Count	1	3	3	3

Graph postponed.

(d) What is the percentile rank of the man's height?

80%

(e) Compute the covariance of the Height and the Weight. Explain what other statistics you need to compute first.

Covariance = 11.93 cm kg

Mean height = 163.07 cm
Mean weight = 62.59 kg
Height variance = 20.11 cm^2
Weight variance = 8.34 kg^2
Height standard deviation = 4.48 cm
Weight standard deviation = 2.89 cm

(f) Compute the correlation coefficient of the Height and the Weight.

Correlation = 0.92

(g) The BMI of a person is that person's weight in kilograms divided by the square of their height in meters: BMI $= W/H^2$. Describe the correct way to compute the average BMI of this group. Why is it correct?

Use the BMI formula for each person, then take the average of those ten BMI results.

BMI is a nonlinear function *of height and weight, so it matters whether you compute the BMI for each person first, then take the average, or take the average height and weight first, then compute BMI using those average. Since the problem is to get the average BMI, the BMI must be computed first.*

23. `[Problem ID #23] statistics logic`
If the covariance of two random variables is zero, the correlation coefficient must be zero. Explain why.

The correlation coefficient is the covariance divided by the product of the standard deviations of the random variables. If either standard deviation is zero, the correlation coefficient cannot be computed. Otherwise, both standard deviations are positive, and dividing zero by a positive number gives zero.

24. `[Problem ID #24] outliers`
In this problem, refer to the distribution depicted below::

```
|_xxxoxoxxxoooxoooxo___...
a                   b
```

The left-to-right scale is quantitative and linear, increasing toward the right. "x" and "o" denote the two types of qualitative outcome. The underscore "_" indicates that there was no observation at that level of the quantitive variable. The bar at the extreme left labelled "a" indicates a theoretical lower limit on the quantitive variable (values lower than a are impossible). The observation labelled "b" will be referred to later. The ellipsis at the right indicates that (1) there is no theoretical upper limit, and (2) the absence of observed values continues forever to the right.

(a) Define *outlier* as used in statistics.

An outlier *is an observed value that is far away from both the "center" of the distribution, and from other values in the distribution.* ***Note:*** *This is not a formal definition. There are no good "generic" definitions of outlier; it's the kind of thing that "you know it when you see it."*

(b) Identify an outlier in the distribution above. (Choose the "worst" outlier.)

The "x" labelled "b" is an outlier. There are perhaps other possibilities, but this is the "worst."

(c) Based on "looking at the picture," give a procedure to compute the outcome to be expected given a position on the line, based on the observed distribution above. You have to predict either "x" or "o" based on the position on the line. It does not have to be based on any particular statistical concept, but do the best you can.

There are three "x"s followed by three "o"s in the middle of the figure. I would take the point dividing that group into "x"s on the left and "o"s on the right, and say anything to the left of that point is going to be "x", and anything to the right of that point is going to be "o".

There are other possibilities, such as 1-nearest neighbor. But all I can think of except 1-nearest neighbor give the same result.

(d) What does your procedure predict in the case of a new observation at the level "b"?

"o"

(e) Explain what *over-fitting* (also called "over-training") is.

Over-fitting *occurs when so many explanatory variables are used that some of them are (accidentally) highly correlated with the random effects in the model. It's called "over-fitting" because the model fits the observed data very well, but will predict the results for new data poorly.*

(f) The "1-nearest neighbor" classifier predicts that if an individual is observed at "b", it will be an "x". Explain how this is an example of "over-fitting".

it's pretty clear that the occasional "x"s on the right and "o"s on the left are random "accidents" of some kind. Predicting that the same accident will happen at "b" twice ignores the general trends in the data, and probably will produce a mistaken prediction.

(g) Explain how "over-fitting" can occur in calculating the mean. (Note: This question was not discussed in class, and is intended to challenge those who consider themselves adept at mathematics. Questions like this that are "difficult extensions" of class discussion will not be asked on the exam.)

Postponed.

25. `[Problem ID #25] statistical model`
Consider collecting data about people.

(a) Give two examples of variables about people that are *qualitative* (also called *categorical*) but not ordered or quantitive (cardinal).

(b) Give two examples of variables about people that are *ordered* (also called *ordinal*) but not quantitive (cardinal).

(c) Give two examples of variables about people that are *quantitative* (also called *cardinal*).

(d) Give the values for yourself of each variable you mentioned in parts a, b, and c.

(e) For each type of variable (qualitative, ordered, quantitative), give one example of a statistical operation that *may* be performed on that kind of data, and one example of a statistical operation that *should not* be performed on that kind of data. (Your answer may be "none" when you believe any statistical operation is valid, respectively invalid, for that kind of data.)

26. `[Problem ID #26] statistics descriptive v. inferential`
Briefly define *descriptive statistics* and *inferential statistics*.

Descriptive statistics *are calculations that summarize observed data, and express "important" properties of the data in compact form.*

Inferential statistics *use the same calculations (with slight variation), but generalize the statement to "laws" about the process generating the data, so that the configuration of future data can be predicted.*

27. `[Problem ID #27] statistics descriptive v. inferential`
The purely mathematical calculations for descriptive statistics and inferential statistics are mostly the same. How can you tell when someone is doing descriptive statistics and when it is inferential statistics?

In descriptive statistics, the name of the statistic and the variables be summarized are simply stated, and only the observed data is characterized. In inferential statistics, the characterization is extended to data that has not yet been observed.

28. `[Problem ID #28] statistics descriptive v. inferential`
There are two kinds of empirical variance, the population variance and the sample variance. The difference is that the sample variance uses the number of observations less one to correct for bias. Is the sample variance a descriptive statistic or an inferential statistic? How do you know? (This question requires some knowledge of statistics not discussed in class, although a native speaker of English can probably get the right answer without knowing about statistics.

This question will not be asked on the test.)

The sample variance is used for inferential statistics, because the notion of bias *only makes sense if you want to predict the outcome of new observations.*

29. `[Problem ID #29] outliers`
Some distributions are said to have "fat tails," and for such distributions the *median* is preferred to the *mean* as a measure of location of the distribution. Many price distributions in empirical finance are said to have fat tails.

    (a) Define *median*.

    The median *of a distribution is the value such that half of the weight of the distribution is on values below the median, and half on values above it. Also,* $F(\textit{median}) = 1/2$.

    (b) Define *mean*.

    The mean is the value such that the sum, weighted by the distribution, of (signed) deviations above and the mean is zero. It is the average, or $\textit{mean} = \sum_X xf(x)$.

    (c) Give an informal definition of *outlier*.

    An outlier *is an observed value that is far away from both the "center" of the distribution, and from other values in the distribution. Outliers are relatively rare (if they weren't, they would pull the "center" to themselves.*

    (d) Explain the practical reason for preferring the median in the case of a fat-tailed distribution. Refer to *the presence of outliers* in your explanation.

    Because outliers are "rare", it's possible that several outliers will all be on the same side of the "true center". Since they are "far away" from the center, this can move the mean a lot. But the median will only move from one central value to the next. Since the central values are clustered together, this distance is small, and more important, doesn't *depend on how "far away" the outlier is. This is useful in fat-tailed distributions since they tend to produce many outliers.*

(e) Explain how outliers affect the mean and median respectively.

An outlier "pulls the mean toward itself" in proportion to 1/(number of observations).

An outlier "pulls the median toward itself" by half the distance from the median to the next observation in the direction of the outlier.

30. `[Problem ID #30] fractiles`
    A certain student scored at the 80th percentile on an exam. Does this mean that the student got 80 points on the exam? Explain.

    No. It means that 80% of the students got the same or lower grades as this student.

31. `[Problem ID #31] fractiles`
    A certain student scored 80 points on an exam. Does this mean that the student is ranked at the 80th percentile in the class? Explain.

    No. To determine the percentile we need to count the fraction of the students who got 80 points or lower.

32. `[Problem ID #32] fractiles`
    A certain student scored 80 points on an exam, placing her at the 80th percentile in the class. Is this ranking appropriate? Explain.

    This is a trick question.

    There's no reason to answer one way or the other. The percentage score on the exam and the percentile rank have no defined relationship to each other.

33. `[Problem ID #33] fractiles`

Note: this is the same data set used in Problem 22. You should do Problem 22 first, and use the results here.

Here is some data about 10 college students:

Person	1	2	3	4	5	6	7	8	9	10
Gender	F	F	F	F	F	M	F	F	F	F
Age	23	20	21	20	19	18	20	22	18	22
Height	161.7	169.5	159.2	159.4	165.1	166.4	153.8	163.3	164.0	168.3
Weight	62.0	65.1	57.2	62.9	66.1	64.9	57.4	62.7	63.6	64.0

(a) What is the percentile rank of student #5 in Height?

70th percentile.

(b) What Weight is the 80th percentile?

64.9 kg

(c) By definition, what percentile is the median?

50th.

(d) By definition, what percentile is the mean?
This is a trick question.

The mean is not defined by a percentile.

(e) Compute the interquartile range.

***This question is broken**, because there are several variables and it doesn't say which one. Sorry! I will do the weight. This is a little tricky because the quartiles are the 25th percentile and the 75th percentile, which are not in the table. There are several possible definitions. I use the average of the next lower and next higher percentiles that are in the table.*
The 25th percentile is (57.4 + 62.0)/2 = 59.7. The 75th percentile is (64.0 + 64.9)/2 = 64.45 kg.

34. `[Problem ID #34] probability algebra independence`

The probability of the event A is a, and the probability the event $A \cap B$ is b. The events are independent. What is the probability of event B?

$P(B) = b/a$.

35. `[Problem ID #35] probability algebra independence`
The probability of the event A is a, and the probability the event $A \cap B$ is b. What is the probability of event B?

This is a trick question. (Compare the preceding problem.)

This cannot be computed without knowing whether A and B are independent.

36. `[Problem ID #36] probability algebra distribution`
The heights of children who want to ride a jet coaster at an amusement park are distributed normally with mean 150cm and standard deviation 15cm. Denote this cumulative distribution function (CDF) by F. The jet coaster has a height restriction: a rider must be at least 140cm tall, or they may not ride.

(a) Express the probability that a child who arrives at the ride is 150cm tall using the CDF F.
This is a trick question.

The normal distribution is continuous, so the probability of getting any height exactly *is zero.*

(b) Express the probability that a child who gets off the ride is 150cm tall using the CDF F.
This is a trick question.

The normal distribution is continuous, so the probability of getting any height exactly *is zero.*

(c) Express the probability that a child who arrives at the ride will be permitted to ride using the CDF F.

$1 - F(140)$.

(d) Express the probability that a child who arrives at the ride is over 160cm tall using the CDF F.

$1 - F(160)$

(e) Express the probability that a child who gets off the ride is over 160cm tall using the CDF F.

This is a conditional probability: $(1 - F(160))/(1 - F(140))$.

(f) Which is greater, the probability that a child who arrives at the ride is over 160cm, or the probability that a child who gets off the ride is over 160cm? Explain using the CDF F.

Using the formula in the previous part, we know that $1 - F(140)$ *is less than 1, so children getting off the ride are more likely to be taller than 160cm.*

(g) What kind of probability is the probability that a child who gets off the ride is over 160cm?

This is a conditional probability. The condition is that the child is allowed to ride, that is, has a height of at least 140cm.

(h) What kind of probability is the probability that a child who arrives the ride is over 160cm?

This is an unconditional probability.

37. `[Problem ID #37] probability algebra conditional`
Consider two events A and B.

(a) Define *conditional probability*.

Conditional probability is the probability of one event given the information that another has definitely occurred. The formula is $P(B \mid A) = P(A \cap B)/P(A)$.

(b) State Bayes' Rule.

$P(A \mid B) = P(B \mid A)P(A)/P(B)$.

(c) Let $P(A) = 1/2$ and $P(B \mid A) = 1/3$. Use Bayes' Rule to compute $P(A \mid B)$.
This is a trick question.

This cannot be done because $P(B)$ *is not given.*

(d) Let $P(A) = 1/2$, $P(B) = 1/6$, and $P(B \mid A) = 1/3$. Use Bayes' Rule to compute $P(A \mid B)$.

(e) Let $P(A) = 1/2$ and $P(A \mid B) = 1/2$. Can you conclude that A is independent of B?

Yes, using the definition of conditional probability. We have

$$P(A \cap B) = P(A \mid B)P(B) = P(A)P(B),$$

and independence is established. Note that although independence is a symmetric relationship between A *and* B *we can use the asymmetric formula* $P(A \mid B) = P(A)$ *to establish independence.*

(f) Suppose $P(B \mid A) = 1$. Can you conclude that as events $A = B$? Explain.

No, we can conclude that $A \subset B$*, but not that* $B \subset A$*.*

Note: *If you are familiar with measure theory, you will note that this is only true "up to" sets of measure zero. If you aren't, don't worry about it. Either way, it doesn't matter here.*